UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Unravelling mechanisms of reproductive isolation between two sister species of Iberian voles

Doutoramento em Biologia
Biologia Evolutiva

Margarida Alexandra de Sousa Carvalho Tavares Duarte

Tese orientada por:
Doutora Cristiane Bastos-Silveira
Professora Doutora Maria da Luz Mathias

Documento especialmente elaborado para a obtenção do grau de Doutor

2016

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Unravelling mechanisms of reproductive isolation between two sister species of Iberian voles

Doutoramento em Biologia
Biologia Evolutiva

Margarida Alexandra de Sousa Carvalho Tavares Duarte

Tese orientada por:
Doutora Cristiane Bastos-Silveira
Professora Doutora Maria da Luz Mathias

Júri:
Presidente:
- Professor Doutor Pedro Miguel Alfaia Barcia Ré

Vogais:
- Doutor Gerald Heckel
- Professor Doutor Carlos Manuel Martins Santos Fonseca
- Professor Doutor António Paulo Pereira de Mira
- Doutora Susana Sá Couto Quelhas Lima Mainen
- Professora Doutora Maria da Luz da Costa Pereira Mathias
- Professor Doutor Octávio Fernando de Sousa Salgueiro Godinho Paulo

Documento especialmente elaborado para a obtenção do grau de Doutor

Fundação para a Ciência e Tecnologia (SFRH/BD/70646/2010; PTDC/BIA-BEC/103729/2008; UID/AMB/50017); FCT/MEC; FEDER

2016

The Ph.D. candidate was awarded with studentship SFRH/BD/70646/2010, financed by Fundação para a Ciência e Tecnologia (FCT).

The present Ph.D. project was developed under PTDC/BIA-BEC/103729/2008 "Understanding the origin and persistence of *Microtus* species in the Iberian Peninsula: the genetical, behavior and ecological forces" (FCT), coordinated by Cristiane Bastos-Silveira. It was also supported by Centro de Estudos de Ambiente e Mar through National (UID/AMB/50017 and FCT/MEC) and European (FEDER, within the PT2020 Partnership Agreement and Compete 2020) funds.

The present Ph.D. dissertation should be cited as:

Duarte MA (2016) Unravelling mechanisms of reproductive isolation between two sister species of Iberian voles. Dissertation, Faculdade de Ciências, Universidade de Lisboa, Lisboa.

“Success is not final, failure is not fatal:
it is the courage to continue that counts.”

Winston Churchill

Preliminary notes

According to the Post-graduate Studies Regulation (Diário da República, 2.ª série, N.º 57, published on the 23th of March, 2015), the present dissertation includes papers published in ISI-indexed scientific journals and manuscripts in preparation for submission.

These papers are included in **Chapters 2**, **3**, **4** and **5**, and were formatted according to the dissertation style, except for specific requirements made by the scientific journals, such as references and measurement units.

The Ph.D. candidate is presented as the first author since it was responsible for the scientific design, data collection, analysis of the results and writing of the papers, under the supervision of Cristiane Bastos-Silveira (primary supervisor) and Professor Maria da Luz Mathias (secondary supervisor).

Lisbon, 25th of July, 2016

Margarida Alexandra de Sousa Carvalho Tavares Duarte

Table of contents

Acknowledgments

Firstly, I would like to express my sincere gratitude to Cristiane Bastos-Silveira, for her unconditional support and ability to challenge me throughout this Ph.D. project. You were always more than a supervisor and our friendship will always be treasured. Secondly, to Professor Maria da Luz Mathias for all the insightful comments and counselling provided these past years.

I will also like to thank all my fellow colleagues, particularly Patrícia Sardinha, my partner in crime at the Museu de História Natural e da Ciência (MUHNAC), for all the support during the good, the bad and the ugly; Joaquim Tapisso and Rita Monarca for all the assistance with behavioural testing and statistical analyses; and to Isabel Queirós Neves, Ana Sofia Quina and Sofia Gabriel for the encouragement during these past years.

For their support and interest I have to thank Catarina Mateus, Judite Alves, Alexandra Cartaxana and Luís Ceríaco from MUHNAC; and Isa Matos, Silke Waap, Mónica Silva and D. Branca from the Faculdade de Ciências da Universidade de Lisboa.

A special thanks is due to Gerald Heckel (Institute of Ecology and Evolution, University of Bern) for all the theoretical discussions throughout the years and for enabling me to enrich my molecular biology data with his precious *Microtus* samples. I would also like to thank Heidi Lischer for all the support with the repeated random haplotype sampling analysis and the sweet Susanne Tellenbach for technical assistance.

At the Center for Neuroscience and Cell Biology (University of Coimbra) I would like to thank Bruno Manadas for his availability and assistance with proteomic analyses.

I also acknowledge the support of my FCULian friends Inês Silva, Ana Filipa Parreira, Ana Mónica Gabriel, Márcia Pinheiro, Fernanda Baptista, Marina Ribeiro and Frederico Rodrigues; and the Doroteias “gang”, particularly the wanderlust Susana Gasalho.

For proofreading the present dissertation I would like to thank John Lee from the Champalimaud Foundation. Thanks are also due to Micael Fonseca for custom drawing all illustrations present in each chapter's cover page.

My sincere thanks also go to my current Champalimaud Research mates from Lima, Mainen and Renart Labs, and Vivarium and Glasswash and Media Preparation Platforms, particularly Silvana "Banas" Araújo, Susana "Sue" Valente, Diogo Matias, Luís Moreira, Raquel Tomás, Ana Margarida Nunes, João Pereira and Patrick Teca. Without their support, interest and insights the task of finishing my Ph.D. dissertation would have been much more difficult. Moreover, I would also like to thank my "handler" Susana Lima and "boss" Zachary Mainen, for giving me a warm welcome to this great institute and for supporting this last stretch of my Ph.D. marathon. It is a pleasure to wake up every day knowing that I am a part of the multidisciplinary Champalimaud Research family.

For supporting my academic and personal life unconditionally I would like to thank my parents. Without them nothing would have been possible. Also, a tight and long hug to the forth member of the family, my *petit chat*, the most beautiful cat that I will ever seen and that left me with eighteen years of treasured memories.

Lastly, the biggest of acknowledgments to Pedro for supporting and encouraging me throughout the Ph.D. journey and in my most recent endeavours. Having you by my side has made all the difference.

Resumo

Este projeto de Doutoramento constituiu um primeiro passo na clarificação de mecanismos de isolamento reprodutor entre duas espécies de roedores que divergiram muito recentemente, apenas há 60,000 anos: o rato-cego *Microtus lusitanicus* e o rato-cego Mediterrânico *Microtus duodecimcostatus*. Este episódio evolutivo constitui um dos eventos de especiação mais recentes em espécies do género *Microtus*.

O isolamento reprodutor é essencial ao processo de especiação entre populações divergentes e à manutenção de unidades taxonómicas distintas, como é o caso de *M. lusitanicus* e *M. duodecimcostatus*. Dois tipos de isolamento, pré- e pós-copulatório, poderão prevenir a hibridação de duas espécies. Enquanto que as barreiras pré-copulatórias previnem o comportamento reprodutor heterospecífico e promovem a cópula conspecífica, i.e. entre indivíduos da mesma espécie; as barreiras pós-copulatórias afectam a fertilização do oócito por um espermatozóide de uma espécie diferente, a viabilidade do híbrido e possível esterilidade do mesmo.

Estas espécies irmãs Ibéricas apresentam uma área de distribuição em alopatria, com *M. lusitanicus* mais a Norte e *M. duodecimcostatus* mais a Sul da Península; e em simpatria, onde ambas as espécie ocorrem, localizada no centro da Península Ibérica, cobrindo parte de Portugal e Espanha.

Dados de citocromo *b* e microssatélites, obtidos no decorrer do projeto PTDC/BIA-BEC/103729/2008, revelaram discordância citonuclear numa grande área de simpatria em Portugal, indicando uma introgressão histórica de DNA mitocondrial de *M. duodecimcostatus* para *M. lusitanicus*.

Um isolamento reprodutor incompleto entre estas espécies irmãs na natureza é sugerido pela existência de apenas dois possíveis híbridos numa amostragem de aproximadamente trezentos indivíduos. Esta observação é complementada por dados de escolha de parceiro, através da urina, que revelaram uma preferência por odores conspecíficos a heterospecíficos, indicando a presença de isolamento

comportamental pré-cópula. Barreiras gaméticas foram igualmente sugeridas entre *M. lusitanicus* e *M. duodecimcostatus* devido a um menor sucesso da reprodução heterospecífica versus conspecífica, em condições laboratoriais. Esta observação revela que a fertilização entre ambas as espécies poderá nem sempre ocorrer após a cópula, provavelmente devido a incompatibilidades no reconhecimento oócito-espermatozóide.

Consequentemente, este projeto de Doutoramento focou-se em mecanismos de isolamento reprodutor entre *M. lusitanicus* e *M. duodecimcostatus*, nomeadamente em barreiras comportamentais pré-copulatórias e em barreiras gaméticas pós-copulatórias. Cinco objetivos específicos foram considerados: 1) identificar genes candidatos relacionados com a comunicação através do odor; 2) analisar a expressão de proteínas na urina de ambas as espécies; 3) inferir se ambas as espécies favorecem a cópula conspecífica em oposição à heterospecífica; 4) determinar se ambas as espécies apresentam uma ligação do casal reprodutor, indicativa de um sistema monogâmico social; 5) investigar o papel da proteína de reconhecimento do espermatozóide, zona pellucida 3 como barreira de isolamento gamético.

Os resultados obtidos permitiram testar as seguintes hipóteses: 1) a comunicação através do odor é uma barreira reprodutora comportamental ativa entre *M. lusitanicus* e *M. duodecimcostatus*; 2) *M. lusitanicus* e *M. duodecimcostatus* preferem copular com indivíduos conspecíficos a heterospecíficos, na presença de potenciais parceiros de ambas as espécies; 3) *M. lusitanicus* e *M. duodecimcostatus* são espécies monogâmicas sociais; e 4) a região putativa de ligação ao espermatozóide do ZP3 é uma barreira de isolamento reprodutor gamética, que afecta o acasalamento heterospecífico de *M. lusitanicus* e *M. duodecimcostatus*.

Relativamente à análise de mecanismos relacionados com a comunicação através do odor, dois tipos de genes candidatos foram analisados: 1) receptores olfactivos, Olfr31 e Olfr57, ao nível das proteínas receptoras de sinal; e 2) MHCI e MHCII ao nível das proteínas emissoras de sinal. Foram ainda examinadas urinas de *M. lusitanicus* e *M. duodecimcostatus* de forma a inferir-se se existem, ou não,

proteínas urinárias específicas de espécie que possam estar a contribuir como barreiras comportamentais na escolha de parceiro. Tendo em conta a hipótese colocada, determinou-se que os receptores olfactivos Olfr31 e Olfr57 provavelmente não estão relacionados com o isolamento reprodutor de ambas as espécies, visto haver baixa variabilidade genética e ausência de seleção positiva em diversos aminoácidos localizados na zona de reconhecimento de partículas odoríferas. Devido a constrangimentos metodológicos, não foi possível aferir se MHCI e MHCII apresentam um papel relevante no isolamento reprodutor entre *M. lusitanicus* e *M. duodecimcostatus*. Adicionalmente, dados de proteómica utilizando a urina de ambas as espécies questionaram o papel das MUPs (major urinary proteins), em particular o MUP20 (Darcin), como barreiras comportamentais. Este poderá ser clarificado através da análise futura de uma maior amostragem de urinas de ambas as espécies e sexos.

Considerando a hipótese de que *M. lusitanicus* e *M. duodecimcostatus* preferem reproduzir-se conspecificamente a heterospecificamente, na presença de potenciais parceiros de ambas as espécies, foram simulados dois ambientes de sintopia artificiais com uma macho e fêmea de cada *taxon*. Num deles houve uma clara dominância e agressividade de *M. duodecimcostatus* para com *M. lusitanicus*, levando à morte do macho *M. lusitanicus* e ao cancelamento desse ambiente artificial. Duas ninhadas conspecíficas de *M. duodecimcostatus* nasceram durante este ambiente. Por oposição, na outra simulação de sintopia foi observada tanto cópula heterospecífica entre a fêmea *M. duodecimcostatus* e o macho *M. lusitanicus*, como conspecífica entre ambos *M. lusitanicus*. Enquanto ambas as fêmeas e o macho *M. lusitanicus* socializavam diariamente e partilhavam o ninho, o macho *M. duodecimcostatus* manteve-se sempre associal e isolado dos restantes animais, havendo criado um ninho próprio. Duas ninhadas foram geradas durante este ensaio. A genotipagem de todos os filhotes revelou que uma das ninhadas era conspecífica de *M. lusitanicus* e a outra heterospecífica resultante do cruzamento entre a fêmea *M. duodecimcostatus* e o macho *M. lusitanicus*. Estes resultados confirmaram parcialmente a hipótese de que *M. lusitanicus* e *M. duodecimcostatus*

preferem reproduzir-se conspecificamente, visto terem sido geradas três ninhadas conspecíficas e apenas uma heterospecífica. Creio que neste caso particular a hibridação entre estas espécies irmãs foi possível devido ao papel da variabilidade comportamental individual na escolha de parceiro. Esta possível barreira comportamental entre *M. lusitanicus* e *M. duodecimcostatus* em sintopia deverá ser considerada em estudos futuros.

A hipótese seguinte considera que *M. lusitanicus* e *M. duodecimcostatus* são espécies monogâmicas sociais. O sistema de acasalamento foi inferido através de ensaios de comportamento recorrendo a um olfactómetro, de forma a testar a ligação entre membros de um mesmo casal reprodutor estável através de uma escolha entre o odor do parceiro ou de um estranho, *naïve* ou experiente ao nível sexual. Em todos os cenários testados, à exceção de um, verificou-se uma preferência pelo odor do parceiro. A exceção foi observada quando os machos tiveram de escolher entre a parceira e uma fêmea *naïve*. Os resultados obtidos confirmaram a presença de uma ligação entre os membros do casal, característica de um sistema monogâmico social, com possibilidade de cópula extra-casal por parte do macho. Essa possibilidade poderá aumentar o sucesso reprodutor dos machos numa situação de aumento de recursos naturais. Assim sendo, coloco a hipótese de que na natureza, em sintopia, a monogamia social poderá atuar como barreia comportamental indireta entre *M. lusitanicus* e *M. duodecimcostatus*.

A barreia gamética foi inferida através da análise evolutiva da região putativa de ligação ao espermatozóide da glicoproteína do oócito zona pellucida 3, baseada em várias subfamílias de roedores da família Cricetidae. Este estudo refutou o papel desta região, localizada no exão 7, como barreia gamética entre várias espécies de cricetídeos, incluindo entre *M. lusitanicus* e *M. duodecimcostatus*. Era expectável encontrar uma grande variação entre os aminoácidos das duas espécies, existindo uma sequência específica de espécie, de forma a impedir fertilizações heterospecíficas. No entanto encontraram-se sequências partilhadas entre diferentes espécies, incluindo *M. lusitanicus* e *M. duodecimcostatus*, e diferentes deleções de aminoácidos na região putativa de ligação ao

espermatozóide e/ou numa zona adjacente, que poderão afectar a estabilidade da ligação entre oócito-espermatozóide e consequentemente a especificidade da fertilização. Assim sendo, estes resultados refutam a postulação que esta região da zona pellucida 3 é uma barreia de isolamento gamético.

Concluindo, os resultados deste projeto de Doutoramento sugerem que o isolamento reprodutor entre estas espécies irmãs está associado a barreiras múltiplas, e não a apenas uma, e que ainda está incompleto, permitindo a ocorrência de hibridações esporádicas na natureza. Os resultados sugerem ainda um possível papel de proteínas urinárias na discriminação ao nível de espécie através do odor; confirmam a existência de um sistema monogâmico social para ambas as espécies, podendo constituir uma barreia de isolamento comportamental indireto entre *M. lusitanicus* e *M. duodecimcostatus*; revelam que a variabilidade comportamental individual poderá desempenhar um papel significativo no isolamento reprodutor entre *M. lusitanicus* e *M. duodecimcostatus*; e refutam a região putativa de ligação ao espermatozóide da zona pellucida 3 como barreia gamética.

Palavras-chave: *Microtus lusitanicus*; *Microtus duodecimcostatus*; isolamento reprodutor; especiação.

Abstract

The present Ph.D. project constituted a first step in understanding mechanisms of reproductive isolation between two recently diverged sister species: the Lusitanian pine vole *Microtus lusitanicus* and the Mediterranean pine vole *Microtus duodecimcostatus*.

Reproductive isolation is essential to speciation, and two types of isolation, pre- and post-mating, may prevent hybridization between two species. While pre-mating barriers prevent copulation and promote conspecific reproduction, post-mating barriers affect the success of heterospecific fertilization and hybrid viability, and potentiate its sterility.

M. lusitanicus diverged from *M. duodecimcostatus* approximately 60,000 years ago, constituting one of the most recent speciation events among *Microtus sp.* voles. While *M. lusitanicus* inhabits the Northern region of the Iberian Peninsula, reaching the French Pyrenees, *M. duodecimcostatus* occupies Southern Iberia and part of the South of France. There is also a sympatry area of distribution, where both species occur, located in the centre of the Iberian Peninsula, covering parts of Portugal and Spain.

Analyses on cytochrome *b* and microsatellites have uncovered a cytonuclear discordance over a large geographic area in Portugal, indicating a historical introgression of mitochondrial DNA from *M. duodecimcostatus* to *M. lusitanicus*. An incomplete reproductive isolation in nature is also suggested between both voles since two possible hybrids were detected in a sample size of nearly three hundred individuals. Moreover, behavioural isolation was hinted at, since there is a preference for conspecific over heterospecific odour cues. The gametic isolation barrier was proposed since heterospecific mating, in laboratory conditions, is less reproductively prolific than conspecific mating. This result suggests that fertilization between *M. lusitanicus* and *M. duodecimcostatus* may not always occur after copulation, probably due to incompatibilities in the sperm-oocyte heterospecific recognition.

Considering these previous findings, the present Ph.D. project focused on *M. lusitanicus* and *M. duodecimcostatus* reproductive isolation, particularly on pre-mating behavioural and post-mating gametic isolation barriers. It comprises five specific aims: 1) identify candidate genes related to odour cues communication; 2) analyse the expression of urinary proteins in both species; 3) infer if both species favour conspecific to heterospecific mating; 4) determine if both species present a pair bond, indicative of a monogamous mating system; and 5) evaluate the role of the sperm-binding protein zona pellucida 3, as a gametic isolation barrier.

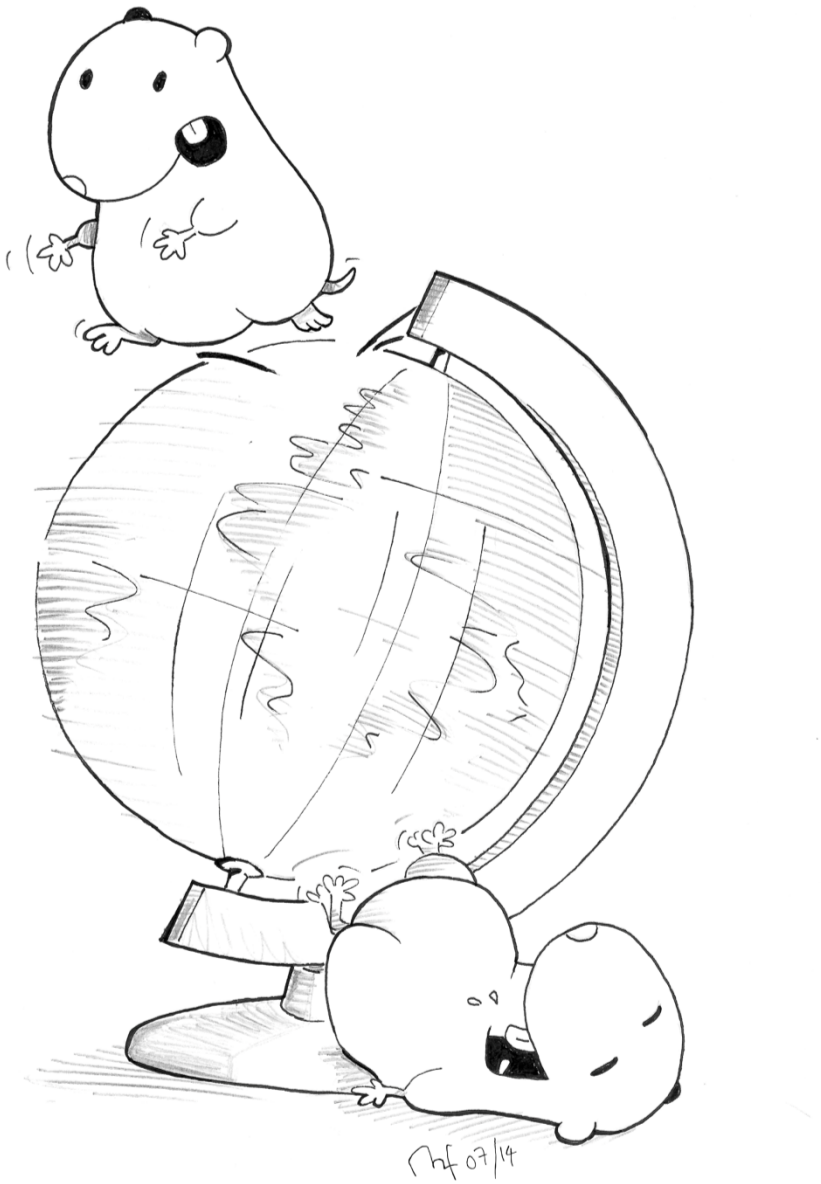
Four hypotheses were tested: 1) odour cues communication is an active behavioural reproductive barrier between *M. lusitanicus* and *M. duodecimcostatus*; 2) *M. lusitanicus* and *M. duodecimcostatus* prefer conspecific to heterospecific mating in the presence of potential mates of both species; 3) *M. lusitanicus* and *M. duodecimcostatus* are socially monogamous; and 4) the putative sperm-binding region of zona pellucida 3 is a gametic isolation barrier that impairs heterospecific mating between *M. lusitanicus* and *M. duodecimcostatus*.

The results of the present Ph.D. project suggest that reproductive isolation between these sister species relies on multiple barriers and is still incomplete, enabling sporadic hybridization in nature. Overall, results also indicate that urinary proteins may play a role in species-specific discrimination; confirm social monogamy as the mating system of both voles, being a possible indirect behavioural isolation barrier at syntopy; reveal that individual behavioural variability may contribute to the behavioural isolation between *M. lusitanicus* and *M. duodecimcostatus*; and refute the putative sperm-binding region of ZP3 as a gametic barrier.

Keywords: *Microtus lusitanicus*; *Microtus duodecimcostatus*; reproductive isolation; speciation.

Chapter 1

General Introduction



1.1 The genus *Microtus*

Voles of the speciose genus *Microtus* Schrank, 1798 are small herbivores that inhabit the Northern Hemisphere, mostly open grasslands, but also forests and highland habitats (**Figure 1**) (Getz, 1985; Hoffmann & Koeppl, 1985; Mitchell-Jones et al., 1999; Nowak, 1999).

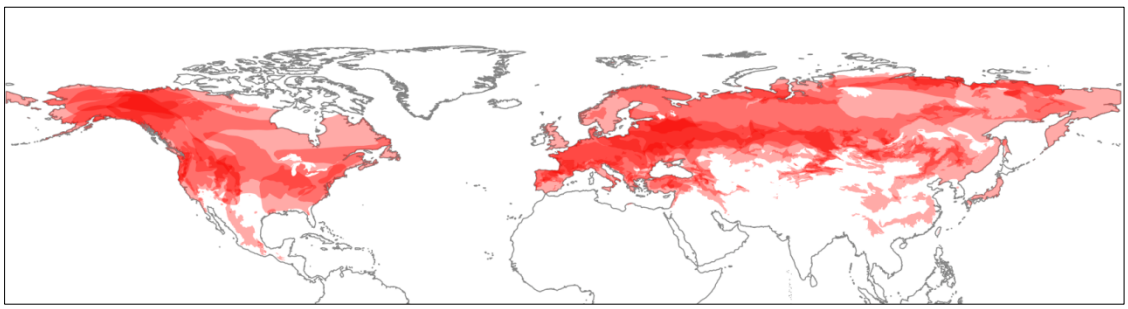


Figure 1 – Species richness distribution of *Microtus* genus extant taxa, across the Holarctic. Darker areas correspond to a higher richness (plotted in Quantum GIS 1.8.0, using digital distribution maps of the IUCN Red List of Threatened Species, 2016).

The genus *Microtus* holds nearly half of the Arvicolinae species (e.g. Musser & Carleton, 1993) and are an example of a recent and rapid radiation, which occurred 1.2-2 million years ago (Mya), resulting in 65 extant species (Musser & Carleton, 1993; Chaline et al., 1999; Nowak, 1999). The only mammalian genus with similar diversity across the Holarctic is *Sorex* Linnaeus, 1758 (Soricidae, Insectivora), which started to differentiate approximately 11.5Mya (Fumagalli et al., 1999), a long time before *Microtus*.

The genus *Allophaiomys* Kormos, 1933, a descendant of *Mimomys* Forsyth-Major, 1902, seems to be the ancestor of *Microtus* voles (Chaline & Graf, 1988). It appeared in Southern Asia during the Late Pliocene and in Europe at the beginning of the Pleistocene (Chaline & Graf, 1988). The original Asian stock diverged into many lineages, some of which migrated to North America, through the Beringian land bridge, during the last glaciation (Chaline & Graf, 1988). European and North American species appear to have diverged directly from *Allophaiomys* (Chaline & Graf, 1988; Repenning, 1992; Chaline et al., 1999) or

indirectly (Repenning, 1992), through a morphological intermediate, similar to *Lasiopodomys* Lataste, 1887. On the other hand, in Southern Asia, voles may have diverged directly from Pliocene *Mimomys* (Chaline & Graf, 1988; Garapich & Nadachowski, 1996; Conroy & Cook, 1999).

Many independent colonization events have originated the current *Microtus* Holarctic distribution (Fink et al., 2010). Ancestors of extant *Microtus* species colonized the European and North American continents repeatedly, in several independent events, on similar colonization routes during their radiation (Fink et al., 2010); instead of only three independent colonization events, one per continent, as previously suggested (Brunet-Lecomte & Chaline, 1991; Chaline et al., 1999).

The genus *Microtus* is characterized by inconsistent systematics. Taxonomic classifications are particularly difficult due to its rapid radiation and frequent gradual variation in morphological and molecular traits between extant taxa (Mitchell-Jones et al., 1999). Some Palearctic and Nearctic fossil specimens are dated from Late Pliocene (Van der Muelen, 1978; Repenning, 1992; McKenna & Bell, 1997; Chaline et al., 1999), however paleontological information is missing for most extant species or appear relatively late (Tamarin, 1985). The oldest fossil records are dated to the Middle Pleistocene, about 0.7-0.5Mya (Rabeder, 1986; Richmond, 1996; Chaline et al., 1999), suggesting that some taxa may have speciated due to the last glaciation (e.g. Chaline & Graf, 1988; Brunet-Lecomte & Chaline, 1990). Although these fossil records are some of the most detailed, considering extant rodent genera (Gromov & Polyakov, 1977; Rabeder, 1981; Rekovets & Nadachowski, 1995), they are still incomplete in order to provide a reliable evolutionary history of this speciose genus.

This genus classification has been mostly based on paleontological and morphological characteristics, particularly through dental criteria, which allow discrimination of most taxa, both extant and already extinct, with the exception of cryptic species (Chaline, 1987). Biochemical and chromosomal data have also helped to enlighten some evolutionary and taxonomic issues (Chaline, 1987). Only

more recently has molecular data been applied in order to infer *Microtus* genus phylogeny and evolutionary history, using both mitochondrial (Conroy & Cook, 1999; Conroy & Cook, 2000; Galewski et al., 2006; Robovsky et al., 2008; Jaarola et al., 2004; Fink et al., 2010) and nuclear data (Galewski et al., 2006; Fink et al., 2006; Fink et al., 2007; Robovsky et al., 2008; Acosta et al., 2010a, b; Fink et al., 2010). Molecular analysis of the speciose *Microtus* radiation demonstrated the importance of geographic isolation, with subradiations in Europe, Asia and North America, and secondary colonizations (Fink et al., 2010). Moreover, phylogeographical studies also uncovered relatively deep divergence between parapatric evolutionary lineages within recognized taxa (Jaarola & Searle, 2002; Brunhoff et al., 2003; Fink et al., 2004; Heckel et al., 2005). These results may indicate that the taxonomic status of some of these current lineages could change to new species in the future (e.g. *Microtus agrestis* (Linnaeus, 1761): Hellborg et al., 2005; Beysard et al., 2012; Paupério et al., 2012; *M. arvalis* (Pallas, 1778): Heckel et al., 2005; Braaker & Heckel, 2009). Hence, in the recent speciose *Microtus* genus, there is evidence of ongoing speciation.

1.1.1 *Microtus* voles of the Iberian Peninsula

During the last glaciation, most of northern and central Europe was inhospitable to temperate species (Dawson, 1992). Nevertheless, there were regions in the Mediterranean peninsulas that presented a temperate climate and vegetation (Huntley, 1988; Bennett et al., 1991); hence, some species, such as *Microtus* ancestors, migrated to these Mediterranean refugia and speciated.

One of these refugia, and a hotspot of endemism, is the Iberian Peninsula. It is comprised of Portugal, Spain and Andorra, and is separated from the rest of Europe by the Pyrenees. The *Microtus* genus is represented in this peninsula by six taxa: *M. agrestis*, *M. arvalis*, *M. cabrerae* Thomas, 1906, *M. duodecimcostatus*de Selys-Longchamps, 1839, *M. gerbei* (Gerbe, 1879) and *M. lusitanicus* (Gerbe, 1879) (IUCN, 2016). From these, only *M. cabrerae* and *M. lusitanicus* are endemic. Two sister species are also present in this refugium, being the closest relatives from their

phylogenetic clade: *M. duodecimcostatus* and *M. lusitanicus* (Jaarola et al., 2004; Bastos-Silveira et al., 2012).

1.2 Iberian sister species

The Lusitanian pine vole *Microtus lusitanicus* and the Mediterranean pine vole *Microtus duodecimcostatus* are sister species from the *Terricola* subgenus, sharing a common ancestor and a very close evolutionary relationship. These small arvicolids are classified as separate taxa based on morphological, ecological and cytogenetic differences (e.g. Cabrera, 1914; Ellerman & Morrison-Scott, 1951; Spitz, 1978; Madureira, 1981; Mathias, 1996; Cotilla & Palomo, 2007; Mira & Mathias, 2007; Santos et al., 2009a, Santos et al., 2009b; Santos et al., 2010; Gornung et al., 2011; Santos et al., 2011).

Persistence of a rhombus in M3 teeth suggested that their ancestor lineage was one of the first to split from *Allophaiomys*, approximately 1.2-1.6Mya (Chaline, 1974; Chaline & Mein, 1979); however, biochemical data do not support this assumption (Chaline & Graf, 1988). Presently, it is considered that *M. duodecimcostatus* probably derived indirectly from an Iberian *Allophaiomys* taxon, probably *Allophaiomys chalinei* Alcalde, Agustí & Villalta, 1981, while *M. lusitanicus* diverged from *M. duodecimcostatus*, approximately 60,000 years ago (Chaline, 1966, 1972; Brunet-Lecomte et al., 1987, Brunet-Lecomte & Chaline, 1991). The intermediate ancestral taxon, from which *M. duodecimcostatus* could have differentiated, was probably *Microtus brecciensis* (Giebel, 1847) (Chaline, 1987).

Geographically, *M. lusitanicus* occupies Northern Iberia, reaching the French Pyrenees (Mira & Mathias, 2007), while *M. duodecimcostatus* inhabits Southern Iberia and part of the South of France (Cotilla & Palomo, 2007) (**Figure 2**). These voles present a sympatry area of distribution, where both species occur, located in the middle of the Iberian Peninsula, covering parts of Portugal and Spain, reaching the Pyrenees (Madureira, 1984; Mitchell-Jones et al., 1999; Cotilla & Palomo, 2007; Mira & Mathias, 2007) (**Figure 2**).

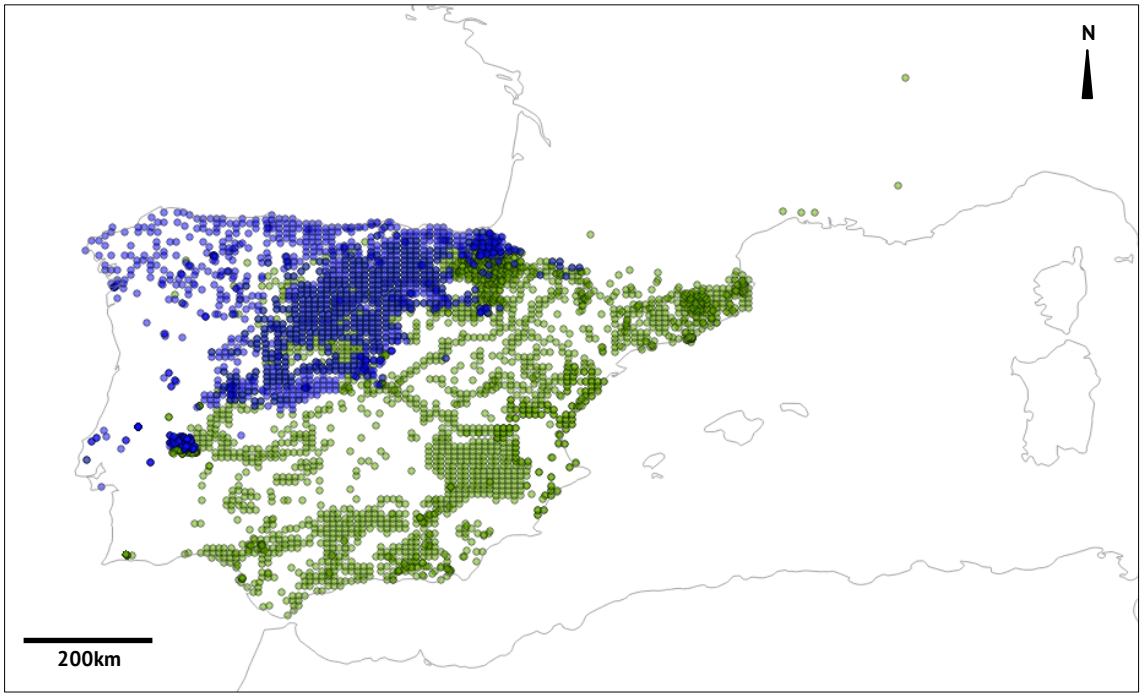


Figure 2– Capture locations of *M. lusitanicus* (blue) and *M. duodecimcostatus* (green) based on geographical coordinated available on the Global Biodiversity Information Facility and Bastos-Silveira and colleagues (2012) (plotted in Quantum GIS 1.8.0).

M. lusitanicus allopatric populations present higher M1 teeth morphological variability than *M. duodecimcostatus*, suggesting that *M. lusitanicus* has possibly occupied a broader range of distribution in the past and that ecological competition may be occurring in the sympatry areas, between both species (Brunet-Lecomte et al., 1987).

M. lusitanicus and *M. duodecimcostatus* present the same karyotype 2n=62, differing in sex chromosome morphology, constitutive heterochromatin, rDNA sites and satDNA patterns (Gornung, 2011).

A mitochondrial cytochrome *b* (Cyt*b*) phylogeny on the *Microtus* genus revealed 4-5% genetic divergence between allopatric *M. lusitanicus* and *M. duodecimcostatus* individuals (Jaarola et al., 2004). This is the lowest genetic divergence found among *Microtus* taxonomically recognized taxa (Jaarola et al., 2004). More recently, Cyt*b* and microsatellites analyses have discovered a cytonuclear discordance over a large geographic area in Portugal, suggesting a historical introgression of mitochondrial DNA from *M. duodecimcostatus* to *M. lusitanicus*

(Bastos-Silveira et al., 2012). This study also disclosed a relatively advanced speciation process, based on two clear microsatellites genetic clusters, composed of allopatric and sympatric individuals, corresponding to each species morphologically identified individuals.

Concerning nuclear molecular markers, it has been suggested that the *p53* gene is involved in the divergence between both sister voles (Quina et al., 2015), possibly due to its association with ecological stress such as hypoxia, and consequently to a fossorial life-style.

1.2.1 The Lusitanian pine vole

Natural populations of *M. lusitanicus* are organized in small family groups, which occupy complex burrow systems, excavated using the feet and incisor teeth (Mira & Mathias, 2007). Underground galleries consist of superficial (≈15cm) and deeper tunnels (<40cm), with cameras for the nest or storing food (Mira & Mathias, 2007).

M. lusitanicus reaches 77.5-105mm and 14-19g (Mira & Mathias, 2007). The body shape of this vole reveals a semi-fossorial life-style. Its cylindrical body is covered by a dark grey to sepia pelage in the back and exhibiting a grey belly (**Figure 3**), where it presents two pairs of inguinal nipples.

Figure 3– Photograph of an adult *M. lusitanicus* from the animal facility colony (see **Chapter 4**) © Duarte, M.A.

The large head terminates with a blunt snout and a small mouth, with slightly projecting upper incisors. Coherently to its subterranean living, *M. lusitanicus* has small eyes, ears (6.5-10mm) and feet (13-16mm) (**Figure 3**).

Females reach sexual maturity at 35 days of age, while the sexual maturation of males is only reached at 50 days (Mira & Mathias, 2007). In nature, the number of embryos per litter varies from one to five (Madureira, 1984). Pups are born naked and blind, weighing about 1.5g and measuring 15mm (Mira & Mathias, 2007). Hair begins to appear within three days (**Figure 4**) and after two weeks they look like miniature adults. Captive births occur every 28 days and a post-partum oestrus and gestation lasts 22-24 days (Mira & Mathias, 2007).

Figure 4– Photograph of three *M. lusitanicus* pups, one week old, from the animal facility colony (see **Chapter 4**) © Duarte, M.A.

The Lusitanian pine vole occupies diverse habitats, ranging from meadows, pastures, riversides and woods to agricultural areas, such as apple orchards and carrot crops (Mathias, 1999; Mira & Mathias, 2007; Santos, 2009). Its diet varies throughout the year. In the winter and spring *M. lusitanicus* eats mostly leaves and stems, while during the summer and autumn it consumes mainly subterranean parts of herbaceous plants, showing a preference for geophytes (Mathias, 1999; Mira & Mathias, 2007).

Studies on population dynamics in natural habitats are inexistent; however, in fruit orchards, common densities range from 100-200 individuals per hectare, exceeding 300 individuals per hectare in extremely favourable conditions (Mira & Mathias, 2007). Occasionally, *M. lusitanicus* is considered a pest when it reaches

high densities, leading to a 10-15% loss in fruit orchards (Bäumler et al., 1984; Mira & Mathias, 2007). This situation is enabled by sprinkler irrigation, a very common practice that promotes the growth of weeds near the tree trunk, leading to an increase of moisture and soil disaggregation, an optimal condition for this vole.

M. lusitanicus is a usual prey of the barn owl *Tyto alba* (Scopoli, 1769), tawny owl *Strix aluco* Linnaeus, 1758 and some small/medium sized terrestrial carnivores (Mira & Mathias, 2007).

1.2.2 The Mediterranean pine vole

M. duodecimcostatus is bigger and more robust than *M. lusitanicus*, reaching 80-110mm and weighing 19-32g (Cotilla & Palomo, 2007). Coherently to its subterranean living, and similarly to *M. lusitanicus*, it presents small eyes, ears (7.5-10mm) and feet (14.5-18.5mm) (Cotilla & Palomo, 2007). Its pelage is yellowish brown tone, with a characteristic ochre edge separating the back from the belly, where it exhibits two pairs of inguinal nipples, like its sister species (**Figure 5**).

Figure 5 – Photograph of an adult *M. duodecimcostatus* from the animal facility colony (see **Chapter 4**) © Duarte, M.A.

Still, the light and dark shades of its pelage vary by area of distribution (Cotilla & Palomo, 2007). The 23-35mm tail is grey, unlike *M. lusitanicus*' which is always bicoloured (Cotilla & Palomo, 2007). *M. duodecimcostatus* also features a strong neck musculature, prominent upper incisors, developed premaxilla and diastema, revealing that it is perfectly adapted to a semi-fossorial life-style (Madureira, 1982; Mathias, 1990).

This vole becomes sexually mature at 60-70 days of age (Mira, 1999). The breeding season is variable and gestation lasts 24 days (Cotilla & Palomo, 2007). Pups are born naked and blind, weighing 2-3g (Cotilla & Palomo, 2007). Two-week old pups present the appearance of an adult, similarly to *M. lusitanicus*. In nature, the number of embryos per litter ranges from one to five (Cotilla & Palomo, 2007) (**Figure 6**).

Figure 6– Photograph of two *M. duodecimcostatus* pups, three days old, from the animal facility colony (see **Chapter 4**) © Duarte, M.A.

In terms of behaviour, *M. duodecimcostatus* is characterised by being more aggressive and less social than *M. gerbei* (Gerbe, 1879). *M. duodecimcostatus* also uses substrate-borne signals more commonly than acoustic repertoires, conversely to less aggressive *M. gerbei* (Giannoni et al., 1997).

This Iberian vole occupies both natural and agricultural areas of Mediterranean influence, being conditioned by the existence of stable, moist, herbaceous and easy to dig soils (Mira & Mathias, 1994; Paradis, 1995; Mira, 1999; Cotilla & Palomo, 2007; Santos, 2009). Its diet is mostly based on subterranean plant parts, although aerial parts may also be consumed (Borghi & Giannoni, 1997; Cotilla &

Palomo, 2007). Analogously to *M. lusitanicus,* there are reports of high consumption of geophytes (Soriguer & Amat, 1980), namely the subterranean parts of *Oxalis pes-caprae* Linnaeus 1753 (Bäumler et al., 1984; Mira 1999).

In normal conditions, *M. duodecimcostatus* presents densities of 100-400 individuals per hectare, whereas with favourable conditions, such as in irrigated crops, it can reach high annual average densities of 390 individuals per hectare (Cotilla & Palomo, 2007). This can be extended to 900 individuals per hectare in extraordinary favourable conditions (Cotilla & Palomo, 2007). In such scenarios *M. duodecimcostatus* can be considered a pest and may lead to 5-10% loss in fruit orchards (Bäumler et al., 1984; Vinhas, 1993; Cotilla & Palomo, 2007).

Its complex burrows comprise tunnels varying between 10-50cm of depth, which may increase to 1m during summer, when *M. duodecimcostatus* searches for soil moisture (Cotilla & Palomo, 2007). Galleries are highly branched and usually have a single nest for the family group and chambers for storing food, as *M. lusitanicus* (Cotilla & Palomo, 2007). Galleries of both voles can be differentiated because *M. duodecimcostatus* leave small monticules of soil near de openings, while *M. lusitanicus* do not (Purroy & Varela, 2005; Santos et al., 2009b).

The underground habits of *M. duodecimcostatus* are a very effective defensive strategy, so that it can only be caught when surfacing, by predators such as *Tyto alba* and small/medium sized carnivores, similarly to *M. lusitanicus* (Cotilla & Palomo, 2007).

1.2.3 Mating system

Natural populations of *M. lusitanicus* and *M. duodecimcostatus* present a balanced sex ratio (Paradis & Guédon, 1993), K selection strategy (Guédon et al., 1991b; Guédon & Pascal, 1993; Ventura et al., 2010) and reduced litter size (1-5 pups) (Guédon et al., 1991a, b).

Spatial overlap and similar home range for both sexes were observed in *M. lusitanicus*, with sexually active males showing restricted daily movements (Madison, 1980; Wolff, 1985; Salvioni, 1988; MacGuire et al., 1990; Lambin &

Krebs, 1991; Santos et al., 2010). This vole lives in small groups, composed of a couple and its pups, and with nests shared by males and females, or one female and one sub-adult (Mira & Mathias, 2007; Santos, 2009).

M. duodecimcostatus is also socially organized in small family groups (Paradis & Guédon, 1993; Cotilla & Palomo, 2007) and, similarly to *M. lusitanicus* (Madureira, 1982; Heske & Ostfeld, 1990; Ventura et al., 2010), presents a sexual monomorphism in terms of adult weight (Mira, 1999) and relatively small testis in adult males (Montoto et al., 2011).

These ecological and reproductive characteristics suggest that both sister species present a monogamous mating system.

1.2.4 Reproductive isolation barriers

Reproductive isolation is essential to speciation (Dobzhansky, 1937; Mayr, 1942; Mayr, 1970). Biological, morphological, phylogenetic and genetic species concepts agree that reproductive isolation mechanisms are fundamental to recognize a diverging population as a new species (Cracraft, 1997; Baker & Bradley, 2006); thus it is important to analyse how heterospecific mating is avoided. Two types of isolation may act against hybridization between two species: pre-mating and post-mating reproductive barriers (Coyne & Orr, 2004).

Pre-mating barriers prevent copulation and promote conspecific reproduction:

- Geographical isolation: species are separated by a physical barrier, such as a river or a mountain;
- Ecological/spatial isolation: species do not meet because they inhabit different habitats, even if they occur in the same geographical region;
- Temporal isolation: species present different sexually active periods;
- Behavioural isolation: potential mates meet, but show a preference for conspecifics over heterospecifics.

On the other hand, post-mating barriers affect fertilization, viability and sterility:

- Mechanical isolation: copulation is attempted, but is physically impossible, due to incompatible genitalia;
- Gametic isolation: the female immune system attacks the heterospecific sperm, after copulation; or gametes are incompatible and fertilization does not occur;
- Zygotic mortality: the egg is fertilized, but the zygote does not develop;
- Hybrid unviability: the hybrid embryo forms, but with a reduced viability;
- Hybrid sterility: the hybrid is viable, but as an adult it is sterile;
- Hybrid breakdown: first generation (F1) is viable and fertile, but further hybrid generations (second generation and backcrosses) may be unviable or sterile.

Sister species *M. lusitanicus* and *M. duodecimcostatus* share a considerable sympatry area (**Figure 2**). Additionally, physical barriers, such as mountains and rivers, do not seem to affect the distribution of these voles, because they inhabit low and high altitude locations (Cotilla & Palomo, 2007; Mira & Mathias 2007) and are proficient swimmers (Giannoni, 1993, 1994). Thus, it is very unlikely that geographical isolation acts as a reproductive barrier between these taxa (**Figure 7**).

Ecological/spatial isolation is also improbable because *M. lusitanicus* and *M. duodecimcostatus* inhabit similar habitats (Mira & Mathias, 1994; Paradis, 1995; Mathias, 1999; Mira, 1999; Cotilla & Palomo, 2007; Mira & Mathias, 2007) and occurred in syntopy in the past, since fossils from both voles were discovered in the Caldeirão cave (Tomar, Portugal) (Póvoas et al., 1992) and in the la Buena Pinta cave (Pinilla del Valle, Spain) (López-García, 2008) (**Figure 7**). Nevertheless, it is unknown if in the present day there are syntopic locations as well.

Furthermore, both species present similar sexually active periods (Cotilla & Palomo, 2007; Mira & Mathias, 2007), showing that temporal isolation does not seem to affect *M. lusitanicus* and *M. duodecimcostatus* reproductive isolation (**Figure 7**).

Conversely to the previous barriers, pre-mating behavioural isolation has been suggested by a preference for conspecific individuals over heterospecific (Soares, 2013) (**Figure 7**).

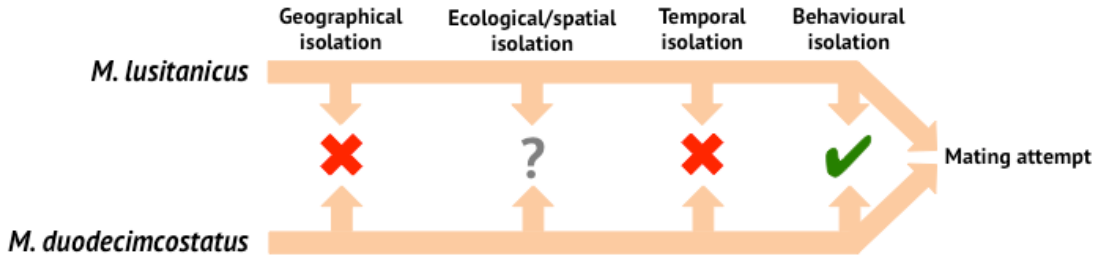


Figure 7 – Pre-mating reproductive barriers involved in *M. lusitanicus* and *M. duodecimcostatus* reproductive isolation. Cross = absent; check = present; ? = undetermined.

Regarding post-mating barriers, *M. lusitanicus* and *M. duodecimcostatus* can produce F1 hybrids in captivity (Wiking, 1976; Soares, 2013) and nature (Bastos-Silveira et al., 2012), reaching adulthood (Soares, 2013), revealing that mechanical isolation and the hybrid unviability barrier are very implausible (**Figure 8**). Zygotic mortality cannot be, however, discarded, since embryonic data is currently unavailable.

Male hybrids are infertile (Soares, 2013), according to the Haldane rule (Haldane, 1922), making further hybrid generations improbable, being coherent to the results presented by Bastos-Silveira and colleagues (2013) (**Figure 8**). These observations indicate that the hybrid sterility barrier is active and, consequently, the hybrid breakdown barrier is absent (**Figure 8**). Concerning the female hybrids, it is known that they are fertile (Soares, 2013), but additional data on second generation and backcrosses fertility is needed to discard the hybrid breakdown barrier for this gender (**Figure 8**).

Lastly, a partial gametic isolation seems to exist, because *M. lusitanicus* and *M. duodecimcostatus* can produce F1 hybrids, both in the lab (Wiking, 1976) and in nature (Bastos-Silveira et al., 2012), however heterospecific mating, in laboratory conditions, is less productive in terms of reproductive success than conspecific

mating (Soares, 2013), suggesting that fertilization may not always occur (**Figure 8**).

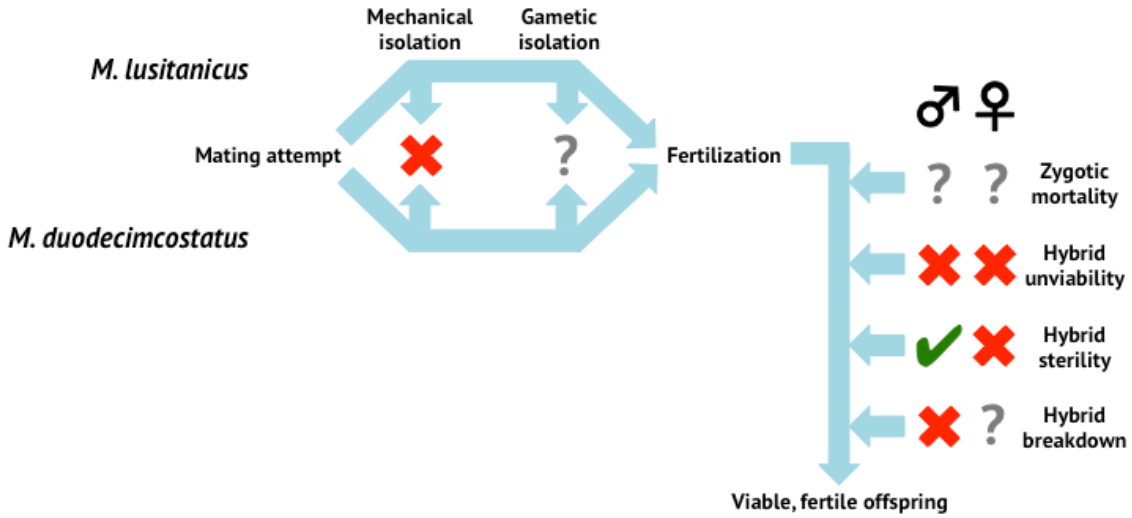


Figure 8 – Post-mating reproductive barriers involved in *M. lusitanicus* and *M. duodecimcostatus* reproductive isolation. Barriers after a successful fertilization are specified per gender. Cross = absent; check = present; ? = undetermined.

Hence, both pre-mating and post-mating barriers seem to be responsible for *M. lusitanicus* and *M. duodecimcostatus* reproductive isolation (**Figure 7** and **8**).

1.3 Aims and hypotheses

Considering that *M. lusitanicus* and *M. duodecimcostatus* incomplete reproductive isolation in nature seems to be associated to more than one barrier, possibly related to behavioural (pre-mating) and gametic isolation (post-mating), the present Ph.D. project comprised five specific aims:

- Identify candidate genes related to odour cues communication (**Chapter 2**);
- Analyse the expression of urinary proteins in both species (**Chapter 3**);
- Infer if both species favour conspecific to heterospecific mating (**Chapter 4 – 4.1**);
- Determine if both species present a pair bond, indicative of a monogamous mating system (**Chapter 4 – 4.2**);

- Evaluate the role of the sperm-binding protein ZP3 (zona pellucida 3), as a gametic isolation barrier (**Chapter 5**).

The results obtained enabled to test the following hypotheses:

- Hypothesis 1: Odour cues communication is an active behavioural reproductive barrier between *M. lusitanicus* and *M. duodecimcostatus*.

  Most mammals rely on chemosensory systems for communicating in a social context, either intra- or inter-specifically, using odour cues excreted in urine, faeces, saliva, sweat or milk (reviewed in Wyatt, 2003; Liberles, 2014). Hence, for the present hypothesis, a double approach was performed and focused both on odour cues (major histocompatibility complex I and II peptides) and respective receptors (olfactory receptors) to infer their potential role in the pre-mating reproductive isolation between *M. lusitanicus* and *M. duodecimcostatus*. In addition, proteomic analyses on the urine of *M. lusitanicus* and *M. duodecimcostatus* were performed in order to test the possible role of particular urinary proteins in odorous communication associated with species-specific mate choice.
- Hypothesis 2: *M. lusitanicus* and *M. duodecimcostatus* prefer conspecific to heterospecific mating in the presence of potential mates of both species.

  It is known that these voles can produce F1 hybrids in captivity (Wiking, 1976; Soares, 2013) and nature (Bastos-Silveira et al., 2012). However, hybridization in nature is a very rare event and in captivity it is only possible because either the subjects mate heterospecifically, "forced", or remain sexually naïve. Thus, to test the present hypothesis two artificial syntopic environments were established, for the first time, and populated with animals from each species and genders. The generated litters were genotyped in order to determine the maternal/paternal origin, and consequently if hybridization occurred or only conspecific mating was favoured.
- Hypothesis 3: *M. lusitanicus* and *M. duodecimcostatus* are socially monogamous.

Ecological and reproductive characteristics suggest that both sister species present a monogamous mating system (e.g. Madureira, 1982; Heske & Ostfeld, 1990; Guédon et al., 1991a, b; Guédon & Pascal, 1993; Paradis & Guédon, 1993; Mira, 1999; Santos, 2009; Ventura et al., 2010; Santos et al., 2010; Montoto et al., 2011). Nevertheless, behaviour assays, such as partner preference and selective aggression tests, and fieldwork paternity inference, had not been performed until the present day in order to clarify the type of monogamy exhibited by both voles. Here, partner preference tests were performed, using urinary and faecal odour cues, in order to determine if *M. lusitanicus* and *M. duodecimcostatus* conspecific couples reveal a pair bond, indicative of social monogamy.

- Hypothesis 4: The putative sperm-binding region of ZP3 is a gametic isolation barrier that impairs heterospecific mating between *M. lusitanicus* and *M. duodecimcostatus*.

  Mating between subjects of both species, in captivity, is less reproductively successful than between conspecifics (Soares, 2013). This observation suggests that fertilization of heterospecific gametes may not always occur. Therefore, in this hypothesis, the putative sperm-binding region of the ZP3 was tested as one of the control mechanisms that affect putative hybridizations after successful mating. This region is historically related to species-specific fertilization (Wassarman & Litscher, 1995; Wassarman, 1999; Wassarman et al., 2005); thus, it presents a potential role as a gametic barrier.

1.4 References

Acosta MJ, Marchal JA, Romero-Fernández I, Megías-Nogales B, Modi WS, Sánchez Baca A (2010a) Sequence analysis and mapping of the Sry gene in species of the subfamily Arvicolinae (rodentia). Sex Dev 4(6):336–347

Acosta MJ, Marchal JA, Fernández-Espartero C, Romero-Fernández I, Rovatsos MT, Giagia-Athanasopoulou EB, Gornung E, Castiglia R, Sánchez A (2010b)

Characterization of the satellite DNA Msat-160 from species of Terricola (Microtus) and Arvicola (Rodentia, Arvicolinae). Genetica 138(9-10):1085–1098

Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. J Mammal 87(4):643–662

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

Baümler VW, Naumann-Etienne C, Godinho J, Grilo C, Sezinando T, Vinhas A (1984) Schadliche feldnager in Portugal. Anz Schadlingskde, Pflazenschutz, Umweltschutz 57:134–138

Bennett KD, Tzedakis PC, Willis KJ (1991) Quaternary refugia of north European trees. J Biogeogr 18:103–115

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

Borghi CE, Giannoni SM (1997) Dispersal of geophytes by mole-voles in the Spanish Pyrenees. J Mammal 78:550–555

Braaker S, Heckel G (2009) Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole *Microtus arvalis*. Mol Ecol 18:2518–2531

Brunet-Lecomte P, Chaline J (1991) Morphological evolution and phylogenetic relationships of the European ground voles (Arvicolidae, Rodentia). Lethaia 24(1):45–53

Brunet-Lecomte P, Brochet G, Chaline J, Delibes M (1987) Morphologie dentaire comparée de *Pitymys lusitanicus* et *Pitymys duodecimcostatus* (Arvicolidae, Rodentia) dans le nord-ouest de l'Espagne. Mammalia 51:145–158

Brunhoff C, Galbreath KE, Fedorov VB, Cook JA, Jaarola M (2003) Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. Mol Ecol 12(4):957–968

Cabrera A (1914) Fauna Ibérica: Mamíferos. Museo Nacional de Ciencias Naturales, Madrid

Chaline J (1966) Une exemple d'évolution chez les arvicolids (Rodentia): les lignées *Allophaiomus-Pitymus* et *Microtus*. C R Acad Sci Hebd Seances Acad Sci D 263(17):1202–1204

Chaline J (1972) Le role des Rongeurs dand l'éboration d'une biostratigraphie et d'une stratigraphie climatique fine du Quaternaire. Mém Bur Rech Géol Min 77:275–379

Chaline J (1974) Un nouveau critère d'étude des *Mimomys*, et les rapports de *Mimomys occitanus*, *Mimomys stehlini* et de *Mimomys polonicus* (Arvicolidae, Rodentia). Acta zool cracov 19:337–356

Chaline J (1987) Arvicolid Data (Arvicolidae, Rodentia) and Evolutionary Concepts. Evol Biol 21:237–310

Chaline J, Graf J-D (1988) Phylogeny of the Arvicolidae (Rodentia): Biochemical and Paleontological Evidence. J Mammal 69(1):22

Chaline J, Mein P (1979) Les Rongeurs et l'évolution. Doin Éditeurs, Paris

Chaline J, Brunet-Lecomte P, Montuire S, Viriot L, Courant F (1999) Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. Ann Zool Fennici 36(4): 239–267

Conroy CJ, Cook JA (1999) MtDNA evidence for repeated pulses of speciation within arvicoline and murid rodents. J Mamm Evol 6(3):221–245

Conroy CJ, Cook JA (2000) Molecular Systematics of a Holarctic Rodent (Microtus: Muridae) J Mammal 81(2):344–359

Cotilla I, Palomo LJ (2007) *Microtus duodecimcostatus* (de Selys-Longchamps, 1839). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 422–425

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

Cracraft J (1997) Species concepts in systematics and conservation biology: an ornithological viewpoint. In: Claridge MF, Dawah HA, Wilson MR (eds) Species: the units of biodiversity. Chapman & Hall; London

Dawson AG (1992) Ice Age Earth: Late Quaternary Geology and Climate. Routledge, London and New York

Dobzhansky T (1937) Genetics and the origin of species. Columbia University Press, New York

Ellerman JR, Morrison-Scott TCS (1951) Check list of Palearctic and Indian mammals. British Museum, London

Fink S, Excoffier L, Heckel G (2004) Mitochondrial gene diversity in the common vole *Microtus arvalis* shaped by historical divergence and local adaptations. Mol Ecol 13(11):3501–3514

Fink S, Excoffier L, Heckel G (2006) Mammalian monogamy is not controlled by a single gene. P Natl Acad Sci USA 103(29):10956–10960

Fink S, Excoffier L, Heckel G (2007) High variability and non-neutral evolution of the mammalian arginine vasopressin receptor. BMC Evol Biol 7, e176

Fink S, Fischer MC, Excoffier L, Heckel G (2010) Genomic Scans Support Repetitive Continental Colonization Events during the Rapid Radiation of Voles (Rodentia: *Microtus*): the Utility of AFLPs versus Mitochondrial and Nuclear Sequence Markers. Syst Biol 59:548–572

Fumagalli L, Taberlet P, Stewart DT, Gielly L, Hausser J, Vogel P. Molecular phylogeny and evolution of *Sorex* shrews (Soricidae: insectivora) inferred from mitochondrial DNA sequence data. Mol Phylogenet Evol 11(2):222–235

Galewski T, Tilak MK, Sanchez S, Chevret P, Paradis E, Douzery EJ (2006) The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. BMC Evol Biol 6:80.

Garapich A, Nadachowski A (1996) A contribution to the origin of *Allophaiomys* (Arvicolidae, Rodentia) in Central Europe: the relationship between *Mimomys* and *Allophaiomys* from Kamyk (Poland). Acta zool craoc 39(1):179–184

Guédon G, Pascal M (1993) Dynamique de population du campagnol provençal (*Pitymys duodecimcostatus* de Sélys-Longchamps, 1839) dans deux agrosystémes de la région montpelliéraine. Rev Ecol – Terre Vie 48:375–398

Guédon G, Pascal M, Mazouin F (1991a) Le campagnol provençal en captivité

(*Pitymys duodecimcostatus* de Sélys-Longchamps, 1839) (Rongeurs, Microtidés) I. La reproduction. Mammalia 55:97–106

Guédon G, Pascal M, Mazouin F (1991b) Le campagnol provençal en captivité (*Pitymys duodecimcostatus* de Sélys-Longchamps, 1839) (Rongeurs, Microtidés) II. La croissance. Mammalia 55:397–406

Getz LL (1985) Habitats. In: Tamarin RH (ed) Biology of New World Microtus. The American Society of Mammalogists, Special Publication 8, pp. 286–309

Giannoni SM, Borghi CE, Martinezrica JP (1993) Swimming ability of a fossorial Iberic vole *Microtus* (*Pitymys*) *lusitanicus* (Rodentia, Microtinae). Mammalia 57(3):337–340

Giannoni SM, Borghi CE, Martinezrica JP (1994) Swimming ability of the Mediterranean pine vole *Microtus* (*Terricola*) *duodecimcostatus*. Acta Theriol 39(3):257–265

Giannoni SM, Marquez R, Borghi CE (1997) Airborne and substrate borne communications of *Microtus* (*Terricola*) *gerbei* and *M.* (*T.*) *duodecimcostatus*. Acta Theriol 42(2):123–141

Gornung E, Castiglia R, Rovatsos M, Marchal JA, Díaz de la Guardia-Quiles R, Sanchez A (2011) Comparative cytogenetic study of two sister species of Iberian ground voles, *Microtus* (*Terricola*) *duodecimcostatus* and *M.* (*T.*) *lusitanicus* (rodentia, cricetidae). Cytogenet Genome Res 132(3):144–150

Gromov V.J, Polyakov IY (1977) Fauna of the SSSR: Mammals III, 8: voles (Microtinae). Nauka, Leningrad

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. J Genet 12(2):101–109

Heckel G, Burri R, Fink S, Desmet J-F & Excoffier L (2005) Genetic structure and colonization processes in European populations of the common vole *Microtus arvalis*. Evolution 59(10):2231–2242

Hellborg L, Gündüz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. Mol Ecol 14(7):2025–2031

Heske EJ, Ostfeld RS (1990) Sexual dimorphism in size, relative size of testes, and mating systems in North American voles. J Mamm 71(4):510–519

Hoffmann RS, Koeppl JW (1985) Zoogeography. In: Tamarin RH (ed) Biology of New World Microtus. The American Society of Mammalogists, Boston, pp 84–115

Huntley B (1988) Glacial and Holocene vegetation history: Europe. In: Huntley B, Webb T III (eds) Vegetation History, Kluwer, Dordrecht

IUCN Red List of Threatened Species (2016) http://www.iucnredlist.org

Jaarola M, Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. Mol Ecol 11(12):2613–2621

Jaarola M, Martínková N, Gündüza I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedakis-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioğlu H, Mathias ML, Villate I, Searle JB (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol 33(3):647–663

Lambin X, Krebs CJ (1991) Spatial organization and mating system of *Microtus townsendii*. Behav Ecol Sociobiol 28:353–363

Liberles SD (2014) Mammalian pheromones. Annu Rev Physiol 76:151-175

López-García JM, Blain HA, Cuenca-Bescós G, Arsuaga JL (2008) Chronological, environmental, and climatic precisions on the Neanderthal site of the Cova del Gegant (Sitges, Barcelona, Spain). J Hum Evol 55(6):1151–1155

MacGuire B, Pissuto T, Getz LL (1990) Patterns of visitation in prairie voles (*Microtus ochrogaster*) reveal a role for males in population regulation. In: Tamarin RH et al. (eds) Social Systems and Population Cycles in Voles. Advances in Life Sciences, Birkhäuser Verlag, Basel, pp 89–100

Madison DM (1980) An integrated view of the social biology of *Microtus pennsylvanicus*. Biologist 62:20–33

Madureira ML (1981) Discriminant analysis in portuguese pine voles: *Pitymys lusitanicus* Gerbe and *Pitymys duodecimcostatus* de Sélys-Longchamps (Mammalia: Rodentia) Arq Mus Bocage Série A I:111–122

Madureira ML (1982) New data on the taxonomy of portuguese *Pitymys*. Arq Mus Bocage A I:341–368

Madureira ML (1984) A biologia de *Microtu*s (*Pitymys*) *duodecimcostatus* de Sélys-Longchamps, 1839 e *M.* (*P.*) *lusitanicus* Gerbe, 1879 em Portugal (Arvicolidae, Rodentia): taxonomia, osteologia, ecologia e adaptações. Dissertation, Faculdade de Ciências, Universidade de Lisboa, Lisboa

Mathias ML (1990) Morphology of the incisors and the burrowing activity of Mediterranean and Lusitanian Pine Voles (Mammalia, Rodentia). Mammalia 54: 301–306

Mathias ML (1996) Skull size variability in adaptation and speciation of the semifossorial pine voles *Microtus duodecimcostatus* and *M. lusitanicus* (Arvicolidae, Rodentia). In: Proceedings of The First Congress of Mammalogy, Museu Nacional de História Natural, Museu Bocage, Lisbon, pp 271–286

Mathias ML (1999) Guia dos Mamíferos Terrestres de Portugal Continental, Açores e Madeira. Instituto da Conservação da Natureza, Lisboa

Mayr E (1942) Systematics and the origin of species. Columbia University Press, New York

Mayr E (1970) Populations, species, and evolution. Harvard University Press, Cambridge

McKenna MC, Bell SK (1997) Classification of mammals; above the species level. Columbia University Press, New York

Mira A (1999) Ecologia do rato-cego-mediterrânico, *Microtus* (*Terricola*) *duodecimcostatus* (Rodentia, Muridae), em pomares de citrinos. Dissertation, Faculdade de Ciências, Universidade de Lisboa, Lisboa

Mira A, Mathias ML (1994) Conditions controlling the colonization of an orange orchard by *Microtus duodecimcostatus* (Rodentia, Arvicolidae). Pol Ecol Stud 20:249–255

Mira A, Mathias ML (2007) *Microtus lusitanicus* (Gerbe, 1879). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 418–421

Mitchel-Jones AJ, Amori G, Bogdanowicz W, Kryšrufek B, Reijnders PJH, Spitzenberger F, Stubbe M, Thissen JBM, Vohralík V, Zima J (1999) The Atlas of European Mammals. T & AD Poyser, London

Montoto LG, Magaña C, Tourmente M, Martín-Coello J, Crespo C, Luque-Larena JJ, Gomendio M, Roldan ERS (2011) Sperm competition, sperm numbers and sperm quality in muroid rodents. PLoS One 6(3):e18173

Musser GM, Carleton MD (1993) Family Cricetidae. In: Wilson DE, Reeder DM (eds) Mammal species of the world: a taxonomic and geographic reference. Smithsonian Institution, Washington, pp 955–1189

Nowak RM (1999) Walker's Mammals of the World. Sixth Edition. Volume II. Johns Hopkins University Press, Baltimore

Paradis E (1995) Survival, immigration and habitat quality in the Mediterranean pine vole. J Anim Ecol 64:579–591

Paradis E, Guédon G (1993) Demography of a Mediterranean microtine: the Mediterranean pine vole, *Microtus duodecimcostatus*. Oecologia 95:47–53

Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012) Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. Mol Ecol 21:6015–6032

Póvoas L, Zilhão J, Chaline J, Brunet-Lecomte P (1992) La faune de rongeurs du Pléistocene supérieur de la frotte de Caldeirão (Tomar, Portugal). Quaternaire 3(1):40–47

Purroy FJ, Varela JM (2005) Mamíferos de España. Península: Baleares y Canarias. Lynx Edicions. Barcelona

Quina AS, Bastos-Silveira C, Miñarro M, Ventura J, Jiménez R, Paulo OS, Mathias ML (2015) P53 gene discriminates two ecologically divergent sister species of pine voles. Heredity 115(5):444–451

Rabeder G (1981) Die Arvicoliden (Rodentia, Mammalia) aus dem Pliozän und dem alterem Pleistozän von Niederosterreich. Beitr Palaont Osterr 8:1–343

Rabeder G (1986) Herkunft und fruhe Evolution der Gattung *Microtus* (Arvicolidae, Rodentia). Z Saugetierkd (Mamm Biol):51:350–367

Rekovets LI, Nadachowski A (1995) Pleistocene voles (Arvicolidae) of the Ukraine. Paleontol Evol 28–29:145–245

Repenning C (1992) *Allophaiomys* and the age of the Olyor Suite, Krestovka Sections, Yakutia. U S Geol Surv Bull 2037:1–98

Richmond GM (1996) The INQUA-approved provisional Lower-Middle Pleistocene boundary. In: Turner c (ed) The Early-Middle Pleistocene in Europe. Balkema, Rotterdm, pp 319–326

Robovský J, Ra Ánková V, Zrzavý J (2008) Phylogeny of Arvicolinae (Mammalia, Cricetidae): utility of morphological and molecular datasets in a recently radiating clade. Zool Scripta 37: 571–590

Salvioni M (1988) Home range and social behaviour of three species of European *Pitymys* (Mammalia, Rodentia). Behav Ecol Sociobiol 22:203–210

Santos SML (2009) Factors influencing the distribution of the lusitanian and the mediterranean pine voles (*Microtus lusitanicus* and *M. duodecimcostatus*) in Portugal: a multiscale approach. Dissertation, Faculdade de Ciências, Universidade de Lisboa, Lisboa

Santos SM, Mira AP, Mathias ML (2009a) Using presence signs to discriminate between similar species. Integr Zool 4:258–264

Santos SM, Mira AP, Mathias ML (2009b) Factors influencing large scale distribution of two sister species of pine voles (*Microtus lusitanicus* and *M. duodecimcostatus*): the importance of spatial autocorrelation. Can J Zool 87: 1227–1240

Santos SM, Lourenço RF, Mathias ML, Mira AP (2010) Spatial and temporal ecology of the Lusitanian pine vole (*Microtus lusitanicus*) in a Mediterranean polyculture. Anim Biol 60:209–227

Santos SM, Mathias ML, Mira AP (2011)The influence of local, landscape and spatial factors on the distribution of the Lusitanian and the Mediterranean pine voles in a Mediterranean landscape. Mammalian Biology - Zeitschrift für Säugetierkunde 76(2):133–142

Soares JMA (2013) Fitness dos híbridos entre duas espécies de ratos fossadores,

Microtus lusitanicus (Gerbe, 1879) e *M. duodecimcostatus* (de Selys-Longchamps, 1839) - uma abordagem multidisciplinary. Dissertation, Faculdade de Ciências da Universidade de Lisboa, Lisboa

Soriguer RS, Amat JA (1980) On the structure and function of the burrows of the Mediterranean vole (*Pitymys duodecimcostatus*). Acta Theriol 25:268–270

Spitz F (1978) Etude craniometrique du genre Pitymys. Mammalia 42:267–304

Tamarin RH (1985) Biology of new world *Microtus*. American Society of Mammalogists, Shippensburg, pp 340–372

Van der Meulen JJ (1978) *Microtus* and *Pitymys* (Arvicolidae) from Cumberland Cave, Maryland with a comparison of some New and Old World species. Annls Carnegie Mus 47:101–145

Ventura J, Jiménez L, Gisbert J (2010) Breeding characteristics of the Lusitanian pine vole, *Microtus lusitanicus*. Anim Biol 60:1–14

Vinhas A (1993) *Microtus lusitanicus* (Rato cego) e *Microtus duodecimcostatus* (Rato cego-mediterrênico), roedores pragas das culturas. Rev Ciên Agrár 17:375–382

Wassarman PM (1999) Mammalian Fertilization: Review Molecular Aspects of Gamete Adhesion, Exocytosis, and Fusion. Cell 96:175–183

Wassarman PM, Litscher ES (1995) Sperm-egg recognition mechanisms in mammals. Curr Top Dev Biol 30:1–19

Wassarman PM, Jovine L, Qi H, Williams Z, Darie C, Litscher ES (2005) Recent aspects of mammalian fertilization research. Mol. Cell. Endocrinol 234:95–103

Wiking VH (1976) Karyologie und Biologie der beiden iberischen Wuhlmausarten *Pitymys mariae* und *Pitymys duodecimcostatus*. Z Zool Syst Evol 14:104–129

Wolff JO (1985) Behaviour. In: Tamarin RH (ed) Biology of new world *Microtus*. American Society of Mammalogists, Shippensburg, pp 340–372

Wyatt TD (2003) Pheromones and animal behaviour: communication by smell and taste. Cambridge University Press, Cambridge

Chapter 2

Candidate genes related to odour cues communication

Subchapters:

2.1 Olfactory receptors

2.2 Major histocompatibility complex I and II

2.1 Olfactory receptors

Type of publication: Short communication

Reference: Duarte MA[1,2,3], Heckel G[4,5], Mathias ML[2,3], Bastos-Silveira C[1] (2016) Olfactory receptors and behavioural isolation: a study on *Microtus* voles. Mammal Res. 61(4):399–407. DOI: 10.1007/s13364-016-0266-0.

[1] Museu Nacional de História Natural e da Ciência, Departamento de Zoologia e Antropologia, Universidade de Lisboa, Rua da Escola Politécnica, 58, Lisbon 1250-102, Portugal.

[2] Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[3] Centro de Estudos de Ambiente e Mar, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[4] Computational and Molecular Population Genetics, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland.

[5] Swiss Institute of Bioinformatics, Genopode, Lausanne, Switzerland.

2.1.1 Abstract

Genetic variation in olfactory receptors may trigger mate choice, suggesting that olfaction has undergone diversifying selection in diverging populations and may contribute to premating reproductive isolation. In the present study, we analysed two olfactory receptor genes as candidate barriers of reproductive isolation between two recently divergent voles: *Microtus lusitanicus* and *Microtus duodecimcostatus*. In addition, evolutionary relationships and signs of positive selection were inferred in a European subgenera context, based on 76 samples from 14 species. DNA sequence analysis revealed the presence of shared haplotypes among various *Microtus* species. Tests of selection detected negatively selected amino acids in the extracellular loops of both olfactory receptors and a majority of negatively selected residues in the transmembrane helices, the most variable regions responsible for the reception of odorants. Our findings suggest that, for several *Microtus* species, including *M. lusitanicus* and *M. duodecimcostatus*, these proteins probably recognise conserved odour cues not related to behavioural isolation.

2.1.2 Keywords

Behavioural isolation; Olfaction; Olfactory receptors; *Microtus*; Sister species.

2.1.3 Introduction

Behaviour is highly influenced by olfaction, the dominant sense in most mammals (reviewed in Arakawa et al. 2008). Behavioural interactions make often use of information on species, sex and identity that is provided to the receiving individual in polymorphic odour cues (e.g. Hurst and Beynon 2004; Petrulis 2013). These odour cues may come from a variety of sources (e.g. urine, faeces and specialised scent gland secretions) that are detected by elaborated olfactory systems mostly specialised in the detection of volatile molecules present in the nasal airstream (Brennan and Kendrick 2006). The main olfactory epithelium typically contains receiver proteins, such as olfactory receptors, which are expressed by olfactory sensory neurons (Zhang et al. 2004; Fleischer et al. 2009). Olfactory receptors are highly variable, consistent with the structural diversity of odour cue molecules (e.g. Emes et al. 2004; Ignatieva et al. 2014). In mammals, olfactory receptors have been mainly analysed not only in expression and repertoire studies (e.g. Feldmesser et al. 2006; Gilad and Lancet 2003; Rouquier et al. 2000; Young et al. 2003; Zhang et al. 2004) but also in evolutionary (e.g. Gaillard et al. 2004; Gilad et al. 2003; Li et al. 2015; Zhuang et al. 2009) and phylogenetic contexts (e.g. McGowen 2011). Furthermore, genetic variation in olfactory receptors may trigger mate choice, suggesting that olfaction has undergone diversifying selection in diverging populations and may contribute to premating reproductive isolation (Li et al. 2015; Smadja and Butlin 2009).

Odour communication has been considered as part of a behavioural barrier of prezygotic reproductive isolation in rodents (e.g. Moore 1965; Nevo et al. 1976; Theiler and Blanco 1996; Kotenkova and Naidenko 1999; Stippel 2009), essential to speciation in the absence of other reproductive barriers. This seems to be the case of sister species Lusitanian pine vole *M. lusitanicus* Gerbe (1879) and

Mediterranean pine vole *M. duodecimcostatus* de Selys-Longchamps (1839), two of the most recent *Microtus* species, estimated to have diverged only 60,000 years ago (Brunet-Lecomte and Chaline 1991). *M. lusitanicus* and *M. duodecimcostatus* share a considerable area of sympatry in the Iberian Peninsula (e.g. Santos 2009; Bastos-Silveira et al. 2012), and physical barriers, such as mountains and rivers, do not affect the distribution of these voles, because they inhabit low and high altitude locations (Cotilla and Palomo 2007; Mira and Mathias 2007) and are proficient swimmers (Giannoni et al. 1993, 1994). Regarding ecological/spatial isolation, *M. lusitanicus* and *M. duodecimcostatus* inhabit similar habitats, e.g. meadows, woods and agricultural areas (e.g. Cotilla and Palomo 2007; Mira and Mathias 2007), and can occur in syntopy (Duarte et al. 2015). These species present similar sexually active periods (Cotilla and Palomo 2007; Mira and Mathias 2007), indicating that temporal isolation does not play a role in preventing heterospecific copulation between both voles.

Odour cues appear to contribute to behavioural isolation between *M. lusitanicus* and *M. duodecimcostatus*. Two-way mate choice assays in a Y-shaped olfactometer with urine and faeces as stimuli revealed a preference for conspecific individuals in these sister species (Soares 2013). Using a similar methodology, odour communication was also associated to pair bonding behaviour in *M. lusitanicus* and *M. duodecimcostatus* (Duarte et al. 2015). Hence, olfactory discrimination is a potential premating reproductive isolation mechanism between these sister vole species as in other Cricetidae taxa (e.g. Moore 1965; Theiler and Blanco 1996) and rodents in general (e.g. Nevo et al. 1976; Pillay et al. 1995; Kotenkova and Naidenko 1999; Smadja and Ganem 2008; Stippel 2009).

In the present study, we chose a candidate gene approach as a first step for a molecular understanding of the potential contribution of olfactory receptors to reproductive isolation in the rapidly speciating *Microtus* genus (Fink et al. 2010; Beysard et al. 2012, 2015). Molecular data for olfactory receptors is not available for *Microtus sp.*; thus, we based our selection of candidate genes on information from *Mus musculus*, the closest animal model. We chose the class II olfactory

receptors Olfr31 and Olfr57 (Glusman et al. 2000; Niimura and Nei 2007) because they are expressed in cell lines of the mouse olfactory placode, which gives rise to olfactory sensory neurons in the olfactory epithelium (Illing et al. 2002; Pathak et al. 2009). Genetic variation in these receptors may thus lead to functionally relevant variation in the body region where odour cues are primarily perceived.

Given very high levels of genetic polymorphism in the *Microtus* genus (Jaarola et al. 2004; Fink et al. 2007, 2010; Fischer et al. 2014; Lischer et al. 2014), we expected high variation in Olfr31 and Olfr57 and possibly segregating receptor types between sibling species such as *M. lusitanicus* and *M. duodecimcostatus*. If these receptors were involved in reproductive isolation through odour communication, this may lead to molecular signals of positive selection in the relevant peptides. In particular, we expected to detect positively selected amino acids in the extracellular loops and extracellular half of the transmembrane helices of Olfr31 and Olfr57, since these variable regions are responsible for the binding of odour molecules (Emes et al. 2004). Molecular signatures of adaptive evolution can be difficult to detect in very recently diverged species (e.g. Fink et al. 2007), thus we extended our analyses to cover a total of 14 European species which span most of the evolutionary divergence in the *Microtus* genus (see Fink et al. 2010).

2.1.4 Materials and methods

Seventy-six tissue samples from 14 European *Microtus* species (**Online Resource 1**) were stored in absolute ethanol at −20 °C. Genomic DNA was isolated using a phenol-chloroform extraction procedure (Sambrook et al. 1989).

Our molecular analyses targeted a part of the single exon each for Olfr31 and Olfr57 based on PCR primer pairs designed for *Mus musculus* (Pathak et al. 2009). Reactions contained 100 ng of template DNA, 0.3 mM of each primer, 1.25 U of GoTaq® Flexi DNA Polymerase (Promega), 1× buffer (Promega), 2.5 mM MgCl2, 0.1 µg of BSA (New England Biolabs), and 0.2 mM of each dNTP (Thermo Scientific), to a final volume of 25 µl. PCR amplifications were performed in a MyCycler thermal

cycler (Bio-Rad Laboratories Inc.) and consisted in denaturation at 95 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 58 °C for 1 min and extension at 72 °C for 1 min. An extension step at 72 °C for 10 min was added at the end. PCR products were verified on 1 % agarose gels and purified using ExoI/FastAP protocol (Fermentas). Sequencing using the amplification primers was carried out by Macrogen Inc. (South Korea and the Netherlands) and at the Institute of Ecology and Evolution, University of Bern, using ABI Prism® 3130 Genetic Analyzer (Applied Biosystems).

DNA sequences were aligned using Sequencher 4.8 (Gene Codes Corporation). JModelTest 0.0.1 (Posada 2008) was used to select the best-fitting model of nucleotide substitution (TPM1uf + I, Kimura 1981) based on the Akaike information criterion (Akaike 1974). We applied a recent approach to integrate heterozygous information in existing phylogenetic programs by repeated random haplotype sampling (Lischer et al. 2014). This method generates haploid sequences for each individual by randomly selecting a haplotype from the detected alternative alleles at each position. A tree is then inferred and the process of haplotype generation and tree inference is repeated multiple times, from which a majority rule consensus tree is generated that covers the full extent of allelic and haplotypic variation. Thus, this approach tries to avoid an underestimation of sequence divergence and branch length in the constructed phylogenetic tree (see Lischer et al. 2014). Both Olfr31 and Olfr57 alignments were subjected to n = 10.000 replicates for the maximum likelihood analysis (RAxML) (Stamatakis 2014), and n = 20 replicates, nchains = 4, ngen = 2.000.000 and mcmc burn-in = 500.000 for the Bayesian inference analysis (MrBayes) (Ronquist and Huelsenbeck 2003). The outgroup chosen for both genes was *Mus musculus*. Consensus trees were edited using FigTree version 1.3.1.

DNA polymorphism parameters were estimated using DnaSP version 5.10.1 (Librado and Rozas 2009). Between species pairwise divergences were calculated using the TrN + I + G (Olfr31) and TrN models (Olfr57) (Tamura and Nei 1993) implemented in MEGA version 5.1 (Tamura et al. 2011), with standard deviations

estimated from 10,000 bootstrap replicates. Recombination was inferred using methods implemented in the HyPhy package (Pond et al. 2005) web interface DataMonkey (Delport et al. 2010) and RDP 4 (Martin et al. 2010).

We tested for positive selection using the CodeML subroutine included in PAML 4.8 (Yang 1997, 2007). Maximum likelihood estimations of ω (non-synonymous/synonymous substitution rates) among codons were generated according to six models: M0 (one ω), M1 (nearly neutral), M2 (positive selection), M3 (discrete), M7 (nearly neutral with beta distribution approximating ω variation) and M8 (positive selection with beta distribution approximating ω variation) (Goldman and Yang 1994; Yang et al. 2000, 2005). Additionally, branch-site models were tested in order to allow ω variation among amino acids in the protein and across branches on the phylogenetic tree and thereby detect possible positive selection affecting a few sites along particular lineages (Yang 1998; Yang and Nielsen 1998). We compared the null (model = 2; NSsites = 2; ω = 1) and neutral M1a (model = 0; NSsites = 1; ω = 1) models to MA1 (model = 2; NSsites = 2; ω estimated). Likelihood ratio tests (LRT) of M0 vs. M3, M1 vs. M2, M7 vs. M8, null model vs. MA1 and M1a vs. MA1 were performed in order to determine the most likely model (Nielsen and Yang 1998; Yang et al. 2000). Positively selected sites under M2, M3, M8 and MA1 were identified using the Naive Empirical Bayes and the Bayes Empirical Bayes analysis (Yang et al.2005). Since the power of CodeML can be affected by the accuracy of the input phylogenetic tree (Anisimova et al. 2003), we combined PAML results with HyPhy selection detection methods: SLAC (Pond and Frost 2005), FEL (Pond and Frost 2005), IFEL (Pond et al. 2006) and MEME (Murrell et al. 2012). Due to alignment size restrictions, it was not possible to test REL (Pond et al. 2005) and branch-site REL (Pond et al. 2011).

2.1.5 Results and discussion

Amplifications were successful for most species, with the exception of *M. gerbei* and *M. tatricus* for which Olfr57 could not be amplified (**Online Resource 2**). This positive outcome suggests that these olfactory receptors may be also fruitful as

molecular markers for other *Microtus* taxa (Cricetidae), or even other eumuroids, considering that the primers used were designed for the mouse model (Pathak et al. 2009), which belongs to a different family (Muridae). For Olfr31, we obtained a 352-bp fragment corresponding to *Mus musculus* Olfr31 position 225–576. For Olfr57, we amplified a 488-bp fragment equivalent to *Mus musculus* Olfr57 position 324–811. A tight homology to *Mus musculus* DNA sequences, including two characteristic sequence motifs (transmembrane domain 3 MAYDRYVAIC for Olfr31 and Olfr57, and transmembrane domain 6 KAFSTCASH for Olfr57), and an absence of stop codons and indels indicate that these gene fragments are functional olfactory receptors and do not correspond to pseudogenes (e.g. Malnic et al. 2004). Olfr31 and Olfr57 sequences were collapsed into 31 and 16 unphased diploid genotypes, respectively (**Online Resource 2**). Considering the full European *Microtus* set, nucleotide diversity and number of variable and parsimony informative sites are higher for Olfr57 than for Olfr31 (**Online Resource 2**). The same does not apply when considering the *M. lusitanicus* and *M. duodecimcostatus* subsets alone (**Online Resource 2**). We have deposited the obtained genotypes into GenBank (accession numbers KU172584-KU172615 for Olfr31 and KU172616-KU172632 for Olfr57). These are the first contributions of DNA sequences of Olfr31 and Olfr57 from non-model vertebrates and of olfactory receptor genes in general for *Microtus sp.*. The limited available data are only from mouse transcriptome repertoire studies (e.g. Young et al. 2003).

A total of seven haplotypes were shared by more than one *Microtus* species, four for Olfr31 and three for Olfr57 (**Figs. 1** and **2**). These repeated random haplotypes were generated in order to integrate Olfr31 and Olfr57 heterozygous sites in our phylogenetic analyses (see Lischer et al. 2014). Considering Olfr31, two haplotypes were shared by the sister species *M. lusitanicus* and *M. duodecimcostatus*, another one by *M. lusitanicus*, *M. duodecimcostatus* and the other *Terricola* voles *M. tatricus* and *M. felteni*, and the fourth by *M. socialis* (*Hyrcanicola*) and *M. schelkovnikovi* (*Microtus*) (**Fig. 1**). Analogously, for Olfr57, *M. lusitanicus* and *M. duodecimcostatus* shared two haplotypes, and *Terricola M. multiplex* and *M.*

subterraneus presented a common haplotype (**Fig. 2**). The presence of shared haplotypes also supports the close evolutionary relationship between the species in the *Terricola* subgenus, particularly the recently diverged *M. lusitanicus* and *M. duodecimcostatus*. Consistent with previous studies (Jaarola et al. 2004; Bastos-Silveira et al. 2012; Barbosa et al. 2013), genetic divergence between both species was relatively low: 0.2 % for Olfr57 and 0.5 % for Olfr31 (**Online Resource 3**). The highest genetic divergence involved species from older *Microtus* lineages (Fink et al. 2010): *M. cabrerae* for Olfr31 (2–3.5 %) and *M. agrestis* for Olfr57 (3.5–4.6 %) (**Online Resource 3**).

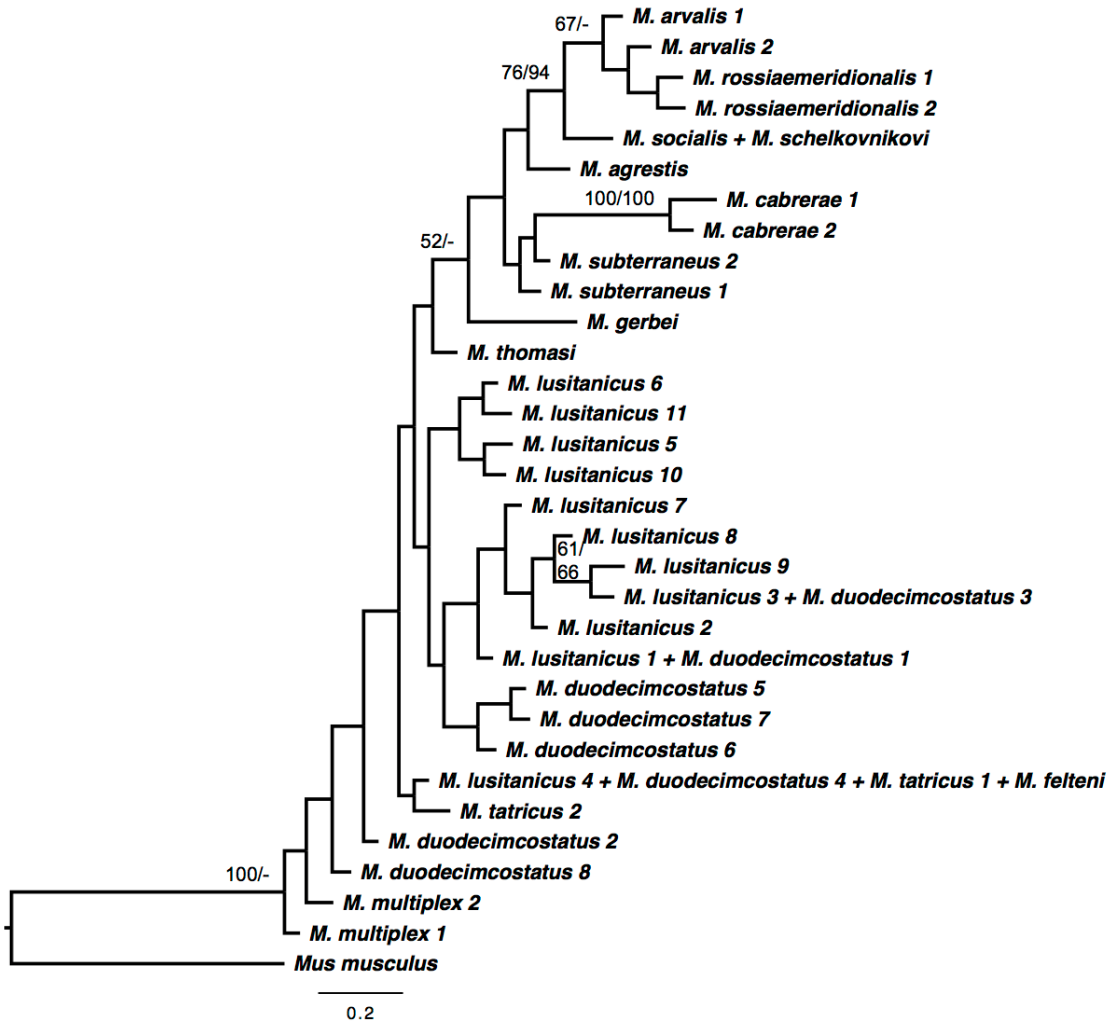


Figure 1 – Bayesian inference phylogenetic tree obtained for the Olfr31 gene fragment. Posterior probability (Bayesian inference) and bootstrap (maximum likelihood) values >50 % are indicated.

Maximum likelihood and Bayesian inference tree topologies were congruent for each of the respective data sets (we only present the latter, **Figs. 1** and **2**). Phylogenetic trees did not reflect the taxonomy attributed at the subgenera level, nor the geographic origin.

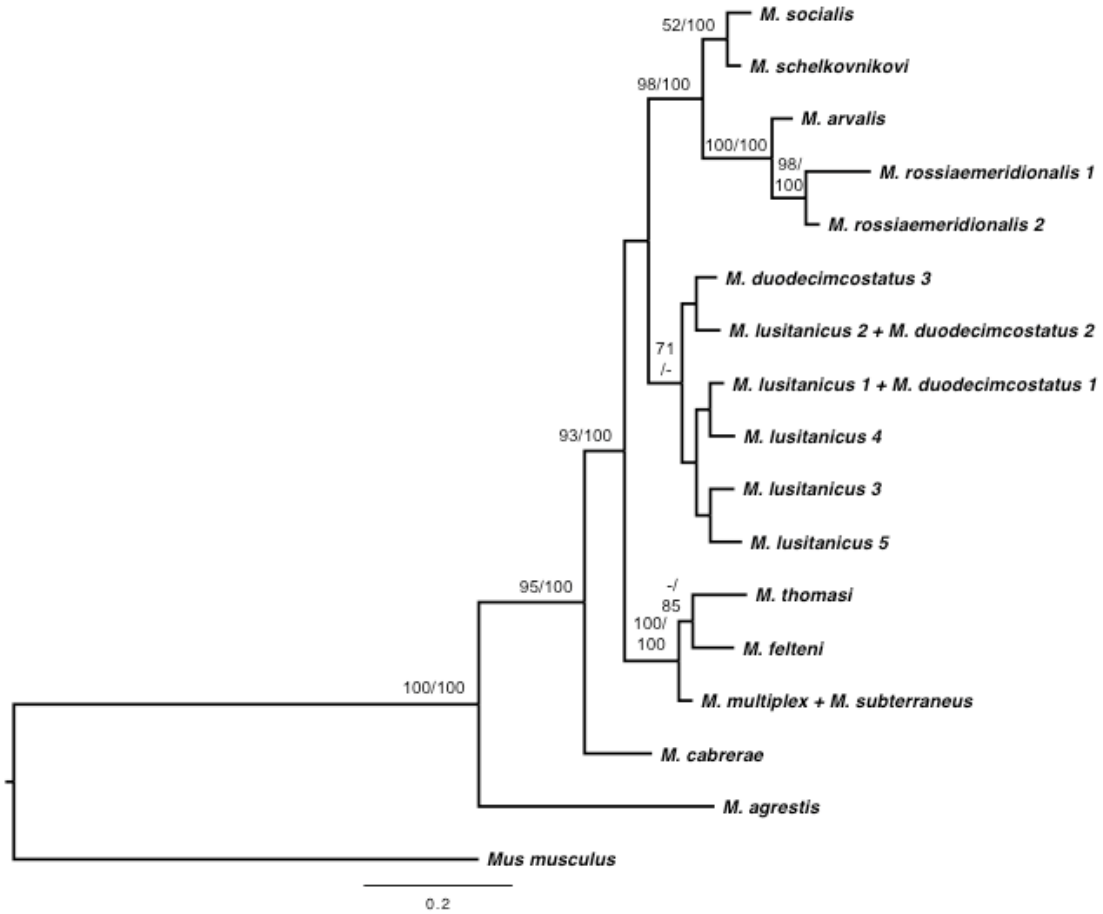


Figure 2 – Bayesian inference phylogenetic tree obtained for the Olfr57 gene fragment. Posterior probability (Bayesian inference) and bootstrap (maximum likelihood) values >50 % are indicated.

We did not find signs of recombination in the analysed Olfr31 and Olfr57 gene fragments. Considering PAML and HyPhy branch-site models, branch-site REL was the only method that indicated a branch under episodic diversifying selection ($p < 0.05$), corresponding to evolutionarily early divergent *M. agrestis* for the Olfr57 fragment gene. For both genes, LRTs of site and branch-site models supported equal substitution rates and ω ratios suggest that the analysed gene fragments are mostly under negative/purifying selection ($\omega < 1$) (**Online Resource 4**). PAML and

HyPhy detected more negatively than positively selected amino acids (**Fig. 3**; **Online Resource 4**). With the *Mus musculus* protein sequence as a reference, models M2, M8 and MEME indicated positive selection for amino acid 145 of Olfr31. For Olfr57, M2, M8 and MEME identified amino acid 220, plus 154 and 227 that were only observed in the M8 model (**Online Resource 4**). Amino acid 145 is located in the fourth transmembrane helix of Olfr31, while residues 154, 220 and 227 of Olfr57 are in the fourth and fifth transmembrane helices and third intracellular loop, respectively (**Fig. 3**). These sites were not significant for Naive Empirical Bayes or Bayes Empirical Bayes analysis. Regarding negatively selected sites in Olfr31, three were indicated by SLAC, seven by REL, 11 by FEL and two by IFEL; however, only amino acids 170 and 180 were common amongst all methods (**Online Resource 4**). Both residues are located in the second extracellular loop, and amino acid 180 is also in motif 3 of the olfactory receptor signature (**Fig. 3**). This signature is composed by five conserved motifs that provide a characteristic fingerprint for olfactory receptors. For Olfr57, one site was indicated by SLAC, 14 by FEL and one by IFEL, with amino acid 241 being detected by all tests (**Online Resource 4**). This residue is located in the sixth transmembrane helix, in motif 4 of the olfactory receptor signature (**Fig. 3**). In the extracellular loops of Olfr31 and Olfr57, only negatively selected sites were detected, whereas on the transmembrane helices, both positively and negatively selected amino acids were revealed (**Fig. 3**). Considering the intracellular loops, only negatively selected sites were identified for Olfr31, whilst for Olfr57, both positively and negatively selected amino acids were found (**Fig. 3**). Nevertheless, we have to consider that these selection tests may have a limited statistical power due to limited size of the DNA sequences analysed (e.g. Yang and dos Reis 2011; Jobling et al. 2014).

A comparison between *Mus musculus* and *Microtus sp.* amino acid sequences revealed a majority of conserved residues between mouse (Muridae) and *Microtus* voles (Cricetidae) (**Online Resource 5**). For Olfr31, only six polymorphic amino acids (with two being *Microtus*-specific), associated to five amino acid sequences, were detected in a total of 117. For Olfr57, we uncovered 18 variable residues

(with 13 being *Microtus*-specific), linked to six amino acid sequences, out of 162 residues (**Online Resource 5**).

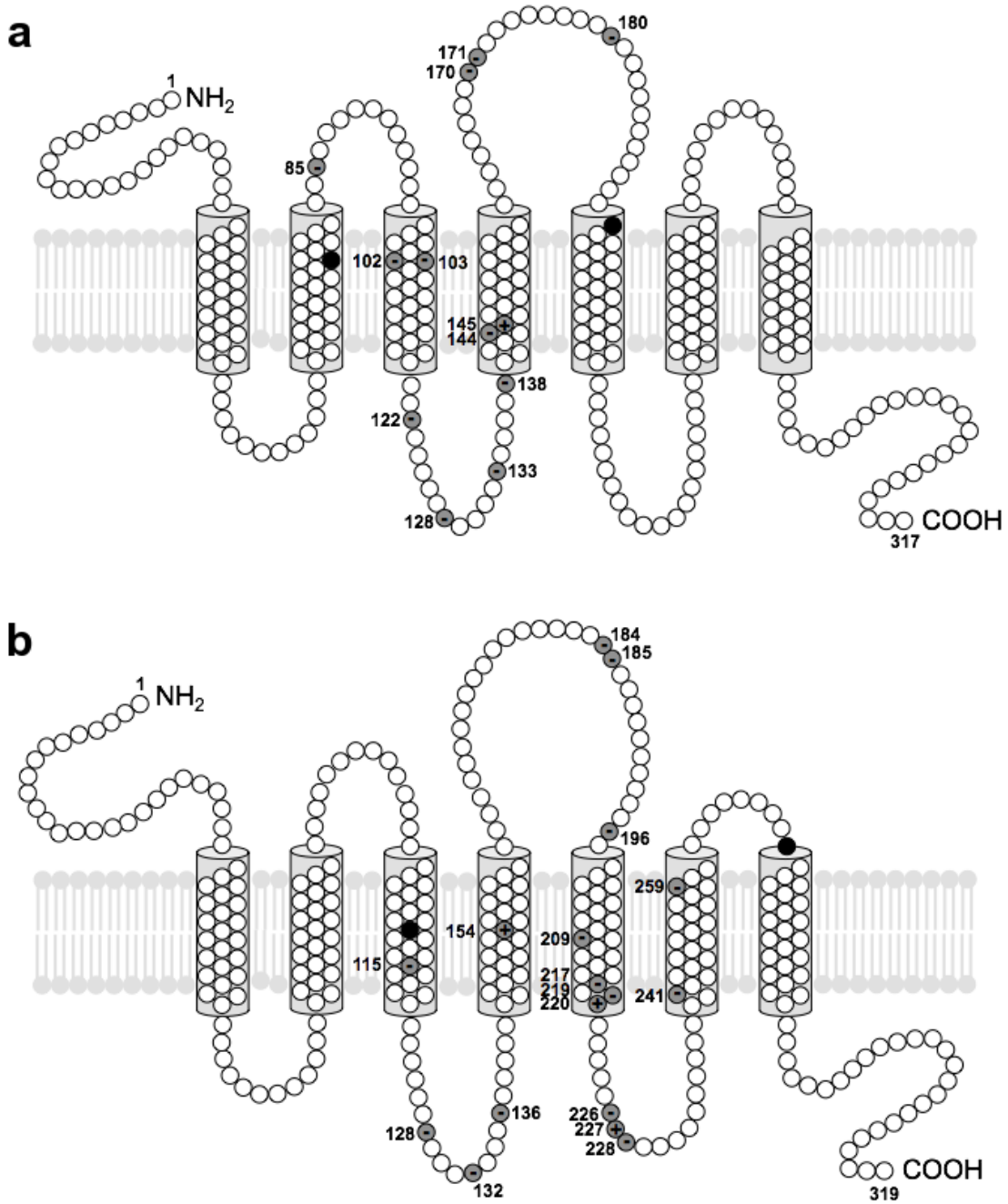


Figure 3 – Schematic amino acid model of Olfr31 (a) and Olfr57 (b) proteins, using *Mus musculus* as reference. Positively and negatively selected amino acids are highlighted as the respective position in the expressed proteins. Beginning and end of the amplified gene fragments (black circle), positively selected amino acid (grey circle with a plus sign), and negatively selected amino acid (grey circle with a minus sign).

All amino acid positions that are polymorphic for Olfr31, beside positively selected 145, and half of Olfr57 polymorphic residues (154—positively selected, 155, 161, 164, 166, 188, 195, 206 and 265) are located at the extracellular loops or extracellular half of the transmembrane helices (**Online Resource 5**).

Contrary to our expectation regarding the location of positive selection in the variable regions responsible for the binding of odour cue molecules (Emes et al. 2004), we only found negatively selected residues in the extracellular loops of Olfr31 and Olfr57, and more negatively selected amino acids than positively selected ones in the transmembrane helices. The present results suggest that Olfr31 and Olfr57 probably recognise conserved odour cues, with very low or inexistent interspecific variation among the analysed *Microtus sp.*.

Our results seem to indicate that Olfr31 and Olfr57 are not related to premating behavioural isolation between *M. lusitanicus* and *M. duodecimcostatus*. Haplotype sharing between these two sister species reduces the chance that sequence polymorphisms in these markers could lead to subtle changes in olfactory perception and influence subsequent specific behaviours. These two markers did also not present species-specific polymorphisms for the other *Microtus* taxa analysed. Considering these results, two hypotheses emerge: (i) Olfr31 and Olfr57 may not contribute to behavioural barriers mediated via odour, and (ii) the expression of Olfr31 and Olfr57 may better reveal the contribution of these receptors than DNA sequence polymorphism data. Thus, considering that hundreds of olfactory receptor genes were detected in the olfactory epithelium of *Mus musculus* (Young et al. 2003), it is pertinent to consider a protein expression approach as the next step. This could be performed in the olfactory epithelium of different *Microtus* taxa, particularly those under ongoing speciation events (e.g. Gileva et al. 2000; Castiglia et al. 2008; Bastos-Silveira et al. 2012; Sutter et al. 2013; Beysard and Heckel 2014). Expression variation of the receptors in the olfactory epithelium under controlled conditions could indicate an interspecific difference of responsiveness of the transduction of chemosignals that are associated with reproductive behaviours, i.e., higher expression levels could

indicate higher sensitivity to small changes in the quantity of odour cues. At present, such analyses are experimentally highly challenging (e.g. Rice et al. 2011; Hohenbrink et al. 2014), particularly if controlled laboratory experiments are combined with ecological testing, but they could provide major insights into the role of olfactory receptors on behavioural isolation.

2.1.6 Acknowledgments

We are grateful to Isabel Rey from the Museo Nacional de Ciencias Naturales (Madrid, Spain) for providing biological samples. We would also like to thank Heidi Lischer for providing the repeated random haplotype sampling analysis software before publication and to Susanne Tellenbach for technical assistance. M. A. Duarte was supported by Ph.D. grant SFRH/BD/70646/2010 from the Foundation for Science and Technology (FCT) and G. Heckel by grant 31003A-149585 from the Swiss National Science Foundation. This study was developed under the project PTDC/BIA-BEC/103729/2008 (FCT), coordinated by C. Bastos-Silveira. Thanks are due, for the financial support to Centro de Estudos de Ambiente e Mar (UID/AMB/50017), to FCT/MEC through national funds and the co-funding by the FEDER, within the PT2020 Partnership Agreement and Compete 2020.

2.1.7 References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723

Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236

Arakawa H, Blanchard DC, Arakawa K, Dunlap C, Blanchard RJ (2008) Scent marking behavior as an odorant communication in mice. Neurosci Biobehav Rev 32(7):1236–1248

Barbosa S, Paupério J, Searle JB, Alves PC (2013) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to

noninvasive sampling. Mol Ecol Resour 13(1):43–56

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

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

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

Beysard M, Krebs-Wheaton R, Heckel G (2015) Tracing reinforcement through asymmetrical partner preference in the European common vole *Microtus arvalis*. BMC Evol Biol 15, e170

Brennan PA, Kendrick KM (2006) Mammalian social odours: attraction and individual recognition. Philos Trans R Soc B 361:2061–2078

Brunet-Lecomte P, Chaline J (1991) Morphological evolution and phylogenetic relationships of the European ground voles (Arvicolidae, Rodentia). Lethaia 24(1):45–53

Castiglia R, Annesi F, Aloise G, Amori G (2008) Systematics of the *Microtus savii* complex (Rodentia, Cricetidae) via mitochondrial DNA analyses: paraphyly and pattern of sex chromosome evolution. Mol Phylogenet Evol 46:1157–1164

Cotilla I, Palomo LJ (2007) *Microtus duodecimcostatus* (de Selys-Longchamps, 1839). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Tragsa, Madrid, pp 422–425

Delport D, Poon AF, Frost SDW, Pond SLK (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26(19):2455–2457

Duarte MA, Mathias ML, Bastos-Silveira C (2015) Pair-bonding behaviour of the sister species *Microtus lusitanicus* and *M. duodecimcostatus*. J Ethol 33:213–223

Emes RD, Beatson SA, Ponting CP, Goodstadt L (2004) Evolution and Comparative Genomics of Odorant- and Pheromone-Associated Genes in Rodents.

Genome Res 14:591–602

Feldmesser E, Olender T, Khen M, Yanai I, Ophir R, Lancet D (2006) Widespread ectopic expression of olfactory receptor genes. BMC Genomics 7:121

FigTree version 1.3.1 http://tree.bio.ed.ac.uk/software/figtree/ Accessed 9 March 2011

Fink S, Excoffier L, Heckel G (2007) High variability and non-neutral evolution of the mammalian arginine vasopressin receptor. BMC Evol Biol 7, e176

Fink S, Fischer MC, Excoffier L, Heckel G (2010) Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. Syst Biol 59(5):548–572

Fischer MC, Foll M, Heckel G, Excoffier L (2014) Continental-scale footprint of balancing and positive selection in a small rodent (*Microtus arvalis*). PLoS ONE 9, e112332

Fleischer J, Breer H, Strotmann J (2009) Mammalian Olfactory Receptors. Front Cell Neurosci 23:9

Gaillard I, Rouquier S, Chavanieu A, Mollard P, Giorgi D (2004) Amino-acid changes acquired during evolution by olfactory receptor 912–93 modify the specificity of odorant recognition. Hum Mol Genet 13(7):771–780

Giannoni SM, Borghi CE, Martinezrica JP (1993) Swimming ability of a fossorial Iberic vole *Microtus* (*Pitymys*) *lusitanicus* (Rodentia, Microtinae). Mammalia 57(3):337–340

Giannoni SM, Borghi CE, Martinezrica JP (1994) Swimming ability of the Mediterranean pine vole *Microtus* (*Terricola*) *duodecimcostatus*. Acta Theriol 39(3):257–265

Gilad Y, Lancet D (2003) Population Differences in the Human Functional Olfactory Repertoire. Mol Biol Evol 20(3):307–314

Gilad Y, Bustamante CD, Lancet D, Pääbo S (2003) Natural Selection on the Olfactory Receptor Gene Family in Humans and Chimpanzees. Am J Hum Genet 73:489–501

Gileva EA, Bol'shakov VN, Polyavina OV, Cheprakov MI (2000) The vole species *Microtus arvalis* and *Microtus rossiaemeridionalis* in the Urals: hybridization in the wild. Dokl Biol Sci 370:47–50

Glusman G, Bahar A, Sharon D, Pilpel Y, White J, Lancet D (2000) The olfactory receptor gene superfamily: data mining, classification, and nomenclature. Mamm Genome 11(11):1016–1023

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11(5):725–36

Hohenbrink P, Dempewolf S, Zimmermann E, Mundy NI, Radespiel U (2014) Functional promiscuity in a mammalian chemosensory system: extensive expression of vomeronasal receptors in the main olfactory epithelium of mouse lemurs. Front Neuroanat 8:102

Hurst JL, Beynon RJ (2004) Scent wars: the chemobiology of competitive signalling in mice. Bioessays 26:1288–1298

Ignatieva EV, Levitsky VG, Yudin NS, Moshkin MP, Kolchanov NA (2014) Genetic basis of olfactory cognition: extremely high level of DNA sequence polymorphism in promoter regions of the human olfactory receptor genes revealed using the 1000 Genomes Project dataset. Front Psychol 5:247

Illing N, Boolay S, Siwoski JS, Casper D, Lucero MT, Roskams AJ (2002) Conditionally Immortalized Clonal Cell Lines from the Mouse Olfactory Placode Differentiate into Olfactory Receptor Neurons. Mol Cell Neurosci 20:225–243

Jaarola M, Martínková N, Gündüz I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedakis-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioğlu H, Mathias ML, Villate I, Searle JB (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol 33(3):647–663

Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C (2014) Human Evolutionary Genetics, 2nd edn. Garland Science, New York Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci U S A 78(1):454–458

Kotenkova EV, Naidenko SV (1999). Discrimination of con- and heterospecific odors in different taxa of the *Mus musculus* species group. In: Johnston RE, Müller-Schwarze D, Sorensen PW (eds) Advances in chemical signals in Vertebrates. Springer, pp 299–308

Li K, Honga W, Jiao H, Wang G-D, Rodriguez KA, Buffenstein R, Zhao Y, Nevo E, Zhao H (2015) Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. Proc Natl Acad Sci U S A 112(38):11905–11910

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

Lischer HE, Excoffier L, Heckel G (2014) Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. Mol Biol Evol 31(4):817–831

Malnic B, Godfrey PA, Buck LB (2004) The human olfactory receptor gene family. Proc Natl Acad Sci U S A 101(8):2584–2589

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26:2462–2463

McGowen MR (2011) Toward the resolution of an explosive radiation – A multilocus phylogeny of oceanic dolphins (Delphinidae). Mol Phylogenet Evol 60:345–357

Mira A, Mathias ML (2007) *Microtus lusitanicus* (Gerbe, 1879). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 418–421

Moore RE (1965) Olfactory Discrimination As an Isolating Mechanism Between *Peromyscus maniculatus* and *Peromyscus polionotus*. Am Midl Nat 73(1):85–100

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK (2012) Detecting Individual Sites Subject to Episodic Diversifying Selection. PLoS Genet 8(7), e1002764

Nevo E, Bodmer M, Heth G (1976) Olfactory discrimination as an isolating mechanism in speciating mole rats. Experientia 32(12):1511–1512

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148(3):929–936

Niimura Y, Nei M (2007) Extensive gains and losses of olfactory receptor genes in mammalian evolution. PLoS ONE 2(8), e708

Pathak N, Johnson P, Getman M, Lane RP (2009) Odorant receptor (OR) gene choice is biased and non-clonal in two olfactory placode cell lines, and OR RNA is nuclear prior to differentiation of these lines. J Neurochem 108(2):486–497

Petrulis A (2013) Chemosignals, Hormones and Mammalian Reproduction. Horm Behav 63(5):723–741

Pillay N, Willan K, Meester J, Cooke J (1995) Evidence of Pre-mating Reproductive Isolation in Two Allopatric Populations of the Vlei Rat, *Otomys irroratus*. Ethol 100:61–71

Pond SLK, Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22(5):1208–1222

Pond SLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21(5):676–679

Pond SLK, Frost SDW, Grossman Z, Gravenor MB, Richman DD, Brown AJL (2006) Adaptation to different human populations by HIV-1 revealed by codon-based analyses. PLoS Comput Biol 2(6), e62

Pond SLK, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K (2011) A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol 28(11):3033–3043

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25(7):1253–1256

Rice AM, Rudh A, Ellegren H, Qvarnstrom A (2011) A guide to the genomics of ecological speciation in natural animal populations. Ecol Lett 14:9–18

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

Rouquier S, Blancher A, Giorgi D (2000) The olfactory receptor gene repertoire

in primates and mouse: evidence for reduction of the functional fraction in primates. Proc Natl Acad Sci U S A 97(6): 2870–2874

Sambrook J, Fritschi EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Santos SML (2009) Factors influencing the distribution of the lusitanian and the mediterranean pine voles (*Microtus lusitanicus* and *M. duodecimcostatus*) in Portugal: a multiscale approach. Dissertation, Faculdade de Ciências, Universidade de Lisboa

Smadja C, Butlin RK (2009) On the scent of speciation: the chemosensory system and its role in premating isolation. Heredity 102:77–97

Smadja C, Ganem G (2008) Divergence of odorant signals within and between the two European subspecies of the house mouse. Behav Ecol 19(1):223–230

Soares JMA (2013) Fitness dos híbridos entre duas espécies de ratos fossadores, *Microtus lusitanicus* (Gerbe, 1879) e *M. duodecimcostatus* (de Selys-Longchamps, 1839) - uma abordagem multidisciplinary. Dissertation, Faculdade de Ciências da Universidade de Lisboa

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312– 1313

Stippel BG (2009) Reproductive isolation in four populations of the striped mouse *Rhabdomys*. Dissertation, Faculty of Science, University of the Witwatersrand

Sutter A, Beysard M, Heckel G (2013) Sex-specific clines support incipient speciation in a common European mammal. Heredity 110:398–404

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3):512–526

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Theiler GR, Blanco A (1996) Patterns of evolution in *Graomys griseoflavus*

(Rodentia, Muridae). III. Olfactory discrimination as a premating isolation mechanism between cytotypes. J Exp Zool 274:346–350

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

Yang Z (2007) PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, dos Reis M (2011) Statistical properties of the branch-site test of positive selection. Mol Biol Evol 28(3):1217–1228

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418

Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Young JM, Shykind BM, Lane RP, Tonnes-Priddy L, Ross JA, Walker M, Williams EM, Trask BJ (2003) Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels. Genome Biol 4(11):R71

Zhang X, Rogers M, Tian H, Zhang X, Zou DJ, Liu J, Ma M, Shepherd GM, Firestein SJ (2004) High-throughput microarray detection of olfactory receptor gene expression in the mouse. Proc Natl Acad Sci U S A 101(39):14168–14173

Zhuang H, Chien M-S, Matsunami H (2009) Dynamic functional evolution of an odorant receptor for sex-steroid-derived odors in primates. Proc Natl Acad Sci U S A 106(50):21247–21251

2.1.8 Supplementary material

Online Resource 1 – *Microtus* species list of samples, subgenus taxonomy, sample size and place of origin. * = samples made available by the Museo Nacional de Ciencias Naturales, (A) = allopatry, (S) = sympatry

Taxonomic classification	Subgenus	Sample size	Origin
M. lusitanicus	*Terricola*	24	Alcobaça, Leiria, Portugal (A) Alijó, Vila Real, Portugal (A) Ervedosa, Bragança, Portugal (A) Germil, Viana do Castelo, Portugal (A) Nogueira, Braga, Portugal (A) Ponte Velha, Portalegre, Portugal (S) Rebordelo, Bragança, Portugal (A) Rio Longo, Braga, Portugal (A) Senhorim, Viseu, Portugal (A) Vale do Peso CP, Portalegre, Portugal (S) Vale Vaqueiros, Portalegre, Portugal (S)
M. duodecimcostatus	*Terricola*	28	Carregueiro, Beja, Portugal (A) Corte Velha, Faro, Portugal (A) Guerreiros do Rio, Faro, Portugal (A) Medelim, Castelo Branco, Portugal (S) Monte Ruas, Beja, Portugal (A) Palmeira, Faro, Portugal (A) Pomar jovem, Faro, Portugal (A) Portalegre, Portugal (S) Porto de Lagos/Portimão, Faro, Portugal (A) Ribeira da Foupana, Faro, Portugal (A) Ribeira, Beja, Portugal (A) Santa Marta, Faro, Portugal (A) Silveira I, Portalegre, Portugal (S) Travesso, Portalegre, Portugal (S)
M. agrestis*	*Agricola*	2	Ardaiz, Spain
M. arvalis	*Microtus*	2	Laa an der Thaya, Niederösterreich, Austria
M. cabrerae	*Iberomys*	5	Grândola, Portugal
M. felteni	*Terricola*	1	Vasilitsa, Thessaly, Greece
M. gerbei*	*Terricola*	1	Sorogain, Spain
M. multiplex	*Terricola*	2	Piotta, Switzerland Molare, Ticino, Switzerland
M. rossiaemeridionalis = M. levis	*Microtus*	2	Tar lake, Tehran province, Iran Gradsko, Macedonia
M. schelkovnikovi	*Hyrcanicola*	1	Talysh, Azerbaijan
M. socialis	*Microtus*	2	Tar lake, Tehran province, Iran Stepanakert, Azerbaijan
M. subterraneus	*Terricola*	2	Gurnigel, Switzerland Bretolet, Switzerland
M. tatricus	*Terricola*	2	Tretie Rohacske pleso lake, High Tatra Mountains, Slovakia
M. thomasi	*Terricola*	2	Arkadia, Peloponnes, Greece Kyllini, Peloponnes, Greece

Online Resource 2 – Fragment size, GC content, number of species and DNA polymorphism results, per olfactory receptor gene, for different sets of species.

Parameters	Olfr31			Olfr57		
	ML	**MD**	**European *Microtus***	**ML**	**MD**	**European *Microtus***
Fragment size (bp)	352			488		
%GC	43.8%±0.2%			52.6±0.2%		
Number of species	14			12		
Number of individuals	25	26	76	24	23	70
Number of unphased haplotypes	11	8	31	6	3	16
Number of variable sites	6	4	28	6	2	36
Number of parsimony informative sites	3	1	19	0	0	21
Nucleotide diversity	0.006	0.005	0.014	0.004	0.003	0.016

Online Resource 3 – Mean pairwise percentage of genetic divergence for Olfr31 (below diagonal) and Olfr57 (above diagonal), per pair of *Microtus* species.

	Taxonomic classification	**Olfr57**													
		Mlus	Mduo	Magr	Marv	Mcab	Mfel	Mger	Mmul	Mros	Msch	Msoc	Msub	Mtat	Mtho
Olfr31	*M. lusitanicus*	-	0.2	3.6	1.0	1.1	1.0	-	0.7	1.1	0.7	0.8	0.7	-	1.1
	M. duodecimcostatus	0.5	-	3.5	1.0	1.1	1.0	-	0.7	1.1	0.7	0.8	0.7	-	1.1
	M. agrestis	1.2	1.1	-	4.2	3.7	4.0	-	3.7	4.6	3.7	3.6	3.7	-	3.7
	M. arvalis	1.4	1.4	0.8	-	2.0	1.9	-	1.6	0.3	0.7	0.8	1.6	-	2.0
	M. cabrerae	2.6	2.5	2.5	2.7	-	1.6	-	1.2	2.1	1.2	1.3	1.2	-	1.7
	M. felteni	0.3	0.3	0.9	1.1	2.2	-	-	0.3	2.0	1.1	1.2	0.3	-	0.7
	M. gerbei	1.9	1.9	1.6	1.8	3.5	1.6	-	-	-	-	-	-	-	-
	M. multiplex	0.6	0.5	1.2	1.4	2.5	0.3	1.9	-	1.7	0.8	0.9	0	-	0.4
	M. rossiaemeridionalis	1.7	1.6	1.0	0.5	2.8	1.3	2.0	1.6	-	0.8	0.9	1.7	-	2.1
	M. schelkovnikovi	1.2	1.1	1.2	0.9	2.5	0.9	2.2	1.2	1.0	-	0.1	0.8	-	1.2
	M. socialis	1.2	1.1	1.2	0.9	2.5	0.9	2.2	1.2	1.0	0	-	0.9	-	1.1
	M. subterraneus	0.7	0.6	0.6	0.9	2.0	0.4	1.7	0.7	1.1	0.6	0.6	-	-	0.4
	M. tatricus	0.6	0.6	1.0	1.2	2.3	0.3	1.7	0.4	1.4	1.0	1.0	0.5	-	-
	M. thomasi	0.5	0.4	1.0	1.2	2.3	0.1	1.7	0.4	1.4	1.0	1.0	0.4	0.3	-

Online Resource 4 – Results of the Olfr31 and Olfr57 gene fragments selection tests. For the PAML results, the log likelihood *l* of each model is given as well as the *Mus musculus* position of positively selected codons (where ω> 1) calculated by the Naive Empirical Bayes analysis and Bayes Empirical Bayes analysis. For the DataMonkey results both positively and negatively selected sites are shown, also having *Mus musculus* as a reference. Sites selected by all the tests performed by PAML or DataMonkey (independently) are underlined. *p0*=proportion of sites where ω<1 (ω0), *p1* = proportion of sites where ω=1 (ω1), *p2* = proportion of sites where ω>1 (ω2), p/q = parameters of the beta distribution, * = probability>95%, ** = probability>99%, - = not applied.

Gene	Software	Model or test	Parameters	κ (ts/tv)	Likelihood *l*	Positively selected sites	Negatively selected sites
Olfr31	**PAML**	M0: one ratio	ω = 0.06345	3.741	-722.258929	Not allowed	-
		M1: nearly neutral	*p0*= 0.96901 ω0 = 0.03689 *p1*= 0.03099 ω1 = 1.00000	3.755	-721.471691		
		M2: selection	*p0*= 0.96901 ω0 = 0.03689 *p1*= 0.01045 ω1 = 1.00000 *p2*= 0.02054 ω2 = 1.00000	3.755	-721.471637	<u>145</u>	
		M3: discrete	*p0*= 0.51347 ω0 = 0.00000 *p1*= 0.28792 ω1 = 0.00000 *p2*= 0.19861 ω2 = 0.32076	3.744	-721.381694	None	
		M7: beta	p = 0.11093 q = 1.48749	3.745	-721.393241	Not allowed	

Online Resource 4 (continued)

Olfr31	**PAML**	M8: beta and ω	*p0*= 0.99999 p = 0.11095 q = 1.48789 *p1*= 0.00001 ω = 1.00000	3.745	-721.393254	145	-
		M0 vs. M3	2Δl=1.755 df=4 p=0.781	-	-	-	
		M1 vs. M2	$2\Delta l=1.1x10^{-4}$ df=2 p=0.999	-	-	-	
		M7 vs. M8	$2\Delta l=2.6x10^{-5}$ df=2 p=0.999	-	-	-	
		MA1, foreground branch = *Terricola*	-	5.908	-572.667032	None	
		MA1, foreground branch = *Microtus* (+*M. schelkovnikovi*)	-	5.908	-572.667032	None	
		MA1, foreground branch = *Iberomys*	-	5.907	-572.666008	None	
		MA1, foreground branch = *Agricola*	-	5.908	-572.667032	None	
		MA1, foreground branch = *M. lusitanicus*	-	5.908	-572.667032	None	
		MA1, foreground branch = *M. duodecimcostatus*	-	5.907	-572.666008	None	
	HyPhy	SLAC	-	-	-	None	122 170* 180*
		REL	-	-	-	None	85** 102** 122** 133** 138** 170** 180**

Online Resource 4 (continued)

Olfr31	**HyPhy**	FEL	-	-	-	None	85 102* 103 122* 128 133* 138 144 170** 171* 180*
		IFEL	-	-	-	None	170 180
		MEME	-	-	-	145	-
Olfr57	**PAML**	M0: one ratio	ω = 0.15760	3.450	-1033.108884	Not allowed	-
		M1: nearly neutral	*p0*= 0.88337 ω0 = 0.06930 *p1*= 0.11663 ω1 = 1.00000	3.522	-1031.666984		
		M2: selection	*p0*= 0.88337 ω0 = 0.06930 *p1*= 0.04844 ω1 = 1.00000 *p2*= 0.06818 ω2 = 1.00000	3.522	-1031.666984	220	
		M3: discrete	*p0*= 0.67602 ω0 = 0.00000 *p1*= 0.10126 ω1 = 0.52164 *p2*= 0.22272 ω2 = 0.52164	3.501	-1031.410893	None	

Online Resource 4 (continued)

Olfr57	**PAML**	M7: beta	p = 0.18374 q = 0.88589	3.502	-1031.502453	Not allowed	-
		M8: beta and ω	*p0*= 0.99999 p = 0.18374 q = 0.88594 *p1*= 0.00001 ω = 1.00000	3.502	-1031.502457	154 220 227	
		M0 vs. M3	2Δl=3.396 df=4 p=0.494	-	-	-	
		M1 vs. M2	2Δl=0 df=2 p=1.00	-	-	-	
		M7 vs. M8	$2\Delta l=8.0\times10^{-6}$ df=2 p=0.999	-	-	-	
		MA1, foreground branch = *Terricola*	-	3.335	-936.672879	None	
		MA1, foreground branch = *Microtus*	-	3.335	-936.672922	None	
		MA1, foreground branch = *Microtus* (+*M. schelkovnikovi*)	-	3.335	-936.672879	None	
		MA1, foreground branch = *Hyrcanicola*	-	3.335	-936.672879	None	
		MA1, foreground branch = *Iberomys*	-	3.330	-936.639514	None	
		MA1, foreground branch = *Agricola*	-	3.335	-936.672879	None	
		MA1, foreground branch = *M. lusitanicus*	-	3.335	-936.672877	None	
		MA1, foreground branch = *M. duodecimcostatus*	-	3.335	-936.672878	None	
	HyPhy	SLAC	-	-	-	None	241*
		REL	-	-	-	None	None

Online Resource 4 (continued)

Olfr57	**HyPhy**	FEL	-	-	-	None	115 128* 132 136 184 185 196 209 217 219 226* 228 241** 259
		IFEL	-	-	-	None	241*
		MEME	-	-	-	220**	-

Olfr31 null model vs. MA1 and M1a vs. MA1:

Terricola: $2\Delta l=0$, df=2, p>0.999 and $2\Delta l=2.9\times10^{-4}$, df=2, p>0.999;

Microtus+Hyrcanicola: $2\Delta l=0$, df=2, p>0.999 and $2\Delta l=2.9\times10^{-4}$, df=2, p>0.999;

Iberomys: $2\Delta l=2.0\times10^{-3}$, df=2, p=0.999 and $2\Delta l=2.3\times10^{-3}$, df=2, p=0.999;

Agricola: $2\Delta l=-1.9\times10^{-3}$, df=2, p=0.999 and $2\Delta l=2.9\times10^{-4}$, df=2, p>0.999;

ML: $2\Delta l=-1.9\times10^{-3}$, df=2, p>0.999 and $2\Delta l=2.9\times10^{-4}$, df=2, p>0.999;

MD: $2\Delta l=2.0\times10^{-3}$, df=2, p=0.999 and $2\Delta l=2.3\times10^{-3}$, df=2, p=0.999.

Olfr57 null model vs. MA1 and M1a vs. MA1:

Terricola: $2\Delta l=0$, df=2, p>0.999 and $2\Delta l=0$, df=2, p>0.999;

Microtus: $2\Delta l=-8.6\times10^{-5}$, df=2, p>0.999 and $2\Delta l=-8.6\times10^{-5}$, df=2, p>0.999;

Microtus+Hyrcanicola: $2\Delta l=0$, df=2, p>0.999 and $2\Delta l=0$, df=2, p>0.999;

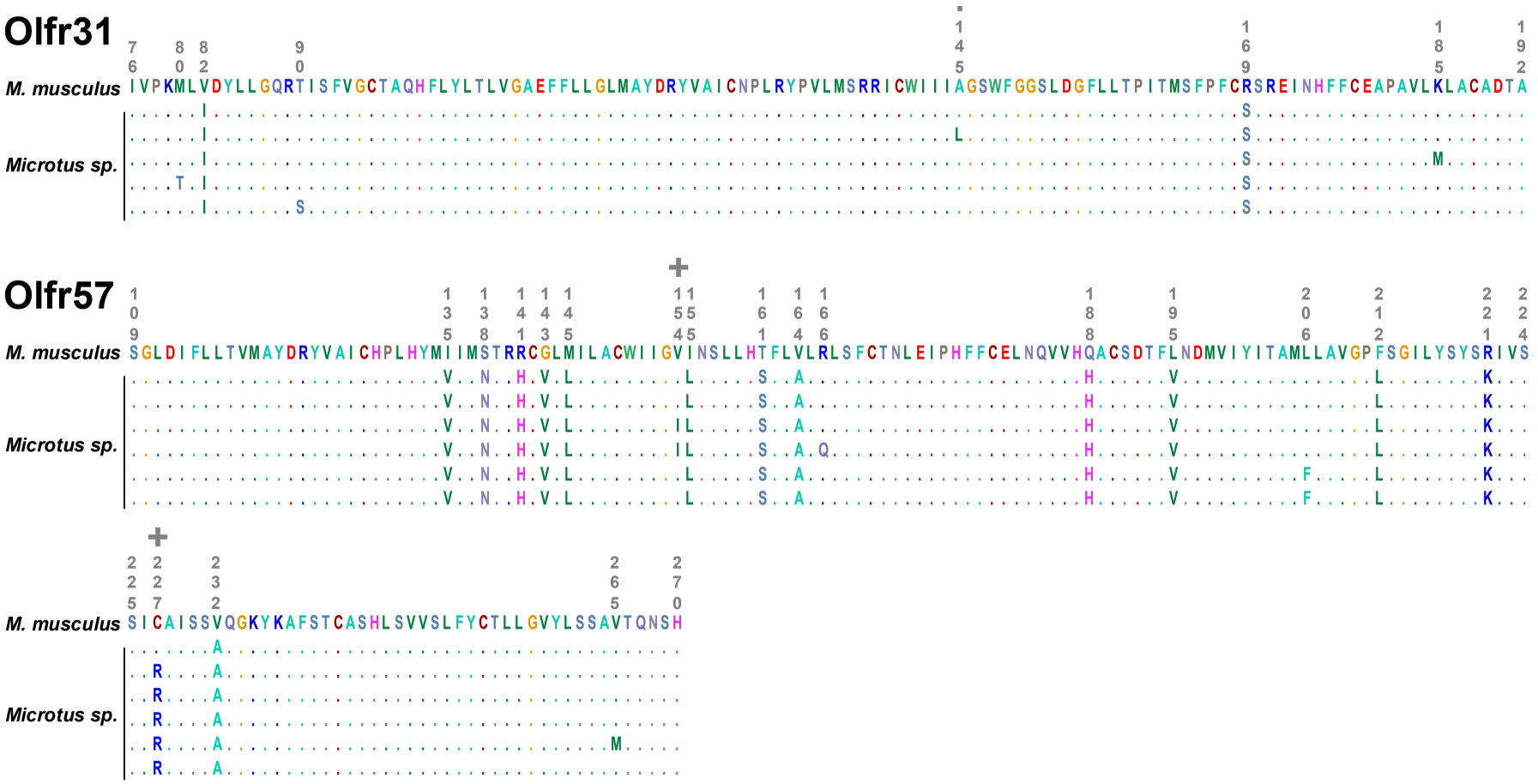
Hyrcanicola: $2\Delta l=0$, df=2, p>0.999 and $2\Delta l=0$, df=2, p>0.999;

Iberomys: $2\Delta l=2.3\times10^{-4}$, df=2, p>0.999 and $2\Delta l=6.6\times10^{-2}$, df=2, p=0.968;

Agricola: $2\Delta l=1.6\times10^{-5}$, df=2, p>0.999 and $2\Delta l=0$, df=2, p>0.999;

ML: $2\Delta l=2.0\times10^{-6}$, df=2, p>0.999 and $2\Delta l=4.0\times10^{-6}$, df=2, p>0.999;

MD: $2\Delta l=4.0\times10^{-6}$, df=2, p>0.999 and $2\Delta l=4.0\times10^{-6}$, df=2, p>0.999.



Online Resource 5 – Comparison of *Mus musculus* and *Microtus sp.* amino acid sequences. Position numbers are provided for the first, last and polymorphic amino acids. Residues that are positively selected are marked with a plus sign.

2.2 Major histocompatibility complex I and II

Type of publication: Short communication

Reference: Duarte MA[1,2,3], Heckel G[4,5], Mathias ML[2,3], Bastos-Silveira C[1] (in preparation) Preliminary data on exon 2 of MHCI and MHCII from different *Microtus* subgenera.

[1] Museu Nacional de História Natural e da Ciência, Departamento de Zoologia e Antropologia, Universidade de Lisboa, Rua da Escola Politécnica, 58, Lisbon 1250-102, Portugal.

[2] Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[3] Centro de Estudos de Ambiente e Mar, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[4] Computational and Molecular Population Genetics, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland.

[5] Swiss Institute of Bioinformatics, Genopode, Lausanne, Switzerland.

2.2.1 Abstract

Major histocompatibility class I and II seem to play a role in intraspecific communication, particularly in mate choice. One of the most important genic regions is exon 2, which is translated into the hyper variable antigen-binding domain of these glycoproteins. In the present study we analysed, for the first time, a wide range of *Microtus* species, from different subgenera. This is a speciose genus with a recent and prolific diversification and mate choice is an important piece on reproductive isolation and speciation of most taxa. The results obtained suggest that *Microtus* probably express more than one MHC locus, similarly to other Arvicolids.

2.2.2 Keywords

Behavioural isolation; odour; MHCI; MHCII; *Microtus* voles.

2.2.3 Introduction

Highly polymorphic genes, such as major histocompatibility I (MHCI) and II (MHCII), are sources of individuality chemosignals. They are communicated through odour cues, since olfaction is the dominant sense in most mammals and heavily influences behaviour (Wyatt, 2003). There is solid evidence that MHC genotypes influence the urine odour of mice (Yamazaki et al., 1979; Yamaguchi et al., 1981; Yamazaki et al., 1990) and rats (Brown et al., 1987). Nevertheless, urinary peptide composition of mice revealed that MHC-derived peptides are present in extremely low concentrations (reviewed in Overath et al., 2014), conversely to MUPs (reviewed in Liberles, 2014).

Concerning mate choice, it is known that mice generally choose to mate with MHC-dissimilar individuals (Jordan & Bruford, 1998), whether in parent–offspring interactions it is important to reveal a MHC-similar genotype between family members (Manning et al., 1992). Moreover, evidence for MHC-disassortative mate choice has been found for *Mus musculus* (Egid & Brown, 1989; Eklund, 1997; Penn & Potts, 1998). It has been positively associated to certain MHCII genotypes, in different mammalian taxa (Ditchkoff et al., 2001; Sauermann et al., 2001); while in others no evidence for MHC-dependent mating preferences was discovered (Beauchamp et al., 1988; Eklund et al., 1991; Manning et al., 1992; Eklund, 1997; Paterson & Pemberton, 1997; Wenink et al., 1998).

Different models describe how a MHC genotype is translated into an odourtype, ranging from the degradation molecules of MHC to volatiles produced by microflora specific for a MHC genotype (Penn & Potts, 1998; Beauchamp & Yamazaki, 2003). Boehm and Zufall (2006) proposed that non-volatile MHCI peptide ligands are an ideal mechanism for signalling MHC genotype. The same binding properties of peptide ligands that are important for transmitting individual identity in the immune system could be appointed by the olfactory system to communicate genetic individuality (Boehm & Zufall, 2006).

In mice, up to 100 alleles are estimated for H-2K and H-2D loci of MHCI (Singh, 2001), and similar values are estimated for other mammalian taxa (Brennan &

Kendrick, 2006). The ability of mice to discriminate MHCI urine odours is related to the degree of amino acid divergence in peptide-binding cleft (Carroll et al. 2002). Peptides 9-20 amino acid long can be recognized by binding/anchor residues, which are recognized by receptors in the same way that MHCI glycoproteins do (**Figure 1**). Brennan and Kendrick (2006) postulated that main olfactory and vomeronasal receptor neurons are able to respond and recognize different to MHCI peptide ligands (**Figure 1**), a mechanism that may contribute to MHC-dependent mate choice in mice.

MHC peptide ligands themselves can function as chemosignals, forming a direct link between individuality at the immunological and behavioural levels (**Figure 1**). Receptor systems with similar binding features as MHCI, e.g. vomeronasal and olfactory receptors, will be activated specifically by particular MHCI peptides (**Figure 1**). Moreover, odorous communication through MHCI peptide ligands seems to require direct physical contact to be detected (Kelliher et al., 2007). What remains to be answered is: does this mechanism also occurs in other mammalian species?

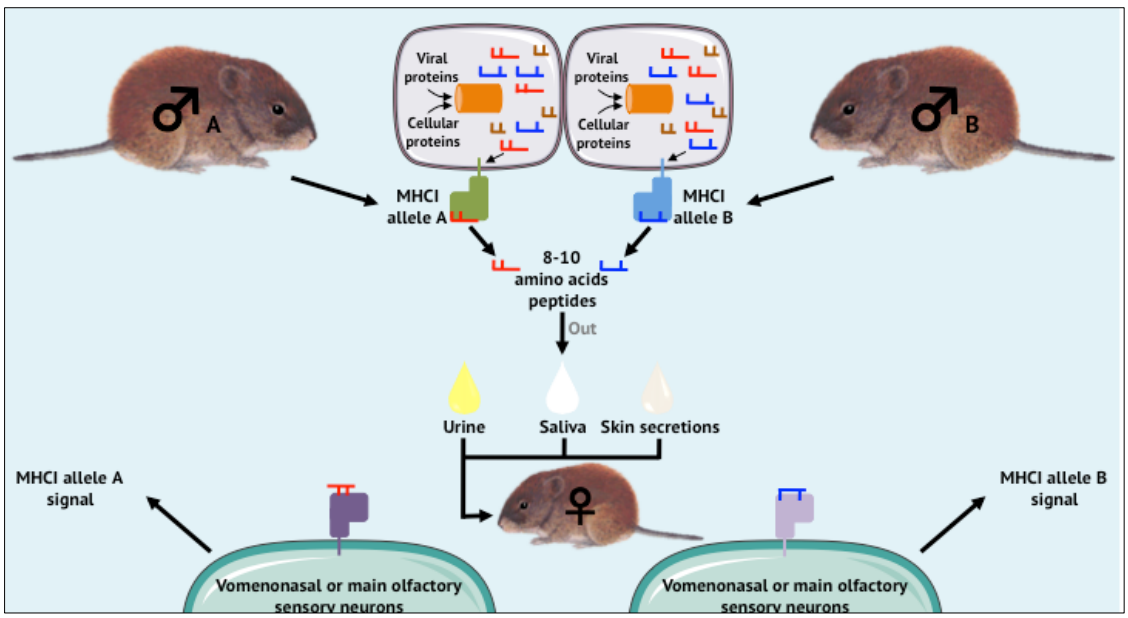


Figure 1 – MHCI peptide ligands communication through biological fluids (based on Brennan & Kendrick, 2006).

In the present study, we analysed for the first time MHCI and MHCII allelic variability and divergence in the speciose *Microtus* Schrank (1798), an evolutionarily young genus that started to radiate 1.2-2 million years ago (Chaline et al., 1999). Molecular variance could indicate a putative role of MHC in mate choice and behavioural pre-mating reproductive isolation on *Microtus* voles, particularly on recently diverged sister species *M. lusitanicus* Gerbe (1879) and *M. duodecimcostatus* de Selys-Longchamps (1839), two of the most recent taxa (Brunet-Lecomte & Chaline, 1991).

2.2.4 Materials and methods

Eighty-four tissue samples from 24 taxa (20 *Microtus sp*., 3 Arvicolinae and 1 Murinae) were stored in absolute ethanol at −20 °C (**Table 1**). Genomic DNA was isolated using a phenol-chloroform extraction procedure (Sambrook et al., 1989).

The number of current *Microtus* subgenera is variable, depending on different authors, but in this dissertation we consider eleven: *Microtus*; *Terricola* Fatio, 1867; *Agricola* Blasius, 1857; *Iberomys* Chaline, 1972; *Alexandromys/Pallasiinus* Ognev, 1914 and Kretzoi, 1964; *Mynomes* Rafinesque, 1817; *Pitymys* McMurtrie, 1831; *Stenocranius* Kastschenko, 1901; *Aulacomys* Rhoads, 1894; *Pedomys* Baird, 1957 and *Hyrcanicola* Nadachowski, 2007 (Jaarola et al., 2004; Wilson & Reeder, 2005; Nadachowski, 2007; Garrido-García & Soriguer-Escofet, 2012).

Exon 2 of the RT1.Ba locus of *Rattus sp*. Fischer de Waldheim, 1803 (DQA homologue) (MHCII) was amplified using published primers (Seddon & Baverstock, 1998). Exon 2 and 3 of H-2K and H-2D loci (MHCI) were amplified using published (Crew et al., 1991; Cao et al., 2003) and newly designed primers M-MHCI-F1 (5'-CAYTCGMTGMGGTATTTC-3'), M-MHCI-F2 (5'-TGTCCCGGCCCGGCCT-3'), M-MHCI-F3 (5'-CAGGCTCYCACACCATCCAG-3'), M-MHCI-R1 (5'-CCAGCTGAGGGTTTCTTCTT-3'), M-MHCI-R2 (5'-ACCCGCGCCCCACGACCC-3'), M-MHCI-R3 (5'-GGAACAGCCCAGTCCC-GAGGCCAC-3'), M-MHCI-R4 (5'-ACCACCTGCGCCTTCTCCG-3'), M-MHCI-R5 (5'-ACCT-GTTCGGCCCCGGGGTC-3'), M-MHCI-R6 (5'-CCAACCCAGTACCTGTGCGC-3'), M-MHCI-R7 (5'-CCTCGCACCTGTGCGC-3'), M-MHCI-R8 (5'-ATGGCCCCGCACCTGTGCGC-3'), M-

MHCI-R9 (5'-TGCGGTCCTGCACCTGTGCGC-3'). Forward primers were designed at the beginning of exon 2 and reverse primers at the beginning of intron 3, based on GenBank sequences from *Mus musculus* Linnaeus, 1758, *Peromyscus sp.* Gloger, 1841 and *Mesocricetus auratus* Waterhouse, 1839 (L29190, M12381, M14825, M18523, M18524, M18525, M23444, M36949, M36950, V00746, X01652, X01815, X03122, X14091). We particularly focused on the exon 2 of MHCI and MHCII since it is translated into the hyper variable antigen-binding domain of these glycoproteins.

Table 1 – List of species analysed for the MHCII study. Sample size and source are also indicated. MNCN = Museo Nacional de Ciencias Naturales, Spain; CMPG = Institute of Ecology and Evolution, Switzerland; MUHNAC = Museu Nacional de História Natural e da Ciência, Portugal.

Species	Sample size	Source
Arvicola sapidus	2	MUHNAC
Lasiopodomys brandti	1	CMPG
Microtus agrestis	2	MNCN
Microtus arvalis	2	CMPG
Microtus cabrerae	4	MUHNAC
Microtus californicus	1	CMPG
Microtus duodecimcostatus	21	MUHNAC
Microtus felteni	1	CMPG
Microtus gerbei	1	MNCN
Microtus kikuchii	1	CMPG
Microtus lusitanicus	25	MUHNAC
Microtus montanus	3	CMPG
Microtus montebelli	1	CMPG
Microtus multiplex	2	CMPG
Microtus ochrogaster	1	CMPG
Microtus oeconomus	2	CMPG
Microtus rossiaemeridionalis	2	CMPG
Microtus schelkovnikovi	1	CMPG
Microtus socialis	2	CMPG
Microtus subterraneus	2	CMPG
Microtus tatricus	2	CMPG
Microtus thomasi	2	CMPG
Mus musculus	2	MUHNAC
Myodes glareolus	1	MNCN

All MHCII reactions contained 100ng of template DNA, 0.3mM of each primer, 1.25U of GoTaq® Flexi DNA Polymerase (Promega), 1x buffer (Promega), 2.5mM

MgCl2, 0.1ug of BSA (New England Biolabs) and 0.2mM of each dNTP (Thermo Scientific), to a final volume of 25µl. DNA amplifications were performed in a MyCycler thermal cycler (Bio-Rad Laboratories Inc.) and consisted of a denaturation at 95°C for 5', followed by 35 cycles of denaturation at 94°C for 1', annealing at 60ºC for 1' and extension at 72ºC for 1'. An extension step at 72 ºC for 10' was added at the end of the reaction. MHCI reactions and amplification conditions were similar to the described above, with the only variable being the annealing temperature (50-66ºC), depending on the primer pairs and species.

PCR products were verified on 1 % agarose gels and purified using ExoI/FastAP protocol (Fermentas). Sequencing using the amplification primers was carried out by Macrogen Inc. (South Korea and the Netherlands) and at the Institute of Ecology and Evolution, University of Bern, using ABI Prism® 3130 Genetic Analyzer (Applied Biosystems).

Sequences were aligned using Sequencher 4.8 (Gene Codes Corporation) and BioEdit 7.2.5 (Hall, 1999). Unphased MHCII sequences were collapsed into haplotypes using DNAcollapser (Villesen, 2007). JModelTest 0.0.1 (Posada, 2008) was used to select the best-fitting model of nucleotide substitution (Total matrix: TIM1+I+G, Posada, 2003; *M. lusitanicus/M. duodecimcostatus* matrix: TPM2+G, Kimura, 1981; Posada, 2008), based on the Akaike information criterion (Akaike, 1974). A recent approach to integrate heterozygous information in existing phylogenetic programs by repeated random haplotype sampling was applied in the present study (Lischer et al., 2014). MHCII alignment was subjected to n=10.000 replicates for the Maximum Likelihood analysis (RAxML) (Stamatakis, 2014); and n=10 replicates, nchains=4, ngen=1.000.000 and mcmc burn-in=250.000 for the Bayesian inference analysis (MrBayes) (Ronquist & Huelsenbeck, 2003). The outgroup chosen was *Mus musculus*. Consensus trees were edited using FigTree version 1.3.1 (Rambaut, 2010).

2.2.5 Results and discussion

The successfulness of amplification was higher for MHCII, than MHCI. For the latter, unsuccessful amplifications for some samples were probably related to DNA sequence variation between *M. lusitanicus* and *M. duodecimcostatus* and the available reference taxa used to design new primers. Primer pair M-MHCI-F1/M-MHCI-R1 combined with a nested PCR using M-MHCI-F1/M-MHCI-R6 presented the best amplification results, resulting in an 818bp fragment corresponding to partial exon 2, intron 2, exon 3 and partial intron 3. For MHCII, 18 and 16 haplotypes were generated for *M. lusitanicus* and *M. duodecimcostatus*, respectively, being constituted of 256bp and comprising only exon 2 (**Figure 2-3**).

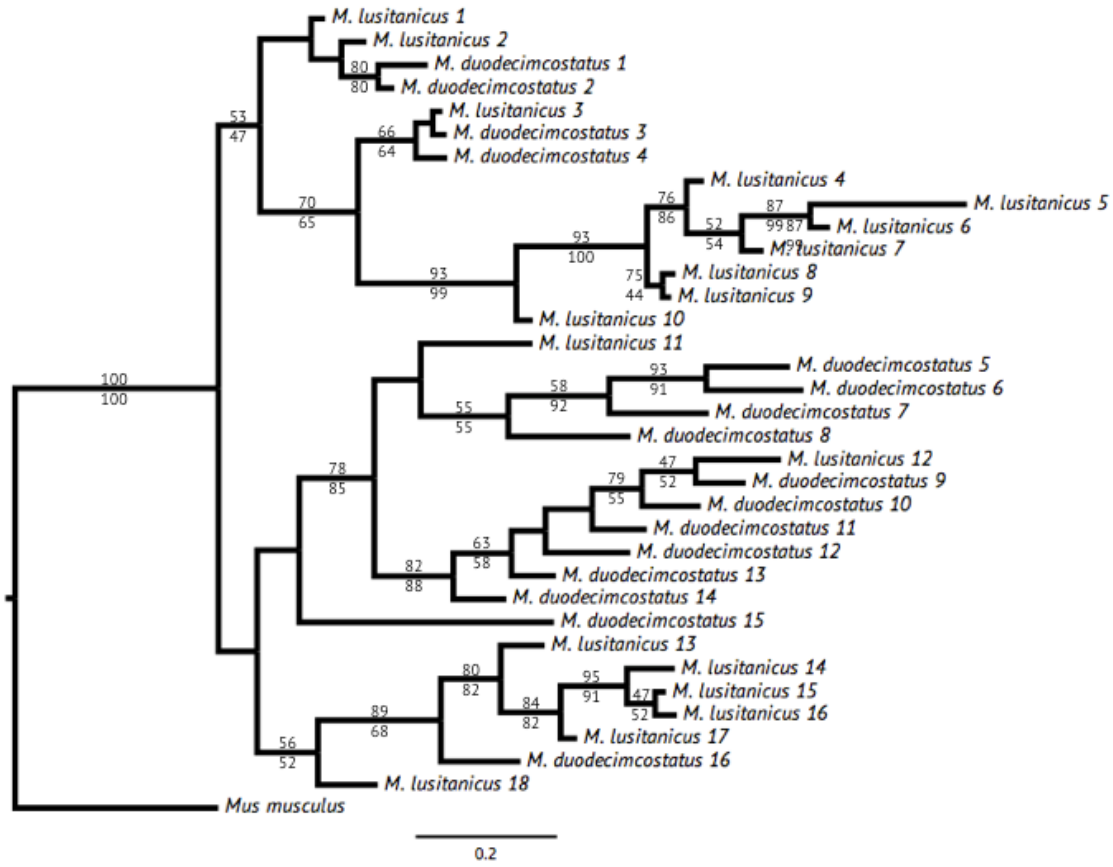


Figure 2– Maximum likelihood phylogenetic tree for *M. lusitanicus* and *M. duodecimcostatus*, considering MHC class II exon 2. RaxML bootstrap values (above) and MrBayes posterior probability values (below) ⩾50% are presented at the respective phylogenetic split.

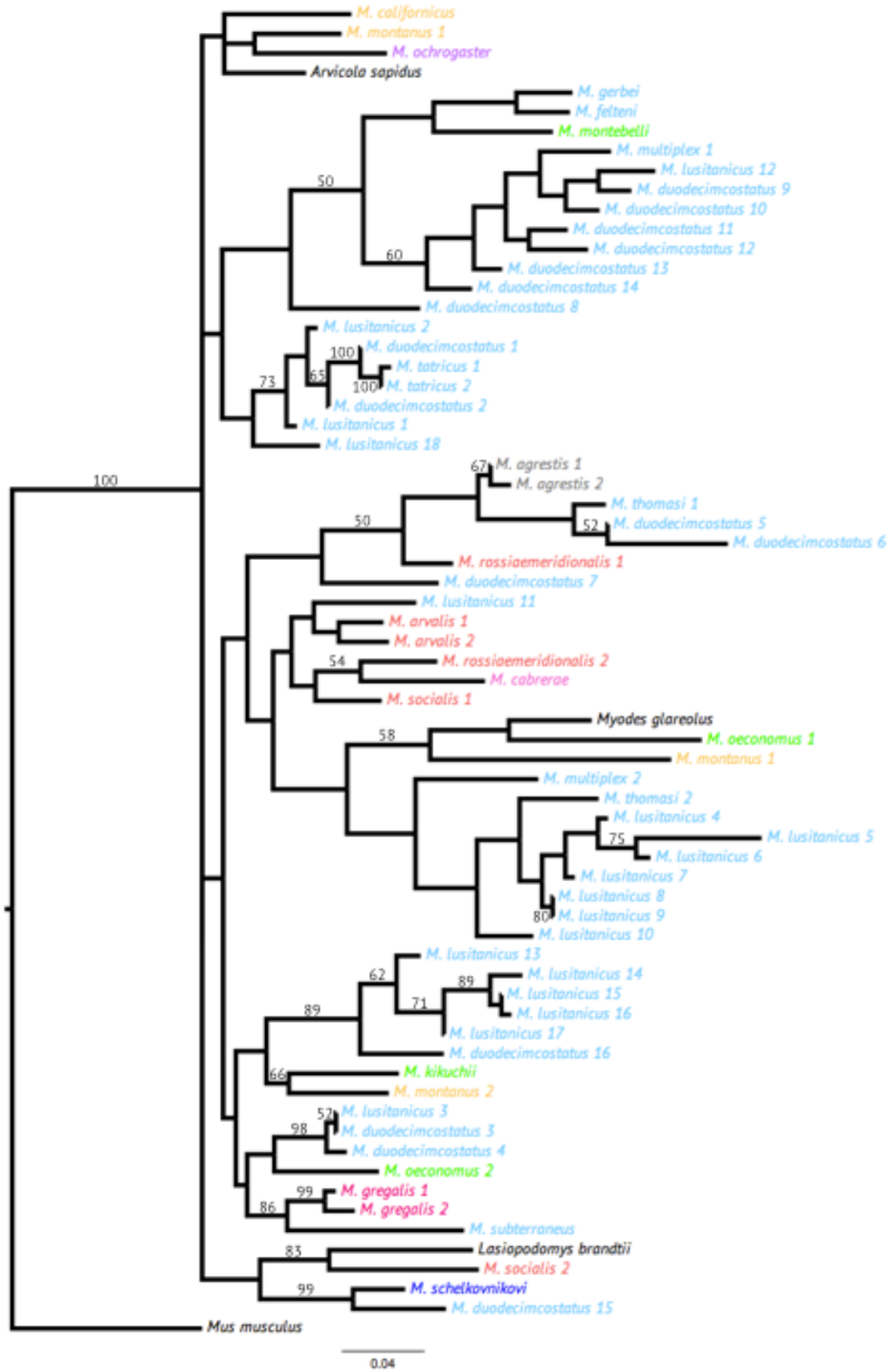


Figure 3– Maximum likelihood phylogenetic tree for all species analysed, considering MHC class II exon 2. RaxML bootstrap values ⩾50% are presented at the respective phylogenetic split.

Figure 3 (continued) – *Microtus* subgenus classification of each haplotype is colour coded as follows: light blue = *Terricola*, bright blue = *Hyrcanicola*, salmon = *Microtus*, pink = *Iberomys*, bright pink = *Stenocranius*, grey = *Agricola*, purple = *Pedomys*, orange = *Mynomes*, and green = *Alexandromys/Pallasiinus*.

DNA sequences obtained were first contributions for all *Microtus* species, but *M. arvalis* and *M. oeconomus* for MHCII (Bryja et al., 2006; Kloch et al., 2013). These standard amplification and sequencing techniques do not enable the identification of the amplified loci, particularly for the less studied MHCI; thus, the following results will only concern MHCII.

Both MHCII phylogenetic trees present clades with a mixture of haplotypes from different species (**Figure 2-3**), different subgenera or even genera (**Figure 3**), indicating an unresolved phylogeny. Moreover, bootstrap and posterior probability values are not ⩾50% for all evolutionary splits (**Figure 2-3**), particularly concerning the broader analysis using more *Microtus sp.* species, which is highly polytomic (**Figure 3**).

The *M. lusitanicus* and *M. duodecimcostatus* phylogenetic tree presents two main phylogenetic groups, separated into six clades (**Figure 2**). From these only two are constituted of single taxon haplotypes: *M. duodecimcostatus 15* and *M. lusitanicus 4-10*.

Concerning the multi-*Microtus sp.* (plus fellow Arvicolinae *Arvicola sapidus* Miller, 1908, *Lasiopodomys brandtii* (Radde, 1861) and *Myodes glareolus* (Schreber, 1780)) phylogenetic tree, four main phylogenetic groups were obtained, divided into six clades (**Figure 3**). None were species-specific or even subgenus-specific, and three were not even genus-specific, due to the inclusion of *Arvicola sapidus*, *Lasiopodomys brandtii* or *Myodes glareolus* haplotypes (**Figure 3**), which did not adopt the expected basal position, after the outgroup *Mus musculus*. Besides *Terricola*, inclusive of *M. lusitanicus* and *M. duodecimcostatus* and present in five clades, other subgenera are also divided: Asian *Alexandromys/Pallasiinus*, European *Microtus* and New World *Mynomes* (**Figure 3**). Hence, phylogenetic relationships in this tree are not, species, subgenus-, genus- nor geographic-specific.

We consider that the next step in MHC studies on these species should include cloning, pyrosequencing or next generation sequencing (reviewed in Wegner, 2009; Babik, 2010). This approach would help to clarify MHCI and MHCII inter- and intraspecific variation for *Microtus sp*., particularly for sympatric and allopatric populations *M. lusitanicus* and *M. duodecimcostatus*. Moreover, these methods would determine the number of expressed loci, present in these taxa, since multiplication of MHC loci has been recorded for other rodents, including Arvicolinae species (e.g. Vincek et al., 1987; Bryja et al., 2006; Axtner & Sommer, 2007; Busch et al., 2008; Penn & Musolf, 2012; Kloch et al., 2013; Winternitz & Wares, 2013).

2.2.6 Acknowledgements

We are grateful to Isabel Rey from the Museo Nacional de Ciencias Naturales (Madrid, Spain) for providing biological samples. We would also like to thank Heidi Lischer for providing the repeated random haplotype sampling analysis software before publication and to Susanne Tellenbach for technical assistance. M. A. Duarte was supported by Ph.D. grant SFRH/BD/70646/2010 from the Foundation for Science and Technology (FCT) and G. Heckel by grant 31003A-149585 from the Swiss National Science Foundation. This study was developed under the project PTDC/BIA-BEC/103729/2008 (FCT), coordinated by C. Bastos-Silveira. Thanks are due, for the financial support to Centro de Estudos de Ambiente e Mar (UID/AMB/50017), to FCT/MEC through national funds and the co-funding by the FEDER, within the PT2020 Partnership Agreement and Compete 2020.

2.2.7 References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723

Axtner J, Sommer S (2007) Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of the bank vole, *Clethrionomys glareolus*. Immunogenetics 59(5):417–426

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. Mol Ecol Resour 10(2):237–251

Beauchamp GK, Yamazaki K, Bard J, Boyse EA (1988) Preweaning experience in the control of mating preferences by genes in the major histocompatibility complex of the mouse. Behav Genet 18:537–547

Beauchamp GK, Yamazaki K (2003) Chemical signaling in mice. Biochem Soc Trans 31:141–151

Boehm T, Zufall F (2006) MHC peptides and the sensory evaluation of genotype. Trends Neurosci 29(2):100–107

Brennan PA, Kendrick KM (2006) Mammalian social odours: attraction and individual recognition. Philos Trans R Soc B 361:2061–2078

Brown RE, Singh PB, Roser B (1987) The major histocompatibility complex and the chemosensory recognition of individuality in rats. Physiol Behav 40(1):65–73

Brunet-Lecomte P, Chaline J (1991) Morphological evolution and phylogenetic relationships of the European ground voles (Arvicolidae, Rodentia). Lethaia 24(1):45–53

Bryja J, Galan M, Charbonnel N, Cosson JF (2006) Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics 58(2-3):191–202

Busch JD, Waser PM, DeWoody JA (2008) Characterization of expressed class II MHC sequences in the banner-tailed kangaroo rat (*Dipodomys spectabilis*) reveals multiple DRB loci. Immunogenetics 60(11):677–688

Cao TM, Lo B, Ranheim EA, Grumet FC, Shizuru JA (2003) Variable hematopoietic graft rejection and graft-versus-host disease in MHC-matched strains of mice. Proc Natl Acad Sci U S A 100(20):11571–11576

Carroll LS, Penn DJ, Potts WK (2002) Discrimination of MHC-derived odors by untrained mice is consistent with divergence in peptide-binding region residues. Proc Natl Acad Sci U S A 99(4):2187–2192

Chaline J, Brunet-Lecomte P, Montuire S, Viriot L, Courant F (1999) Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. Ann Zool Fennici 36(4): 239–267

Crew MD, Filipowsky ME, Neshat MS, Smith GS, Walford RL (1991) Transmembrane domain length variation in the evolution of major histocompatibility complex class I genes. Proc Natl Acad Sci U S A 88(11):4666–4670.

Ditchkoff SS, Lochmiller RL, Masters RE, Hoofer SR, van Bussche RA (2001) Major-histocompatibility-complex-associated variation in secondary sexual traits of white-tailed deer (*Odocoileus virginianus*): evidence for good-genes advertisements. Evolution 55:616–625

Egid K, Brown JL (1989) The major histocompatibility complex and female mating preferences in mice. Anim Behav 38:548–549

Eklund A (1997) The major histocompatibility complex and mating preferences in wild house mice (*Mus domesticus*). Behav Ecol 8:630–634

Eklund A, Egid K, Brown JL (1991) The major histocompatibility complex and mating preferences of male mice. Anim Behav 42:693–694

Garrido-García JA, Soriguer-Escofet RC (2012) Cabrera's Vole *Microtus cabrerae* Thomas, 1906 and the subgenus *Iberomys* during the Quaternary: Evolutionary implications and conservation. Geobios 45(5):437–444

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Jaarola M, Martínková N, Gündüza I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedakis-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioğlu H, Mathias ML, Villate I, Searle JB (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol 33(3):647–663

Jordan WC, Bruford MW (1998) New perspectives on mate choice and the MHC. Heredity 81:127–133

Kelliher KR (2007) The combined role of the main olfactory and vomeronasal systems in social communication in mammals. Horm Behav 52(5):561–570

Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci U S A 78(1):454–458

Kloch A, Baran K, Buczek M, Konarzewski M, Radwan J (2013) MHC influences infection with parasites and winter survival in the root vole *Microtus oeconomus*. Evol Ecol 27:635–653

Liberles SD (2014) Mammalian pheromones. Annu Rev Physiol 76:151-175

Lischer HE, Excoffier L, Heckel G (2014) Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. Mol Biol Evol 31(4):817–831

Manning C, Potts WK, Wakeland EK, Dewsbury DA (1992) What's wrong with MHC mate choice experiments? In: Doty RL, Muller-Schwarze D (eds) Chemical signals in vertebrates. Plenum, New York, pp 229–235

Nadachowski A (2007) The taxonomic status of Schelkovnikov's Pine Vole *Microtus schelkovnikovi* (Rodentia, Mammalia). Acta zool cracov 50A(1-2):67–72

Overath P, Sturm T, Rammensee H-G (2014) Of volatiles and peptides: in search for MHC-dependent olfactory signals in social communication. Cell Mol Life Sci 71(13): 2429–2442

Paterson S, Pemberton JM (1997) No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. Proc R Soc Lond B 264:1813–1819

Penn DJ, Musolf K (2012) The evolution of MHC diversity in house mice. In: Macholán M, Baird SJE, Munclinger P, Piálek J (eds) Evolution of the house mouse. Cambridge University Press, Cambridge, pp 221–252

Penn D, Potts WK (1998) Untrained mice discriminate MHC-determined odors. Physiol Behav 64(3):235–243

Posada, D (2003) Using Modeltest and PAUP* to select a model of nucleotide substitution. In: Baxevanis AD, Davison DB, Page RDM, Petsko GA, Stein LD, Stormo

GD (eds) Current Protocols in Bioinformatics. John Wiley & Sons, Inc. pp 6.5.1–6.5.14

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25(7):1253–1256

Rambaut A (2010) FigTree v1.3.1: Tree figure drawing tool. Available: http://tree.bio.ed.ac.uk/software/figtree/

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

Sambrook J, Fritschi EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Sauermann U, Nürnberg P, Bercovitch FB, Berard JD, Trefilov A, Widdig A, Kessler M, Schmidtke J, Krawczak M (2001) Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. Hum Genet 108(3):249–254

Seddon JM, Baverstock PR (1998) Eight rat RT1Ba sequences. Immunogenetics 48(2):161–162

Singh PB (2001) Chemosensation and genetic individuality. Reproduction 121(4):529–539

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313

Villesen P (2007) FaBox: an online toolbox for fasta sequences. Mol Ecol Notes 7:965–968

Vincek V, Nizetić D, Golubić M, Figueroa F, Nevo E, Klein J (1987) Evolutionary expansion of Mhc class I loci in the mole-rat, *Spalax ehrenbergi*. Mol Biol Evol 4(5):483–491

Wegner KM (2009) Massive parallel MHC genotyping: titanium that shines. Mol Ecol 18(9):1818–1820

Wenink PW, Groen AF, Roelke-Parker ME, Prins HHT (1998) African buffalo maintain high genetic diversity in the major histocompatibility complex in spite of historically known population bottlenecks. Mol Ecol 7:1315–1322

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

Winternitz JC, Wares JP (2013) Duplication and population dynamics shape historic patterns of selection and genetic variation at the major histocompatibility complex in rodents. Ecol Evol 3(6):1552–1568

Wyatt TD (2003) Pheromones and animal behaviour: communication by smell and taste. Cambridge University Press, Cambridge

Yamaguchi M, Yamazaki K, Beauchamp GK, Bard J, Thomas L, Boyse EA (1981) Distinctive urinary odors governed by the major histocompatibility locus of the mouse. Proc Natl Acad Sci U S A 78(9):5817–5820

Yamazaki K, Yamaguchi M, Baranoski L, Bard J, Boyse EA, Thomas L (1979) Recognition among mice. Evidence from the use of a Y-maze differentially scented by congenic mice of different major histocompatibility types. J Exp Med 150(4):755–760

Yamazaki K, Beauchamp GK, Imai Y, Bard J, Phelan SP, Thomas L, Boyse EA (1990) Odor types determined by the major histocompatibility complex in germfree mice. Proc Natl Acad Sci U S A 87(21):8413–8416

Chapter 3

Urinary protein expression

3.1 Urinary protein expression

Type of publication: Short communication

Reference: Duarte MA[1,2,3], Mathias ML[2,3], Bastos-Silveira C[1], Manadas B[4] (in preparation) A first approach on urinary protein expression of sister voles *Microtus lusitanicus* and *M. duodecimcostatus*.

[1] Museu Nacional de História Natural e da Ciência, Departamento de Zoologia e Antropologia, Universidade de Lisboa, Rua da Escola Politécnica, 58, Lisbon 1250-102, Portugal.

[2] Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[3] Centro de Estudos de Ambiente e Mar, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[4] Center for Neuroscience and Cell Biology, University of Coimbra, Portugal.

3.1.1 Abstract

Communication through odour cues, such as urine, plays an important role in social behaviour, particularly mating. Thus, urine may affect mate choice and constitute a behavioural isolation barrier. In this study we analyse, for the first time, the urine of semi-fossorial sister voles *M. lusitanicus* and *M. duodecimcostatus*. An absence of major urinary proteins expression in the urine of *M. duodecimcostatus* and only one (MUP20) in *M. lusitanicus* partially support the postulation of Hagemeyer and colleagues (2011) that major urinary expression is probable absent in fossorial genera. These results also seem to indicate that major urinary proteins are not essential to intra- or interspecific communication among *M. lusitanicus* and *M. duodecimcostatus* and that a different communication pathway, using other type of odour cues, may exist.

3.1.2 Keywords

Behavioural isolation; odour; urinary proteins; sister species; *Microtus lusitanicus*; *Microtus duodecimcostatus*.

3.1.3 Introduction

Urine plays a central role in mammalian communication, both at the intra- and interspecific level, and consequently to survival, reproduction and social behaviour (reviewed in Brennan & Keverne, 2004; Arakawa et al., 2008). Urinary odour cues are structurally diverse and include steroid derivatives, peptides, volatile molecules and large protein-ligand complexes (reviewed in Liberles, 2014).

Externally secreted steroids, e.g. through urine, can provide direct information about the hormonal state of the individual. Steroid derivatives in mouse urine follow the addition of polar moieties (glycine, taurine, and sulfate), which promote water solubility (reviewed in Liberles, 2014). For instance, sulfated steroids are non-volatile cues that induce responses in a high percentage of vomeronasal sensory neurons and accessory olfactory bulb cells (reviewed in Liberles, 2014).

Concerning major histocompatibility complex I (MHCI) peptides, it has been suggested that the mouse olfactory system exploits MHC locus heterogeneity to discriminate individuality during social interactions, by recognizing either receptor fragments, bound peptides or other associated odours (reviewed in Liberles, 2014). MHCI peptides stimulate high-affinity electrical and calcium responses in vomeronasal and main olfactory epithelium sensory neurons (reviewed in Liberles, 2014).

Sex-specific urinary volatiles apparently regulate reproductive physiology, scent attraction and aggressive behaviour. In male urine it can be present 2-(sec-butyl)-dihydrothiazole, dehydro-exo-brevicomin, (methylio)methanethiol, trimethylamine and (Z)-5-tetradecen-1-ol; while in the female urine it can be found 2,5-dimethylpyrazine, aliphatic ketones and acetates (reviewed in Liberles, 2014). They evoke high-affinity responses in both major olfactory epithelium and vomeronasal organ sensory neurons (reviewed in Liberles, 2014). These volatile urinary pheromones may require a lipocalin through the nasopalatine duct to access the vomeronasal organ lumen (reviewed in Liberles, 2014).

Lipocalin is a large and diverse family of small extracellular β-barrel proteins, which present a hydrophobic calyx, appropriate for the binding and transportation

of small hydrophobic molecules (reviewed in Beynon et al., 2007). The most common lipocalins were described for *Mus musculus*: major urinary proteins (MUPs), secreted primarily in the urine and saliva; and odorant binding proteins (OBPs), expressed predominantly in the nasal tissue.

MUPs are responsible for binding and slow release of low-molecular-mass volatile pheromones, controlled by the rate of dissociation from these proteins (reviewed in Beynon & Hurst, 2003). MUPs are extremely polymorphic in the animal model, enabling a chemical signalling at the individual and distinguishing this class from other lipocalins (reviewed in Beynon & Hurst, 2003). Urinary MUPs correspond to >99% of mouse urinary proteins and are synthesized in the liver, where they can be sex specific (Liberles, 2014). MUP genes are highly polymorphic in outbred mouse populations, with different individuals producing unique MUP signatures, enabling genotype discrimination (reviewed in Liberles, 2014). MUPs, proposed to prevent the degradation of volatile molecules from scent marks or to transport them to hydrophilic environments, e.g. urine, or the vomeronasal organ lumen, present the highest binding affinity for mouse 2-(sec-butyl)-dihydrothiazole and lower affinity for other urinary volatiles (reviewed in Liberles, 2014). MUPs alone are now thought to function independently as mouse chemical cues, stimulating responses in vomeronasal organ sensory neurons (reviewed in Liberles, 2014). They are associated to individuality recognition, sexual attraction, aggression, hormone modulation, spatial learning and predator odour-induced fear (reviewed in Liberles, 2014).

Furthermore, individual lipocalins have been identified in different rodent species. Roborovskin is expressed in the urine of the Roborovski hamster *Phodopus roborovskii* (Satunin, 1903) and shares significant homology with OBPs from *Myodes glareolus* and with aphrodisin, a submandibular protein from the golden hamster *Mesocricetus auratus* that is also a lipocalin. Lower levels of homology were detected between roborovskin and other lipocalins, including MUPs from *Mus musculus* and *Rattus norvegicus* (Berkenhout, 1769). Roberts and colleagues (2010) detected a male specific peripheral MUP, named darcin, which acts as a

pheromone in female attraction and learning (Roberts et al., 2010; Phelan et al., 2014). Darcin binds to 2-sec-butyl 4,5 dihydrothiazole, a male-specific pheromone and one of the most abundant volatile molecules in the urine of male *Mus musculus* (reviewed in Phelan et al., 2014). It is encoded by a peripheral gene in the MUPs cluster and presents higher levels of tissue expression and function than those encoded by central genes (reviewed in Phelan et al., 2014).

Nevertheless, biochemical analyses on the urine of different genera of fossorial rodent, eusocial Zambian mole-rats *Fukomys* Kock et al., 2006, solitary Israeli blind mole rats *Spalax* Guldenstaedt, 1770, and social Chilean coruros *Spalacopus* Wagler, 1832 indicated an absence of MUPs expression (Hagemeyer et al., 2011). 2D gel electrophoresis also revealed low levels of proteins and the detection of a possible homologue in *Fukomys* mole-rat urine to the hamster aphrodisin (Hagemeyer et al., 2011). These results seem to indicate that fossorial rodents can successfully communicate using a non-MUPs route.

Considering these inconsistent findings, the present study aimed to analyse protein expression in the urine of semi-fossorial sister species *M. lusitanicus* Gerbe (1879) and *M. duodecimcostatus* de Selys-Longchamps (1839), due to its possible relevance in odorous communication associated with species-specific mate choice and, consequently, to behavioural isolation.

3.1.4 Materials and methods

M. lusitanicus and *M. duodecimcostatus* subjects were first- and second-generation adults selected from the animal facility colony (Faculty of Sciences, University of Lisbon). This colony was established by wild-caught individuals from Portalegre (Portugal) under project PTDC/BIA-BEC/103729/2008 (FCT).

Animals were kept under controlled temperature (22±2°C) and photoperiod (12:12h light:dark cycle, with light available from 7:00 to 19:00) conditions in Makrolon polycarbonate type III cages containing >4cm of wood shavings. Hay, pine cones and small branches were also provided for environmental enrichment. The diet consisted of carrots and apples, available *ad libitum*.

Urine from two males of each species was collected during a period of 10 months (September, 2013 to June, 2014) to sterile tubes. Sampling of each subject has taken 1-4 months. Before storing at -20ºC, each tube was subjected to centrifugation at 4000×g for 10 minutes at room temperature. The precipitation salts were discarded and the supernatant was pipetted to a new tube. Samples from the same subject were mixed and centrifuged at 8000×g in a Nanosep® 10K Omega™ filter (Pall). The retentate was re-suspended in 100µl of Tris 50mM (pH=6.8) solution with 0.1% SDS (sodium dodecyl sulfate), homogenised and transferred to a new sterile tube, before being stored at -20ºC.

Samples were analyzed using a short-GeLC-MS approach as previously described (Anjo et al., 2014) on an AB Sciex 5600 TripleTOF in information-dependent acquisition (IDA). Peptides separation was performed using liquid chromatography (nanoLC Ultra 2D, Eksigent) on a Halo Fused-Core™ C18 reverse phase column (300µm x 15cm, 3µm, 120Å, Eksigent®) at 5µL/min with linear gradient of 2% to 30% acetonitrile in 0.1%FA for 45 minutes gradient. Peptides were eluted into the mass spectrometer using an electrospray ionization source (DuoSpray™ Source, ABSciex).

IDA experiments were performed for each sample. The mass spectrometer was set for IDA scanning full spectra (350-1250m/z) for 250ms, followed by up to 20 MS/MS scans (100–1500m/z for 100ms each). Candidate ions with a charge state between +2 and +5 and counts above a minimum threshold of 70 counts per second were isolated for fragmentation and one MS/MS spectra was collected before adding those ions to the exclusion list for 20 seconds (mass spectrometer operated by Analyst® TF 1.6, ABSciex). Rolling collision energy was used with a collision energy spread of 5.

Protein identification for each IDA method was obtained using ProteinPilot™ software (v4.5, ABSciex®) with the following search parameters: search against SwissProt database (release 2014_02); trypsin digestion; acrylamide as cysteine alkylating reagent; special focus option for gel based, followed by an FDR analysis.

3.1.5 Results and discussion

Single-session volume collection varied according to the species. For *M. lusitanicus* samples varied from 0-750µl, with an average of 85.5µl. On the other hand, *M. duodecimcostatus* urine volumes per session varied from 0-1100µl, with an average of 158.49µl. We consider that these differences are probably related to size differences between both taxa (Cotilla & Palomo, 2002; Mira & Mathias, 2007).

Analyses detected a variety of peptides in the urine of both taxa (**Table 1**), since proteinuria is a normal condition in rodents contrarily to other mammals (Nutr Rev, 1958). Overall, more peptides were found in the urine of *M. lusitanicus* due to a methodological optimization between samples (**Table 1 and Annex**). Most peptides found are associated to catalytic and/or binding functions, and the most common were trypsins, immunoglobulins, serum albumin and keratins (**Table 1** and **Annex**).

The presence of trypsins is related to the methodological digestion of the urinary proteins; whereas keratins, important constituents of the skin and hair of mammals, such as *M. lusitanicus* and *M. duodecimcostatus,* are expected contaminants. Serum albumin peptides were found in high quantities, a normal feature in rodents, as Igs, which contribute to immunity by preventing pathogens from entering and damaging cells; by coating pathogens to be further digested; and by stimulating other immune responses, such as the complement pathway (reviewed in Schroeder & Cavacini, 2010). Nevertheless, abnormal immunoglobulins, such as monoclonal protein (M protein) suggest the presence of disease. We believe that it is not the case of *M. lusitanicus* and *M. duodecimcostatus*, since M protein was not found and Igs peptide homology found could be related to MHC peptide ligands that function as individuality chemosignals. In *Mus musculus*, and possibly other mammals, these MHC peptide ligands form a direct link between individuality at the immunological and behavioural levels (reviewed in Brennan & Kendrick, 2006).

Table 1 – Number of peptides, grouped by general function, detected in the urine of *M. lusitanicus* and *M. duodecimcostatus*. For more detailed information see **Annex**.

General function	Present in ML	Present in MD
Catalytic activity	133	46
Binding	95	40
Antigen binding	34	12
Structural molecule	20	5
Receptor activity	10	11
Transporter activity	16	4
Structural constituent of cytoskeleton	11	6
Inhibitor activity	13	3
Ubiquitous protein	3	1
Activator activity	3	0
Folding	0	2
Structural constituent of epidermis	2	1
Antioxidant activity	1	1
Binding and catabolism of lipoproteins	1	1
Protease inhibitor	1	1
Cytoprotector triggered by oxygen deprivation	1	1
Cytokine activity	1	1
Contributes to colloid osmotic pressure	1	1
Prevents urinary tract infection	1	1
Growth factor activity	1	1
Male pheromone	1	0
Binds most of the male pheromone, 2-sec-butyl-4,5-dihydrothiazole, in urine	1	0
Lipocalin	1	0
Odorant binding	1	0
Copulatory plug formation	1	0
Hormone	1	0
Acetylcholine biosynthesis and secretion	1	0
Regulation of cell growth	1	0
Autophagy	1	0
Apoptotic process	1	0
Antioxidant	1	0
Immune response	1	0
Structural constituent of muscle	1	0
Natriuresis and diuresis	1	0
Ocular mucus homeostasis	1	0
Kidney homeostasis	1	0
Carrier activity	1	0
Folding	1	0
Homodimerization activity	1	0
Structural constituent of hair and nails	0	1
Transcriptional control	0	1

We consider that such hypothesis should be pondered since MHC-derived peptides are present in *M. lusitanicus* and *M. duodecimcostatus* urine in much higher quantities than lipocalins, contrariwise to previous *Mus musculus* studies (reviewed in Overath et al., 2014).

Three distinct peptides with ⩾95% confidence for hypoxia up-regulated protein 1 were also detected in the urine of *M. lusitanicus* and *M. duodecimcostatus* (**Table 1** and **Annex**). This protein is associated to cytoprotective cellular mechanisms triggered by oxygen deprivation, thus suppression may lead to accelerated apoptosis.

Analyses did not detect, however, a range of peptides homologues to known lipocalins, related to odorous communication. The only exception was one distinct peptide with ⩾95% confidence for probasin and another for MUP20, both in the urine of *M. lusitanicus* (**Annex**).

Probasin, an androgen-regulated prostate-specific protein, is synthesized in the rodent prostate and secreted into the seminal fluid (Johnson et al., 2000; Kasper & Matusik, 2000). Probasin secretion is probably a result of copulation and not a pheromone *per se* with a pre-mating role (Kasper & Matusik, 2000). Nevertheless, it is possible that probasin marks females to ward off other males (Kasper & Matusik, 2000). In rodent evolution, there is a mouse-rat split preceded by a mouse-rat common ancestor split from cricetids (Stopková et al., 2010), which include *Microtus*. Such evolutionary history questions the role of probasin in murids (Muridae) versus cricetids (Cricetidae). Is probasin a result of copulation and not a pheromone with a pre-mating role? Does it mark cricetid females to ward off other males? Those questions should be addressed in future analyses on the urine of cricetid taxa.

On the other hand, MUP20, also known as darcin or MUP20, is a male pheromone which stimulates female sexual attraction to male urinary scent and promotes a strong learned attraction to the airborne urinary odour of an individual male (Roberts et al., 2010). It promotes male aggressive behaviour and binds most

of the male pheromone, 2-sec-butyl-4,5-dihydrothiazole (SBT), in urine (Roberts et al., 2010).

An absence for MUPs expression in the urine of *M. duodecimcostatus* and only one (MUP20) in *M. lusitanicus* partially support the postulation of Hagemeyer and colleagues (2011) that major urinary expression is probable absent in fossorial genera. Moreover, Darcin (MUP20), the lipocalin detected for *M. lusitanicus*, corresponded to only one peptide with ⩾95% confidence for MUP20, in a spectrum of hundreds of peptides detected per individual urine. A proportion up to 99% (Humphries et al. 1999) or at least a dominance of MUPs expression when compared to other urinary proteins was expected. These results seem to indicate that MUPs are not essential to intra- or interspecific communication among *M. lusitanicus* and *M. duodecimcostatus*. Probably, these semi-fossorial species use a non-MUPs route to communicate through odour cues, e.g. volatile molecules transported by other lipocalins or carriers. We consider that this hypothesis should be pondered in forthcoming analyses on the urine of both taxa. Furthermore, extensive MUP gene polymorphism has been only found in *Mus musculus* and moderately in *Rattus norvegicus* (Logan et al., 2008), suggesting that the expansion of the MUP gene cluster occurred separately in these genera, being atypical among mammals (Hagemeyer et al., 2011). Our results partially support the postulation of Hagemeyer and colleagues (2011), which considers that there is a universal process that transmits semiochemicals from a sender to a receiver animal across Rodentia, and that the search for such process must continue, since MUPs are clearly not sufficiently prevalent to achieve this mission.

We consider that future analyses based on more *M. lusitanicus* and *M. duodecimcostatus* specimens, including female subjects and quantification of the peptides detected, would enlighten the role of urinary proteins in these species apparent reproductive behavioural isolation, particularly regarding the expression of MHC-derived peptides, MUPs or other lipocalins in these semi-fossorial taxa.

3.1.6 Compliance with ethical standards

Procedures were performed in accordance with the guidelines set by Portuguese (nº. 113/2013 Decree-law) and European (nº. 63/2010/CE Directive) legislation. Animals were handled by a Federation of Laboratory Animal Science Associations category C licensed biologist (M. A. Duarte).

3.1.7 Acknowledgments

M. A. Duarte was supported by Ph.D. grant SFRH/BD/70646/2010 from the Foundation for Science and Technology (FCT). This study was financed by project PTDC/BIA-BEC/103729/2008 (FCT), PTDC/NEU-NMC/0205/2012 (FCT), UID/NEU/04539/2013 (FCT), by the National Mass Spectrometry Network (RNEM, REDE/1506/REM/2005) and QREN - Quadro de Referência Estratégica Nacional. Thanks are due, for the financial support to Centro de Estudos de Ambiente e Mar (UID/AMB/50017) and to Center for Neuroscience and Cell Biology, to FCT/MEC through national funds and the co-funding by the FEDER - Fundo Europeu de Desenvolvimento Regional, within the PT2020 Partnership Agreement and COMPETE - Programa Operacional Factores de Competitividade 2020.

3.1.8 References

Anjo SI, Santa C, Manadas B (2015) Short GeLC-SWATH: A fast and reliable quantitative approach for proteomic screenings. Proteomics, 15(4):757–762

Arakawa H, Blanchard DC, Arakawa K, Dunlap C, Blanchard RJ (2008) Scent marking behavior as an odorant communication in mice. Neurosci Biobehav Rev 32(7):1236–1248

Beynon RJ, Hurst JL (2003) Multiple roles of major urinary proteins in the house mouse, *Mus domesticus*. Biochem Soc Trans 31:142–146

Beynon RJ, Hurst JL, Turton MJ, Robertson DHL, Armstrong SD, Cheetham SA, Simpson D, MacNicoll A, Humphries RE (2007) Urinary Lipocalins in Rodenta: is there a Generic Model? In: Hurst JL, Beynon RJ, Roberts SC, Wyatt TD (eds)

Chemical Signals in Vertebrates 11 (Part I). Springer New York, New York, pp 37–49

Brennan PA, Kendrick KM (2006) Mammalian social odours: attraction and individual recognition. Philos Trans R Soc B 361:2061–2078

Brennan PA, Keverne EB (2004) Something in the air? New insights into mammalian pheromones. Curr Biol 14(2):R81–R89

Cotilla I, Palomo LJ (2007) *Microtus duodecimcostatus* (de Selys-Longchamps, 1839). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 422–425

Fuller RH, Morris GE (2012) Quantitative Proteomics Using iTRAQ Labeling and Mass Spectrometry. In: Leung H-C (ed) Integrative Proteomics. InTech, Rijeka, pp 347–362

Hagemeyer P, Begall S, Janotova K, Todrank J, Heth G, Jedelsky PL, Burda H, Stopka PJ (2011) Searching for major urinary proteins (MUPs) as chemosignals in urine of subterranean rodents. Chem Ecol 37(7):687–694

Humphries RE, Robertson DH, Beynon RJ, Hurst JL (1999) Unravelling the chemical basis of competitive scent marking in house mice. Anim Behav 58(6):1177–1190

Johnson MA, Hernandez I, Wei Y, Greenberg N (2000) Isolation and characterization of mouse probasin: An androgen-regulated protein specifically expressed in the differentiated prostate. Prostate 43(4):255–262

Kasper S, Matusik RJ (2000) Rat probasin: structure and function of an outlier lipocalin. Biochim Biophys Acta 1482(1-2):249–258

Liberles SD (2014) Mammalian pheromones. Annu Rev Physiol 76:151-175

Logan DW, Marton TF, Stowers L (2008) Species specificity in major urinary proteins by parallel evolution. PLoS One 3(9):e3280

Mira A, Mathias ML (2007) *Microtus lusitanicus* (Gerbe, 1879). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 418–421

Nutr Rev (1958) Proteinuria in rats and mice. Nutr Rev 16:337–340

Overath P, Sturm T, Rammensee H-G (2014) Of volatiles and peptides: in search for MHC-dependent olfactory signals in social communication. Cell Mol Life Sci 71(13): 2429–2442

Phelan MM, McLean L, Armstrong SD, Hurst JL, Beynon RJ, Lian LY (2014) The structure, stability and pheromone binding of the male mouse protein sex pheromone darcin. PLoS One 9(10):e108415

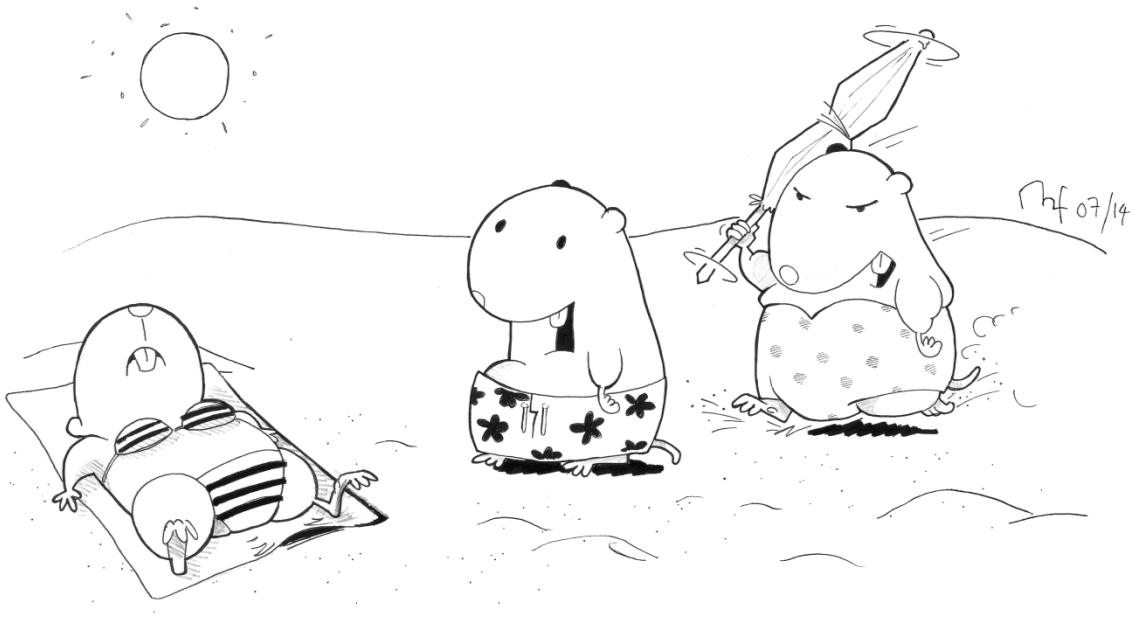
Roberts SA, Simpson DM, Armstrong SD, Davidson AJ, Robertson DH, McLean L, Beynon RJ, Hurst JL (2010) Darcin: a male pheromone that stimulates female memory and sexual attraction to an individual male's odour. BMC Biol 8:75

Schroeder HW, Cavacini L (2010) Structure and Function of Immunoglobulins. J Allergy Clin Immunol 125(2):S41–S52

Stopková R, Stopka P, Janotová K, Jedelský PL (2007) Species-specific expression of major urinary proteins in the house mice (*Mus musculus musculus* and *Mus musculus domesticus*). J Chem Ecol 33(4):861–869

Chapter 4

Partner preference



Subchapters:

4.1 Partner preference in an artificial syntopic environment

4.2 Pair bonding behaviour

4.1 Partner preference in an artificial syntopic environment

Type of publication: Short communication

Reference: Duarte MA[1,2,3], Bastos-Silveira C[1], Heckel G[4,5], Mathias ML[2,3] (in preparation) First recorded evidence of interspecific mating and hybridization between *Microtus lusitanicus* and *M. duodecimcostatus* in an artificial syntopic environment.

[1] Museu Nacional de História Natural e da Ciência, Departamento de Zoologia e Antropologia, Universidade de Lisboa, Rua da Escola Politécnica, 58, Lisbon 1250-102, Portugal.

[2] Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[3] Centro de Estudos de Ambiente e Mar, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[4] Computational and Molecular Population Genetics, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland.

[5] Swiss Institute of Bioinformatics, Genopode, Lausanne, Switzerland.

4.1.1 Abstract

Recently diverged *Microtus lusitanicus* and *Microtus duodecimcostatus* are sister species of the speciose *Microtus* genus. These voles can hybridize in captivity when housed together; while in nature, at the sympatry area of distribution, hybridization seems to be a rare event. Moreover, *M. lusitanicus* and *M. duodecimcostatus* seem to prefer conspecific over heterospecific odour cues, suggesting that behavioural reproductive barriers may be playing an important role in these species reproductive isolation. Thus, in the present study, we investigated *M. lusitanicus* and *M. duodecimcostatus* mate choice in the presence of both potential mates simultaneously, a conspecific and a heterospecific. To do so, we constituted an artificial syntopic environment, inhabited by both species and sexes. Unexpectedly, we recorded, for the first time, spontaneous interspecific mating between a male *M. lusitanicus* and a female *M. duodecimcostatus* in the presence of conspecific mates. We also observed other unforeseen social behaviours between *M. lusitanicus* subjects and the female *M. duodecimcostatus*,

such as grooming and sharing everyday activities. The male *M. duodecimcostatus* presented a monk-like behaviour, not sharing everyday activities with subjects of either species. Moreover, two litters were generated during this assay. Genotyping confirmed that one was intraspecific *M. lusitanicus* and the other a hybrid litter between the male *M. lusitanicus* and the female *M. duodecimcostatus*. These observations suggest that individual behavioural variability may be associated to *M. lusitanicus* and *M. duodecimcostatus* mate choice and support previous findings reporting that both species are social monogamous with the possibility of male extra-pair mating.

4.1.2 Keywords

Speciation; Mate choice; Hybridization; *Microtus lusitanicus*; *Microtus duodecimcostatus*; Sister species

4.1.3 Introduction

The Lusitanian pine vole *Microtus lusitanicus* (Gerbe, 1879) and the Mediterranean pine vole *Microtus duodecimcostatus* de Selys-Longchamps, 1839 are small arvicolids from the speciose *Microtus* genus Schrank, 1798. *M. lusitanicus* diverged from its sister species, *M. duodecimcostatus*, very recently, approximately 60,000 years ago (Brunet-Lecomte & Chaline, 1991). In terms of geographical distribution, *M. lusitanicus* inhabits Northern Iberia and part of the French Pyrenees, while *M. duodecimcostatus* occupies a bigger area covering Southern Iberia and part of the South of France (reviewed in Santos, 2009). These voles present a sympatry area of distribution located in the middle of the Iberian Peninsula, ranging from Portugal, throughout Spain, reaching the Pyrenees (reviewed in Santos, 2009). So far, a single locality (Vale Vaqueiros, Portalegre) in central Portugal constitutes an extant syntopic location (Duarte et al., 2015).

Morphological, reproductive and ecological characteristics suggested that *M. lusitanicus* and *M. duodecimcostatus* would be monogamous voles (Madureira 1982; Heske and Ostfeld 1990; Guédon et al. 1991a,b; Guédon and Pascal 1993; Paradis

and Guédon 1993; Mira 1999; Cotilla and Palomo 2007; Mira and Mathias 2007; Santos 2009; Ventura et al. 2010). Recently, partner preference tests supported such postulation and revealed the possibility of male extra-pair mating for both species, indicating that *M. lusitanicus* and *M. duodecimcostatus* are mostly social monogamous (Duarte et al. 2015).

Furthermore, *M. lusitanicus* and *M. duodecimcostatus* seem to prefer urinary and faecal odour cues of conspecific over heterospecific potential mates (Soares, 2013). Nonetheless, it is known that these sister species hybridize both in captivity (Wiking, 1976; Soares, 2013; Cerveira, in preparation) and nature (Bastos-Silveira et al., 2012). Captive hybridization has been recorded when heterospecific couples are housed together, thus either the subjects mate heterospecifically or remain sexually naïve; while in nature it seems to be a rare event (two possible hybrids in a sample size of 295 individuals).

Considering these findings, we questioned if in the presence of both potential mates simultaneously, a conspecific and a heterospecific, *M. lusitanicus* and *M. duodecimcostatus* would still prefer to mate with the subject of its own species. To address this question we constituted an artificial syntopic environment where both species and sexes co-existed.

4.1.4 Materials and methods

M. lusitanicus and *M. duodecimcostatus* subjects used in this study were second-generation sexually naïve adults (3.5-6 months old) selected from an experimental colony (Faculty of Sciences, University of Lisbon) established by wild-caught individuals. Animals were kept under controlled temperature (22±2°C) and photoperiod (12:12h light:dark cycle, with light available from 7:00 to 19:00) conditions in Makrolon polycarbonate type III cages containing >4cm of wood shavings. Hay, pine cones and small branches were also provided for environmental enrichment. The diet consisted of carrots and apples, available *ad libitum*.

Genomic DNA was extracted from tail tips stored at in absolute ethanol using a standard phenol-chlorophorm protocol (Sambrook et al., 1989). We genotyped 30 *M. lusitanicus* (14 females and 16 males) and 18 *M. duodecimcostatus* (7 females and 11 males). An *a priori* genotyping was indispensable to the selection of potential mates to be used in this assay, so that the parentage of the generated offspring could be clearly determined. Amplification was performed using QIAGEN Multiplex Kit (QIAGEN) according to the protocol described in Braaker & Heckel (2009). Twelve microsatellite loci were analysed: MM1 and MM2 (Ishibashi et al., 1999); CRB5 and CRB7 (Ishibashi et al., 1995); MAG6 and MAG25 (Jaarola et al., 2007); and MAR3, 12, 16, 63, 76, 80 (Walser & Heckel 2008). Fragment separation was carried out on an ABI 3100 sequencer (Applied Biosystems™, Thermo Fisher Scientific Inc.). Genotypes were scored using GeneMapper® software, version 3.7 (Applied Biosystems™, Thermo Fisher Scientific Inc.) against the internal LIZ 500 size standard.

All potential mates were implanted with PIT tags, (FDX-B, 7x1.35mm, AB10320, Loligo® Systems) to be easily identified by a PIT reader (AB10625, Loligo® Systems) without the need of handling during the assay, which could induce stress.

The artificial syntopic environment consisted in a 1m x 1m x 40cm glass terrarium, 6mm thick, enriched with >6cm of wood shavings, hay, pine cones, toilet paper rolls, tree barks and small branches. Animals were kept under the same diet, temperature and photoperiod as the rest of the colony. The setup was composed by one male and one female from each species (N=4). Interaction with test subjects was only performed once a week to provide food and check for the presence of pups. Social behaviours were randomly videotaped using a high definition (1080p) camera.

4.1.5 Results

The artificial syntopic environment was kept for 2 months. After an adjustment period, sharing of everyday activities (**Online Resource 1**) and grooming behaviour was observed between the female *M. duodecimcostatus* and both *M. lusitanicus*

subjects. Conversely, the male *M. duodecimcostatus* acted monk-like, being most often alone and secluded in its nest (**Online Resource 2**). It only appeared at the surface to search for food or nesting materials.

For the first time, interspecific mating between *M. lusitanicus* and *M. duodecimcostatus* was recorded in the presence of both species and sexes (**Online Resource 3**). The *M. duodecimcostatus* female was observed more than one time copulating with the male *M. lusitanicus*, exhibiting normal copulatory behaviour (**Fig. 1a** and **1b**; **Online Resource 3**). Mating behaviour partially occurred in the presence of the female *M. lusitanicus*, which groomed the female *M. duodecimcostatus* by the end of copulation (**Fig. 1c** and **1d**; **Online Resource 3**).

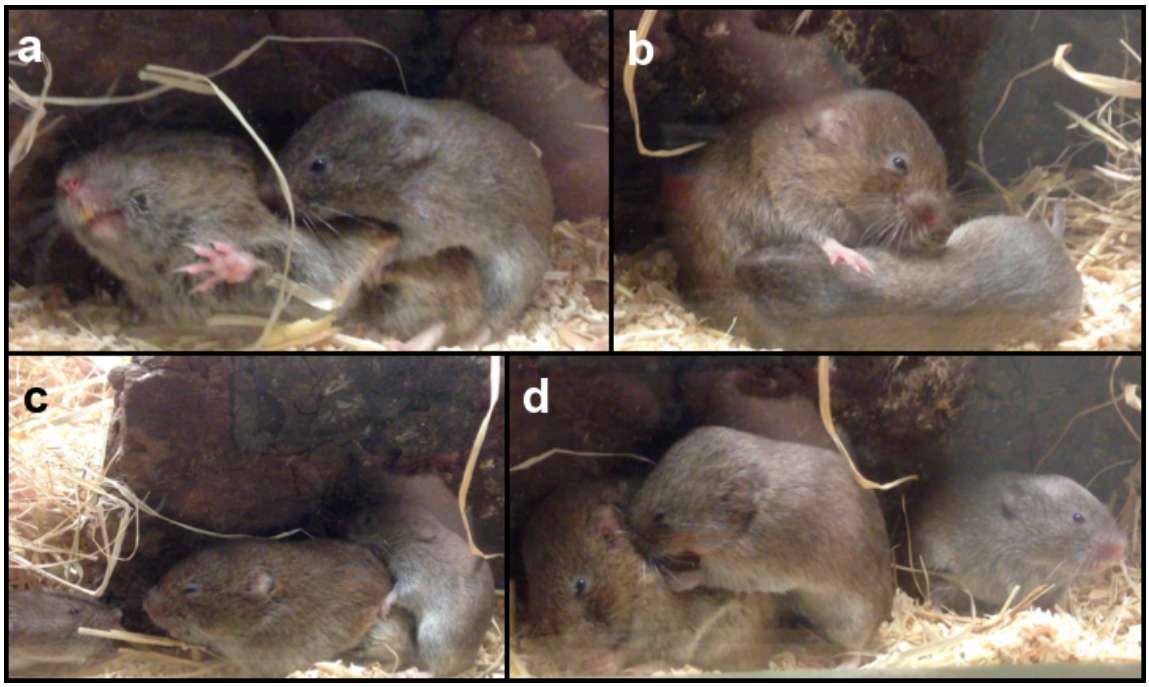


Figure 1 – Video frames of interspecific mating behaviour between a male *M. lusitanicus* and a female *M. duodecimcostatus* (see Online Resource 3): a – the male *M. lusitanicus* (right) copulates with the female *M. duodecimcostatus* (left); b – the female *M. duodecimcostatus* tries to resist copulation, similarly to conspecific mating behaviour; c – the female *M. lusitanicus* approaches (left); d – the female *M. lusitanicus* (middle) grooms the female *M. duodecimcostatus* after copulatory behaviour ends.

During the fifth week of the assay, two litters were found (A and B). A tail sample was collected and genotyped in order to determine the parentage of each

pup (**Table 1**). The genotypes obtained clearly indicate that litter A is the outcome of the recorded interspecific mating between the female *M. duodecimcostatus* female and the male *M. lusitanicus*, demonstrating that both species not only mated but also successfully hybridized (**Table 1**). On the other hand, litter B is an intraspecific *M. lusitanicus* litter (**Table 1**). The obtained genotypes also reveal that both litters are single fathered, i.e. heteropaternal superfecundation was non-existent.

Inter- or intraspecific aggressions were not observed during the duration of the assay.

4.1.6 Discussion

The present behaviour assay recorded, for the first time, spontaneous interspecific mating and hybridization between *M. lusitanicus* and *M. duodecimcostatus* in the presence of both species and sexes. Not only the male *M. lusitanicus* copulated with the female *M. duodecimcostatus*, but also a successful viable hybrid litter was generated.

Furthermore, our results indicate that in the presence of receptive conspecific and heterospecific potential mates the male *M. lusitanicus* can mate with either female subject. This outcome supports social monogamy with the possibility of extra-pair mating, even when pair bonding is formed (Duarte et al., 2015).

Considering the secluded behaviour exhibited by the male *M. duodecimcostatus*, we hypothesise that individual behavioural variability ('animal personality') may be a factor to consider in *M. lusitanicus* and *M. duodecimcostatus* mate choice, and consequently behavioural isolation. The female *M. duodecimcostatus* and *M. lusitanicus* preferred to copulate with the social and sexually available *M. lusitanicus* male, rather than the asocial and secluded *M. duodecimcostatus* male. Individual behavioural variability has been termed 'animal personality', following the human personality psychology research tradition (Gosling and John, 1999; Gosling, 2001; Drent et al., 2003; Dall et al., 2004; Dingemanse and Reale, 2005).

Table 1 – Genotype of *M. lusitanicus* (ML) and *M. duodecimcostatus* (MD) test subjects and born litters (per pup). Parental distinctive alleles inherited by pups are highlighted in bold. N.D. = not determined.

Microsatellite	Genotypes											
	Parental subjects				Litter A				Litter B			
	Male MD	Female MD	Male ML	Female ML	1	2	3	4	1	2	3	4
MM1	51/51	51/**53**	51/51	51/51	51/**53**	51/**53**	51/**53**	51/**53**	51/51	51/51	51/51	51/51
MAR63	293/303	**283**/293	**273**/293	293/293	**273**/293	**273**/**283**	N.D.	N.D.	293/293	N.D.	**273**/293	**273**/293
MAG25	157/157	157/157	155/155	155/**169**	155/157	155/157	155/157	155/157	155/155	155/155	155/155	155/**169**
MAG6	171/171	165/165	**177**/**177**	165/165	165/**177**	165/**177**	N.D.	N.D.	165/**177**	165/**177**	165/**177**	165/**177**
CRB5	126/130	126/126	104/104	104/104	104/126	104/126	104/126	104/126	104/104	104/104	104/104	104/104
MAR16	155/155	155/155	153/153	153/**167**	153/155	153/155	153/155	153/155	153/153	153/153	153/153	153/**167**
MAR80	207/211	223/**225**	**209**/**215**	**199**/**203**	**209**/**225**	**215**/**225**	**209**/**225**	**209**/**225**	**199**/**215**	**199**/**215**	**203**/**215**	**199**/**215**
MM2	178/178	174/178	174/**190**	N.D	178/**190**	174/178	174/178	178/**190**	168/**190**	174/**190**	168/**190**	168/174
CRB7	97/99	97/97	93/**95**	93/95	**95**/97	**95**/97	**95**/97	**95**/97	93/**95**	93/93	93/**95**	93/**95**
MAR76	118/188	110/118	**120**/**126**	108/**116**	118/**126**	118/**126**	118/**126**	118/**126**	**116**/**120**	**116**/**120**	**116**/**126**	**116**/**126**
MAR12	86/88	88/88	86/88	88/**94**	88/88	88/88	88/88	88/88	86/88	88/**94**	88/**94**	88/**94**
MAR3	160/164	**146**/**146**	144/144	144/144	144/**146**	144/**146**	144/**146**	144/**146**	144/144	144/144	144/144	144/144

It has been observed in many species, suggesting that the phenomenon is widespread (e.g. Gosling and John, 1999; Koolhaas et al., 1999; Carere and Maestripieri, 2003; Sih et al., 2004; Reale et al., 2007). 'Animal personality' has been described for different behaviours, including mate choice (reviewed in Shuett et al., 2010); thus, we believe that the bolder and social behaviour of the male *M. lusitanicus* was more attractive to the female *M. duodecimcostatus*, than the shier and asocial behaviour of its conspecific male (e.g. Wilson et al., 1994). Moreover, individual behavioural compatibility between partners could have an important role in mate choice (e.g. Trivers, 1972; Burley, 1983; Barlow, 1992; Shuett et al., 2010), supporting the observations recorded in the present assay. This scenario can be an advantage if similar partners are able to coordinate their behaviour, positively affecting parental care and foraging in cases of predation risk (Shuett et al., 2010), as *M. lusitanicus* and *M. duodecimcostatus* in nature (Cotilla and Palomo, 2007; Mira and Mathias, 2007). This behavioural assortative mate preference has been shown to contribute to the premating reproductive isolation of sister species (e.g. Bolnick and Kirkpatrick, 2012; Mavarez et al., 2006; Rundle and Schluter, 1998; Grant and Grant, 2008; Kozak et al., 2011).

We cannot, however, exclude other possibilities that could explain the present observations, such as female receptivity, male mate choice and environmental conditions. Since both females successfully mated, female receptivity, which is very important in mating behaviour (e.g. Zinck and Lima, 2013**),** and environmental conditions were clearly not factors to consider here**.** Conversely, mutual mate choice, that seems to be common in monogamous species (e.g. Andersson, 1994; Amundsen, 2000; Nolan et al., 2010), such as *M. lusitanicus* and *M. duodecimcostatus*, could have affected male sexual behaviour indirectly, since the *M. duodecimcostatus* male could be simply not sexually interested in the available subject females and preferred to remain sexually naïve.

Methodologically, this work supports the use of the analysed microsatellite loci for *M. lusitanicus* and *M. duodecimcostatus* parentage testing.

Pondering the present results we also hypothesize that hybridization is a rare event in nature not because of behavioural isolation barriers, but probably due to an infrequency of syntopic populations, averting interspecific encounters. Otherwise, more cases of potential hybrids would have been found in earlier studies (Bastos-Silveira et al. 2012). Fieldwork in the Spanish sympatry area, complementing the extant sampled Portuguese locations, could help to clarify this hypothesis.

4.1.7 Compliance with ethical standards

Authors declare no conflicts of interest. Procedures were performed in accordance with the guidelines set by Portuguese (nº. 113/2013 Decree-law) and European (nº. 63/2010/CE Directive) legislation. Animals were handled by a Federation of Laboratory Animal Science Associations category C licensed biologist (M. A. Duarte).

4.1.8 References

Amundsen T (2000) Why are female birds ornamented? Trends Ecol Evol 15(4):149–155

Andersson M (1994) Sexual selection. Princeton University Press, Princeton

Barlow GW (1992) Is Mating Different in Monogamous Species? The Midas Cichlid Fish as a Case Study. Amer Zool 32**:**91–99

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

Bolnick DI, Kirkpatrick M (2012) The relationship between intraspecific assortative mating and reproductive isolation between divergent populations. Curr Zool 58: 484–492

Braaker S, Heckel G (2009) Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole *Microtus arvalis*. Mol Ecol 18:2518–2531

Brunet-Lecomte P, Chaline J (1991) Morphological evolution and phylogenetic relationships of the European ground voles (Arvicolidae, Rodentia). Lethaia 24(1):45–53

Burley N (1983) The meaning of assortative mating. Ethol Sociobiol 4:191–203

Carere C, Maestripieri D (2013) Animal personalities: behavior, physiology and evolution. Chicago University Press, Chicago

Cotilla I, Palomo LJ (2007) *Microtus duodecimcostatus* (de Selys-Longchamps, 1839). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 422–425

Dall SRX, Houston AI, McNamara JM (2004) The behavioural ecology of personality: consistent individual differences from an adaptive perspective. Ecol Lett 7:734–739

Dingemanse NJ, Réale D (2005) Natural selection and animal personality. Behaviour 142:1165–1190.

Drent PJ, van Oers K, van Noordwijk AJ (2003) Realized heritability of personalities in the great tit (*Parus major*). Proc R Soc Lond [Biol] 270:45–51

Duarte MA, Mathias ML, Bastos-Silveira C (2015) Pair-bonding behaviour of the sister species *Microtus lusitanicus* and *M. duodecimcostatus*. J Ethol 33:213–223

Gosling SD (2001) From mice to men: What can we learn about personality from animal research? Psychol Bull 127:45–86

Gosling SD, John OP (1999) Personality dimensions in nonhuman animals: A cross-species review. Curr Dir Psichol Sci 8:69–75

Grant PR, Grant BR (2008) Pedigrees, assortative mating and speciation in Darwin's finches. Proc R Soc Lond [Biol] 275:661–668

Guédon G, Pascal M, Mazouin F (1991a) Le campagnol provençal en captivité (*Pitymys duodecimcostatus* de Sélys-Longchamps, 1839) (Rongeurs, Microtidés) I. La reproduction. Mammalia 55:97–106

Guédon G, Pascal M, Mazouin F (1991b) Le campagnol provençal en captivité (*Pitymys duodecimcostatus* de Sélys-Longchamps, 1839) (Rongeurs, Microtidés) II. La croissance. Mammalia 55:397–406

Guédon, G., Pascal, M., 1993. Dynamique de population du campagnol provençal (*Pitymys duodecimcostatus* de Selys-Longchamps, 1839) dans deux agrosystémes de la région montpelliéraine. Rev Ecol - Terre Vie 48:375–398

Heske EJ, Ostfeld RS (1990) Sexual dimorphism in size, relative size of testes, and mating systems in North American voles. J Mammal 71(4):510–519

Ishibashi Y, Saitoh T, Abe S, Yoshida MC (1995) Polymorphic microsatellite DNA markers in the grey red-backed vole *Clethrionomys rufocanus bedfordiae*. Mol Ecol 4:127–128

Ishibashi Y, Yoshinaga Y, Saitoh T et al. (1999) Polymorphic microsatellite DNA markers in the field vole *Microtus montebelli*. Mol Ecol 8:163–164

Jaarola M, Ashford RT, Ratkiewicz M, Brunhoff C, Borkowska A (2007) Isolation and characterization of polymorphic microsatellite loci in the field vole, *Microtus agrestis*, and their cross-utility in the common vole, *M. arvalis*. Mol Ecol Notes 7:1029–1031

Koolhaas JM, Korte SM, De Boer, SF, Van Der Vegt BJ, Van Reenen CG, Hopster H, De Jong IC, Ruis MAW, Blokhuis HJ (1999) Coping styles in animals: current status in behavior and stress-physiology. Neurosci Biobehav Rev 23:925–935

Kozak GM, Head ML, Boughman JW (2011) Sexual imprinting on ecologically divergent traits leads to sexual isolation in sticklebacks. Proc R Soc Lond [Biol] 278: 2604–2610

Madureira ML (1982) New data on the taxonomy of portuguese *Pitymys*. Arq Mus Bocage A I:341–368

Mavárez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M (2006) Speciation by hybridization in *Heliconius* butterflies. Nature 441:868–871

Mira A (1999) Ecologia do rato-cego-mediterrânico, *Microtus* (*Terricola*) *duodecimcostatus* (Rodentia, Muridae), em pomares de citrinos. Dissertation, Faculdade de Ciências, Universidade de Lisboa

Mira A, Mathias ML (2007) *Microtus lusitanicus* (Gerbe, 1879). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 418–421

Nolan PM, Dobson FS, Nicolaus M, Karels TJ, McGraw KJ, Jouventin P (2010) Mutual mate choice for colorful traits in king penguins. Ethology 116:635–644

Paradis E, Guédon G (1993) Demography of a Mediterranean microtine: the Mediterranean pine vole, *Microtus duodecimcostatus*. Oecol 95:47–53

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. Biol Rev 82:291–318

Rundle HD, Schluter D (1998) Reinforcement of stickleback mate preferences: Sympatry breeds contempt. Evolution 52:200–208

Sambrook J, Fritschi EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Santos SML (2009) Factors influencing the distribution of the lusitanian and the mediterranean pine voles (*Microtus lusitanicus* and *M. duodecimcostatus*) in Portugal: a multiscale approach. Dissertation, Faculdade de Ciências, Universidade de Lisboa

Santos SM, Lourenço RF, Mathias ML, Mira AP (2010) Spatial and temporal ecology of the Lusitanian pine vole (*Microtus lusitanicus*) in a Mediterranean polyculture. Anim Biol 60:209–227

Schuett W, Tregenza T, Dall SRX (2010) Sexual selection and animal personality. Biol Rev 85:217–246

Sih A, Bell AM, Johnson JC, Ziemba RE (2004) Behavioral syndromes: an integrative overview. Q Rev Biol 79:241–277

Soares JMA (2013) Fitness dos híbridos entre duas espécies de ratos fossadores, *Microtus lusitanicus* (Gerbe, 1879) e *M. duodecimcostatus* (de Selys-Longchamps, 1839) - uma abordagem multidisciplinary. Dissertation, Faculdade de Ciências da Universidade de Lisboa

Trivers RL (1972) Parental investment and sexual selection. In: Campbell B (ed) Sexual selection and the descent of the man 1871-1971. Aldine Publishing Company, Chicago, pp 136–197

Ventura J, Jiménez L, Gisbert J (2010) Breeding characteristics of the Lusitanian pine vole, *Microtus lusitanicus*. Anim Biol 60:1–14

Walser B, Heckel G (2008) Microsatellite markers for the common vole (*Microtus arvalis*) and their cross-species utility. Conserv Genet 9:479–481

Wiking VH (1976) Karyologie und Biologie der beiden iberischen Wuhlmausarten *Pitymys mariae* und *Pitymys duodecimcostatus*. Z Zool Syst Evol 14:104–129

Wilson DS, Clark AB, Coleman K, Dearstyne T (1994) Shyness and boldness in humans and other animals. Trends Ecol Evol 9:442–446

Zinck L, Lima SQ (2013) Mate choice in *Mus musculus* is relative and dependent on the estrus state. PLoS ONE 8(6):e66064

4.2 Pair bonding behaviour

Type of publication: Article

Reference: Duarte MA[1,2,3], Mathias ML[2,3], Bastos-Silveira C[1] (2015) Pair-bonding behaviour of the sister species *Microtus lusitanicus* and *M. duodecimcostatus*. J Ethol 33:213–223. DOI: 10.1007/s10164-015-0434-8.

[1] Museu Nacional de História Natural e da Ciência, Departamento de Zoologia e Antropologia, Universidade de Lisboa, Rua da Escola Politécnica, 58, Lisbon 1250-102, Portugal.

[2] Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[3] Centro de Estudos de Ambiente e Mar, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

4.2.1 Abstract

Monogamous mating systems can be inferred by analysis of social behaviour, for example pair-bonding. We studied the Iberian sister species *Microtus lusitanicus* and *Microtus duodecimcostatus*. On the basis of morphological, reproductive, and ecological characteristics, but not behavioural studies, these voles are regarded as monogamous species. Pair-bonding behaviour was inferred by use of partner preference tests using chemical stimuli. Two scenarios were considered: in the first we examined whether the member of a breeding pair prefers chemical stimuli from its partner or from a sexually experienced stranger, simulating a widow/widower of the population, while in the second scenario we assessed whether there was a preference for chemical stimuli from its partner or from a sexually naive stranger, mimicking an immigrant individual. Results support a social monogamous mating system for both species and reveal a significant female preference for the male partner, rather than a stranger. Conversely, male preference differed, depending on the sexual status of the female strangers. When the stranger was sexually experienced, a significant preference for its partner was observed, whereas in the presence of a sexually naive female stranger no significant preference for the female partner was revealed. These results suggest rare male extra-pair mating in *M. lusitanicus* and *M. duodecimcostatus*.

4.2.2 Keywords

Monogamy; Partner preference; Pair-bonding; Male extra-pair mating; *Microtus*; Iberian voles.

4.2.3 Introduction

Monogamy is an exclusive socio-sexual relationship exhibited by the members of a stable breeding pair which share common territory and parental duties. The bonded pair usually remains together throughout several breeding seasons, covering most of their lifespan, although the death of one member may lead to the development of a new pair- bond with another conspecific individual (Kleiman 1977). Monogamous species are characterized by reduced physical or behavioural sexual dimorphism, low reproductive potential, delayed sexual maturation of the young in the presence of the parents (breeding pair), juvenile assistance with rearing of younger siblings, socio-sexual interactions after pair-bond formation, and mating preferences towards the pair mate (Kleiman 1977).

In a scenario in which group living is favoured over a solitary existence, monogamy is advantageous, because less energy is required for sexual or social interactions (Kleiman 1977; Lukas and Clutton-Brock 2013). In addition, monogamous breeding enhances the ability to defend territory containing scarce and valuable resources, for example food and resting areas, and reduces susceptibility to predation (Alexander 1974; Kleiman 1977).

Advances in molecular techniques have enabled biologists to distinguish between a purely social monogamous mating system and genetic monogamy (reviewed by Reichard 2003 and MacManes 2013). Social monogamy is solely a social living arrangement of a female and male characterized by a pair-bond independent of courtship and copulation and by a socio-sexual relationship that does not preclude the possibility of extra-pair mating by either sex. On the other hand, genetic monogamy is characterized by exclusive parentage, leading to the absence of extra-pair offspring (identified by genotyping all the litters produced by a breeding pair). This genetic outcome is not exclusive to genetic monogamy,

because it is also observed for some social monogamous species (Brotherton et al. 1997). Furthermore, the behaviour of strongly pair-bonded animals can vary substantially, including extra-pair mating or genetic promiscuity (Dixon et al. 1994; Mulder et al. 1994; MacManes 2013).

Among mammals monogamy is rare (3–5 %) (Dewsbury 1987) and primarily described for canids, primates, and rodents (Kleiman 1977; Clutton-Brock 1989). *Microtus ochrogaster*, a monogamous vole, has been thoroughly studied with regard to the neurobiology of pair-bonding (reviewed by Young and Wang 2004 and Young et al. 2008, 2011). *M. ochrogaster* males contribute to parental behaviour, for example nest building and guarding, huddling, and retrieval of wandering pups (Thomas and Birney 1979; Gruder-Adams and Getz 1985; Getz and Carter 1996), although the frequency and duration of this behaviour are typically lower than for females (Solomon 1993; McGuire and Getz 2012). Moreover, the monogamous behaviour of *M. ochrogaster* is observed in both nature (Getz et al. 1981; Carter and Getz 1993) and captivity (Williams et al. 1992; Solomon 1993), which enables researchers to infer whether pair-bonded animals recognize and choose their partner over unfamiliar conspecifics. By use of two-way choice tests performed after mating or co-habitation, Williams et al. (1992) observed that *M. ochrogaster* spent significantly more time with its mate, rather than a conspecific stranger, indicating a preference for the partner. This behaviour is referred to as partner preference, and the assays as partner preference tests (PPT).

Similarly to *M. ochrogaster*, monogamous behaviour is also observed for the voles *M. pinetorum* and *M. kikuchii*. Partner preference, aggression toward strangers (Parker et al. 2001; Back et al. 2002), paternal care (Oliveras and Novak 1986), and family cohabitation with cooperative breeding (FitzGerald and Madison 1983; Powell and Fried 1992) are observed for *M. pinetorum*. Home range overlap for both sexes and indications of reproductive exclusiveness are observed for *M. kikuchii* (Wu et al. 2012). Nevertheless, monogamy is not a behavioural trait shared by all *Microtus sp*., because a variety of mating systems have been characterised for microtines, ranging from genetic monogamy (*M. pinetorum*, Marfori et al. 1997),

through social monogamy (*M. ochrogaster*, Solomon et al. 2004, Ophir et al. 2008; *M. kikuchii*, Wu et al. 2012) to genetic promiscuity for most species (e.g. *M. oeconomus,* Tast 1966; *M. pennsylvanicus*, Berteaux et al. 1999; *Microtus arvalis*, Fink et al. 2006). Because patterns of nonsocial behaviour are similar for monogamous and non-monogamous voles (Wolff 1985), there is much potential for comparative studies on their social behaviour, for example pair-bonding.

The Lusitanian pine vole *M. lusitanicus* Gerbe (1879) and the Mediterranean pine vole *M. duodecimcostatus* de Selys-Longchamps (1839), are two semi-fossorial sister species that inhabit the Iberian Peninsula, living both in allopatry and sympatry (Santos 2009). Syntopy was observed in one sympatric location (Vale Vaqueiros, Portalegre, Portugal). A balanced sex ratio (Paradis and Guédon 1993), reduced litter size (1–5pups) (Guédon et al. 1991a, b) and K selection strategy (Guédon et al. 1991b; Guédon and Pascal 1993; Ventura et al. 2010) are observed for natural populations of *M. lusitanicus* and *M. duodecimcostatus*. Findings among captive animals were similar (personal observations). Spatial overlap and similar home range for both sexes have been observed for *M. lusitanicus*, with restricted daily movement observed for sexually active males (Santos et al. 2010). *M. lusitanicus* voles live in small groups composed of a breeding pair and its pups, and nests shared between males and females or between one female and one sub-adult (Mira and Mathias 2007; Santos 2009). *M. duodecimcostatus* is also socially organized in small family groups (Paradis and Guédon 1993; Cotilla and Palomo 2007) and, similarly to *M. lusitanicus* (Madureira 1982; Heske and Ostfeld 1990; Ventura et al. 2010), is sexually monomorphic in terms of adult weight (Mira 1999), with adult males having relatively small testes (Montoto et al. 2011). Hence, these morphological, reproductive, and ecological characteristics suggest that monogamous mating systems might occur for both species.

In this study, the pair-bonding behaviour of *M. lusitanicus* and *M. duodecimcostatus* was inferred, for the first time, by conducting PPT using chemical stimuli on laboratory-established breeding pairs. Two scenarios were considered: in the first we examined whether the female or male of a breeding pair prefers its

partner or a sexually experienced stranger, simulating a widow/widower in the population, while in the second scenario we assessed whether the individual preferred its partner or a sexually naive stranger, mimicking an immigrant individual. On the basis of their morphological, reproductive, and ecological characteristics we hypothesized that preference for the partner would be observed for both species. Our PPT findings shed light on the type of monogamy of these semi-fossorial species.

4.2.4 Materials and Methods

The voles used in this study were first-generation individuals selected from an animal facility colony (Faculty of Sciences of the University of Lisbon) established by wild-caught individuals from Portalegre (Portugal). Animals were kept under controlled temperature (22 ± 2 °C) and photoperiod (12:12 h light:dark cycle, with light available from 7:00 to 19:00) conditions in Makrolon polycarbonate type II (single individuals) and type III (breeding pairs and adult same sex litters) cages containing >4 cm of wood shavings. Hay, pine cones, and small branches were also provided for environmental enrichment. The diet consisted of carrots and apples, available ad libitum.

Twelve breeding pairs from each species were established from unrelated individuals and maintained together until the end of their PPT. Each breeding pair had produced at least one litter before the tests and was not rearing a litter during the PPT. Four scenarios were considered for each breeding pair: the female or male choosing between its partner and a sexually experienced stranger and the female or male choosing between its partner and a sexually naive stranger. Forty-eight individuals per species were used in these behavioural assays, with each animal being tested once a day only. From the 24 females and 24 males, per species, 12 were individuals from the established breeding pairs (also used as sexually experienced strangers, after their PPT as test animals, to reduce the total number of animals) and 12 were sexually naive individuals, maintained in groups of adult siblings of the same sex and litter (sexually naive strangers).

PPT were conducted as two-way choice tests, with use of chemical stimuli. We adapted original procedures, because *M. lusitanicus* and *M. duodecimcostatus* are very sensitive to constraining devices (personal observations), for example being tethered in separate boxes, in contrast with other *Microtus* voles (Williams et al. 1992; Berteaux et al. 1999; Aragona and Wang 2004; Ahern et al. 2009). Tests were conducted from 11:00 to 17:00, when both species are active (Madureira 1984), in a separate room with similar temperature and photoperiod conditions as the colony room. The experimental apparatus consisted of a Y-shaped glass olfactometer, composed of a "start box" (30 x 20 x 15 cm), connected to the main arm (35 cm) of a Y-shaped glass tube (5 cm diameter). Each secondary arm (30 cm) was connected to a "stimulus box" (20 x 20 x 15 cm), containing the chemical stimuli from the stimulus individual (partner or experienced or naive stranger). These chemical stimuli (e.g., urine and faeces) were collected over 1 h, by placing the stimulus individual in its "stimulus box". When one of the stimulus individuals excreted substantially more urine and faeces than the other the PPT was aborted and repeated later. Meanwhile, each test animal was allowed a 5-min habituation period, in the "start box". After chemical stimuli collection, each stimulus individual was returned to its cage and "stimulus boxes" were attached to the Y-tube. The "start box", with the test animal, was then connected to the Y-tube and the PPT began as soon as the vole entered the Y-tube, and lasted 30 min. An individual was only considered on one particular side of the apparatus when it crossed the entrance of the secondary arm with its head.

The duration of our PPT was determined after conducting trials, for both species and sexes, because these voles are very sensitive and easily stressed, which affects the welfare of the animals and partner preference results. We performed 10, 30, and 60-min trials. Some test animals stayed in the "start box" for a few minutes, so 10-min tests were insufficient to reveal partner preference. In contrast, 60-min trials were too long, because the animals became stressed or lethargic, depending on the individual, after approximately 40–45 min in the apparatus. Hence, we chose the 30-min duration for our *M. lusitanicus* and *M.*

duodecimcostatus PPT, because trials of this length gave the test animals enough time to reveal its partner preference, without stress or taking long naps.

All PPT were video-recorded (Canon Legria HF S200) for later analysis. Two variables, time spent in each stimulus secondary arm and respective "stimulus box", and time spent investigating chemical stimuli, by sniffing or licking the floor of the "stimulus box", were measured, in seconds, by a single observer by use of a stopwatch. If the test individual failed to enter the Y-tube spontaneously after 5 min the PPT was aborted and repeated later. Alternation of the right and left position of the partner stimuli between tests was used to control laterality. After each PPT, the entire apparatus (Y-tube and boxes) was washed thoroughly with water then 96 % ethanol solution and dried.

Control tests were performed to determine the effect of other natural behaviours (e.g., exploration of the apparatus, grooming, napping, trying to escape), on the measured variables. Moreover, responses to familiar (partner) and unfamiliar (stranger) chemical stimuli were also assessed separately, to enable understanding of whether test animals were actually preferring the partner or rejecting and/or avoiding the chemical stimuli of the stranger. Six categories of control test were conducted: the female or male choosing between its partner and a blank box; the female or male choosing between a sexually experienced stranger and a blank box; and the female or male choosing between a sexually naive stranger and a blank box. The blank box consisted of a clean "stimulus box", without any chemical cues. These control tests were performed by use of four individuals randomly selected from the different breeding pairs, 1 week after their PPT, in accordance with the ethical principles of the Federation of European Laboratory Animal Science Associations (Guillen 2012), which encourage, when possible, reduction of the number of animals used in scientific research. Laterality was controlled by alternating the left or right position of the stimulus (versus the blank) between control tests.

Procedures were performed in accordance with the guidelines set by Portuguese (no. 113/2013 Decree-law) and European (no. 63/2010/CE Directive)

legislation. Voles were maintained and tested by a Federation of Laboratory Animal Science Associations category C licensed biologist (MA Duarte).

Sexual preference was first assessed by use of an estimator (R) which was used to compare the log-ratio of time spent in contact with the partner versus sexually experienced or naive stranger chemical stimuli (Ganem et al. 2008). R indicates the direction of a choice, with positive values indicating a preference for partner stimuli whereas negative values indicate a preference for the stimuli of the stranger.

A paired-sample two-tailed t test was also performed, using IBM SPSS Statistics version 22, to analyse mean differences between time spent in each stimulus secondary arm and "stimulus box" and time spent investigating chemical stimuli. Tests were significant at the $p<0.05$ level.

4.2.5 Results

A total of 94 PPT were recorded and analysed, 48 PPT on *M. lusitanicus* and 46 on *M. duodecimcostatus*. Forty-eight control PPT were conducted, 24 on each species (**Table 1**). Only 2 PPT were not performed, because of the sudden death of a female *M. duodecimcostatus*.

Control tests, for both species and sexes, revealed a significant preference for sniffing and/or licking stimuli deposited in a "stimulus box", irrespective its origin (partner or stranger), rather than an empty box (control blank "stimulus box") (**Table 1**; **Fig. 1**). Only for control tests of female *M. duodecimcostatus* choosing between a sexually naive male stranger and a blank control did the difference not reach significance ($p = 0.061$, **Table 1**). Overall, our results suggest that *M. lusitanicus* and *M. duodecimcostatus* do not avoid the chemical stimuli of strangers, either from a sexually experienced or sexually naive stranger (**Table 1**; **Fig. 1**). When considering the total time spent in the respective side of the Y-olfactometer (arm and "stimulus box") (**Table 1**; **Fig. 1**), control tests did not reveal significant preference for the chemical stimuli, except for *M. lusitanicus* (**Table 1**).

Table 1 - Results of partner preference tests for *M. lusitanicus* (ML) and *M. duodecimcostatus* (MD), and the respective control tests. Sex, type of test and number of animals tested are indicated. Mean values and standard deviations are presented in seconds, for both variables: total time spent in the partner/stranger/blank arm+box (total time) and total time sniffing the partner/stranger chemical stimuli present in its "stimulus box" (sniffing time). The R estimator and the overall partner preference (OPP, i.e. number of tests revealing partner preference per total number of tests) are also given. Significance was accomplished at $p < 0.05$ level. F=female, M=male, MP = male partner, FP = female partner, EMS = sexually experienced male stranger, EFS = sexually experienced female stranger, NMS = sexually naive male stranger, and NFS = sexually naive female stranger.

Species	Test	Number of test animals	Total time					Sniffing time				
			Partner	Stranger or blank	*p* value	*R* estimator	OPP	Partner	Stranger or blank	*p* value	*R* estimator	OPP
ML	F: MP × EMS	12	679 ± 348	457 ± 324	0.127	0.544 ± 0.982	9/12	57 ± 29	14 ± 8	0.000	1.378 ± 0.523	12/12
	M: FP × EFS	12	692 ± 278	362 ± 187	0.010	0.671 ± 0.647	11/12	69 ± 33	22 ± 8	0.000	1.076 ± 0.351	12/12
	F: MP × NMS	12	679 ± 384	407 ± 210	0.085	0.486 ± 1.045	8/12	54 ± 15	16 ± 7	0.000	1.221 ± 0.305	12/12
	M: FP × NFS	12	501 ± 322	394 ± 237	0.393	0.188 ± 0.795	4/12	45 ± 21	45 ± 28	0.990	0.027 ± 0.629	7/12
MD	F: MP × EMS	11	569 ± 165	504 ± 167	0.471	0.135 ± 0.525	7/11	86 ± 30	51 ± 19	0.000	0.545 ± 0.294	11/11
	M: FP × EFS	12	656 ± 235	488 ± 147	0.057	0.267 ± 0.517	8/12	79 ± 44	48 ± 39	0.002	0.537 ± 0.334	12/12
	F: MP × NMS	11	611 ± 255	469 ± 209	0.228	0.291 ± 0.621	7/11	89 ± 40	51 ± 24	0.001	0.563 ± 0.315	11/11
	M: FP × NFS	12	640 ± 369	437 ± 237	0.232	0.371 ± 1.070	7/12	70 ± 30	49 ± 31	0.070	0.463 ± 0.694	12/12
ML (controls)	F: MP × blank	4	773 ± 186	312 ± 227	0.037	1.248 ± 1.049	4/4	49 ± 24	5 ± 4	0.041	2.106 ± 0.743	4/4
	M: FP × blank	4	670 ± 334	245 ± 83	0.060	0.935 ± 0.416	4/4	89 ± 41	12 ± 8	0.027	2.125 ± 0.650	4/4
	F: EMS × blank	4	803 ± 591	281 ± 342	0.228	1.284 ± 1.236	3/4	35 ± 20	3 ± 4	0.043	2.221 ± 0.627	4/4
	M: EFS × blank	4	534 ± 214	327 ± 67	0.100	0.445 ± 0.334	4/4	101 ± 25	6 ± 5	0.003	2.796 ± 0.498	4/4
	F: NMS × blank	4	599 ± 350	337 ± 263	0.046	0.690 ± 0.304	4/4	40 ± 16	2 ± 1	0.019	2.496 ± 0.631	4/4
	M: NFS × blank	4	377 ± 73	301 ± 60	0.055	0.228 ± 0.141	4/4	72 ± 8	3 ± 2	0.000	1.414 ± 0.445	4/4
MD (controls)	F: MP × blank	4	431 ± 227	327 ± 90	0.349	0.185 ± 0.410	2/4	95 ± 37	8 ± 3	0.014	2.341 ± 0.168	4/4
	M: FP × blank	4	855 ± 299	327 ± 180	0.109	1.042 ± 0.955	4/4	73 ± 22	5 ± 2	0.007	2.624 ± 0.430	4/4
	F: EMS × blank	4	758 ± 495	297 ± 102	0.151	0.798 ± 0.670	4/4	85 ± 39	11 ± 6	0.022	1.974 ± 0.103	4/4
	M: EFS × blank	4	902 ± 264	397 ± 295	0.129	1.100 ± 1.195	4/4	104 ± 47	5 ± 2	0.022	2.862 ± 0.226	4/4
	F: NMS × blank	4	578 ± 425	193 ± 97	0.135	0.968 ± 0.649	4/4	94 ± 63	7 ± 4	0.061	2.469 ± 0.390	4/4
	M: NFS × blank	4	804 ± 411	298 ± 51	0.074	0.892 ± 0.450	4/4	94 ± 21	6 ± 3	0.003	2.693 ± 0.346	4/4

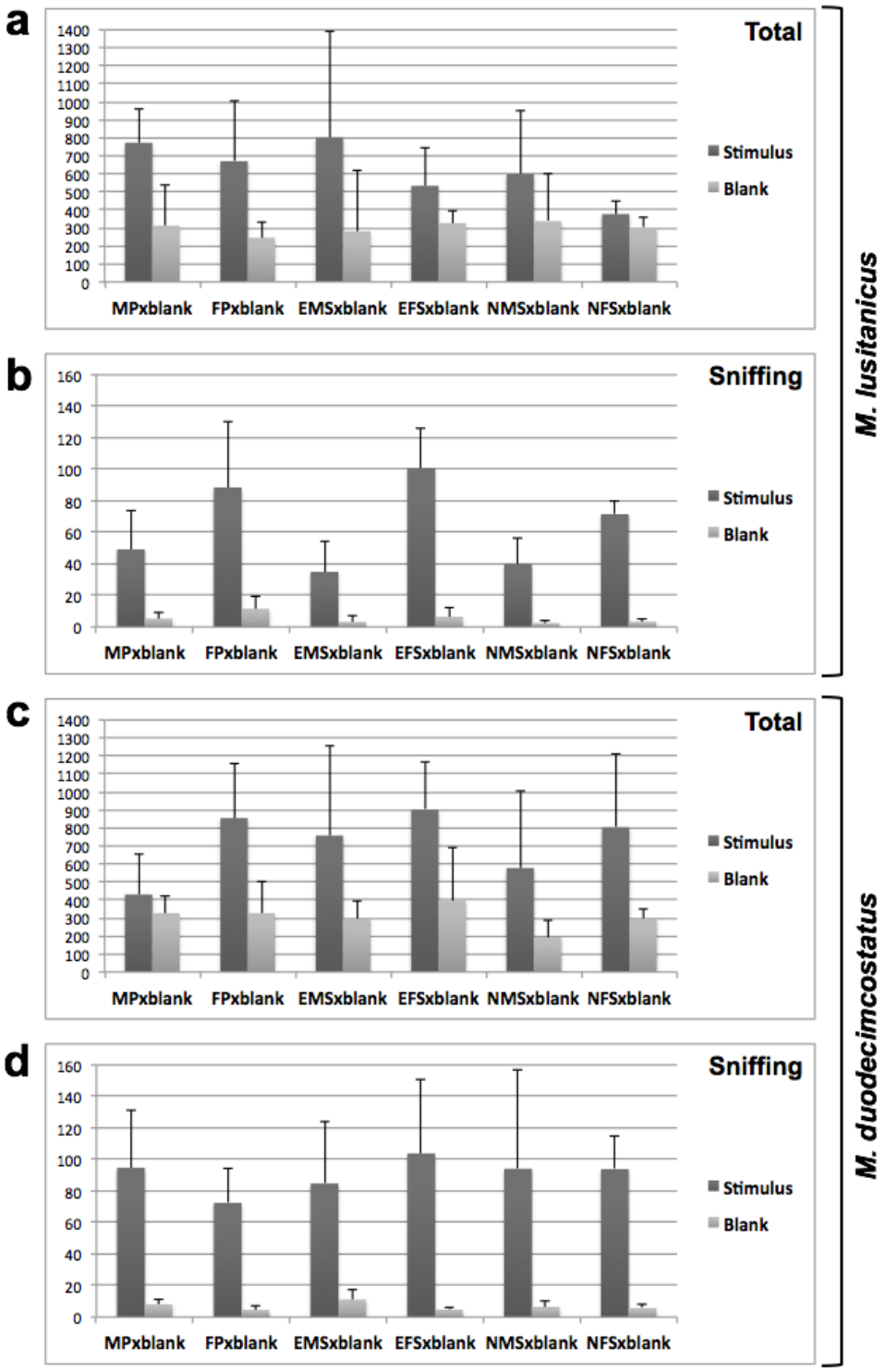


Figure 1 – Results of the control tests. Mean total time spent in the stimulus arm and box (a, c) and mean total time sniffing the "stimulus box" versus a blank (b, d) are indicated for *M. lusitanicus* (a, b) and *M. duodecimcostatus* (c, d). Standard deviations are also shown. MP = male partner, FP = female partner, EMS = sexually experienced male stranger, EFS = sexually experienced female stranger, NMS = sexually naive male stranger, NFS = sexually naive female stranger.

This result can be explained by other behaviours, not related to partner preference, which were detected sporadically on both sides of the Y-olfactometer, for example extensive exploration of the apparatus (these voles are very curious regarding new spaces), grooming, short naps, and trying to escape the apparatus by jumping in the corners and near the walls of the "stimulus box".

Results from *M. lusitanicus* and *M. duodecimcostatus* PPT were congruent for the control tests (**Table 1**; **Fig. 2**), demonstrating the absence of significant preference for the partner chemical stimuli in terms of the total time spent in the respective side of the Y-olfactometer (arm and "stimulus box" variable; **Table 1** and **Fig. 2**). The exception was the choice of *M. lusitanicus* males between the female partner and a sexually experienced female stranger ($p = 0.010$, **Table 1**).

In contrast, the total time spent sniffing the "stimulus box" of the partner was significant for females of both species ($p < 0.001$, **Table 1**), whereas male preference varied according to the sexual status of the female stranger. When male *M. lusitanicus* and *M. duodecimcostatus* had to choose between their female partner and a sexually experienced female stranger, males had a significant preference for their partner ($p < 0.002$, **Table 1**), similarly to females. However, when males had to choose between their female partner and a sexually naive female stranger, no significant partner preference was observed (**Table 1**).

4.2.6 Discussion

Use of modified PPT enabled inference of pair-bonding behaviour in Iberian sister species *M. lusitanicus* and *M. duodecimcostatus*. Our results show there is a significant preference by both voles for the chemical cues of the partner. The only exception was when male individuals had to choose between their partner and a sexually naive female, when no significant partner preference was observed. These observations partially confirm our hypothesis and support a social monogamous mating system for these voles, with the possibility of rare male extra-pair mating.

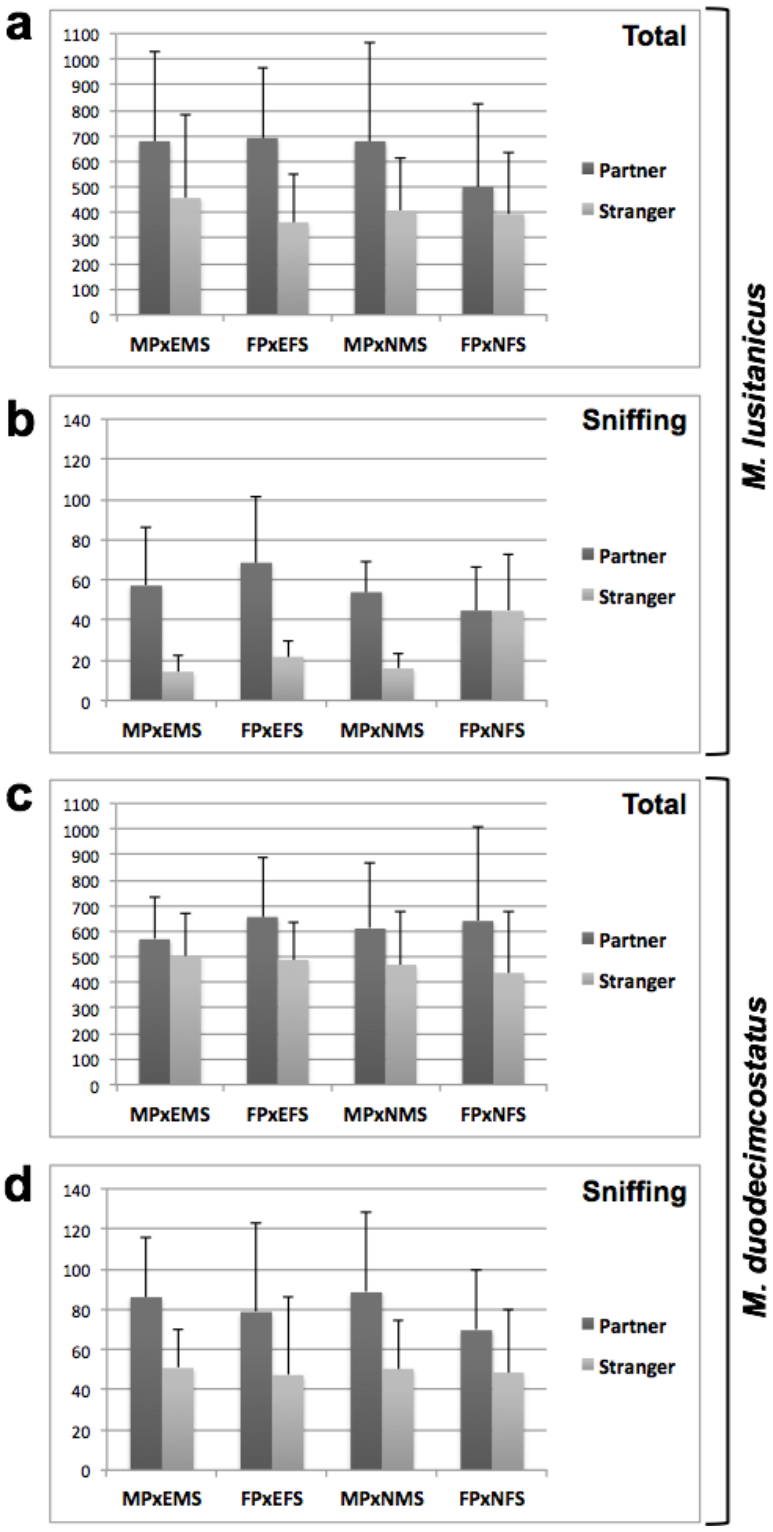


Figure 2 – Results of the partner preference tests. Mean total time spent in the stimulus (partner or stranger) arm and box (a, c) and mean total time sniffing the "stimulus box" (b, d), are indicated for *M. lusitanicus* (a, b), and *M. duodecimcostatus* (c, d). Standard deviations are also shown. MP = male partner, FP = female partner, EMS = sexually experienced male stranger, EFS = sexually experienced female stranger, NMS = sexually naive male stranger, NFS = sexually naive female stranger.

Although social monogamy is characterized by a socio-sexual relationship, pair-bonded females or males can mate with extra-pair individuals (Bishop et al. 2007; Munshi-South 2007; Borkowska et al. 2009; Cohas and Allainé 2009; Barelli et al. 2013). This behaviour has been observed among *Microtus* voles *M. ochrogaster* (Solomon et al. 2004; Ophir et al. 2008) and *M. kikuchii* (Wu et al. 2012).

We hypothesise that by engaging in extra-pair mating, *M. lusitanicus* and *M. duodecimcostatus* males may have a chance to increase their reproductive success if an extra-pair sexually naive female conceives. Our results support the male extra-pair mating hypothesis, because of the absence of significant partner preference for partner chemical cues when a sexually naive female is available. In nature, the frequency of extra-pair mating may be directly correlated with population density (Say et al. 1999; Dean et al. 2006; Bryja et al. 2008), possibly because of increased food supply. Variation in population density has been described for both species (Cotilla and Palomo 2007; Mira and Mathias 2007); we therefore suggest genotyping and parentage analysis of wild-caught animals from populations with different densities, to determine whether this affects the extent of male extra-pair mating by *M. lusitanicus* and *M. duodecimcostatus*. An alternative hypothesis could explain male preference, however. It is possible that, occasionally, males prefer chemical cues of sexually naive females because of their higher levels of ovarian oestradiol compared with recently mated or pregnant partner females (Carter and Getz 1985). It is also known that some sexually naive females can be induced into oestrus when physically disturbed (e.g., moved; Richmond and Conaway 1969), a situation that could have occurred in this study, enabling attraction of some of the pair-bonded males through their oestrus odour cues.

Social monogamous species are also characterized by shared use of territory between the members of a bonded breeding pair. Concurrently with this observation, Santos et al. (2010) observed home range overlap between females and males during a spatial and temporal ecology study on *M. lusitanicus* during the breeding season, supporting social monogamy as the mating system for this

species. Unfortunately, territory use is currently unknown for *M. duodecimcostatus*. Under laboratory conditions the behaviour of both species is coherent with social monogamy, because paternal care and juvenile assistance in rearing younger siblings was observed (personal observations). Also, female intolerance of strangers was witnessed for both naive and experienced individuals, through aggression and rejection (personal observations), behaviour already described as part of social monogamy (Lukas and Clutton-Brock 2013).

As mentioned above, these sister species are semi-fossorial, constructing complex underground burrow systems (Vericad 1970; Soriguer and Amat 1980; Borghi 1992; Giannoni 1994). The subterranean niche has advantages, particularly protection from predators, e.g., *Tyto alba* (Cotilla and Palomo 2007; Mira and Mathias 2007). Nevertheless, energy costs are associated with their fossorial lifestyle, for example digging new burrows (Lovegrove 1989; Powell and Fried 1992; Ebensperger and Bozinovic 2000; Luna et al. 2002; Faulkes and Bennett 2013). We hypothesize that such constraints contributed to the social monogamous mating system of *M. lusitanicus* and *M. duodecimcostatus*, reducing the energy costs associated with searching for, copulating with, and protecting multiple mates, and with parental investment, leading to a K selection strategy.

PPT are conducted under laboratory conditions to indirectly assess the mating system of small, secretive animals, for example *M. lusitanicus* and *M. duodecimcostatus*. Our extensive experience working with these voles has revealed they are not easily trapped or monitored in nature, because of their semi-fossorial lifestyle, which has contributed to the small number of field records of their social behaviour. Live-trapping and radiotelemetry, performed by our workgroup on *M. lusitanicus*, was used to assess male and female home-range overlap and nest sharing (Santos et al. 2010). Although these methods have been widely performed on voles (Getz et al. 1981; Carter and Getz 1993; McGuire and Getz 2012), partner preference cannot be determined by use of such an approach. Thus, laboratory PPT, such as those performed in this study, are a viable alternative.

We demonstrated it is possible to perform PPT on *M. lusitanicus* and *M. duodecimcostatus* by using urinary and faecal chemical cues to infer pair-bonding behaviour. The decision to use this modified approach, instead of tethered or confined animals, was because these sister voles are easily stressed if constrained with a collar or confined to a "stimulus box" (personal observations). By using chemical cues alone we also controlled the effect of the behaviour of the stimulus animals, for example vocalizations and exacerbated movements, which could affect the choice made by the test animals as a result of restriction inside the Y-olfactometer apparatus. Nevertheless, acoustic and visual signals, and olfactory traits may be important in mate choice (reviewed by Chenoweth and Blows 2006 and Charlton 2013), but are unaccounted for in PPT based on chemical cues alone. Still, use of chemical cues alone has been validated for rodents (Christophe and Baudoin 1998; Smadja and Ganem 2002; Ganem et al. 2008; Cutrera et al. 2012), including *Microtus* species (Newman and Halpin 1988; Ferkin et al. 1997; Kruczek and Golas 2003), and now for *M. lusitanicus* and *M. duodecimcostatus*.

Other PPT have furnished results similar to those presented herein. Carr et al. (1979, 1980) showed for monogamous rats that pair-bonded females significantly prefer the odour of their male partner to that of a stranger, in contrast with polygamous rats (Carr et al. 1979), whereas male pair-bonded rats prefer the odour of a female stranger over that of their female partner (Carr 1980). Regarding voles, Newman and Halpin (1988) reported a significant preference of *M. ochrogaster* pair-bonded females for cues from their partner, irrespective of the sexual status of the animal stranger. Conversely, whereas males significantly preferred their female partner to a sexually experienced female, no significant preference could be distinguished when males were tested with their partner and sexually naive female odour cues (Newman and Halpin 1988). Moreover, DeVries and Carter (1999) revealed that female *M. ochrogaster* formed partner preferences more quickly than males, and that this preference was longer-lasting, indicating sexual dimorphism in the development and maintenance of social preferences, probably as a result of different reproductive strategies of the sexes.

On the basis of these results we foresee that, among natural populations of *M. lusitanicus* and *M. duodecimcostatus*, paired males may encounter sexually naive females, as simulated by the PPT. These females may be members of the population, members of a nearby population, or immigrants from a different population. The first two scenarios are more probable, because male-biased dispersal is more common among arvicolines (Le Galliard et al. 2012), nevertheless, the third scenario may be plausible if natal dispersal of *M. lusitanicus* and *M. duodecimcostatus* is non-sex-biased, similar to that of monogamous *M. ochrogaster* (McGuire and Getz 2012).

Our results, and these species' ecological, reproductive, and behavioural characteristics, support the existence of this mating system among both species. In fact, according to Lukas and Clutton-Brock (2013), such closely related species as *M. lusitanicus* and *M. duodecimcostatus*, usually have the same social system. These sister voles diverged very recently (60,000 years ago; Brunet-Lecomte and Chaline 1991), have low cytochrome b divergence (Jaarola et al. 2004), and have no mechanical or gametic barrier, because they can produce F1 hybrids under laboratory conditions (Wiking 1976; personal observations). Hybridization between *M. lusitanicus* and *M. duodecimcostatus* is, however, very rare in the wild (Bastos-Silveira et al. 2012), suggesting the existence of behavioural mechanisms of reproductive isolation. Thus, it is plausible to consider that the social monogamous mating system may also contribute to their reproductive isolation. We also believe that our findings can be extrapolated to natural populations, by considering the possibility of heterospecific encounters between a male and a sexually naive female, probably from a nearby population (e.g., syntopic locations). These encounters may challenge the pair-bond and copulation could exist outside the established breeding pair, enabling rare hybridization between the two species, in agreement with the results of Bastos-Silveira et al. (2012). Heterospecific PPT could be a means of testing this hypothesis, helping to reveal effect of extra-pair mating in the interspecific population dynamics of these two sister species.

4.2.7 Acknowledgments

M.A. Duarte was supported by Ph.D. grant SFRH/BD/70646/2010 from the Foundation for Science and Technology (FCT). We are grateful to Tiago Alves for assisting M.A. Duarte with maintenance of the animal facility colony and to Samuel J. Walker for helpful suggestions on the manuscript. This study was developed under the project PTDC/BIA-BEC/103729/2008 (FCT), coordinated by C. Bastos-Silveira. It was also supported by European Funds, through COMPETE, and by National Funds through the FCT, within project PEst-C/MAR/LA0017/2013.

4.2.8 References

Ahern TH, Modia ME, Burketta JP, Younga LJ (2009) Evaluation of two automated metrics for analyzing partner preference tests. J Neurosci Meth 182(2):180–188

Alexander RD (1974) The evolution of social behaviour. Ann Rev Ecol Syst 5:325–383

Aragona BJ, Wang Z (2004) The prairie vole (*Microtus ochrogaster*): an animal model for behavioral neuroendocrine research on pair bonding. ILAR J 45(1):35–45

Back SR, Beeler LA, Schaefer RL, Solomon NG (2002) Testing functional hypotheses for the behavior of resident pine voles, *Microtus pinetorum*, toward non-residents. Ethology 108(11):1023–1039

Barelli C, Matsudaira K, Wolf T, Roos C, Heistermann M, Hodges K, Ishida T, Malaivijitnond S, Reichard UH (2013) Extra-pair paternity confirmed in wild white-handed gibbons. Am J Primatol 75:1185–1195

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

Berteaux D, Bêty J, Rengifo E, Bergeron J-M (1999) Multiple paternity in meadow voles (*Microtus pennsylvanicus*): investigating the role of the female. Behav Ecol Sociobiol 45:283–291

Bishop JM, O'Ryan C, Jarvis JUM (2007) Social common mole-rats enhance outbreeding via extra-pair mating. Biol Lett 3(2):176–179

Borghi CE (1992) Los micromamíferos subterráeos del Pirineo Aragonés (Subgénero *Pitymys*): Impacto sobre la vegetación y papel bioerosivo. Ph.D. Thesis, Universidad Autónoma de Madrid, Madrid, pp 1–167

Borkowska A, Borowskib Z, Krysiukc K (2009) Multiple paternity in free-living root voles (*Microtus oeconomus*). Behav Process 82:211–213

Brotherton PNM, Pemberton JM, Komers PE, Malarky G (1997) Genetic and behavioural evidence of monogamy in a mammal, Kirk's dik-dik (*Madoqua kirkii*). Proc R Soc Lond B 264:675–681

Brunet-Lecomte P, Chaline J (1991) Morphological evolution and phylogenetic relationships of the European ground voles (Arvicolidae, Rodentia). Lethaia 24(1):45–53

Bryja J, Patzenhauerová H, Albrecht T, Mosansky L, Stanko M, Stopka P (2008) Varying levels of female promiscuity in four *Apodemus* mice species. Behav Ecol Sociobiol 63:251–260

Carr WJ, Demesquita-Wander M, Sachs SR, Maconi P (1979) Responses of female rats to odors from familiar versus novel males. Bull Psychon Soc 14:118–120

Carr WJ, Hirsch JT, Balazs JM (1980) Responses of male rats to odors from familiar vs novel females. Behav Neur Biol 29:331–337

Carter CS, Getz LL (1985) Social and hormonal determinants of reproductive patterns in the prairie vole. In: Giles R, Balthazart J (eds) Neurobiology. Springer-Verlag, Berlin, pp 18–36

Carter CS, Getz LL (1993) Monogamy and the prairie vole. Sci Am 268:100–106

Charlton BD (2013) Experimental tests of mate choice in nonhuman mammals: the need for an integrative approach. J Exp Biol 216:1127–1130

Chenoweth SF, Blows MW (2006) Dissecting the complex genetic basis of mate choice. Nat Rev Genet 7:681–692

Christophe N, Baudoin C (1998) Olfactory preferences in two strains of wild mice, *Mus musculus musculus* and *Mus musculus domesticus*, and their hybrids. Anim Behav 56(2):365–369

Clutton-Brock T (1989) Mammalian mating systems. P Roy Soc Lond B Bio 236:339–372

Cohas A, Allainé D (2009) Social structure influences extra-pair paternity in socially monogamous mammals. Biol Lett 5:313–316

Cotilla I, Palomo LJ (2007) *Microtus duodecimcostatus* (de Selys-Longchamps, 1839). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Tragsa, Madrid, pp 422–425

Cutrera AP, Fanjul MS, Zenuto RR (2012) Females prefer good genes: MHC-associated mate choice in wild and captive tuco tucos. Anim Behav 83:847–856

Dean MD, Ardlie KG, Nachman MW (2006) The frequency of multiple paternity suggests that sperm competition is common in house mice (*Mus domesticus*). Mol Ecol 15(13):1441–1451

DeVries AC, Carter SC (1999) Sex differences in temporal parameters of partner preference in prairie voles (*Microtus ochrogaster*). Can J Zool 77(6):885–889

Dewsbury DA (1987) The comparative psychology of monogamy. Nebr Sym Motiv 35:1–50

Dixon A, Ross D, O'Malley S, Burke T (1994) Paternal investment inversely related to degree of extra-pair paternity in the reed bunting. Nature 371:698–700

Ebensperger LA, Bozinovic F (2000) Energetics and burrowing behaviour in the semifossorial degu *Octodon degus* (Rodentia: Octodontidae). J Zool 252(2):179–186

Faulkes CG, Bennett NC (2013) Plasticity and constraints on social evolution in African mole-rats: ultimate and proximate factors. Phil Trans R Soc B 368:20120347

Ferkin MH, Sorokin ES, Johnston RE, Lee CJ (1997) Attractiveness of scents varies with protein content of the diet in meadow voles. Anim Behav 53(1):133–141

Fink S, Excoffier L, Heckel G (2006) Mammalian monogamy is not controlled by a single gene. P Natl Acad Sci USA 103(29):10956–10960

FitzGerald RW, Madison DM (1983) Social organization of a free-ranging population of pine voles, *Microtus pinetorum*. Behav Ecol Sociobiol 13(3):183–187

Ganem G, Litel C, Lenormand T (2008) Variation in mate preference across a house mouse hybrid zone. Heredity 100:594–601

Getz LL, Carter CS (1996) Prairie-vole partnerships. Am Sci 84(1):56–62

Getz LL, Carter CS, Gavish L (1981) The mating system of the prairie vole, *Microtus ochrogaster*: field and laboratory evidence for pair bonding. Behav Ecol Sociobiol 8:189–194

Giannoni SM (1994) Ecoerología comparada de dos especies de roedores subterráneos: *Microtus* (*Terricola*) *duodecimcostatus* y *M.* (*T.*) *pyrenaicus*. Ph.D. Thesis, Universidad Autónoma de Madrid, Madrid, pp 1–237

Gruder-Adams S, Getz LL (1985) Comparison of the mating system and paternal behavior in *Microtus ochrogaster* and *Microtus pennsylvanicus*. J Mammal 66:165–167

Guédon G, Pascal M (1993) Dynamique de population du campagnol provençal (*Pitymys duodecimcostatus* de Selys-Longchamps, 1839) dans deux agrosystémes de la région montpelliéraine. Rev Ecol – Terre Vie 48:375–398

Guédon G, Pascal M, Mazouin F (1991a) Le campagnol provençal en captivité (*Pitymys duodecimcostatus* de Sélys-Longchamps, 1839) (Rongeurs, Microtidés) I. La reproduction. Mammalia 55:97–106

Guédon G, Pascal M, Mazouin F (1991b) Le campagnol provençal en captivité (*Pitymys duodecimcostatus* de Sélys-Longchamps, 1839) (Rongeurs, Microtidés) II. La croissance. Mammalia 55:397–406

Guillen J (2012) FELASA guidelines and recommendations. JAALAS 51(3):311–321

Heske EJ, Ostfeld RS (1990) Sexual dimorphism in size, relative size of testes, and mating systems in North American voles. J Mamm 71(4):510–519

Jaarola M, Martínková N, Gündüza I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedakis-Tsolis S, González-Esteban J, López-Fuster MJ, Kandaurov AS, Kefelioğlu H, Mathias ML, Villate I, Searle JB (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol 33(3):647–663

Kleiman DG (1977) Monogamy in mammals. Q Rev Biol 52(1):39–69

Kruczek M, Golas A (2003) Behavioural development of conspecific odour preferences in bank voles *Clethrionomys glareolus*. Behav Process 64(1):31–39

Le Galliard JF, Remy A, Ims RA, Lambin X (2012) Patterns and processes of dispersal behaviour in arvicoline rodents. Mol Ecol 21:505–523

Lovegrove BG (1989) The cost of burrowing by the social mole rats (Bathyergidae) *Cryptomys damarensis* and *Heterocephalus glaber*: the role of soil moisture. Physiol Zool 62(2):449–469

Lukas D, Clutton-Brock TH (2013) The evolution of social monogamy in mammals. Science 341(6145):526–530

Luna F, Antinuchi CD, Busch C (2002) Digging energetics in the South American rodent *Ctenomys talarum* (Rodentia, Ctenomyidae). Can J Zool 80(12):2144–2149

MacManes MD (2013) On the accurate description of social and genetic mating systems. Peer J Preprints 1:e10v1. http://dx.doi. org/10.7287/peerj.preprints.10v1. Accessed 6 Sept 2013

Madureira ML (1982) New data on the taxonomy of portuguese *Pitymys*. Arq Mus Bocage A I:341–368

Madureira ML (1984) A biologia de *Microtus* (*Pitymys*) *duodecimcostatus* de Sélys-Longchamps, 1839 e *M.* (*P.*) *lusitanicus* Gerbe, 1879 em Portugal (Arvicolidae, Rodentia): taxonomia, osteologia, ecologia e adaptações. Ph.D. thesis, Faculdade de Ciências, Universidade de Lisboa, Lisboa

Marfori MA, Parker PG, Gregg TG, Vandenbergh JG, Solomon NG (1997) Using DNA fingerprinting to estimate relatedness within social groups of pine voles. J Mamm 78:715–724

McGuire B, Getz LL (2012) Multiple captures provide evidence of small mammal social behaviour. Acta Theriol 57(1):49–57

Mira A (1999) Ecologia do rato-cego-mediterrânico, *Microtus* (*Terricola*) *duodecimcostatus* (Rodentia, Muridae), em pomares de citrinos. Ph.D. Thesis, Faculdade de Ciências, Universidade de Lisboa, Lisboa

Mira A, Mathias ML (2007) *Microtus lusitanicus* (Gerbe, 1879). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Tragsa, Madrid, pp 418–421

Montoto LG, Magaña C, Tourmente M, Martín-Coello J, Crespo C, Luque-Larena JJ, Gomendio M, Roldan ERS (2011) Sperm competition, sperm numbers and sperm quality in muroid rodents. PLoS One 6(3):e18173

Mulder R, Dunn P, Cockburn A, Lazenbycohen K, Howell M (1994) Helpers liberate female fairy-wrens from constraints on extra-pair mate choice. P Roy Soc Lond B Bio 255:223–229

Munshi-South J (2007) Extra-pair paternity and the evolution of testis size in a behaviorally monogamous tropical mammal, the large treeshrew (*Tupaia tana*). Behav Ecol Sociobiol 62:201–212

Newman KS, Halpin ZT (1988) Individual odours and mate recognition in the prairie vole, *Microtus ochrogaster*. Anim Behav 36:1779–1787

Oliveras D, Novak M (1986) A comparison of paternal behaviour in the meadow vole *Microtus pennsylvanicus*, the pine vole *M. pinetorum* and the prairie vole *M. ochrogaster*. Anim Behav 34(2):519–526

Ophir AG, Phelps SM, Sorin AB, Wolff JO (2008) Social but not genetic monogamy is associated with greater breeding success in prairie voles. Anim Behav 75(3):1143–1154

Paradis E, Guédon G (1993) Demography of a Mediterranean microtine: the Mediterranean pine vole, *Microtus duodecimcostatus*. Oecologia 95:47–53

Parker KJ, Phillips KM, Lee TM (2001) Development of selective partner preferences in captive male and female meadow voles, *Microtus pennsylvanicus*. Anim Behav 61:1217–1226

Powell R, Fried J (1992) Helping by juvenile pine voles (*Microtus pinetorum*), growth and survival of younger siblings, and the evolution of pine vole sociality. Behav Ecol 3:325–333

Reichard UH (2003) Monogamy: past and present. In: Reichard UH, Boesch E (eds) Monogamy: mating strategies and partnerships in birds, humans and other mammals. University Press, Cambridge, pp 3–25

Richmond ME, Conaway CH (1969) Induced ovulation and oestrus in *Microtus orchrogaster*. J Rep Fer S 6:357–376

Santos SML (2009) Factors influencing the distribution of the lusitanian and the mediterranean pine voles (*Microtus lusitanicus* and *M. duodecimcostatus*) in Portugal: a multiscale approach. Ph.D. Thesis, Faculdade de Cieˆncias, Universidade de Lisboa, Lisboa

Santos SM, Lourenço RF, Mathias ML, Mira AP (2010) Spatial and temporal ecology of the Lusitanian pine vole (*Microtus lusitanicus*) in a Mediterranean polyculture. Anim Biol 60:209–227

Say L, Pontier D, Natoli E (1999) High variation in multiple paternity of domestic cats (*Felis catus* L.) in relation to environmental conditions. Proc R Soc Lond B 266:2071–2074

Smadja C, Ganem G (2002) Subspecies recognition in the house mouse: a study of two populations from the border of a hybrid zone. Behav Ecol 13(3):312–320

Solomon NG (1993) Comparison of parental behaviour in male and female prairie voles (*Microtus ochrogaster*). Can J Zool 71:434–437

Solomon NG, Keane B, Knoch LR, Hogan PJ (2004) Multiple paternity in socially monogamous prairie voles (*Microtus ochrogaster*). Can J Zool 82:1667–1671

Soriguer RS, Amat JA (1980) On the structure and function of the burrows of the Mediterranean vole (*Pitymys duodecimcostatus*). Acta Theriol 25:268–270

Tast J (1966) The root vole, *Microtus oeconomus* (Pallas), as an inhabitant of seasonally flooded land. Ann Zool Fenn 3:127–171

Thomas JA, Birney EC (1979) Parental care and mating system of the prairie vole, *Microtus ochrogaster*. Behav Ecol Sociobiol 5:171–186

Ventura J, Jiménez L, Gisbert J (2010) Breeding characteristics of the Lusitanian pine vole, *Microtus lusitanicus*. Anim Biol 60:1–14

Vericad JR (1970) Estudio faunístico y biológico de los mamíferos del Pirineo. Publ del Cent Piren de Biol Exp 4:1–258

Wiking VH (1976) Karyologie und Biologie der beiden iberischen Wuhlmausarten *Pitymys mariae* und *Pitymys duodecimcostatus*. Z Zool Syst Evol 14:104–129

Williams JR, Catania KC, Carter CS (1992) Development of partner preferences in female prairie voles (*Microtus ochrogaster*): the role of social and sexual experience. Horm Behav 26:339–349

Wolff JO (1985) Behaviour. In: Tamarin RH (ed) Biology of new world *Microtus*. American Society of Mammalogists, Shippensburg, pp 340–372

Wu J-S, Chiang P-J, Lin L-K (2012) Monogamous system in the Taiwan Vole *Microtus kikuchii* inferred from microsatellite DNA and home ranges. Zool Stud 51(2):204–212

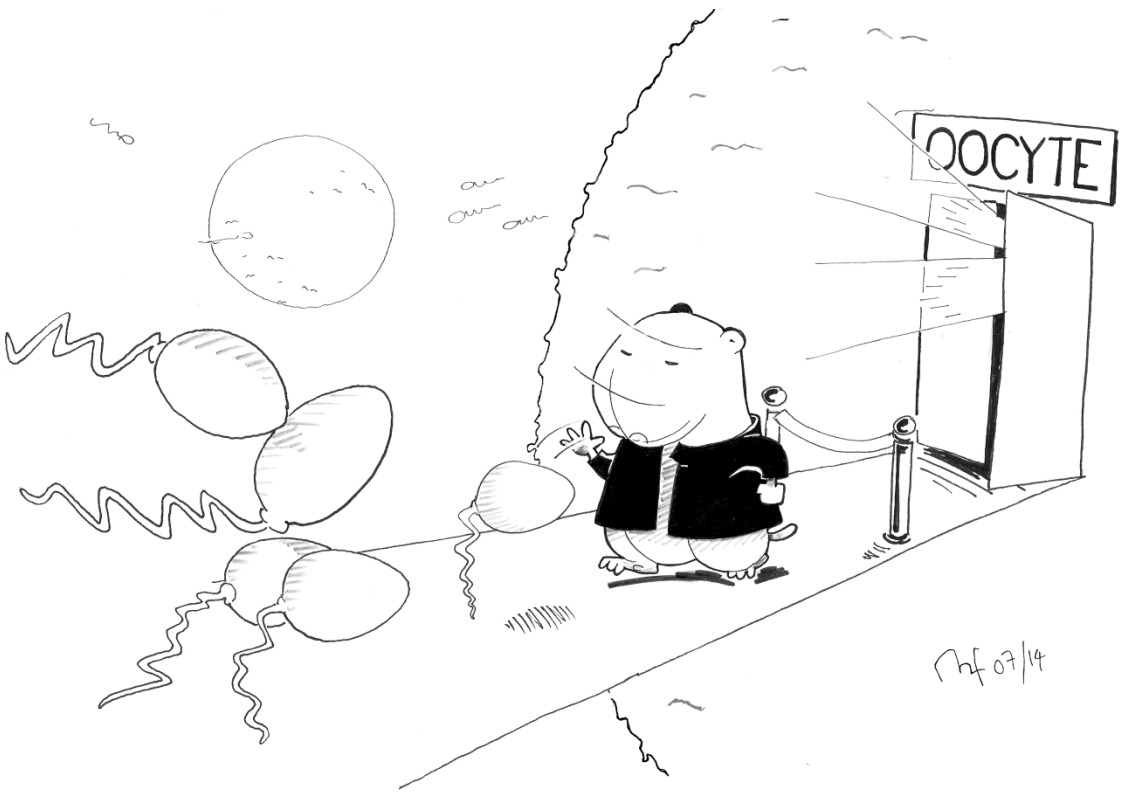
Young LJ, Wang Z (2004) The neurobiology of pair bonding. Nat Neurosci 7:1048–1054

Young KA, Liu Y, Wang Z (2008) The neurobiology of social attachment: a comparative approach to behavioral, neuroanatomical, and neurochemical studies. Comp Biochem Phys C 148(4):401–410

Young KA, Gobrogge KL, Liu Y, Wang Z (2011) The neurobiology of pair bonding: insights from a socially monogamous rodent. Front Neuroendocr 32:53–69

Chapter 5

Putative sperm-binding region of zona pellucida 3



5.1 Putative sperm-binding region of zona pellucida 3

Type of publication: Article

Reference: Duarte MA[1,2,3], Heckel G[4,5], Mathias ML[2,3], Bastos-Silveira C[1] (in preparation) Loss of amino acid residues in the putative sperm-binding region of ZP3 in cricetids.

[1] Museu Nacional de História Natural e da Ciência, Departamento de Zoologia e Antropologia, Universidade de Lisboa, Rua da Escola Politécnica, 58, Lisbon 1250-102, Portugal.

[2] Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[3] Centro de Estudos de Ambiente e Mar, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

[4] Computational and Molecular Population Genetics, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland.

[5] Swiss Institute of Bioinformatics, Genopode, Lausanne, Switzerland.

5.1.1 Abstract

Gamete surface proteins are essential to fertilization and, consequently, to reproductive isolation. In mammals, glycoprotein ZP3, located at the zona pellucida of the oocyte, is described as the primary receptor during fertilization. It has undergone rapid molecular divergence and particularly the putative sperm-binding region, located in exon 7, exhibits considerable amino acid variation. In the present study, we analysed evolutionary patterns in the putative sperm-binding region of rodent ZP3, focusing on the spectacular radiation of the cricetids (Cricetidae, Rodentia) and the speciose *Microtus* genus (Arvicolinae, Cricetidae), in comparison to the also highly species-rich Muridae. Our DNA sequence analyses revealed extensive genetic variation in the sperm-binding region of ZP3. Shared amino acid haplotypes between multiple Cricetidae species indicate that this region does not constitute a species-specific barrier as previously suggested. Furthermore, we uncovered the deletion of one amino acid residue common to Neotominae, six shared among Arvicolinae and four/nine in Sigmodontinae taxa relative to the ZP3 sequence found in the Muridae. Considering these findings, we

hypothesize that amino acid deletions in and around the ZP3 putative sperm-binding region may compromise the interaction stability between oocyte and sperm head, potentially impairing the species-specificity of fertilization and partially disrupting this isolation barrier.

5.1.2 Keywords

ZP3; reproductive isolation barriers; fertilization; cricetids; *Microtus* voles; amino acid deletion.

5.1.3 Introduction

Gamete surface proteins play an important role in reproductive isolation (reviewed in Turner & Hoekstra, 2008b). They constitute species-specific barriers to fertilization i.e. through post-mating gametic isolation, and potentially contribute to speciation (Swanson & Vacquier, 2002; Findlay & Swanson, 2010). In mammals, zona pellucida and sperm-head interacting proteins have co-evolved at a fast rate, often as a result of adaptive evolution, leading to species-specific fertilization and the establishment of genetic distinctiveness (reviewed in Swanson & Vacquier, 2002; Coyne & Orr, 2004; Seehausen *et al.*, 2014). This adaptive co-evolution is necessary to maintain gametic interaction and it potentially contributes also to amino acid differences between diverging populations (reviewed in Swanson & Vacquier, 2002; Clark *et al.*, 2009). Subsequently, gametic incompatibility may arise, leading to the differentiation of genomes by genetic drift and potentially the establishment of new species. Adaptive evolution of reproduction proteins happening at early stages of species divergence would thus constitute a driver of diversification, and not a consequence of speciation (reviewed in Swanson & Vacquier, 2002).

One of the best-characterized reproduction proteins, both functionally and evolutionarily, is ZP3, a glycoprotein of the oocyte zona pellucida of mammals (reviewed in Wassarman & Litscher, 1995). ZP3 consists of a polypeptide chain

connected to O- and N-linked glycans, and is described as the primary receptor during fertilization (Kinloch & Wassarman, 1989), because it binds directly to sperm and inhibits further binding of other sperm to the oocyte (Bleil & Wassarman, 1980, 1986). The putative sperm-binding region, located in exon 7, exhibits considerable amino acid variation between species which may together with modifications in the structure of the O-linked glycans enable a species-specific binding of the sperm to the oocyte (Wassarman & Litscher, 1995; Wassarman, 1999; Wassarman *et al*., 2005). In mice, the best-studied system, sperm-oocyte interactions have been associated in particular to a five serine (S) rich region, comprising S-332 and S-334 (Florman & Wassarman, 1985; Rosiere & Wassarman, 1992; Kinloch *et al*., 1995; Chen at al., 1998). The classical model of sperm-oocyte binding proposes that gametic interactions occur via O-linked glycans attached to S-332 and Ser-334, and that after fertilization these residues are deglycosylated to prevent further sperm adhesion (Florman & Wassarman, 1985; Chen *et al*., 1998). More recently, studies using genetically manipulated mouse models have challenged this classical model of sperm-oocyte binding and proposed alternative scenarios (reviewed in Redgrove *et al*., 2012). Moreover, it has been suggested that both conserved O-linked glycosylation sites outside exon 7 and the putative sperm-binding region are exposed on the same 3D protein surface, indicating that multiple distinct binding sites may be involved in sperm-oocyte recognition (Chalabi *et al*., 2006; Monné *et al*., 2011). Although the exact role of the putative sperm-binding region of ZP3 remains uncertain, it is clear that this glycoprotein, together with other zona pellucida and sperm head ligands, mediates sperm-oocyte binding, regardless of its specific molecular mechanism of action.

Evolutionarily, ZP3 is among the 10% most divergent proteins in mammals (Swanson *et al*., 2001). It has undergone rapid divergence and adaptive evolution driven by positive natural selection (Swanson *et al*., 2001; Swanson & Vacquier, 2002; Swanson *et al*., 2003; Turner & Hoekstra, 2008b; Palumbi, 2009; Morgan *et al*., 2010). Until recently, studies on the evolution of mammalian reproductive

proteins have mainly focused on comparing DNA sequences of distantly related species (e.g. Swanson *et al*., 2001; Morgan *et al*., 2010). This has shifted to shorter evolutionary timescales, since fertilization mechanisms within species and among closely related taxa are highly relevant to understand how amino acid changes affect reproductive isolation. For example, Turner and Hoekstra (2006, 2008a) documented positive selection acting on the putative sperm-binding region of ZP3 in various *Peromyscus* species Gloger, 1841 (Neotominae, Cricetidae, Rodentia), suggesting that divergence within the genus is adaptive. However, studies on murine rodents (Murinae, Muridae, Rodentia) (Swann *et al*., 2002, 2007), bovines (Bovinae, Bovidae, Artiodactyla) (Chen *et al*., 2011) and cetaceans (Cetacea) (Amaral *et al*., 2011) have questioned the action of positive selection on this locus and the species-specificity of the sperm-binding region of ZP3.

Consequently, to unravel evolutionary patterns of the putative sperm-binding region of ZP3 within Rodentia, the most diverse group of extant mammals, we performed a comprehensive comparative analysis on representatives of its most diverse families: Muridae and Cricetidae. We focused on cricetids and particularly on the speciose genus *Microtus* Schrank (1798), an evolutionarily young group that started to radiate 1.2-2 million years ago (Chaline *et al*., 1999) and has given rise to 65 extant species (e.g. Musser & Carleton, 2005) many of which are undergoing further diversification (e.g. Fink *et al*., 2010; Bastos-Silveira *et al*., 2012; Paupério *et al*., 2012; Beysard & Heckel, 2014). We found no evidence of positive selection overall and report the existence of extensive amino acid deletions and several shared amino acid haplotypes between taxa. These findings provide new insights into the role of the putative sperm-binding region in sperm recognition and question its stand-alone species-specificity feature.

5.1.4 Material and methods

Samples, DNA extraction, amplification and sequencing

The present study is based on 93 taxa from the two most species-rich families in extant mammals, Muridae and Cricetidae. Fifty Cricetidae species comprise 25

Arvicolinae (20 *Microtus sp.*), 17 Neotominae, 4 Cricetinae, 2 Tylomyinae and 2 Sigmodontinae taxa; all forty-three Muridae species are from the Murinae subfamily (**Table S1**). Tissue samples were stored in absolute ethanol at -20 ºC and made available by different research institutions (**Table S1**). Genomic DNA was extracted using standard protocols, requiring incubation with SDS and digestion with proteinase K, followed by a phenol-chlorophorm DNA extraction (Sambrook *et al.*, 1989).

Exon 6, intron 6 and exon 7 of the ZP3 gene were amplified using newly designed primers M-ZP3-F2 (5'-ATCACCTGTCATCTCAAAGTCA-3') and M-ZP3-R1 (5'-CATGCCTGCGGTTTCTAGAAGC-3'). All reactions contained 100 ng of template DNA, 0.3 mM of each primer, 1.25 U of GoTaq® Flexi DNA Polymerase (Promega), 1x buffer (Promega), 2.5 mM $MgCl_2$, 0.1 ug of BSA (New England Biolabs) and 0.2 mM of each dNTP (Thermo Scientific), to a final volume of 25 μl. DNA amplifications were performed in a MyCycler thermal cycler (Bio-Rad Laboratories Inc.) and consisted of a denaturation at 95 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 58 ºC for 1 min and extension at 72 ºC for 1 min. An extension step at 72 ºC for 10 min was added at the end of the reaction.

The length of the PCR products was verified on 1% agarose gels by comparing them with GeneRuler™ 100 bp Plus DNA Ladder (Fermentas). Products were purified using a ExoI/FastAP protocol (Fermentas). Sequencing in both directions with the same primers as the amplification was carried out by Macrogen Inc. (South Korea and the Netherlands) using an ABI Prism® 3100 Genetic Analyzer (Applied Biosystems).

Sequence analyses

Sequences were aligned using Sequencher 4.8 (Gene Codes Corporation) and BioEdit 7.2.5 (Hall, 1999). Our sequence dataset was supplemented with seventy-seven ZP3 sequences from the Murinae, Neotominae and Arvicolinae subfamilies, published in GenBank (**Table S1**), in order to compare patterns of ZP3 molecular

evolution among the two most species-rich rodent families. One representative species per Murinae genus was included in the analyses.

For further analyses, we focused on the coding region in exon 6 and 7, because of its importance for the expressed ZP3 glycoprotein related to reproductive isolation. Sequences were collapsed into unphased haplotypes using DNAcollapser (Villesen, 2007). Heterozygous positions of larger intraspecific datasets (*M. lusitanicus* and *M. duodecimcostatus*) were phased by Phase 2.1.1 (Stephens *et al.*, 2001; Stephens & Donnelly, 2003) implemented in DNAsp 5.10 (Librado & Rozas, 2009). Five independent runs were conducted using default values, and after checking for concordance a final run with 10 times more iterations (1000 iterations and 1000 burn-in iterations) was performed. Heterozygous positions of smaller intraspecific datasets were phased manually. DNA polymorphism parameters, such as the number of variable sites, number of parsimony informative sites and GC content were calculated using DnaSP 5.10.1. The translation of DNA into amino acid sequences was also performed with BioEdit 7.2.5. Amino acid sequence variation was visualized using the WebLogo application (Schneider & Stephens, 1990; Crooks *et al.*, 2004), available at the ExPASy SIB Bioinformatics Resource Portal.

JModelTest 2.1.7 (Darriba *et al.*, 2012) was used to select the best-fitting model of nucleotide substitution TVM+G (Posada, 2003; Yang, 1993), based on the Akaike information criterion (Akaike, 1974). Due to the presence of alignment gaps in exon 7 of some species, it was important to consider this information in the phylogenetic analysis. Bayesian inference with MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) allows the incorporation of gaps as binary characters in a different partition using a phylogenetic mixed model. Binary matrices were constructed by SeqState 1.4.1 (Müller, 2005), using two types of coding for the gaps: the simple indel coding (SIC, Simmons & Ochoterena, 2000) and modified complex indel coding (MCIC, Müller, 2006). Bayesian inference on each matrix consisted of two runs with four chains, one cold and three heated, and 4.000.000 generations, with every 100th generation sampled. The average

standard deviation of split frequencies between the independent runs was checked for convergence to ensure that a value < 0.01 was achieved (Huelsenbeck & Ronquist 2001). The first 25% trees were discarded (burn-in) and the remaining trees were used to construct a consensus tree and estimate Bayesian posterior probabilities. The Bayesian Inference consensus tree obtained was drawn using FigTree 1.3.1 (Rambaut, 2010).

Recombination in our ZP3 fragment could mislead selection analyses, thus we tested for its presence using RDP (Martin & Rybicki, 2000), BOOTSCAN (Salminen, 1995; Martin *et al*., 2005), GENECONV (Padidam *et al*., 1999), MAXCHI (Smith, 1992; Posada & Crandall, 2001), CHIMAERA (Posada & Crandall, 2001), SISCAN (Gibbs, 2000), and 3SEQ (Boni *et al*., 2007) methods implemented in RDP 4 (Martin *et al*., 2010).

We tested for positive selection using the CodeML subroutine included in PAML 4.8 (Yang, 1997, 2007). Maximum likelihood estimations of ω (non-synonymous(dN)/synonymous(dS) substitution rates) among codons were generated according to six models for distribution patterns of ω: M0 (one ω), M1 (nearly neutral, one ω, two classes of sites), M2 (positive selection, three classes of sites), M3 (discrete, three classes of sites); M7 (nearly neutral with the beta distribution approximating ω variation, 10 classes of sites) and M8 (positive selection with the beta distribution approximating ω variation, 11 classes of sites) (Goldman & Yang, 1994; Yang *et al*., 2000; Yang *et al*., 2005). The ω ratio is a sensitive measure of selective pressure, enabling the detection of positive selection when ω> 1 (Yang & Nielsen, 2002). Additionally, we used branch-site models that allow ω variation among amino acids in the protein and across branches on the phylogenetic tree in order to detect possible positive selection affecting a few sites along particular lineages (foreground branches) (Yang, 1998; Yang & Nielsen, 1998). This approach enables the detection of positive selection in specific families and/or subfamilies in the phylogenetic tree that may affect only a few codons in the analyzed ZP3 protein fragment (Yang & Nielsen, 2002; Yang *et al*., 2005; Zhang *et al*., 2005). Thus, it can be a statistically more powerful

method than site-based tests, which average over all branches of the phylogeny (Nielsen & Yang, 1998; Suzuki & Gojobori 1999; Yang *et al*., 2000). The null (model = 2; NSsites = 2; ω = 1) and neutral M1a (model = 0; NSsites = 1; ω = 1) models were compared to the MA1 (model = 2; NSsites = 2; ω estimated), the alternative model in the branch-site test of positive selection. Likelihood ratio tests (LRTs) of M0 vs. M3, M1 vs. M2, M7 vs. M8, null model vs. MA1 and M1a vs. MA1 were performed in order to test for evidence of positive selection (Nielsen & Yang, 1998; Yang *et al*., 2000). Two times the log-likelihood difference between models (2Δl) is compared to a chi-quare distribution with the number of degrees of freedom (dF) corresponding to the difference in the number of parameters between both models (Yang *et al*., 2000). Positively selected sites under M2, M3, M8 and MA1 were identified using the Naive Empirical Bayes (NEB) and the Bayes Empirical Bayes (BEB) analysis (Yang *et al*., 2005).

The power of CodeML can be affected by the accuracy of the input phylogenetic tree (Anisimova *et al*., 2003). Therefore, we combined the PAML results with selection detection methods implemented in the HyPhy package (Pond *et al*., 2005) web interface DataMonkey (Delport *et al*., 2010): SLAC (single likelihood ancestor counting, Pond & Frost, 2005), FEL (fixed effects likelihood, Pond & Frost, 2005), IFEL (internal fixed effects likelihood, Pond *et al*., 2006b) and MEME (mixed effects model of evolution, Murrell *et al*., 2012). Due to alignment size restriction it was not possible to test the REL method (random effects likelihood, Pond *et al*., 2005) and branch-site REL (Pond *et al*., 2011). All methods were tested under a significance threshold of 0.05.

5.1.5 Results

Genetic variation and phylogeny

Our analyses of 43 Muridae and 103 Cricetidae DNA sequences (**Table S1**) revealed extensive length and sequence variation in ZP3, not only in intron 6 but also in exon 6 and 7, including the putative sperm-binding region (see below) (**Fig. S1**-**S2**). The final data matrix containing only the coding region was 228 bp long,

corresponding to positions 835-1063 in the reference mouse ZP3 gene. Most Cricetidae taxa analysed in the present study are new DNA sequence contributions to public databases (GenBank accession numbers: exon 6 – XXxxxxx-YYyyyyy, exon 7 – WWwwwww-KKkkkkk, **Table S1**). There was no evidence of length variation between the two alleles of an individual, or recombination in our ZP3 data. Twenty DNA sequences, four from public databases and 16 of our new sequences, comprised heterozygous positions (16 at one position; three two positions; one at three positions). The phased dataset contained a total of 78 variable sites of which 63 were parsimony informative, and a GC content of 54.3%. These polymorphisms defined a total of 111 haplotypes, 40 in the Muridae and 71 in the Cricetidae: 35 Arvicolinae, 23 Neotominae, 4 Cricetinae, 6 Tylomyinae and 3 Sigmodontinae. None of these haplotypes was shared between families or subfamilies but all were shared between species in the same subfamily. In total, there were ten haplotypes shared between 15 Cricetidae taxa (14.1% of all Cricetidae haplotypes), whereas the Muridae only presented two (5% of all Muridae haplotypes) observed in five species.

Tree topologies obtained for the standard matrix were congruent with the ones constructed using gaps as binary characters (here we only present the phylogenetic tree with SIC method, **Fig. 1**). Branch length of some haplotypes varied between trees, due to the usage of gap data (data not shown). No differences were found between both Bayesian Inference runs performed per matrix (data not shown).

Phylogenetic reconstructions based on the coding sequences separated all species at the family level, and set a topology within the Cricetidae that was largely, but not fully consistent with the current taxonomy (**Fig. 1**). Phylogenies showed high support for most nodes, irrespective of the coding method for the indel positions. Taxa in the Arvicolinae and Sigmodontinae subfamilies were monophyletic, but several nodes were not resolved at the subfamily level since not all haplotypes from Cricetinae and Tylomyinae species clustered together, respectively (**Fig. 1**).

Amino acid variation

The translation of the DNA sequence of exon 6 and 7 yielded 74 amino acids (positions 279-354 according to the reference mouse ZP3 protein). Forty-five variable amino acid sites (60.8%) and 14 indel positions defined a total of 72 amino acid sequence types (**Fig. S1**-**S2**).

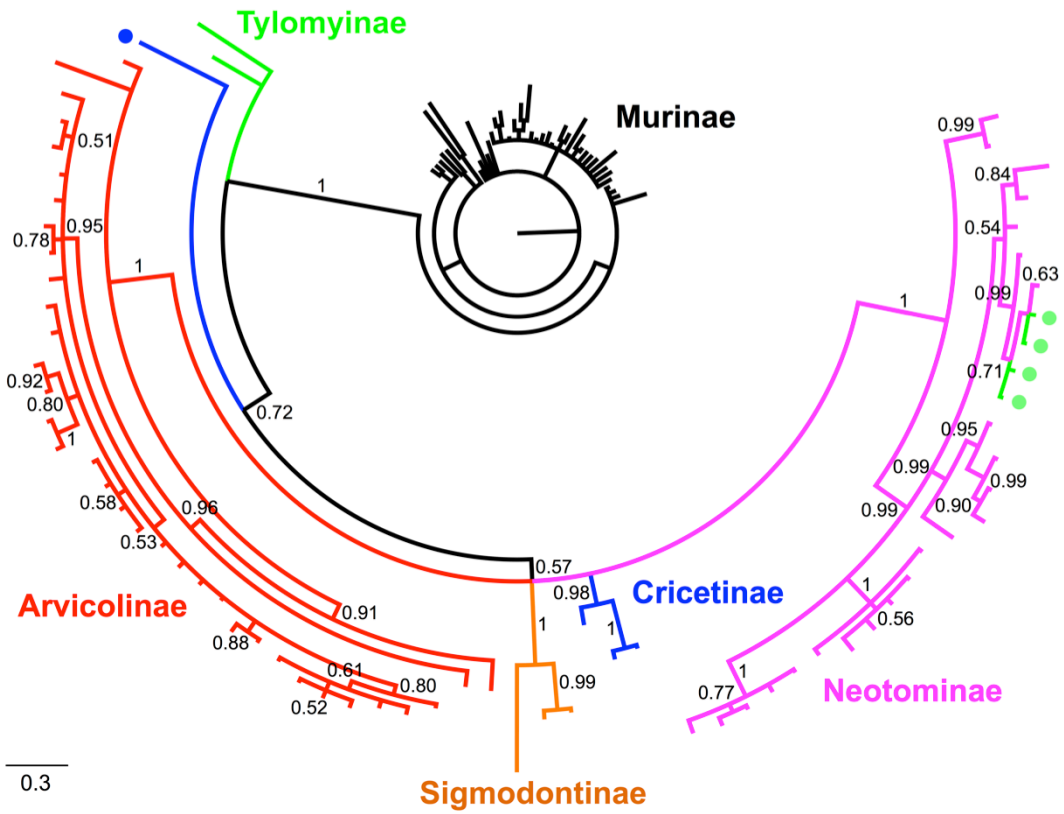


Figure 1 – Bayesian inference phylogenetic tree obtained for exon 6 and 7 of the ZP3 gene of Muroid rodents, using the SIC method of gap coding. Posterior probabilities higher than 0.50 for nodes within the Cricetidae family are shown. Circles indicate haplotypes separated from the respective subgenus cluster: Cricetinae *Mesocricetus auratus* and Tylomyinae *Tylomys watsoni* and *Nyctomys sumichrasti*.

Considerable length variation was observed due to the deletion of amino acids predominantly in the putative sperm-binding region relative to mouse (328-343, Rossiere & Wassarman, 1992). In all Arvicolinae species, six amino acids at position 342-347 were missing relative to mouse ZP3, and all Sigmodontinae lacked amino acid position 330 (together with Neotominae *Onychomys torridus*)

and 336-338 (**Fig. 2** and **Fig. S2**). Additional deletions were observed in *Sigmodon arizonae* (331-334 and 344). Interestingly, taxa belonging to species-poor subfamilies Cricetinae and Tylomyinae did not present amino acid deletions, contrasting with species-rich Sigmodontinae, Arvicolinae and Neotominae. In the Muridae family, only *Lemniscomys griselda* presented amino acid deletions within the sperm-binding region, at positions 336-337 (**Fig. S1**).

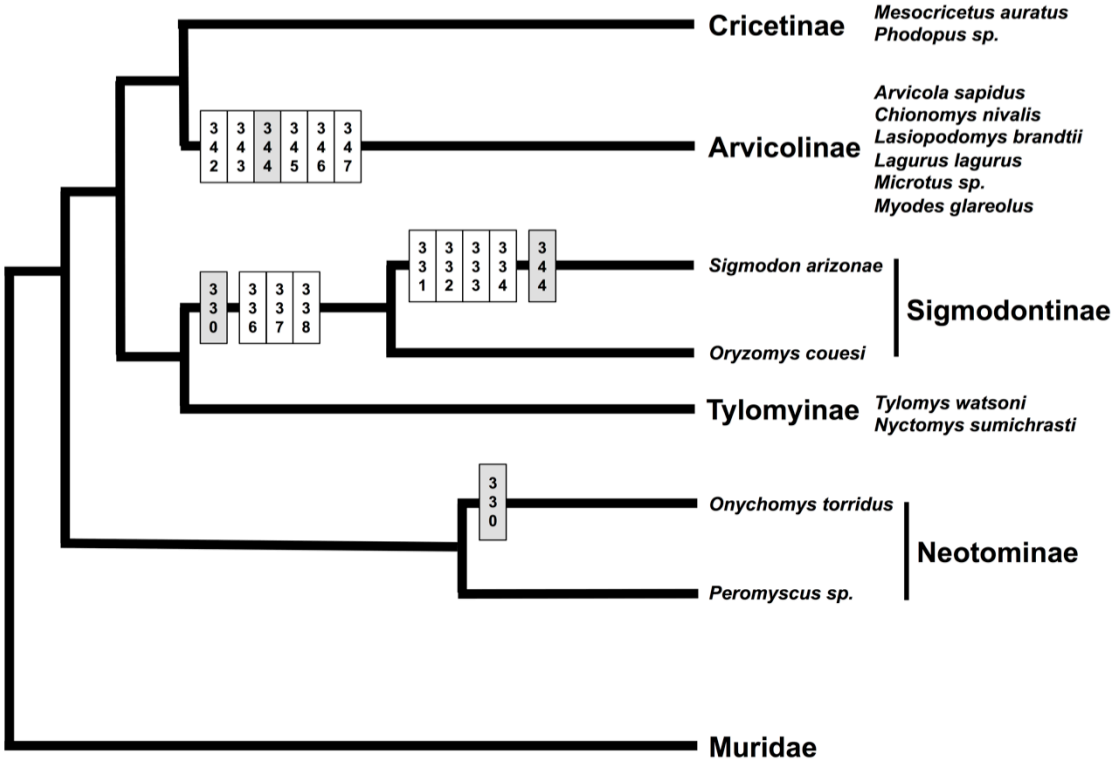


Figure 2 – Schematic representation of length variation in the sperm-binding region of ZP3 along a phylogenetic tree of the Cricetidae (modified after Fabre *et al.* 2012). The amino acid deletions in the Cricetidae relative to the position in the mouse reference sequence are indicated on the respective branches, with grey specifying deletions being shared by two subfamilies.

Amino acid positions 288 and 348 were diagnostic for Muridae and Cricetidae species (**Fig. S1**-**S2**). Sharing of amino acid sequence types was found between members of the same genus but also between genera in the same subfamily (**Fig. S1**-**S2**). Different levels of intra-genus variability were found within Cricetidae: 16 *Peromyscus* (Neotominae) taxa revealed 14 amino acid sequence types and 20 *Microtus* (Arvicolinae) species harboured eleven (**Fig. S2**). We also recorded some

cases of intra-species variability, concerning Neotominae *Onychomys torridus*, *Peromyscus difficilis* and *Peromyscus mexicanus*; Tylomyinae *Nyctomys sumichrasti*; and Arvicolinae *Microtus arvalis*, *Microtus duodecimcostatus*, *M. rossiaemeridionalis* and *M. socialis*, which presented more than one amino acid sequence type due to non-synonymous polymorphisms (**Fig. S2**).

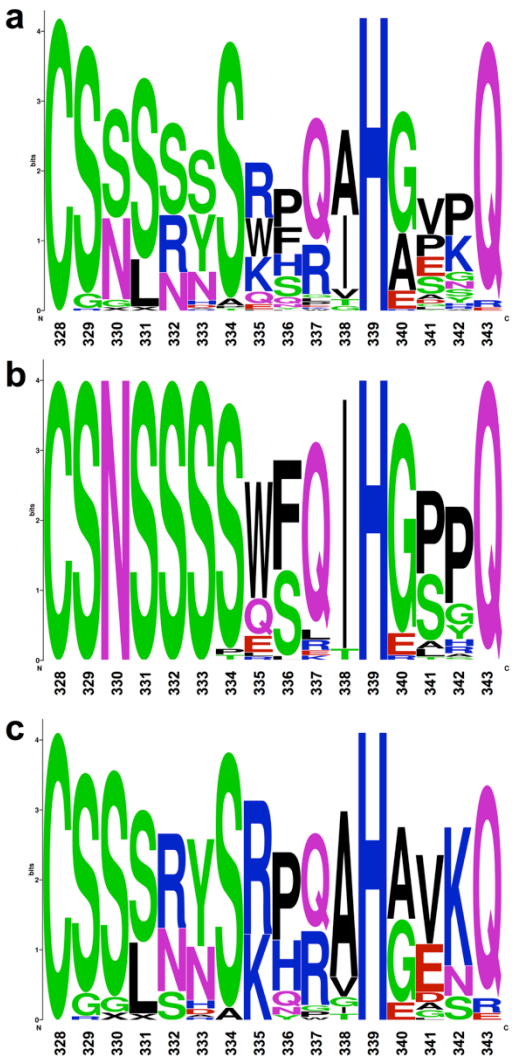


Figure 3 – Amino acid sequence logos showing variation in the putative sperm-binding region of ZP3: a – consensus of all Muroid taxa, b – Muridae and c - Cricetidae. Overall height indicates sequence conservation at the position, while symbol height within a stack indicates the relative frequency of each amino acid.

High variability was also found in the ZP3 putative sperm-binding region with only sites 328 and 339 being invariable across all species (**Fig. 3** and **Fig. S1**-**S2**).

Most Muridae present the characteristic serine-asparagine-serine-serine-serine-serine sequence at positions 329-334, with the exception of *Conilurus penicullatus* and *Pseudomys laborifex*, while the Cricetidae show extensive (**Fig. 3** and **Fig. S1**-**S2**).

Selection tests

Statistical tests of positive selection suggest that the analysed ZP3 data is under variable selective pressure (**Table 1** and **Table S2**). PAML LTRs rejected the null hypotheses models M0, M1 and M7 in favour of the alternative M3, M2 and M8 ($P < 0.001$), particularly for the M3 model which is indicative of variable selective pressure (**Table 1** and **Table S2**). Moreover, ω values < 1 for the different site models suggest that most codons are under negative/purifying selection. Model M0 obtained a $\omega = 0.456$, suggesting mostly purifying selection, a result also supported by M1 with 69% of $\omega < 1$ sites, and M8, with 88% of $\omega < 1$ sites (**Table S2**). Also, M2 indicated 9% positively selected sites with $\omega 2 = 2.889$ and 64% of $\omega < 1$ sites; and model M3 a 14% positively selected sites, with $\omega 2 = 2.176$ and 38% of $\omega < 1$ sites and 48% of $\omega = 1$ sites (**Table S2**).

PAML M2, M3 and M8 models and HyPhy SLAC, FEL, IFEL and MEME tests detected positively and negatively selected sites, distributed throughout exon 6 and 7 of the ZP3 gene (**Fig. S3** and **Table S2**). From these, HyPhy and PAML both identified two amino acid sites under positive selection (336 and 337) located in the sperm-binding region, while other positively selected positions were only identified by PAML or HyPhy methods, respectively (**Fig. S3** and **Table S2**). Purifying selection was detected at 14 shared amino acid sites both at the 95% and 99% cut-off which include the serine-rich site S-334 and the two invariable positions in the sperm-binding region (328 and 339; **Fig. S3** and **Table S2**).

PAML branch-site comparison of the null model vs. MA1 and M1a vs. MA1 revealed variable selective pressure, depending on the family/subfamily analyzed as the foreground branch (**Table 1** and **Table S2**). The null hypothesis of neutrality was rejected for both families, Muridae and Cricetidae, and for two subfamilies of

cricetids: Arvicolinae and Tylomyinae ($p < 0.05$). Moreover, MA1 model uncovered positively selected sites for the Muridae (315 and 322) and Cricetidae (287, 307 and 311) families, and for three subfamilies of cricetids: Arvicolinae (287), Cricetinae (296 and 309) and Tylomyinae (307 and 311). All positively selected sites are located outside the putative sperm-binding region of ZP3 (**Table S2**).

Table 1 – Results of the likelihood ratio tests (LRT) considering site- and branch-site models implemented by PAML on exon 6 and 7 of the ZP3 gene of the analysed taxa. LRT values are significant at $p < 0.05$. $2\Delta l$ = twice the log likelihood difference between the two compared models, df = degrees of freedom.

Type	LRT	$2\Delta l$	df	*P* value
Site-models	M0 vs. M3	273.415	4	< 0.001
	M1 vs. M2	41.840	2	< 0.001
	M7 vs. M8	46.232	2	< 0.001
Branch-site models: Muridae	null vs. MA1	0.796	2	< 0.001
	M1a vs. MA1	1.420	2	< 0.001
Branch-site models: Cricetidae	null vs. MA1	1.077	2	< 0.001
	M1a vs. MA1	12.208	2	< 0.001
Branch-site models: Arvicolinae	null vs. MA1	6.280	2	< 0.001
	M1a vs. MA1	6.560	2	< 0.001
Branch-site models: Cricetinae	null vs. MA1	0	2	1.000
	M1a vs. MA1	0.927	2	0.629
Branch-site models: Neotominae	null vs. MA1	0	2	1.000
	M1a vs. MA1	0	2	1.000
Branch-site models: Sigmodontinae	null vs. MA1	0	2	1.000
	M1a vs. MA1	0.004	2	0.998
Branch-site models: Tylomyinae	null vs. MA1	0.087	2	< 0.001
	M1a vs. MA1	0.242	2	0.242

5.1.6 Discussion

The present study constitutes a step forward in an evolutionary understanding of the role of the ZP3 putative sperm-binding region as a gametic reproductive isolation barrier in mammals. The presence of amino acid deletions in and around the putative sperm-binding region of most Cricetidae species analysed, together with an absence of conservation in positions 332 and 334 lead to the refutation of the classical model of ZP3 O-linked glycan sperm-oocyte binding. Our results also indicate that this region does not constitute a species-specific barrier as

previously suggested (Florman & Wassarman, 1985; Chen *et al*., 1998) since shared amino acid sequence types were found between different taxa. Thus, the putative sperm-binding region of ZP3 *per se* is not a gametic isolation barrier. Nevertheless, our findings do not question the role of ZP3 in sperm recognition together with other proteins from the zona pellucida and sperm head. The present results rebut the role of the putative sperm-binding region alone in the species-specificity of sperm-oocyte binding.

Regarding exon 7, comprising the putative sperm-binding region, we found extensive length variation in Arvicolinae and Sigmodontinae species. The presence/absence of amino acid deletions is not directly associated with the evolutionary relationships of sister subfamilies Arvicolinae+Cricetinae and Sigmodontinae+Tylomyinae proposed by Fabre *et al*. (2012), since the analysed Arvicolinae and Sigmodontinae species presented amino acid deletions, whereas Cricetinae and Tylomyinae did not (Fig. 2).

Furthermore, positions 332 and 334 of the five serine-rich region were not conserved as expected (Fig. 2 and Fig. S1-S2). According to the classical model, this would affect gametic recognition since S-332 and S-334 are supposed to carry O-linked glycans essential to sperm-oocyte binding. Since fertilization is not impaired in cricetids with different amino acids at these positions, S-332 and S-344 are not vital to gametic recognition in these Cricetidae and probably other mammals. Given that at least two species in the Muridae do not present S-334 (Fig. S1-S2), this may be a relatively general feature of rodents.

The species-specificity function of ZP3 might be related to protein sections other than exon 7, since our analyses of a relatively large number of rodent taxa detected shared sequence types. This is consistent with the propositions that: i) sperm binds to ZP3 by interacting with O-linked glycans not connected to S-332 and S-334 (Visconti & Florman, 2010) or to N-linked glycans and accessible protein regions located within the C-terminal domain of ZP3 (Clark *et al*., 2011); ii) two conserved O-linked glycosylation sites (residues T-155 and T-162/S-164/S-165) shared by mouse and human ZP3 may be the actual attachment sites of the

sperm-binding glycans (Chalabi *et al*., 2006); and iii) these two conserved O-linked glycosylation sites and the putative sperm-binding region in exon 7 are exposed on the same 3D protein surface, indicating that multiple distinct binding sites might be involved in sperm-oocyte recognition (Monné *et al*., 2011). Analysis of the complete ZP3 protein including other putative sperm-binding regions (Chalabi *et al*., 2006; Monné *et al*., 2011) could enlighten such conjectures.

Moreover, we could not find an overall positive selection signature based on site-models. This observation is congruent with previous studies on different mammalian groups (murines: Swann *et al*., 2002; cetaceans: Amaral *et al*., 2011; bovines: Chen *et al*., 2011). Nevertheless, branch models detected variable selective pressure acting on the phylogenetic branches of the Muridae and Cricetidae families and Arvicolinae and Tylomyinae subfamilies. This incongruence may be caused by: 1) functional and structural constraints on ZP3 sequence, as it is possible that amino acid changes may disturb ZP3 glycosylation and affect sperm-oocyte recognition (Kinloch *et al*., 1995); and 2) evolutionary divergence of particular subfamilies, which may lead to or be the consequence of reproductive isolation and speciation.

In consideration of the present findings, we hypothesize that amino acid deletions in and around the ZP3 putative sperm-binding region may compromise interaction stability between gametes, potentially impairing the species-specificity of fertilization and partially disrupting this isolation barrier. This scenario suggests a possible role of ZP3 in the speciation of cricetids and highlights the importance of reproductive barriers development to speciation and biodiversity (reviewed in Butlin et al., 2009; Langerhans & Riesch, 2013).

5.1.7 Acknowledgements

We are grateful to Robert Baker, Heath Garner and Kathy MacDonald from the Museum of Texas Tech University (Lubbock, USA), Peter Fritzsche from the Institute of Zoology of the Martin-Luther University (Halle (Salle), Germany), and Isabel Rey from the Museo Nacional de Ciencias Naturales (Madrid, Spain) for

providing biological samples. We would also like to thank Susanne Tellenbach for technical assistance. M. A. Duarte was supported by PhD grant SFRH/BD/70646/2010 from the Foundation for Science and Technology (FCT) and G. Heckel by grant 31003A-149585 from the Swiss National Science Foundation. This study was developed under the project PTDC/BIA-BEC/103729/2008 (FCT), coordinated by C. Bastos-Silveira. Thanks are due, for the financial support to Centro de Estudos de Ambiente e Mar (UID/AMB/50017), to FCT/MEC through national funds, and the co-funding by the FEDER, within the PT2020 Partnership Agreement and Compete 2020.

5.1.8 References

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J., Bierne, N. *et al.* 2013. Hybridization and speciation. *J. Evol. Biol.* **26**: 229-246.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**: 716-723.

Amaral, A.R., Möller, L.M., Beheregaray, L.B., Coelho, M.M. 2011. Evolution of 2 reproductive proteins, ZP3 and PKDREJ, in cetaceans. *J. Hered.* **102**: 275-282.

Anisimova, M., Nielsen, R., Yang, Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**: 1229-1236.

Bastos-Silveira, C., Santos, S.M., Monarca, R., Mathias, M.L., Heckel, G. 2012. Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Mol. Ecol.* **21**: 5309-5323.

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

Beysard M., Krebs-Wheaton, R., Heckel, G. 2015. Tracing reinforcement through asymmetrical partner preference in the European common vole *Microtus arvalis*. *BMC Evol. Biol.* **15**: 170.

Bleil, J.D., Wassarman, P.M. 1980. Structure and Function of the Zona Pellucida: Identification and Characterization of the Proteins of the Mouse Oocyte's Zona Pellucida. *Dev. Biol.* **76**: 185-202.

Bleil, J.D., Wassarman, P.M. 1986. Autoradiographic visualization of the mouse egg's sperm receptor bound to sperm. *J. Cell Biol.* **102**: 1363-1369.

Boja, E.S., Hoodbhoy, T., Fales, H.M., Dean, J. 2003. Structural Characterization of Native Mouse Zona Pellucida Proteins Using Mass Spectrometry. *J. Biol. Chem.* **278**: 34189-34202.

Boni, M.F., Posada, D., Feldman, M.W. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **176**: 1035-1047.

Castiglia, R., Annesi, F., Aloise, G., Amori, G. 2008. Systematics of the *Microtus savii* complex (Rodentia, Cricetidae) via mitochondrial DNA analyses: paraphyly and pattern of sex chromosome evolution. *Mol. Phylogenet. Evol.* **46**: 1157-1164.

Chalabi, S., Panico, M., Sutton-Smith, M., Haslam, S.M., Patankar, M.S., Lattanzio, F.A. *et al.* 2006. Differential O-glycosylation of a conserved domain expressed in murine and human ZP3. *Biochem.* **45**: 637-647.

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

Chen, J., Litscher, E.S., Wassarman, P.M. 1998. Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. *Proc. Natl. Acad. Sci. USA* **95**: 6193-6197.

Chen, S., Costa, V., Beja-Pereira, A. 2011. Evolutionary patterns of two major reproduction candidate genes (Zp2 and Zp3) reveal no contribution to reproductive isolation between bovine species. *BMC Evol. Biol.* **11**: 24.

Clark, N.L., Gasper, J., Sekino, M., Springer, S.A., Aquadro, C.F., Swanson, W.J. 2009. Coevolution of Interacting Fertilization Proteins. *PLoS Genet.* **5**: e1000570.

Clark, G.F. 2011. The molecular basis of mouse sperm–zona pellucida binding: a still unresolved issue in developmental biology. *Reproduction* **142**: 377-381.

Coyne, J.A., Orr, H.A. 2004. *Speciation*. Sinauer Associates, Sunderland.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. 2004. WebLogo: A sequence logo generator. *Genome Res.* **14**: 1188-1190.

Delport, W., Poon, A.F., Frost, S.D.W., Pond, S.L.K. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**: 2455-2457.

Ellies, L.G., Tsuboi, S., Petryniak, B., Lowe, J.B., Fukuda, M., Marth, J.D. 1998. Core 2 oligosaccharide biosynthesis distinguishes between selectin ligands essential for leukocyte homing and inflammation. *Immunity* **9**: 881-890.

ExPASy SIB Bioinformatics Resource Portal. See http://www.expasy.org

Fabre, P.-H., Hautier, L., Dimitrov, D., Douzery, E.J.P. 2012. A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evol. Biol.* **12**: 88.

Findlay, G.D., Swanson, W.J. 2010. Proteomics enhances evolutionary and functional analysis of reproductive proteins. *BioEssays* **32**: 26-36.

Fink, S., Fischer, M.C., Excoffier, L., Heckel, G. 2010. Genomic Scans Support Repetitive Continental Colonization Events during the Rapid Radiation of Voles (Rodentia: *Microtus*): the Utility of AFLPs versus Mitochondrial and Nuclear Sequence Markers. *Syst Biol.* **59**: 548-572.

Florman, H.M., Wassarman, P.M. 1985. O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. *Cell* **41**: 313-324.

Gahlay, G., Gauthier, L., Baibakov, B., Epifano, O., Dean, J. 2010. Gamete Recognition in Mice Depends on the Cleavage Status of an Egg's Zona Pellucida Protein. *Science* **329**: 216-219.

Galleni, L., Tellini, A., Stanyon, R. 1994. Taxonomy of *Microtus savii* (Rodentia, Arvicolidae) in Italy: cytogenetic and hybridization data. *J. Mammal.* **75**: 40-44.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J. 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**: 573-582.

Gileva, E.A., Bol'shakov, V.N., Polyavina, O.V., Cheprakov, M.I. 2000. The vole species *Microtus arvalis* and *Microtus rossiaemeridionalis* in the Urals: hybridization in the wild. *Dokl. Biol. Sci.* **370**: 47-50.

Goldman, N., Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**: 725-736.

Grant, P.R. 1974. Reproductive compatibility of voles from separate continents (Mammalia: *Clethrionomys*). *J. Zool., Lond.* **174**: 245-254.

Grant, P.R. 1976. An 11-year study of small mammal populations at Mount St. Hilaire, Quebec. *Can. J. Zool.* **54**: 2156-2173.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95-98.

Han, L., Monné, M., Okumura, H., Schwend, T., Cherry, A.L., Flot, D. *et al.* 2010. Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell* **143**: 404-415.

Harrison, R.G., Larson, E.R. 2014. Hybridization, Introgression, and the Nature of Species Boundaries. *J. Hered.* **105**: 795-809.

Huelsenbeck, J.P., Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.

Kinloch, R.A., Sakai, Y., Wassarman, P.M. 1995. Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis. *Proc. Natl. Acad. Sci. USA.* **92**: 263-267.

Kinloch, R.A., Wassarman, P.M. 1989. Nucleotide sequence of the gene encoding zona pellucida glycoprotein ZP3--the mouse sperm receptor. *Nucleic Acids Res.* **17**: 2861-2863.

Lavrenchenko, L.A. 2014. Hybrid Speciation in Mammals: Illusion or Reality? *Biol. Bull. Rev.* **4**: 198–209.

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

Martin, D., Rybicki, E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**: 562-563.

Martin, D.P., Posada, D., Crandall, K.A., Williamson, C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retrovir.* **21**: 98-102.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**: 2462-2463.

Monné, M., Jovine, L. 2011. A Structural View of Egg Coat Architecture and Function in Fertilization. *Biol. Reprod.* **85**: 661-669.

Morgan, C.C., Loughran, N.B., Walsh, T.A., Harrison, A.J., O'Connell, M.J. 2010. Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. *BMC Evol. Biol.* **10**: 39.

Müller, K. 2005. SeqState - primer design and sequence statistics for phylogenetic DNA data sets. *Appl. Bioinformatics* **4**: 65-69.

Müller, K. 2006. Incorporating information from length-mutational events into phylogenetic analysis. *Mol. Phylogenet. Evol.* **38**: 667-676.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., Pond, S.L.K. 2012. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genet.* **8**: e1002764.

Musser, G.M., Carleton, M.D. 1993. Family Cricetidae, In: *Mammal species of the world: a taxonomic and geographic reference* (D.E. Wilson & D.M. Reeder, eds), pp. 955-1189. Smithsonian Institution, Washington.

Nielsen, R., Yang, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**: 929-936.

Padidam, M., Sawyer, S., Fauquet, C.M. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**: 218-225.

Palumbi, S.R. 2009. Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity* **102**: 66-76.

Paupério J., Herman, J.S., Melo-Ferreira, J., Jaarola, M., Alves, P.C., Searle, J.B. 2012. Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Mol. Ecol.* **21**: 6015-6032.

Pond, S.L.K., Frost, S.D.W. 2005. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol. Biol. Evol.* **22**: 1208-1222.

Pond, S.L.K., Frost, S.D.K., Muse, S.V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**: 676-679.

Pond, S.L.K., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W. 2006a. Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. *Mol. Biol. Evol.* **23**: 1891-1901.

Pond, S.L.K., Frost, S.D.W., Grossman, Z., Gravenor, M.B., Richman, D.D., Brown, A.J.L. 2006b. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comput. Biol.* **2**: e62.

Pond, S.L.K., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W., Scheffler, K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* **28**: 3033-3043.

Posada, D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**: 1253-1256.

Posada, D., Crandall, K.A. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. USA* **98**: 13757-13762.

Rauschert, K. 1963. Sexuelle affinitat zwischen arten und unterarten von Rotelmausen (*Clethrionomys*). *Biol. Zbl.* **82**: 653-664.

Rambaut, A. 2010. FigTree. See http://tree.bio.ed.ac.uk/software/figtree/

Redgrove, K.A., Aitken, R.J., Nixon, B. 2012. More than a simple lock and key mechanism: unraveling the intricacies of sperm-zona pellucida binding. In: *Binding Protein* (K. Abdelmohsen, ed), pp. 73-122. InTech, Rijeka.

Ronquist, F., Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.

Rosière, T.K., Wassarman, P.M. 1992. Identification of a Region of Mouse Zona Pellucida Glycoprotein mZP3 That Possesses Sperm Receptor Activity. *Dev. Biol.* **154**: 309-317.

Safronova, L.D., Golenishchev, F.N., Cherepanova, E.V., Baskevich, M.I. 2011. Meiosis in Gray Voles of the Subgenus *Microtus* (Rodentia, Arvicolinae) and in Their Hybrids. *Genetika* **47**: 968-974.

Salminen, M. 1995. Identification of breakpoints in intergenotypic recombinants of HIV type I by bootscanning. *AIDS Res. Hum. Retrovir.* **11**: 1423-1425.

Sambrook, J., Fritschi, E.F., Maniatis, T. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.

Schneider, T.D., Stephens, R.M. 1990. Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res.* **18**: 6097-6100.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A. *et al.* 2014. Genomics and the origin of species. *Nat. Rev. Genet.* **15**: 176-192.

Shorter, K.R., Crossland, J.P., Webb, D., Szalai, G., Felder, M.R., Vrana, P.B. 2012. *Peromyscus* as a Mammalian Epigenetic Model. *Genet. Res. Int.* **2012**: 179159.

Simmons, M.P., Ochoterena, H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**: 369-381.

Smith, J.M. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**: 126-129.

Spannhof, L. 1960. Histochemische Untersuchungen zur Sterilitat bei mannlichen Saugerbastarden (Artkreuzung der Rotelmause *Clethrionomys glareolus* x *C. rutilus*). *Zool. Anz. Suppl.* **23**: 99-107.

Stephens, M., Donnelly, P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**: 1162-1169.

Stephens, M., Smith, N., Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**: 978-989.

Sutter, A., Beysard, M., Heckel, G. 2013. Sex-specific clines support incipient speciation in a common European mammal. *Heredity* **110**: 398-404.

Swann, C.A., Hope, R.M., Breed, W.G. 2002. cDNA nucleotide sequence encoding the ZPC protein of Australian hydromyine rodents: a novel sequence of the putative sperm-combining site within the family Muridae. *Zygote* **10**: 291-299.

Swann, C.A., Cooper, S.J.B., Breed, W.G. 2007. Molecular evolution of the carboxy terminal region of the zona pellucida 3 glycoprotein in murine rodents. *Reproduction* **133**: 697-708.

Swanson, W.J., Vacquier, V.D. 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3**: 137-144.

Swanson, W.J., Yang, Z., Wolfner, M.F., Aquadro, C.F. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *PNAS* **98**: 2509-2514.

Swanson, W.J., Nielsen, R., Yang, Q.F. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**: 18-20.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of dna sequences. *Lectures Math. Life Sci.* **17**: 57-86.

Turner, L.M., Hoekstra, H.E. 2006. Adaptive Evolution of Fertilization Proteins within a Genus: Variation in ZP2 and ZP3 in Deer Mice (*Peromyscus*). *Mol. Biol. Evol.* **23**: 1656-1669.

Turner, L.M., Hoekstra, H.E. 2008a. Reproductive protein evolution within and between species: maintenance of divergent ZP3 alleles in *Peromyscus*. *Mol. Ecol.* **17**: 2616-2628.

Turner, L.M., Hoekstra, H.E. 2008b. Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev. Biol.* **52**: 769-780.

Villesen, P. 2007. FaBox: an online toolbox for fasta sequences. *Mol. Ecol. Notes* **7**: 965-968.

Visconti, P.E., Florman, H.E. 2010. Mechanisms of Sperm-Egg Interactions: Between Sugars and Broken Bonds. *Sci. Signal.* **3**: pe35.

Walker, L., Rojas, M., Flores, S., Spotorno, A., Manríquez, G. 1999. Genomic compatibility between two phyllotine rodent species evaluated through their hybrids. *Hereditas* **131**: 227-238.

Wassarman, P.M. 1999. Mammalian Fertilization: Review Molecular Aspects of Gamete Adhesion, Exocytosis, and Fusion. *Cell* **96**: 175-183.

Wassarman, P.M., Jovine, L., Qi, H., Williams, Z., Darie, C., Litscher, E.S. 2005. Recent aspects of mammalian fertilization research. *Mol. Cell. Endocrinol.* **234**: 95-103.

Wassarman, P.M., Litscher, E.S. 1995. Sperm-egg recognition mechanisms in mammals. *Curr. Top. Dev. Biol.* **30**: 1-19.

Wassarman, P.M., Litscher, E.S. 2010. Egg's ZP3 structure speaks volumes. *Cell* **143**: 337-338.

Wiking, V.H. 1976. Karyologie und Biologie der beiden iberischen Wuhlmausarten *Pitymys mariae* und *Pitymys duodecimcostatus*. *Z. Zool. Syst. Evol.* **14**: 104-129.

Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* **10**: 1396-1401.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555-556.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**: 568-573.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**:1586-1591.

Yang, Z., Nielsen, R. 1998 Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**: 409-418.

Yang, Z., Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**: 908-917.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.-M.K. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431-449.

Yang, Z., Wong, W.S.W., Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**: 1107-1118.

Zimmermann, K. 1965. Art-Hybriden bei Rotelmausen. *Z . Saugetierk.* **30**: 315-320.

5.1.9 Supporting information

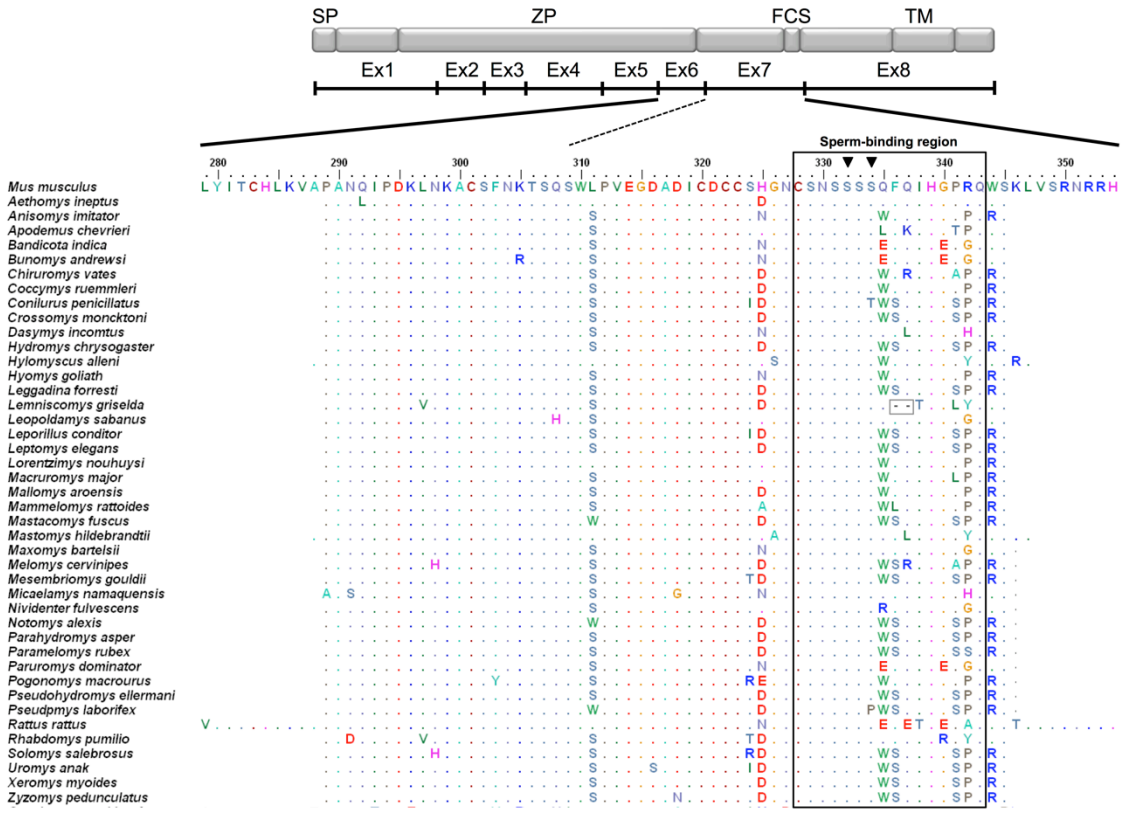


Figure S1 – ZP3 exon 6 and 7 amino acid sequence alignment for Muridae with the schematic position in the mouse ZP3 protein on top. Dots represent amino acids equal to the reference *Mus musculus*. The black box marks the putative sperm-binding region according to Rossiere & Wassarman (1992). Grey squares highlight deletions. SP = signal peptide, ZP = zona domain, FCS = furin cleavage site, TM = transmembrane domain.

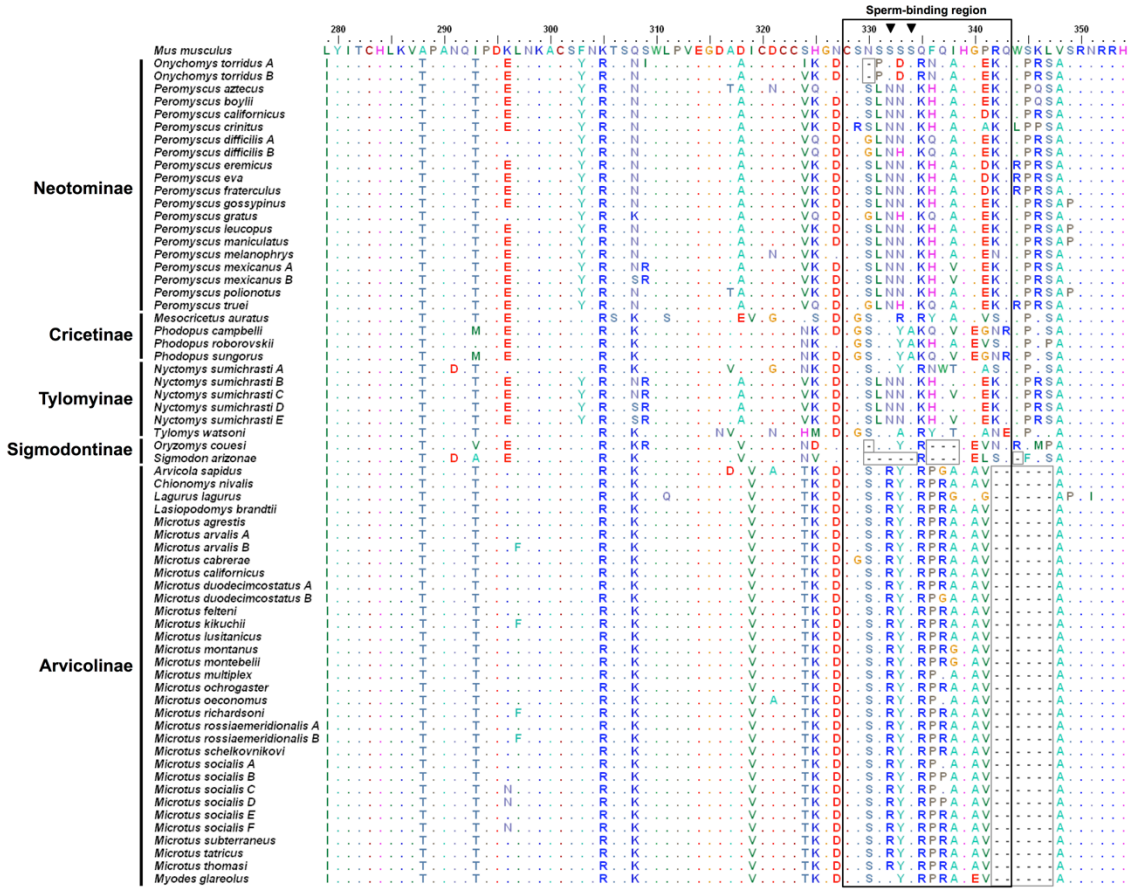


Figure S2 – ZP3 exon 6 and 7 amino acid alignment for Cricetidae from different subfamilies. Dots represent amino acids equal to the reference *Mus musculus*. The black box indicates the putative sperm-binding region. Grey squares highlight deletions relative to *Mus musculus*.

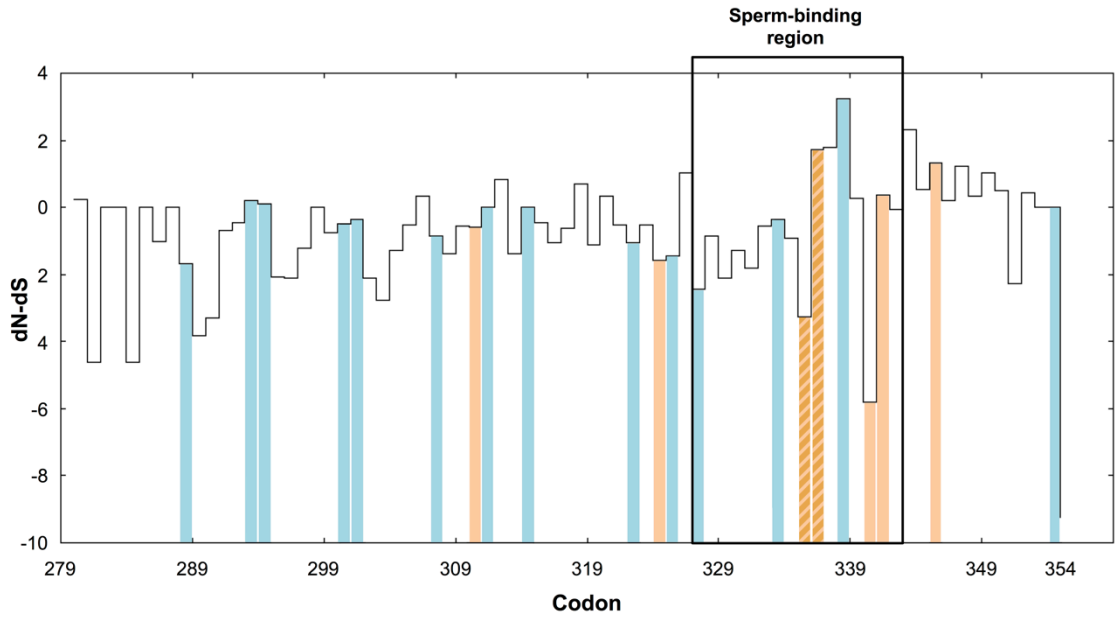


Figure S3 – Distribution of the common sites under selection identified by PAML M2, M3 and M8 models and by HyPhy SLAC, FEL, IFEL and MEME tests ($P < 0.05$), independently. The step line is the normalized dN-dS calculated by SLAC per codon. Positively selected sites according to Hyphy are presented in orange (orange stripes indicate the common sites between PAML and HyPhy methods), while negatively selected sites are in blue. The black box indicates the putative sperm-binding region.

Table S1 – List of species analysed for this study. The family and subfamily taxonomic classification, origin (and voucher code, if applied), sample size for new contributions (N) and respective GenBank accession number are also indicated. MTTU = Museum of Texas Tech University, USA; MNCN = Museo Nacional de Ciencias Naturales, Spain; CMPG = Institute of Ecology and Evolution, University of Bern, Switzerland; IZ-MLU = Institute of Zoology, Martin-Luther University, Germany; MUHNAC = Museu Nacional de História Natural e da Ciência, Portugal.

Taxonomic classification	Subfamily	Origin	N	GenBank accession number
Arvicola sapidus	Arvicolinae, Cricetidae	MUHNAC	2	
Chionomys nivalis		CMPG	2	
Lagurus lagurus		GenBank	-	AF515621
Lasiopodomys brandti		CMPG	1	
Microtus agrestis		MNCN (ES-20909)	1	
Microtus arvalis		CMPG	2	
Microtus cabrerae		MUHNAC	3	
Microtus californicus		CMPG	1	
Microtus duodecimcostatus		MUHNAC	25	
Microtus felteni		CMPG	1	
Microtus kikuchii		CMPG	1	
Microtus lusitanicus		MUHNAC	29	
Microtus montanus		CMPG	2	
Microtus montebelli		CMPG	1	
Microtus multiplex		CMPG	2	
Microtus ochrogaster		CMPG	2	
Microtus oeconomus		CMPG	1	
Microtus richardsoni		CMPG	1	
Microtus rossiaemeridionalis		CMPG	2	
Microtus schelkovnikovi		CMPG	1	
Microtus socialis		CMPG	2	
Microtus subterraneus		CMPG	1	
Microtus tatricus		CMPG	2	
Microtus thomasi		CMPG	2	
Myodes glareolus		MNCN (MNCN-144800 / MNCN-144801)	2	
Mesocricetus auratus	Cricetinae, Cricetidae	IZ-MLU	2	
Phodopus campbelli			1	
Phodopus roborovskii			3	
Phodopus sungorus			3	
Nyctomys sumichrasti	Tylomyinae, Cricetidae	MTTU (TK 19590 / TK 113595)	2	
Tylomys watsoni		MTTU (TK 136061)	1	

Table S1 (continued)

Oryzomys couesi	Sigmodontinae, Cricetidae	MTTU (TK 92437)	1	
Sigmodon arizonae		MTTU (TK 112658)	1	
Onychomys torridus	Neotominae, Cricetidae	GenBank	-	DQ668293+DQ668343 DQ668292+DQ668342
Peromyscus aztecus				DQ668245+DQ668332
Peromyscus boylii				DQ668250+DQ668323
Peromyscus californicus				DQ668253+DQ668341
Peromyscus crinitus				DQ668255+DQ668306
Peromyscus difficilis				DQ668259+DQ668313
Peromyscus eremicus				DQ668262+DQ668310
Peromyscus eva				DQ668263+DQ668337
Peromyscus fraterculus				DQ668264+DQ668338
Peromyscus gossypinus				DQ668267+DQ668319
Peromyscus gratus				EU568656
Peromyscus leucopus				DQ668270+DQ668316
Peromyscus maniculatus				DQ668276+DQ668299
Peromyscus melanophrys				DQ668279+DQ668336
Peromyscus mexicanus				DQ668283+DQ668330
Peromyscus polionotus				EU489722
Peromyscus truei				EU568744

Table S1 (continued)

Aethomys ineptus	Murinae, Muridae	GenBank	-	EU004038
Anisomys imitator				EF364448
Apodemus chevrieri				EU004040
Bandicota indica				EU004041
Bunomys andrewsi				EU004042
Chiruromys vates				EF364449
Coccymys ruemmleri				EF364450
Conilurus penicillatus				EF364451
Crossomys moncktoni				EF364452
Dasymys incomtus				EU004043
Hydromys chrysogaster				EF364453
Hylomyscus alleni				AY057789
Hyomys goliath				EF364454
Leggadina forresti				EF364455
Lemniscomys griselda				EU004044
Leopoldamys sabanus				EU004046
Leporillus conditor				EF364457
Leptomys elegans				EF364458
Lorentzimys nouhuysi				EF364459
Macruromys major				EF364460
Mallomys aroensis				EF364461
Mammelomys rattoides				EF364464
Mastacomys fuscus				EF364465
Mastomys hildebrandtii				AY057790
Maxomys bartelsii				EU004047
Melomys cervinipes				EF364469
Mesembriomys gouldii				EF364472
Micaelamys namaquensis				EU004039
Mus musculus				M20026
Niviventer fulvescens				EU004049
Notomys alexis				EF364474
Parahydromys asper				EF364479
Paramelomys rubex				EF364482
Paruromys dominator				EU004050
Pogonomys macrourus				EF364484
Pseudohydromys ellermani				EF364466
Pseudomys laborifex				EF364499
Rhabdomys pumilio				EU004064
Rattus rattus				Y10823
Solomys salebrosus				EF364507
Uromys anak				EF364508
Xeromys myoides				EF364510
Zyzomys pedunculatus				EF364514

Table S2 – Results of the ZP3 selection tests. For the PAML results, the log likelihood *l* of each model is given as well as the position of positively selected codons (where ω > 1) calculated by the Naive Empirical Bayes (NEB) analysis and Bayes Empirical Bayes (BEB) analysis. For the HyPhy results both positively and negatively selected sites are shown. Sites selected by all the tests performed by PAML or HyPhy (independently) are underlined, while sites selected both by PAML and HyPhy methods are double underlined. *p0* = proportion of sites where ω < 1 (ω0), *p1* = proportion of sites where ω = 1 (ω1), *p2* = proportion of sites where ω > 1 (ω2), p/q = parameters of the beta distribution, [n] = NEB positively selected site, [b] = BEB positively selected site, * = probability > 95%, ** = probability > 99%, - = not applied.

Soft-ware	Model or test	Parameters	κ (ts/tv)	Likelihood *l*	Positively selected sites	Negatively selected sites
PAML	M0: one ratio	ω = 0.45555	4.378	-2800.235353	Not allowed	-
	M1: nearly neutral	*p0* = 0.68617 ω0 = 0.10441 *p1* = 0.31383 ω1 = 1.00000	4.337	-2696.353408		
	M2: selection	*p0* = 0.63784 ω0 = 0.10427 *p1* = 0.26917 ω1 = 1.00000 *p2* = 0.09299 ω2 = 2.88845	4.916	-2675.433307	311[bn] 325[bn**] 336[bn] 337[bn**] 341[bn**] 342[bn**] 346[bn]	
	M3: discrete	*p0* = 0.38025 ω0 = 0.00000 *p1* = 0.47508 ω1 = 0.41204 *p2* = 0.14467 ω2 = 2.17610	4.707	-2663.527989	311[n**] 324[n**] 325[n**] 335[n*] 336[n**] 337[n**] 338[n*] 341[n**] 342[n**] 346[n**] 347[n]	
	M7: beta	p = 0.22961 q = 0.43401	4.212	-2688.433216	Not allowed	
	M8: beta and ω	*p0* = 0.88268 p = 0.33564 q = 0.93167 *p1* = 0.11732 ω = 2.38402	4.735	-2665.317104	311[bn] 324[bn] 325[bn**] 335[bn] 336[bn**] 337[bn**] 341[bn**] 342[bn**] 346[bn]	

Table S2 (continued)

PAML	MA1, foreground branch = Muridae	-	4.383	-1631.463527	315^{bn} 322^{bn}	-
	MA1, foreground branch = Cricetidae	-	4.082	-1626.069192	287^{bn} 307$^{bn*/**}$ 311^{bn*}	-
	MA1, foreground branch = Arvicolinae	-	4.226	-1628.893840	287^{bn**}	-
	Foreground branch = Cricetinae	-	4.242	-1631.709580	296^{b} 309^{b}	-
	MA1, foreground branch = Neotominae	-	4.237	-1632.173401	None	-
	MA1, foreground branch = Sigmodontinae	-	4.231	-1632.171253	None	-
	MA1, foreground branch = Tylomyinae	-	4.162	-1632.052340	307^{b} 311^{bn}	-

Table S2 (continued)

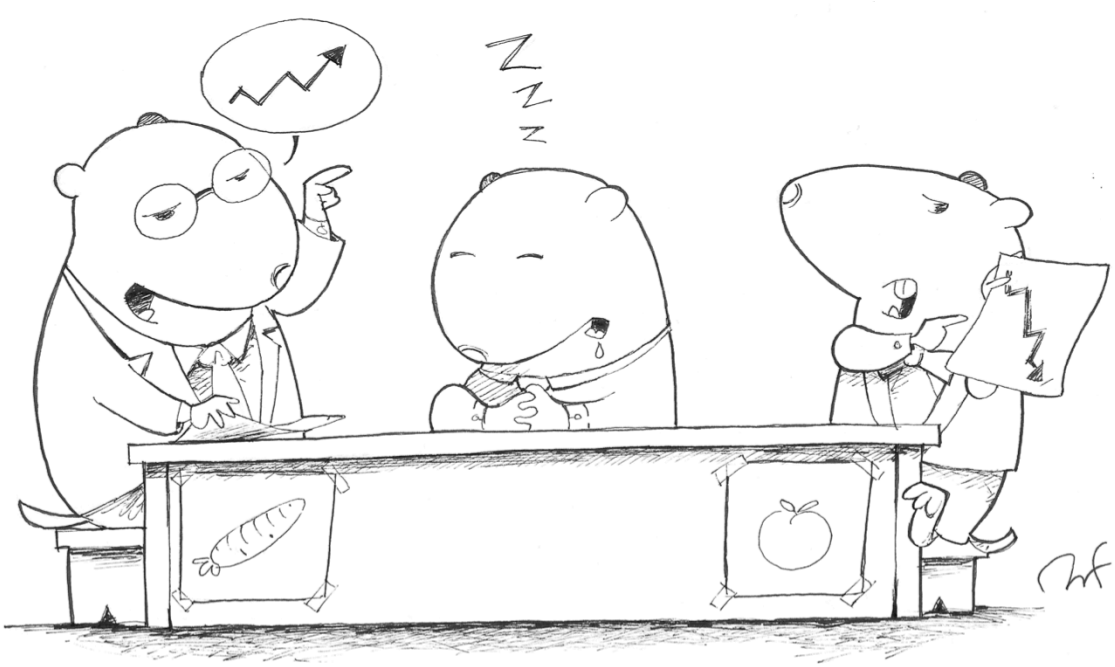
HyPhy	SLAC	-	-	-	337*	280* 283* 288* 289** 294** 295** 301** 302** 307* 312* 315* 323* 326** 328** 330* 334** 339** 354**
	FEL	-	-	-	317* 336* 337**	280* 283* 287** 289** 290* 294** 295** 296* 301** 302** 307** 312** 314* 315** 323** 326** 328** 334** 339** 354**

Table S2 (continued)

HyPhy	IFEL	-	-	-	336** 337* 342**	287** 289** 294** 295* 301* 302** 307* 312* 315* 321* 323* 326** 328* 334* 339** 354**
	MEME	-	-	-	317* 335* 336* 337**	-

Chapter 6

General Conclusion



6.1 Discussion & Conclusion

The present Ph.D. project constitutes a first step in unravelling mechanisms of reproductive isolation between recently diverged sister species *M. lusitanicus* and *M. duodecimcostatus*. I took a multidisciplinary approach based on genetic, proteomic and behavioural data in order to shed a light into this complex and multifactorial subject.

Two pre-zygotic reproductive barriers, behavioural (pre-mating) and gametic (post-mating) isolation, were investigated. Overall, the obtained results 1) indicate that urinary proteins may play a role in species-specific discrimination; 2) confirm social monogamy as the mating system of both voles, being a possible indirect behavioural isolation barrier at syntopy; 3) reveal that individual behavioural variability may contribute to the behavioural isolation between *M. lusitanicus* and *M. duodecimcostatus*; and 4) refute the putative sperm-binding region of ZP3 as a gametic barrier.

Henceforth, I present a discussion of the results disclosed in each Chapter and if they were able to confirm or refute the proposed hypotheses.

6.1.1 Hypothesis 1: Odour cues communication is an active behavioural reproductive barrier between *M. lusitanicus* and *M. duodecimcostatus*.

To test this hypothesis, both receptor (olfactory receptors) and emitted odour molecules (MHCI and MHCII) were analysed in order to infer their potential role as behavioural barriers between *M. lusitanicus* and *M. duodecimcostatus*.

Regarding olfactory receptors, it is known that these proteins are highly variable, consistent with the structural diversity of odour cue molecules (e.g. Emes et al. 2004; Ignatieva et al. 2014). DNA sequences of candidate genes Olfr31 and Olfr57 were analysed (**Chapter 2 – 2.1**), expecting that positively selected amino acids would be found in the extracellular loops and extracellular half of the transmembrane helices of both olfactory receptors, since these variable regions are responsible for the binding of odorous molecules (Emes et al. 2004). DNA

sequences from both olfactory receptor genes revealed the presence of shared haplotypes among various *Microtus* species, particularly in the sister taxa *M. lusitanicus* and *M. duodecimcostatus* (**Chapter 2** – **2.1**). Haplotype sharing and the presence of a majority of negatively selected residues in the extracellular loops of Olfr31 and Olfr57 suggest that these olfactory receptors, concerning the analysed *Microtus* taxa, probably recognize conserved odour cues, with very low or inexistent interspecific variation, not related to behavioural isolation (**Chapter 2** – **2.1**).

MHCI and MHCII were also analysed (**Chapter 2** – **2.2**). These highly polymorphic genes are sources of individuality chemosignals and MHC peptide ligands can function as chemosignals by themselves, forming a direct link between individuality at the immunological and behavioural levels. Standard amplification and sequencing techniques were unable to determine the identity of the amplified loci, either for MHCI or MHCII. Nevertheless, MHCII results indicated a loci multiplication in the analysed *Microtus* taxa, similarly to other Arvicolinae species (e.g. Vincek et al., 1987; Bryja et al., 2006; Axtner & Sommer, 2007; Busch et al., 2008; Penn & Musolf, 2012; Kloch et al., 2013; Winternitz & Wares, 2013).

Additionally, other candidate genes were considered due to their putative role in mating behaviour: urinary MUPs (major urinary proteins, e.g. Hurst et al., 2001; Stockley et al., 2013), lacrimal ESP1 (exocrine gland-secreted peptide 1, e.g. Haga et al., 2010) and ESP36 (exocrine gland-secreted peptide 36, e.g. Kimoto et al., 2007), and lacrimal/salivary ABPa (androgen-binding protein alpha, e.g. Karn & Dlouhy, 1991; Hwang et al., 1997; Laukaitis et al., 1997; Talley et al., 2001). For these candidate genes different pairs of *Mus musculus* published primers were tested (MUP1-25: Stopková et al., 2007; Logan et al., 2008; ESP1 and ESP36: Kimoto et al., 2007; ABPa: Hwang et al., 1997). Unfortunately, amplification was not successful for *M. lusitanicus* and *M. duodecimcostatus*, revealing that even though cross-species amplification has been a valuable methodological approach for a variety of taxa and molecular markers (e.g. Meusnier et al., 2008; Dubut et al., 2010; Hoffmann et al., 2015), including the olfactory receptors analysed here, it is

not always successful due to evolutionary divergence or absence of available genetic data.

Complementarily to these genomic data, protein expression was analysed in the urine of *M. lusitanicus* and *M. duodecimcostatus* (**Chapter 3**). Results partially supported the findings of Hagemeyer and colleagues (2011), which suggest an absence of MUP expression in fossorial taxa. I did not found MUPs in the urine of *M. duodecimcostatus* and only one, Darcin (MUP20) was detected in *M. lusitanicus*. These results seem to indicate that MUPs are not essential to intra- or interspecific communication among these sister voles, Probably, these semi-fossorial species use a non-MUPs route to communicate through odour cues. Thus, I agree with the postulation of Hagemeyer and colleagues (2011), which considers that there is a universal process that transmits semiochemicals across rodents, and that the search must continue, since MUPs are clearly not sufficiently prevalent to achieve this role.

Overall, these results suggest that Olfr31 and Olfr57 are probably not related to behavioural isolation between *M. lusitanicus* and *M. duodecimcostatus* and question the role of urinary proteins in odour cues communication in these species. Consequently, these findings are insufficient to support or refute my hypothesis that odour cues communication is an active behavioural reproductive barrier between *M. lusitanicus* and *M. duodecimcostatus*.

6.1.2 Hypothesis 2: *M. lusitanicus* and *M. duodecimcostatus* prefer conspecific to heterospecific mating in the presence of potential mates of both species.

M. lusitanicus and *M. duodecimcostatus* seem to prefer urinary and faecal odour cues of conspecific over heterospecific potential mates (Soares, 2013). Nonetheless, it is known that these voles can produce F1 hybrids, both in captive heterospecific breeding (Wiking, 1976; Soares, 2013) and in nature (Bastos-Silveira et al., 2012). The former is achievable because only one mate is available, hence

either the subjects mate heterospecifically or remain naïve; while the latter is a rare event due to the apparent presence of behavioural isolation barriers.

Artificial syntopic environments were simulated in order to infer if in the presence of both potential mates simultaneously, a conspecific and a heterospecific, *M. lusitanicus* and *M. duodecimcostatus* would still prefer to mate with the subject of its own species. These simulations also benefited my understanding of the interspecific dynamics between both voles, which may include aggressive behaviour in syntopic locations. This scenario was sustained by fieldwork data, recorded during the course of this Ph.D. project, since we found at Vale Vaqueiros (Portalegre, Portugal) the first present-day syntopic location described for *M. lusitanicus* and *M. duodecimcostatus* (**Chapter 4** – **4.2**). Until here, only two records of syntopy were registered, nevertheless in the past and in different regions of Portugal and Spain (Póvoas et al., 1992; López-García, 2008).

Two environments were established, being composed by one female and one male of each species. The first was kept for 41 days, while the second for 49 days. The former was aborted because the male *M. lusitanicus* was apparently killed by the male *M. duodecimcostatus*. This assumption was corroborated by the presence of fight wounds both on the deceased and on the surviving male. Additionally, constantly bullying towards the male *M. lusitanicus*, either physically or with vocalizations, was exhibited by both *M. duodecimcostatus* subjects (male and female). During such encounters the female *M. lusitanicus* was usually hiding in their nest. Aggressive behaviours were interventioned prior to this death. In this artificial syntopic environment, there was no nest sharing between both taxa and *M. duodecimcostatus* reared two litters.

The second environment, described in detail **Chapter 4** (**4.1**), revealed contrasting results. The male *M. duodecimcostatus* acted monk-like, being most often alone and secluded in its nest, while the female *M. duodecimcostatus* and both *M. lusitanicus* shared daily activities and presented social behaviours. Moreover, the female *M. duodecimcostatus* was videotaped copulating with the male *M. lusitanicus*. Heterospecific copulatory behaviour partially occurred in the

presence of the female *M. lusitanicus*, which groomed the female *M. duodecimcostatus* by the end of copulation. This was the first recorded evidence of spontaneous heterospecific mating behaviour between *M. lusitanicus* and *M. duodecimcostatus* in the presence of both species and sexes. Two litters were born during the duration of this assay, one *M. lusitanicus* conspecific and the other heterospecific, as a result of the copulation between the female *M. duodecimcostatus* and the male *M. lusitanicus*. This outcome questions the role of odour cues communication in these species apparent behavioural isolation and suggests that individual behavioural variability may affect mate choice.

The present results also indicate an aggressive behaviour by *M. duodecimcostatus* towards *M. lusitanicus*, exhibited by and directed to both sexes. *M. duodecimcostatus* was previously considered more aggressive than *M. lusitanicus* (Vinhas 1993), possibly contributing to behavioural reproductive isolation between both species, being the dominant species (Santos, 2009). Moreover, interactions between members of different *M. duodecimcostatus* groups exhibit high levels of aggressiveness (Giannoni, 1994). *M. duodecimcostatus* seems to use substrate-borne signals more often than an acoustic repertoire (Giannoni et al., 1997). Giannoni and colleagues (1997) speculated that the emission of substrate-borne signals in mole voles, and probably in fossorial rodents, would be negatively associated with sociability, and positively associated with individual dominance and/or aggressiveness. This postulation was made when comparing *M. duodecimcostatus* with fellow subgenus *Terricola M. gerbei* (Gerbe, 1879). Since *M. duodecimcostatus*, *M. lusitanicus* and *M. gerbei* are semi-fossorial species and sister species *M. duodecimcostatus* and *M. lusitanicus* are socially monogamous, I do not agree with this speculation. From my personal observations aggression is not negatively related to sociability and seems to be a trait in the presence of less dominant species or individuals.

Considering the present results, my initial hypothesis was partially confirmed since conspecific mating was more common (three conspecific litters) than heterospecific (one litter) copulation. The latter was only possible in the absence

of *M. duodecimcostatus*' aggressive behaviours towards *M. lusitanicus* and when one of the test subjects was unavailable for mating.

6.1.3 Hypothesis 3: *M. lusitanicus* and *M. duodecimcostatus* are socially monogamous.

Ecological and reproduction characteristics, such as balanced sex ratio, K selection strategy, reduced litter size, home range dimension and social organization suggested that *M. lusitanicus* and *M. duodecimcostatus* present a monogamous mating system (Madison, 1980; Madureira, 1982; Wolff, 1985; Salvioni, 1988; Heske & Ostfeld, 1990; MacGuire et al., 1990; Guédon et al., 1991ab; Lambin & Krebs, 1991; Guédon & Pascal, 1993; Paradis & Guédon, 1993; Mira, 1999; Mira & Mathias, 2007; Santos, 2009; Santos et al. 2010; Ventura et al. 2010; Montoto et al., 2011) .

This hypothesis was tested through modified partner preference tests, using urinary and faecal chemical cues alone. This alternative to tethered or confined animals is a viable approach and seems to reduce stress in sensitive animals such as *M. lusitanicus* and *M. duodecimcostatus*; thus, I highly recommend it when performing partner preference tests on these taxa or other *Microtus* voles.

Pair bonding behaviour (**Chapter 4** – **4.2**), exhibited by *M. lusitanicus* and *M. duodecimcostatus* supports a monogamous mating system for these voles, as suggested by ecological and reproduction characteristics. The only absence of partner preference significance was when male individuals had to choose between their partner and a sexually naive female. These observations supported social monogamy with the possibility of rare male extra-pair copulation as the mating system of both sister voles.

In nature, by engaging in extra-pair mating, *M. lusitanicus* and *M. duodecimcostatus* males may have a chance to increase their reproductive success if an extra-pair sexually naive female conceives. The frequency of extra-pair mating may be directly correlated with population density (Say et al., 1999; Dean

et al., 2006; Bryja et al., 2008), possibly because of increased food supply. Considering that variation in population density has been described for both species (Cotilla & Palomo, 2007; Mira & Mathias, 2007), it is plausible to consider that genetic to social monogamy occurs during episodes of low food resources, while social monogamy with the possibility of rare male extra-pair mating surfaces when food supplies are vast.

Based on these partner preference tests, I expect that among natural populations of *M. lusitanicus* and *M. duodecimcostatus* paired males may encounter sexually naive females. Thus, it is reasonable to consider that the social monogamous mating system exhibited by *M. lusitanicus* and *M. duodecimcostatus* may indirectly contribute to their reproductive isolation in syntopic locations. These putative heterospecific encounters could challenge the pair bond and copulation may occur outside the established breeding pair, enabling rare hybridization between the two species, in agreement with the results of Bastos-Silveira and colleagues (2012).

6.1.4 Hypothesis 4: The putative sperm-binding region of ZP3 is a gametic isolation barrier that impairs heterospecific mating between *M. lusitanicus* and *M. duodecimcostatus*.

Ooocyte and sperm surface proteins play an important role in reproductive isolation (reviewed in Turner & Hoekstra, 2008). They constitute species-specific barriers to fertilization, through post-mating gametic isolation, and potentially contribute to speciation (Swanson & Vacquier, 2002; Findlay & Swanson, 2010).

One of them is ZP3 (zona pellucida 3), a glycoprotein of the oocyte zona pellucida of mammals. Its putative sperm-binding region exhibits considerable amino acid variation between species, which may together with modifications in the structure of the O-linked glycans enable a species-specific binding of the sperm to the oocyte (Wassarman & Litscher, 1995; Wassarman, 1999; Wassarman et al., 2005).

In **Chapter 5** evolutionary patterns in the putative sperm-binding region were analysed, in order to determine if species-specific amino acid sequences would be found in *M. lusitanicus* and *M. duodecimcostatus* individuals. This study focused on the hyperdiverse Cricetidae family and particularly on the speciose *Microtus* genus, aiming to contribute with a comprehensive comparative analysis comprising the two most important radiations of rodents.

The presence of amino acid deletions in and around the putative sperm-binding region of most Cricetidae species analysed, together with an absence of conservation in positions 332 and 334 lead to the refutation of the classical model of ZP3 O-linked glycan sperm-oocyte binding. Results also indicate that this region does not constitute a species-specific barrier as previously suggested (Florman & Wassarman, 1985; Chen et al., 1998) since shared amino acid sequence types were found between different taxa. Thus, the putative sperm-binding region of ZP3 *per se* is not a gametic isolation barrier. Nonetheless, our findings do not question the role of ZP3 in sperm recognition together with other protein(s) from the zona pellucida and sperm head. These unforeseen results constituted a step forward in an evolutionary understanding of the role of the putative sperm-binding region of ZP3 in mammals.

The species-specificity function of ZP3 might be related to protein sections other than the putative sperm-binding region. This is consistent with the propositions that: i) sperm binds to ZP3 by interacting with O-linked glycans not connected to S-332 and S-334 (Visconti & Florman, 2010) or to N-linked glycans and accessible protein regions located within the C-terminal domain of ZP3 (Clark et al., 2011); ii) two conserved O-linked glycosylation sites (residues T-155 and T-162/S-164/S-165) shared by mouse and human ZP3 may be the actual attachment sites of the sperm-binding glycans (Chalabi et al., 2006); and iii) these two conserved O-linked glycosylation sites and the putative sperm-binding region are exposed on the same 3D protein surface, indicating that multiple distinct binding sites might be involved in sperm-oocyte recognition (Monné et al., 2011).

In consideration of the present findings, I postulate that amino acid deletions in and around the ZP3 putative sperm-binding region may compromise interaction stability between gametes, potentially impairing the species-specificity of fertilization and partially disrupting this isolation barrier.

Concluding, the hypothesis that the putative sperm-binding region of ZP3 is a gametic isolation barrier between *M. lusitanicus* and M. *duodecimcostatus* is refuted.

6.2 Future directions

The present Ph.D. project clarified some mechanisms of reproductive isolation between *M. lusitanicus* and M. *duodecimcostatus* and raised new questions and directions for future works. I consider that complementing genetic, proteomic and behavioural approaches should be considered in forthcoming studies.

Forthcoming genetic analyses should include transcriptome analysis (e.g. RNA-seq: Hoeijmakers et al., 2013; Mutz et al., 2013; Wolf, 2013), to infer which proteins are being differently expressed in *M. lusitanicus* and *M. duodecimcostatus*, both at the qualitative and quantitative level. Such methodology would indicate which proteins related to odour cues communication and sperm-oocyte interactions may be related to behavioural and gametic isolation barriers between *M. lusitanicus* and *M. duodecimcostatus*, respectively. Moreover, pyrosequencing or next generation sequencing (reviewed in Wegner, 2009; Babik, 2010), both from sympatric and allopatric individuals would clarify the role of genic complexes such as MHCI and MHCII, including the number of expressed loci, since multiplication has been recorded for other rodents, including Arvicolinae species (e.g. Vincek et al., 1987; Bryja et al., 2006; Axtner & Sommer, 2007; Busch et al., 2008; Penn & Musolf, 2012; Winternitz & Wares, 2013).

At the proteomic level, I consider that the next step would be to analyse both male and female *M. lusitanicus* and *M. duodecimcostatus* urine, saliva and tears for the presence of species-specific odour cues proteins, both at the qualitative or quantitative level.

In terms of behavioural assays, I believe that it would be interesting to test the same premises inferred by the artificial syntopic environments but in a larger scale, e.g. follow a confined heterospecific population living under natural conditions, which could better simulate a natural syntopic population. This direction could help to clarify if conspecific breeding is favoured over heterospecific copulation; and if aggressive behaviours are expected from *M. duodecimcostatus* towards *M. lusitanicus* or if they were enhanced in one of the setups due to the close proximity between both taxa. A valid alternative would be to capture, mark, genotype, release and re-capture Vale Vaqueiros syntopic adult individuals and determine if hybrid juveniles emerge in the population.

Furthermore, to elucidate a probable aggressive and dominance status of *M. duodecimcostatus* over *M. lusitanicus*, I suggest the implementation of intrasexual aggression assays, using both species and sexes, in order to understand which species presents a higher amount of aggressive behaviour (e.g. Randall, 1978; Wolff et al., 1983; Dempster & Perrin, 1990; Courtalon et al., 2003; Lancaster & Pillay, 2010; Dupre et al., 2015). It will be also interesting to complement such a study with a comparison of scent-marking, to infer if isolated and adjacent marking are different between taxa, a sign of subordinacy, or if over-marking exists, a sign of dominance (e.g. Ferkin, 1999; Becker et al., 2012; Hurst, 2005).

Concerning the putative role of social monogamy as an indirect behavioural reproductive isolation barrier, heterospecific PPT could be performed to test this hypothesis. It would enable the clarification of the male extra-pair mating scenario in the presence of both taxa, such as in syntopic locations and would contribute to the understanding of interspecific population dynamics between these sister voles.

6.3 References

Axtner J, Sommer S (2007) Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of the bank vole, *Clethrionomys glareolus*. Immunogenetics 59(5):417–426

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

Becker EA, Petruno S, Marler CA (2012) A Comparison of Scent Marking between a Monogamous and Promiscuous Species of *Peromyscus*: Pair Bonded Males Do Not Advertise to Novel Females. PLoS ONE 7(2):e32002

Bryja J, Galan M, Charbonnel N, Cosson JF (2006) Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics 58(2-3):191–202

Bryja J, Patzenhauerová H, Albrecht T, Mosansky L, Stanko M, Stopka P (2008) Varying levels of female promiscuity in four *Apodemus* mice species. Behav Ecol Sociobiol 63:251–260

Busch JD, Waser PM, DeWoody JA (2008) Characterization of expressed class II MHC sequences in the banner-tailed kangaroo rat (*Dipodomys spectabilis*) reveals multiple DRB loci. Immunogenetics 60(11):677–688

Chalabi S, Panico M, Sutton-Smith M, Haslam SM, Patankar MS, Lattanzio FA et al. (2006) Differential O-glycosylation of a conserved domain expressed in murine and human ZP3. Biochem 45:637–647

Chen J, Litscher ES, Wassarman PM (1998) Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. Proc Natl Acad Sci U S A 95:6193–6197

Clark GF (2011) The molecular basis of mouse sperm–zona pellucida binding: a still unresolved issue in developmental biology. Reproduction 142:377–381

Cotilla I, Palomo LJ (2007) *Microtus duodecimcostatus* (de Selys-Longchamps, 1839). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 422–425

Courtalon P, Dolcemascolo A, Trioano V, Alvarez MR, Busch M (2003) Inter and intraspecific relationships in *Akodon azarae* and *Calomys laucha* (Rodentia, Sigmodontinae) in Pampean agroecosystems. Mastozool Neotrop 10(1):27–39

Dean MD, Ardlie KG, Nachman MW (2006) The frequency of multiple paternity suggests that sperm competition is common in house mice (*Mus domesticus*). Mol Ecol 15(13):1441–1451

Dempster ER, Perrin MR (1990) Interspecific aggression in sympatric *Gerbillurus* species. Z Säeugetierkunde 55(6):392–398

Duarte MA, Mathias ML, Bastos-Silveira C (2015) Pair-bonding behaviour of the sister species *Microtus lusitanicus* and *M. duodecimcostatus*. J Ethol 33:213–223

Duarte MA, Heckel G, Mathias ML, Bastos-Silveira C (2016) Olfactory receptors and behavioural isolation: a study on *Microtus* voles. Mammal Res

Dubut V, Sinama M, Martin JF, Meglécz E, Fernandez J, Chappaz R, Gilles A, Costedoat C (2010) Cross-species amplification of 41 microsatellites in European cyprinids: A tool for evolutionary, population genetics and hybridization studies. BMC Res Notes 3:135

Dupre SI, Lambert TD, Adler GH, Hegde LM, Kennedy EH (2015) Aggressive and investigative behaviors of sympatric species of echimyid rodents, *Proechimys semispinosus* and *Hoplomys gymnurus*, in Central Panama. Ethol Ecol Evol

Emes RD, Beatson SA, Ponting CP, Goodstadt L (2004) Evolution and Comparative Genomics of Odorant- and Pheromone-Associated Genes in Rodents. Genome Res 14:591–602

Findlay GD, Swanson WJ (2010) Proteomics enhances evolutionary and functional analysis of reproductive proteins. BioEssays 32:26–36

Ferkin MH (1999) Meadow Voles (*Microtus pennsylvanicus*, Arvicolidae) Over-mark and Adjacent-mark the Scent Marks of Same-sex Conspecifics. Ethology 105:825–837

Florman HM, Wassarman PM (1985) O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. Cell 41:313–324

Giannoni SM, Borghi CE, Martinezrica JP (1994) Swimming ability of the Mediterranean pine vole *Microtus* (*Terricola*) *duodecimcostatus*. Acta Theriol 39(3):257–265

Giannoni SM, Marquez R, Borghi CE (1997) Airborne and substrate borne communications of *Microtus* (*Terricola*) *gerbei* and *M.* (*T.*) *duodecimcostatus*. Acta Theriol 42(2):123–141

Guédon G, Pascal M (1993) Dynamique de population du campagnol provençal (*Pitymys duodecimcostatus* de Sélys-Longchamps, 1839) dans deux agrosystémes de la région montpelliéraine. Rev Ecol – Terre Vie 48:375–398

Haga S, Hattori T, Sato T, Sato K, Matsuda S, Kobayakawa R, Sakano H, Yoshihara Y, Kikusui T, Touhara K (2010) The male mouse pheromone ESP1 enhances female sexual receptive behaviour through a specific vomeronasal receptor. Nature 466(7302):118–222

Hagemeyer P, Begall S, Janotova K, Todrank J, Heth G, Jedelsky PL, Burda H, Stopka PJ (2011) Searching for major urinary proteins (MUPs) as chemosignals in urine of subterranean rodents. Chem Ecol 37(7):687–694

Heske EJ, Ostfeld RS (1990) Sexual dimorphism in size, relative size of testes, and mating systems in North American voles. J Mamm 71(4):510–519

Hoeijmakers WA, Bártfai R, Stunnenberg HG (2013) Transcriptome analysis using RNA-Seq. Methods Mol Biol. 923:221–239

Hoffmann GS, Johannesen J, Griebeler EM (2015) Species cross-amplification, identification and genetic variation of 17 species of deer (Cervidae) with microsatellite and mitochondrial DNA from antlers. Mol Biol Rep 42(6):1059–1067

Hurst JL, Payne CE, Nevison CM, Marie AD, Humphries RE, Robertson DHL, Cavaggioni A, Beynon RJ (2001) Individual recognition in mice mediated by major urinary proteins. Nature 414:631–634

Hurst J (2005) Making sense of scents: reducing aggression and uncontrolled variation in laboratory mice. NC3Rs 2:1–8

Hwang JM, Hofstetter JR, Bonhomme F, Karn RC (1997) The microevolution of mouse salivary androgen-binding protein (ABP) paralleled subspeciation of Mus musculus. J Hered 88(2):93–97

Ignatieva EV, Levitsky VG, Yudin NS, Moshkin MP, Kolchanov NA (2014) Genetic basis of olfactory cognition: extremely high level of DNA sequence polymorphism

in promoter regions of the human olfactory receptor genes revealed using the 1000 Genomes Project dataset. Front Psychol 5:247

Karn RC, Dlouhy SR (1991) Salivary androgen-binding protein variation in Mus and other rodents. J Hered 82(6):453–458

Kimoto H, Sato K, Nodari F, Haga S, Holy TE, Touhara K (2007) Sex- and strain-specific expression and vomeronasal activity of mouse ESP family peptides. Curr Biol 17(21):1879–1884

Kloch A, Baran K, Buczek M, Konarzewski M, Radwan J (2013) MHC influences infection with parasites and winter survival in the root vole *Microtus oeconomus*. Evol Ecol 27:635–653

Lambin X, Krebs CJ (1991) Spatial organization and mating system of *Microtus townsendii*. Behav Ecol Sociobiol 28:353–363

Lancaster J, Pillay N (2010) Behavioral interactions between a coexisting rodent *Micaelamys namaquensis* and macroscelid *Elephantulus myurus*. Curr Zool 56(4):395–400

Laukaitis CM, Critser ES, Karn RC (1997) Salivary Androgen-Binding Protein (ABP) Mediates Sexual Isolation in *Mus musculus*. Evolution 51(6):2000–2005

Logan DW, Marton TF, Stowers L (2008) Species specificity in major urinary proteins by parallel evolution. PLoS One 3(9):e3280

López-García JM, Blain HA, Cuenca-Bescós G, Arsuaga JL (2008) Chronological, environmental, and climatic precisions on the Neanderthal site of the Cova del Gegant (Sitges, Barcelona, Spain). J Hum Evol 55(6):1151–1155

MacGuire B, Pissuto T, Getz LL (1990) Patterns of visitation in prairie voles (*Microtus ochrogaster*) reveal a role for males in population regulation. In: Tamarin RH et al. (eds) Social Systems and Population Cycles in Voles. Advances in Life Sciences, Birkhäuser Verlag, Basel, pp 89–100

Madison DM (1980) An integrated view of the social biology of *Microtus pennsylvanicus*. Biologist 62:20–33

Madureira ML (1982) New data on the taxonomy of portuguese *Pitymys*. Arq Mus Bocage A I:341–368

Meusnier I, Singer GA, Landry JF, Hickey DA, Hebert PD, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9:214

Mira A (1999) Ecologia do rato-cego-mediterrânico, *Microtus* (*Terricola*) *duodecimcostatus* (Rodentia, Muridae), em pomares de citrinos. Dissertation, Faculdade de Ciências, Universidade de Lisboa, Lisboa

Mira A, Mathias ML (2007) *Microtus lusitanicus* (Gerbe, 1879). In: Palomo LJ, Gisbert J, Blanco JC (eds) Atlas y libro rojo de los mamíferos terrestres de España. Ministerio de Medio Ambiente, Madrid, pp 418–421

Monné M, Jovine L (2011) A Structural View of Egg Coat Architecture and Function in Fertilization. Biol Reprod 85:661–669

Montoto LG, Magaña C, Tourmente M, Martín-Coello J, Crespo C, Luque-Larena JJ, Gomendio M, Roldan ERS (2011) Sperm competition, sperm numbers and sperm quality in muroid rodents. PLoS One 6(3):e18173

Mutz KO, Heilkenbrinker A, Lönne M, Walter JG, Stahl F (2013) Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol 24(1):22-30

Paradis E, Guédon G (1993) Demography of a Mediterranean microtine: the Mediterranean pine vole, *Microtus duodecimcostatus*. Oecologia 95:47–53

Penn DJ, Musolf K (2012) The evolution of MHC diversity in house mice. In: Macholán M, Baird SJE, Munclinger P, Piálek J (eds) Evolution of the house mouse. Cambridge University Press, Cambridge, pp 221–252

Póvoas L, Zilhão J, Chaline J, Brunet-Lecomte P (1992) La faune de rongeurs du Pléistocene supérieur de la frotte de Caldeirão (Tomar, Portugal). Quaternaire 3(1):40–47

Randall JA (1978) Behavioral Mechanisms of Habitat Segregation between Sympatric Species of *Microtus*: Habitat Preference and Interspecific Dominance. Behav Ecol Sociobiol 3:187–202

Robertson DHL, Hurst JL, Searle JB, Gunduz I, Beynon RJ (2007) Characterization and comparison of major urinary proteins from the house mouse, *Mus musculus domesticus*, and the aboriginal mouse, *Mus macedonicus*. J Chem Ecol 33:613–630

Salvioni M (1988) Home range and social behaviour of three species of European *Pitymys* (Mammalia, Rodentia). Behav Ecol Sociobiol 22:203–210

Santos SML (2009) Factors influencing the distribution of the lusitanian and the mediterranean pine voles (*Microtus lusitanicus* and *M. duodecimcostatus*) in Portugal: a multiscale approach. Dissertation, Faculdade de Ciências, Universidade de Lisboa, Lisboa

Santos SM, Lourenço RF, Mathias ML, Mira AP (2010) Spatial and temporal ecology of the Lusitanian pine vole (*Microtus lusitanicus*) in a Mediterranean polyculture. Anim Biol 60:209–227

Say L, Pontier D, Natoli E (1999) High variation in multiple paternity of domestic cats (*Felis catus* L.) in relation to environmental conditions. Proc R Soc Lond B 266:2071–2074

Soares JMA (2013) Fitness dos híbridos entre duas espécies de ratos fossadores, *Microtus lusitanicus* (Gerbe, 1879) e *M. duodecimcostatus* (de Selys-Longchamps, 1839) - uma abordagem multidisciplinary. Dissertation, Faculdade de Ciências da Universidade de Lisboa, Lisboa

Stockley P, Bottell L, Hurst JL (2013) Wake up and smell the conflict: odour signals in female competition. Philos Trans R Soc Lond B Biol Sci 368: 20130082

Stopková R, Stopka P, Janotová K, Jedelský PL (2007) Species-specific expression of major urinary proteins in the house mice (*Mus musculus musculus* and *Mus musculus domesticus*). J Chem Ecol 33(4):861–869

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. Nat Rev Genet 3:137–144

Talley HM, Laukaitis CM, Karn RC (2001) Female preference for male saliva: implications for sexual isolation of *Mus musculus* subspecies. Evolution 55(3):631–634

Turner LM, Hoekstra HE (2008) Causes and consequences of the evolution of reproductive proteins. Int J Dev Biol 52:769–780

Ventura J, Jiménez L, Gisbert J (2010) Breeding characteristics of the Lusitanian pine vole, *Microtus lusitanicus*. Anim Biol 60:1–14

Vincek V, Nizetić D, Golubić M, Figueroa F, Nevo E, Klein J (1987) Evolutionary expansion of Mhc class I loci in the mole-rat, *Spalax ehrenbergi*. Mol Biol Evol 4(5):483–491

Vinhas A (1993) *Microtus lusitanicus* (Rato cego) e *Microtus duodecimcostatus* (Rato cego-mediterrânico), roedores pragas das culturas. Rev Ciên Agrár 17:375–382

Visconti PE, Florman HE (2010) Mechanisms of Sperm-Egg Interactions: Between Sugars and Broken Bonds. Sci Signal 3:pe35

Wassarman PM (1999) Mammalian Fertilization: Review Molecular Aspects of Gamete Adhesion, Exocytosis, and Fusion. Cell 96:175–183

Wassarman PM, Litscher ES (1995) Sperm-egg recognition mechanisms in mammals. Curr Top Dev Biol 30:1–19

Wassarman PM, Jovine L, Qi H, Williams Z, Darie C, Litscher ES (2005) Recent aspects of mammalian fertilization research. Mol. Cell. Endocrinol 234:95–103

Wegner KM (2009) Massive parallel MHC genotyping: titanium that shines. Mol Ecol 18(9):1818–1820

Wiking VH (1976) Karyologie und Biologie der beiden iberischen Wuhlmausarten *Pitymys mariae* und *Pitymys duodecimcostatus*. Z Zool Syst Evol 14:104–129

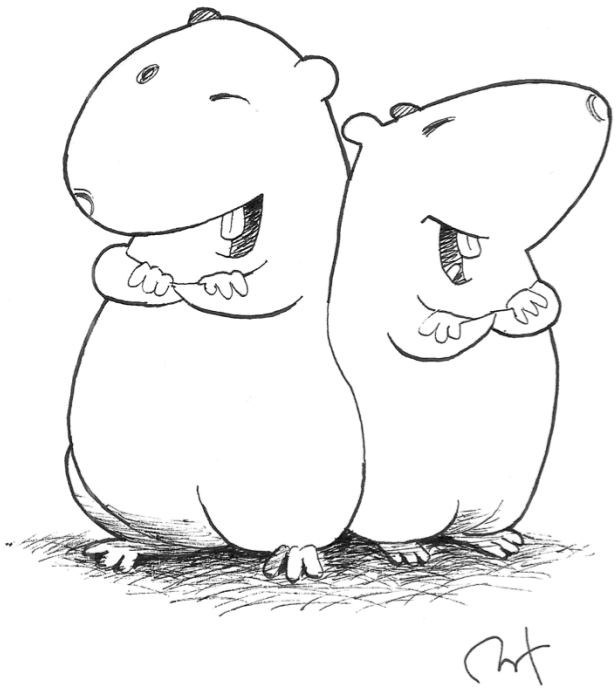
Winternitz JC, Wares JP (2013) Duplication and population dynamics shape historic patterns of selection and genetic variation at the major histocompatibility complex in rodents. Ecol Evol 3(6):1552–1568

Wolf JB (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Mol Ecol Resour 13(4):559–572

Wolff JO, Freeberg MH, Dueser RD (1983) Interspecific territoriality in two sympatric species of *Peromyscus* (Rodentia: Cricetidae). Behav Ecol Sociobiol 12(3):237–242

Wolff JO (1985) Behaviour. In: Tamarin RH (ed) Biology of new world *Microtus*. American Society of Mammalogists, Shippensburg, pp 340–372

Annex



Annex – Corresponding proteins to the labelled peptides detected in the urine of *M. lusitanicus* (ML) and *M. duodecimcostatus* (MD), in alphabetical order. General function(s) and UniProt accession number(s) (http://www.uniprot.org) are also indicated.

Protein	Accession number(s)	General function(s)	Identified in ML	Identified in MD
3-isopropylmalate dehydrogenase	P50455	Catalytic activity	✔	
Acid ceramidase	Q17QB3 Q6P7S1	Catalytic activity	✔	✔
Actin-10	Q54GX7	Structural constituent of cytoskeleton	✔	
Actin, cytoplasmic 1	P84336	Structural molecule	✔	✔
Actin, cytoplasmic 2	Q5ZMQ2	Structural constituent of cytoskeleton	✔	
Acylcarnitine hydrolase	Q91WG0	Catalytic activity	✔	
Acylphosphatase-2	P00821	Catalytic activity	✔	
Adenosylhomocysteinase	Q3MHL4	Catalytic activity	✔	
Adiponectin	Q60994	Hormone Binding	✔	
Alcohol dehydrogenase [NADP(+)]	P14550	Catalytic activity	✔	
Aldehyde oxidase 3	Q5QE80	Catalytic activity	✔	
Aldehyde oxidase 4	Q3TYQ9 Q5QE79	Catalytic activity	✔	✔
Aldo-keto reductase family 1 member C21	Q91WR5	Catalytic activity	✔	
Alkaline phosphatase, tissue-nonspecific isozyme	P09242	Binding	✔	
Alpha-1-antiproteinase 2	P38029	Inhibitor activity	✔	✔
Alpha-1-antitrypsin	P97277	Inhibitor activity	✔	✔
Alpha-2-antiplasmin	Q61247	Inhibitor activity Binding	✔	
Alpha-amylase 1	P00687 P04745	Catalytic activity	✔	✔
Alpha-amylase 2B	P19961	Catalytic activity	✔	
Alpha-N-acetylgalactosaminidase	P17050 Q66H12	Catalytic activity		✔
Alpha-S1-casein	P02662	Antioxidant activity Transporter activity	✔	✔
Alpha-S2-casein	P02663	Binding Transporter activity	✔	✔
Aminopeptidase N	O57579 P15144 P15684	Catalytic activity Binding	✔	✔
Angiopoietin-related protein 2	Q9UKU9	Binding	✔	
Anionic trypsin-1	P00762	Catalytic activity Binding	✔	
Annexin A2	P07356 Q6TEQ7	Inhibitor activity Binding	✔	

Annex (continued)

Annexin A5	Q5R1W0	Binding	✔	
Annexin A11	P27214	Binding		✔
Antithrombin-III	P32262 Q5R5A3	Catalytic activity Binding	✔	
Apolipoprotein D	P51910	Transporter activity Binding	✔	
Apolipoprotein E	P02650 P08226	Binding and catabolism of lipoproteins	✔	✔
Aquaporin-1	Q02013	Transporter activity	✔	
Arginase-1	Q2KJ64	Catalytic activity	✔	
Arginine--tRNA ligase	Q492L0	Catalytic activity Binding	✔	
Aspartate aminotransferase, cytoplasmic	P13221	Catalytic activity Binding	✔	
Aspartate carbamoyltransferase	Q9PNJ6	Catalytic activity	✔	
Aspartate carbamoyltransferase regulatory chain	O58452	Binding	✔	
Attractin	Q9WU60	Binding Receptor activity		✔
Basal cell adhesion molecule	Q9ESS6	Receptor activity	✔	
Beta-1,4-glucuronyltransferase 1	Q8BWP8	Catalytic activity	✔	
Beta-2-glycoprotein 1	Q01339	Binding	✔	
Beta-2-microglobulin	Q9WV24	Antigen binding	✔	
Beta-casein	P02666 Q9TSI0	Transporter activity	✔	✔
Beta-glucuronidase	P06760 P12265	Catalytic activity Binding	✔	✔
Beta-hexosaminidase subunit alpha	P06865 Q641X3	Catalytic activity	✔	✔
Beta-hexosaminidase subunit beta	Q6AXR4	Catalytic activity	✔	✔
Beta-lactoglobulin	P02754 P02755	Binding		✔
Beta-mannosidase	Q95327	Catalytic activity		✔
Biotinidase	A6QQ07 Q8CIF4	Catalytic activity	✔	
Bleomycin hydrolase	Q13867	Catalytic activity	✔	
Cadherin-1	F1PAA9 P09803	Binding	✔	✔
Calbindin	Q5R4V1	Binding	✔	
Carboxylesterase 1C	P10959	Catalytic activity	✔	✔
Carboxylesterase 1D	P16303 Q8VCT4	Catalytic activity		✔
Carboxylesterase 1E	Q64176	Catalytic activity	✔	✔
Carboxypeptidase A4	Q6P8K8 Q9UI42	Catalytic activity Binding	✔	

Annex (continued)

Carboxypeptidase E	Q00493	Catalytic activity Binding		✔
Carboxypeptidase Q	Q5RDN7 Q6IRK9 Q9Y646	Catalytic activity	✔	
Carcinoembryonic antigen-related cell adhesion molecule 5	Q3UKK2	Structural molecule	✔	
Caspase-14	P31944	Catalytic activity	✔	
Cathepsin B	P07858 P10605	Catalytic activity Binding	✔	
Cathepsin D	P00795 P24268 Q4LAL9	Catalytic activity	✔	✔
Cathepsin E	P25796	Catalytic activity		✔
Cathepsin L1	P07154	Catalytic activity Binding		✔
Cathepsin L2	O60911	Catalytic activity	✔	
Cathepsin Z	Q9R1T3	Catalytic activity	✔	
Cation-independent mannose-6-phosphate receptor	Q07113	Binding Transporter activity	✔	
Cationic trypsin-3	P08426	Catalytic activity Binding	✔	✔
Cell adhesion molecule 1	Q9BY67	Binding	✔	
CD44 antigen	Q05078	Receptor activity	✔	
Choline transporter-like protein 4	Q6MG71	Acetylcholine biosynthesis and secretion Regulation of cell growth	✔	
Chondroadherin	O15335	Binding		✔
Clusterin	P05371 P14683 Q06890 Q29549 Q9XSC5	Binding	✔	✔
Collagen alpha-3(VI) chain	P12111	Structural molecule Inhibitor activity	✔	
Collectrin	Q9HBJ8	Catalytic activity	✔	
Complement C3	P01025 P01026 Q2UVX4	Inflammatory response Binding	✔	✔
Copper transport protein ATOX1	O08997	Binding Transporter activity	✔	
Cubilin	O70244 Q9JLB4	Binding Transporter activity Receptor activity	✔	
Cystatin-A	P01040	Inhibitor activity Binding Structural molecule	✔	

Annex (continued)

Cystatin-C	O19093	Inhibitor activity	✔	
Cysteine-rich and transmembrane domain-containing protein 1	Q9H1C7	Structural molecule	✔	
Dehydrogenase/reductase SDR family member 4	Q9GKX2	Catalytic activity	✔	
Deoxyribonuclease-1	P21704	Catalytic activity	✔	✔
Desmoplakin	P15924	Binding Structural constituent of cytoskeleton	✔	
Dipeptidase 1	P31428	Catalytic activity Binding	✔	
Dipeptidyl peptidase 2	Q9ET22	Catalytic activity	✔	✔
Disease resistance protein RGA2	Q7XBQ9	Binding	✔	
DnaJ homolog subfamily B member 11	Q9UBS4	Binding Folding		✔
Ectopic P granules protein 5 homolog	Q0IEK6	Autophagy	✔	
EGF-containing fibulin-like extracellular matrix protein 1	O35568	Binding	✔	
EGF-containing fibulin-like extracellular matrix protein 2	Q9WVJ9	Binding	✔	
Endoplasmic reticulum aminopeptidase 1	Q9NZ08	Catalytic activity Binding		✔
Endoplasmic reticulum resident protein 44	Q9D1Q6	Catalytic activity	✔	
Epididymal secretory protein E1	P61918	Binding	✔	
Fatty acid-binding protein, heart	Q99P61	Binding Transporter activity	✔	
Ferritin heavy chain	Q2MHN2	Catalytic activity Binding		✔
Fibronectin	P11276 Q91740	Binding	✔	
Filaggrin	P20930	Structural molecule	✔	
Filaggrin-2	Q5D862	Structural molecule	✔	
Flagellin	P80583 Q05203	Structural molecule	✔	✔
Fructose-bisphosphate aldolase B	Q91Y97	Catalytic activity Binding	✔	
Fumarylacetoacetase	P35505	Catalytic activity	✔	
Galectin-3-binding protein	P70117	Receptor activity	✔	
Gamma-glutamylcyclotransferase	O75223	Catalytic activity	✔	
Gamma-glutamyl hydrolase	Q92820	Catalytic activity	✔	
Gamma-glutamyl phosphate reductase	Q8YV15	Catalytic activity	✔	
Gamma-glutamyltranspeptidase 1	P07314 Q60928	Catalytic activity	✔	

Annex (continued)

Ganglioside GM2 activator	Q60648	Catalytic activity Activator activity Transporter activity	✔	
Gasdermin-A	Q9EST1	Apoptotic process	✔	
Gelsolin	P06396 Q3SX14 Q68FP1	Binding	✔	✔
Glandular kallikrein-7, submandibular/renal	P36373	Catalytic activity	✔	
Glia-derived nexin	P07093 Q07235	Catalytic activity Binding Protease inhibitor	✔	✔
Glucosylceramidase	P17439	Catalytic activity	✔	
Glutathione peroxidase 6	Q64625	Catalytic activity	✔	
Glutathione S-transferase P	P46424	Catalytic activity	✔	
Glutathione synthetase	P46413	Catalytic activity Binding	✔	
Glyceraldehyde-3-phosphate dehydrogenase	P04406 P16858	Catalytic activity Binding	✔	
Golgi apparatus protein 1	Q9Z1E9	Binding	✔	
Granulins	P28798	Binding	✔	
Group XV phospholipase A2	Q8NCC3	Catalytic activity Binding	✔	
Haptoglobin	O35086 P06866	Antioxidant	✔	
Hemoglobin subunit beta	B3EWE4	Transporter activity		✔
Hemopexin	P20059 P50828 Q5R543	Binding	✔	
Hephaestin	Q920H8	Catalytic activity Binding	✔	
High-molecular weight cobalt-containing nitrile hydratase subunit alpha	P21219	Catalytic activity	✔	✔
Hypoxia up-regulated protein 1	Q63617	Cytoprotector triggered by oxygen deprivation	✔	✔
Ig alpha-1 chain C region	P20758	Antigen binding		✔
Ig gamma-1 chain C region	P01857	Antigen binding		✔
Ig gamma-2A chain C region	P20760	Antigen binding	✔	
Ig heavy chain V region 345	P18526	Antigen binding	✔	
Ig heavy chain V region 5A	P19181	Antigen binding	✔	
Ig heavy chain V region S43	P01755	Antigen binding	✔	
Ig heavy chain V region T601	P01808	Antigen binding	✔	
Ig heavy chain V-I region HG3	P01743	Antigen binding	✔	
Ig heavy chain V-I region V35	P23083	Antigen binding		✔
Ig heavy chain V-III region KOL	P01772	Antigen binding		✔
Ig kappa chain C region	P01834	Antigen binding		✔
Ig kappa chain V-I region Gal	P01599	Antigen binding	✔	
Ig kappa chain V-I region Ni	P01613	Antigen binding	✔	✔

Annex (continued)

Ig kappa chain V-I region Roy	P01608	Antigen binding	✔	
Ig kappa chain V-I region S107A	P01632	Antigen binding	✔	
Ig kappa chain V-II region 26-10	P01631	Antigen binding	✔	
Ig kappa chain V-II region 2S1.3	P01629	Antigen binding	✔	
Ig kappa chain V-II region Cum	P01614	Antigen binding		✔
Ig kappa chain V-II region GM607	P06309	Antigen binding	✔	
Ig kappa chain V-II region MOPC 167	P01626	Antigen binding	✔	
Ig kappa chain V-II region MOPC 511	P01628	Antigen binding	✔	
Ig kappa chain V-II region RPMI 6410	P06310	Antigen binding	✔	✔
Ig kappa chain V-III region CLL	P04207	Antigen binding	✔	
Ig kappa chain V-III region HIC	P18136	Antigen binding	✔	
Ig kappa chain V-III region IARC/BL41	P06311	Antigen binding	✔	
Ig kappa chain V-III region PC 4050	P01663	Antigen binding	✔	
Ig kappa chain V-III region PC 7175	P01671	Antigen binding	✔	
Ig kappa chain V-III region PC 7183	P01666	Antigen binding	✔	
Ig kappa chain V-III region PC 7940	P01672	Antigen binding	✔	
Ig kappa chain V-IV region B17	P06314	Antigen binding	✔	
Ig kappa chain V-IV region STH	P83593	Antigen binding	✔	
Ig kappa chain V-V region J606	P01652	Antigen binding	✔	
Ig kappa chain V-V region K2	P01635	Antigen binding	✔	✔
Ig kappa chain V-V region MOPC 149	P01636	Antigen binding	✔	✔
Ig kappa chain V-V region T1	P01637	Antigen binding	✔	
Ig kappa chain V-VI region NQ2-48.2.2	P04941	Antigen binding	✔	
Ig lambda-3 chain C regions	P0CG06	Antigen binding		✔
Ig lambda chain V-I region MEM	P06887	Antigen binding	✔	
Ig mu chain C region	P01872	Antigen binding Receptor activity	✔	
Immunoglobulin J chain	P01591	Antigen binding	✔	

Annex (continued)

Inhibitor of carbonic anhydrase	Q9DBD0	Inhibitor activity		✔
Insulin-like growth factor-binding protein 7	Q16270 Q61581	Binding	✔	
Interleukin-1 receptor accessory protein	Q61730	Receptor activity	✔	
Interleukin-18-binding protein	Q9Z0M9	Immune response	✔	
Isocitrate dehydrogenase [NADP] cytoplasmic	O88844	Catalytic activity Binding	✔	
Kappa-casein	P02668 Q28417	Binding	✔	✔
Keratin, type I cuticular Ha1	Q15323	Structural constituent of cytoskeleton	✔	
Keratin, type I cytoskeletal 9	P35527	Structural constituent of cytoskeleton	✔	✔
Keratin, type I cytoskeletal 10	P13645	Structural constituent of epidermis	✔	✔
Keratin, type I cytoskeletal 14	P02533 Q61781	Structural constituent of cytoskeleton	✔	✔
Keratin, type I cytoskeletal 16	P08779	Structural constituent of cytoskeleton	✔	
Keratin, type I cytoskeletal 17	Q9QWL7	Structural molecule Binding Receptor activity	✔	
Keratin, type I cytoskeletal 19	P19001	Structural constituent of muscle	✔	
Keratin, type II cuticular Hb1	Q9ERE2	Structural molecule	✔	
Keratin, type II cuticular Hb4	Q9NSB2	Structural constituent of cytoskeleton and epidermis	✔	
Keratin, type II cytoskeletal 1	P04264	Structural molecule Binding Receptor activity	✔	✔
Keratin, type II cytoskeletal 1b	Q6IG01 Q7Z794	Structural molecule	✔	
Keratin, type II cytoskeletal 2 epidermal	P35908	Structural constituent of cytoskeleton	✔	✔
Keratin, type II cytoskeletal 5	A5A6M8 P13647	Structural constituent of cytoskeleton	✔	✔
Keratin, type II cytoskeletal 6A	P02538 P50446	Structural constituent of cytoskeleton	✔	
Keratin, type II cytoskeletal 6B	P04259 Q9Z331	Structural molecule	✔	
Keratin, type II cytoskeletal 6C	P48668	Structural molecule	✔	
Keratin, type II cytoskeletal 75	Q6IG05	Structural constituent of hair and nails		✔
Keratin, type II cytoskeletal 78	Q8N1N4	Structural molecule	✔	

Annex (continued)

Keratin, type II cytoskeletal 79	Q5XKE5	Structural molecule	✔	
Kininogen-1	O08677	Inhibitor activity Binding Natriuresis and diuresis	✔	
L-lactate dehydrogenase B chain	Q9PW05	Catalytic activity	✔	
Lactadherin	P21956	Binding	✔	✔
Lactoperoxidase	P22079 P80025	Catalytic activity Binding	✔	✔
Lactotransferrin	Q9TUM0	Catalytic activity Binding	✔	
Latent-transforming growth factor beta-binding protein 1	Q8CG19	Binding	✔	
Leukocyte elastase inhibitor A	Q4G075	Inhibitor activity	✔	
Lipase member K	Q5VXJ0	Catalytic activity	✔	
Liver carboxylesterase	Q29550	Catalytic activity		✔
Liver carboxylesterase 1	Q8VCC2	Catalytic activity	✔	✔
Liver carboxylesterase 4	Q64573	Catalytic activity	✔	✔
Low-density lipoprotein receptor-related protein 2	A2ARV4 P98158 P98164	Binding	✔	
Lysosomal Pro-X carboxypeptidase	Q7TMR0	Catalytic activity	✔	
Lysosomal protective protein	P10619 P16675	Catalytic activity	✔	✔
Major outer membrane lipoprotein	Q6D622	Structural molecule	✔	✔
Major urinary protein 20	Q5FW60	Male pheromone. Binds most of the male pheromone, 2-sec-butyl-4,5-dihydrothiazole, in urine.	✔	
Malate dehydrogenase, cytoplasmic	P14152	Catalytic activity	✔	
Maltase-glucoamylase, intestinal	O43451	Catalytic activity	✔	
Mannan-binding lectin serine protease 2	Q9JJS8	Catalytic activity	✔	
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	P45701	Catalytic activity	✔	
Mannosyl-oligosaccharide glucosidase	Q13724	Catalytic activity		✔
Matrix-remodeling-associated protein 8	Q148M6 Q9DBV4	Structural molecule	✔	
Meprin A subunit alpha	P28825 Q16819 Q64230	Catalytic activity	✔	
Meprin A subunit beta	P28826	Catalytic activity	✔	

Annex (continued)

Metalloproteinase inhibitor 2	Q9TTY1 Q9WUC6	Catalytic activity Inhibitor activity	✔	
Methylmalonate semialdehyde dehydrogenase [acylating] 1	Q5L025	Catalytic activity	✔	
Microsomal glutathione S-transferase 3	O14880	Catalytic activity	✔	
Monocyte differentiation antigen CD14	P10810	Binding	✔	
Mucin-19	Q6PZE0	Ocular mucus homeostasis	✔	
Multiple inositol polyphosphate phosphatase 1	Q9Z2L6	Catalytic activity	✔	✔
Myosin light chain kinase, smooth muscle	Q6PDN3	Catalytic activity Binding	✔	
N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase	Q64191	Catalytic activity	✔	
N-sulphoglucosamine sulphohydrolase	P51688	Catalytic activity	✔	
Napsin-A	O09043	Catalytic activity	✔	
Nectin-2	P32507	Binding	✔	
Neuroplastin	P97546	Catalytic activity Binding	✔	
Neutral alpha-glucosidase AB	Q14697	Catalytic activity	✔	
Neutral and basic amino acid transport protein rBAT	Q64319	Catalytic activity Binding	✔	
Nucleobindin-1	Q02818 Q0P569 Q63083	Binding	✔	✔
Nucleobindin-2	P81117 Q9JI85	Binding	✔	✔
Nucleoside diphosphate kinase B	P22392	Catalytic activity Binding	✔	
Nucleotide exchange factor SIL1	Q32KV6 Q9H173	Folding		✔
Ornithine decarboxylase	P49725	Catalytic activity	✔	
Pancreatic alpha-amylase	P00688 P00689 P00690 P83053	Catalytic activity Binding	✔	✔
Pantetheinase	Q9BDJ5	Catalytic activity	✔	
Parvalbumin alpha	P20472 P80080	Binding	✔	
Peptidase inhibitor 16	Q9ET66	Inhibitor activity	✔	
Peptidyl-glycine alpha-amidating monooxygenase	P10731 P14925	Catalytic activity Binding	✔	
Peptidyl-prolyl cis-trans isomerase B	P24369	Catalytic activity Binding	✔	✔
Peroxiredoxin-1	Q6B4U9	Catalytic activity	✔	

Annex (continued)

Peroxiredoxin-2	Q5RC63	Catalytic activity	✔	
Phosphatidylethanolamine-binding protein 1	P70296	Inhibitor activity Binding	✔	
Phosphatidylethanolamine-binding protein 4	Q9D9G2	Binding	✔	
Phosphoserine aminotransferase	Q99K85 Q9Y617	Catalytic activity	✔	
Plasminogen	P20918 Q5R8X6	Catalytic activity Binding	✔	✔
Platelet-activating factor acetylhydrolase	Q28262 Q60963	Catalytic activity	✔	
Polymeric immunoglobulin receptor	O70570	Receptor activity	✔	
Polyubiquitin-B	Q8MKD1	Ubiquitous protein	✔	✔
Pro-cathepsin H	P00786 Q3T0I2	Catalytic activity		✔
Pro-epidermal growth factor	P07522 P01132 P01133 Q95ND4	Binding Growth factor activity	✔	✔
Probasin	O08976	Lipocalin Odorant binding Transporter activity	✔	
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	Q5R9N3	Catalytic activity Binding		✔
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	Q811A3 Q9R0B9	Catalytic activity Binding	✔	✔
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	Q9R0E1	Catalytic activity Binding		✔
Prostaglandin-H2 D-isomerase	O09114 P22057	Catalytic activity Binding Transporter actiivty	✔	
Prostasin	Q9ES87	Catalytic activity	✔	
Proteasome subunit alpha type-6	Q9QUM9	Catalytic activity Binding	✔	
Proteasome subunit alpha type-7	Q9Z2U0	Catalytic activity	✔	
Proteasome subunit alpha type-7-A	Q9PVY6	Catalytic activity	✔	
Proteasome subunit beta type-2	Q5E9K0	Catalytic activity	✔	
Proteasome subunit beta type-5	Q5R8S2	Catalytic activity	✔	
Proteasome subunit beta type-6	P28072 Q60692	Catalytic activity	✔	
Protein abnormal spindle	Q9VC45	Binding	✔	
Protein-glutamine gamma-glutamyltransferase 4	Q8BZH1 Q99041	Catalytic activity Copulatory plug formation	✔	

Annex (continued)

Protein-L-isoaspartate O-methyltransferase	Q3IDD2	Catalytic activity	✔	
Protein AMBP	P02760 Q64240	Catalytic activity Binding	✔	✔
Protein CREG1	O75629	Transcriptional control Catalytic activity Binding		✔
Protein disulfide-isomerase A3	P10731 P27773 P86235	Catalytic activity	✔	✔
Protein FAM151A	Q642A7 Q8QZW3	Structural molecule	✔	
Protein FAM3B	Q9D309	Cytokine activity	✔	✔
Protein OS-9	Q8K2C7	Binding		✔
Protocadherin-12	O55134	Binding	✔	
Putative phospholipase B-like 2	Q3TCN2	Catalytic activity	✔	
Pyrethroid hydrolase Ces2e	Q8BK48	Catalytic activity	✔	
Radixin	Q32LP2	Binding	✔	
Retinal dehydrogenase 1	P24549 P86886	Catalytic activity	✔	
Retinoid-inducible serine carboxypeptidase	Q920A5	Catalytic activity Kidney homeostasis	✔	
Retinol-binding protein 4	P04916	Binding Transporter actiivty	✔	
Ribonuclease pancreatic	Q9WUV3	Catalytic activity	✔	
Ribonuclease UK114	Q3T114	Catalytic activity	✔	
Semaphorin-7A	Q9QUR8	Binding		✔
Serine protease inhibitor A3C	P29621	Inhibitor activity	✔	
Serine protease inhibitor A3F	Q80X76	Inhibitor activity	✔	
Serine protease inhibitor A3M	Q03734	Inhibitor activity	✔	
Serotransferrin	P09571 P12346 Q921I1	Binding	✔	✔
Serum albumin	A6YF56 O35090 P02768 P02770 P07724 P14639 P49822 Q5XLE4	Binding	✔	✔
Serum amyloid P-component	P02743	Binding	✔	
SH3 domain-binding glutamic acid-rich-like protein 3	Q91VW3	Carrier activity Activator activity Catalytic activity Binding	✔	
Sialate O-acetylesterase	P70665 P82450	Catalytic activity	✔	
Sialidase-1	O35657	Catalytic activity	✔	

Annex (continued)

Signal peptide peptidase-like 4	Q0DWA9	Catalytic activity	✔	
Sodium-dependent neutral amino acid transporter B(0)AT1	Q2A865	Transporter activity	✔	
Sodium/glucose cotransporter 2	P53792	Transporter activity	✔	
Solute carrier family 12 member 1	P55016	Transporter activity	✔	
Solute carrier family 12 member 3	P55018	Transporter activity	✔	
Sphingomyelin phosphodiesterase	Q04519	Catalytic activity	✔	✔
Sulfated glycoprotein 1	P10960 Q61207	Transporter activity	✔	
Sulfhydryl oxidase 1	O00391 Q8BND5	Catalytic activity	✔	✔
Superoxide dismutase [Cu-Zn]	P08228 Q96VL0	Catalytic activity Binding	✔	✔
T-complex protein 1 subunit epsilon	P47209	Folding Binding	✔	
Tenascin	Q80YX1	Binding	✔	
Tetranectin	Q2KIS7	Binding	✔	
Thioredoxin	P11232 Q5R9M3	Catalytic activity	✔	
Thioredoxin-1	P0AA30	Catalytic activity	✔	
Thy-1 membrane glycoprotein	P01831	Binding Activator activity	✔	
Tissue alpha-L-fucosidase	Q99LJ1	Catalytic activity	✔	
Transcobalamin-2	O88968	Binding	✔	
Transmembrane protease serine 13	Q9BYE2	Catalytic activity Receptor activity	✔	
Transthyretin	P02766 P02767 P07309	Binding	✔	✔
Trypsin	P00761	Catalytic activity	✔	✔
Trypsin-1	P07477	Catalytic activity	✔	
Trypsin-2	P07478	Catalytic activity	✔	
Tubulin alpha-1C chain	Q9BQE3	Structural molecule Binding		✔
Tubulin beta chain	Q91575	Structural constituent of cytoskeleton		✔
Tyrosine-protein kinase receptor UFO	Q00993	Catalytic activity Binding	✔	
Ubiquitin-40S ribosomal protein S27a	P68203	Ubiquitous protein	✔	
Ubiquitin-60S ribosomal protein L40	P68205	Ubiquitous protein	✔	
UPF0764 protein C16orf89 homolog	Q3UST5	Homodimerization activity	✔	

Annex (continued)

Urokinase-type plasminogen activator	P04185 P06869	Catalytic activity	✔	
Uromodulin	P27590 P48733 Q862Z3	Binding Contributes to colloid osmotic pressure Prevents urinary tract infection	✔	✔
Uroplakin-2	P38575	Structural molecule	✔	
Vascular cell adhesion protein 1	P29534	Binding	✔	
Vesicular integral-membrane protein V	Q9DBH5	Binding	✔	
Vesicular integral-membrane protein VIP36	Q9DBH5	Binding	✔	
Vitamin D-binding protein	P02774 P21614 P53789	Binding	✔	
Zinc-alpha-2-glycoprotein	P25311 Q63678	Antigen binding		✔