Evolutionary genetics and incipient speciation in the European rabbit (*Oryctolagus cuniculus*)

Insights from the analysis of nucleotide polymorphism at mtDNA and sex chromosomes

Armando Geraldes

Dissertação apresentada à Faculdade de Ciências da Universidade do Porto para a obtenção do grau de Doutor em Biologia

Universidade do Porto

2006

Evolutionary genetics and incipient speciation in the European rabbit (*Oryctolagus cuniculus*)

Insights from the analysis of nucleotide polymorphism at mtDNA and sex chromosomes

Armando Geraldes

Universidade do Porto

2006

À minha irmã

Acknowledgements

I would like to thank all of those that in some way helped me in this five year long journey. You were many and in no way I will be able to refer to you all, but be sure that every now and then I will think of you and a smile will lighten my face.

First of all I would like to acknowledge my advisor, Professor Nuno Ferrand. His first, and perhaps most decisive, contribution to this thesis occurred during my undergraduate studies at Faculdade de Ciências da Universidade do Porto where he taught "Zoofilogenia e Evolução". His passion for Evolution changed forever the way I look at science and made me want to be a scientist. We have worked together ever since. All through these years, he played a crucial role in my work, especially by pointing new directions for my research and by helping me to find the right collaborators and institutions to pursue these new directions. One special night will forever be remembered. It happened three years ago, in Ciudad Real, Spain. That night we realized that the most parsimonious explanation for a series of "aberrant" results from my Y chromosome work was that the SRY gene was duplicated in the rabbit Y chromosome and that both copies were evolving in concert. It was the most rewarding scientific discussion I have ever experienced.

I would also like to acknowledge Dr. Michael Nachman for welcoming me in his lab at the Department of Ecology and Evolutionary Biology at the University of Arizona. In the last two years his contribution to this work was unique. He helped me to have a global perspective of my work and realize that what I was studying was speciation and most importantly, he made me understand that my research was not only a positive contribution to the study of the European rabbit, but also to the study of Evolution. I feel that my two year stay at his lab has allowed me to become a more independent and critical researcher. Finally I would like to thank him for giving me the opportunity to do a post-doc in his lab.

I would like to thank Dr. Rafael Villafuerte (IREC, Spain) and Dr. Claire Rogell-Gaillard (INRA-LREG, France) for the chance to work in their labs, and Dr. Rod Wing and Teri Rambo for the sequencing and assembly of the BAC clone containing the SRY gene. I would also like to thank Celine Chantry-Darmon (INRA-LREG) for her help with trying to subclone microsatellites from the rabbit Y

chromosome, and members of IREC for all the rabbit samples from Spanish populations that they kindly allowed me to use in this work.

To all my colleagues and friends at CIBIO, thank you for being such a nice group of people, and for helping me through so many scientific difficulties. For all the information on the European rabbit I would like to thank Rita Campos and Pedro Esteves. Madalena Branco for providing the raw data of the allozyme work. José Melo-Ferreira and Paulo Célio for sharing *Lepus* samples and data. Sara Rocha for help with estimation of models of nucleotide substitution. For hours of intensive data and theory discussion I would like to thank Miguel Carneiro, Rui Faria, Xana Sá Pinto and Fernando Sequeira. I would like to thank Sara Ferreira for her outstanding help with so many bureaucracies and for her ever present tenderness and support. A special thank you to Catarina Lopes Pinho for being a great friend, a very insightful thinker and for all kinds of bureaucratic problem solving that she has helped me with.

I would like to thank the members of the Nachman lab and other labs at EEB for making me a part of this very exciting group of researchers. A special thanks to Matt Dean, Jeff Good, Mike Hammer, Tovah Salcedo, Matt Saunders, Gabriela Wlasiuk and Elizabeth Wood for their interest in my research and for hours of discussion and fun.

For help with the writing of the Résumé of this thesis I would like to thank Guillaume Queney and Patrick Basset.

I would like to thank my parents and sister for their love, emotional and financial support. My father for supporting my choices, my mother for always pushing me one step further and for all the late dinners and snacks and my sister for always encouraging and supporting me unconditionally.

A very special thank you to Tom Dodd for teaching me a new way of living and for helping me to deal with an alien culture. Thank you for all your support and strength in the bad times and specially for making the good times happen.

Finally this work would not have been possible without the PhD fellowship SFRH/BD/4621/2001 and research project POCTI/BSE/40280/2001 from Fundação para a Ciência e a Tecnologia.

Declaração

Na elaboração desta dissertação, e nos termos do nº 2 do Artigo 8º do Decreto Lei nº 388/70, os resultados de trabalhos já publicados foram totalmente aproveitados e fazem parte integrante de alguns Capítulos desta dissertação.

Em todos estes trabalhos, o candidato participou na obtenção, interpretação e discussão dos resultados e na elaboração das suas formas publicadas.

Resumo

O coelho Europeu (*Oryctolagus cuniculus*) é uma espécie de grande interesse agronómico e biomédico e desempenha um papel crucial em ecossistemas mediterrânicos uma vez que é a principal presa de espécies emblemáticas tais como a Águia Imperial Ibérica (*Aquila adalberti*) e o Lince Ibérico (*Lynx pardinus*). Embora o coelho seja um modelo muito interessante para o estudo de processos evolutivos, o seu genoma e a sua história evolutiva são ainda pouco conhecidos. Estudos de genética populacional mostraram que esta espécie alberga duas linhagens mitocondriais muito divergentes que se sobrepõem ao longo de uma estreita zona de contacto e que correspondem às subespécies *O. c. cuniculus* e *O. c. algirus*. Recentemente, o interesse nesta espécie aumentou de forma considerável e, como resultado, foram desenvolvidas ferramentas importantes para o estudo do seu genoma. Neste momento estão já localizados cerca de 300 genes num mapa citogenético, a elaboração do mapa genético está em curso e uma primeira versão da sequência do seu genoma foi disponibilizada publicamente.

O objectivo principal deste trabalho foi aprofundar o conhecimento da história evolutiva desta espécie através do estudo de variabilidade nucleotídica de *loci* mitocondriais (Cytb), ligados ao cromossoma Y (SRY) e ao cromossoma X (os genes PHKA2 e HPRT1, localizados junto aos telómeros, e os genes SMCX e MSN localizados junto ao centrómero). Os níveis de polimorfismo detectados foram muito variáveis, sendo os mais baixos encontrados no gene SRY e os mais elevados encontrados no gene Cytb. Em qualquer dos casos, os níveis de polimorfismo encontrado foram consistentemente mais elevados do que os descritos para outras espécies de mamíferos. Este resultado mantém-se mesmo quando os níveis de polimorfismo são calculados por subespécie e não na espécie como um todo. Estes resultados confirmam a existência de uma marcada dicotomia no genoma do coelho, uma vez que se encontraram duas linhagens divergentes em todos os *loci* estudados. As estimativas da taxa de recombinação nos dois *loci* teloméricos do cromossoma X (PHKA2 e HPRT1) são elevadas e nestes *loci* observaram-se haplótipos recombinantes entre as duas linhagens. O uso de simulações do processo evolutivo condicionadas pelas estimativas de polimorfismo e recombinação para os quatro *loci* do cromossoma X permitiram rejeitar a hipótese nula de que estas linhagens poderiam ter evoluído no seio de uma única população panmítica. Estimativas de máxima

verosimilhança do tempo a partir do qual estas linhagens começaram a divergir são variáveis, mas sugerem um evento com cerca de 2 milhões de anos. A excepção foi o *locus* SRY, para o qual o tempo estimado de divergência entre linhagens foi de apenas 1 milhão de anos. Estes resultados são consistentes com a evolução das duas subespécies de coelho em diferentes refúgios glaciares. Em todos os *loci* estudados, as duas linhagens coexistem em algumas populações, mas observou-se uma clara diferença entre *loci* (SRY, Cytb, SMCX e MSN) para os quais esse facto se restringe a populações da zona de contacto e, *loci* (PHKA2 e HPRT1) para os quais as duas linhagens coexistem em quase todas as populações estudadas. Isto é, as duas subespécies de coelho estão altamente diferenciadas para o cromossoma Y, o DNA-mitocondrial e os dois *loci* centroméricos do cromossoma X, enquanto que para os dois *loci* teloméricos do cromossoma X as duas subespécies são muito similares. Compararam-se as estimativas de diferenciação populacional (Fst) para estes seis *loci* com os níveis de diferenciação para *loci* proteicos e microssatélites anteriormente estudados, e verificou-se a ausência de valores de Fst entre 0,4 e 0,6 num espectro que de outro modo é quase contínuo. Apenas os *loci* SRY, Cytb, SMCX e MSN apresentam valores superiores a 0,6. Devido ao facto de estes *loci* estarem em regiões genómicas não recombinantes (DNA-mitocondrial e cromossoma Y), ou então em regiões em que a taxa de recombinação estimada é muito baixa (região centromérica do cromossoma X), propõe-se que no genoma do coelho os níveis de diferenciação genética e de recombinação podem estar inversamente correlacionados. Estas observações sugerem que regiões genómicas onde os níveis de desequilíbrio de ligação são muito elevados constituem bons alvos para a detecção de incompatibilidades entre grupos de populações que estão num processo de especiação.

Os dados de polimorfismo nucleotídico obtidos para o gene SRY, que é transmitido exclusivamente por via paterna, e para o gene Cytb, que é transmitido exclusivamente por via materna, foram utilizados para comparar as contribuições do sexo masculino e feminino para processos demográficos. Em primeiro lugar, compararam-se os níveis de diferenciação populacional em cada subespécie para os *loci* SRY e Cytb. Foi possível mostrar que as populações de ambas as subespécies estão menos diferenciadas para o SRY, que define linhagens paternas, do que para o Cytb, que define linhagens maternas. Esta observação está de acordo com a conhecida filopatria das fêmeas e a maior tendência dos machos para dispersarem antes da primeira época reprodutiva. Em segundo lugar, analisaram-se estes dados à luz da

teoria coalescente e mostra-se que, em cada subespécie, o tamanho efectivo da população de fêmeas estimado é muito superior ao tamanho efectivo da população de machos estimado, o que está de acordo com a forte hierarquia entre machos nos grupos sociais de coelho bravo. Em qualquer dos casos, para cada um dos sexos, o efectivo polulacional estimado é muito elevado. Com base nos dados de polimorfismo do cromossoma X, calculou-se um tamanho populacional efectivo total para a subespécie *algirus* de aproximadamente 880 mil indivíduos, e um tamanho efectivo populational total para a subespécie *cuniculus* de aproximadamente 420 mil indivíduos. Os padrões de polimorfismo nucleotídico para os *loci* SRY e Cytb em populações de *algirus* são indicativos de uma forte expansão populacional, enquanto que os padrões observados nas populações de *cuniculus* são mais sugestivos de populações que se mantiveram relativamente estáveis por longos períodos de tempo. Por fim, enquanto que o tempo até ao ancestral comum mais recente (TMRCA) é cerca do dobro para o gene Cytb (1,9 milhões de anos) em relação ao gene SRY (0,9 milhões de anos) quando se analisa a espécie como um todo, o mesmo já não se verifica quando o TMRCA é calculado separadamente para cada uma das subespécies, notando-se que os valores obtidos se aproxima dos 300 mil anos e são semelhantes para ambas as subespécies e marcadores genéticos. No entanto, os desvios padrão destas estimativas são bastante elevados.

Quando este trabalho teve início apenas estava descrito um fragmento muito curto do gene SRY. De forma a poder comparar directamente a contribuição de ambos os sexos para os padrões demográficos da espécie foi necessário desenvolver um marcador no cromossoma Y que apresentasse os níveis de polimorfismo necessários para esta análise. Para tal fez-se a sequenciação directa de um clone BAC previamente localizado no cromossoma Y do coelho. Ao longo deste trabalho encontrou-se evidência de que o gene SRY está duplicado no cromossoma Y do coelho e que as duas cópias estão a evoluir concertadamente através de abundante conversão génica. A sequenciação completa do referido clone mostrou que, de facto, este gene está duplicado e que se encontra nos braços de um palindroma. Esta organização e modo de evolução molecular são semelhantes à dos genes amplicónicos do cromossoma Y da espécie humana. No entanto, na espécie humana, o SRY está presente em cópia única. Esta observação está de acordo com outros estudos que indicam que o cromossoma Y de mamíferos é um cromossoma muito dinâmico e que o conhecimento aprofundado desta região do genoma só será possível com inclusão

do cromossoma Y nos projectos de sequenciação do genoma de outras espécies de mamíferos.

Résumé

Le lapin de garenne (*Oryctolagus cuniculus*) est un animal important pour la agriculture et la recherche. Il joue un rôle crucial dans l'écosystème méditerranean oú il est la proie principale por des espéces emblématiques telles que l'Aigle Impérial Espagnole (*Aquila adalberti*) et le Lynx Ibérique (*Lynx pardinus*). Bien que ce soit un modèle trés intéressant pour l'étude des processus évolutifs, son génome et son histoire évolutive sont encore très peu connus. Les études génétiques précédentes ont montré que cette espèce se caractérise par la présence de deux lignées mitochondriales très divergentes (qui correspondent aux sous-espèces *O. c. cuniculus* et *O .c. algirus*) qui se chevauchent dans une zone d'hybridation étroite. Récemment, l'intérêt pour le lapin s'est énormément accru et, en conséquence, des outils importants pour l'étude de son génome ont été développés. La première carte cytogénétique a été publiée et la première version de la séquence du génome est maintenant disponible.

Dans ce travail, notre objectif était d'augmenter l'état des connaissances sur l'histoire évolutive de cette espèce à travers l'étude de la variabilité nucléotidique au niveau des loci mitochondriaux (Cytb), liés au chromosome Y (SRY) et au chromosome X (PHKA2 et HPRT1, proches des télomères, et SMCX et MSN, proches du centromère). Les niveaux de polymorphisme observé étaient très variables, les plus faibles étant dans le gène SRY et les plus élevés pour Cytb. De toutes façons, toutes les valeurs étaient fortement plus élevées que celles déjà décrites pour d'autres espèces de mammifères. Ce résultat est aussi confirmé quand les résultats sont présentés par sous-espèce et non plus pour l'espèce dans sa globalité. Nos résultats confirment l'existence d'une très forte dichotomie dans le génome du lapin, qui se révèle évidente dans l'observation de deux lignées très divergentes à chaque locus étudié. Les estimations de recombinaison sur deux loci proches des télomères du chromosome X (PHKA2 et HPRT1) se sont montrées élevées et des haplotypes recombinants ont été observés entre les deux lignées. Les simulations basée sur les niveaux de polymorphisme et de recombinaison estimés pour les quatre loci du chromosome X a permis de rejeter l'hypothèse nulle que les deux lignées auraient pu évoluer au sein d'une seule population panmictique. Les estimations de maximum du vraisemblance du temps initial de divergence entre les deux lignées sont variables mais suggèrent un événement proche de 2 MA. La seule exception a été le

locus SRY, pour lequel l'estimation du temps de divergence entre les deux lignées est de seulement 1 MA. Ce résultat est cohérent avec l'évolution de deux sous-espèces de lapin localisées au niveau de deux refuges glaciaires différents. Pour les six loci étudiés, les deux lignées coexistent dans quelques populations, mais il y a une différence très forte entre ceux qui coexistent uniquement dans les populations de la région de contact (Cytb, SRY, SMCX et MSN) et ceux qui se distribuent pratiquement dans toutes les populations (PHKA2 et HPRT1). Autrement dit, les deus sous-espèces sont très différenciées pour le chromosome Y, l' ADN mitochondrial et les deux loci centromériques du chromosome X. Par contre, les deux sous-espèces sont très homogènes pour les deux loci telomériques du chromosome X. Les comparaisons des estimations de la différenciation populationnelle (Fst) entre ces six loci et ceux décrits dans d'autres travaux sur les allozymes et les microsatellites, montre l'absence de valeurs comprises entre 0,4 et 0,6 au sein d'un spectre qui se montre presque continu. Seuls les loci Cytb, SRY, SMCX et MSN ont montré des valeurs supérieurs à 0,6. En tenant compte du fait que ces loci se localisent dans des régions génomiques non recombinantes (Cytb et SRY) ou dans des régions avec de très faibles taux de recombinaison (SMCX et MSN), nous suggérons que les niveaux de recombinaison et les niveaux de différentiation entre les deux sous-espèces sont inversement corrélés dans le génome du lapin. Cette observation suggère que les régions du génome qui subissent un fort niveau de disequilibre de liaison sont des bonnes cibles pour localiser des incompatibilities entre des groupes de populations en cours de spéciation.

Nous avons utilisé les données de polymorphisme nucléotidique obtenues pour SRY, qui se transmet exclusivement à travers les mâles, et pour Cytb, qui se transmet exclusivement à travers les femelles, pour comparer les deux contributions dans les processus démographiques. Tout d'abord, nous avons comparé les niveaux de différentiation populationnelle dans chaque sous-espèce pour les loci SRY et Cytb. Nous avons montré que, dans les deux sous-espèces, les populations sont plus différenciées pour le gène Cytb, qui définie les lignées maternelles, que pour le gène SRY, qui définie les lignées paternelles. Cette observation est en accord avec la dispersion biaisée par les mâles, bien connue chez le lapin. Deuxièmement, nous avons analysé les mêmes donnés par coalescence pour montrer que, dans chaque sous-espèce, les estimations de la taille populationnelle effective des femelles est toujours plus large que les estimations obtenues pour les mâles, ce qui est en accord

avec la forte hiérarchie sociale décrite chez les males dans les populations de cette espèce. Dans l'ensemble, les estimations de la taille populationnelle effective étaient toujours élevées, et nous avons déterminé des valeurs proches de 880.000 individus pour *algirus* et de 420.000 pour *cuniculus* en utilisant les données des loci du chromosome X. Les patrons de polymorphisme nucléotidique observés pour Cytb et SRY dans les populations d' *algirus* indiquent une forte expansion populationnelle, et contrastent avec l'inférence de populations essentiellement stables pour *cuniculus*. Finalement, nous avons observé que le temps de l'ancêtre commun le plus récent (TMRCA) est deux fois plus élevé pour le Cytb (1,9 MA) que pour le SRY (0,9 MA) si on analyse globalement l'espèce, mais cette observation ne tient pas quand on analyse les deux sous-espèces séparément: en fait, le TMRCA des deux gènes s'avère très similaire (300.000 ans). Il faut tout de même noter que les écart-types de ces estimations sont très élevés.

Au début de ce travail, seulement un petit fragment du gène SRY était connu. Pour comparer la démographie des mâles et des femelles, nous avons séquencé directement un clone BAC qui avait été cartographié dans le chromosome Y du lapin afin de développer des marqueurs spécifiques des mâles qui montraient un polymorphisme suffisant pour faire ce type d'analyses. Pendant ce travail, nous avons découvert que le gène SRY était dupliqué dans le chromosome Y du lapin, et aussi qu'il se localisait dans les bras d'un palindrome. De plus, nous avons montré que les deux copies du gène SRY évoluaient de façon concertée à travers des mécanismes de conversion génique. Cette observation ressemble à l'organisation et au mode d'évolution moléculaire des gènes ampliconiques dans le chromosome Y humain. Pourtant, une seule copie du gène SRY est présente dans l'espèce humaine. Nos résultats suggèrent ainsi, avec d'autres rapports récents, que le chromosome Y est une entité très dynamique et renforcent la nécessité de l'inclure dans des projets de séquençage de génomes de façon à mieux comprendre son contenu en gènes et son évolution.

Summary

The European rabbit (*Oryctolagus cuniculus*) is an important agricultural and research animal and plays a crucial role in mediterranean ecosystems where it is the main prey for emblematic species such as the Spanish Imperial Eagle (*Aquila adalberti*) and the Iberian Lynx (*Lynx pardinus*). Although it is a very interesting model to study evolutionary processes, its genome and evolutionary history are still poorly characterized. Previous population genetic studies have show that this species is characterized by the presence of two deeply divergent mtDNA lineages that overlap across a narrow contact zone and correspond to two subspecies *O. c. cuniculus* and *O. c. algirus*. Recently interest in this species has seen a dramatic increase and as a consequence important tools for the study of its genome have been developed. At this point about 300 genes are mapped onto its cytogenetic map, a genetic map is being originated and the first draft of its genome sequence has been made publicly available.

The main goal of this work was to deepen the knowledge of this species evolutionary history by looking at nucleotide variability at mitochondrial (Cytb), Y-linked (SRY) and X-linked loci (PHKA2 and HPRT1, near the telomeres, and SMCX and MSN near the centromere). Levels of polymorphism were highly variable being the lowest at SRY and highest at Cytb. Nonetheless, they were consistently higher than levels of polymorphism observed at other mammalian species. This result holds up even when levels of polymorphism are considered per subspecies instead of the entire species. Our results confirm the existence of a marked dichotomy in the rabbit genome as we have found two deeply divergent lineages at each locus. Estimates of recombination at the two X-linked telomeric loci (PHKA2 and HPRT1) were high and recombinant haplotypes between the two lineages were found. Computer simulations of the evolutionary process conditioned on the estimated levels of polymorphism and recombination at the four X-linked loci rejected the null hypothesis that these lineages could have evolved in a single panmitc population. Maximum likelihood estimates of the time since these lineages started to diverge are variable, but point to an event around 2 MYA. The exception was the SRY locus where the estimated time of divergence between the two lineages was only 1 MYA. This is consistent with the evolution of the two rabbit subspecies in different glacial refugia. At all six loci these lineages coexist in some populations. While at four loci

(SRY, Cytb, SMCX and MSN) this was mostly restricted to populations from the contact zone, at the other two (PHKA2 and HPRT1) this was the case in almost every population. In other words, the two subspecies are highly differentiated for the Y chromosome, the mitochondria and the two centromeric X-linked loci, but at the two telomeric X-linked loci the two subspecies are almost identical. The comparison of estimates of population differentiation (Fst) at these six loci to previously published data on protein and microsatellite polymorphism revealed the absence of Fst values between 0.4 and 0.6 in an otherwise quasi continuum spectrum. Only SRY, Cytb, SMCX and MSN showed Fst values above 0.6. Given that these genes are either in non-recombining regions of the genome (SRY and Cytb) or in areas of low estimated recombination rate (SMCX and MSN) we suggest that levels of recombination and levels of differentiation between subspecies might be inversely correlated in the rabbit genome. This observation suggests that genomic regions of high linkage disequilibrium are good targets to find incompatibilities between groups of populations that are undergoing speciation.

We used our nucleotide polymorphism data at the paternally inherited SRY and the maternally inherited Cytb to compare male and female contributions to demographic processes. First, we compared levels of population differentiation within each subspecies for the SRY and Cytb genes. We found that in both subspecies populations are more differentiated for the maternally inherited Cytb gene than for the paternally inherited SRY gene. This is in agreement with the known female philopatry and male biased natal dispersal of rabbit populations. Second, we analyzed this data set under a coalescent framework to show that, within each subspecies, estimated female effective population sizes are always much larger than male effective population sizes as expected given the strong male hierarchies observed in rabbit populations. In any case, estimated effective population sizes were high and we estimated the effective population size of *algirus* populations to be ~880,000 individuals and the effective population size of *cuniculus* populations to be ~420,000 using our X-chromosome data. Patterns of nucleotide polymorphism at SRY and Cytb in *algirus* populations are indicative of strong population expansion, while the patterns observed at *cuniculus* populations conform better to long term stable population sizes. Finally, while the time to the most recent common ancestor (TMRCA) for the entire rabbit sample is twice as high for Cytb (1.9 MY) as for SRY (0.9 MY), the same is not true for the TMRCA of each subspecies: the TMRCA for

both markers and subspecies is very similar (~0.3 MY). It is worth noting though that the standard deviations on these estimates are quite large.

At the beginning of this work only a very short fragment of the SRY gene was described. In order to directly compare male and female demographic processes we directly sequenced a BAC clone that mapped to the Y chromosome in order to develop a male specific marker that harbored enough polymorphism for this analysis. During this work we found that the SRY gene is duplicated in the rabbit Y chromosome and that it is located in the arms of a palindrome. We further found that these gene copies are evolving in concert through gene conversion. This resembles the organization and mode of molecular evolution of ampliconic genes in the human Y chromosome, yet in humans the SRY gene is a single copy gene. Our observation suggests, in agreement with other recent studies, that the Y chromosome is a highly dynamic chromosome and reinforces the need to include this chromosome in genome sequencing projects in order to understand its gene content and evolution.

Contents

1. GENERAL INTRODUCTION

1.1 The European rabbit as a model to study evolutionary processes

The European rabbit (*Oryctolagus cuniculus*) has been widely used for laboratory research. More than 500,000 rabbits are used per year in biomedical research and for testing the safety of chemicals and consumer products (Banks 1989). A simple literature search in public databases like PubMed shows that it is used as a model to study numerous human diseases, namely arteriosclerosis, cardiac, viral and immunological diseases. Rabbit placentae allow an unusually high degree of contact between maternal and fetal bloodstreams, a condition they share with humans. Thus, they are useful models for the study of human pregnancy and fetal development (Banks 1989). The use of domestic rabbits as research animals for genetic studies dates from the dawn of Mendelism, with the various coat colors affording ready material for confirmation of the generalities of Mendelian inheritance (Castle 1924).

Rabbits are regarded as extremely successful colonizers because they are present in every continent, except for Antarctica, and in more than 800 islands (Flux 1994). Together with our own species and house mice it has adapted to a wider range of ecological settings than any other mammal. It has done so, most likely in the last 3,000 years, mediated by man, in what constitutes a remarkable population expansion (Thompson and King 1994). Last, but not least, it is the only known example of an animal domesticated in Western Europe (Callou 2003). All of the above make it an ideal subject for evolutionary studies, but advance in this area has been slow. Nonetheless, three major conclusions can be drawn from the few recent molecular population genetics studies in this species. First, high levels of polymorphism are detected in wild populations. Second, there are two population groups corresponding to the subspecies *Oryctolagus cuniculus algirus* in southwestern Iberia and *O. c. cuniculus* in the rest of the world, and some natural hybridization occurs between them in central Iberia. Third, French and domestic populations have low levels of nucleotide polymorphism and are derived from *O. c. cuniculus*. Domestication has occurred very recently most likely in Mediterranean France.

1.1.1 Taxonomy of the LEPORIDAE and of the European rabbit

The European rabbit (*Oryctolagus cuniculus* L., 1758), belongs to the order Lagomorpha. Although the higher level relationships of the Lagomorpha are controversial, most authors include them in Glires implying a sister taxon relationship with the order Rodentia (Murphy *et al.* 2001). Within the Lagomorpha there are two families, the LEPORIDAE and the OCHOTONIDAE. The later is represented by a single genus, *Ochotona*, and the former by eleven genera of rabbits and hares. Seven of these eleven genera are monotypic (*Brachylagus*, *Pentalagus*, *Caprolagus*, *Bunolagus*, *Poelagus*, *Romerolagus* and *Oryctolagus*), and four (*Lepus*, *Nesolagus*, *Sylvilagus* and *Pronolagus*) are polytypic. The evolutionary relationships between LEPORIDAE genera are hard to trace and obtaining a robust phylogeny has proved to be difficult. Recently, a molecular supermatrix (Matthee *et al.* 2004) based on nucleotide variability at two mitochondrial (1882 bp) and five nuclear (3601 bp) genes was constructed for 27 taxa representing all 11 genera of the family. Based on these analyses, the closest species to *Oryctolagus cuniculus* are: *Caprolagus hispidus* (Hispid hare an Asian species), *Bunolagus monticularis* (Riverine rabbit endemic to South Africa) and *Pentalagus furnessi* (Amami rabbit found exclusively on two archipelagos south of Japan) (Figure 1). All species in this group, but *Caprolagus hispidus* are burrowers, a not so common trait in this family. Species in this group are thought to have begun to diverge approximately 9.4 (±1.15) MYA in Asia. This happened shortly after two other important events: the separation of the *Lepus* lineage 11.8 (±1.24) MYA in North America and the separation 10.28 (±1.18) MYA of the lineage leading to the American *Sylvilagus* and *Brachylagus* species from the Asian group that included the ancestor of *Oryctolagus* (Figure 1). The fossil record of the European rabbit is scarce. The oldest know fossil is a tooth found in southern Spain dated at about 6 MYA (Lopez-Martinez 1989). This fossil marks the appearance of the genus *Oryctolagus*. Fossil records exist for two extinct species of *Oryctolagus*. *Oryctolagus laynensis* was widely distributed in the Iberian Peninsula 2-3 MYA and *Oryctolagus lacosti* extended to southern France and existed about 1.8-2 MYA (Lopez-Martinez 1989).

Oryctolagus cuniculus was first recognized as a distinct species by Linnaeus in 1758. The name *Lepus cuniculus* was intended to combine the Latin names for hare (lepus) and rabbit (cuniculus). The name *Oryctolagus* was first used by Lilljeborg

(1874) as a subgenus of *Lepus* as a means of emphasizing the differences between rabbits and hares. Despite remarkable morphological resemblance, the American cottontails were from start correctly recognized as distinct from the European rabbit

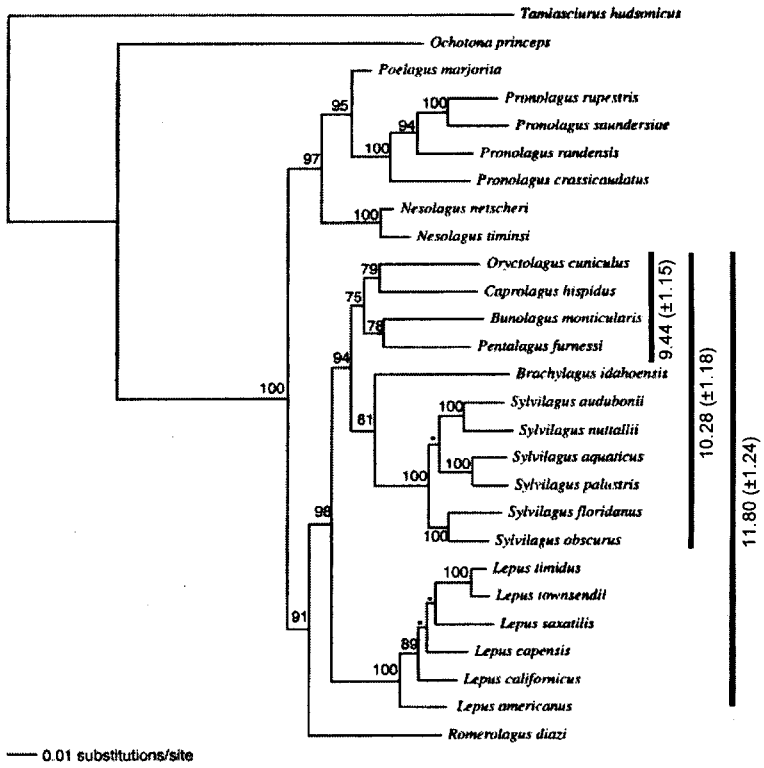


Figure 1- Supermatrix Maximum Likelihood phylogeny based on sequences from five autossomal and two mtDNA fragments. Bootstrap values are given for each node and branch lengths are proportional to the number of substitutions. Numbers shown after the vertical thick bars indicate the estimated time of divergence and the standard deviation in millions of years. Adapted from Matthee *et al.* (2004) and Robinson and Matthee (2005).

and named *Sylvilagus*. In 1858 Loche named the rabbits from North Africa as *Cuniculus algirus* and in 1874 Haeckel named the rabbits from Porto Santo (Madeira Archipelago) as *Lepus huxleyi*. Both are now included in *Oryctolagus cuniculus*.

Although there is some morphological variation in wild rabbits (size and coat color) there is no consensus regarding intraspecific taxonomy. Sharples *et al.* (1996) revised this problem. They took craniometric data from 260 skulls of *Oryctolagus cuniculus* from Western Europe and North Africa. Their samples were divided into ten groups roughly corresponding to ten morphotypes or subspecies previously described (Figure 2). The distribution of groups SSp and NSp roughly corresponds to the distribution of *O. c. algirus* and *O. c. cuniculus* in the Iberian Peninsula as proposed by Ferrand (1995), Branco *et al.* (2000) and Ferrand and Branco (2006).

Sharples *et. al.* (1996) found that there are no reasons to separate the rabbit groups that they analyzed into different subspecies based on skull measurements although they found significant differences between groups. Rabbits in the extreme north of the European distribution were always larger than animals in the extreme south. This variation was not discrete though, and the authors argue that the high degree of association between body size and geoclimatic variables simply obeys Bergmann's rule: body size increases with increasing latitude.

Figure 2- Geographic range of the ten groups used in the study of Sharples *et al.* (1996). SSp roughly corresponds to the distribution in Iberia of *O. c. algirus* proposed by Ferrand (1995), Branco *et al.* (2000) and Ferrand and Branco (2006). Similarly NSp roughly corresponds to the Iberian distribution of *O. c. cuniculus* proposed by the same authors. From Sharples *et al.* (1996).

Population genetic analysis gives us a different picture of the intraspecific taxonomy of *O. cuniculus*. For some genetic markers there are clear differences within Iberia with rabbits from southwestern Iberia being clearly differentiated from rabbits from northeastern Iberia. This will be discussed in further detail in the next chapters.

1.1.2 Ecology and Social structure of European rabbit populations

The European rabbit, *Oryctolagus cuniculus*, is a small herbivore that feeds mainly of grasses (*Graminae*) but can adjust its diet to the available vegetation (Martins *et al.* 2002). Rabbits have a brown/grey pelage and reach lengths of 34-35cm

weighing between 1.5-2kg (*O. c. cuniculus*) and 0.9-1.35kg (*O. c. algirus*) when adults (Villafuerte 2002). Rabbit abundance is restricted in grassland because of high predation and restricted in scrubland by food availability. The highest abundance of rabbits (40 individuals per hectare) is thus reached in ecotone areas where both habitats are readily available (Lombardi *et al.* 2003). Rabbits are usually found bellow 1500 m in soft and sandy soils that better allow them to build warrens. Territories have typical home ranges of one to two hectares and might encompass more than one warren. Rabbits from one social group can be found in different warrens within their group but are very rarely found on warrens of other groups (Daly 1981). In Australia, each group contains up to 20 adults with approximately equal numbers of males and females, but this number is highly variable in different habitats.

Rabbit's social groups are known to form hierarchies which determine an individual reproductive success (Myers and Mykytowicz 1958). For example, Daly (1981) found that in a social group, at least 16% of all kittens could not have been sired by the dominant male, and 7% were not the progeny of any of the bucks in the group. Given the low number of genetic markers employed in this study, it is likely that these numbers are even higher.

Rabbit dispersal distances in Iberia are usually low, with a record of two kilometers reported (Calvete and Estrada 2000). But, natal dispersal has mainly been studied in recently colonized areas: Australia (e.g. Daly 1981; Richardson *et al.* 2002), Germany (Künkele and VonHolst, 1996) and UK (e.g. Webb *et al.* 1995). In every case, significantly more males than females moved to a new social group before the start of their first breeding season. The percentage of males and females undergoing natal dispersal ranged from 18 to 95. It seems that although males disperse more than females, this behavior is highly variable and depends on the density of the population and several ecological variables (Richardson *et al.* 2002), *i.e.*, soil hardness. Webb *et al.* (1995), using allozymes and DNA fingerprinting, found that, in England, this differential dispersal resulted in populations where females within each group are more closely related than males.

1.1.3 Two divergent population groups within the Iberian Peninsula

Although molecular studies of wild rabbit population polymorphism date back to at least 1980 (Richardson *et al.* 1980), it was not until 1995 (Ferrand 1995) that

wild populations from Iberia were thoroughly screened. The first study to include samples from wild populations in the Iberian Peninsula (Arana *et al.* 1987) was a single protein locus study in six populations from central and northern Spain. Two polymorphic alleles were observed but no significant differences between populations were detected. The study of Biju-Duval and colleagues (1991) was the first to include a population from southwestern Spain. These authors studied nucleotide variability at the mtDNA molecule in Iberian wild populations and found two deeply divergent (4% nucleotide divergence) lineages (Biju-Duval *et al.* 1991). This study was surprising in two aspects. First, the divergence observed between the two mtDNA lineages placed their separation in the Pliocene/Pleistocene boundary (1.6-2 MYA). Second, one of the lineages observed (lineage A) was restricted to the population from southwestern Spain, while the other (lineage B) was present in all the samples from northern Spain, France, Tunisia and all domestic populations. Later, Ferrand (1995) and Ferrand and Branco (2006) used, respectively, 16 and 20 polymorphic protein loci in 12 and 10 wild populations from the Iberian Peninsula and France (plus two Atlantic insular populations and one domestic population) and found high levels of polymorphism. In accordance with the mtDNA results from Biju-Duval *et al.* (1991) these works found two distinct population groups, one present in southwestern Iberian Peninsula, associated with *O. c. algirus*, and the other in northeastern Iberian Peninsula, France and domestic populations, associated with *O. c. cuniculus*. A detailed Restriction Fragment Length Polymorphism (RFLP) survey of the mitochondrial Cytb gene, revealed high levels of nucleotide variability within each lineage (1.1% within *O. c. algirus* and 1.4% within *O. c. cuniculus*) and 11.9% nucleotide divergence between the two mtDNA lineages in Iberia (Branco *et al.* 2000). This value is three times higher than the one reported for the entire mitochondria, but the authors argue that this could be due to the use of different nucleotide sampling regimes and mutation rate heterogeneity within the mitochondria. This same study helped to define the contact zone between the two mitochondrial lineages as a narrow region bisecting the Iberian Peninsula in a NW-SE axis (Figure 3). Introgression of mtDNA can be observed in some populations outside the contact zone (defined by Bragança, Toledo, Ciudad Real and Las Amoladeras populations) but these were considered to be recent events most likely due to human intervention.

1.1.4 Postglacial expansion and secondary contact

Major paleogeographical events in Europe are likely to have had a great effect on today's species genetic structure. From the onset of the Quaternary (2.4 MYA) until 0.9 MYA ice sheets advanced and retreated in 41,000 year cycles after which more dramatic cycles of 100,000 years were observed (Hewitt 2000). During these major Pleistocene glaciations most European species were restricted to the southern parts of Iberia, Italy, the Balkans and the Caucasus. It is expected that, as a consequence of population retraction to southern refugia during glacial maxima, populations became repeatedly subdivided and isolated. With climatic amelioration during interglacial periods these populations expanded their range northwards to colonize newly available habitats (e.g. Santucci *et al.* 1998).

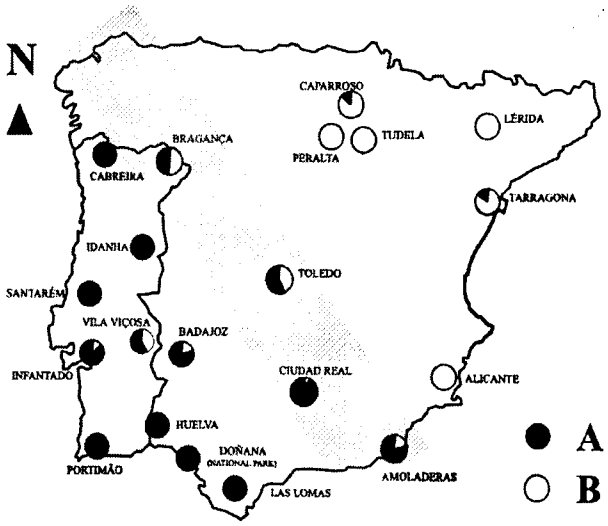


Figure 3- Distribution of mitochondrial lineages A and B in Iberia as inferred from PCR-RFLP survey of Cytb variation. The inferred contact zone is shaded. Adapted from Branco *et al.* (2000).

The Iberian Peninsula is thought to have been one of the most important southern refugia in Europe (Weiss and Ferrand 2006). Several lines of evidence make it likely that there were several refugia within this refugium (Gomez and Lunt 2006). First, the Iberian Peninsula possesses high physiographic complexity. Second, it is under the climatic influence of both the North Atlantic and the Mediterranean and presently enjoys a wide range of climates from desert to alpine. Third, it has a large area. Fourth, more than 30 contact zones between different taxa have been described in Iberia and can be most easily explained by survival throughout the Pleistocene ice ages in different Iberian refugia (Gomez and Lunt 2006).

The European rabbit is one good example of a species that likely had more than one single refugium in Iberia. As stated before, two well differentiated population groups exist in Iberia. At the mtDNA level two lineages are observed. Nucleotide divergence between the two mitochondria is 4.2% (Biju-Duval *et al.* 1991) and 11.9% if only the Cytb gene is considered (Branco *et al.* 2000). Using an average rate of mammalian mtDNA divergence, Biju-Duval *et al.* (1991) estimated a divergence time of 2 MY between them. This places their origin on a Pliocene/Pleistocene boundary, predating more recent glaciations. Branco *et al.* (2002) suggest that these lineages have remained in allopatry in two glacial refugia, one in the southernmost tip of Iberia (Andalusia, Spain and the south of Portugal) and the other in the eastern Mediterranean coast of Spain and the Ebro valley. The authors stress that during that period, rabbits have most likely been able to maintain stable population sizes as major losses of genetic variability are not observed, especially in southwestern populations. Expansion from these refugia is estimated to have occurred 200,000 YA (Branco *et al.* 2000). A glacial refugium in the Ebro valley has been suggested for other groups such as rotifers (Gomez *et al.* 2000), oaks (Lumaret *et al.* 2002) and pines (Salvador *et al.* 2000).

Nested clade analyses of PCR-RFLP Cytb data (Branco *et al.* 2002) revealed that the expansion of *Oryctolagus cuniculus algirus* from a southern Iberian refugium occurred mainly through restricted gene flow, which can be explained by wavelike population movements as rabbit populations slowly spread northward with a decrease in genetic variability. Two main dispersal routes from this southern refugium were detected (Figure 4i). One is placed along the Portuguese coast, and the other has a northern and eastern direction. Isolation by distance seems to have been the main determining process of present day genetic variability between populations, but some long dispersal movement along the Atlantic coast is apparent. Statistical power to detect the main demographic events that shaped the genetic variability of *Oryctolagus cuniculus cuniculus* is lower, but again, isolation by distance seems to have had a major impact on the expansion of rabbit populations from the Ebro valley/Mediterranean refugium. This expansion followed two main routes, a western route in the direction of central Iberia and a northern route towards the Pyrenees. As a result of the expansion from these refugia, a narrow contact zone in Iberia was formed. This contact zone is recent and has a northwest/southeast orientation bisecting the Iberian Peninsula. Mismatch distribution analysis of mitochondrial

control region sequence data along three transects in Iberia further consolidated the previous findings and supported the idea that the contact zone is recent, because, amongst other reasons, genetic divergence within each lineage is lowest in populations near the contact zone (Branco *et al.* 2002).

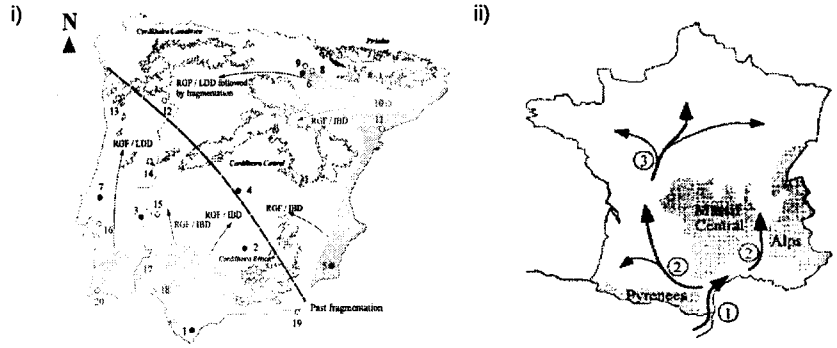


Figure 4- Main population demographic events that have shaped the postglacial expansion of rabbit populations in Iberia and France. i) Postglacial expansion of rabbit populations in Iberia, as depicted by Nested clade analyses and mismatch pairwise distributions of mtDNA variability. Shaded areas represent the inferred glacial refugia of rabbit populations and full lines the main mountain ranges. Arrows indicate main routes of dispersal and dashed line represents the past fragmentation event between the two subspecies. This line also coincides with the mitochondrial contact zone. RGF- restricted gene flow, LDD- long distance dispersal, IBD- isolation by distance. From Branco *et al.* (2002). ii) Colonization scenario of the rabbit into and in France as inferred from microsatellite variability. Numbers indicate the successive stages of colonization and shaded areas the main mountain ranges. From Queney *et al.* (2001).

The study of genetic variability at microsatellite loci in Iberia and France (Queney *et al.* 2001) showed that the levels of variability found in the later is only a subset of that found in Iberia implying two important conclusions: i) the colonization of France was mediated through a very small number of individuals, and ii) since no new mutations were detected at these hyper variable loci, this colonization event is likely to be recent. Within France three population groups were detected: southwest, southeast and north of France. The observed pattern is suggestive of a three step colonization of France (Figure 4ii). The first step of this process consists in the crossing of the Pyrenees by a small number of founding individuals. From here, rabbits followed two routes colonizing the southwest and the southeast regions divided by the Massif Central. In the last step the north of the country was colonized by rabbits belonging only to the southwest group. The analyses of 20 polymorphic protein loci in Iberian and French populations (Ferrand and Branco 2006), further supported the idea that French populations are recent, and that they are derived from a small number of founding individuals belonging to *O. c. cuniculus* only.The study of microsatellite markers also showed that domestication likely occurred recently in the South of France and that this was a single domestication event (Queney *et al.* 2002).

1.1.5 Conflicting portraits as inferred from mtDNA and nuclear markers

As discussed above, there are two divergent population groups of rabbits in the Iberia Peninsula that correspond to the subspecies *O. c. algirus* and *O. c. cuniculus*. At the mtDNA level, there is very restricted overlap between the distribution of the two lineages, and this occurs mainly across a narrow contact zone that bisects the Iberian Peninsula in a NW-SE direction (Figures 3 and 4i). Some introgression of mitochondrial lineage B haplotypes is observed in Southern Portugal as well as some introgression of lineage A haplotypes in Northern Spain (Figure 3) (Branco *et al.* 2000). The same is not true for the 20 allozyme loci studied by Ferrand and Branco (2006). These authors estimate an average Nei's genetic distance (D_N) of 0.11 (well within the usual range for subspecies) between *algirus* and *cuniculus* when only populations from SW and NE corners of Iberia are studied. At most loci, there are no fixed differences between subspecies, only a smooth cline in allele frequencies. The extent of introgression is much larger at autosomal loci than at the mtDNA. The contrast between autosomal and cytoplasmic loci is even greater when microsatellites are considered. Queney *et al.* (2001) found little or no differentiation between *cuniculus* and *algirus* populations from NE and SW Iberia. Although in this case absence of differentiation is likely due to homoplasy (Queney *et al.* 2001), the same argument does not hold up for the allozyme loci because neutral mutation rates at protein coding regions are usually much lower than at microsatellite loci.

There are at least three alternative hypotheses to explain the differences between autosomal and cytoplasmic loci: effective population size, male mediated gene flow and Dobzhansky-Muller incompatibilities involving the mitochondria. First: mitochondria are maternally inherited without recombination while autosomal loci are biparentally inherited and undergo recombination. This makes the effective population size of mitochondria one fourth that of the autosomes and therefore the mitochondria is more sensitive to the effects of genetic drift. Given this, we would expect to see greater differentiation between subspecies at cytoplasmic loci by chance alone. Second: as discussed above male rabbits undergo natal dispersal at much higher rates than female rabbits do (e.g. Kunkele *et al.* 1996). It is thus expected that gene flow between rabbit populations is mainly driven by long range male dispersal. If so, and since mitochondria are transmitted exclusively through the maternal lineage, we expect to observe the highest gene flow at the Y chromosome, followed

by the autosomes, the X chromosome and only then mitochondria. These two hypotheses are purely demographic. Alternatively selection against hybrid genotypes could be involved. Mitochondrial lineages A (*algirus*) and B (*cuniculus*) are greatly diverged. Mutations that have arisen in each lineage have never been tested on the genomic background of the other subspecies. It is possible that some of these mutations have epistatic interactions with other loci that cause hybrid inviability, sterility or reduced fitness, preventing the introgression of mitochondria between subspecies. These types of interactions are known as Dobzhansky-Muller incompatibilities and are further discussed below.

1.2 Recent advances in mammalian population genetics and phylogeography

Population genetics is the study of how evolution works as a genetic process in natural populations. It is highly mathematical in theory and approximate in application. The first and still most common models used in population genetics are very simplistic and probably every natural population violates at least one model assumption. In recent years, the field of population genetics has seen dramatic changes. These were boosted by a wealth of new theoretical models, statistical frameworks and technological innovations.

It has long been realized that populations are often subdivided in nature, and that these structured populations need new methodological tools to be analyzed. With the advent of DNA sequencing the field has seen a great shift and two schools of thought developed: one overlays evolutionary trees with geography, phylogeography (Avise 2000), and the other uses information that is shown in the patterns of DNA variation to estimate important population parameters, coalescent theory (Kingman 1982; Hudson 1990). While the first offers the advantage of not being model based and being very graphical and intuitive, it lacks the statistical support that complex model based approaches have. In recent years, population geneticists have begun to develop new approaches that in some ways unite the tree-based and summary-statistic-based methods (Hey and Machado 2003).

The advent of PCR and automated DNA sequencing has revolutionized the way population genetics is done. It has allowed direct access to nucleotide variability instead of relying solely on amino acid substitutions at the protein level and patterns of enzyme restriction. This has culminated in what became known as the genome era where drafts of complete genome sequences of several species have became available allowing for completely new avenues of research.

1.2.1 The Y chromosome: the male contribution to the evolutionary history of a species

Sex determination is a widely variable and labile phenomenon in nature. Probably the most common mechanism for sex determination is chromosomal sex determination. In many species, one sex has two identical sex chromosomes (homogametic sex) and the other has two distinct sex chromosomes (heterogametic sex). In most mammals, males are the heterogametic sex (with a Y and an X chromosome) and females are the homogametic sex (with two X chromosomes). Many insects have this same system of sex determination but the insect X and Y chromosomes have evolved from a different pair of autosomes and are not homologous to the mammalian X and Y. In butterflies (Lepidoptera) chromosomal sexual determination arose yet another time and males are the homogametic sex (ZZ) and females the heterogametic sex (ZW). Birds also have a ZW sex determination system. In this review we will focus mainly on the evolution of the mammalian Y chromosome and specifically the human Y chromosome because it is the only fully sequenced and annotated Y chromosome to date (Skaletsky *et al.* 2003).

The Y chromosome is a highly peculiar chromosome. First, it is present in males only. Second, it is the only mammalian chromosome that does not undergo heterologous recombination [in this section we will refer only to the Male Specific Region (MSY), and mostly ignore the small telomeric pseudoautosomal regions of the Y that undergo heterologous recombination with the X chromosome]. Third, it is probably the most functionally coherent chromosome. Finally, it has a peculiar mode of molecular evolution.

The mammalian sex chromosomes evolved from an ancestral homologous pair of autosomes 300 MYA (Lahn and Page 1999). Differentiation into heterologous chromosomes was possible due to suppression of recombination between this pair of chromosomes. Lahn and Page (1999) measured Ks (estimated mean number of synonymous substitutions per synonymous site) between 19 homologous genes on the X and Y chromosomes. Because synonymous sites are considered to evolve neutrally (Li 1993) and in a clockwise fashion, the authors were able to estimate the time since these gene pairs started to diverge, and hence determine when the sex chromosomes started to differentiate. They found that these gene pairs could be grouped into four distinct clusters according to the time since they started to diverge. The most striking

observation was that genes belonging to each of these clusters are arranged orderly in the X chromosome but not on the Y. The authors hypothesized that during evolution, differentiation (caused by cessation of X-Y recombination) was initiated one region, or stratum, at a time. This indicates that X-Y recombination was suppressed regionally beginning in stratum 1 and expanding subsequently to the other strata. Cessation of recombination was obtained through the inversion of each of the strata in the Y chromosome. Sandstedt and Tucker (2004) applied the same reasoning to the house mouse (*Mus musculus*) and found good correspondence to the results for the human sex chromosomes. Two major differences were observed. First, since the mouse X chromosome is rearranged in relation to other mammalian X chromosomes, the strata are not arranged in physical or genetic order in the mouse X. Second, only three strata were found. This is in agreement with the prediction from Lahn and Page (1999) that the fourth and most recent stratum (dated to 50 MY) is exclusive of simian mammals. In both studies the gene pair with the highest Ks was SRY/SOX3, indicating that the sex determining region was the first to differentiate on the sex chromosomes of mammals.

Because the Y chromosome does not participate in sexual recombination, it has degenerated substantially, both in size and gene content, in comparison with the X chromosome. The complete sequencing of the human MSY has revealed that it is composed of three classes of sequences: X-degenerate sequences, X-transposed sequences and ampliconic sequences (Skaletsky *et al.* 2003). The X-transposed sequences are 99% identical to DNA sequences in the human X chromosome, and are thought to have been transposed recently (less than 4 MYA) from the X. There are only two genes in this region which is the least gene dense region on the MSY (Table 1). The X-degenerate segments of the Y are composed of single copy genes or pseudogene homologues of 27 X-linked genes. They exhibit only 60 to 96% homology with their homologues and are thought to be the true relics of the ancient pair of autosomes from which sex chromosomes arose (as discussed above). The X-degenerate sequences still hold 16 functional genes, very similar, but distinct, to their X counterparts. All of the 12 ubiquitously expressed genes found in the MSY are X-degenerate genes. From all the 11 genes that are exclusively or predominantly expressed in the testis, only one is X-degenerate, the sex determining gene (SRY). The third class of sequences found in the Y is called ampliconic and harbors genes that are mainly expressed in testis. These ampliconic segments are characterized by

having intrachromosomal similarities greater than 99.9%. There are seven such blocks of sequences and they are organized in palindromes separated by a small region of unique sequence called the spacer. Palindrome arm length ranges from 9 kb to 1.45 Mb and spacer length ranges from 2 to 170 kb in length.

Table 1- Three sequence classes found in the human MSY

Sequence class	Defining characteristics	Evolutionary origins	Aggregate length (Mb)	No. of coding genes	No. of non-coding transcription units	No. of transcription units per Mb
X-transposed	99% identity to X	single transposition from X	3.4	2	0	0.6
X-degenerate	Single-copy genes or pseudogenes homologues of X-linked genes	Relics of ancient autosomes from which X and Y evolved	8.6	16 (most expressed widely)	4	2.2
Ampliconic	≥ 99.9% similarity to other regions on MSY	Acquired from diverse sources and then amplified	10.2	60 (in 9 families, expressed mainly or only in testis)	74 (9 single-copy; 65 in 15 families, expressed mainly or only in testis)	13.3

Adapted from Skaletsky *et al.* (2003).

The finding of these ampliconic regions led to the recognition that there are two independent mechanisms of Y chromosome evolution. One applies to the X-degenerate genes and the other to the ampliconic regions. The evolution of X-degenerate genes, where no recombination whatsoever occurs has been marked by a progressive decline in functional genes through the accumulation of deleterious mutations that have progressively turned the Y homologues into pseudogenes. This evolutionary model provides no explanation for some of MSY's defining characteristics. These include: testis-specific gene expression, near perfect palindromes with multi-copy genes and an abundance of autosomal as well as X-chromosomal sequence similarities. Ampliconic regions were acquired through transposition and retrotransposition of genes from other genomic regions.

Independent of the genomic source of these genes they have all come to be expressed (almost exclusively) in the spermatogenic cells of the testis, making the Y probably the most coherent of all human chromosomes. The emergence of a male-specific domain in the Y chromosome, created a genomic niche where selection could operate to enhance male germ-cell development. Subsequently, amplification of these newly acquired genes allowed higher levels of their expression and production of sperm. This amplification also created the opportunity for the occurrence of gene conversion between gene copies. The identification and sequencing of chimpanzee Y-linked orthologues of human MSY palindromes established that gene conversion between palindrome arms has occurred both in the human and chimp lineages, and has continued to occur in human populations (Rozen *et al.* 2003). The rate of gene conversion between palindrome arms of the human MSY is estimated to be several orders of magnitude higher than the mutation rate. In each generation at least one such event is estimated to occur. Pervasive gene conversion has allowed these ampliconic genes to escape inactivating mutations and ensures their persistence in the Y chromosome. This may be a common feature of Y and W chromosomes, because gene conversion has recently been shown to also drive the evolution of the avian sex chromosome (Backstrom *et al.* 2005).

Curiously, the Drosophila Y chromosome, although having evolved independently from the human Y chromosome, also shows unusual functional specialization. It is enriched for genes involved in male fertility (Carvalho *et al.* 2001). Interestingly most of these genes have autosomal, instead of X-linked, homologues. These two features suggest that these genes were not present in the hypothetical pair that gave rise to the sex chromosomes. Instead it points to the possibility that that they were acquired from the autosomes and were retained because they confer a specific fitness advantage to their carriers (Carvalho *et al.* 2000).

Another interesting consequence of the absence of heterologous recombination and the presence in exclusivity of the Y chromosome in males is that it allows us to reconstruct the male history of a species. These same properties (and the fact that it exists in multiple copies in cells) have made the mitochondria the most widely used marker in molecular population genetics and specifically in phylogeographical studies. They both define lineages (the mitochondria defines matrilineages and the Y patrilineages) and if one assumes a sex ratio of 1:1 they both have an effective population size of ¼ that of the autosomes. If we imagine a man and a woman in the

population they have four copies of each autosome, three of the X chromosome and only one of the Y chromosome (and although they have two mitochondria only the female one will contribute to the next generation). As a direct consequence of its effective population size, the Y chromosome harbors less polymorphism (Hammer 2003) and is more prone to the effects of genetic drift.

Contrary to the mitochondria, use of the Y chromosome in phylogeographical studies is rare and restricted to a handful of organisms. Several factors are responsible for this. The first is historical and is due to the fact that while the Y chromosome is present in a single copy in each cell. On the other hand, mitochondria exist in thousands of copies and as a consequence were the first genetic material that researchers were able to extract from cells in quantities necessary for genetic analyses. With the advent of Polymerase Chain Reaction (PCR) other markers (as autosomes and the X-chromosome) started to be used more commonly. The same did not happen as fast with Y-linked markers because they are not as conserved making cross-species amplification hard or even impossible and thus restricting its use to model organisms. Another important factor is that genes in the Y chromosome usually harbor very low levels of polymorphism [for an extreme example see Lindgren *et al.* (2004)]. This is likely to be the result of: unequal sex ratios, strong social hierarchies where only a subset of males are able to reproduce (Petit *et al.* 2002), pervasive positive directional selection that eliminates standing genetic variation through successive selective sweeps (Tucker and Lundrigan 1993; Whitfield *et al.* 1993; Wyckoff et al. 2002; Gerrard and Filatov 2005) and purifying selection that eliminates deleterious mutations (Gerrard and Filatov 2005).

Recently, Y chromosome polymorphism has started to be used in non model organisms allowing direct analyses of the male contribution to a species natural history (e.g. Feng *et al.* 2001; Lindgren *et al.* 2004; Hellborg *et al.* 2005).

1.2.2 The X chromosome: a peculiar mode of inheritance

The X chromosome is present in a single copy in males and in two copies in females, because of this, every existing X chromosome has spent 2/3 of its history in females, and has an effective population size of ¾ that of the autosomes (if we again assume a sex ratio of 1:1). Since mutation rate is lower in females than in males, the X chromosome is expected to have the lowest mutation rate of the nuclear genome

(and the Y chromosome the highest) (Li *et al.* 2002). It is thus expected that the X chromosome presents lower polymorphism than autosomal markers. Polymorphism is further reduced because of its smaller effective population size. Another consequence of its smaller effective population size is an increased sensitivity to the effects of genetic drift. With this reasoning, population structure should be the most pronounced in the Y chromosome and mtDNA and the least pronounced in the autosomes, with the X chromosome showing an intermediate picture. The measured recombination rate for the human X chromosome (Kong *et al.* 2002) is roughly 2/3 of the genomic average as would be expected since it can not recombine while it is present in males. We can therefore expect increased linkage disequilibrium in X-linked markers.

Apart from the above features, its presence in hemizygosity in males allows newly arisen recessive advantageous mutations to be immediately exposed to selection. While a recessive advantageous mutation is most likely lost from the population by drift soon after its appearance in an autosome, it might be driven to higher frequencies in the X chromosome because it is exposed in males. Theory predicts that recessive mutations that are advantageous to males should thus concentrate in the X chromosome. This has important implications both for the evolutionary fate of genes present in the X chromosome and for its gene repertoire. Emerson *et al.* (2004) performed a genomic analysis of duplicate genes produced by retroposition, whereby a mature messenger RNA is reverse transcribed and integrated into the genome. This study revealed extensive gene traffic on the X chromosome. The X chromosome harbors a disproportionately large number of genes that originate new retroposed copies in the autosomes. The human X chromosome originated 299% more retroposed genes than what would be expected if all chromosomes generated the same number of retroposed genes, and the mouse X 309% more. Similarly the human X is a genomic outlier in terms of acceptance of retrotransposed genes from the autosomes. The human X has accepted a 260% excess of genes and the mouse X a 246% excess. If this is not due to a mechanistic bias, then it can only be explained by natural selection. This chromosome-biased gene recruitment appears to be an important process actively driving the differentiation of the X chromosome in mammals and suggests that this differentiation is still in progress. Below, we discuss the role of the X chromosome in speciation.

1.3 The study of speciation

The field of speciation is concerned with the study of the natural mechanisms that lead to the origin of new species. Irrespectively of the speciation concept that one adopts, the most common goal is to understand why sympatric, sexually reproducing organisms fall into different clusters. One of the most widely accepted species concepts, and the one that probably best addresses the above question, is the Biological Species Concept (Coyne and Orr 2004). This concept was first introduced in 1942 by Ernst Mayr, and in its present version states that, species are groups of interbreeding natural populations that are reproductively isolated from other such groups (Mayr 1995). If one accepts some version of this concept, then one fundamental problem becomes understanding the genetic basis of reproductive isolation.

Two general approaches have been used to address this problem: controlled interspecific crosses in the laboratory and the analyses of hybrid zones. Both have advantages and problems, and provide different and complementary insights into the speciation process. Laboratory crosses offer, amongst others, the advantages of controlling both the genetic and environmental background and of being repeatable. Hybrid zones however allow us to look at the result of introgression after many generations of recombination, making the fine-scale mapping of genomic regions involved in the speciation process possible. This can be done simply by looking at patterns of gene flow and introgression, and without any prior knowledge of the phenotypes involved. Another advantage is that the fitness of hybrids is tested under natural conditions.

A century of research has allowed us to define two rules of speciation. First, barriers to gene flow often derive from epistatic interactions between two or more loci (Bateson 1909; Dobzhansky 1936; Muller 1940, 1942), and second, the X-chromosome seems to have a disproportionately large effect on hybrid sterility and inviability (Coyne and Orr 1989). We discuss each of these rules in the next two sections. Then we discuss the role of the Y chromosome in speciation by reviewing evidence that in some systems it plays a crucial role and is involved in epistatic interactions that cause intrinsic postzygotic isolation, while in others its role is not very clear.

1.3.1 Epistatic interactions and the accumulation of Dobzhansky-Muller incompatibilities

One of the paradigms of speciation is the unfitness of hybrids. How could natural selection favor the evolution of unfit offspring? Four general situations have been identified that could explain this phenomenon: different ploidy levels (thought to be widespread in plants, but not common in animals), different chromosomal rearrangements, infection by different endosymbionts, and different alleles that do not work well in hybrids (Coyne and Orr 2004).

Although chromosomal rearrangements and cytoplasmic incompatibilities (caused by infections with different symbionts) could cause reproductive isolation, epistatic interactions between alleles at different loci are believed to be the most common cause of intrinsic postzygotic isolation. A one locus model (within-locus incompatibilities), in which allele *A* from one species is incompatible with allele *a* from the other species is unlikely because we can not evolve unfit *Aa* hybrids without one of the species passing through that same unfit state. This one locus model is however possible if one of the species has successfully passed such an adaptive valley, or if multiple substitutions at a locus occur. Neither is very likely. Alternatively, intrinsic postzygotic isolation could result from between-locus incompatibilities. This is known as the Dobzhansky-Muller model (as proposed by Bateson 1909; Dobzhansky 1936; Muller 1940, 1942). In this model we assume that an ancestral population splits into two that then diverge in allopatry. Let's imagine that the ancestral population had a two locus genotype *aabb* (Figure 5i). With time substitutions occur in both populations and one of them becomes fixed for *aaBB* and the other for *AAbb*. When they meet *aAbB* hybrids are formed. This is the first time that alleles A and *B* "see" each other, this is, this genotype had never been exposed to natural selection before.

A wealth of theoretical and empirical work has shown that this simple model explains most of the hybrid sterility phenotypes observed (Coyne and Orr 2004), and helped realize that the Dobzhansky-Muller model predicts the following patterns. First, substitutions must occur at both loci (Orr 1995). In figure 5ii arrows indicating incompatibilities never run up to loci that have not yet mutated. Second, derived alleles cause incompatibilities more often than ancestral alleles, because newly derived alleles can cause incompatibilities with a derived or an ancestral allele, but an

ancestral allele can only be incompatible with a derived one (Figure 5ii). Orr (1995) has shown that with equal rates of substitution in both lineages, derived alleles are three times more likely to cause incompatibilities than ancestral ones. Third, incompatibilities are asymmetric (Muller 1942). In figure 5ii, although *B* might be incompatible with *A*, *a* can not be incompatible with *b*. Fourth, the number of hybrid

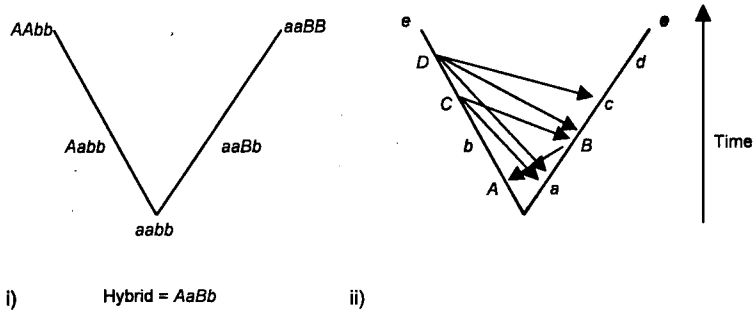


Figure 5- Epistatic interactions between-loci. i) The Dobzhansky-Muller model, showing how genic incompatibilities between two loci (A and B) can evolve without any of the populations crossing an adaptive valley. ii) The evolution of genic incompatibilities at multiple loci. Possible genic incompatibilities in hybrids are shown as arrows. Adapted from Orr (1995).

sterility and inviability genes "snowballs" through time. Not only later substitutions are more likely to cause incompatibilities than earlier ones, but the number of incompatibilities separating two taxa rises with the square of time (Orr 1995; Orr and Turelli 2001). Finally, complex incompatibilities should be common. In fact the more loci involved, the easier it is to evolve incompatibilities.

1.3.2 The large X-chromosome effect

The second "rule" of speciation is the large X-effect. Evidence for this effect comes from theoretical and empirical work.

Probably the observation from the speciation field that has received the most attention during the last century of research is the so called "Haldane's rule". In 1922, Haldane stated:

> *"When in the F_1 offspring of two animal races* (usually distinct species) *one sex is absent, rare, or sterile, that sex is the heterozygous* (heterogametic, i.e., XY or ZW) *sex."*

This phenomenon was found to be obeyed in almost every interspecific cross observed in animals from a wide range of groups (Coyne 1992). Which are then the genetic causes of Haldane's rule?

Many explanations of Haldane's rule have been offered, and although some of them might work from time to time, they can not explain the ubiquity of the rule. Amongst these are, epistatic interactions between X and Y-linked genes (but the rule is also obeyed in X0 taxa) and disruption of dosage compensation or sex determination. Four explanations remain prominent in the literature, and for two of them there is strong empirical and theoretical evidence: the dominance theory (Muller 1940, 1942; Orr 1987) and faster male evolution (Wu and Davis 1993; Wu *et al.* 1996). Both explanations imply a crucial role for the X chromosome. For a detailed discussion on the genetics of intrinsic postzygotic isolation see chapter 8 of Coyne and Orr (2004).

In laboratory crosses between closely related taxa, especially in D*rosophila*, genes having a large effect on postzygotic isolation often map to the X chromosome. This pattern has also been observed in mammals. The X-linked *Hst-3* gene causes male sterility in hybrids between *Mus domesticus* and *Mus spretus* (Guenet *et al.* 1990). Moreover, backcross analysis in *Drosophila* show that the substitution of an X chromosome from one species versus the other has a larger effect on postzygotic isolation than any autosome (Coyne and Orr 1989). Studies on hybrid zones in a wide array of taxa (*e.g.* butterflies, Hagen 1990; *Drosophila*, Llopart *et al.* 2005; *Mus*, Tucker *et al.* 1992) repeatedly show reduced introgression of the X chromosome relatively to the autosomes.

In almost every empirical study on the genetics of postzygotic isolation the greatest effect is attributed to the X chromosome. Not only more genes involved map to the X chromosome, but also the genes with the biggest effect are X-linked. To date only four genes have been identified that cause intrinsic postzygotic isolation. Although we do not know if these were the first genes involved in the speciation process between the taxa studied, we know that today they cause reproductive isolation. They have become known as speciation genes. Three of these genes have been identified in *Drosophila* (Barbash *et al.* 2003; Presgraves *et al.* 2003; Ting *et al.* 1998) and one in *Xiphophorous* (Malitschek *et al.* 1995). All of them are either sex-linked or are involved in epistatic interactions with sex-linked loci, all of them are rapidly evolving and all of them have evolved by positive Darwinian selection.

1.3.3. The conflicting role of the Y-chromosome

While it has been well established that the X chromosome plays a disproportionately large role in speciation, the role of the Y chromosome remains much more elusive. There is strong evidence though, that at least in some cases Y-linked loci play a determinant role in intrinsic postzygotic isolation.

There are good reasons to suspect that loci on the Y chromosome might be involved in hybrid male sterility. First, the Y chromosome is, in many species, enriched for genes that affect spermatogenesis (humans, Skaletsky *et al.* 2003; *Drosophila*, Carvalho *et al.* 2000, 2001). Second, in *Drosophila*, genetic interactions affecting spermatogenesis have been found between homologous loci on the X and Y chromosomes (Livak 1984). Third, genes on the Y chromosome evolve rapidly between species and some have been hypothesized to evolve through Darwinian positive selection (Tucker and Lundrigan 1993; Whitfield *et al.* 1993; Whyckoff *et al.* 2002; Gerrard and Filatov 2005).

Evidence for an important role of the Y in speciation comes mainly from empirical data, although, throughout the last century at least two theories to explain the genetic basis of Haldane's rule involve the Y chromosome. One of the first such theories (Haldane 1932), suggested that the sterility of heterogametic hybrids results from incompatibilities between loci on the X and Y chromosome. It has also been proposed that translocations of Y and X portions into the autosomes could explain Haldane's rule. The later has been discredited (Coyne 1985), but the former, X-Y interactions, has received considerable attention and although interactions between sex chromosomes have been shown to exist in some cases, they are certainly not the general cause of Haldane's rule.

In Drosophila, many crosses have been performed to investigate the role of the Y and/or X-Y interactions in hybrid male sterility and their causal role in Haldane's rule. In reciprocal crosses between *D. mauritiana* and *D. simulans* it has been shown that not only the Y has a significant effect, but also that interactions between the X and the Y are important. Coyne (1985) observed that certain portions of the X caused a greater reduction in sperm motility when in presence of a heterospecific Y than others. These results have been confirmed, and the incompatibility of the *D. mauritiana* Y chromosome has been extended to crosses with *D. sechellia*, where it

causes complete male sterility (Johnson *et al.* 1993). Similarly, *D. simulans* Y cannot be introgressed into *D. sechellia* (Zeng and Singh 1993). The Y chromosome of *D. sechellia* has no effect in sterility when introgressed into *D. simulans* (Zeng and Singh 1993). There is large asymmetry in the fitness of Y introgressions in crosses involving *D. sechellia* but the effects are symmetrical in the crosses between the other two species.

Theodosius Dobzhansky pioneered the study of hybrid incompatibilities in *D. pseudoobscura*. He did not find any effect of the Y chromosome (Dobzhansky 1933, 1936). In 1987 Orr revisited these studies on hybrid male and female sterility between *D. pseudoobscura pseudoobscura* and *D. persimilis*, amongst other reasons, because the proxy for male sterility previously used was dubious. The main findings regarding the role of the Y were: i) males with heterospecific sex chromosomes are sterile regardless of the autosomal background, ii) substitution of the Y chromosome has a large effect on the fertility of hybrid males in the direction predicted by X-Y interaction theory, and iii) hybrid females having an autosomal imbalance as severe as that of F_1 males are highly fertile. Later this study was extended to the cross between subspecies of *D. pseudoobscura* (*D. p. pseudoobscura* and *D. p. bogotana*) but no Y chromosome effect was detected (Orr 1989).

The strongest evidence for a crucial Y chromosome effect comes from *D. mojavensis* and *D. arizonae* hybrids. This seems to be one case of very incipient speciation, since both reciprocal crosses produce fertile females and only the cross between *D. arizonae* females and *D. mojavensis* males produces sterile males. In this cross the Y chromosome from *D. arizonae* interacts with *D. mojavensis* alleles at autosomal loci to cause male hybrid sterility (Vigneault and Zouros 1986). Pantazidis *et al.* (1993) observed that when *D. arizonae* Y chromosome is co-introgressed with the *SMF* locus of *D. arizonae*, into *D. mojavensis*, fertility is restored. In this case of very recent speciation, no X chromosome effect has been detected only an interaction between the Y chromosome and an autosome (Vigneault and Zouros 1986; Pantazidis *et al.* 1993).

Two other studies using laboratory crosses in *Drosophila* are worth mentioning. First, a significant effect of the Y chromosome was not detected in crosses between *D. virilis* and *D. novomexicana*, but a 15% reduction in fertility was observed in the cross between *D. virilis* and *D. texana*. This effect was half the effect of the X chromosome (Orr and Coyne 1989). Second, at least three genes on the X

and a gene on the Y chromosome have been found to cause hybrid male sterility between *D. yakuba* and *D. santomea* (Coyne *et al.* 2004).

The genetics of postzygotic isolation have received much less attention in mammals, but it is known that in two specific laboratory crosses between different strains of house mice, sex reversal is sometimes observed (Eicher and Washburn 1986). In one of this crosses, laboratory strains carrying an Y^{POS} chromosome (from *Mus domesticus poschiavinus*) produce males with ovaries when crossed with laboratory strains that are autosomaly mostly of *Mus musculus* origin (such as C57BL/B6). Eicher *et al.* (1996) determined that Y^{POS} interacts at least with *tda1* (chromosome 4) and *tda2* (chromosome 2) to cause the observed phenotype. A third factor, *tda3* (chromosome 5) also seemed to be implicated but the evidence was not strong. In the other sex reversing cross, incompatibilities arise when a *Mus domesticus* (from laboratory strain ARK/J) Y chromosome is introgressed into a predominantly *M. musculus* (from laboratory strain B6) genomic background. Using a transgenic rescue approach Washburn *et al.* (2001) identified the *Sry* locus as the factor causing sex reversal on a *Mus musculus* background. Sequencing of *Sry* in *Mus musculus* and *Mus domesticus* individuals allowed them to sort them into two groups. These were differentiated only by a 10 bp deletion in the 5' region of *Sry* and a linked glutamine repeat in the *Sry* open reading frame. The genetic basis of sex reversal (and thus male sterility) in this cross is caused by one or both these changes in *Sry*.

Further evidence for a Y chromosome effect on intrinsic reproductive isolation in mammals comes from the fact that the Y (and the X) chromosome shows reduced levels on introgression (relative to the autosomes) across two transects of the European hybrid zone between *Mus musculus* and *Mus domesticus* (Vanlerberghe *et al.* 1986; Tucker *et al.* 1992). Again the Y chromosome has a conflicting role in reproductive isolation since introgression of the Y across a third transect of this hybrid zone is not reduced (Munclinger *et al.* 2002). In Macaque species, phylogenetic studies have shown that the Y chromosome is always monophyletic, while mtDNA is polyphyletic. This is thought to be due to non introgression of Y chromosomes relative to other genomic regions (Tosi *et al.* 2003). In the hybrid zone between sister species of the *Drosophila melanogaster* subgroup, *D. santomea* and *D. yakuba*, there is absence of introgression at three Y linked genes and reduced introgression of the X chromosome relative to regions of the autosomes with similar levels of recombination (Llopart *et al.* 2005).

1.4 Objectives and organization of the thesis

This thesis is organized in six chapters. The first one is a general introduction to the European rabbit and to some areas of modern population genetics that will be instrumental in understanding the research presented here. The three following chapters reflect the main goals of this thesis: development of informative male specific molecular markers and the study of their peculiar mode of molecular evolution, the study of subspecies specific processes based on sex specific markers and the study of the speciation process between subspecies of the European rabbit. Chapter five is the general discussion where we will try to summarize and put into context the three previous chapters under the theory presented in chapter one and, finally, in chapter six we will highlight some of our findings and present future directions that we would like to see happening with this line of research. We will now describe chapters two, three and four of the thesis.

The second chapter deals with the development of male specific molecular markers in the European rabbit and their mode of molecular evolution and includes two papers already published. The first one is entitled "High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) SRY gene" and was published in Animal Genetics. This manuscript describes how we were able to sequence for the first time more than 2 Kb of the Y chromosome in a Leporid species, the European rabbit. This work was possible because of the existence of a European rabbit BAC (Bacterial Artificial Chromosome) library (Rogel-Gaillard *et al.* 2001), where a BAC clone containing the SRY gene had previously been mapped to the Y chromosome (Hayes *et al.* 2002). We performed direct sequencing of this clone starting on the conserved HMG (High Mobility Group) box of the SRY gene and moving outwards using a primer walking approach. We were able to sequence the entire open reading frame, the entire 3' UTR, and part of the 5' region of SRY where two microsatellite repeats were found. In this manuscript we also performed a preliminary survey of nucleotide polymorphism in this region (the microsatellite repeats were excluded) in four wild caught rabbits and eight domestic breeds. We found high levels of nucleotide polymorphism in wild caught animals, deep divergence between sequences from animals belonging to each subspecies and only one haplotype (similar to one haplotype from *O. c. cuniculus*) in domestic animals. The second paper is entitled "A 7-bp insertion in the 3' untranslated region suggests

the duplication and concerted evolution of the rabbit SRY gene" and was published in Genetics, Selection and Evolution. In this manuscript we survey more than 300 rabbits to investigate the geographic distribution of a 7 bp insertion in the SRY gene region. We found that the insertion is present only in the *algirus* lineage. Interestingly, we found not two genotypes (as expected for an haplotypic marker) but three, corresponding to a two locus system in which individuals either have the insertion in both gene copies, in only one gene copy, or they have no insertion. We suggest that this observation can best be explained by a duplication of the SRY gene in the rabbit Y chromosome where the two gene copies are evolving in concert through gene conversion. In the general discussion section, we present additional compelling data to support this claim. We also show evidence that this duplication is ancient and is present in other Lagomorphs and compare these findings with the finding of extensive gene conversion between palindrome arms of the human MSY.

In the third chapter of this thesis we use the Y-linked SRY gene as a male specific genetic marker and the mitochondrial Cytb gene as a female specific genetic marker to investigate sex specific demographic processes in each of the European rabbit subspecies. The findings from this work are present in a manuscript entitled "Sex-specific differences in effective population size and gene flow in subspecies of the European rabbit (*Oryctolagus cuniculus*) inferred from Y chromosome and mtDNA sequence data". This manuscript is still in preparation and will be submitted to the journal Molecular Ecology. The motivation for these analyses comes from the realization that we could exploit this newly available genetic marker (SRY) to directly compare male and female specific behaviors. The European rabbit is a species with a well documented behavior and social structure and presents two particularly interesting behaviors mentioned in the introduction of this thesis. First, females are phylopatric and males tend to disperse, to either join other social groups or form new ones, before their first breeding season. Second, in each social group, a dominant male controls the group's females and gets most of the matings. Part of this work deals with the direct comparison of male and female inherited markers to study these sex-specific behaviors and their impact on levels of male and female mediated gene flow and sex specific effective population sizes. We sequenced a portion of the SRY and Cytb genes in sample of approximately 100 male rabbits and asked: i) what is the divergence time between rabbit subspecies based on these markers?, ii) what is the time to the most recent common ancestor of the alleles found in each lineage?, iii) are

there differences in male and female effective population sizes?, and iv) are within subspecies levels of gene flow different for the Y chromosome and mtDNA?

Finally, in the fourth chapter we used the European rabbit as a model to study the phenomena of population divergence and isolation. The chapter is composed of two manuscripts. The first paper in this chapter paper has now been published in the journal Genetics and is entitled "Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*)". Motivated by the large X-effect in speciation documented in other species we investigated patterns of nucleotide polymorphism and introgression at four X-linked genes in 43 wild caught rabbits from the Iberian Peninsula. We addressed three main questions: i) what are the levels and patterns of nucleotide variability at genes on the rabbit X chromosome?, ii) are patterns of variation and introgression heterogeneous among loci and if so, do the differences observed correlate with genomic location?, and iii) are the data compatible with a model of divergence without gene flow? The second paper is entitled "Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula" and it has now been submitted for publication in the journal Evolution. In this work we investigate the distribution of the two previously found divergent Y chromosome lineages and compare the results to previously published data on other loci with the hope of providing strong evidence for a role of the Y chromosome in the speciation process between subspecies of rabbit. Our specific questions for this work were: i) what is the level of introgression between subspecies for the Y chromosome?, and ii) how do they relate to the levels of introgression found at other loci?

1.5 References

Arana A, Zaragoza P, Rodellar C and Amorena B. 1987. Evidence for transferrin polymorphism in Spanish wild rabbits. *Animal Genetics*, 18: 125-132.

Avise J C. 2000. Phylogeography: The history and formation of species. Harvard University Press. Harvard, US.

Backstrom N, Ceplitis H, Berlin S and Ellegren H. 2005. Gene conversion drives the evolution of *HINTW*, an ampliconic gene on the female-specific avian W chromosome. *Molecular Biology and Evolution*, 22: 1992-1999.

Banks, R. 1989. "Rabbits: Models and Research Applications (USAMRIID Seminar Series)" (On-line at http://netvet.wustl.edu/species/rabbits/rabtmodl.txt)

Barbash D A, Siino D F, Tarone A M and Roote J. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proceedings of the National Academy of Sciences USA*, 100: 5302-5307.

Bateson W. 1909. Heredity and variation in modern lights. Pp. 85-101 *in Darwin and modern science*. Cambridge Univ. Press. Cambridge, U.K.

Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, Soriguer R C, Gaaied A, Hili A and Mounolou J C. 1991. Mitochondrial DNA evolution in Lagomorphs: Origin of systematic heteroplasmy and organization of diversity in european rabbits. *Journal of Molecular Evolution*, 33: 92-102.

Branco M, Ferrand N and Monnerot M. 2000. Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome *b* gene. *Heredity*, 85: 307-317.

Branco M, Monnerot M, Ferrand N and Templeton A R. 2002. Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution*, 56:792-803.

Callou C. 2003. De la garenne au clapier: Etude archeozoologique du lapin en Europe occidentale. Muséum National d'Histoire Naturelle. Paris, France.

Calvete C and Estrada R. 2000. Epidemiología de enfermedad hemorrágica viral (VHD) y mixomatosis en el conejo silvestre en el valle médio del Ebro. *Herramientas de gestión*. Consejo de Protección de la Naturaleza de Aragón. Zaragoza, Spain.

Carvalho A B, Lazzaro B P and Clark A G. 2000. Y chromosomal fertility factors *kl-2* and *kl-3* of *Drosophila melanogaster* encode dynein heavy chain polypeptides. *Proceedings of the National Academy of Sciences USA*, 97: 13239-13244.

Carvalho A B, Dobo B A, Vibranovski M D and Clark A G. 2001. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences USA*, 98: 13225-13230.

Castle W E. 1924. On the occurrence in rabbits of linkage in inheritance between albinism and brown pigmentation. *Proceedings of the National Academy of Sciences USA*, 10: 486-488.

Coyne J A. 1985. The genetic basis of Haldane's rule. *Nature*, 314: 736-738.

Coyne J A. 1992. Genetics and speciation. *Nature*, 355: 511-515.

Coyne J A and Orr H A. 1989. Two rules of speciation. Pp. 180-207 *in Speciation and its consequences*. Sinauer Associates. Maryland, US.

Coyne J A and Orr H A. 1997. "Patterns of speciation in *Drosophila*" revisited. *Evolution*, 51: 295-303.

Coyne J A and Orr H A. 2004. Speciation. Sinauer Associates. Maryland, US.

Coyne J A, Elwyn S, Kim S Y and Llopart A. 2004. Genetic studies of two sister species in the *Drosophila melanogaster* subgroup, *D. yakuba* and *D. santomea*. *Genetical Research (Cambrige)*, 84: 11-26.

Daly J C. 1981. Effects of social organization and environmental diversity on determining the genetic structure of a population of the wild rabbit, *Oryctolagus cuniculus*. *Evolution*, 35: 689-706.

Dobzhansky T H. 1933. On the sterility of interracial hybrids in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences (US)*, 19: 397-403.

Dobzhansky T H. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, 21: 113-135.

Eicher E M and Washburn L L. 1986. Genetic control of primary sex determination in mice. *Annual Reviews in Genetics*, 20: 327-360.

Eicher E M, Washburn L L, Schork N J, Lee B K, Shown E P, Xu X, Dredge R D, Pringle M J and Page D C. 1996. Sex-determining genes on mouse autosomes identified by linkage analysis of C57BL/6J-Y^{POS} sex reversal. *Nature Genetics*, 14: 206-209.

Emmerson J J, Kaessmann H, Betran E, Long M. 2004. Extensive gene traffic on the mammalian X chromosome. *Science*, 303: 537-540.

Feng J, Laija C, Taylor D J and Webster M S. 2001. Genetic distinctivness of endangered Dwarf Blue Sheep (*Pseudois nayaur schaeferi*): Evidence from mitochondrial control region and Y-linked *Zfy* intron sequences. *The Journal of Heredity*, 92: 9-15.

Ferrand N. 1995. Variação genética de proteínas em populações de coelho (*Oryctolagus cuniculus*) (PhD Dissertation). Universidade do Porto. Porto, Portugal.

Ferrand N and Branco M. 2006. The evolutionary history of the European rabbit (*Oryctolagus cuniculus*), pp. 207-235, *in* Phylogeography in Southern European Refugia. Edited by S. Weiss and N. Ferrand. Springer. Amsterdam, The Netherlands. (in press).

Flux, J E C. 1994. World distribution. Pp. 8-21 *in The European rabbit: the history of a successful colonizer.* Oxford Science Publications. Oxford, UK.

Gerrard D T and Filatov D A. 2005. Positive and negative selection on mammalian Y chromosomes. *Molecular Biology and Evolution*, 22: 1423-1432.

Gomez A, Carvalho G R and Lunt H. 2000. Phylogeography and regional endemism of a passively dispersing zooplankter: mitochondrial DNA variation in rotifer resting egg banks. *Proceedings of the Royal Society of London Serie B-Biological Sciences*, 267: 2189.2197.

Gomez A and Lunt D H. 2006. Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula, pp. 155-188, *in* Phylogeography in Southern European Refugia. Edited by S. Weiss and N. Ferrand. Springer. Amsterdam, The Netherlands. (in press).

Guenet J L, Naganime C, Simon-Chazottes D, Montagutelli X and Bonhomme F. 1990. *Hst-3*: an X-linked hybrid sterility gene. *Genetical Research*, 56: 163-165.

Hagen R H. 1990. Population structure and host use in hybridizing subspecies of *Papilio glaucus* (Lepidoptera: Papilionidae). *Evolution*, 44: 1914-1930.

Haldane J B S. 1922. Sex ratio and the unisexual sterility in animal hybrids. *Journal of Genetics*, 12: 101-109.

Haldane J B S. 1932. The causes of evolution. Longmans. New York, US.

Hammer M F, Blackmer F, Garrigan D, Nachman M W and Wilder J A. 2003. Human population structure and its effects on sampling Y chromosome variation. *Genetics*, 164:1495-1509.

Hayes H, Rogel-Gaillard C, Zijlstra C, de Haan N A, Urien C *et al.*. 2002. Establishment of an R-banded rabbit karyotype nomenclature by FISH localization of 23 chromosome-specific genes on both g- and R-banded chromosomes. *Cytogenetic and Genome Research*, 98: 199-205.

Hellborg L, Gunduz I and Jaarola M. 2005. Analyses of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, 14: 2025-2031.

Hey J and Machado C. 2003. The study of structured populations: new hope for a divided science. Nature Reviews in Genetics, 4:535-543.

Hewitt G M. 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405: 907-913.

Hudson R R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, 7: 1-44.

Johnson N A, Hollocher H, Noonburg E and Wu C I. 1993. The effects of interspecific Y chromosome replacement on hybrid sterility within the *Drosophila simulans* clade. *Genetics*, 135: 443-453.

Kingman J F C. 1982. The coalescent. *Stochastic Processes and their Applications*, 235-248.

Kong A, Gudbjartsson D F, Sainz J, Jonsdottir G M, Gudjonsson S A *et al.*. 2002. A high-resolution recombination map of the human genome. *Nature Genetics*, 31: 241-247.

Künkele J and VonHolst D. 1996. Natal dispersal in the European wild rabbit. *Animal Behaviour*, 51: 1047-1059.

Lahn B T and Page D C. 1999. Four evolutionary strata on the human X chromosome. *Science*, 286: 964-967.

Li W H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitutions. *Journal of Molecular Evolution*, 36: 96-99.

Li W H, Yi S and Makova K. 2002. Male-driven evolution. *Current opinion in Genetics and Development*, 12: 650-656.

Lindgren G, Backstrom N, Swinburne J, Hellborg L, Einarsson A *et al.*. 2004. Limited number of patrilines in horse domestication. *Nature Genetics*, 36: 335-336.

Livak K J. 1984. Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics*, 107: 611-634.

Llopart A, Lachaise D and Coyne J A. 2005. Multilocus analysis of introgression between two sympatric species of Drosophila: *D. yakuba* and *D. santomea*. *Genetics*, 171: 197-210.

Lombardi L, Fernandez N, Moreno S and R. Villafuerte. 2003. Habitat-related differences in rabbit (*Oryctolagus cuniculus*) abundance, distribution, and activity. *Journal of Mammalogy*, 84: 26-36.

Lopez-Maritnez N. 1989. Revision sistematica y biostratigrafica de los lagomorphos (Mammalia) del neogeno y cuaternario de España. Memoria del Museo Paleontológico de la Universidade de Zaragoza, Diputación General de Aragón.

Lumaret R, Mir C, Michaud H and Raynal V. 2002. Phylogeographical variation of chloroplast DNA in holm oak (*Quercus ilex* L.). *Molecular Ecology*, 11: 2327-2336.

Malitschek B, Fornzler D and Schartl M. 1995. Melanoma formation in *Xiphophorus*: A model system for the role of receptor kinases in tumorigenesis. *BioEssays*, 17: 1017-1023.

Martins H, Milne J A and Rego F. 2002. Seasonal and spatial variation in the diet of the wild rabbit (*Oryctolagus cuniculus* L.) in Portugal. *Journal of Zoology*, 258: 395-404.

Matthee C A, van Vuuren B J, Bell D and Robinson T J. 2004. A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the miocene. *Systematic Biology*, 53: 433-447.

Mayr E. 1942. *Systematics and the origin of Species*. Columbia University Press. New York, US.

Mayr E. 1995. Species, classification, and evolution. Pp. 3-12 *in Biodiversity and evolution*. National Science Museum Foundation. Tokyo, Japan.

Muller H J. 1940. Bearing of the *Drosophila* work on systematics. Pp.185-268 *in The new systematics*. Clarendon press. Oxford, U.K.

Muller H J. 1942. Isolating mechanisms, evolution and temperature. *Biological. Symposium*, 6: 71-125.

Munclinger P, Bozikova E, Sugerkova M, Pialek J and Macholan M. 2002. Genetic variation in house mice (*Mus*, Muridae, Rodentia) from the Czech and Slovak Republics. *Folia Zoologica*, 51: 81-92.

Murphy W J, Eizirik E, O'Brien S J, Madsen O, Scally M *et al.*. 2001. Resolution of the early Placental Mammal radiation using bayesian phylogenetics. *Science*, 294:2348-2351.

Myers K and Mykytowicz R. 1958. Social behaviour in the wild rabbit. *Nature*, 181: 1515-1516.

Orr H A. 1987. Genetics of male and female sterility in hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, 116: 555-563.

Orr H A. 1989. Genetics of sterility in hybrids between two subspecies of Drosophila. *Evolution*, 43: 180-189.

Orr H A. 1995. The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics*, 139: 1805-1813.

Orr H A and Coyne J A. 1989. The genetics of postzygotic isolation in the *Drosophila virilis* group. *Genetics*, 121: 527-537.

Orr H A and Turelli M. 2001. The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. *Evolution*, 55: 1085-1094.

Pantazidis A C, Galanopoulos V K and Zouros E. 1993. An autosomal factor from *Drosophila arizonae* restores normal spermatogenesis in *D. mojavensis* males carrying a *D. arizonae* Y chromosome. Genetics, 134: 309-318.

Petit E, Balloux F and Excoffier L. 2002. Mammalian population genetics: Why not Y? *Trends in Ecology and Evolution*, 17: 28-33.

Presgraves D C, Balagopalan L, Abmayr S M and Orr H A. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature*, 423: 715-719.

Queney G, Ferrand N, Weiss S, Mougel F and Monnerot M. 2001. Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution*, 18: 2169-2178.

Queney G, Vachot M A, Brun J M, Dennebuoy N, Mulsant P and Monnerot M. 2002. Different levels of human intervention in domestic rabbits: effects on genetic drift. *Journal of Heredity*, 93: 205-209.

Richardson B J, Rogers P M and Hewitt G M. 1980. Ecological genetics of the wild rabbit in Australia. II. Protein variation in British, French and Australian rabbits and the geographical distribution of the variation in Australia. *Australian Journal of Biological Sciences*, 33: 371-383.

Richardson B J, Hayes R A, Wheeler S H and Yardin M R. 2002. Social structures, genetic structures and dispersal strategies in Australian rabbit (*Oryctolagus cuniculus*) populations. *Behavioral Ecology and Sociobiology*, 51:113-121.

Robinson T J and Matthee C A. 2005. Phylogeny and evolutionary origins of the Leporidae: a review of cytogenetics, molecular analyses and a supermatrix analyses. *Mammal Review*, 35:231-247.

Rogel-Gaillard C, Piumi F, Billaut A, Borgeaux N, Save J C *et al.*. 2001. Construction of a rabbit bacterial artificial chromosome (BAC) library: application to the mapping of the major histocompatibility complex to position 12.q.1.1. *Mammalian Genome*, 12: 235-255.

Rozen S, Skaletsky H, Marszalek J D, Minx P J *et al.*. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, 423: 873-876.

Salvador L, Alia R, Agundez R and Gil L. 2000. Genetic variation and migration pathways of maritime pine (Pinus pinaster Ait.) in the Iberian Peninsula. *Theoretical Applied Genetics*, 100: 89-95.

Sandstedt S A and Tucker P K. 2004. Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome. *Genome Research*, 14: 267-272.

Santucci F, Emerson B C and Hewitt G M. 1998. Mitochondrial DNA phylogeography of European hedgehogs. *Molecular Ecology*, 7: 1163-1172.

Sharples C M, Fa J E and Bell D J. 1996. Geogreaphical variatoion in size in the European rabbit *Oryctolagus cuniculus* (Lagomorpha: Leporidae) in western Europe and North Africa. *Zoological Journal of the Linnean Society*, 117: 141-158.

Skaletksy H, Kuroda-Kawaguchi T, Minx P J, Cordum H S *et al.*. 2003. The male-specific region of the Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423: 825-837.

Thompson H V and King C M. 1994. *The European rabbit: the history of a successful colonizer.* Oxford Science Publications. Oxford, UK.

Ting C-T, Tsaur C-S, Wu M-L and Wu C-I. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science*, 282: 1501-1504.

Tosi A J, Morales J C and Melnick D J. 2003. Paternal, maternal and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution*, 57:1419-1435.

Tucker P K and Lundrigan B L. 1993. Rapid evolution of the sex determining locus in Old World mice and rats. *Nature*, 364: 715-717.

Tucker P K, Sage R D, Warner G, Wilson A C and Eicher E M. 1992. Abrupt cline for sex chromosomes between two species of mice. *Evolution*, 46:1146-1163.

Vanlerberghe F, Dod B, Boursot P, Bellis M, and Bonhomme F. 1986. Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. *Genetical Research (Cambrige)*, 48: 191-197.

Vigneault G and Zouros E. 1986. The genetics of asymmetrical male sterility in *Drosophila mojavensis* and *D. arizonensis* hybrids: interactions between the Y-chromosome and autossomes. *Evolution*, 40: 1160-1170.

Villafuerte R. 2002. *Oryctolagus cuniculus in Atlas de los Mamíferos terrestres de España*, pp. 464-467, Dirección General de Conservación de la Naturaleza-SECEM-SECEMU. Madrid, Spain.

Washburn L L, Albrecht K H and Eicher E E. 2001. C57BL/6J-T-associated sex reversal in mice is caused by reduced expression of a *Mus domesticus* SRY allele. *Genetics*, 158: 1675-1681.

Webb N J, Ibrahim K M, Bell D J and Hewitt G M. 1995. Natal dispersal and genetic structure in a population of the European wild rabbit (*Oryctolagus cuniculus*). *Molecular Ecology*, 4: 239-247.

Weiss S and Ferrand N. 2006. Phylogeography in Southern European Refugia. Springer. Amsterdam, The Netherlands. (in press).

Whitfield L S, Lovell-Badge and Goodfellow P N. 1993. Rapid sequence evolution of the mammalian sex-determining gene *SRY*. *Nature*, 364: 713-715.

Wyckoff G J, Li J and Wu C I. 2002. Molecular evolution of functional genes on the mammalian Y chromosome. *Molecular Biology and Evolution*, 19: 1633-1636.

Wu C I and Davis A W. 1993. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. *The American Naturalist*, 142:187-212.

Wu C I, Johnson N A and Palopoli M F. 1996. Haldane's rule and its legacy: Why are there so many sterile males? *Trends in Ecology and Evolution*, 11:281-284.

Zeng L W and Singh R S. 1993. The genetic basis of Haldane's rule and the nature of asymmetric hybrid male sterility among *Drosophila simulans, D. mauritiana and D. sechellia. Genetics*, 134: 251-260.

2. MALE SPECIFIC MOLECULAR MARKERS AND INSIGHTS INTO THEIR PECULIAR MODE OF MOLECULAR EVOLUTION

2.1 Paper I

Published in, Animal Genetics (2005), 36: 349-351.

High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) SRY gene

A. Geraldes*, C. Rogel-Gaillard† and N. Ferrand*

*CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661, Vairão, and Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, Portugal. †Laboratoire de Radiobiologie et d'Etude du Génome, UMR INRA-CEA 13.314, INRA, Jouy en Josas, France

Summary

We have sequenced 2388 bp of the European rabbit *sex determining region Y* (SRY) gene. These data provide a 10-increase in the coverage of the Y chromosome in this species, including the entire open reading frame of the SRY, the polyadenylation signal, and two repetitive sequences in the 5'-region. A survey of 2021 bp of this gene in eight domestic breeds and four wild individuals revealed a total of nine single nucleotide polymorphisms and one indel, defining two deeply divergent lineages. The resulting estimation of nucleotide diversity (π= $1.34X10^{-3}$) is very high when compared to other species, but no variability was detected among the domestic breeds. This study represents a first step in the characterization of the European rabbit Y chromosome and its variability. These sequences can be used in additional phylogeographical analyses of the European rabbit and other Leporid species, as well as in evolutionary studies of sex determination and the Y chromosome in wild species.

Keywords: European rabbit, SRY gene, single nucleotide polymorphism, Y chromosome

In the last decade, many studies have focused on the study of the mammalian Y chromosome, namely in the fields of public health (Jobling & Tyler-Smith 2000) and the evolution of sex determination (Graves 2002). The Y chromosome is the only portion of the mammalian genome that is exclusively paternally transmitted and therefore, defines patrilineages. Because of that, the Y chromosome is a unique tool in studying the male contribution to the evolutionary history of a species. For example, Y-linked genetic markers have provided considerable insight into the phylogeographic history of humans (Hammer *et al.* 1997). However, application of Y chromosome data to non-primate population genetic studies is still rare because of lack of Y chromosome markers and sequence information (Petit *et al.* 2002) and of low levels of (Hellborg & Ellegren 2004; Meadows *et al.* 2004).

The European rabbit, *Oryctolagus cuniculus*, is a Leporid species native to the Iberian Peninsula. Partially because of its high economic interest both as a game species and as a source of meat and fur, domestic and wild animals have been introduced across the world. Many genetic studies have been carried out in the rabbit, but at the genomic level it is still a very poorly characterized species. Some of these studies have focused on describing its genetic diversity and geographical distribution. The results have been somewhat discordant, ranging from a very strong phylogeographical pattern of two highly divergent mtDNA clades that overlap in a narrow contact zone in central Iberian Peninsula (Branco *et al.* 2000), to a complete lack of population structure as depicted in the study of autosomal microsatellites (Queney *et al.* 2001).

Sinclair *et al.* (1990) described a 35 Kb region of the human Y chromosome including the SRY gene. They also reported that the gene is present in several other mammals including the European rabbit, from which they sequenced 357 bp. Recently, Wallner *et al.* (2001) successfully amplified the conserved HMG box of SRY both from *O. cuniculus* and a related Leporid species, the brown hare (*Lepus europaeus*), and Hellborg & Ellegren (2003) amplified > 3.5 Kb from two Y-linked genes of *Lepus europaeus*.

In this work, a European rabbit BAC library (Rogel-Gaillard *et al.* 2001) was screened by PCR with the primers described by Wallner and colleagues (2001), and the BAC clone 828D7 was retrieved. The mapping of the rabbit SRY gene to position Yp12 by FISH of that rabbit BAC clone (Rogel-Gaillard *et al.* 2001) was previously reported by Hayes *et al.* (2002). To expand the sequenced region of the SRY, DNA

from this BAC clone was first prepared using classical alkaline lyses method followed by purification with the Qiaex II gel extraction kit under conditions recommended by the manufacturer (Qiagen, Valencia, CA, USA). Two micrograms of DNA were used for each sequencing reaction. The first set of sequencing reactions was performed using the primers SRYf and SRYr (Wallner *et al.* 2001). Sequences were aligned to each other and a total of five primers, f2-1107 (5'- AGT GAA CAC AGG CAG CTC GC -3'), f3-1552 (5'- GCA TCT CCC TGT AAA CTA CCC -3'), f4-1773 (5'- GTC CTC CTT CAA CAG TGG GG -3'), r2-710 (5'- CGT GTT CCT CCT GTA TTG C -3') and r3-378 (5'- GCC CTG GAG CTC AAG TGC CC -3'), were designed to sequence outwards. With this primer walking approach a concatenated sequence of 2388 bp was obtained (GenBank Accession number AY785433). Using BioEdit (Hall, 1999), we identified the 621 bp SRY ORF (Fig. 1) between nucleotides 613 and 1233. The conserved HMG box lies between nucleotides 736 and 981 and the polyadenylation (AATAAA) signal between nucleotides 2040 and 2045. In the 5' region of the sequence we found two repetitive sequences, an $A_{(14)}$ (nucleotides 36 to 49) and a $TG_{(13)}$ (nucleotides 204 to 229).

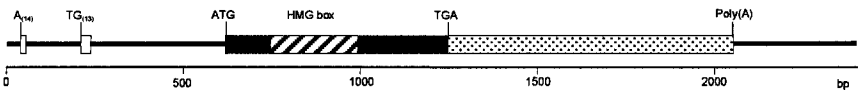


Figure 1 Schematic representation of the European rabbit (*Oryctolagus cuniculus*) SRY genomic locus and flanking regions. Portions of the SRY transcription unit are indicated: 3' untranslated region (dotted), HMG box encoding region (hatched) and the other coding regions bounded by initiation (ATG) and stop/TGA codons (solid). Also indicated are the two repetitive regions (open boxes).

Twelve samples (Table 1) were PCR amplified using primers YTF-272 (5'- GCT GTG ATT TTC AAA GGC AAC -3') and YTR-2343 (5'- GTG TTG CTT GTG GCT GAG AA -3'). Four wild rabbits were chosen to represent the two very divergent lineages found at the mtDNA level (Branco *et al.* 2000) along with eight domestic rabbits belonging to eight domestic breeds. Fifty-microlitre PCR was performed using 50 ng of genomic DNA, 0.4 μM of each primer, 2 mM $MgCl_2$, 0.8mM dNTP, 1 U Platinum *Taq* DNA Polymerase High Fidelity and reaction buffer (600 mM Tris-SO_4 and 180 mM Ammonium Sulfate) (Invitrogen, Carlsbad, CA, USA). Reactions consisted of 2 min initial denaturation followed by 40 cycles of 94°C for 20 s, 58°C for 20 s and 68°C for 3 min. The products were purified using the QIAquick PCR Purification Kit following conditions recommended by the manufacturer (Qiagen), and then sequenced on an ABI 3100 automated sequencer

Table1 Individual samples and polymorphic sites at SRY.

	493	596	597	1233	1490[1]	1594	1652	1708	2098	2153
New Zealand (828D7)[2]	C	T	C	G	-	A	G	A	T	A
Argenté de Champagne[2]	.	.	.	.	.	.	.	.	.	.
Fauve de Bourgogne[2]	.	.	.	.	.	.	.	.	.	.
Chinchilla[2]	.	.	.	.	.	.	.	.	.	.
French Lop[2]	.	.	.	.	.	.	.	.	.	.
Vienna White[2]	.	.	.	.	.	.	.	.	.	.
English[2]	.	.	.	.	.	.	.	.	.	.
Flemish Giant[2]	.	.	.	.	.	.	.	.	.	.
Zrg16[3]	.	.	.	.	.	.	.	.	.	.
Alt120[4]	.	.	.	.	.	.	C	G	.	.
Luc9[5]	G	A	A	A	.	C	C	.	A	G
Lom3[6]	G	A	A	A	G	C	C	.	A	G

Sites are numbered according to the reference sequence (GenBank accession no. AY785433) of clone 828D7.
[1]Sample Lom3 had the following 7 bp indel after nucleotide 1490: GAATTAA.
[2]Domestic breeds.
[3,4]Sample Zrg16 was collected near Zaragoza, Northeast Spain. Sample Alt120 was collected near Alicante, Eastern Spain. These samples were chosen to reflect mtDNA lineage B (Branco *et al.* 2000).
[5,6] Sample Luc16 was collected near Cordoba, Southwestern Spain. Sample Lom3 was collected near Las Lomas, Southwestern Spain. These samples were chosen to reflect mtDNA lineage A (Branco *et al.* 2000).

(Applied Biosystems, Foster City, CA, USA) using primers SY1034R (5'- CGT CCA TAT GCA CCT CGC TGC -3'), SY949F (5'- GAC TAC AAG TAC AGA CCT CG -3'), SY1517F (5' –GAC TAA GTT GGT GAT TAG TG -3') and SY1991F (5' – AGA TGA AGT AGA AAC TGG CAG -3'). A total of 2021 bp, between nucleotides 318 and 2331 of the reference sequence (AY785433), were sequenced and aligned by eye using the BioEdit program (Hall 1999), resulting in the identification of nine segregating sites and a 7 bp indel that define four haplotypes (Table1). Only one out of the four polymorphic positions is located within the SRY ORF and it is a silent substitution. Nucleotide variability for the 2021 bp sequenced in these 12 individuals was $\pi = 1.34X10^{-3}$ (SD = $5.6X10^{-4}$) (Nei & Li 1979). We observed a deep split between Southwest Iberian samples and all others, defining two lineages that differed by a total of seven fixed positions.

To our knowledge this estimate of nucleotide diversity in the rabbit Y chromosome ($\pi = 1.34X10^{-3}$) is the largest reported so far in any mammalian species (for a review see Meadows *et al.* 2004). This results from the occurrence of two highly divergent Y chromosome lineages that most probably reflect the same molecular signature of strong population subdivision described for the mtDNA molecule in Iberian populations (Branco *et al.* 2000). It is possible that the persistence of wild rabbit populations in different refugia during the Pleistocene climatic

oscillations resulted in long coalescent times for the SRY genealogy and associated high levels of polymorphism. In contrast, the screening of rabbits from eight different domestic breeds failed to reveal any SNP, in agreement with the conclusion of Meadows *et al.* (2004) suggesting that very low levels of nucleotide diversity in the Y chromosome of mammals results from a sharp reduction in male effective population size.

Further research using the Y chromosome markers now described may be of help in illuminating the evolutionary history of the European rabbit. The data may also be used to contrast the apparent high nucleotide diversity of wild populations with the low levels of polymorphism that characterize domestic breeds. We also expect that the description of > 2,3 Kb of sequence of the rabbit SRY gene may prove useful not only for those working on the evolutionary history and phylogeny of Leporid species, but also for those studying the evolution of sex determination and the Y chromosome in general.

Acknowledgments

We thank Noelle Bourgeaux for preparing the DNA used for direct sequencing of the clone 828D7. The comments of two anonymous reviewers and Dr. Noelle E. Cockett considerably improved the quality of a previous version of this manuscript. Financial support was obtained from Fundação para a Ciência e a Tecnologia (SFRH/BD/4621/2001 PhD grant to A. Geraldes and research project POCTI/BSE/40280/2001).

References

Branco M., Ferrand N. and Monnerot M. (2000) Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome *b* gene. *Heredity* **85**, 307-17.

Graves J.A.M. (2002) The rise and fall of the SRY. *Trends In Genetics* **18**, 259-64.

Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-8.

Hammer M.F., Spurdle A.B., Karafet T. *et al.* (1997) The geographic distribution of human Y chromosome variation. *Genetics* **145**, 787-805.

Hayes H., Rogel-Gaillard C., Zijlstra C., de Haan N.A., Urien C., Bourgeaux N., Bertaud M. and Bosma A.A. (2002) Establishment of an R-banded rabbit karyotype nomenclature by FISH localization of 23 chromosome-specific genes on both G- and R-banded chromosomes. *Cytogenetic and Genome Research* **98**, 199-205.

Hellborg L. and Ellegren H. (2003) Y chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male-specific DNA. *Molecular Ecology* **12**, 283-91.

Hellborg L. and Ellegren H. (2004) Low levels of nucleotide diversity in mammalian Y chromosomes. *Molecular Biology and Evolution* **21**, 158-63.

Jobling M.A. and Tyler-Smith C. (2000) New uses for new haplotypes the human Y chromosome, disease and selection. *Trends In Genetics* **16**, 356-62.

Meadows J.R.S., Hawken R.J. and Kijas J.W. (2004) Nucleotide diversity on the ovine Y chromosome. *Animal Genetics* **35**, 379-85.

Nei M. and Li W.H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76**, 5269-73.

Petit E., Balloux F. and Excoffier L. (2002) Mammalian population genetics: why not Y? *Trends in Ecology & Evolution* **17**, 28-33.

Queney G., Ferrand N., Weiss S., Mougel F. and Monnerot M. (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution* **18**, 2169-78.

Rogel-Gaillard C., Piumi F., Billault A., Bourgeaux N., Save J.C., Urien C., Salmon J., Chardon P. (2001) Construction of a rabbit bacterial artificial chromosome (BAC) library: application to the mapping of the major histocompatibility complex to position 12q1.1. *Mammalian Genome* **12**, 253-5.

Sinclair A.H., Berta P., Palmer M.S., Hawkins J.R., Griffiths B.L., Smith M.J., Foster J.W., Frishaud A., Lovell-Badge R. and Goodfellow P.N. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240-44.

Wallner B., Huber S. and Achmann R. (2001) Non-invasive PCR sexing of rabbits (*Oryctolagus cuniculus*) and hares (*Lepus europaeus*). *Mammalian Biology* **66**, 190-92.

2.2 Paper II

Published in, Genetics, Selection, Evolution (2006), 38: 313-320

A 7-bp insertion in the 3' untranslated region suggests the duplication and concerted evolution of the rabbit SRY gene

A. Geraldes[*,†] and N. Ferrand[*]

[*]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, e Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal. [†]IREC, Instituto de Investigación en Recursos Cinegéticos (CSIC/UCLM/JCCLM), Spain

Abstract

In this work we report the genetic polymorphism of a 7-bp insertion in the 3' untranslated region of the rabbit SRY gene. The polymorphic GAATTAA motif was found exclusively in one of the two divergent rabbit Y-chromosomal lineages, suggesting that its origin is more recent than the separation of the *O. c. algirus* and *O. c. cuniculus* Y-chromosomes. In addition, the remarkable observation of haplotypes exhibiting 0, 1 and 2 7-bp inserts in essentially all *algirus* populations suggests that the rabbit SRY gene is duplicated and evolving under concerted evolution.

Keywords: rabbit, SRY gene, polymorphic insertion, duplication, concerted evolution

Introduction

The recent sequencing of the human male-specific region of the Y chromosome (MSY) revealed a mosaic of X-transposed, X-degenerate and ampliconic sequence classes [15]. While the first two classes showed a total of 18 single-copy genes, the latter exhibited multiple copies of an additional nine genes that were mostly associated with palindromic regions. All these multi-copy genes are testis-specific and their evolution is shaped by abundant gene conversion as demonstrated by very high intra-palindromic sequence identity in humans and apes [14]. A notable exception to this pattern is the sex-determining gene, SRY, which is predominantly expressed in testes but is X-degenerate and represented by a single copy. Whereas a similar observation has been made in the house mouse, *Mus musculus*, evidence for two or more copies of the SRY gene in various rodent species is accumulating[5,7,9,11].

The European rabbit (*Oryctolagus cuniculus*) is a mammalian species native to the Iberian Peninsula that successfully colonized many regions of the world by a combination of natural and human-mediated processes [10,12] and which domestication was achieved during the Middle Ages [4]. Biju-Duval *et al.* [1] first showed the occurrence of two highly divergent mtDNA lineages in Iberian wild rabbit populations by using RFLP of the whole molecule and tentatively dated their divergence at 1-2 million years ago. Later, Branco *et al.* [2,3] described phylogeographical evidence for a recent contact zone bisecting the Iberian Peninsula from the Northwest to the Southeast and suggested that postglacial population expansions from two different refugia could explain the observed patterns. Recently, we reported high levels of nucleotide diversity in the rabbit SRY gene and suggested that this result could be explained by strong population subdivision [6]. In fact, the subsequent analysis of a comprehensive sample of wild rabbits from the Iberian Peninsula showed the existence of a relatively sharp contact zone between two divergent Y-chromosome lineages, in complete concordance with mtDNA data (Geraldes *et al.*, unpublished data).

In this study we describe the polymorphism and geographical distribution of a 7-bp insertion in the 3' untranslated region of the rabbit SRY that we previously identified in only one of the divergent Y-chromosome lineages [6] and use the strong

population structure of Iberian wild rabbits to better understand the evolution of this sex-determining gene in Lagomorphs.

Materials and Methods

Ear tissue was obtained from a total of 335 field-collected wild rabbits originating from 30 different populations in the whole Iberian Peninsula and southern France. In addition, a sample of 37 domestic rabbits from various breeds was also studied. Genomic DNA was isolated following standard protocols.

Initial PCR amplifications of the rabbit SRY gene followed the protocols described in Geraldes *et al.* [6] and confirmed the occurrence of a GAATTAA insertion between nucleotide positions 1490 and 1496, in the 3' untranslated region. Subsequently, the forward (5' to 3': CGG TGA TGT GAA ACA CAC AA) and reverse primers (5' to 3': TAC AGG GAG ATG CAC AAA CG) were developed and amplified fragments of 125 bp (absence of insertion) and 132 bp (presence of insertion). The amplification protocol included an initial denaturation at 94°C for 5 min, 35 cycles, with denaturation at 94 °C for 20 s, annealing at 54°C for 20 s, and elongation at 72 °C for 20 s, and a final extension step at 72 °C for 5 min. PCR products were visualized on 4.5% Metaphor agarose gels.

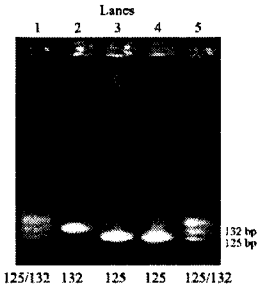


Figure 1- Agarose gel electrophoresis (Metaphor 4.5% w/v) showing the three banded patterns observed in the PCR amplification of the 3' region of the SRY gene. Lanes 1 and 5 exhibit PCR products of 125 and 132 bp and an additional heteroduplex band (corresponding to individuals with haplotype 125/132), lane 2 shows a single 132 bp band (corresponding to individuals with haplotype 132/132), and lanes 3 and 4 display a single 125 bp band (corresponding to individuals with haplotype 125/125).

Results and Discussion

Routine separation of PCR products designed to amplify the GAATTAA insertion showed clear band patterns: one-banded patterns were easily interpreted and corresponded to individuals with or without the 7-bp insertion, but three-banded patterns were unexpected and more difficult to explain (Fig. 1). These patterns were formed by both the 125 bp and 132 bp bands, as well as by an additional third band with lower mobility that was interpreted as a heteroduplex. Sequencing of PCR products displaying all three patterns confirmed this interpretation and overlapped sequences comparable to those observed in heterozygotes for autosomal markers were obtained for individuals showing three-banded patterns (results not shown). The most likely explanation for this observation is the presence of two SRY copies in the rabbit Y chromosome with a polymorphic 7 bp insertion.

The large scale application of our PCR protocol to screen the polymorphic 7-bp insertion in a comprehensive sample of Iberian wild rabbits showed notably that this polymorphism is restricted to rabbits possessing the Y chromosome lineage A (Fig. 2). These populations, that also exhibit mtDNA lineage A [3] and correspond to the subspecies *Oryctolagus cuniculus algirus*, are confined to Southwest Iberia and show relatively similar amounts of haplotypes 125/125 (no insertion), 125/132 (insertion in only one SRY copy) and 132/132 (insertion in both SRY copies) (Tab. I). The distribution of these haplotypes does not show any obvious geographical pattern. In contrast, rabbit populations from Northeastern Iberia, that exhibit mtDNA lineage B and correspond to the subspecies *Oryctolagus cuniculus cuniculus*, do not show the occurrence of the 7-bp insertion. In addition, all domestic rabbits originated from various breeds also lacked the insertion, confirming their recent derivation from the single *O. c. cuniculus* subspecies. These data suggest that the 7-bp insertion arose exclusively in a rabbit population from Southwestern Iberia, after the divergence of the *algirus* and *cuniculus* Y chromosomes. An alternative hypothesis would imply a recent SRY duplication occurring only in *algirus* Y chromosomes. However, the observation that some individuals from both subspecies exhibit "heterozygosity" at a few nucleotide positions as well as at linked microsatellites [6] while bearing haplotypes 125/125 or 132/132 clearly favors the first hypothesis (Geraldes *et al.*, unpublished data).

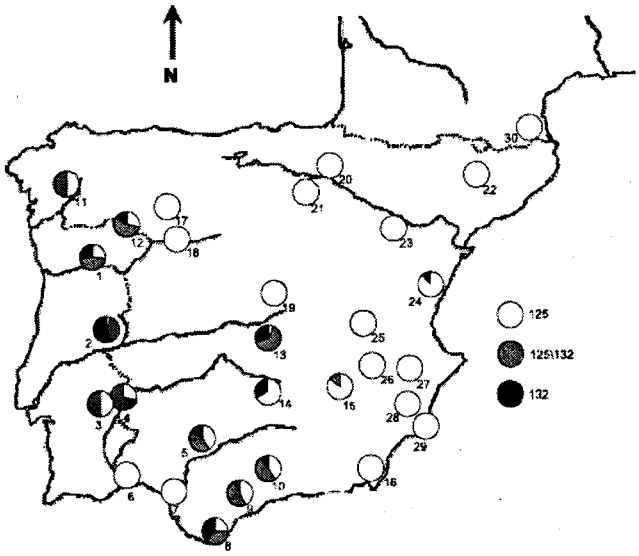


Figure 2- Geographical distribution of the 7-bp polymorphic insertion in wild rabbit populations from the Iberian Peninsula and southern France. Haplotypes 125/125, 125/132 and 132/132 are represented in white, grey and black, respectively. Partitions in the pie charts represent the haplotype frequencies for each population. 1 – Vila Real; 2 – Idanha; 3 – Vila Viçosa; 4 – Elvas; 5 – Sevilla; 6 – Huelva; 7 – Doñana; 8 – Las Lomas; 9 – Fuente Piedra; 10 – Córdoba; 11 – Verin; 12 – Bragança; 13 – Toledo; 14 – Ciudad Real; 15 – Albacete SW; 16 – Las Amoladeras; 17 – Benavente; 18 – Zamora; 19 – Madrid; 20 – Tudela; 21 – La Rioja; 22 – Lleida; 23 – Zaragoza; 24 – Rosell; 25 – Cuenca; 26 – Albacete N; 27 – Valencia; 28 – Alicante; 29 – Cartagena; 30 – Perpignan.

A remarkable pattern within *algirus* Y-chromosomes is the observation in almost all populations of haplotypes 125/125, 125/132 and 132/132. There are at least two possible explanations for this observation: (i) the 7 bp insertion arose independently in the two SRY copies, or (ii) the insertion originated in a single SRY copy and gene conversion events have subsequently been responsible for its transference between copies. While the first hypothesis seems highly unlikely because the independent insertion of two identical motifs is necessarily an extremely rare event, abundant gene conversion between paralogous sequences has recently been described in the Y chromosome of humans and the great apes [14,15]. We thus hypothesize that the 7 bp insertion occurred only once in an *algirus* Y chromosome, 3' to the HMG box of one SRY copy (giving rise to the 125/132 haplotype), and was afterwards copied by gene conversion to the second SRY gene (giving rise to the 132/132 haplotype) (Fig. 3). Given the non-equilibrium status of wild rabbit populations in Iberia due to recent demographic expansions in response to post-glacial climatic amelioration [3], we further suggest that the patterns we observe today for the 7-bp insertion polymorphism in *algirus* populations are the likely result of a balance between drift (that tends to increase differences between populations) and

migration and gene conversion (that tends to homogenize differences between populations).

Table I- Geographical distribution of the 7 bp polymorphic insertion in the rabbit SRY gene observed in wild and domestic rabbits from the Iberian Peninsula and Southern France.

		algirus individuals			*cuniculus* individuals		
		125/125	125/132	132/132	125/125	125/132	132/132
No.*	Population						
1	Vila Real	3	5	3			
2	Idanha	1	7	10			
3	Vila Viçosa	5	5				
4	Elvas	2	5				
5	Sevilla	11	12	4			
6	Huelva	6					
7	Doñana	8					
8	Las Lomas	2	3	3			
9	Fuente Piedra	6	7	1			
10	Córdoba	4	5	1			
11	Verin	3	3				
12	Bragança	1	4	1	1		
13	Toledo	1	16	8	1		
14	Ciudad Real	12		6			
15	Albacete SW	6	1		1		
16	Amoladeras				4		
17	Benavente				21		
18	Zamora				14		
19	Madrid				17		
20	Tudela				8		
21	La Rioja				4		
22	Lleida				5		
23	Zaragoza				9		
24	Rosell	1		1	6		
25	Cuenca	1			4		
26	Albacete N				8		
27	Valencia	4			5		
28	Alicante				18		
29	Cartagena				10		
30	Perpignan				11		
	Domestic				37		

*Population numbers as in Figure 2

The recent availability of the human Y chromosome sequence revealed a number of notable features of which the occurrence of multi-copy testis genes in palindromic sequences that may promote gene conversion and maintain sequence identity is probably the most remarkable. In the near future, comparisons with other Y chromosome sequences (e.g. mice as well as other species) are expected to reveal unprecedented biological insights due to the dramatic differences known to occur in the organization of this chromosome in mammals. The SRY gene may be an example of this because (i) is a single copy gene surrounded by unique sequence in humans, (ii) is a single copy gene flanked by long inverted repeats in mice [7], and (iii) is a multi-copy gene in several African [9] and European rodents [5]. Our data on the rabbit SRY suggests that this sex-determining gene is probably duplicated and

evolving under concerted evolution in Lagomorphs. The recent identification of a clone that contains this gene [8] in a rabbit BAC library [13] will hopefully allow its complete sequencing in our laboratories and the examination of the hypotheses described above.

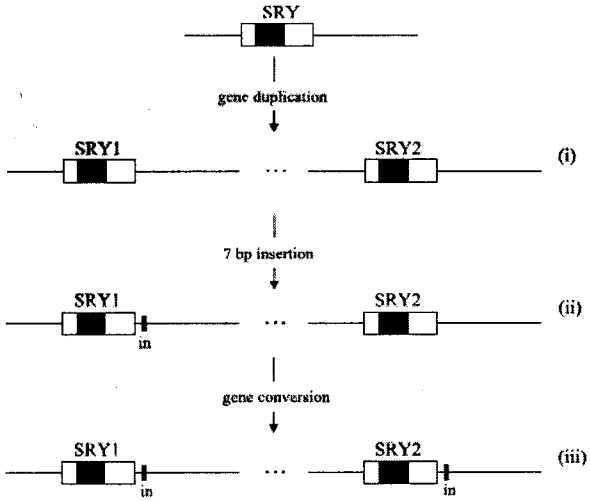


Figure 3- A schematic hypothesis proposed to explain the 7-bp polymorphism at the rabbit *SRY* locus. The coding region of *SRY* is represented by a white box, and the HMG domain is represented by a black box. The 7 bp insertion is located downstream from the coding region and is represented by a black bar and the letters "in". (i) the two *SRY* copies resulted from an ancestral gene duplication; (ii) the GAATTAA motif is inserted in the 3' untranslated region of one *SRY* gene exclusively in *algirus* Y-chromosomes; (iii) this 7-bp motif is transferred to the second *SRY* copy by a mechanism of gene conversion. Note that subsequently (i) can also be derived from (ii) depending on the converted *SRY* copy.

Acknowledgments

This work was partly supported by Fundação para a Ciência e a Tecnologia (SFRH/BD/4621/2001 PhD grant to Armando Geraldes and Research Project POCTI/BSE/40280/2001). We also thank Rafael Villafuerte, Christian Gortazar and José Dávila (IREC, Instituto de Investigación en Recursos Cinegéticos, CSIC/UCLM, España) for their help in Spanish sampling campaigns. Hélène Hayes and two anonymous referees provided constructive criticisms that improved a previous version of this paper.

References

[1] Biju-Duval C., Ennafaa H., Dennebouy N., Monnerot M., Mignotte F., Soriguer R. C., Gaaied A., Hili A., Mounolou J.-C., Mitochondrial DNA evolution in lagomorphs: origin of systematic heteroplasmy and organization of diversity in European rabbits, J. Mol. Evol. 33 (1991) 92-102

[2] Branco M., Ferrand N., Monnerot M., Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene, Heredity 85 (2000) 307-317.

[3] Branco M., Monnerot M., Ferrand N., Templeton A., Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation, Evolution 56 (2002) 792-803.

[4] Callou C., De la garenne au clapier. Étude archéozoologique du lapin en Europe occidentale, Mémoires du Muséum national d'Histoire naturelle, Paris, 2003.

[5] Fernández R., Barragán M.J.L., Bullejos M., Marchal J.A., Martínez S., Guardia R.D., Sánchez A., Mapping the SRY gene in *Microtus cabrerae*: a vole species with multiple SRY copies in males and females, Genome 45 (2002) 600-603.

[6] Geraldes A., Rogel-Gaillard C., Ferrand N., High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) *SRY* gene, Animal Genetics 36 (2005) 349-351.

[7] Gubbay J., Vivian N., Economou A., Jackson D., Goodfellow P., Lovell-Badge R., Inverted repeat structure of the *SRY* locus in mice, Proc. Natl. Acad. Sci. USA 89 (1992) 7953-7957.

[8] Hayes H., Rogel-Gaillard C., Zijlstra C., de Haan N. A., Urien C., Bourgeaux N., Bertaud M., Bosma A. A., Establishment of an R-banded rabbit karyotype nomenclature by FISH localization of 23 chromosome-specific genes on both G- and R- banded chromosomes, Cytogenet. Genome Res. 98 (2002) 199-205.

[9] Lundrigan B.L., Tucker P.K., Evidence for multiple functional copies of the male sex-determining locus, *SRY*, in African murine rodents, J. Mol. Evol. 45 (1997) 60-65.

[10] Monnerot M., Vigne J.-D., Biju-Duval C., Casane D., Callou C., Hardy C., Mougel F., Soriguer R., Dennebouy N., Mounolou J.-C., Rabbit and man: genetic and historic approach, Genet. Sel. Evol. 26 (Suppl 1) (1994) 167-182.

[11] Nagamine C.M., The testis-determining gene, SRY, exists in multiple copies in Old World rodents, Genet. Res. Camb. 64 (1994) 151-159.

[12] Queney G., Ferrand N., Weiss S., Mougel F., Monnerot M., Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). Mol. Biol. Evol. 18 (2001) 2169-2178.

[13] Rogel-Gaillard C., Piumi F., Billault A., Bourgeaux N., Save J. C., Urien C., Salmon J., Chardon P., Construction of a rabbit bacterial artificial chromosome (BAC) library: application to the mapping of the major histocompatibility complex to position 12q.1.1., Mammalian Genome 12 (2001) 253-255.

[14] Rozen S., Skaletsky H., Marszalek J.D., Minx P.J., Cordum H.S., Waterston R.H., Wilson R.K., Page D.P., Abundant gene conversion between arms of palindromes in human and ape Y chromosomes, Nature 423 (2003) 873-876.

[15] Skaletsky H., Kuroda-Kawaguchi T., Minx P.J., Cordum H.S., Hiller L., Brown L.G., Repping S., Pyntikova T., Ali J., Bieri T., Chinwalla A., Delehaunty A., Delehaunty K., Du H., Fewell G., Fulton L., Fulton R., Graves T., Hou S.-F., Latrielle P., Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W., Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K., Schultz B., Strong C., Tin-Wollam A., Yang S.-P., Waterston R.H., Wilson R.K., Rozen S., Page D.P., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes, Nature 423 (2003) 825-837.

3. COMPARING MALE AND FEMALE CONTRIBUTIONS TO DEMOGRAPHIC PROCESSES

3.1 Paper III

In preparation

Sex-specific differences in effective population size and gene flow in subspecies of the European rabbit (*Oryctolagus cuniculus*) inferred from Y chromosome and mtDNA data

A. Geraldes[*,†] and N. Ferrand[*]

[*]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, e Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal. [†] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721, USA

Abstract

In the animal kingdom there are marked behavioral differences between sexes and the impact of these differences in levels and patterns of nucleotide variability remain largely underappreciated. In the European rabbit (*Oryctolagus cuniculus*) females are phylopatric and males undergo natal dispersal. Rabbit social groups are characterized by having strong male hierarchies where the dominant male has preferential access to females. Here, we investigated levels of nucleotide variability at the SRY and Cytb genes to estimate the impact of these sex specific behaviors in the European rabbit. We found that within each subspecies populations are more differentiated for the Cytb gene than for the SRY which is consistent with biased male dispersal. Our estimation of effective population size is two times higher for females than for males of *O. c. cuniculus* and five times higher for females than for males of *O. c. algirus*. Again, this is in agreement with strong male hierarchies in rabbit social groups. Accordingly, the TMRCA of the entire species was about 2 MY for Cytb and only 1 MY for SRY, but the TMRCA of each subspecies was approximately 300,000 years for both markers. Our results imply that great caution should be taken when making population inferences based solely on patterns of nucleotide variability of mtDNA.

Keywords: European Rabbit, Y chromosome, mtDNA, Sex-biased dispersal, Social hierarchies, TMRCA

Introduction

Sex-specific differences in ecology, behavior and migration are abundant in the animal kingdom (Avise 1993). The social structure of a species can have important impacts on its genetic composition. A major impact is on the effective number of breeding individuals. Even with 1:1 sex ratios at birth, the effective population size can be greatly skewed towards one or the other sex depending on the species breeding system. Sex-specific differences in natal dispersal, the permanent movement of a young individual from its natal site to its first breeding site, are commonly observed in mammals, and males are usually the dispersing sex (Pusey 1987). Surprisingly these sex-specific behaviors are still mostly appreciated qualitatively rather than quantitatively.

In molecular ecology studies of most mammalian species, information about males has been mostly neglected, and phylogeographical studies based solely on mtDNA variation are still common. Whenever the male contribution to the evolutionary history of a species is taken into account, it is usually deduced from the differences between patterns of diversity at mtDNA and autosomal markers (*e.g.*, Nyakaana and Arctander 1999; Hooft *et al.* 2003; Cassens *et al.* 2004). But, differences between males and females can be directly assessed by comparing patterns of genetic variation at the maternally inherited mtDNA and the paternally inherited Y chromosome. Studies where both markers were looked at are mainly restricted to humans (*e.g.*, Seielstad *et al.* 1998; Wilder *et al.* 2004 a,b), primates (*e.g.*, Stone *et al.* 2002) and model species such as the house mouse (*e.g.*, Boissinot and Boursot 1997). Recently, the sequencing of regions of the human mitochondria and Y chromosome in the same sample has allowed the direct comparison of male and female contributions to our species evolutionary history in terms of sex-specific migration patterns (Wilder *et al.* 2004a) and effective population size (Wilder *et al.* 2004b). The use of the Y chromosome in population genetics and phylogeographical studies of other mammals has been hampered by several reasons, the most important being the lack of described markers and the reduced levels of nucleotide variability found in the Y chromosome (Petit *et al.* 2002). To our knowledge, only very recently Y chromosome molecular markers have been used in population studies of other mammals (*e.g.*, Feng *et al.* 2001; Hellborg *et al.* 2005).

The European rabbit (*Oryctolagus cuniculus*) is an ideal system to study the effects of sex specific behavior at the molecular level. First it has well characterized social and breeding systems (Myers and Mykytowycz 1958), where a dominant male seems to father most of the progeny (Dally 1981). Second, females are known to be phylopatric, while males undergo natal dispersal (Webb *et al.* 1995; Kunkele and Holst 1996; Surridge *et al.* 1999; Richardson *et al.* 2002). And finally, both maternally and paternally inherited polymorphic markers are available (Gissi *et al.* 1998; Branco *et al.* 2000; Geraldes *et al.* 2005). The origin of this species is in southwestern Europe (Corbet 1994) and until the Middle Ages it was restricted to the Iberian Peninsula and southern France (Callou 2003). In the Iberian Peninsula two deeply divergent mitochondrial (Biju-Duval *et al.* 1991; Branco *et al.* 2000), Y-linked (Geraldes *et al.* 2005) and X-linked (Geraldes *et al.* 2006) genetic lineages can be found. The distribution of these lineages is parapatric (except for two of the four X-linked loci). Lineage A is predominant in southwestern Iberia and is associated with *O. c. algirus*, and Lineage B is mostly found in northeastern Iberia and is associated with *O. c. cuniculus* (Branco *et al.* 2000; Geraldes *et al.* 2006; Geraldes *et al.* evolution). These two lineages coexist in a narrow contact zone that bisects the Iberian Peninsula along a NW-SE axis (Branco *et al.* 2000; Geraldes *et al.* 2006). Branco and colleagues (2000) argued that the divergence observed between the two lineages places their origin in a Pliocene/Pleistocene boundary, prior to the more recent Quaternary glaciations, and the results from X chromosome loci fully corroborate this view (Geraldes *et al.* 2006). They further hypothesized that the two rabbit populations would have started to expand their range from their glacial refugia in the southwest tip of the Iberian Peninsula and the Mediterranean coast of Spain approximately 200,000 years ago after climatic amelioration (Branco *et al.* 2000).

Here, we look at nucleotide variability at the paternally inherited SRY gene and the maternally inherited Cytb gene in, respectively, 107 and 94 male rabbits from the Iberian Peninsula and France, representing the distribution of *O. c. algirus*, *O. c. cuniculus* and their contact zone. With this data set we address the following questions: i) what is the divergence time between the lineages observed at SRY and Cytb?, ii) what is the TMRCA of all the alleles observed in the European rabbit at SRY and Cytb?, iii) what is the TMRCA of each subspecies based on these same markers?, iv) are the effective population sizes of males and females similar?, v) are

the levels of introgression between subspecies similar at both markers?, and iv) are within subspecies levels of gene flow different for the Y chromosome and mtDNA?

Materials and Methods

Sampling

We sampled 103 male rabbits (*Oryctolagus cuniculus*) from 28 populations. Populations were divided into five groups (Table 1): southwest Iberian Peninsula refugium (SWR) and southwest Iberian Peninsula expansion (SWE), both corresponding to *O. c. algirus*; northeast Iberian Peninsula refugium (NER) and northwest Iberian Peninsula expansion and France (NEE), both corresponding to *O. c. cuniculus*; and the contact zone (CZ), following the findings of Branco *et al.* (2000; 2002) and Geraldes *et al.* (evolution) based on mtDNA and Y chromosome variation. The approximate geographic location of the populations is shown in Figure 1. Additionally we sampled six closely related species (Table 1): *Lepus granatensis* (Iberian hare), *Lepus europaeus* (European hare), *Lepus timidus* (Mountain hare), *Lepus alleni* (Antelope jackrabbit), *Lepus californicus* (Black-tailed jackrabbit) and *Sylvilagus spp.* (Cottontail).

Amplification and sequencing

DNA extraction, PCR amplification and sequencing of the SRY gene in the European rabbit samples were performed following Geraldes *et al.* (2005). PCR amplification of the outgroup samples was performed using primers LGSRYF (CAT GCT TTG AGG CAA ATG AAT AAC) and LGSRYR (TTT TGA ACC TTG AAC TTG GCA TC). Amplifications were carried out in 50 µL volumes using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, CA, USA) following manufacturer recommendations. Cycling temperatures were as follows: an initial denaturation step at 94°c for 1′and 20′′ followed by 35 cycles of 94°C for 20′′, annealing at 58°C for 20′′ and extension at 68°C for 2′. PCR products were purified using the QIAquick PCR Purification Kit following conditions recommended by the manufacturer (Qiagen, CA, USA). Sequencing was carried out on an ABI 3100 (Applied Byosystems, CA, USA), using primers SRY1054R (CGT CCA TAT GCA CCT CGC TGC), SRY949F (GAC TAC AAG TAC AGA CCT CG), SRY1516F (AGA CTA AGT TGG TGA TTA GTG) and SRY1991F (AGA TGA AGT AGA AAC TGG

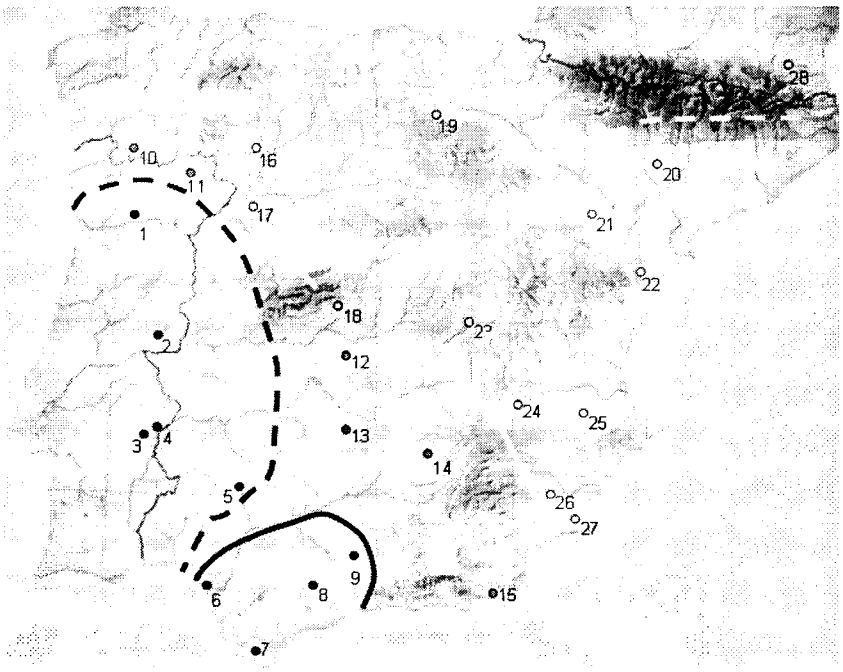


Figure 1 - Geographic location of the European rabbit populations sampled. Black dots indicate SW populations (*O. c. algirus*), white dots indicate NE populations (*O. c. cuniculus*) and grey dots indicate CZ populations. Dashed lines delimit the populations that belong to the areas of recent expansion of each subspecies and solid lines the inferred refugial areas (Branco *et al.* 2002). 1- Vila Real, 2- Idanha-a-nova, 3-Vila Viçosa, 4- Elvas, 5- Sevilla, 6- Doñana, 7- Las Lomas, 8- Fuente Piedra, 9- Córdoba, 10- Verín, 11- Bragança, 12- Toledo, 13- Ciudad Real, 14- Albacete SW, 15- Las Amoladeras, 16- Benavente, 17- Zamora, 18- Madrid, 19- La Rioja, 20- Lerida, 21- Zaragoza, 22- Castelló, 23- Cuenca, 24- Albacete NE, 25- Valencia, 26- Alicante, 27- Cartagena and 28- Perpignan.

CAG). A short fragment of the mitochondrial Cytb gene was also sequenced in 94 of the European rabbit samples and in the outgroup (Table 1). PCR amplification was performed using primers CYTB14167F (AGA ACC TAA TGA CCA ACA TTC G) and CYTB14858R (GTC TTT GAT TGT GTA GTA GGG GTG G). Amplifications were carried out in 50 µL volumes using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, CA, USA), following manufacturer recommendations. Cycling temperatures were as follows: an initial denaturation step at 94ºC for 1′and 20′′ followed by 35 cycles of 94ºC for 20′′, annealing at 56ºC for 20′′ and extension at 68ºC for 1′. PCR cleaning and sequencing followed the protocol used for the SRY fragment, but using primer CYTB14858R.

Data analyses

Alignments and phylogenetic relationship between alleles

Sequences were visually inspected and concatenated using the computer program Sequencher (Gene Codes, MI, US). Samples were aligned to each other

manually using the BioEdit software (Hall 1999). Four previously published European rabbit SRY sequences (Geraldes *et al.* 2005) were also used. The phylogenetic relationships between all European rabbit samples at SRY and Cytb were calculated using the Median Joining (MJ) algorithm (Bandelt *et al.* 1999) as implemented in Network v4.1.0.8 (http://www.fluxus-technology.com/). Sites segregating for three nucleotides were excluded, as well as one 7 bp insertion found at SRY as it has been discussed in greater detail elsewhere (Geraldes and Ferrand 2006).

Table 1 - Population localities and samples used in this study

Species	Population number[a]	Sry		Cyt*b*	
		n	Sample names	n	Sample names
Oryctolagus cuniculus	1	4	Vrl1, Vrl4, Vrl7, Vrl 10	4	Vrl1, Vrl4, Vrl7, Vrl10
Oryctolagus cuniculus	2	4	Id73, Id74, Id84, Id85	4	Id73, Id74, Id84, Id85
Oryctolagus cuniculus	3	3	VV1_1/94, VV1_4/94, VV5_9/93	2	VV1_1/94, VV5_9/93
Oryctolagus cuniculus	4	4	Elv3, Elv5, Elv6, Elv9	4	Elv3, Elv5, Elv6, Elv9
Oryctolagus cuniculus	5	4	Pfr1, Pfr5, Pfr6, Pfr7	4	Pfr1, Pfr5, Pfr6, Pfr7
Oryctolagus cuniculus	6	4	Don2, Don6, Don5822, Don6170	1	Don6
Oryctolagus cuniculus	7	4	Lom3[b], Lom8, Lom13, Lom18	4	Lom3, Lom8, Lom13, Lom18
Oryctolagus cuniculus	8	2	Ftp1, Ftp38	1	Ftp38
Oryctolagus cuniculus	9	4	Luc2, Luc4, Luc9[b], Luc17	4	Luc2, Luc4, Luc9, Luc17
Oryctolagus cuniculus	10	4	Gal25c1, Gal25c3, Gal25c7, Gal25c8	2	Gal25c3, Gal25c7
Oryctolagus cuniculus	11	4	Bra1, Bra10, Bra11, Bra13	2	Bra1, Bra13
Oryctolagus cuniculus	12	5	Tol22, Tol25, Tol50, Tol129, Tol162	5	Tol22, Tol25, Tol50, Tol129, Tol162
Oryctolagus cuniculus	13	4	Cre1, Cre7, Vdm12, Vdm27	3	Cre7, Vdm12, Vdm27
Oryctolagus cuniculus	14	5	Pvd3, Pvd4, Pvd5, Pvd6, Pvd7	5	Pvd3, Pvd4, Pvd5, Pvd6, Pvd7
Oryctolagus cuniculus	15	1	Amo2	1	Amo2
Oryctolagus cuniculus	16	5	Bnv3, Bnv14, Bnv22, Bnv25, Bnv33	5	Bnv3, Bnv14, Bnv22, Bnv25, Bnv33
Oryctolagus cuniculus	17	4	Zam1, Zam2, Zam20, Zam21	4	Zam1, Zam2, Zam20, Zam21
Oryctolagus cuniculus	18	4	Mdr2, Mdr4, Mdr6, Mdr7	4	Mdr2, Mdr4, Mdr6, Mdr7
Oryctolagus cuniculus	19	4	Lrj2, Lrj3, Lrj4, Lrj6	3	Lrj2, Lrj4, Lrj6
Oryctolagus cuniculus	20	1	Ler16	0	
Oryctolagus cuniculus	21	4	Zrg16[b], Zrg18, Zrg19, Zrg20	4	Zrg16, Zrg18, Zrg19, Zrg20
Oryctolagus cuniculus	22	4	Rsl2, Rsl4, Rsl9, Rsl10	4	Rsl2, Rsl4, Rsl9, Rsl10
Oryctolagus cuniculus	23	5	Cue1, Cue2, Cue3, Cue4, Cue5	5	Cue1, Cue2, Cue3, Cue4, Cue5
Oryctolagus cuniculus	24	4	Mdg1, Mdg2, Mdg3, Mdg5	4	Mdg1, Mdg2, Mdg3, Mdg5
Oryctolagus cuniculus	25	5	Val3, Val4, Val9, Val16, Val18	5	Val3, Val4, Val9, Val16, Val18
Oryctolagus cuniculus	26	6	Alt1, Alt15, Alt103, Alt107, Alt120[b], Alt124	5	Alt15, Alt103, Alt107, Alt120, Alt124
Oryctolagus cuniculus	27	4	Cat3, Cat12, Cat13, Cat14	4	Cat3, Cat12, Cat13, Cat14
Oryctolagus cuniculus	28	1	Pep18	1	Pep18
Lepus granatensis		1	LgMad1	1	LgMad1
Lepus europaeus		1	LeVille28	1	LeVille28
Lepus timidus		1	Ltural	1	Ltural
Lepus alleni		1	Kl	1	Kl
Lepus californicus		1	Aj4	1	Aj4
Sylvilagus spp.		1	Ac4	1	Ac4

[a] Population numbers and groups from Figure 1. SWR - Populations from south-west Iberian Peninsula refugium and SWE - Populations from south-west Iberian Peninsula expansion, corresponding to the distribution of *Oryctolagus cuniculus algirus*. NER - Populations from north-east Iberian Peninsula refugium and NEE - Populations from north-east Iberian Peninsula expansion and France, corresponding to the distribution of *Oryctolagus cuniculus cuniculus*. CZ - Populations from the contact zone as previously defined by mtDNA variation (Branco *et al.* 2000; Branco *et al.* 2002). O-Outgroups.

[b] Sequences from Geraldes *et al.* 2005

Polymorphism

The number of polymorphic sites (s), number of haplotypes (H), and two estimates of the population mutation parameter, π (Nei and Li 1979) and θ_w (Watterson 1975) were calculated using DnaSP 4.10.3 (Rozas *et al.* 2003). Nucleotide

diversity, π, is based on the average number of nucleotide differences between two sequences randomly drawn from a sample, and θ_w, is based on the proportion of segregating sites in a sample. Under equilibrium conditions with respect to mutation and drift, both estimate the neutral parameter $\theta = 4Ne\mu$ for autosomal loci, where Ne is the effective population size and μ is the neutral mutation rate. Since the effective population size of Y-linked and mtDNA loci is ¼ the effective population size of autosomes, both π and θ_w are estimators of $Ne\mu$.

Tests of Neutrality

Tajima's D (Tajima 1989), Fu and Li's D (Fu and Li 1993) and Fu's Fs (Fu 1997) statistics were calculated to test for deviations from a neutral site frequency distribution using DnaSP 4.10.3 (Rozas *et al.* 2003). Tajima's D compares the average number of nucleotide differences between two randomly drawn sequences from a sample (π), with the proportion of polymorphic sites (θ_w), Fu and Li's D statistic is based on the difference between the number of singletons in a sample and the total number of mutations and finally Fu's Fs is based on the haplotype frequency distribution conditioned on the value of θ. Ratios of polymorphism to divergence at SRY and Cytb were compared with the expectations under a neutral model using the HKA test (Hudson *et al.* 1987) with the HKA software (Hey and Kliman 1993). The ratio of nonsynonymous polymorphisms to divergence was compared to the ratio of synonymous polymorphisms to divergence using the MK test (McDonald and Kreitman 1991) in DnaSP 4.10.3 (Rozas *et al.* 2003). If SRY and Cytb in the European rabbit are evolving under neutrality, then these ratios should be similar. For the MK and HKA tests divergence was measured to *Lepus granatensis*. Additionally the HKA test was also performed between subspecies of *Oryctolagus cuniculus*.

Sex-specific patterns of gene flow

Patterns of gene flow were inspected at different levels. First we calculated Fst (Wright 1951), between SW and NE populations, to asses the degree of mitochondrial and Y chromosome introgression between subspecies of rabbit. To access patterns of gene flow and genetic differentiation within each subspecies we calculated Fst between SWR and SWE (*O. c. algirus*) and between NER and NEE (*O. c. cuniculus*) for SRY and Cytb. For these analyses we excluded individuals with introgressed Y chromosomes or mitochondria from the other subspecies. Fst was calculated using the

program DnaSP 4.10.3 (Rozas *et al.* 2003) following the method of Hudson *et al.* (1992). When the number of populations studied is high and the sampling size in each population is low, the statistic Snn (Hudson 2000) was shown to have the most power to detect genetic differentiation between populations. To avoid the effects of pooling populations into population groups, we also estimated genetic differentiation between all populations from each subspecies with the statistic Snn using DnaSP 4.10.3 (Rozas *et al.* 2003). Four populations that only had one sample each were excluded. Significance was accessed with 1000 permutation tests.

Sex-specific effective population sizes and TMRCA

The estimation of crucial population parameters as the Time to the Most Recent Common Ancestor (TMRCA) in years, and the effective population size (N_e) requires the use of locus-specific mutation rates. Due to the high polymorphism observed at Cytb and the deep divergence to the outgroups used in this study, high levels of homoplasy are expected to occur at this locus. The Akaike Information Criterion in Modeltest 3.06 (Posada and Crandall 1998; Posada and Buckley 2004) was used to select an appropriate model of nucleotide substitution for each locus. For this estimation, we used all the European rabbit haplotypes and all the outgroup samples. Maximum Likelihood (ML) net nucleotide distances between *L. granatensis* and *O. cuniculus,* and between the two lineages found in *O. cuniculus,* were calculated using the estimated model of substitution using PAUP v 4.0 (Swofford 2002). The neutral mutation rate for each gene was calculated assuming a divergence time between *L. granatensis* and *O. cuniculus* of 11.8 million years (My) (Matthee *et al.* 2004). We used the mutation rates for each gene to estimate the time since the split of the two lineages observed at each locus.

Additionally, ML estimates of TMRCA and sex-specific effective population sizes were generated under a coalescent framework using GENETREE version 9.0 (R.C. Grifiths http://www.stats.ox.ac.uk/~griff/software.html). An infinite sites model is assumed by the program. For the SRY gene there was no evidence of either recombination or homoplasy. For the Cytb data set, there were several pairs of sites showing all four gametic types most likely due to recurrent mutations. The methodology used to obtain a nonreticulating tree for this locus is presented in the results section. We first estimated the TMRCA of all rabbit samples for each locus. The null demographic model incorporated in GENETREE is one of panmixia. Since

it is apparent from the observed genealogies of SRY and Cytb that this is not the case in *O. cuniculus* we also performed GENETREE analyses on each of the observed lineages independently, because these probably conform better to the model used by GENETREE. This analysis requires the estimation of a single population parameter θ_{ml} ($2N_e\mu$ where N_e is the sex specific effective population size and μ is the neutral mutation rate). We estimated θ_{ml} by generating a single-likelihood curve over a wide range of possible values. GENETREE also allows for a population exponential growth model. This model requires the estimation of an extra parameter, β, the intrinsic population growth parameter:

$$N(t) = N_0 \exp(-\beta t/N_0)$$

where N_0 represents the present day population size and N(t) represents the population size t generations in the past. Likelihood estimates of multiple parameters are not independent using this coalescent approach, they have to be estimated jointly. We iteratively estimated the likelihood of these parameters across a narrow range of values until a local maximum was reached. The fit of our data to a constant population size model or an exponential growth model was assessed by a likelihood ratio test of the trees obtained with both models. The TMRCA for each locus and subspecies was estimated conditioned on the obtained values of θ_{ml} (when a constant population size model could not be rejected) or θ_{ml} and β_{ml} (when a constant population size model was rejected).

Results

Sequence alignments

We examined 107 and 94 European male rabbits for nucleotide diversity at SRY and Cytb genes respectively. For the SRY gene the alignment of all the European rabbit sequences has 2014 bp (from position 318 to 2331 of the reference sequence Gen Bank accession number AY785433) after a polymorphic 7 bp insertion was removed. When the six outgroup sequences were considered and all indel polymorphisms were removed the alignment is reduced to 1791 bp (nucleotides 368 to 2173 of the reference sequence GenBank accession number AY785433). For the Cytb no indels were observed, but two sites (nucleotides 14282 and 14683 of the complete mitochondrial sequence GenBank accession number AJ001588) segregating

for three nucleotides within the European rabbit samples were removed from the alignments. The total size of the aligned region in the European rabbit and outgroups is 616bp (from nucleotides 14181 to 14798 of the complete mitochondrial sequence GenBank accession number AJ001588). All Y chromosome and mtDNA sequences were deposited in GenBank (Accession nos. XXXXXXX to XXXXXXX).

Levels of polymorphism

When observing the chromatograms for SRY, we detected that several individuals had two different nucleotides at the same site (similar to a heterozygote individual at an autosomal locus). Sequencing of European rabbit BAC clone 828D7 containing the SRY gene revealed that this locus is duplicated in the arms of a palindrome (A. Geraldes unpublished results). The arms of the palindrome are at least 23 Kb long and are separated by a spacer region of approx. 19 Kb. The two copies are 99.96% identical (99.94% if indel polymorphisms are included). Mapping of this clone by Fluorescent in situ Hybridization (FISH) revealed that this clone is homologous only to the Y chromosome (Hayes *et al.* 2002). This observation coupled with the fact that no amplification was observed in female rabbits assures us that the observed heterozygote positions are due to the fact that we are amplifying two copies (hereafter referred to as SRY1 and SRY2) of the SRY gene and that both are located in the Y chromosome and are male specific. Since in our sequencing of the SRY locus we could not separate the two copies, we decided to consider not 2014 bp but 4028 bp and every time we had a mutation in "homozygous state", we introduced it in the two copies (two mutations in 4028 bp is the same as one mutation in 2014 bp) and whenever a mutation was present in "heterozygous state" we considered it only once (one mutation in 4028 bp is the same as half mutation in 2014 bp). All analyses hereafter were performed considering the two copies of SRY. For GENETREE analyses all "heterozygote" haplotypes were removed and only one copy of SRY was analyzed.

Table 2 shows all the polymorphic positions found at SRY in the European rabbit and the populations at which each of the resulting haplotypes were found. We found 18 nucleotide positions that were variable, 11 were present in both copies (*i.e.*, homozygous in every haplotype), 4 were only found in one of the gene copies (*i.e.*, always in an heterozygote sate) and three were found both in heterozygosity and homozygosity. In the concatenated sequence of SRY1 and SRY2 a total of 32

Table 2 -Polymorphic sites at both copies of *Sry* and geographic distribution of haplotypes

	Sry 1														*Sry* 2																SW populations									CZ populations						NEIP populations																
									1	1	1	1	2	2	2									1	1	1	1	2	2	2	2	2	2																																				
	3	3	4	5	5	6	9	2	5	6	7	0	0	1	3	3	4	5	5	6	7	9	2	5	6	7	0	0	1	1	2	2																																
	5	6	9	9	9	0	9	3	9	5	0	7	9	5	5	6	9	9	9	0	6	9	3	9	5	0	7	9	1	5	2	4																																
Haplotype	8	6	3	6	7	3	2	3	4	2	8	2	8	3	5	6	3	6	7	3	8	2	3	4	2	8	2	8	6	3	4	9	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	n
Outgroup	A	A	G	T	C	C	A	G	A	C	A	C	A	A	A	A	G	T	C	C	C	A	G	A	C	A	C	A	C	A	T	A																														
YA1	.	.	.	A	A	.	.	A	.	.	.	.	.	G	.	.	.	A	A	.	.	.	A	.	.	.	.	.	.	G	.	.				1	1	4	1			3			1	3									1		1				16
YA2	.	.	.	A	A	.	.	A	.	.	.	.	.	G	.	.	.	A	A	.	T	.	A	.	.	.	.	.	.	G	.	.	1																													1
YA3	.	.	.	A	A	.	.	A	.	.	.	.	.	G	.	.	.	A	A	.	.	.	A	.	.	.	.	.	T	G	.	.										1																			1
YA4	.	.	.	A	A	.	.	A	.	.	.	.	.	G	.	.	.	A	A	.	.	.	A	C	.	.	.	.	.	G	.	.	1		1		1		1																						4
YA5	.	.	.	A	A	.	.	A	C	.	.	.	.	G	.	.	.	A	A	.	.	.	A	C	.	.	.	.	.	G	.	.	2	4	2	2	2		2	2	4	1	2	2	1	1													27
YA6	.	.	.	A	A	.	.	A	C	.	.	.	.	G	.	.	.	A	A	T	.	.	A	C	.	.	.	.	.	G	.	.												2																	2
YA7	.	.	.	A	A	T	.	A	C	.	.	.	.	G	.	.	.	A	A	T	.	.	A	C	.	.	.	.	.	G	.	.											2																	2
YA8	T	.	.	A	A	.	T	A	C	.	.	.	.	G	T	.	.	A	A	.	.	T	A	C	.	.	.	.	.	G	.	.				1																									1
YB1	.	G	C	.	.	.	.	.	G	.	.	T	.	.	.	G	C	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.																													1
YB2	.	.	C	.	.	.	.	.	G	.	.	T	.	.	.	.	C	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.											1					4		3		4		1				1	14
YB3	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.														1							1	1			1		4
YB4	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.																4		2			2	2	3	1	1			15
YB5	.	.	C	.	.	.	.	.	.	.	G	.	T	.	.	.	C	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.											1					1		2	1		1		1	3	3	1		14
YB6	.	.	C	.	.	.	.	.	.	.	G	G	T	.	.	.	C	.	.	.	.	.	.	.	.	G	G	T	.	.	.	.																										1		2
YB7	.	.	C	.	.	.	.	.	.	.	G	.	T	.	.	.	C	.	.	.	.	.	.	.	.	G	.	T	.	.	C	C																									2	1		3
n																																	4	4	3	4	4	4	4	2	4	4	4	5	4	5	1	5	4	4	4	1	4	4	5	4	5	6	4	1	

Population numbers are from Figure 1 and Table 1.

polymorphic positions were found defining a total of 15 haplotypes. Twelve of the

observed mutations separate two groups of haplotypes thus defining two genetic lineages. Lineage A (eight haplotypes) is found in all *algirus* populations and all CZ populations but population 15 from which only one individual was surveyed. Lineage A was also detected in two individuals from *cuniculus* populations 23 and 25. In contrast Lineage B (seven haplotypes) is absent from *algirus* populations, from CZ populations 10 and 13 and is present in all *cuniculus* populations (Figure 2).

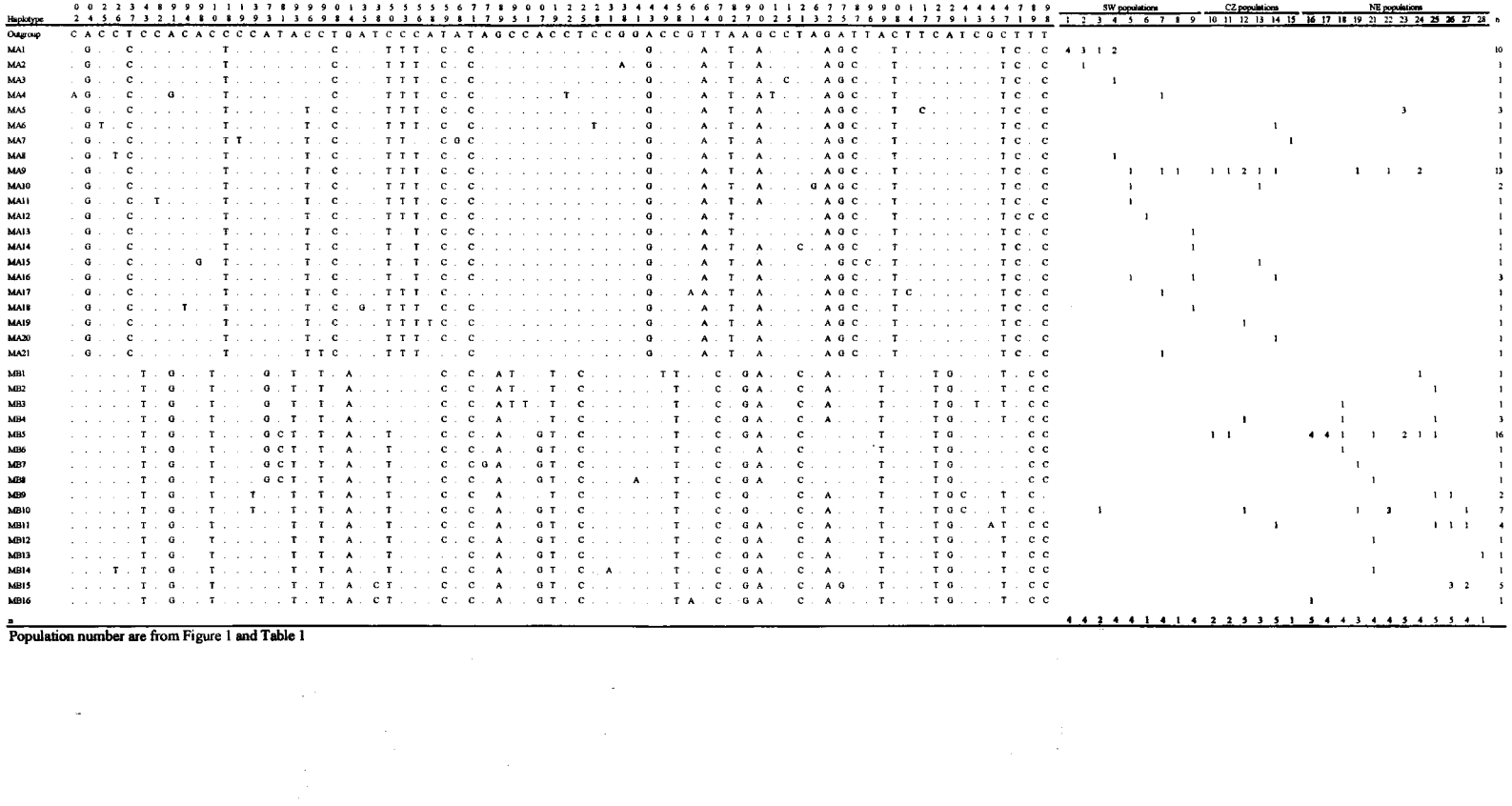
At Cytb we found a total of 74 polymorphic positions (37 haplotypes), 22 of which separate Lineages A and B (Table 3). Lineage A was found in all *algirus* and CZ populations and also in *cuniculus* populations 19, 22, 23 and 24. Conversely, Lineage B was found in all *cuniculus* populations, all CZ populations but population 15 (only one sample was sequenced from this population) and *algirus* population 3 (Figure 2). At both markers very little introgression is found outside of the contact zone.

At SRY, when all the European rabbit samples are considered, π=0.213% while at Cytb π=3.272% (Table 4). Within each of the two genetic lineages nucleotide polymorphism is much lower than in the total sample, at SRY π=0.034% and 0.055% at Lineages A and B respectively. At Cytb a similar pattern is observed with polymorphism being twice as high in Lineage B (π=0.336% and 0.881% at lineages A and B respectively). When comparing the two estimators of the population mutation parameter θ, θ_w is lower than π when the entire data set is considered, and the opposite is true when we divide the data into the two lineages observed. Also, θ_w is higher for lineage A of Cytb than for lineage B, which is the opposite of what was found for π.

Tests of Neutrality

From the neutrality tests based on the neutral frequency spectrum of mutations, significant deviations from the expectations under a neutral model of molecular evolution were only detected at Cytb Lineage A (*O. c. algirus*) (Table 4). This deviation was significant for the three tests performed (Tajima's D, Fu and Li's D and Fu's Fs). Although Lineage A at SRY also showed consistently negative values for these tests, statistical significance was never reached. To see weather the test was not significant at SRY just because there were less polymorphic positions, we also calculated $D/|D_{min}|$. When Tajima's D is divided by its minimum value for a given

Table 3- Polymorphic sites at Cytb and geographic distribution of the observed haplotypes



Population number are from Figure 1 and Table 1

number of segregating sites ($|D_{min}|$), we have a measure of the skew from the neutral frequency spectrum (Schaeffer 2002). $D/|D_{min}|$ is more extreme at Cytb than at SRY (results not shown), meaning that the skew is greater at Cytb and the difference observed between markers is probably not due to the power of the test, but to a more pronounced female population expansion.

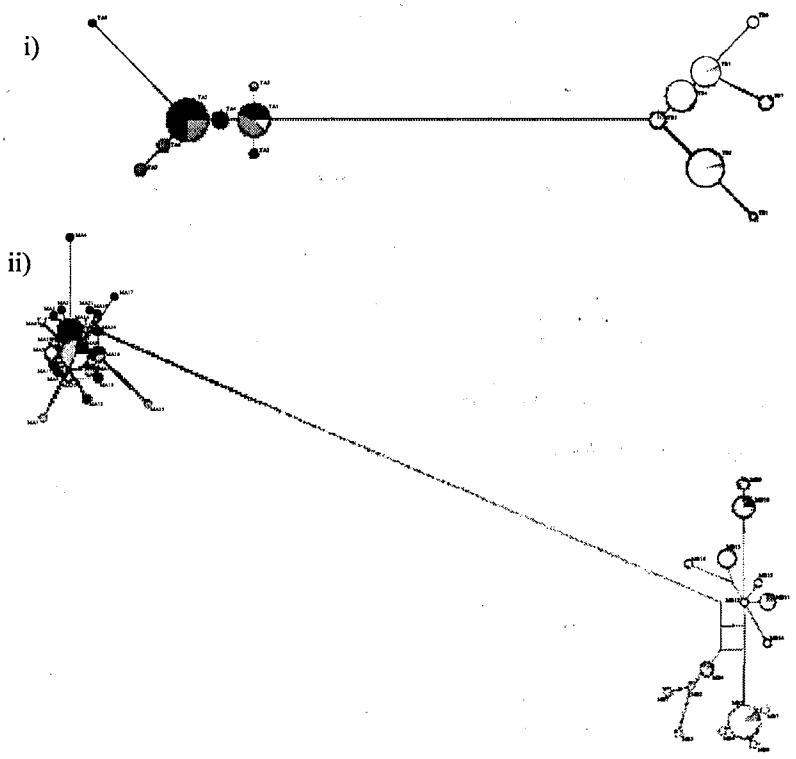


Figure 2 - MJ (Medium Joining) haplotype networks. i) *Sry* gene and ii) Cyt*b* gene. Each haplotype is represented by a circle. The diameter of the circle is proportional to the frequency of each haplotype in the sample. Each haplotype is a pie chart indicating the percentage of individuals from each of the three population groups that has the given haplotype. SW individuals are represented in black, CZ individuals in grey and NE individuals in white. Haplotype names in i) and ii) are from Tables 2 and 3.

The HKA test compares levels of polymorphism to divergence at two or more loci. First, levels of polymorphism at SRY and Cytb were compared to levels of divergence to *Lepus granatensis* (Table 5). A significant deviation ($P<0.001$) from neutral expectations was observed. Second, we performed the same test, but comparing levels of polymorphism within each subspecies to divergence between subspecies. When this comparison is done, no deviation from neutral expectations is observed. Finally, the MK test compares levels of polymorphism and divergence at synonymous and nonsynonymous sites. At SRY we observe more replacement than

silent substitutions fixed between species while the opposite is true for polymorphic sites in rabbit, but this deviation was statistically different from neutral expectations (Table 6). Significance was assessed with G test after Yate's correction.

Table 4 – Levels of nucleotide polymorphism and tests of the neutral model of molecular evolution based on the frequency spectrum of mutations and the number of haplotypes in a sample.

					Sry			
Population	n	s	H	π (%)	θW (%)	TD	FLD	Fs
O. cuniculus	107	32	15	0,213	0,151	1,241	-0,594	4,521
O. c. algirus	54	10	8	0,034	0,054	-1,047	-2,351	-1,633
O. c. cuniculus	53	10	7	0,055	0,055	0,003	0,153	0,758
					Cyt b			
Population	n	s	H	π (%)	θW (%)	TD	FLD	Fs
O. cuniculus	94	74	37	3,272	2,341	1,183	-1,535	-0,493
O. c. algirus	47	32	21	0,336	1,172	-2,41**	-4,868**	-16,886***
O. c. cuniculus	47	26	16	0,881	0,953	-0,248	-1,469	-1,662

Sex-specific gene flow

We calculated Fst between *algirus* and *cuniculus* populations to assess the degree of introgression between subspecies of the European rabbit. Both markers indicate significant ($p<0.001$) levels of genetic differentiation between subspecies but Fst was higher (0.89) at SRY than at Cytb (0.71).

Table 5 - HKA test between *Sry* and Cyt*b*

	chi-square	HKA statistic
O. cuniculus/L. granatensis	0.0002	0.0003
O. c. cuniculus/O. c. algirus	0.2446	0.2129

We measured gene flow within each of the subspecies of *O. cuniculus* by calculating Fst values between refugia and expansion areas (Branco *et al.* 2002) of each subspecies. Results are shown in table 7. In *O. c. algirus* the Fst value between these two areas is 0.01 at SRY and 0.31 at Cytb. In *O. c. cuniculus* Fst at SRY is 0.06 and 0.29 at Cytb. At both subspecies there is significant differentiation between the populations that are distributed in the glacial refugium areas and the recently colonized areas for Cytb but not for SRY. Fst is, in both subspecies, an order of magnitude higher at Cytb than at SRY.

Because the pooling of different populations into population groups can introduce a bias in these analyses, we also estimated genetic differentiation between all pairs of populations within each subspecies using the test statistic Snn (Hudson 2000). These analyses were concordant with Fst in showing at both subspecies more

genetic differentiation at Cytb than at SRY, but the magnitude of this difference was much lower. Snn between *algirus* populations was 0.12 at SRY and 0.25 at Cytb, and Snn between *cuniculus* populations was 0.14 at SRY and 0.17 at Cytb ($p<0.05$ in all tests).

Table 6 – MK test comparing levels of polymorphism within *O. cuniculus* to divergence to *L. granatensis* for silent and replacement changes

	Sry		*Cyt* b	
	Polymorphism	Divergence	Polymorphism	Divergence
Synonymous	3	23	64	67
Replacement	2	27	9	2

Mutation rates and divergence times

For each locus, we estimated maximum likelihood models of nucleotide substitutions to correct for multiple substitutions using MODELTEST 3.06 (Posada and Crandall 1998) with the Akaike Information Criterion (Posada and Buckley, 2004). A TVM model of nucleotide substitution was found to better explain the evolutionary dynamics at each locus (with Gamma distribution for Cytb). Maximum likelihood net nucleotide divergences between *L. granatensis* and *O. cuniculus* were used to estimate locus specific substitution rates, assuming a divergence time of 11.8 My (Matthee *et al.* 2004) between these species (Table 8). Based on these calculations we estimated the mutation rate per site per year to be $1.92X10^{-9}$ for SRY, while for Cytb it was an order of magnitude higher, $1.42X10^{-8}$. Using these estimated mutation rates and the net nucleotide divergence between lineages of rabbits at each locus, a divergence time of 0.95 My for SRY and 2.45 My for Cytb is obtained.

Table 7 - Fst values between population groups of *O. c. algirus* and *O. c. cuniculus* at Sry and Cyt*b*

	Fst	
	Sry	*Cyt* b
Between *O. c. algirus* and *O. c. cuniculus*	0.89***	0.71***
Between refugium and expansion areas of *O. c. algirus*	0.02	0.31***
Between refugium and expansion areas of *O. c. cuniculus*	0.06	0.33***

*Details on population groups are given in Materials and Methods.

Effective population size and TMRCA

Nonreticulating gene trees for the SRY and Cytb genes are shown in figure 3. For SRY we used information from one gene copy only. We did this by removing all individuals presenting "heterozygous" positions leaving 13 polymorphic sites and 9 haplotypes in a total of 81 individuals for analyses. For the Cytb data set there were

190 pairs of sites showing four gametic types. These were mostly removed by eliminating individuals from the analyses. For two sites, several haplotypes with intermediate frequencies would have to be removed from the analyses to make it compatible with an infinite sites model. In these two cases recurrent mutations were treated as separate sites following Wilder *et al.* (2004*b*). From the initial 74 segregating sites and 37 haplotypes, 67 segregating sites and 26 haplotypes were used in a total of 76 individuals. The ancestral state for each SRY polymorphism was unambiguously determined by comparison with *Lepus granatensis*. For Cytb, this was often impossible because *Lepus granatenis* showed a third nucleotide segregating at that position. Using other species as outgroups did not eliminate the problem. In these cases, whenever a position was polymorphic in only one of the lineages, the base present in the other lineage was taken to be the ancestral state. For the few sites for which the ancestral state could not be determined by any of these methods, one of the states was chosen by chance.

Table 8 – Maximum Likelihood models of substitution for *Sry* and *Cyt* b, neutral mutation rates per site per year and net nucleotide divergences (%) between *O. cuniculus* and *L. granatensis* and between subspecies of rabbits and, divergence time in millions of Years (My) between subspecies of rabbit.

		O. cuniculus/L. granatensis		*O. c. algirus/O. c. cuniculus*	
Locus	Model	% Divergence	Mutation rate	% Divergence	Divergence Time (My)
Sry	TVM[1]	4.531	$1.92X10^{-8}$	0.365	0.95
Cyt b	TVM+G[2]	33.428	$1.42X10^{-9}$	6.927	2.45

[1]TVM model with estimated α parameter describing the gamma distribution of 0.2213.

Maximum likelihood estimates of population parameters were obtained with GENETREE. We first estimated the value of θ_{ml} of the entire data set for each locus, assuming a constant population size model. We then estimated β_{ml} for these data sets to see if a model with exponential grow would conform better with the data but a value of β_{ml} close to zero was obtained in each case and no significant improvement of the likelihood of the trees was observed. We then estimated the N_e and TMRCA of each locus for a constant sized population model (Table 9). Our results indicate that the coalescence time for the Y chromosome is approximately 880,000 years (range from 634,000 to 1,126,000 years) and this is roughly half the coalescence time of the mitochondria, which is approximately 1,886,000 years (range from 1,742,000 to 2,030,000 years). Also our estimates indicate that have an effective population size half that of females (Ne for Y chromosome is 323,000 and for the mitochondria is 760,000).

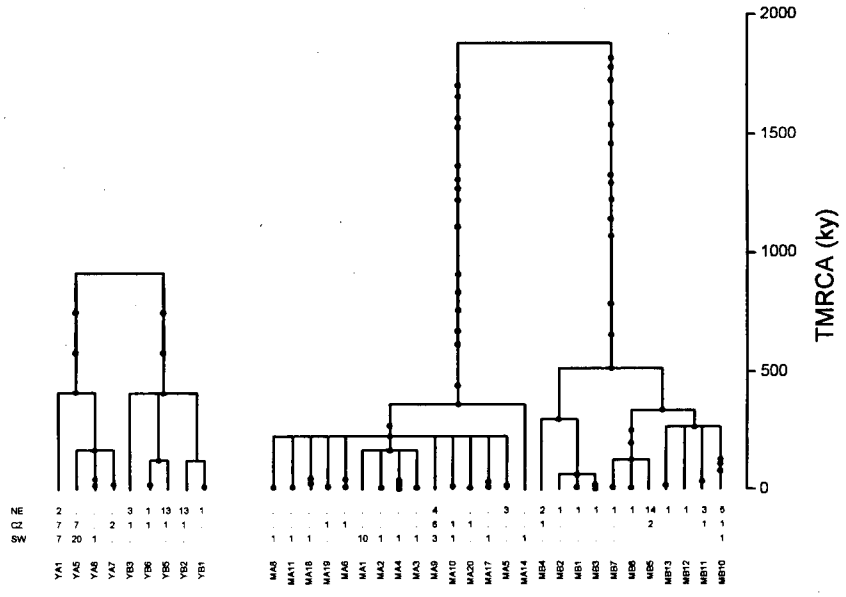


Figure 3 – Non reticulating genealogies for *Sry* (left) and *Cytb* (right). Dots indicate polymorphic positions. The distribution of each haplotype in each population group and the haplotype name from Tables 2 and 3 are indicated at the bottom of each tree. Trees are scaled according to their TMRCAs of the total tree. Note that the reported TMRCA for each subspecies was calculated in separate analyses where only the haplotypes from the corresponding subspecies were included. Thus trees are scaled with respect to the TMRCA of the entire sample only and not of each subspecies.

From the observed gene trees (Figure 3) it is obvious that our data is not compatible with a single panmitic population. This violates the population model used by GENETREE. In order to more reliably estimate N_e and the TMRCA the same analysis was applied to each subspecies independently. First we analyzed our data assuming a constant sized population model. These estimates were compared to those obtained when a model that allows for population size changes is used (by jointly estimating θ_{ml} and β_{ml}) by performing a likelihood ratio test. Our results show that for *O. c. algirus* the observed genealogies for SRY and Cytb are better explained when a model allowing for population size growth is used. On the other hand, a constant population size model could not be rejected for *O. c. cuniculus*. Results are shown in table 9. Similarly to what we estimated for the entire data set, male population sizes seem to be lower for males than for females in each subspecies (for *algirus* populations N_e is 244,000 for the Y chromosome and 1,220,000 for the mitochondria and for *cuniculus* populations N_e is 194,000 for the Y and 376,000 for the mitochondria). Coalescence times are close for the Y chromosome and mtDNA in both subspecies and their ranges largely overlap. We estimated a TMRCA for *algirus* populations of 323,000 years (range from 280,000 to 366,000 years) for Cytb and

250,000 years (range from 173,000 to 327,000) for the SRY gene. For *cuniculus* populations the TMRCA of Cytb is 383,000 years (range from 294,000 to 472,000 years) and the TMRCA for SRY is 342,000 years (range from 183,000 to 465,000 years).

Table 9 – Population parameters estimated using GENETREE for constant size (upper row for each data set) versus exponential growth (lower row) demographic models

Locus	Data set	n[1]	s[2]	θ_{ml}	β_{ml}	N_e or N_0 (10^{-3})[3]	TMRCA (SD) in kya	Likelihood Score (SE)	LRS
Cyt b	*O. cuniculus*	76	67	13.3	-	760	1886 (144)	$3.31X10^{-35}$ $(2.33X10^{-35})$	
	O. c. algirus	47	30	11.3	-	644	361 (66)	$2.15X10^{-13}$ $(1.06X10^{-35})$	6.62**
				21.5	10.70	1220	323 (43)	$7.86X10^{-13}$ $(3.26X10^{-13})$	
	O. c. cuniculus	47	23	6.6	-	376	383 (89)	$4.49X10^{-20}$ $(2.72X10^{-22})$	2.63[n.s.]
				8.6	1.73	490	343 (56)	$1.67X10^{-20}$ $(8.14X10^{-21})$	
Sry	*O. cuniculus*	81	13	2.51	-	323	880 (246)	$8.12X10^{-16}$ $(1.39X10^{-17})$	
	O. c. algirus	46	4	0.96	-	124	203 (89)	$9.61X10^{-6}$ $(1.32X10^{-8})$	5.02*
				1.27	0.91	244	250 (77)	$1.18X10^{-4}$ $(1.85X10^{-8})$	
	O. c. cuniculus	35	4	1.01	-	194	342 (141)	$1.52X10^{-6}$ $(1.13X10^{-9})$	0.20[n.s.]
				1.21	0.51	233	305 (98)	$1.68X10^{-6}$ $(3.91X10^{-9})$	

[1] Sample size

[2] Number of segregating sites

[3] N_e for constant size population model; N_0, the present day population size for exponential growth model

Discussion

The direct comparison of patterns of nucleotide variability at mitochondrial and Y linked genes has allowed us to disentangle the female and male contribution to demographic processes in the European rabbit. We found that in some instances, male and female specific markers reveal genetic signatures of similar population processes, but in some cases, they have shown interesting differences that can be related to well known social behaviors in this species.

Divergence and gene flow between subspecies of rabbit

Previous studies on mitochondrial DNA (Biju-Duval *et al.* 1991; Branco *et al.* 2000), four X chromosome loci (Geraldes *et al.* 2006) and allozyme variation (Ferrand and Branco 2006) have concluded that the European rabbit is composed of two highly divergent lineages that correspond to two subspecies *O. c. algirus* in southwestern Iberian Peninsula and *O. c. cuniculus* in northeastern Iberian Peninsula. In a preliminary screening of SRY variability in the European rabbit, Geraldes *et al.* (2005) sequenced four wild rabbits and eight domestic rabbits and argued that their data was concordant with mtDNA data in showing two divergent lineages. Here, our extensive sequencing of 2 kb of the SRY gene in 107 wild rabbits confirmed the

existence of only two rabbit lineages in the Iberian Peninsula (Figure 2). Furthermore, our results confirm previous reports of high levels of genetic differentiation between subspecies for Y and mtDNA loci (Geraldes *et al.* 2006 evolution). Our estimates of genetic differentiation are slightly lower than those from Geraldes *et al.* (2006 evolution) but this might be accounted for by different methods for estimating Fst, different sampling regimes and the use of nucleotide variability at the sequence level instead of RFLPs diagnostic for the two lineages observed.

Divergence between the two mitochondrial lineages based on RFLP of the entire molecule is aprox. 4% (Biju-Duval *et al.* 1991). Assuming a mammalian molecular clock of 2% divergence per MY, Biju-Duval *et al.* (1991) estimated that the two mitochondrial lineages have started to diverge 2 MYA. Using RFLPs of the Cytb gene alone, Branco *et al.* (2000) estimated a much larger nucleotide divergence (11.9%) between the two lineages found at Cytb. Support for this old divergence between rabbit subspecies was obtained by Geraldes *et al.* (2006) with the sequencing of four X-linked loci. Maximum likelihood estimates of divergence for these gene ranged between 1.76 and 5.79 MY. Here, we used the same approach to estimate divergence at SRY and Cytb. Our ML estimate of divergence for Cytb (6.9%) is almost half the estimated divergence from Cytb RFLP, but higher than the estimated divergence from RFLP data from the entire mitochondria. The use of *Lepus granatensis* to estimate the mutation rate for the Cytb fragment used instead of relying on general mammalian rates lead us to estimate a similar time frame for the divergence of the two lineages. While the estimated divergence at Cytb is concordant with previous estimates for the entire mitochondria (Biju-Duval *et al.* 1991) and X-linked loci (Geraldes *et al.* 2006) in depicting an old split between subspecies of rabbits, for SRY our estimate of divergence was only 0.95 MY. This same pattern was observed when we estimated the TMRCA of all the alleles observed in the wild rabbit (Figure 3). The TMRCA for the SRY gene is 0.88 MY at SRY and 1.89 MY at Cytb. Again, although the TMRCA estimates are slightly more recent than the estimates of divergence between lineages they are concordant in showing a much older split for the mitochondria than the Y chromosome. A similar pattern is observed in Humans where coalescence times for the Y chromosome are half the coalescence times for mtDNA (Tang *et al.* 2002; Wilder *et al.* 2004b). The most likely explanation for this pattern is accelerated divergence of this gene between rabbits and hares, resulting in an overestimation of the mutation rate for this gene between subspecies of rabbit. It is

known that the SRY gene has evolved rapidly by positive directional selection between species of rodents (Tucker and Lundrigan 1993) and between species of primates (Whitfield *et al.* 1993). Recently, Gerrard and Filatov (2005) showed that positive darwininan selection had greatly impacted the evolution of two other mammalian Y-linked genes. In the rabbit, we observed a significant deviation from neutral expectations when ratios of polymorphism and divergence at SRY and Cytb were compared with the HKA test. Also, the MK test, although not significant, revealed that at SRY the number of nonsynonymous substitutions between rabbits and hares (27) is higher than the number of synonymous substitutions (23), while the opposite is true for nonsynonymous (2) and synonymous (3) polymorphisms. This suggests that the SRY gene might show accelerated rates of evolution between Leporid species and this could account for the lower than usual estimate of divergence between subspecies of rabbits when using this marker. The situation seems to be different within the European rabbit since no departure from neutral expectations was observed when ratios of polymorphism to divergence are measured within and between subspecies of rabbit in an HKA framework.

TMRCA of rabbit subspecies

Branco *et al.* (2000) argued that the variability observed at mtDNA within each subspecies of rabbit was consistent with a most recent common mitochondrial ancestor at about 183,000 years and 233,000 years for *O. c. algirus* and *O. c. cuniculus* respectively. We revisited this problem here using a coalescent framework and to estimate the TMRCA for each subspecies at SRY and Cytb (Figure 3). After accounting for the standard deviation of our estimates, we found that the TMRCA of each subspecies is similar for the SRY and Cytb genes, and also that both subspecies have similar TMRCAs (around 300,000 years). These results are slightly lower but consistent with the estimate of Branco *et al.* (2000) and are especially interesting given that the TMRCA of all rabbit alleles for the SRY gene was half the TMRCA of the mitochondria.

Effective population size of males and females

An important aspect of rabbit's behavior is their social structure (Myers and Mykytowycz 1958; Daly 1981). Within each social group a hierarchy between males

is formed. Usually there is a dominant male and several subordinate males per group. To our knowledge the only study that looked at the genetic effects of this behavior was that of Daly (1981) that found that dominant males sire between 37% and 74% of the progeny on warrens.

In our study, we found high levels of nucleotide variability both at SRY and Cytb, when compared to other mammalian species. To our knowledge the highest level of nucleotide variability for Y linked markers on mammalian species had been detected in *Microtus agrestis* (Hellborg *et al.* 2005). The values observed for *O. cuniculus* are even higher (0,213% versus 0.190% for *Microtus agrestis*). Estimates for these two species are very high and are probably due to the fact that in both cases the genealogies have a very long branch that corresponds to two independent evolutionary lineages. Levels of polymorphism for each rabbit subspecies alone are one order of magnitude lower at SRY and Cytb but still higher than values reported for other species. At least two reasons could account for these high levels of nucleotide polymorphism: higher mutation rates in the European rabbit, or higher long term effective population sizes in rabbit populations. Since our estimates of the mutation rate for each loci are highly concordant with estimates for the Y chromosome and mitochondria in species with much lower effective population sizes such as humans (Wilder *et al.* 2004*b*), we have to discard differences in mutation rates as the explanation for the higher levels of nucleotide polymorphism observed in rabbits. This leaves us with higher effective population sizes as the likely explanation for this observation. Indeed, Geraldes *et al.* (2006) estimated the effective population size of *algirus* to be approximately 882,675 and the effective population size of *cuniculus* to be aprox. 422,149. These values are high, and higher than values estimated for most mammalian species. Here, our estimates derived from the Cytb gene are close to this, but the estimates for the SRY gene are much lower (Table 9). Our results show that the present day effective population size of females is five times higher than that of males in *algirus* populations while in *cuniculus* populations historical effective population size of females is only approximately two times higher than that of males. In principle, two explanations could account for these differences: stronger hierarchies in *algirus* than *cuniculus* populations, or different contributions of males and females to population expansions. Since to our knowledge, there have been no comparative behavioral studies between subspecies, we do not have any means to know whether social hierarchies between *algirus* males are stronger than

those in *cuniculus* males. On the other hand, in our coalescent analyses of Cytb and SRY we found that the patterns of nucleotide variability in algirus *populations* were not compatible with long term constant population sizes. Similarly, we observed significant deviations from a neutral model of molecular evolution at Cytb in *algirus* populations for the three tests of neutrality based on the frequency spectrum. Taking these results together, we suggest that *algirus* populations recently underwent large population increases and that possibly females contributed more to these increase than males.

Male biased natal dispersal

Rabbit populations exhibit marked differences in dispersal rates between sexes (Kunkele and Holst 1996; Surridge *et al.* 1999; Richardson *et al.* 2002). Genetic evidence for this behavior was found in populations from the UK (Surridge *et al.* 1999) where the relatedness of females within social groups was at least twice the average relatedness of males within social groups. This pattern held within the entire fenced studied population. Our results clearly support these conclusions. Within each subspecies there was no significant population differentiation for the Y chromosome, while for the female inherited mitochondria there was substantial genetic structuring. Assuming an island model of population structure (Wright 1951), within each subspecies the effective number of migrant females between populations per generation is close to 2, while the effective number of migrant males per generation is 99 between *algirus* populations and 16 between *cuniculus* populations (results not shown).

To our knowledge this is the first study of a wild species that compares the contribution of both sexes to demographic processes by directly comparing patterns of nucleotide variability in the same individuals for the Y chromosome and mitochondrial DNA under a coalescent framework. Our data clearly shows the extent to which the evolutionary history of a species can be much more complex than what is depicted by the study of nucleotide variability at the maternally inherited mtDNA alone. This is clearly exemplified in terms effective population sizes and gene flow. We show that in rabbit populations gene flow is mostly male driven and the effective population sizes of males and females can differ up to five times. These differences

are likely to be due to female philopatry and male social hierarchies and both are know to be widespread phenomena in mammalian populations.

Acknowledgments

We would like to acknowledge Rafael Villafuerte for help in obtaining samples from Spanish populations and Rafael Villafuerte and Michael W. Nachman for facilitating access to their laboratories where this work was partially done. We would also like to thank Catarina Pinho for help with GENETREE and Sara Rocha for help with estimation of maximum likelihood divergence estimates. This work was funded by PhD fellowship SFRH/BD/4621/2001 to Armando Geraldes and research project 40280/BSE/2001 from Fundação para a Ciência e a Tecnologia.

Literature Cited

Avise JC (1993) *Molecular Markers, Natural History and Evolution.* Kluwer Academic Publishers, Boston.

Bandelt H J, Forster P, Rohl A (1999) Median-Joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37-48.

Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, *et al.* (1991) Mitochondrial DNA evolution in Lagomorphs: origin of systematic heteroplasmy and organization of diversity in European rabbits. *Journal of Molecular Evolution*, 33, 92-102.

Boissinot S, Boursot P (1997) Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house Mouse: Selection on the Y? *Genetics*, 146, 1019-1034.

Branco M, Ferrand N, Monnerot M (2000) Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analyses of the cytochrome *b* gene. *Heredity*, 85, 307-317.

Branco M, Monnerot M, Ferrand N, Templeton A R (2002) postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution*, 56, 792-803.

Callou C (2003) *De la garenne au clapier. Etude archeozoologique du lapin en Europe occidentale.* Memoires du Museum national d'Histoire naturelle, Paris.

Cassens I, Waerebeek K V, Best P B, Tzika A, Helden A L V, Crespo E A, Milinkovitch M C (2005) Evidence for male dispersal along the coasts but no migration in pelagic waters in dusky dolphins (*Lagenorhynchus obscurus*). *Molecular Ecology*, 14, 107-121.

Corbet G B (1994) Taxonomy and Origins. In: The European rabbit: the history and biology of a successful colonizer (ed Thompson H V and King C M), pp. 1-7. Oxford Science Publications, Oxford.

Dally J C (1981) Effects of social organization and environmental diversity on determining the genetic structure of a population of the wild rabbit, *Oryctolagus cuniculus*. *Evolution*, 35, 689-706.

Feng J, Lajia C, Taylor D J, Webster M S (2001) Genetic distinctiveness of endangered Dwarf Blue Sheep (*Pseudois nayaur schaeferi*): Evidence from

mitochondrial control region and y-linked *Zfy* intron sequences. *The Journal of Heredity*, 92, 9-15.

Ferrand N, Branco M. (2006) The evolutionary history of the European rabbit (*Oryctolagus cuniculus*), *in* Phylogeography in Southern European Refugia. Edited by S. Weiss and N. Ferrand. Springer. Amsterdam, The Netherlands. (in press).

Fu X Y, Li W H (1993) Statistical tests of neutrality of mutations. *Genetics*, 133, 696-709.

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

Geraldes A, Ferrand N (2006) A 7-bp insertion in the 3' untranslated region suggests the duplication and concerted evolution of the rabbit *SRY* gene. *Genetics, Selection, Evolution* 38, 313-320.

Geraldes A, Rogel-Gaillard C, Ferrand N (2005) High levels of nucleotide diversity in the European rabbit (Oryctolagus cuniculus) *SRY* gene. *Animal Genetics*, 36, 349-351.

Geraldes A, Nachman MW, Ferrand N (submitted 2005) Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. Submitted to *Evolution.*

Geraldes A, Ferrand N, Nachman MW (2006) Contrasting patterns of introgression at X–linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, DOI: 10.1534/gentics.105.054106.

Gerard DT, Filatov DA (2005) Positive and negative selection on mammalian Y chromosomes. *Molecular Biology and Evolution*, 22, 1423-1432.

Gissi C, Gullberg A, Arnason U (1998) The complete mitochondrial DNA sequence of the rabbit, *Oryctolagus cuniculus*. *Genomics*, 50, 161-169.

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

Hayes H, Rogel-Gaillard C, Zijlstra C, de Haan NA, Urien C, Bourgeaux N, Bertaud M, Bosma AA (2002) Establishment of an R-banded rabbit karyotype nomenclature by FISH localization of 23 chromosome-specific genes on both G- and R-banded chromosomes. *Cytogenetic and Genome Research*, 98, 199-205.

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

Hey J, Kliman R M (1993) Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Molecular Biology and Evolution*, 10, 804-822.

Hooft W F V, Groen A F, Prins H H T (2003) Genetic structure of African buffalo herds based on variation at the mitochondrial D-loop and autossomal microsatellite loci: Evidence for male-biased gene flow. *Conservation Genetics*, 4, 467-477.

Hudson RR (2000) A new statistic for detecting genetic differentiation. *Genetics*, 155, 2011-2014.

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, 116, 153-159.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, 132, 583-589.

Kunkele J, Holst D V (1996) Natal dispersal in the European wild rabbit. *Animal Behaviour*, 51, 1047-1059.

McDonald JH, Kreitman M (1991) Adaptive evolution at the *Adh* locus in *Drosophila*. *Nature*, 351, 652-654.

Matthee C A, van Vuuren B J, Bell D, Robinson T J (2004) A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Systematic Biology*, 53, 433-447.

Myers K, Mykytowycz R (1958) Social behaviour in the wild rabbit. *Nature*, 181, 1515-1516.

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, 76, 5269-5273.

Nyakaana S, Arctander P (1999) Population genetic structure of the African Elephant in Uganda based on variation at mitochondrial and nuclear loci: evidence for male-biased gene flow. *Molecular Ecology*, 8, 1105-1115.

Petit E, Balloux F, Excoffier L (2002) Mammalian population genetics: why not Y? *Trends in Ecology & Evolution*, 17, 28-33.

Posada D and Crandall K A (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 14: 817–818.

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

Pusey A E (1987) Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends in Ecology & Evolution*, 2, 295-299.

Richardson B J, Hayes R A, Wheeler S H, Yardin M R (2002) Social structures, genetic structures and dispersal strategies in Australian rabbit (*Oryctolagus cuniculus*) populations. *Behavioral Ecology and Sociobiology*, 51, 113-121.

Rozas J, Sanchez-DelBarrio J C, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19, 2496-2497.

Schaeffer, SW (2002) Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura. Genetical Research*, 80, 163-175.

Seielstad M T, Minch E, Cavalli-Sforza L L (1998) Genetic evidence for a higher female migration rate in humans. *Nature Genetics*, 20, 278-280.

Stone, A C, Griffiths R C, Zegura S L, Hammer M F (2002) High levels of Y-chromosome nucleotide diversity in the genus *Pan. Proceedings of the National Academy of Sciences of the United States of America*, 99, 43-48.

Surridge A K, Ibrahim K M, Bell D J, Webb N J, Rico C, Hewitt G M (1999) Fine-scale structuring in a natural population of European wild rabbits (*Oryctolagus cuniculus*). *Molecular Ecology*, 8, 299-307.

Swofford D L (2002) PAUP* ver 4.0.b10. Phylogenetic Analysis Using Parsimony and other methods. Sinauer Associates, Sunderland, Sunderland MA.

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

Tang H, Siegmund D O, Shen P, Oefner P J, Feldman M W (2002) Frequentist estimation of coalescence times from nucleotide sequencing data using a tree-based partition. *Genetics*, 161, 447-459.

Tucker P K, Lundrigan B L (1993) Rapid evolution of the sex determining locus in Old World mice and rats. *Nature*, 364, 715-717.

Watterson G A (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7, 256-276.

Webb N J, Ibrahim K M, Bell D J, Hewitt G M (1995) Natal dispersal and genetic structure in a population of the European wild rabbit (*Oryctolagus cuniculus*). *Molecular Ecology*, 4, 239-247.

Whitfield L S, lovell-Badge R, Goodfellow P N (1993) rapid sequence evolution of the mammalian sex-determining region gene SRY. *Nature*, 364, 713-715.

Wilder J A, Kingan S B, Mobasher Z, Pilkington M M, Hammer M F (2004a) Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nature Genetics*, 36, 1122-1125.

Wilder J A, Mobasher Z and Hammer M F (2004b) Genetic evidence for unequal effective population sizes of human females and males. *Molecular Biology and Evolution*, 21, 2047-2057.

Wright S 1951 The genetical structure of populations. *Annals of Eugenics*, 15, 323-354.

4. INSIGHTS INTO THE SPECIATION PROCESS OF THE TWO SUBSPECIES OF THE EUROPEAN RABBIT (*Oryctolagus cuniculus*)

4.2 Paper IV

Published in Genetics (2006), June

Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*)

A. Geraldes[*,†,‡], N. Ferrand[*,†] and M. W. Nachman[‡]

[*]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. [†]Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal. [‡]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721, USA

Abstract

Hybrid zones provide an excellent opportunity for studying the consequences of genetic changes between closely related taxa. Here we investigate patterns of genetic variability and gene flow at four X-linked loci within and between the two subspecies of European rabbit (*Oryctolagus cuniculus cuniculus* and *O. c. algirus*). Two of these genes are located near the centromere and two are located near the telomeres. We observed a deep split in the genealogy of each gene with the root located along the deepest branch in each case, consistent with the evolution of these subspecies in allopatry. The two centromeric loci showed low levels of variability, high levels of linkage disequilibrium and little introgression between subspecies. In contrast, the two telomeric loci showed high levels of variability, low levels of linkage disequilibrium, and considerable introgression between subspecies. These data are consistent with suppression of recombination near the centromere of the rabbit X chromosome. These observations support a view of speciation where genomic incompatibilities at different loci in the genome create localized differences in levels of gene flow between nascent species.

Keywords: X chromosome, Introgression, Nucleotide variability, Linkage disequilibrium, Speciation

Introduction

A key problem in evolutionary genetics concerns the origin of reproductive isolation between incipient species (COYNE and ORR 2004). Two important conclusions come from previous studies of the genetics of reproductive isolation. First, barriers to gene flow often derive from incompatibilities between allelic variants at two or more loci, *i.e.*, epistasis (BATESON 1909; DOBZHANSKY 1936; MULLER 1940, 1942). Empirical support for epistasis comes from a large body of work in Drosophila, beginning with DOBZHANSKY (1936). More recently, specific genes underlying reproductive isolation have been identified, and all involve epistatic interactions (MALITSCHEK *et al.* 1995; TING *et al.* 1998; BARBASH *et al.* 2003; PRESGRAVES *et al.* 2003). Second, loci contributing to reproductive isolation tend to be overrepresented on the X chromosome in groups in which males are heterogametic (COYNE and ORR 1989). Evidence for the "large X effect" comes from mapping studies of hybrid sterility and hybrid inviability (e.g. DOBZHANSKY 1936; GRULA and TAYLOR 1980; TRUE *et al.* 1996; PRESGRAVES 2003; PRESGRAVES *et al.* 2003; TAO *et al.* 2003). Moreover, HALDANE's (1922) rule (the sterility or inviability of heterogametic hybrids) seems to be due largely to incompatibilities involving recessive X-linked mutations (TURELLI and ORR 1995, 2000). Finally, in a number of cases where sister species hybridize in nature, X-linked loci introgress less than autosomal loci (e.g. HAGEN 1990; SPERLING and SPENCE 1991; TUCKER *et al.* 1992).

The genetic basis of reproductive isolation has been studied both with laboratory crosses and in natural hybrid zones, and both approaches have advantages and disadvantages. For example, laboratory crosses make it possible to control the genetic background as well as the environment, and they are repeatable. Hybrid zones offer the advantage of many generations of recombination, making fine-scale mapping more feasible. In hybrid zones, it is possible to identify genes contributing to isolation simply from patterns of gene flow without prior knowledge of the phenotype. Hybrid zones also allow us to study species that cannot be crossed in the laboratory. Finally, hybrid zones provide a picture of the fitness of hybrid genotypes under natural conditions.

The European rabbit (*Oryctolagus cuniculus*) provides an opportunity to study the genetic basis of reproductive isolation between recently evolved taxa. This

species consists of two subspecies, *O. c. algirus* in the southwestern portion of the Iberian Peninsula and *O. c. cuniculus* in the northeast of the Iberian Peninsula and France. These two groups diverged in allopatry during the early Pleistocene and have subsequently come into secondary contact in central Iberia, forming a contact zone that runs in a NW-SE direction (Figure 1) (BRANCO *et al.* 2000, 2002). The two subspecies are well differentiated with respect to mtDNA (BRANCO *et al.* 2000), the Y chromosome (GERALDES *et al.* 2005), and some allozyme loci (FERRAND and BRANCO 2006).

Motivated by the large X-effect documented in other species, here we focus on four X-linked loci to understand the nature of reproductive isolation in rabbits. Two of these loci are near the centromere and two are near the telomeres. We address three main questions. First, what are the levels and patterns of genetic variation at genes on the rabbit X chromosome? Second, are patterns of variation and introgression heterogeneous among loci, and if so, do the differences correlate with the physical location of genes on the X chromosome? Third, are the data compatible with a model of divergence without gene flow? We surveyed nucleotide variability at four X-linked loci, in a sample of 43 male rabbits representing both subspecies and the area of contact. All four loci showed two divergent lineages. Despite this deep divergence, there is still evidence of gene flow between subspecies. Patterns of gene flow and nucleotide variability were heterogeneous among loci, being low at the centromeric loci and high at the telomeric loci. We hypothesize that the centromeric region of the X chromosome of the European rabbit may be involved in reproductive isolation between these two subspecies.

Materials and Methods

Samples: Forty-three male European rabbits were sampled (Table 1). The samples were divided into three groups: 20 from the northeastern region of the Iberian Peninsula and from France, corresponding to *O. c. cuniculus* (NE), 14 from the southwestern region of the Iberian Peninsula corresponding to *O. c. algirus* (SW) and nine from the contact zone (CZ) as defined by mtDNA variation (BRANCO *et al.* 2000). The geographic locations of the populations sampled are shown in Figure 1,

and collecting localities are given in Table 1. Additionally, one male *Lepus granatensis* was used as an outgroup.

Table 1- Individuals sampled and their geographic locations

Population	Sample Size	Population no.[a]	Group	Individual ID
Versailles	1	1	NE	Ver1827
Vaulx-en-Velin	1	2	NE	Vau1
Carlucet	1	3	NE	Cau19
Perpignan	1	4	NE	Pep18
Zaragoza	2	5	NE	Zrg16, Zrg20
Castelló	2	6	NE	Rsl4, Rsl10
Benavente	1	7	NE	Bnv3
Zamora	2	8	NE	Zam1, Zam20
La Rioja	2	9	NE	Lrj3, Lrj6
Madrid	1	10	NE	Mdr7
Alicante	3	11	NE	Alic1, Alt107, Alt120
Cartagena	1	12	NE	Cat12
Cuenca	2	13	NE	Cue1, Cue3
Galicia	1	14	CZ	Gal25c3
Bragança	2	15	CZ	Bra1, Bra13
Toledo	3	16	CZ	Tol25, Tol50, Tol64
Ciudad Real	2	17	CZ	Cre1, Vdm12
Las Amoladeras	1	18	CZ	Amo2
Córdoba	3	19	SW	Luc4, Luc9, Luc17
Sevilla	3	20	SW	Pfr1, Pfr5, Pfr7
Doñana	1	21	SW	Don6
Vila Real	3	22	SW	Vrl1, Vrl4, Vrl7
Idanha-a-nova	1	23	SW	Id85
Elvas	2	24	SW	Elv3, Elv6
Vila Viçosa	1	25	SW	Vv1_1/94

[a]Population numbers from Figure 1.

PCR amplification and sequencing: Genomic DNA was extracted either from blood, muscle, or liver following SAMBROOK and RUSSEL (2001). Introns of four X-linked loci were PCR amplified; two are located near the centromere and two are near the telomeres (Figure 2). Amplification and sequencing primers are listed in Supplemental Table 1. For each locus, two pairs of amplification primers were designed. The first was based either on published rabbit sequences (*Phka2*) (DAVIDSON *et al.* 1992) or on conserved exonic regions among human, mouse and rat. Nested primers were then designed specifically for the rabbit based on the first sequences obtained. Amplifications were carried out in 50-μL volumes using

Platinum *Taq* High Fidelity DNA Polymerase (Invitrogen, San Diego, CA) following manufacturer recommendations. Cycling temperatures were as follows: an initial denaturation step at 94° for 1min and 20 sec followed by 35 cycles of 94° for 20 sec, annealing for 20 sec and extension at 68° for 4 min. Annealing temperatures for each PCR are specified in Supplemental Table 1. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Chatsworth, CA) prior to sequencing. Sequencing was carried out using an ABI 3700 automated sequencer. All sequences have been deposited in GenBank under accession numbers DQ306315-DQ306490.

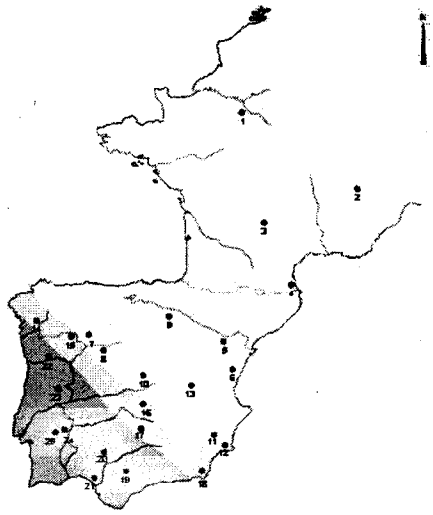


Figure 1- Populations of European rabbit sampled and their geographic locations. Dark shading indicates SW populations, light shading indicates CZ populations and no shading indicates NE populations. The name and number of samples from each population, and population names are specified in Table 1.

Data Analyses: Sequences were inspected and concatenated using the computer program Sequencher (Gene Codes, Ann Arbor, MI) and then aligned manually using the BioEdit software (HALL 1999). By sequencing the X chromosome in males we were able to recover haplotypes directly. The analyses below were based on single nucleotide polymorphisms in introns only.

Basic population genetic parameters, including the number of segregating sites, number of haplotypes, levels of nucleotide diversity, π (NEI and LI 1979), and the proportion of segregating sites, θ (WATTERSON 1975), were estimated using the program DnaSP 4.00 (ROZAS *et al.* 2003) for the entire data set and also for the NE, CZ and SW groups (Figure 1). Phylogenetic relationships among alleles were

estimated using the Median Joining (MJ) algorithm (BANDELT *et al.* 1999) as implemented in Network v4.1.0.8 (http://www.fluxus-technology.com/).

We estimated divergence three ways. First, divergence between all *O. cuniculus* alleles and *L. granatensis* was calculated as the average pairwise distance per nucleotide site, D_{xy} (Nei 1987), and as the number of net nucleotide substitutions per site, D_a (NEI 1987). D_a is defined as $D_{xy} - 0.5 (D_x + D_y)$, where D_{xy} is the average pairwise distance between groups and D_x and D_y are the average pairwise distances within groups. Second, D_{xy} and D_a were calculated between the NE and SW groups of *O. cuniculus*. Finally, to estimate the divergence time of the two subspecies of *O. cuniculus*, maximum likelihood (ML) net nucleotide distances between *L. granatensis* and *O. cuniculus*, and between the two main lineages found in *O. cuniculus* (see RESULTS), were calculated using PAUP v 4.0 (SWOFFORD, 2002). Divergence time between subspecies of *O. cuniculus* was calculated assuming a divergence time between *L. granatensis* and *O. cuniculus* of 11.8 million years (MY) (MATTHEE *et al.* 2004).

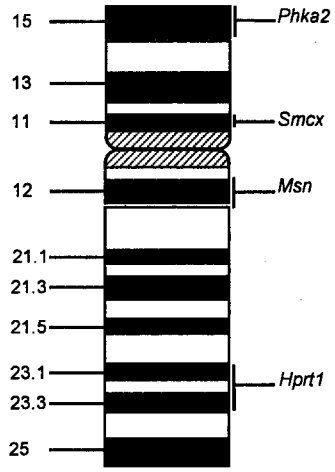


Figure 2 - Chromosomal location of the four X-linked loci used in the present study. Modified from CHANTRY-DARMON *et al.* (2003) and HAYES *et al.*(2002).

The population recombination parameter, R (R= 3Nc for X-linked loci, where c is the recombination rate per generation and N is the population size) between adjacent sites (HUDSON 1987), the minimum number of recombination events, Rm (HUDSON and KAPLAN 1985), and the number of pairs of sites showing four gametic types were calculated using DnaSP 4.00 (ROZAS *et al.* 2003). Another estimator of the population recombination parameter, γ (HEY and WAKELEY 1997), was

calculated using the software SITES. While Hudson's R is based on the variance of the number of base pair differences between DNA sequences, γ is a maximum likelihood estimator developed using a coalescent model for a sample of four DNA sequences with recombination. Linkage disequilibrium (LD) between pairs of polymorphic sites present at a frequency of at least 10% was calculated within and between all loci, using the statistics D' (LEWONTIN 1964) and r^2 (HILL and ROBERTSON 1968) as implemented in DnaSP 4.00 (ROZAS *et al.* 2003).

Tajima's D (TAJIMA 1989) and Fu and Li's D (FU and LI 1993) were calculated to test for deviations from a neutral equilibrium frequency distribution using DnaSP 4.00 (ROZAS *et al.* 2003). Ratios of polymorphism within *O. cuniculus* to divergence between *O. cuniculus* and *L. granatensis* were compared with the expectations under a neutral model using the Hudson-Kreitman-Aguadé (HKA) test (HUDSON *et al.* 1987). We performed one four-locus test and six pairwise comparisons between loci using the HKA software (HEY and KLIMAN 1993).

At each of the four loci we detected a deep split in the genealogy (see RESULTS). We asked if the observed pattern of nucleotide polymorphism is compatible with a single panmitic population, as opposed to some form of population subdivision. If two populations have evolved in allopatry, the basal branch of a gene genealogy may be longer than in a panmitic population. Furthermore, mutations arising in an isolated subpopulation are unable to recombine with mutations in a different subpopulation, resulting in higher levels of LD. WALL (2000) suggested two measures based on LD that could be powerful indicators of population subdivision. The first, l_b, is the number of congruent sites, defined as the number of mutations that, on a pairwise basis, result in only two haplotypes. The second, g_d, is the maximum physical distance between congruent sites. Coalescent simulations of panmixia were performed with the computer program ms (Hudson 2002). For each locus 50,000 genealogies of 43 individuals were simulated conditioned on the estimated values of θ and γ. Additionally, for each locus, coalescent simulations were performed using two different values of the population recombination parameter (3Nc=0.0015 and 3Nc=0.015, per site) chosen to reflect the range of recombination rates known for other mammals (e.g. DIETRICH *et al.* 1994; KONG *et al.* 2002; JENSEN-SEAMAN *et al.* 2004). A computer program (GARRIGAN *et al.* 2005) was used to calculate l_b and g_d from the simulated datasets, and the distribution of the two statistics for each set of conditions were plotted against each

other. The probability of obtaining the observed values of l_b and g_d was calculated as the proportion of simulated genealogies for which the values of l_d and g_d were greater than the observed values.

Fst and Nm were calculated using the method of HUDSON *et al.* (1992a) implemented in DnaSP 4.00 (ROZAS *et al.* 2003). Genetic differentiation was also calculated using the test statistic Ks* (HUDSON *et al.* 1992b), and significance was assessed by performing 1000 permutations. To test for significant population structure among populations and among groups of populations, analyses of molecular variance (AMOVA; EXCOFFIER *et al.* 1992) between the SW and NE groups were performed using ARLEQUIN (SCHNEIDER *et al.* 2000).

One simple model of divergence is an isolation model in which two populations become separated with no subsequent gene exchange. The HKA model (HUDSON *et al.* 1987) takes this form and further assumes that the ancestral species has a population size that is the average of the two descendent species. More recent models relax this assumption. For example, WAKELEY and HEY (1997) proposed a model that is similar to the HKA model but includes an additional parameter, θ_A, the population mutation parameter for the ancestral species. While the HKA test only uses the number of polymorphic sites and divergence, this model also incorporates the total number of polymorphic positions in the two groups (S), the number of polymorphisms exclusive to one group (Sx_{NE} and Sx_{SW}), the number of shared polymorphisms (Ss) and the number of fixed differences (Sf). We tested the fit of our data to these two models in two different ways. First we performed pairwise comparisons between all loci, and second we performed tests with all four loci together. The fit of our data to the WAKELEY and HEY model of divergence without gene flow was tested using the program WH (WANG *et al.* 1997).

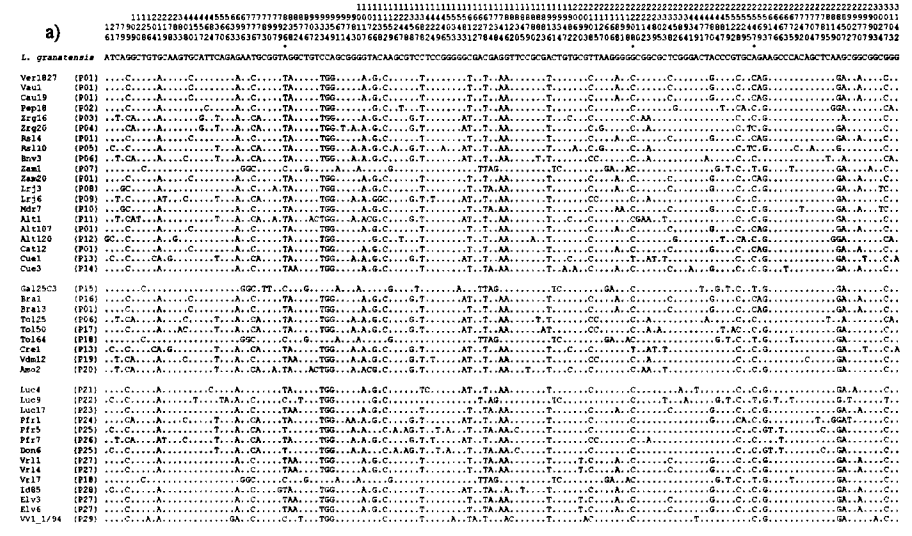
These models assume that there has been no gene flow between the two populations since the initial split. In many cases this is an unrealistic assumption. HEY and NIELSEN (2004) developed a model of population divergence that allows for genetic drift (increasing population divergence) and gene flow (preventing population divergence) to act together, which they call the Isolation with Migration model. The computer program IM is an implementation of the Markov Chain Monte Carlo method for the analyses of genetic data under this model. We used IM to estimate the effective population size of *O. c. cuniculus* and of *O. c. algirus* and to estimate migration rates for each locus between subspecies in each direction. IM

assumes that there is no recombination within loci. For *Phka2* and *Hprt1* we used the largest region showing no evidence of recombination, following WON and HEY (2005). For *Phka2* a portion of 687 bp containing 23 polymorphic sites was used, and for *Hprt1* a region of 501 bp with 20 polymorphic sites was used. For *Smcx* all 20 NE and 14 SW individuals were used since the data are free of recombination. For *Msn* we removed three recombinant individuals (Vau1, Rsl4 and Rsl10) from the NE group. We assigned wide prior distributions of the parameters based on preliminary trial runs. We ran the program under Metropolis Coupled Monte Carlo Markov Chains (MC3), using 10 chains with linear heating. We used a burn-in period of 1,000,000 steps and recorded results every 40 steps. To test that the chains were mixing well we ran the program with different random seed numbers and the results were similar. We ran the program for 25,625,601 steps after the burn-in period and recorded the results of 625,014 steps.

Results

Levels and patterns of variation: We observed considerable variation at all four genes. Polymorphic sites for each gene are shown in Figure 3, and summaries of variation are given in Table 2. The number of polymorphic sites varied from 56 at *Msn* to 151 at *Phka2*, but the number of haplotypes observed at each gene was much more constant (from 23 at *Smcx* to 29 at *Phka2*). Levels of polymorphism

a)



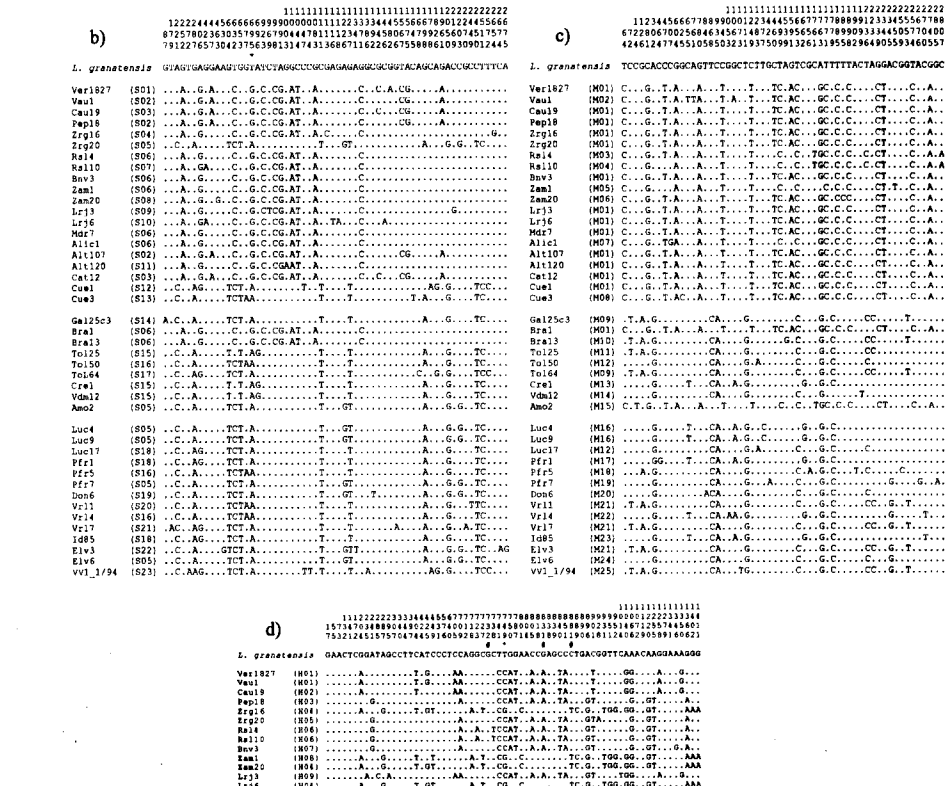


Figure 3- Polymorphisms for the 4 X-linked loci. a) *Phka2*, b) *Smcx*, c) *Msn* and d) *Hprt1*. The nucleotide found in the *Lepus granatensis* sample was used to infer the ancestral state for each position. Numbers in the columns indicate the position of the polymorphism in the alignment. Each row represents one individual. A dot (.) represents the ancestral state, and letters (A, C, G, and T) represent the derived nucleotide. The name of the haplotype defined by each sequence is shown in parentheses. Positions polymorphic for bases other than the one found in *Lepus granatensis* are indicated with an asterisk (*) and positions segregating for three different bases in *O. cuniculus* are indicated with the symbol (#). Whenever a polymorphic site in *O. cuniculus* corresponded to a deletion in *L. granatensis* we treated the most frequent base in our sample of *O. cuniculus* as the ancestral state.

were high, both in the total sample and within each subspecies. However, nucleotide diversity (π) at the two centromeric loci was considerably lower (0.52% at *Smcx* and 0.55% at *Msn*) than at the two telomeric loci (0.70% at *Phka2* and 1.26% at *Hprt1*). This contrast was even more striking in the proportion of segregating sites, where θ at the centromeric loci was roughly half the value seen at the telomeric loci.

Table 2- Levels of polymorphism, allele frequency spectrum tests of neutrality, divergence and recombination

									Divergence (%)								
			Polymorphism					Frequency Spectrum Tests of Neutrality		Between NE and SW		Between *O. cuniculus* and *L. granatensis*		Recombination			
Locus	Sample	n	L[a]	K[b]	S[c]	π (%)	θ (%)	Tajima's D	Fu and Li's D	Da[d]	Dxy[e]	Da[f]	Dxy[g]	γ[h]	R[i]	Rm[j]	4gt[k]
Phka2	All	43	3168	29	151	0.699	1.102	-1.337	-0.995	0.008	0.659	5.719	6.059	0.0051	0.0025	17	390
	NE	20	3168	14	105	0.574	0.934	-1.584	-2.106					0.0019	0.0018	10	95
	CZ	9	3168	9	86	0.924	0.999	-0.385	0.121					0.0052	0.0028	3	25
	SW	14	3168	10	95	0.708	0.945	-1.109	-1.109					0.0074	0.0032	10	137
Smcx	All	43	2709	23	60	0.517	0.512	0.035	-1.499	0.531	0.782	1.475	1.735	0	0.0012	0	0
	NE	20	2709	13	44	0.341	0.458	-1.021	-0.647					0	0	0	0
	CZ	9	2709	6	29	0.390	0.394	-0.054	0.667					0	0.0003	0	0
	SW	14	2709	8	21	0.160	0.244	-1.438	-2.172					0	0.0020	0	0
Msn	All	43	2825	25	56	0.553	0.458	0.733	-0.849	0.786	0.949	3.803	4.083	0.0023	0.0011	7	77
	NE	20	2825	8	14	0.079	0.140	-1.591	-1.390					0.0003	0.0002	1	3
	CZ	9	2825	8	36	0.462	0.469	-0.073	0.423					0	0.0005	0	0
	SW	14	2825	11	23	0.246	0.256	-0.161	-0.649					0.0016	0.0117	2	7
Hprt1	All	43	1473	28	68	1.256	1.115	0.454	-0.381	0.027	1.256	3.848	4.428	0.0088	0.0086	6	180
	NE	20	1473	13	47	1.129	0.899	1.025	0.978					0.0071	0.0034	2	58
	CZ	9	1473	8	49	1.228	1.249	-0.087	0.134					0.0094	0.0109	4	104
	SW	14	1473	12	50	1.328	1.132	0.760	0.642					0.0079	0.0120	4	111

[a]Length of the sequence in base pairs.
[b]Number of haplotypes.
[c]Number of polymorphic sites.
[d]Net nucleotide divergence per site (NEI 1987) between NE and SW (CZ was excluded from this analysis).
[e]Average pairwise nucleotide substitutions per site (NEI 1987) between NE and SW (CZ was excluded from this analysis).
[f]Net nucleotide divergence (NEI 1987) between the haplotype found in *Lepus granatensis* and all *O. cuniculus* sequences.
[g]Average pairwise nucleotide substitutions per site (NEI 1987) between the haplotype found in *L. granatensis* and all *O. cuniculus* sequences.
[h]Maximum likelihood estimate of the population recombination parameter between adjacent sites (HEY and WAKELEY 1997).
[i]Hudson's (1987) estimator of the population recombination parameter between adjacent sites.
[j]Minimum number of recombination events in the history of the sample (HUDSON and KAPLAN 1985).
[k]No. of the pairs of sites that show all four gametic types.

We assessed the amount of LD in our sample in several ways, and all were consistent in revealing more recombination (less LD) at the telomeric loci than at the centromeric loci (Table 2), consistent with suppression of recombination near the centromere. The number of pairs of sites showing all four gametic types was zero at *Smcx*, 77 at *Msn*, 390 at *Phka2*, and 180 at *Hprt1*. Rm, the minimum number of recombination events in the history of the sample (HUDSON and KAPLAN 1985) was zero at *Smcx*, intermediate at *Msn* and *Hprt1* (7 and 6 respectively) and highest at *Phka2* (17). Similarly, R (HUDSON 1987) between adjacent sites was low at the two centromeric loci (*Smcx* and *Msn*) and much higher at the telomeric loci (*Phka2* and *Hprt1*). The values for γ (HEY and WAKELEY 1997), a maximum likelihood estimator of the population recombination parameter, are concordant with R in showing more recombination at the telomeric loci than at the centromeric loci. Thus, although there are small differences among estimators, levels of recombination are higher at the telomeric loci than at the centromeric loci. We assessed the significance of LD through pairwise comparisons of polymorphic sites present at a

frequency of at least 10% using a Fisher's Exact Test. From the 8515 pairwise comparisons performed, 2497 were significant in a two-tailed test, and 980 remained significant after a Bonferroni correction for multiple tests. Of these, 197 were between polymorphic sites at *Smcx*, 219 between sites at *Msn*, 93 between sites at *Phka2*, and 130 between sites at *Hprt1*. Interlocus LD was only detected between *Smcx* and *Msn* where there were 341 pairs of sites that showed significant LD.

The distribution of allele frequencies as measured by Tajima's D and Fu and Li's D generally conformed to expectations under a neutral model of molecular evolution (Table 2). For example, in the total sample, Tajima's D was either positive (*Msn* and *Hprt1*), very close to zero (*Smcx*) or negative (*Phka2*), but not significantly different from zero. When the population groups were analyzed separately, Tajima's D was negative (except for *Hprt1* in the NE and SW groups), but not significantly so ($P > 0.05$ for all tests). We also tested a neutral model of molecular evolution by comparing ratios of polymorphism within *O. cuniculus* (in the total sample) to divergence between *O. cuniculus* and *L. granatensis*, using the HKA test. We performed one four-locus comparison as well as six pairwise comparisons between loci, and none of these were significant ($P > 0.05$ for each).

Divergence and gene flow between subspecies: Fst estimates between *O. c. cuniculus* and *O. c. algirus* are shown in Table 3. Fst was very high at the centromeric loci, *Smcx* (0.680) and *Msn* (0.829), and one order of magnitude lower at the telomeric loci, *Phka2* (0.027) and *Hprt1* (0.022). The two subspecies were significantly differentiated, using the Ks* test statistic, at all loci but *Hprt1* (Table 3). Similarly, the AMOVA analyses revealed that at the two centromeric loci, *Smcx* and *Msn*, most of the genetic variation is partitioned among the two subspecies (64% and 84% respectively) while at the telomeric loci, *Phka2* and *Hprt1*, most of the observed variation was partitioned among populations within each subspecies (93% at both loci), and only a marginal proportion (2%) of the variation was segregating between subspecies.

This differentiation can also be seen in the phylogeny of alleles for each gene (Figure 4). At each locus there were two divergent groups of haplotypes, and in each case the root fell along the deep branch separating these two groups. In this analysis, only *Smcx* was free of homoplasy. The locus with the most homoplasy was *Phka2*. This homoplasy may be due to recombination or recurrent mutation. Evidence for recurrent mutation comes from the observation that at *Hprt1* three different positions

have three nucleotides segregating (Figure 3d). Other evidence of recurrent mutation is the fact that the amount of homoplasy is slightly reduced if CpG sites, which are known to be hypermutable, are excluded. For example, for *Phka2*, the consistency index (CI) increased from 0.778 to 0.803 when CpG sites were removed. However, much of the homoplasy is probably due to recombination, as evidenced by the fact that the CI was 1.0 for *Smcx*, 0.889 at *Msn*, but was 0.778 and 0.717 at *Phka2* and *Hprt1*, respectively. Moreover, at *Phka2* and *Hprt1* one and 16 individuals, respectively, were identified as recombinants between the two divergent lineages based on their position on the haplotype network and by visual inspection of the table of polymorphism (Figure 3 a and d).

Table 3- Genetic differentiation between NE and SW groups at four X-linked loci

	Fst[a]	Nm[b]	ϕ_{ct}[c]	ϕ_{st}[d]	ϕ_{sc}[e]	Da(%)[f]
Phka2	0.0266*	12.22	0.02	0.07	0.05	0.008
Smcx	0.6796***	0.16	0.64	0.80	0.43	0.531
Msn	0.8286***	0.07	0.84	0.86	0.11	0.786
Hprt1	0.0218	14.99	0.02	0.07	0.05	0.027

*P<0.05, **P<0.01, ***P<0.005.

[a]Fst was calculated using the method proposed by HUDSON *et al.* (1992). Statistical significance for the estimation of Fst between the two groups was obtained with the Kst* statistic (HUDSON et al. 1992).

[b]Nm was calculated according to WRIGHT's (1951) island model of population structure, using the expression (Fst=1/(1+3Nm)) for X-linked loci.

[c]ϕ_{ct} is the fixation index for the amount of variation segregating between NE and SW groups, calculated using the AMOVA framework (EXCOFIER *et al.* 1992). For the NE group the thirteen populations studied were pooled into seven subgroups (NE1 - populations 1, 2, 3 and 4; NE2 - population 5; NE3 - population 6; NE4 - populations 7 and 8; NE5- populations 9 and 10; NE6 - populations 11 and 12; and NE7 - population 13). For the SW group, the seven populations studied were pooled into four subgroups (SW1 - population 19; SW2 - populations 20 and 21; SW3 - population 22; and SW4 - populations 23, 24 and 25).

[d]ϕ_{st} is the fixation index for the amount of variation segregating within each subgroup, calculated using the AMOVA framework (EXCOFIER *et al.* 1992).

[e]ϕ_{sc} is the fixation index for the amount of variation segregating among subgroups within each group, calculated using the AMOVA framework (EXCOFIER *et al.* 1992).

[f]Da is the net nucleotide distance per base pair between populations (NEI 1987).

The degree of introgression between subspecies can also be seen by the concordance (or lack thereof) between geography and phylogeny. For the two centromeric loci (*Smcx* and *Msn*) there was good concordance between phylogeny and geography; *i.e.*, the two major lineages correspond well with each subspecies (Figure 4). At *Msn* we did not detect any introgressed haplotypes and at *Smcx* we only observed three NE individuals with haplotypes from the lineage that is otherwise restricted to the SW and CZ groups. At the two telomeric genes (*Phka2* and *Hprt1*), in contrast, there seems to be little or no concordance between

phylogeny and geography. At all four genes, individuals from the CZ group are scattered throughout the haplotype networks.

The proportion of congruent sites, l_b, is greater at the two centromeric loci (representing 30% and 32% of all polymorphic sites at *Smcx* and at *Msn* respectively), than at the telomeric loci (11% at *Phka2* and 16% at *Hprt1*).

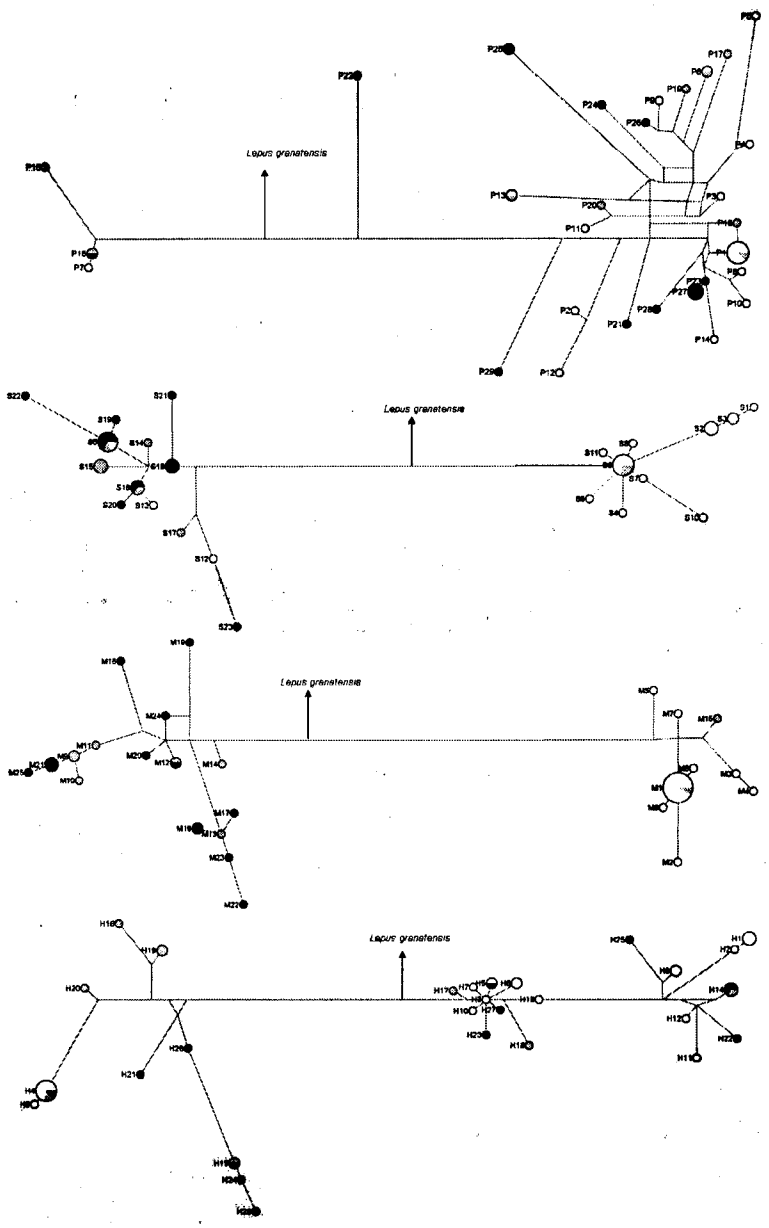


Figure 4- Median-Joining haplotype networks representing the phylogenetic relationships between all the alleles found in the European rabbit. From top to bottom: *Phka2*, *Smcx*, *Msn* and *Hprt1*. The size of the circles is proportional to the frequency of each haplotype. The population group of the individuals that are represented in each haplotype is denoted by black (SW), grey (CZ) and white (NE). The point in the network from which the outgroup sequence of *L. granatensis* stems is represented by an arrow. Haplotype IDs correspond to Figure 3.

Similarly, the maximum distance between congruent sites, g_d, is greater at *Smcx* (85% of the total locus length) and at *Msn* (95%) than at *Phka2* (65%) and *Hprt1* (16%). We calculated the probability of observing these values of l_b and g_d using coalescent simulations of 50,000 genealogies of 43 individuals evolving neutrally under panmixia with mutation (θ) and recombination (γ) parameters estimated from the data. Results are shown in Table 4. Under these conditions, the null model was rejected for *Msn* (P=0.00214). This test is quite conservative using γ estimated from the data since population subdivision will increase LD and thus underestimate the true value of recombination. Therefore, we also conducted simulations with a population size of 10^5 and per site recombination rates of $0.5X10^{-8}$ and $5X10^{-8}$, reflecting the range of recombination rates seen in other mammals (e.g. DIETRICH *et al.* 1994; KONG *et al.* 2002; JENSEN-SEAMAN *et al.* 2004). At the two centromeric loci, *Msn* and *Smcx*, the null model was rejected using either value of recombination. For *Phka2* the null hypothesis was rejected only with the higher recombination rate, and *Hprt1* was marginally significant (P=0.066) only for the higher recombination rate (Table 4).

Table 4- Probabilities of observing the number of congruent sites, l_b, and maximum distance between congruent sites, g_d, under a single panmitic population.

	γ estimated from data[a]			γ =0.0015[b]			γ =0.015[c]		
	l_b	g_d	l_b and g_d	l_b	g_d	l_b and g_d	l_b	g_d	l_b and g_d
'hka 2	0.35364	0.96288	0.34002	0.76470	0.99352	0.76296	0.04484	0.89190	0.04032
mcx	0.39566	0.79952	0.37298	0.09558	0.33292	0.04248	0.00010	0.10748	0.00004
tsn	0.03470	0.03466	0.00214	0.06722	0.04908	0.00638	0.00006	0.01082	0.00000
lprt 1	0.17324	0.99916	0.17302	0.62354	0.99992	0.62354	0.06626	0.99742	0.06626

Probabilities were calculated as the proportion of simulated genealogies with values of l_b, g_d, or both, equal or greater than those observed in our data.

[a] For *Phka2* γ=0.0051 per site, for *Smcx* γ=0 per site, for *Msn* γ=0.0023 per site and for *Hprt1* γ=0.0088 per site.

[b] γ=3Nc, where $N=1X10^5$ and $c=0.5X10^{-8}$ per site.

[c] γ=3Nc, where $N=1X10^5$ and $c=5X10^{-8}$ per site Table 4 - Shared and fixed variation between NE and SW groups at 4 X-linked loci.

Another way of looking at divergence is to quantify the amount of shared and fixed variation between the two groups (Table 5). The number of shared polymorphisms was low at the centromeric loci (16% and 6% of all polymorphisms at *Smcx* and *Msn* respectively), and high at the telomeric loci (42% and 64% at *Phka2* and *Hprt1* respectively). Only *Msn* showed fixed differences between the two groups. These patterns of variation suggest that there has been gene flow between *O. c. cuniculus* and *O. c. algirus* at some, but not all, loci. To further test this, we performed an HKA test between NE and SW population groups. A multilocus test between all four loci failed to reject the null model, but in one pairwise comparison

(between *Phka2* and *Msn*) the model was rejected (P=0.035). This result seems to be mainly driven by the fact that divergence at *Msn* was much higher than expected (observed D=26.80, expected D=10.68). The isolation without migration model (WAKELEY and HEY, 1997) shares most of the assumptions with the HKA model, but estimates the ancestral population size instead of assuming that it is the average of the population size of the extant populations. A four locus test using this model also failed to reject the null hypothesis. We also tested the fit of the data using all pairwise comparisons to *Msn*. Only the comparisons to *Msn* were performed because in the other comparisons there are no fixed differences and the program is unable to simulate the distribution of the expected values. The comparison between *Smcx* and *Msn* failed to reject the null model while the other two comparisons did reject the null model (*Phka2/Msn* $P_{(\chi 2)}$=0.032 and $P_{(WH)}$=0.030; *Msn/Hprt1* $P_{(\chi 2)}$=0.024 and $P_{(WH)}$=0.009).

We used IM (HEY and NIELSEN 2004) to obtain Maximum Likelihood Estimates (MLE) of the effective population size for each subspecies. We also

Table 5- Shared and fixed variation between NE and SW groups at four X-linked loci[a]

	S[b]	Sx_{NE}[c]	Sx_{SW}[d]	Ss[e]	Sf[f]
Phka2	140	45 *(64.6)*	35 *(56.5)*	60 *(41.1)*	0 *(10.9)*
Smcx	56	35 *(16.9)*	12 *(14.7)*	9 *(10.7)*	0 *(2.8)*
Msn	53	11 *(15.2)*	20 *(13.3)*	3 *(9.7)*	19 *(2.6)*
Hprt1	59	9 *(20.3)*	15 *(17.7)*	38 *(12.9)*	0 (3.4)

[a]The expected values under the population parameters estimated with software WH (WAKELEY and HEY 1997) are shown between parentheses
[b]S- Number of polymorphic positions.
[c]Sx_{NE}- number of exclusive polymorphisms in the NE group.
[d]Sx_{SW}- number of exclusive polymorphisms in the SW group.
[e]Ss- Number of shared polymorphisms.
[f]Sf- number of fixed differences.

estimated migration rates for each locus in each direction. The average estimate of the effective population size was approximately 882,000 for *O. c. algirus* and 422,000 for *O. c. cuniculus* (Table 6). The probability distribution of the ancestral population parameter was flat (not shown), as expected if the ancestral population existed long ago (WON and HEY 2005). Similarly the probability distribution of *t*, the time since divergence, was flat, but non zero (not shown). This suggests that the two subspecies were isolated in the past, but this analysis does not provide a reliable estimate of the time of isolation. Gene flow at the telomeric loci was higher from NE to SW than from SW to NE. For the centromeric loci, introgression of *Msn* is quite low in both directions, while *Smcx* shows some unidirectional introgression

from SW to NE. Thus it seems that levels and patterns of gene flow are very different between centromeric and telomeric loci.

Table 6 - Maximum-Likelihood Estimates (MLE) and the 90% Highest Posterior Density (HPD) Intervals[a] of demographic parameters and migration rates from NE to SW (top) and from SW to NE (bottom).

			Nm from NE to SW			
	Ne_{SW}[b]	Ne_{NE}[c]	*Phka2*	*Smcx*	*Msn*	*Hprt1*
MLE	882,675	422,149	10.215	0.008*	0.008*	3.470
HPD	520,833-1,737,939	202,851-685,307	2.600-16.092	0.008-6.480	0.008-0.411	0.733-14.695

			Nm from SW to NE			
	Ne_{SW}	Ne_{NE}	*Phka2*	*Smcx*	*Msn*	*Hprt1*
MLE	882,675	422,149	0.281	0.404	0.012	0.951
HPD	520,833-1,737,939	202,851-685,307	0.004-7.696[+]	0.004-3.569	0.004-0.358	0.004-6.687

*The estimated value of Nm is at the lower limit of resolution; *i. e.* 0.008 corresponds to the first bin of the parameter space surveyed for the migration parameter from SW to NE.

[+]The actual interval was larger than this and could not be estimated reliably, because the likelihood surface was relatively flat.

[a]The 90% HPD intervals contain 90% of the probability density for each estimate.

[b]The estimated effective population size of the SW population group.

[c]The estimated effective population size of the NE population group.

We estimated divergence time between the two subspecies of *O. cuniculus* using a phylogenetic approach. Assuming a divergence time of 11.8 MY (MATTHEE *et al.* 2004) between *O. cuniculus* and *L. granatensis*, divergence time between *O. c. cuniculus* and *O. c. algirus* was estimated to be on the order of 2-5 MY ago (Table 7).

Table 7- Uncorrected and corrected[a] net nucleotide (Da)[b] divergences between *O. cuniculus* and *L. granatensis* and between subspecies of *O. cuniculus*, and estimates of divergence time (MY) between subspecies of *O. cuniculus*.

	O. cuniculus/L. granatensis		*O. c. cuniculus/O. c. algirus*			
	Uncorrected	Corrected	Uncorrected		Corrected	
	Da (%)	Da (%)	Da (%)	Divergence Time (MY)	Da (%)	Divergence Time (MY)
Phka2	5.719	8.483[c]	1.153	2.38	1.263[c]	1.76
Smcx	1.475	1.529[d]	0.735	5.88	0.750[d]	5.79
Msn	3.803	4.634[e]	0.711	2.21	0.819[e]	2.08
Hprt1	3.848	5.313[f]	1.633	5.01	1.911[f]	4.24

[a] Appropriate models of nucleotide substitution to correct for multiple hits were selected for each gene using MODELTEST 3.06 (POSADA and CRANDALL 1998) with the Akaike Information Criterion (POSADA and BUCKLEY, 2004).

[b] Pairwise distances (Dxy) per site were calculated using PAUP v 4.0 (SWOFFORD, 2002) with locus specific estimated models of substitution. Net nucleotide divergences (Da) per site were calculated as Dxy - 0.5 (Dx +Dy). Recombinant haplotypes were excluded from this analysis.

[c] TAMURA-NEI model (1993) with proportion of invariable sites of 0.6578 and estimated α parameter describing the gamma distribution of 1.0022.

[d] TAMURA-NEI model (1993) with proportion of invariable sites of 0.5536.

[e] Transversion Model with estimated α parameter describing the gamma distribution of 0.2204.

[f] TAMURA-NEI model (1993) with proportion of invariable sites of 0.8524.

Discussion

We documented genetic variation at four X-linked loci in natural populations of the European rabbit, *O. cuniculus*. At each locus, we observed a deep split in the phylogeny with the root lying along the long internal branch. This pattern is consistent with the evolution of each subspecies in allopatry and subsequent secondary contact. Despite this broad similarity among loci, we detected heterogeneity among loci in terms of levels of nucleotide polymorphism, recombination, and introgression between the two subspecies. This heterogeneity corresponds well with the physical location of the loci on the rabbit X chromosome. The two centromeric loci had lower levels of nucleotide polymorphism, higher levels of LD and reduced introgression in comparison to the two telomeric loci. Although we do not have direct estimates of the frequency of crossing-over in rabbits, these observations are consistent with suppression of recombination near the centromere, as has been observed in other species (e.g. KONG *et al.* 2002).

Levels and patterns of variation: Across the entire sample (i.e. including both subspecies), the average heterozygosity among all loci (π=0.76%) was high and roughly one order of magnitude higher than heterozygosity at X-linked loci in humans (π= 0.081%; HAMMER *et al.* 2004) and mice (π=0.078%; NACHMAN 1997). Clearly, this high level of nucleotide variability reflects not only nucleotide polymorphism within each subspecies but also the divergence between subspecies. One gene (*Msn*) showed no introgression between subspecies. Levels of nucleotide polymorphism at this gene were 0.14% for *O. c. cuniculus* and 0.26% for *O. c. algirus*, closer to values observed in humans (HAMMER *et al.* 2004) and mice (NACHMAN 1997).

We also observed variation in levels of polymorphism among loci. Interestingly, the two centromeric loci had lower levels of π and θ than observed at the two telomeric loci, both for the entire dataset and for each subspecies considered separately. Within each subspecies, this difference may be explained by different levels of introgression. In other words, *Smcx* and *Msn* may be less variable within each subspecies because they contain relatively few introgressed haplotypes, compared to *Phka2* and *Hprt1*. However, we also observe less variation at *Smcx* and *Msn* in the total sample. This may be due in part to lower mutation rates at these genes. For example, divergence between *Oryctolagus* and *Lepus* is lower at *Smcx*

(Dxy=1.74%) and *Msn* (Dxy=4.08%) than at *Phka2* (Dxy=6.06%) or *Hprt1* (Dxy=4.43%) (Table 2). If recombination is suppressed near the centromere, these differences in mutation rate may reflect an association between mutation and recombination (e.g. HELLMANN *et al.* 2003). It is also possible that reduced variation at *Smcx* and *Msn* may be due partly to the effect of either positive or negative selection at linked sites (MAYNARD-SMITH and HAIGH 1974; CHARLESWORTH *et al.* 1993).

Indirect evidence that recombination is suppressed near the centromere comes from our observation of increased LD at *Smcx* and *Msn* compared to *Phka2* and *Hprt1*. Patterns of LD are affected by many factors, including selection, mutation, recombination, and changes in population size (e.g. ARDLIE *et al.* 2002). However, in humans, there is good evidence that levels of LD are inversely correlated with recombination rate over much of the genome (e.g. REICH *et al.* 2001; MCVEAN *et al.* 2004; MYERS *et al.* 2005). Moreover, in many organisms, recombination is suppressed near the centromeres, particularly in metacentric chromosomes (e.g. KONG *et al.* 2002). Thus, our observation of increased LD at *Smcx* and *Msn* relative to *Phka2* and *Hprt1* is consistent with, but not proof of, reduced recombination near the rabbit X centromere.

Divergence and gene flow between subspecies: RFLP surveys of mtDNA polymorphism in the Iberian Peninsula and France have shown that *O. cuniculus* is composed of two deeply divergent mtDNA lineages that are thought to have diverged about 2 MYA (BIJU-DUVAL *et al.* 1991; BRANCO *et al.* 2000). A survey of nucleotide variability at *Sry* also found evidence for the existence of two divergent lineages in the Y chromosome (GERALDES *et al.* 2005). These two lineages are associated with *O. c. algirus* and *O. c. cuniculus* (BRANCO *et al.* 2000), and are thought to have evolved in allopatry. Our X chromosome data confirm the existence of two divergent evolutionary units in *O. cuniculus*, and we show that in general the data reject the evolution of the two lineages under panmixia. The divergence time estimated from these loci is in good agreement with divergence time estimated from mitochondrial genes, and places the origin of these two subspecies at the Pliocene/Pleistocene boundary. We observed high levels of population differentiation at the two centromeric loci, but not at the telomeric loci. At the centromeric loci the two divergent lineages correspond well with the described subspecies and are broadly concordant with the patterns of differentiation seen at the

Y chromosome and at the mtDNA. The same was not observed at the two telomeric loci, where geography and phylogeny are largely decoupled.

If two populations evolve in allopatry for a sufficiently long time and then come into secondary contact with little or no gene flow, a high percentage of fixed differences and a small number of shared polymorphisms are expected. In our data this is seen only at *Msn* where 36% of all polymorphisms correspond to fixed differences between groups, and 6% correspond to shared polymorphisms. At all other loci, there are no fixed differences between subspecies and the percentage of shared polymorphisms varies from 16% at *Smcx* (centromeric) to 64% at *Hprt1* (telomeric). This heterogeneity among loci is also reflected in the rejection of an isolation without gene flow model using the HKA test between *Phka2* and *Msn* and the rejection of the null model using the WH test between *Phka2* and *Msn*, and between *Hprt1* and *Msn*.

It is noteworthy that the patterns of reduced introgression seen at *Smcx* and *Msn*, which may experience reduced recombination, are similar to the patterns seen previously at the mtDNA (BRANCO *et al.* 2000) and the Y chromosome (Geraldes, unpublished data), genomic regions with no recombination. In contrast, a survey of 14 allozyme loci revealed higher, but variable, levels of introgression (Fst between subspecies ranged from 0 to 0.46), comparable to the patterns observed at X chromosome loci. Differences among loci in levels of introgression have also been documented in other organisms. For example, genomic regions with suppressed recombination as a result of chromosomal rearrangements introgress less than co-linear regions in comparisons between *Drosophila pseudoobscura* and *D. persimilis* (NOOR *et al.* 2001; MACHADO *et al.* 2002) and between hybridizing sunflowers of the genus *Helianthus* (RIESEBERG *et al.* 1999). Based on such observations, NOOR *et al.* (2001) and RIESEBERG (2001) have argued that chromosomal rearrangements may promote speciation, not through underdominance directly as in traditional models (e.g. WHITE 1978), but by suppressing recombination and thereby extending the effects of isolation genes to linked sites. Our finding of low levels of introgression in an area of high LD near the X chromosome centromere of the rabbit is consistent with similar observations in fruit flies and sunflowers. Similarly, in *Anopheles* mosquitoes, two (out of three) areas of reduced introgression map to centromeres (TURNER *et al.* 2005).

Our observations also have some interesting parallels with studies of hybridization in the house mice, *Mus musculus* and *M. domesticus*. In the house mouse hybrid zone in Western Europe, the Y chromosome shows reduced introgression (VANLERBERGHE *et al.* 1986; TUCKER *et al.* 1992; DOD *et al.* 1993), and the X chromosome shows lower levels of introgression than do the autosomes (TUCKER *et al.* 1992; DOD *et al.* 1993; MUNCLINGER *et al.* 2002), although there is also considerable variability in levels of introgression among loci on the X chromosome (PAYSEUR *et al.* 2004). In a similar fashion, we observe some X linked loci with much reduced introgression in rabbits, providing further support for the importance of the X chromosome in reproductive isolation. Interestingly, the differences among loci in migration estimates (Table 6), may provide some clues to the nature of incompatibilities underlying reproductive isolation. In particular we note that estimates of the number of migrants from NE to SW for the centromeric loci are slightly lower than in the opposite direction. This is in agreement with the expected asymmetric behavior of young Dobzhansky-Muller interactions (ORR 1995) and suggests that incompatibilities may derive from interactions between the *cuniculus* X chromosome and an *algirus* genetic background.

One ultimate goal of speciation studies is to determine the identity of genes involved in reproductive isolation between nascent species. With the completion of the sequence of the rabbit genome expected in the next few years, it may soon be possible to identify candidate genes for reproductive isolation in this species. The results presented here suggest that some of these genes may lie near the centromere of the X chromosome.

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (SFRH/BD/4621/2001 PhD grant to AG and Research Project POCTI/BSE/40280/2001, and by a National Science Foundation grant to MWN. We thank R. Villafuerte for help in collecting rabbit samples, J. Good and C. Pinho for valuable discussions, D. Garrigan, T. Salcedo and M. Dean for help with ms simulations, and J. Hey for help with IM. We would also like to thank two anonymous reviewers for suggestions on a previous version of this manuscript.

References

ARDLIE, K. G., L. KRUGLYAK and M. SEIELSTAD, 2002 PATTERNS of linkage disequilibrium in the human genome. Nat. Rev. Genet. **3:** 299-309.

BANDELT, H. J., P. FORSTER, and A. ROHL, 1999 Median-Joining networks for inferring intraspecific phylogenies. Mol. Bio. Evol. **16:** 37-48

BARBASH , D. A., D. F. SIINO, A. M. TARONE and J. ROOTE, 2003 A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. Proc. natl. Acad. Sci. USA **100:** 5302-5307.

BATESON, W., 1909 Heredity and variation in modern lights, pp. 85-101 in *Darwin and modern science*, edited by A. C. Seward. Cambridge Univ. Press, Cambridge, U.K.

BIJU-DUVAL, C., H. ENNAFAA, N. DENNEBOUY, M. MONNEROT, F. MIGNOTTE *et al.*, 1991 Mitochondrial DNA evolution in Lagomorphs: origin of systematic heteroplasmy and organization of diversity in European rabbits. J. Mol. Evol. **33:** 92-102.

BRANCO, M., N. FERRAND and M. MONNEROT, 2000 Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analyses of the cytochrome *b* gene. Heredity **85:** 307-317.

BRANCO, M., M. MONNEROT, N. FERRAND and A. R. TEMPLETON, 2002 Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA variation. Evolution **56:** 792-803.

CHANTRY-DARMON, C., C. ROGEL-GAILLARD, M. BERTAUD, C. URIEN, M. PERROCHEAU *et al.*, 2003 133 new gene localizations on the rabbit cytogenetic map. Cytogenet. Genome Res. **103:** 192-201.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289-1903.

COYNE, J. A. and H. A. ORR, 1989 Two rules of speciation, pp. 180-207 in *Speciation and its consequences*, edited by D. Otte and J. Endler. Sinauer Associates, Sunderland, MA.

COYNE, J. A. and H. A. ORR, 2004 Speciation. Sinauer Associates, Sunderland, MA.

DAVIDSON, J. J., T. OZCELIK, C. HAMACHER, P. J. WILLEMS, U. FRANCKE, *et al.*, 1992 cDNA cloning of a liver isoform of the Phosphorylase-kinase alpha-subunit and mapping of the gene to Xp22.2-p22.1, the region of human X-linked liver glycogenosis. Proc. Natl. Acad. Sci. USA **89:** 2096-2100.

DOBZHANSKY, T. H., 1936 Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics **21:** 113-135.

DIETRICH, W. F., J. C. MILLER, R. G. STEEN, M. MERCHANT, D. DAMRON *et al.*, 1994 A genetic map of the mouse with 4,006 single sequence length polymorphisms. Nature Genetics

DOD, B., L. S. JERMIIN, P. BOURSOT, V. H. CHAPMAN, J. TONNES-NIELSEN *et al.*, 1993 Counterselection on sex-chromosomes in the *Mus musculus* European hybrid zone. J. Evol. Biol. **6:** 529-546.

EXCOFFIER, L., P. E. SMOUSE and J. M. QUATTRO, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to Human mitochondrial DNA restriction data. Genetics **131:** 479-491

FERRAND, N. and M. BRANCO, 2006 The evolutionary history of the European rabbit (*Oryctolagus cuniculus*), pp. 207-235, in *Phylogeography of Southern European refugia*, edited by S. Weiss and N. Ferrand. Springer, Amsterdam, Netherlands.

FU, Y. X. and W. H. LI, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 696-709.

GARRIGAN, D., Z. MOBASHER, S. B., KINGAN, J. A. WILDER AND M. F. HAMMER, 2005 Deep haplotype divergence and long-range linkage disequilibrium at Xp21.1 provide evidence that humans descended from a structured ancestral population. Genetics **170:** 1849-1856.

GERALDES, A., C. ROGEL-GAILLARD and N. FERRAND, 2005 High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) SRY gene. Anim. Genet. **36:** 349-351.

GRULA, J. W. and O. R. TAYLOR, 1980 Some characteristics of hybrids derived from the sulphur butterflies, *Colias eurytheme* and *Colias philodice*: phenotypic effects of the X chromosome. Evolution **34:** 673-687.

HAGEN, R. H., 1990 Population structure and host use in hybridizing subspecies of *Papilio glaucus* (Lepidoptera: Papilionidae). Evolution **44:** 1914-1930.

HALDANE, J. B. S., 1922 Sex-ratio and unisexual sterility in hybrid animals. J. Genet. **12:** 101-109.

HALL, T. A., 1999 BioEdit: a user friendly biological sequence alignment editor and analyses program for Windows 95/98/NT. Nucl. Acids Symp. Ser. **41:** 95-98.

HAMMER, M. F., D. GARRIGAN, E. WOOD, J. A. WILDER, Z. MOBASHER *et al.*, 2004 Heterogeneous patterns of variation among multiple X-linked loci: the possible role of diversity-reducing selection in Non-Africans. Genetics **167:** 1841-1853.

HAYES, H., C. ROGEL-GAILLARD, C. ZIJLSTRA, N. A. DE HAAN, C. URIEN *et. al.*, 2002 Establishment of an R-banded rabbit karyotype nomenclature by FISH localization of 23 chromosome-specific genes on both G- and R-banded chromosomes. Cytogenet. Genome. Res. **98:** 199-205.

HELLMAN, I., I. EBERSBERGER, S. E. PTAK, S. PAABO and M. PRZEWORSKI, 2003 A neutral explanation for the correlation of diversity with recombination rates in humans. Am. J. Hum. Genet. **72:** 1527-1535.

HEY, J. and R. M. KLIMAN, 1993 Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Bio. Evol. **10:** 804-822.

HEY, J. and J. WAKELEY, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833-846.

HEY, J. and R. NIELSEN, 2004 Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis.* Genetics **167:** 747-760.

HILL, W. G. and A. ROBERTSON, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226-231.

HUDSON, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245-250.

HUDSON, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18:** 337-338.

HUDSON, R. R. and N. L. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147-164.

HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992 A statistical test for detecting geographic subdivision. Mol. Bio.Evol. **9:** 138-151.

HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153-159.

HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583-589.

JENSEN-SEAMAN, M. I., T. S. FUREY, B. A. PAYSEUR, Y. LU, K. M. ROSKIN *et al.*, Comparative recombination rates in the Rat, Mouse and Human genomes. Genome Research **14:** 528-538.

LEWONTIN, R. C., 1964 Interaction of selection + Linkage. I. General considerations - Heterotic models. Genetics **49:** 49-67.

KONG, A., D. F. GUDBJARTSSON, J. SAINZ, G. M. JONSDOTTIR, S. A. GUDJONSSON *et al.*, 2002 A high-resolution recombination map of the human genome. Nat. Genet. **31:** 241-247.

MACHADO, C. A., R. M. KLIMAN, J. A. MARKERT and J. HEY, 2002 Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. Mol. Bio. Evol. **19:** 472-488.

MALITSCHEK, B., D. FORNZLER and M. SCHARTL, 1995 Melanoma formation in *Xiphophorus*: a model system for the role of receptor tyrosine kinases in tumorigenesis. Bioessays **17:** 1017-1023.

MATTHEE , C. A., B. J. VAN VUUREN, D. BELL AND T. J. ROBINSON, 2004 A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. Syst. Biol. **53**: 433-447.

MAYNARD-SMITH, J. and J. HAIGH, 1974 The hitch-hiking effect of a favorable gene. Genet. Res. **23:** 23-35.

MCVEAN, G. A. T., S. R. MYERS, S. HUNT, P. DELOUKAS, D. R. BENTLEY and P. DONNELLY, 2004 The fine-scale structure of recombination rate variation in the human genome. Science **304:** 581-584.

MULLER, H. J., 1940 Bearing of the *Drosophila* work on systematics, pp.185-268 in *The new systematics*, edited by J. S. Huxley. Clarendon press, Oxford, U.K.

MULLER, H. J., 1942 Isolating mechanisms, evolution and temperature. Biol. Symp. **6:** 71-125.

MUNCLINGER, P., E. BOZIKOVA, M. SUGERKOVA, J. PIALEK and M. MACHOLAN, 2002 Genetic variation in house mice (Mus, Muridae, Rodentia) from the Czech and Slovak Republics. Folia Zool. **51:** 81-92.

MYERS, S., L. BOTTOLO, C. FREEMAN, G. MCVEAN and P. DONNELLY, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science **310:** 321-324.

NACHMAN, M. W., 1997 Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics **147:** 1303-1316.

NEI, M., 1987 Molecular Evolutionary Genetics. Columbia University Press, New York.

NEI, M. and W. H. LI, 1979 Mathematical model for studying genetic-variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269-5273.

NOOR, M. A. F., K. L. GRAMS, L. A. BERTUCCI and J. REILAND, 2001 Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA. **98:** 12084-12088.

ORR, H. A., 1995 The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics **139:** 1805-1813.

PAYSEUR, B. A., J. G. KRENZ and M. W. NACHMAN, 2004 Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution **58:** 2064-2078.

POSADA, D. and K. A. CRANDALL, 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14**: 817–818.

POSADA, D. and T. R . BUCKLEY, 2004 Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over Likelihood ratio tests. Syst. Biol. **53**: 793–808.

PRESGRAVES, D. C., 2003 A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. Genetics **163:** 955-972.

PRESGRAVES, D. C., L. BALAGOPALAN, S. M. ABYMAYR and H. A. ORR, 2003 Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature **423:** 715-719.

REICH , D. E., M. CARGILL, S. BOLK, J. IRELAND, P. C. SABETI *et al.*, 2001 Linkage disequilibrium in the human genome. Nature **411:** 199-204.

RIESEBERG, L. H., J. WHITTON and K. GARDNER, 1999 Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics **152:** 713-727.

RIESEBERG, L. H., 2001 Chromosomal rearrangements and speciation. Trends Ecol. Evol. **16:** 351-358.

ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**: 2496-2497.

SAMBROOK, J. and D. W. RUSSEL, 2001 *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

SCHNEIDER, S., D. ROESSLI and L. EXCOFFIER, 2000 Arlequin: a software program for population genetics data analysis. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

SPERLING, F. A. H. and J. R. SPENCE, 1991 Structure of an asymmetric hybrid zone between two water strider species (Hemiptera: Gerridae: Limnoporus). Evolution **45:** 1370-1383.

SWOFFORD, D. L., 2002 PAUP* ver 2.0.b10. Phylogenetic Analysis Using Parsimony and other methods. Sinauers Associates, Sunderland, Sunderlan MA.

TAJIMA, F., 1989 Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585-595.

TAMURA, K. and M. NEI, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Bio. Evol. **10:** 512-526.

TAO, Y., S. CHEN, D. L. HARTL and C. C. LAURIE, 2003 Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana.* I. Differential accumulation of hybrid male sterility effects on the X and autosomes. Genetics **164:** 1383-1397.

TING, C. T., S. C. TSAUR, M. L. WU and C. I. WU, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282:** 1501-1504.

TRUE, J. R., B. S. WEIR and C. C. LAURIE, 1996 A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. Genetics. **142:** 819-837.

TUCKER, P. K., R. D. SAGE, J. WARNER, A. C. WILSON and E. M. EICHER, 1992 Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. Evolution **46:** 1146-1163.

TURELLI, M. and H. A. ORR, 1995 The dominance theory of Haldane's rule. Genetics **140:** 389-342.

TURELLI, M. and H. A. ORR, 2000 Dominance, epistasis and the genetic of postzygotic isolation. Genetics **154:** 1663-1679.

TURNER, T. L., M. W. HAHN and S. V. NUZHDIN, 2005 Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3: e285.

VANLERBERGHE, F., B. DOD, P. BOURSOT, M. BELLIS and F. BONHOMME, 1986 Absence of Y chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. Genet. Res. **48:** 191-197.

WANG, R. L., J. WAKELEY and J. HEY, 1997 Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics **147:** 1091-1106.

WAKELEY, J. and J. HEY, 1997 Estimating ancestral population parameters. Genetics **145:** 847-855.

WALL, J. D., 2000 Detecting ancient admixture in Humans using sequence polymorphism data. Genetics **154:** 1271-1279.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256-276.

WHITE, M. J. D., 1978 *Modes of speciation*. W. H. Freeman, San Francisco.

WON, Y. J. and J. HEY, 2005 Divergence population genetics of chimpanzees. Mol. Bio. Evol. **22:** 297-307.

WRIGHT, S., 1951 The genetical structure of populations. Ann. Eugenics **15:** 323-354.

Supplemental Table 1- X-chromosome loci details, PCR and Sequencing Primers[a]

Primer no.	Primer name	Type[b]	Primer sequence (5'-3')	Temp.[c]
1	Phka2e2f	Pcr	CCTGGGTGCGAGACAACATCTACAG	
2	Phka2e4r	Pcr	CCAGGAACAGGAGAAAGAGGGAGGT	66°
3	NPhka2e2f	NPcr	CATGGCCTACCGCAAGAATG	
4	NPhka2e4r	NPcr	GCAGGTGGCAGTGTTGTACTT	62°
5	P730r	S	TTTCGAGGCTCCCTACTTCA	
6	P640f	S	GATTCCCACCAAAGCAAAGA	
7	P1900f	S	ACCTCCTCCCCTCAACCTTA	
8	P2015r	S	TGTGGAAAACACAACCTGGA	
9	P2630f	S	TAGAGGGTCACAGTCTCGGG	
10	SPhka2_e2f2	S	TCAGAAGCCAAGATTCCCAC	
11	Sphka2_e3r	S	CCTCATCATGCACTGGAGAA	
12	Sphka2_e3f	S	GAAGCTGATGCGAGGTCTTC	
13	Sphka2_e3r2	S	GCACATTACTTCCCGTGTCC	
14	Sphkf	S	ACGAACACCTCACCTGCTTAT	
15	Slehf	S	GCCATCCTCCACTGCACTC	
16	SP2919r	S	ATCAGTGTGTTCTTTGCCCC	
17	SP851f	S	TGTCAGGTCTAGGAAGAGT	
18	LP2000r	S	CTCCTGCACCAGCCCGCC	
19	SP1410r	S	TCACTGGGAGATGGCCACGT	
20	Smcxe2f	Pcr	GGACTGGCAGCCACCCTTTGCTGTG	
21	Smcxe4r	Pcr	CACAAGGTTGGCTCCAGACTGGTACA	64°
22	NSmcxe2f	NPcr/S	GTAACAGGGTTGGGAAAACGGATACC	
23	Nsmcxe4r	NPcr/S	TGTTTTGCCTGGTGGATAGTTGAGG	62°
24	SSmcxf2	S	GGGAAGTGAGAGTGAGTGCC	
25	SS1580r	S	TTCACTTCCTCATGCACCAG	
26	Ssmcxr2	S	TCTCCAAATGAACCGAAACC	
27	Msne4f2	Pcr	CCCAGCGCCTGTTCTTTCTGCAAG	
28	Msne7r3	Pcr/S	CCTGATTTCACTCCAAGGGAAGCC	62°
29	M105f	NPcr	GGTCTGTGCCTCTCCATTAGGG	
30	M3226r	NPcr/S	TGGTTAGTTGCCAACACTGAGGG	62°
31	M847r	S	ACCCAGATCTACTACACTGT	
32	M769f	S	TAGGGACAGTAAGCTGTGC	
33	M1513f	S	ACAGAGATGATGCTGGCACT	
34	M2099f	S	TATCTGAAGATTGCTCAAGA	
35	M3300r	S	AATAGAATAGAGTAAGAGT	
36	Msne4f3	S	TCAACAAGGAAGTGCACAAGTGTGG	
37	Msne5f1	S	CACAAACTCAACAAGGACCAGTGGG	
38	Msne6r1	S	ATTCTGCTCATAGATGTTGAGACCC	
39	Smsnf	S	AGTGAACTACTTCAGCATTA	
40	Smsnr	S	AGTGCCAGCATCATCTCTGT	
41	Hprt11e2f	Pcr	GCGATGATGAACCAGGTTATGACC	
42	Hprt11e3r	Pcr	TAGCTCTTCAGTCTGATAAAATCTACAG	64°
43	NpHprt1_f	NPcr/S	ATTACGTCGAGGACTTGGAAAGG	
44	NpHprt1_r	NPcr/S	AGCAGGTCAGCAAAGAACTTATAGCC	68°
45	Hprt1r2	S	GGATTCTAGTCCCGGTTGCT	
46	Sh654r	S	TCCAATGAATGTTAGAACTA	
47	Sh520f	S	CTCTCCTCAGCTTCCTGTA	

[a]Primers 1 through 19 were used for the region between exons two and four of *Phka2* [Phosphorylase kinase alpha 2 (liver)] located on *O. cuniculus* Xp15 (CHANTRY-DARMON *et al.* 2003). Primers 20-26 were used for the region between exons two and four of *Smcx* [Smcy homolog, X-linked (mouse)] located on *O. cuniculus* Xp11 (CHANTRY-DARMON *et al.* 2003). Primers 27 through 40 were used in the study of the region between exons six and nine of *Msn* (Moesin) located on *O. cuniculus* Xq12prox (CHANTRY-DARMON *et al.* 2003). For this gene the number of exons in the three species compared is not the same. The primers used for the amplifications are in exons six and nine of humans, which correspond to exons three and six in the house mouse and exons four and seven in the rat. Finally, primers 41 through 47 were used for the region between exons two and three of *Hprt1* [Hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome)] located on *O. cuniculus* Xq23 (HAYES *et al.* 2002).

[b]Purpose for which the primers were used: PCR- primers used for PCR, NPCR- primers used for Nested PCR and S- primers used for most sequencing reactions, SA- alternative sequencing primers used to sequence individuals in which the S primers did not work.

[c]Annealing temperature used in the respective PCR or Nested PCR reactions.

4.2 Paper V

Submitted to Evolution

Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula

A Geraldes[*,†,‡], M W Nachman[‡] and N Ferrand[*,†]

[*]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

[†]Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal

[‡]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721, USA

Abstract

The role of the Y chromosome in speciation is unclear. Hybrid zones provide natural arenas for studying speciation, as differential introgression of markers may reveal selection acting against incompatibilities. Two subspecies of the European rabbit (*Oryctolagus cuniculus*) form a contact zone in the Iberian Peninsula, and these subspecies are well differentiated with respect to mtDNA. Autosomal markers show varying degrees of introgression between subspecies. Here we explore the role of the Y chromosome by genotyping two diagnostic Y chromosome single nucleotide polymorphisms (SNP) in a sample of 353 male rabbits representing both subspecies. We found that Y chromosome lineages are largely subspecies-specific, suggesting that selection may be preventing Y chromosome introgression following secondary contact.

Keywords: European rabbit, Y chromosome, hybrid zone, gene flow, RFLP, speciation

Introduction

The genetic basis of reproductive isolation is a key problem in evolutionary biology and it has been studied both with laboratory crosses and with natural hybrid populations. Laboratory crosses have the advantage of providing a controlled setting, and wild populations have the advantage of providing more realistic biological conditions for detecting fitness differences. Both kinds of studies have provided abundant evidence for the importance of the X chromosome in reproductive isolation in taxa in which the male is the heterogametic sex, such as *Drosophila* and mammals (Coyne and Orr 2004). For example, Haldane's rule appears to be largely a consequence of epistatic interactions involving recessive X-linked mutations (Turelli and Orr 1995).

The role of the Y chromosome in speciation is far less clear (Coyne and Orr 2004). In crosses involving species in the *Drosophila virilis* group, for example, some Y-chromosome introgression experiments reveal a strong phenotypic effect, while others do not (Orr and Coyne 1989). In house mice, primary sex determination is disrupted in consomic strains in which the *Mus domesticus* Y chromosome is introduced onto the genetic background of some laboratory strains containing *M. musculus* alleles (Eicher and Washburn 1986; Washburn et al. 2001). The phenotypic effect depends on the genetic composition of both the *M. domesticus* Y chromosome and of the laboratory strains; in some cases, males show complete sterility, while in other cases males are fully fertile. In two different transects of the *M. musculus – M. domesticus* hybrid zone, the Y chromosome introgresses less than other chromosomes (Vanlerberghe et al. 1986; Tucker et al. 1992). However, in another transect, the Y chromosome shows a fair amount of introgression (Munclinger et al. 2002). Remarkably, the role of the Y chromosome in reproductive isolation has not been studied in other mammalian systems.

The European rabbit (*Oryctolagus cuniculus*) provides an opportunity to study the Y chromosome in the context of hybridizing taxa. This species is native to the Iberian Peninsula and has two recognized subspecies, *O. c. algirus* in the southwest and *O. c. cuniculus* in the northeast. The two subspecies form a contact zone that runs in a northwest/southeast direction. These taxa are characterized by two divergent mtDNA lineages (11.9% nucleotide divergence based on *cytb* RFLP data; Branco et al. 2000), suggesting that they diverged approximately 2 million years ago.

Branco et al. (2002) suggested that the subspecies have recently come into secondary contact following Pleistocene climate changes. Analyses of allozyme (Branco and Ferrand 1998, 2002, 2003; Branco et al. 1999; Ferrand and Branco 2005) and immunoglobulin (van der Loo et al. 1991; van der Loo et al. 1999; Esteves et al. 2004) variability in the Iberian Peninsula are consistent with the mtDNA data in showing two major groups. However, the level of differentiation between these groups is generally low for these markers, suggesting that these autosomal genes have introgressed extensively following secondary contact. Geraldes et al. (2005) sequenced approximately 2 Kb of the Y chromosome, including the *Sry* gene, in four wild and eight domestic (*O. c. cuniculus*) rabbits and found two divergent lineages separated by seven nucleotide differences (0.40% average divergence). One lineage was present in *O. c. cuniculus* (represented by the eight domestic rabbits and two individuals from the northeast Iberian Peninsula), and the other lineage was present in *O. c. algirus* (represented by two individuals from the southwest Iberian Peninsula).

Here, we investigate geographic variation in the Y chromosome in a large sample of rabbits representing both subspecies. We genotyped two diagnostic Y-specific single nucleotide polymorphisms (SNPs) in a sample of 353 male rabbits from the Iberian Peninsula and south of France. Our results show that levels of introgression for the Y chromosome are very low, suggesting that selection has acted to prevent Y chromosome gene flow between subspecies of rabbits.

Materials and Methods

Sampling. We sampled 353 male rabbits from 30 natural populations. Populations were divided into three groups (Table 1): southwest Iberian Peninsula (SW), corresponding to the distribution of the subspecies *O. c. algirus*; northeast Iberian Peninsula and France (NE), where *O. c. cuniculus* occurs; and the contact zone (CZ), as defined previously by Branco and colleagues (2000) based on mtDNA. The approximate geographic location of the populations is shown in Figure 1.

Molecular methods. Genomic DNA was extracted from blood, liver, kidney or muscle following standard protocols (Sambrook and Russell 2001). PCR was performed on two different fragments of the *Sry* gene region described by Geraldes et al. (2005). The primers TSPYF267 (5'- GCA AAG CTG TGA TTT TCA AAG GC - 3') and TSPYR716 (5'-GTA TTG CAC TGG TGG TTT GTG C-3') were used to

amplify a 450bp fragment, with 35 cycles of 25 s at 94 °C, 25 s at 59 °C and 25 s at 72 °C, preceded by an initial denaturation step at 94 °C for 2 min, and followed by a final extension of 5 min at 72 °C. The primers MAEYF1086 (5'- GCA GCT AAT CTG CTC ACA GCC -3') and MAEYR1376 (5'-AAC AAT CAT ACC CAT TGG TCG AG-3') were used to amplify a 291bp fragment using the PCR conditions above but with an annealing temperature of 58 °C. For both assays, primer names indicate their location in the *Sry* sequence (GenBank AY785433).

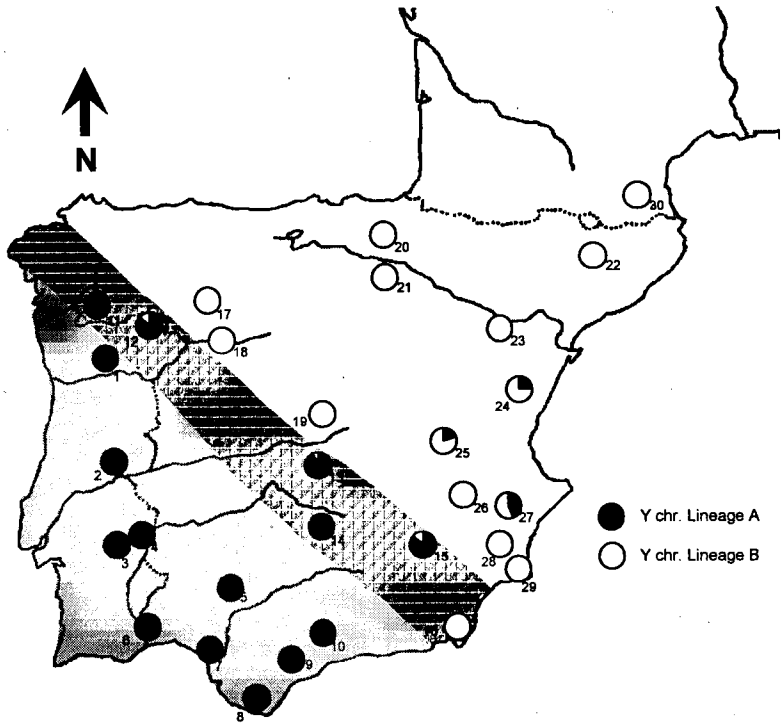


Figure 1- Geographical distribution of Y chromosme lineages A and B. Pie charts indicate the frequency of each lineage in the population. Grey indicates the range of SW populations, stripped pattern indicates CZ populations and white NE. Sample sizes for each population are indicated in Table 1. 1- Vila Real, 2- Idanha, 3- Vila Viçosa, 4- Elvas, 5- Sevilla, 6- Huelva, 7- Doñana, 8- Las Lomas, 9- Fuente Piedra, 10- Córdoba, 11- Verin, 12- Bragança, 13- Toledo, 14- Ciudad Real, 15- Albacete SW, 16- Las Amoladeras, 17- Benavente, 18- Zamora, 19- Madrid, 20- Tudela, 21- La Rioja, 22- Lleida, 23- Zaragoza, 24- Rosell, 25- Cuenca, 26- Albacete N, 27- Valencia, 28- Alicante, 29- Cartagena and 30- Perpignan.

Genotyping was performed using restriction fragment length polymorphisms (RFLP). The first PCR fragment was digested with the restriction enzyme *Tsp*509I (New England Biolabs, MA, USA), and the second PCR fragment was digested with the restriction enzyme *Mae*III (Roche Diagnostics, CA, USA). Restriction maps for both fragments were generated using the BioEdit software (Hall 1999). *Tsp*509I either cuts the 450bp fragment of the *Sry* gene three times, at nucleotides 319, 516 and 596,

producing four different fragments of 53, 197, 80 and 120bp (profile *Tsp*509I-A), or two times, at nucleotides 319 and 516, producing three different fragments of 53, 197, 200 bp (profile *Tsp*509I-B). *Mae*III was used to digest the other fragment of *Sry* (291 bp), producing either one cut at nucleotide 1177 with two fragments of 92 and 199 bp (profile *Mae*III-A), or two cuts at nucleotides 1177 and 1233 with fragments of 92, 56 and 143 bp (profile *Mae*III-B). Restriction products were visualised by silver staining after nondenaturing electrophoresis separation in 9% polyacrylamide gels.

Data Analysis. Haplotype frequencies were calculated for each population. To test for population structure among groups of populations, analyses of molecular variance AMOVA (Excoffier et al. 1992) between the SW and NE groups were performed using ARLEQUIN (Schneider et al. 2000). The same analysis was performed for the mtDNA data set from Branco and colleagues (2000), for 14 allozyme loci (Ferrand and Branco 2005) and for nine microsatellite loci reported by Queney et al. (2001).

Results

PCR and restriction digests were performed in all 353 samples and no novel restriction profiles were observed. The restriction profile *Tsp*509I-A was always associated with the restriction profile *Mae*III-A, indicating complete linkage disequilibrium between the mutations at *Sry* nucleotides 596 and 1233. In the total sample, 203 individuals had the A profile and 150 individuals had the B profile, referred to below as lineage A and lineage B. The frequency of lineage A in the different populations is shown in Table 1 and its geographic distribution is shown in Figure 1.

Table1- Frequency of Lineage A in each of the populations studied

		Total	Lineage A	
Population Group	Population nr.	n	n	Frequency
NE	17	21	0	0
NE	18	14	0	0
NE	19	17	0	0
NE	20	8	0	0
NE	21	4	0	0
NE	22	7	0	0
NE	23	9	0	0
NE	24	8	2	0.25
NE	25	5	1	0.2
NE	26	8	0	0
NE	27	9	4	0.44
NE	28	18	0	0
NE	29	10	0	0
NE	30	11	0	0
CZ	11	7	7	1
CZ	12	7	6	0.86
CZ	13	32	30	0.94
CZ	14	18	18	1
CZ	15	8	7	0.97
CZ	16	4	0	0
SW	1	11	11	1
SW	2	18	18	1
SW	3	12	12	1
SW	4	7	7	1
SW	5	27	27	1
SW	6	11	11	1
SW	7	8	8	1
SW	8	10	10	1
SW	9	14	14	1
SW	10	10	10	1

*Population numbers from Figure 1.

All ten populations in the SW group were fixed for lineage A. Of the 11 populations in the NE group, eight were fixed for lineage B and three were polymorphic, with frequencies of lineage A of 0.2, 0.25 and 0.44, respectively. In the contact zone two populations were fixed for linage A, one for lineage B, and three were polymorphic (frequency of lineage A was 0.68, 0.94 and 0.97 in these populations). Haplotype diversity was significantly partitioned between subspecies (ϕ_{CT}= 0.95 between SW and NE; ϕ_{CT}= 0.88 among SW, NE, and CZ). This reveals a high level of genetic differentiation between the two major groups. Thus, the two differentiated Y chromosome lineages are highly structured geographically and

probably reflect a long period of evolution of the two groups in isolation with recent contact following a range expansion.

Discussion

We observed substantial differentiation between subspecies of rabbit for the Y chromosome, with lineage A found predominantly in *O. c. algirus* and lineage B found predominantly in *O. c. cuniculus*. Several polymorphic populations were found in the contact zone, as well as in NE Spain. The latter are likely due to recent introductions associated with hunting. Similar patterns were observed for mtDNA (Branco et al. 2000). In general, however, the degree of Y chromosome introgression between subspecies is low. We also calculated ϕ_{CT} using the mtDNA data from (Branco et al. 2000) for the same geographic regions surveyed here and found that the two subspecies are similarly differentiated for the Y chromosome and for mtDNA (Table 2).

This stands in contrast to autosomal markers which show little (allozymes) or no differentiation (microsatellites) between NE and SW populations (Table 2). The lower levels of population differentiation observed at autosomal loci might have

Table 2- Values of population differentiation for different markers

	Φ_{CT}[1]	Reference
Y chromosome	**0.95**	*This study*
mtDNA	**0.87**	Branco *et al.*, 2000
Allozymes	**0.26**	Ferrand and Branco, 2006
Autosomal microsatellites	0.02	Queney *et al.*, 2001

[1] Φ_{CT} as calculated in Arlequin v2.000 (Schneider et al. 2000)

several different explanations. In the case of microsatellites, Queney et al. (2001) argued that their data revealed "severe allele-size homoplasy which vastly underestimates divergence between the main groups of populations in Iberia." This effect is specific to microsatellite loci and is due to their high mutation rate. The low average level of population differentiation at other autosomal loci is likely a consequence of either unsorted ancestral polymorphism or on-going introgression of these markers across the contact zone.

As two populations diverge, gene genealogies will typically proceed from polyphyly to paraphyly to reciprocal monophyly, and the rate at which this occurs will depend on population size (Avise 1994). Assuming a sex-ratio of one, the effective population size of mtDNA and the Y chromosome is 1/4 that of the

autosomes, and therefore the autosomes are expected to have a coalescence time four times greater than for mtDNA or the Y chromosome, although the variance in coalescence times is large (Hudson and Turelli 2003). In the European rabbit, the ratio of *cytb* mtDNA variation within subspecies (1.1 – 1.4%) to differences between subspecies (11.9%) is approximately 1:10 (Branco et al. 2000). This suggests that sufficient time has elapsed since the divergence of these subspecies for many autosomal and X-linked genes to have attained reciprocal monophyly. This is well supported by our recent sequencing of three autosomal and four X-linked genes; at each gene we observe two very divergent lineages, consistent with evolution in allopatry, although at some genes these divergent lineages have introgressed across the contact zone (Ferrand et al. unpublished data). Thus, it seems unlikely that the variation in ϕ_{CT} among markers in Table 2 can be explained mainly by differential lineage sorting.

Instead, the variation in ϕ_{CT} among markers is likely due to differential introgression of genes following secondary contact. In fact, the low ϕ_{CT} value presented for the allozymes is an average of 14 different loci that show a high degree of variation between them (range: 0 to 0.46). These results are consistent with a view of speciation in which different portions of the genome show different levels of genetic isolation (e.g. Rieseberg et al. 1999; Machado et al. 2002; Wu and Ting 2004). Evidence from many studies indicates that reproductive isolation between divergent taxa is often due to negative epistatic interactions (i.e. "Dobzhansky-Muller incompatibilities," Coyne and Orr 2004). Presumably, genomic regions that introgress less are those that contain genes involved in such incompatibilities. The data presented here support the idea that the Y chromosome harbors such genes.

Our finding of reduced introgression for the Y chromosome across the European rabbit hybrid zone is consistent with both laboratory experiments (Eicher and Washburn 1986; Washburn et al. 2001) and hybrid zone studies (Vanlerberghe et al. 1986; Tucker et al. 1992) involving house mice, the only other mammal for which the role of the Y chromosome in reproductive isolation has been studied. Our results are also consistent with the observation that the Y chromosome includes a disproportionate number of genes involved in male reproduction (Skaletsky et al. 2003). The lack of Y chromosome introgression between subspecies of rabbits thus provides further support for the hypothesis that incompatibilities involving the Y chromosome may be important in speciation.

Acknowledgements

This work was supported by a PhD grant to AG (SFRH/BD/4621/2001) and Research Project POCTI/BSE/40280/2001 from Fundação para a Ciência e a Tecnologia, and by an National Science Foundation grant to MWN. We thank R. Villafuerte for help in collecting rabbit samples. M. Branco, R. Campos, J. Good, C. Pinho, G. Wlasiuk and E. Wood made valuable comments to the manuscript.

References

Avise, J.C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York.

Branco, M., and Ferrand, N. 1998. Genetic polymorphism of rabbit (*Oryctolagus cuniculus*) tissue acid phosphatases (ACP2 and ACP3). Comp. Biochem. Physiol. 120B:405-409.

___. 2002. Genetic polymorphism of antithrombin III, haptoglobin, and haemopexin in wild and domestic European rabbits. Biochem. Genet. 40:387-393.

___. 2003. Biochemichal and population genetics of the rabbit, *Oryctolagus cuniculus*, carbonic anhydrases I and II, from the Iberian Peninsula and France. Biochem. Genet. 41:391-404.

Branco, M., Machado, J. C., and Ferrand, N. 1999. Extensive genetic polymorphism of peptidases A, B, C, and D, in wild rabbit (*Oryctolagus cuniculus*) populations from the Iberian Peninsula. Biochem. Genet. 37:237-249.

Branco, M., Ferrand, N., and Monnerot, M. 2000. Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome *b* gene. Heredity 85:307-317.

Branco, M., Monnerot, M., Ferrand, N., and Templeton, A. R. 2002. Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA variation. Evolution. 56:792-803.

Coyne, J. A., and Orr, H. A. 2004. Speciation. Sinauer Associates, Inc. Sunderland.

Eicher, E. M., and Washburn, L. L. 1986. Genetic control of primary sex determination in mice. Ann. Rev. Genet. 20:327-360.

Esteves, P. J., Lanning, D., Ferrand, N., Knight, K. L., Zhai S. K., and van der Loo, W. 2004. Allelic variation at the Vha locus in natural populations of rabbit (*Oryctolagus cuniculus,* L.). J. Imunnol. 172:1044-1053.

Excoffier, L., Smouse, P.E., Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to Human mitochondrial DNA rectriction data. Genetics. 131:479-491

Ferrand, N., and Branco, M. 2006. The evolutionary history of the European rabbit (*Oryctolagus cuniculus*): major patterns of population differentiation and geographic expansion inferred from protein polymorphism. *In* Phylogeography

in Southern European Refugia. Edited by S. Weiss and N. Ferrand. Springer Publishers. Amsterdam, The Netherlands.

Geraldes, A., Rogel-Gaillard, C., and Ferrand, N. 2005. High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) SRY gene. Anim. Genet. (in press)

Hall, T. A. 1999. BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41:95-98.

Hudson, R. R., and Turelli, M. 2003. Stochasticity overrules the "three times rule": genetic drift, genetic draft and coalescence times for nuclear loci versus mitochondrial DNA. Evolution. 57:182-190.

Machado C. A., Kliman R. M., Markert, J., Hey, J. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. Mol. Biol. Evol. 19: 472-488.

Munclinger, P., Bozikova, E., Sugerkova, M., Pialek, J., and Macholan, M. 2002. Genetic variation in house mice (*Mus*, Muridae, Rodentia) from the Czech and Slovak Republics. Folia Zool. 51:81-92.

Orr, H. A., and Coyne, J. A. 1989. The genetics of postzygotic isolation in the *Drosophila virilis* group. Genetics. 121:527-537.

Queney, G., Ferrand, N., Weiss, S., Mougel, F., and Monnerot, M. 2001. Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). Mol. Bio. Evol. 18:2169-2178.

Rieseberg, L. H.,Whitton, J., and Gardner, K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two wild sunflower species. Genetics. 152:713-727.

Sambrook J., and Russel, D. W. 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Schneider, S., Roessli, D., and Excoffier, L. 2000. Arlequin: a software program for population genetics data analysis. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hiller, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S. F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S. P., Waterston, R. H., Wilson, R. K., Rozen, S., Page, D. P. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 423:825-837.

Tucker, P. K., Sage, R. D., Warner, J., Wilson, A. C., Eicher, E. M. 1992. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. Evolution. 46:1146-1163.

Turelli, M., and Orr, J. C. 1995. The dominance theory of Haldane's rule. Genetics. 140:389-402.

van der Loo, W., Ferrand, N., and Soriguer R. C. 1991. Estimation of the gene diversity at the b locus of the constant region of the immunoglobulin light chain in natural populations of the European rabbit (*Oryctolagus cuniculus*) in Portugal, Andalusia and the Azorean Islands. Genetics. 127: 789-799.

van der Loo, W., Mougel, F., Sanchez, M. S., Bouton, C., Castien, E., Fonseca, A., Ferrand, N., Soriguer, R. and Monnerot, M. 1999. Cytonuclear disequilibria in wild populations of rabbit (*Oryctolagus cuniculus* L.) suggest unequal allele turnover rates at the b locus (IGKC1). Immunogenetics. 49:629-643.

Vanlerberghe, F., Dod, B., Boursot, P., Bellis, M., and Bonhomme, F. 1986. Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. Genet. Res., Camb. 48:191-197.

Washburn, L. L., Albrecht, K. H., and Eicher, E. M. 2001. C57BL/6J-T-associated sex reversal in mice is caused by reduced expression of a *Mus domesticus Sry* allele. Genetics. 158:1675-1681.

Wu, C. I., and Ting, C. T. 2004. Genes and speciation. Nat. Rev. Genet. 5:114-122.

5. GENERAL DISCUSSION

5.1 The evolutionary history of the European rabbit: a perspective based on sexual chromosome and mitochondrial markers

5.1.1 Evidence for a marked dichotomy in the sexual component of the rabbit genome

Fifteen years ago Biju-Duval *et al.* (1991) compared for the first time restriction patterns of rabbit mitochondrial DNA between southwestern Iberian populations and others. These authors estimated a divergence time between these molecules of 2 MY based on the 4% (uncorrected for multiple hits) nucleotide divergence found and a 2% divergence per MY average rate for mammalian mitochondria. Branco *et al.* (2000) found a much larger divergence (11.9% also uncorrected for multiple hits) between these mitochondrial lineages in an RFLP survey of the Cytb gene alone. They argued that this discrepancy could be due either to the different nucleotide sampling regimes employed or to rate heterogeneity of mitochondrial genes.

In this work we revisited this question by taking a slightly different approach. First, we surveyed six genetic loci instead of just one. Four of them are X-linked (PHKA2, SMCX, MSN and HPRT1), one is Y-linked (SRY) and one is cytoplasmic (Cytb). Second, we estimated divergences based not on restriction analyses but with sequence data. Third, we estimated nucleotide divergences corrected for multiple substitutions. Finally, we used the observed divergence to *Lepus granatensis* and the previously estimated divergence time of 11.8 MY (Matthee *et al.* 2004) between these species to calculate the rate of evolution of each locus instead of relying on general mammalian rates of evolution.

We showed that at each locus two divergent genetic lineages are observed (Figure 1). Two of the six loci studied (the two telomeric X-linked loci PHKA2 and HPRT1) show evidence for high levels of recombination. Although here two lineages are also observed, several recombinant haplotypes between the two lineages were detected. We performed coalescent simulations of neutral genealogies based on each of the four X-linked loci to ask whether levels of linkage disequilibrium (LD) observed conform to a scenario of evolution of these lineages under panmixia. For three of these loci we were able to reject this null hypothesis with statistical significance ($P<0.05$). For HPRT1 only marginal significance was obtained ($P=0.066$). Since the Y chromosome and the mitochondria are non-recombining the approach described above is not applicable to these loci, but from the genealogies alone it is clear that these two lineages are highly diverged and mostly subspecies

specific. This result implies that this dichotomy observed in the rabbit genome is likely the result of long term evolution of two populations in allopatry.

Assuming a molecular clock, we estimated the time since these populations

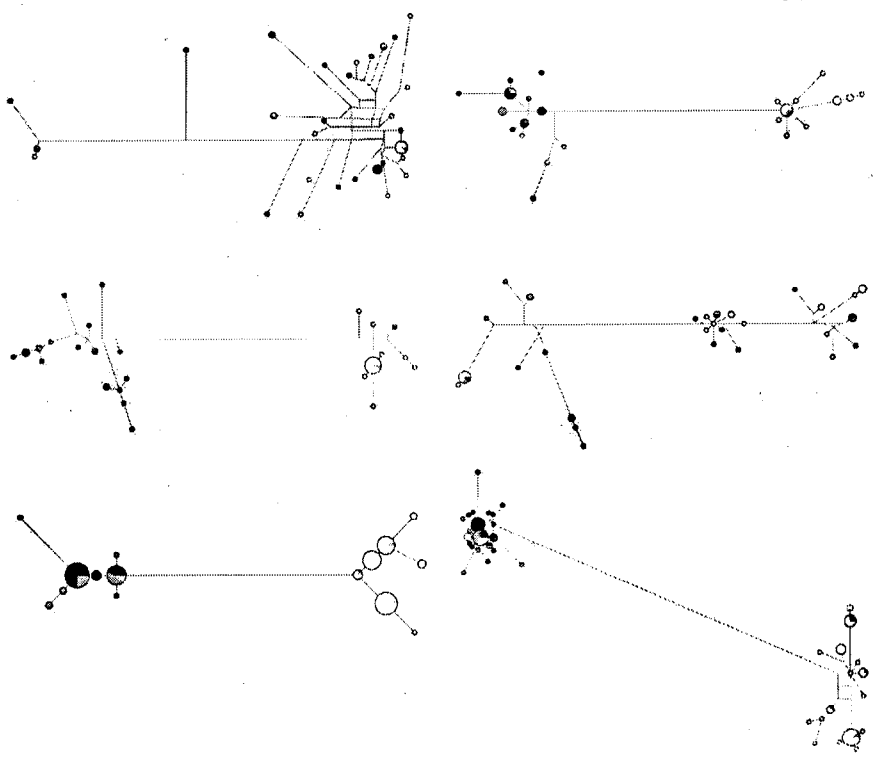


Figure 1- Median-Joining haplotype networks representing the phylogenetic relationships between all the alleles found in the European rabbit. From top to bottom and left to right: PHKA2, SMCX, MSN, HPRT1, SRY and Cytb. The size of the circles is proportional to the frequency of each haplotype. Black corresponds to populations from southwest Ibrian Peninsula, grey corresponds to populations from the contact zone and white to populations from northeast Iberian Peninsula and France. Modified from Geraldes *et al.* (2006) and Geraldes and Ferrand (*in prep*).

began to evolve under allopatry by taking the observed lineages as a proxy for the populations that they once evolved in (haplotypes inferred to be recombinants between the two lineages were excluded from this analysis). We calibrated our molecular clock for each locus by considering that the observed divergence between all alleles found in *O. cuniculus* and the allele found in *L. granatensis* corresponded to 11.8 MY of divergence (Matthee *et al.* 2004). Results for each locus are shown in Table 1. Using net nucleotide distances corrected for multiple substitutions we obtained divergence times for each locus that were at least as old as the previously suggested 2 MY of divergence between subspecies (Biju-Duval *et al.* 1991). The

exceptions were PHKA2 and the SRY gene. For PHKA2, this difference is very small and likely within the standard deviation of the estimates of divergence, but for SRY the estimated divergence time is only 0.95 MY. This is less than half the value estimated using RFLPs of the entire mitochondria and the majority of other loci studied here. At least two explanations could account for this result. First, positive Darwinian selection is known to drive the evolution of the SRY gene in Rodents (Tucker *et al.* 1991) and Primates (Whitfield *et al.* 1991). If positive Darwinian selection has driven the evolution of this locus one would expect increased divergence to *Lepus spp.* and reduced polymorphism within *Oryctolagus spp.*. Even if positive Darwinian selection has also driven the evolution of this locus between the two subspecies of rabbit, the time elapsed here is much smaller and the net result could be an inflation of the divergence between genera in relation to divergence between subspecies of rabbit. We find some support for this hypothesis in our data. When we compare ratios of polymorphism to divergence for synonymous and non-synonymous substitutions at SRY we find much higher levels of replacement substitutions than silent substitutions. Also, when we compare the ratio of polymorphism to divergence between Cytb and SRY a significant deviation from neutral expectations is observed. There is also a possible purely neutral explanation for this result. There is strong evidence that SRY is evolving under concerted evolution in the European rabbit. This

Table 1- Estimates of net nucleotide divergence[1] at six sex-linked loci between *O. cuniculus* and *L. granatensis* and between lineages of *O. cuniculus* corrected[2] and uncorrected for multiple hits and estimate of the time[3] since the split between subspecies of *O. cuniculus*.

	O. cuniculus/L. granatensis		*O. c. cuniculus/O. c. algirus*			
	Uncorrected	Corrected	Uncorrected		Corrected	
	Da (%)	Da (%)	Da (%)	Divergence Time (MY)	Da (%)	Divergence Time (MY)
Cytb	13.454	33.428	5.139	4.51	6.927	2.45
SRY	4.373	4.531	0.362	0.98	0.365	0.95
PHKA2	5.719	8.483	1.153	2.38	1.263	1.76
SMCX	1.475	1.529	0.735	5.88	0.643	4.96
MSN	3.803	4.634	0.711	2.21	0.819	2.08
HPRT1	3.848	5.313	1.633	5.01	1.911	4.24

[1]Per cent net nucleotide distances per nucleotide site (Da %; Nei 1987) were calculated as Da=Dxy-0.5(Dx-Dy), where Dxy is the per cent average pairwise distance per site between all samples, and Dx and Dy are the average per cent pairwise distances per site within each group. Average pairwise distances were calculated using PAUP (Swofford 2002).

[2]Appropriate models of nucleotide substitution to account for multiple substitutions were chosen under a maximum likelihood frame using MODELTEST (Posada and Crandall 1998) with the Akaike Information Criterion (Posada and Buckley 2004).

[3]Assuming a divergence time between *O. cuniculus* and *L. granatensis* of 11.8 MY (Matthee *et al.* 2004) and using the divergence between the observed lineages of *O. cuniculus* as a surrogate for the divergence of the subspecies *O. c. cuniculus* and *O. c. algirus*.

is driven by gene conversion between copies of this gene. If gene conversion is biased, rates of molecular evolution could be either lowered or fastened. This topic will be discussed in greater detail in the section "Abundant gene conversion drives the evolution of the SRY gene in Leporids".

5.1.2 High levels of nucleotide polymorphism suggest long term effective population sizes

At all six genes we found very high levels of nucleotide polymorphism when all European rabbit samples are considered together. When compared to published data on other taxa, levels of nucleotide polymorphism in the European rabbit clearly stand out. For example, at the Y chromosome, π in the European rabbit (0.213% per site) is one order of magnitude higher than that observed in humans (0.014%, Hammer *et al.* 2003), chimpanzee (0.067%, Stone *et al.* 2002) and house mice (0.041%, Nachman 1998). The same is true even for the less polymorphic X chromosome loci (SMCX and MSN). One of the reasons for the large levels of nucleotide polymorphism observed is the existence of population substructure. In fact, as discussed in the previous section, there is a marked dichotomy in the rabbit genome where at each gene two very divergent lineages are detected. This leads to an overestimate of the population mutation parameter (θ). This is also the case in the chimpanzee data set where several subspecies were pooled together and in a recent report of nucleotide variability in the field vole, *Microtus agrestis* (π = 0.190% for the Y chromosome Hellborg *et. al* 2005), where two deeply divergent lineages are found. These comparisons show that even taking into account the pooling of two divergent subspecies, levels of nucleotide variability in rabbit populations are still higher than what is usually observed in other mammalian species. There are two possible explanations for these surprisingly high levels of nucleotide variability. Mutation rates can be higher in rabbits than in the other mammals used for comparison; alternatively, the European rabbit could have been able to maintain larger than average effective population sizes for long periods of evolution.

We estimated mutation rates per nucleotide site per year for all six loci (Table 2) by calculating divergence to *Lepus granatensis* and assuming that the observed divergence corresponds to 11.8 MY (Matthee *et al.* 2004). Divergence was calculated

after an appropriate model of nucleotide substitution was selected to account for multiple substitutions.

The mutation rates estimated from our data conform well to recent genome wide mutation rate estimates per site per year from humans $2.2X10^{-9}$ and house mice $4.5X10^{-9}$ (Waterston *et al.* 2002), but are in general slightly lower. The average European rabbit mutation rate when only the X chromosome loci are considered is $2.1X10^{-9}$ but this value is greatly affected by the low mutation rate estimated for *Smcx*. If this gene is removed then the mutation rate for X-linked loci is $2.6X10^{-9}$. For the Y chromosome and mitochondria our estimates are also very close to those reported for humans ($1.33X10^{-9}$ for the Y and $1.58X10^{-8}$ for mtDNA; Wilder *et al.* 2004). If an accelerated mutation rate in the rabbit lineage can not explain the high levels of nucleotide substitution, one must assume that rabbit populations have maintained high effective population sizes for extended periods of time.

Table 2- Estimation of mutation rates per site per year and selected model of nucleotide substitution at six loci in *Oryctolagus cuniculus*.

	Model of substitution[a]	Mutation rate[b]
Cytb	TVM+G[c]	$1.42X10^{-8}$
SRY	TVM[d]	$1.92X10^{-9}$
PHKA2	TrN+I+G[e]	$3.59X10^{-9}$
SMCX	TrN+I[f]	$6.48X10^{-10}$
MSN	TVM+G[g]	$1.96X10^{-9}$
HPRT1	TrN+I[h]	$2.25X10^{-9}$

[a]Appropriate models of nucleotide substitution to correct for multiple hits were selected for each gene using MODELTEST 3.06 (POSADA and CRANDALL 1998) with the Akaike Information Criterion (POSADA and BUCKLEY, 2004).
[b]Mutation rate per site per year.
[c]Transversion Model with estimated α parameter describing the gamma distribution of 0.2213.
[d]Transversion Model.
[e]TAMURA-NEI model (1993) with proportion of invariable sites of 0.6578 and estimated α parameter describing the gamma distribution of 1.0022.
[f]TAMURA-NEI model (1993) with proportion of invariable sites of 0.5536.
[g]Transversion Model with estimated α parameter describing the gamma distribution of 0.2204.
[h]TAMURA-NEI model (1993) with proportion of invariable sites of 0.8524.

We analyzed our data under a coalescent framework to estimate rabbit effective population sizes. From our X chromosome data we estimated an effective rabbit population size of $1.3X10^{6}$ (Geraldes *et al.* 2006), from the mitochondria data we estimated a female effective population size of $1.6X10^{6}$ and from Y chromosome data the effective male population size is $4.4X10^{5}$. The estimated effective population size of males is less than half the estimated effective female population size. This difference is not at all surprising since we know that rabbit populations display a marked social structure with a dominant male siring most of the progeny from a given

social group (Myers and Mykytowycz 1958). This social behavior greatly impacts the number of patrilineages that are passed on to the next generation, and as a result diminishes the number of effectively reproducing males. These estimates of effective population size are very large and are likely the main reason for the high levels of nucleotide variability observed.

5.1.3 Subspecies specific demographic processes and Post Pleistocene range expansions

Given the marked dichotomy observed in the rabbit genome, we were interested in studying some population phenomena at the subspecific level. Among these are population dynamics and gene flow.

The European rabbit is a classic ecological example of a species with very high reproductive capacity, with several litters per year. Here we asked how did the post glacial colonization of the Iberia Peninsula occurred, and particularly we were interested in looking for differences or similarities between the expansion of *cuniculus* from its eastern refugium and *algirus* from its southwestern refugium. Did these population expansions leave similar genetic signatures on the mitochondria and Y chromosome, *i.e.*, did males and females contributed similarly to this population expansion?

The European rabbit is usually referred to as a successful colonizer (Thompson and King 1994). This is partially due to its remarkably fast colonization of Australia and hundreds of islands. This suggests that rabbits have high dispersal capabilities. Previous behavior and molecular studies have shown that in general males undergo natal dispersal and females are phylopatric (e.g. Richardson *et al.* 2002). The magnitude of this behavior is not consistent across studies, and to our knowledge no such study was performed in the Iberian Peninsula. Here we directly compared, in each subspecies, patterns of gene flow for the mtDNA and Y chromosome to test the hypothesis that populations are more differentiated for the mtDNA than for the Y chromosome.

In the previous section we reported the estimates of effective male and female population sizes for the European rabbit. We also estimated effective population sizes for each subspecies based on our mtDNA and Y chromosome polymorphism data. These were analyzed under a coalescent framework using a model that allows for the

estimation the population mutation parameter and of an additional parameter, the population growth parameter. Likelihood ratio tests were performed to determine if each data set is best described by a constant population size model, population growth parameter equals zero, or an exponential growth model. Results are presented in Table 3.

Our results show some interesting differences and similarities between subspecies of *Oryctolagus cuniculus*. First, for *algirus* a model incorporating exponential population growth statistically fits the Y chromosome and mtDNA data better, while for *cuniculus* the same is not true. This signal of population growth was also detected for mtDNA using tests of the neutral model based on the frequency spectrum of polymorphisms (Tajima's D, Fu and Li's D and Fu's Fs). For the Y chromosome, these tests were also negative in the *algirus* data set, implying either a population expansion or the action of positive directional selection, but they were not statistically different from zero. This could simply be due to the lack of power of these tests to detect deviations from neutrality with low numbers of segregating sites. It thus seems that both males and females of *algirus* have experienced exponential population growth during the colonization of the Iberian Peninsula from the southern refugium while neither sex in *cuniculus* shows the same pattern. In fact, the data from this subspecies are more compatible with long term constant population sizes.

Table 3 – Estimation of the Time to the Most Recent Common Ancestor (TMRCA) and other population parameters for each subspecies of the European rabbit using male (Y chromosome) and female (mtDNA) specific markers estimated under a coalescent maximum likelihood (ML) framework using GENETREE with constant size (upper row for each data set) and exponential growth (lower row) demographic models

		θ_{ml}[1]	β_{ml}[2]	N_e or N_0 (10^3)[3]	TMRCA (10^3)(SD)	Likelihood Score (SE)	LRS[4]
O. c. algirus	Cytb	11.3	-	644	361 (66)	$2.2X10^{-15}$ ($1.1X10^{-35}$)	6.62**
		21.5	10.70	1220	323 (43)	$7.9X10^{-13}$ ($3.3X10^{-13}$)	
	SRY	0.96	-	124	203 (89)	$9.6X10^{-6}$ ($1.3X10^{-8}$)	5.02*
		1.27	0.91	244	250 (77)	$1.2X10^{-4}$ ($1.9X10^{-8}$)	
O. c. cuniculus	Cytb	6.6	-	376	383 (89)	$4.5X10^{-20}$ ($2.7X10^{-22}$)	2.63[n.s.]
		8.6	1.73	490	343 (56)	$1.7X10^{-20}$ ($8.1X10^{-21}$)	
	SRY	1.01	-	194	342 (141)	$1.5X10^{-6}$ ($1.1X10^{-9}$)	0.20[n.s.]
		1.21	0.51	233	305 (98)	$1.7X10^{-6}$ ($3.9X10^{-9}$)	

[1]Per gene population parameter.
[2]Per gene population exponential growth parameter.
[3]Ne – effective population size, for constant population size model and N_0 – present day population size, for exponential population growth model.
[4]P values are indicated as follows: (n.s.) P>0.05, (*) 0.01<P<0.05, (**) 0.005<P<0.01, (***) P<0.005.

The estimated time to the most recent common ancestor (TMRCA) is similar for both subspecies and markers (Figure 2 and Table 3). Although the TMRCA of

cuniculus is slightly older than the TMRCA of *algirus*, the value of the TMRCAs +/- one standard deviation are overlapping. This implies that the variation observed in each subspecies dates back to approximately 300,000 years ago, *i.e.*, all the alleles that we observe in our sample of each subspecies coalesce to a single allele in each of the populations that was present around 300,000 years ago.

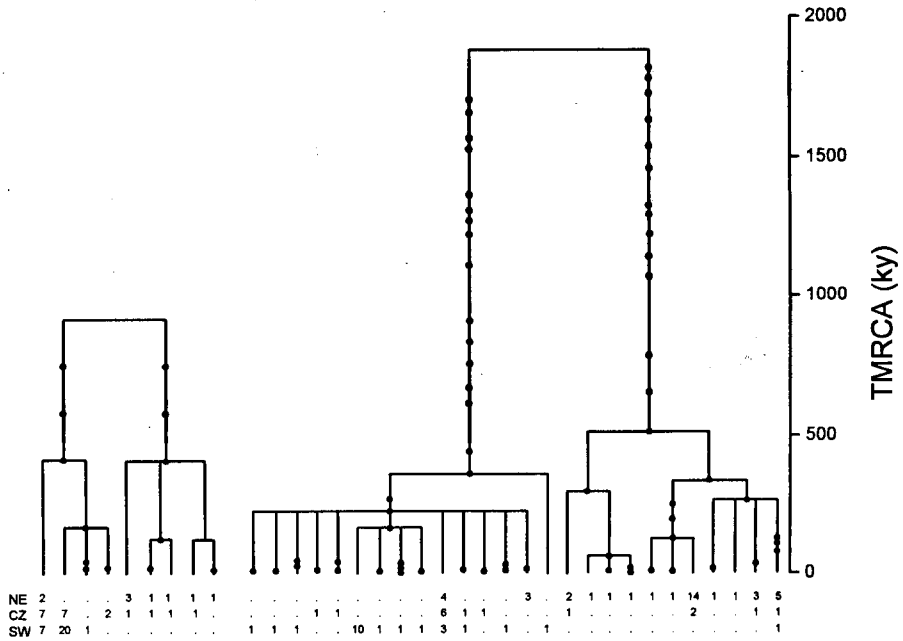


Figure 2- Non reticulating genealogies for SRY and Cytb. Dots in the trees indicate mutations. The geographic distribution of each haplotype is indicated at the bottom. NE corresponds to northeastern Iberian Peninsula (*cuniculus* populations), CZ corresponds to the contact zone that bissects the Iberian Peninsula in a northwest/southeast axis, and SW corresponds to southwestern Iberian Peninsula (*algirus* populations). Trees were obtained using the software Genetree and are scaled according to the TMRCA of the total sample for each gene.

Finally, in both subspecies the estimated population size of males is much smaller than the estimated population size of females. In *cuniculus* there are approximately two reproducing females for each male, while in *algirus* there are five reproducing females per male. These data are in good agreement with the known social structure of rabbits. The fact that the difference between sexes is much bigger in *algirus* populations points to possible stronger social hierarchies in *algirus* populations. Alternatively, females might have contributed more than males for the population expansion experienced by *algirus* populations.

In the same study we compared male and female levels of gene flow within subspecies. Particularly we compared levels of population differentiation at the maternally inherited mtDNA and the paternally inherited Y chromosome between

populations from the two glacial refugia and the areas of recent colonization. If females are philopatric we would expect there to be less female biased gene flow and hence greater population differentiation at the mtDNA than the Y chromosome. Two population groups per subspecies were defined, the populations from the glacial refugia and those present in post-glacially colonized areas (Figure 3).

From Table 4 it is clear that in both subspecies gene flow is mostly male driven. Estimates of population differentiation (Fst) are indeed extremely low for the Y chromosome and fairly high for mtDNA. If we assume an island model of population structure the Fst estimates of population differentiation can be converted

Table 4– Fst[1] and Nm[2] estimates between refugium and expansion areas of each subspecies of *Oryctolagus cuniculus* at SRY and Cytb

	Fst		Nm	
	SRY	Cytb	SRY	Cytb
O. c. algirus	0.02	0.31	53.40	1.14
O. c. cuniculus	0.06	0.33	15.11	1.03

[1]Fst based on nucleotide data (Hudson, Slatkin & Maddison 1992).
[2]Nm is the effective number of migrantas per generation between populations assuming an island model of population structure.

into effective number of migrants per generation (Nm) if one assumes that the populations are at mutation drift equilibrium. According to these estimates in each generation there is approximately one female being exchanged between population groups of each subspecies, while there are 15 *cuniculus* and 53 *algirus* males being exchanged between populations. This results show that, as previously described, male rabbits tend to disperse more than female rabbits do. Again, the pattern observed is the same in both populations, but the magnitude of the difference is bigger in *algirus* populations. A possible problem with this analysis is the assumption that each population group is a panmitc population in mutation drift equilibrium. In fact, it is likely that they are not. Another possible source of error in this analyses is the assumption that migration dynamics are the same between existing populations (as assumed in the above calculations) and between an existing population and unoccupied territory. In order to account for these potential problems, we compared gene flow between all pairs of populations within each subspecies. When we do this, we have a high number of populations being compared and very small sample sizes per population. When the number of populations studied is high and the sampling size in each population is low, the statistic Snn (Hudson 2000) was shown to have the most power to detect genetic differentiation between populations. Using this statistic

we still see basically no population differentiation for the SRY gene and higher population differentiation for Cytb. This difference is much smaller than that depicted in our previous analysis but still points to higher male than female migration between populations.

5.1.4 Comparing recombining with non-recombining regions of the rabbit genome: implications for the study of mammalian hybrid zones

In the previous section we compared male and female gene flow within each subspecies of the European rabbit. We interpreted our estimates of gene flow as a proxy for male and female migration between population groups of the same subspecies. In this section we will focus on differentiation between subspecies of rabbit.

Figure 3- Map of the Iberian Peninsula. Black dots represent popualtions from the algirus range and white dots populations from the cuniculus range. Solid lines indicate glacial refugia and dashed lines indicate áreas of post glacial expansion of each subspecies.

The comparison of the high levels of genetic differentiation between subspecies of the European rabbit as depicted by RFLP analyses of the mitochondrial Cytb gene (Branco *et al.* 2000) with the extremely low levels of differentiation depicted by nine autosomal microsatellite loci (Queney *et al.* 2001) raised the hypothesis that gene flow between subspecies of European rabbit may be mainly male driven. In support of this hypothesis is the natal dispersal male behavior of rabbit populations. As we have seen in the previous section this is well supported by our estimates of within subspecies gene flow for the Y chromosome and mtDNA.

A direct test of the former hypothesis is the comparison of levels of gene flow for the Y chromosome and mtDNA across the contact zone between *cuniculus* and *algirus*. If male driven gene flow is responsible for the low levels of differentiation observed at autosomal loci then estimates of gene flow should be higher for the Y chromosome than for the mitochondria. Figure 4 shows Fst values for these two loci between populations from southwestern and northeastern Iberian Peninsula. Fst values for both loci are very high and are surprisingly higher for the Y chromosome than the mitochondria. The obvious interpretation of these results would be that dispersal between rabbit populations is reduced and that males migrate even less than females. Several lines of evidence argue against this. First, behavioral observations indicate that rabbits can disperse long distances in short periods of time. An obvious example is the colonization of Australia (Zenger *et al.* 2003). Also in Australia it has been found that gene flow occurs over very large distances (Fuller *et al.* 1997). Second, our estimates of gene flow indicate that within each subspecies levels of gene flow are high and that males disperse more than females. Given this, the low levels of differentiation observed for microsatellites could simply be a result of "severe allele size homoplasy" resulting from the high mutation rate of these loci coupled with high long term effective size of rabbit populations (Queney *et al.* 2001). Here we propose that the action of natural selection could reconcile these contradictory observations.

In Figure 4 we included estimates of population differentiation between subspecies of *O. cuniculus* for several loci. In each case only populations outside of the contact zone were included. Fst estimates for Y chromosome, mtDNA and X-linked loci were obtained during the course of this study and are derived from nucleotide polymorphism data (Geraldes *et al. in prep.*; Geraldes *et al.* 2006). Fst values for microsatellite loci are as reported in Queney *et al.* (2001) and Fst values for 16 allozyme loci were calculated from the data in Ferrand and Branco (2006). Three striking observations emerge from Figure 4. First, the high variability (from 0 to approximately 0.9) in Fst estimates from different loci. Second, the very high Fst values of SRY, Cytb, SMCX and MSN. Finally, the absence of Fst values between 0.4 and 0.6 in an otherwise quasi continuum spectrum.

The four genes with high Fst values have some features in common: they were genotyped with sequence data (instead of fragment size polymorphism as for microsatellites and electrophoretic mobility as for the protein loci), they are sex linked, and they are either devoid of homologous recombination (SRY and Cytb), or

show low levels of recombination (SMCX and MSN). We find that the genotyping method cannot account for the differences observed since two X-linked genes (PHKA2 and HPRT1) were also genotyped by sequencing and they show some of the lowest Fst values. Accordingly, nucleotide polymorphism data from three other autosomal loci (HBB personal communication R. Campos; CSN3, Carneiro 2005; and CAII, Campos 2006) also show very low levels of differentiations (0<Fst<0.1). The same reasoning applies to discard the mode of inheritance as a possible source of the discrepancies (again PHKA2 and HPRT1, just as SMCX and MSN are X-linked). Finally, we propose that the joint action of recombination and natural selection could explain the differences observed.

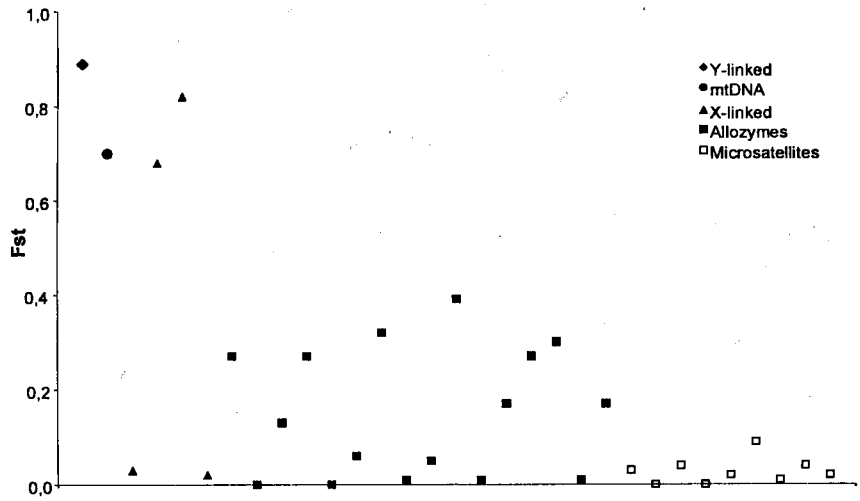


Figure 4– Estimates of population differentiation (Fst) between subspecies of European rabbit for Y-linked (Geraldes and Ferrand *in prep.*), mtDNA (Geraldes and Ferrand *in prep.*), X-linked (Geraldes *et al.* 2006), allozyme (Geraldes *et al. submitted*, calculated with data from Branco and Ferrand 2006) and microsatellite loci (Queney *et al.* 2001). Fst values are presented in the Y axis and all loci are ordered along the X axis. From left to right along the X axis, a diamond (◊) indicates SRY (Y-linked), a circle (•) indicates Cytb (mtDNA), triangles (▲) indicate PHKA2, SMCX, MSN and HPRT1 (X-linked), black squares (■) indicate ADA, ALB, CAI, CAII, GALT, GC, HBA, HBB, NP, PEPA, PEPB, PEPC, PEPD, PGD, SOD and TF (autosomal allozymes and other protein loci) and finally white squares (□) indicate SAT2, SAT3, SAT4, SAT5, SAT7, SAT8, SAT12, SAT13 and SAT16 (autosomal microsatellites).

In several populations from the contact zone, the two subspecies co-exist. This is evident from the presence of both lineages of the Y chromosome (Geraldes *et al. submitted*), MSN and SMCX (Geraldes *et al.* 2006) and mtDNA (Branco et al. 2000) in several populations from this region. This indicates that there is potential for hybridization and introgression between the genomes of the two subspecies. We argue that genes exhibiting low Fst values introgress freely across the contact zone, while genes showing high Fst cannot introgress. The fact that these loci are in regions where

linkage disequilibrium extends very far increases the probability that they are linked to genes involved in reproductive isolation between both subspecies, *i.e.*, they are linked to putative speciation genes. The idea that regions of low recombination could be particularly prone to be involved in reproductive isolation is not new. Several authors have argued that speciation frequently occurs when a population becomes fixed for one or more chromosomal rearrangements that reduce fitness in heterozygous individuals (e.g., White 1978). This would occur because these rearrangements could originate unbalanced gametes as a result of recombination during meiosis. These theories face an important problem: for a rearrangement to get fixed in a population it would first have to pass through a heterozygote state. If it were to cause unbalanced gametes that would result in inviable or infertile hybrids then it would never reach fixation in a population because it would need to pass through an adaptive valley. Recently, Rieseberg (2001) and Noor *et al.* (2001) proposed that chromosomal rearrangements could promote speciation by reducing recombination and extending the effect of speciation genes to linked sites. Looking at our data, and specifically to the loci for which we have an idea of the recombination rate (X-linked, Y-linked and cytoplasmic loci), we see that the genes that show reduced introgression are the ones that are devoid of recombination (SRY and Cytb) and those for which the estimated recombination rate is very low (SMCX and MSN; Geraldes *et al.* 2006). For genes showing low linkage disequilibrium, *i.e.*, with high estimated recombination rates (PHKA2 and HPRT1; Geraldes *et al.* 2006) introgression is high and no genetic differentiation between subspecies is observed. This same phenomenon has been documented in other species such as fruit flies (e.g. Noor *et al.* 2001, Machado *et al.* 2002), and sunflowers (Rieseberg *et al.* 1999) and suggests that this could be a common pattern between hybridizing taxa. To explain the generalized low levels of genetic differentiation depicted by the allozyme loci we suggest the action of three different factors: genotyping method, recombination landscape and selection. First, it is possible that the low resolution techniques employed to score variation at some of the loci was unable to detected fixed differences between subspecies. Second, some of the genes that code for these proteins could be in regions of high recombination that decouple them from loci involved in reproductive isolation. Finally, some of these loci can introgress because mutations acquired in allopatry in one lineage could increase the fitness of individuals from both subspecies. If so, one could predict that they would introgress adaptively across the contact zone.

5.2 Evolution of the SRY gene in Leporids

5.2.1 Duplication of the SRY gene in the European rabbit and in other Leporids

SRY is the sex determining region on the Y chromosome. This single exon gene controls mammalian sex determination by initiating testis development from the undifferentiated gonadal primordium (Gubbay *et al.* 1990; Sinclair *et al.* 1990). It is a single copy Y-linked gene in humans, house mice and most mammalian species. In some rodent species it has been shown to exist in multiple copies (Fernandez *et al.* 2002; Lundrigan and Tucker 1997; Naganime *et al.* 1994).

In Leporids, the SRY gene was first identified in the European rabbit in 1990 by Sinclair and colleagues. In 2001, Wallner *et al.*, described a PCR protocol for sexing Leporids. In this paper they amplified a fragment of the SRY from *Oryctolagus cuniculus* and *Lepus europaeus*. The following year SRY was mapped to position Yp12 of the rabbit using FISH (Fluorescent in-Situ Hybridization) mapping of a BAC clone containing the SRY gene (Hayes *et al.* 2002).

In this work we have sequenced approximately 2.4 Kb of the SRY gene of the European rabbit (Geraldes *et al.* 2005). To do so, we used a BAC clone, containing the SRY gene, from a rabbit genomic library (Rogel-Gaillard *et al.* 2001). We used a primer walking coupled with direct sequencing approach. This is described in detail in Geraldes *et al.* (2005). In the course of this study we found strong evidence that this gene is duplicated in the rabbit Y chromosome. Next we will summarize the evidence that we found for the duplication of this gene in the European rabbit and we will conclude by showing the results of the sequencing and assembly of the entire rabbit BAC clone 828D7.

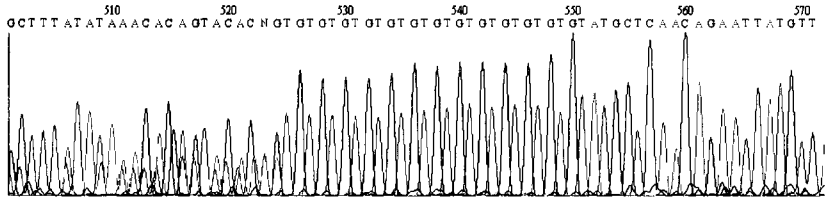


Figure 5- Chromatogram of a sequence from rabbit BAC clone 828D7. This is a reverse sequence showing that (from right to left, direction of sequencing reaction) after the $TG_{(13)}$ repeat two overlapping sequences are obtained.

The first evidence for duplication of the SRY gene in the European rabbit was obtained through the direct sequencing of this locus. During our primer walking approach to expand the length of characterized SRY sequence, we observed that after

the sequencing of a $TG_{(13)}$ repeat we obtained a multiple sequence pattern consistent with the presence of size polymorphism. In Figure 5 we can see an example of one such sequence where after the repeat two overlapping sequences differing by a TG insertion are observed. This pattern could either be obtained by a duplication of the gene within this clone where the two copies would differ by one TG insertion or by a replication error during the culture of this clone in bacteria.

The sequenced fragment of SRY contains another tandem repeat, a mononucleotide $A_{(14)}$ array (Geraldes *et al.* 2005). The $TG_{(13)}$ repeat is located approximately 400 bp upstream of the coding region of SRY and the $A_{(14)}$ repeat is located approximately 500 bp upstream the coding region of SRY. Specific primers for each of these repeats were designed, and these regions were amplified in a sample of rabbits from the Iberian Peninsula and France. We genotyped these samples with vertical polyacrilamide gel electrophoresis to look for fragment size polymorphism. At both loci, several individuals showed two different size alleles, indicating the amplification of two different loci (unpublished data). Since no amplification from female DNA was ever achieved we concluded that the observed pattern was due to the duplication of both these loci in the Y chromosome.

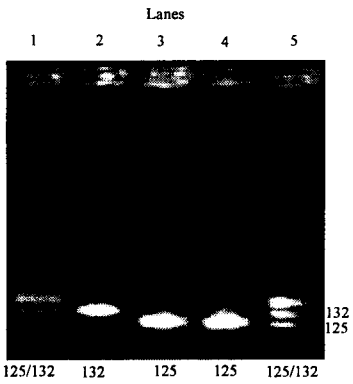


Figure 6- Agarose gel electrophoresis (Metaphor 4.5% w/v) showing the three banded patterns observed in the PCR amplification of the 3' region of the SRY gene. Lanes 1 and 5 exhibit PCR products of 125 and 132 bp and an additional heteroduplex band (corresponding to individuals with haplotype 125/132), lane 2 shows a single 132 bp band (corresponding to individuals with haplotype 132/132), and lanes 3 and 4 display a single 125 bp band (corresponding to individuals with haplotype 125/125) From Geraldes and Ferrand (2006).

Further evidence for the duplication of SRY was obtained in Geraldes and Ferrand (2006). In this study we again surveyed several populations from Iberia. This time we were interested in studying the distribution of a seven base pair insertion in the 3' UTR of SRY described in Geraldes *et al.* (2005). This polymorphism was scored on 4.5% high resolution agarose after PCR amplification of a 125 bp product

corresponding to absence of insertion and a 132 bp fragment corresponding to presence of insertion. Three patterns were observed (Figure 6). They correspond to genotypes 125, 132 and surprisingly 125/132. This last pattern was interpreted as corresponding to individuals where one copy of SRY had the insertion and the other did not. We sequenced individuals corresponding to each of the patterns observed to confirm this result. Exemplifying chromatograms are shown in Figure 7. We can see that clean sequences are only obtained for individuals with or without the insertion, but in some individuals, after the insertion a pattern corresponding to two overlapping sequences differing by seven base pairs is observed.

With the sequencing of most of the region described in Geraldes *et al.* (2005),

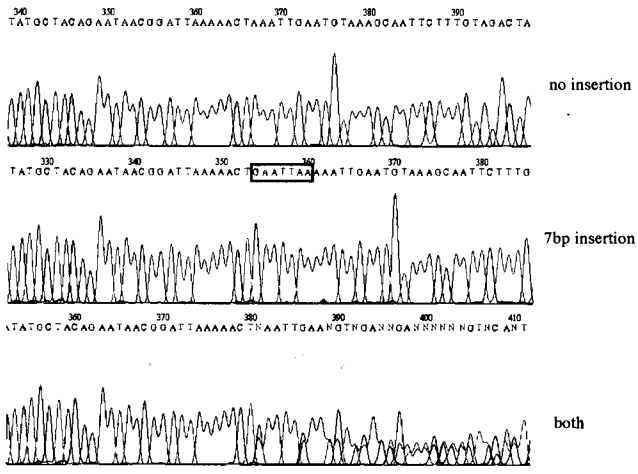


Figure 7- Chromatograms showing the absence of a 7 bp insertion (top) in the 3' UTR region of SRY, the presence of the insertion (middle), and one individual with two overlapping sequences after the 7 bp insertion.

in a sample of more than 100 male rabbits from the Iberian Peninsula (Geraldes and Ferrand *in preparation*) we detected five positions that where "heterozygous" for a nucleotide substitution. One position was heterozygous in a single individual, one in two individuals, one in four individuals and finally three individuals had the same two heterozygous positions.

This work culminated with the complete sequencing and assembly of the entire 92 kb rabbit 828D7 clone containing the SRY gene. This work was done in collaboration with the Arizona Genomics Institute and the Department of Ecology and Evolutionary Biology, both at the University of Arizona. Figure 8 shows the organization of the BAC clone in relation to the SRY gene. Two copies of the SRY gene are present in this clone. One copy of the SRY gene is located approximately 2 kb after the beginning of the clone, and the other is present at approximately 62 kb

after the beginning of the clone. These two copies are present in opposite directions because they are located in an inverted repeat. This inverted repeat sequence of at least 23 kb is present in this clone twice and it is separated by a 19 kb long region of unique sequence. At this point we cannot determine the actual size of the repeat. We can only say that it is at least 23 kb, but it is possible that it extends further.

Figure 8- Organization of BAC clone 828D7 in relation to the SRY gene. The sequenced region has approximately 92 kb. The 2.4 kb region of the SRY gene previously sequenced and reported in Geraldes *et al.* (2005) is represented by a black box. Two copies of the SRY gene were found. They are oriented in opposite directions and are present in an inverted repeat (represented by a white box) of at least 23 kb. This palindrome is separated by a spacer of 19 kb consisting of single copy sequence.

The sequencing and assembly of this clone allowed us to show that the SRY gene is duplicated in the rabbit Y chromosome. It also allowed us to understand the nature of this duplication. The gene is contained in a palindrome of at least 23 kb separated by 19 kb of unique sequence. This organization resembles that of most of the euchromatin of the human Y chromosome (Skaletsky *et al.* 2003). As outlined in the introduction of this work, the complete sequencing of the human Y chromosome revealed, among other things, that most of the Y chromosome is composed of large palindromes with a single copy sequence spacer region between them. These palindromes have the highest gene density of the Y chromosome and the genes located in these palindromes are present in multiple copies and are expressed mainly or only in testis and are mostly related to spermatogenesis. The SRY gene is also expressed in testis, but mostly during embryogenesis, and is not involved in spermatogenesis. In humans it is a single copy gene (Skaletsky *et al.* 2003). Here we show that in rabbits the rabbit the SRY gene is duplicated and present in a palindrome.

5.2.2 Abundant gene conversion drives the evolution of the SRY gene in the European rabbit

When the human Y chromosome sequence was completed and published (Skaletsky *et al.* 2003), a companion paper (Rozen *et al.* 2003) described in detail the organization of the newly found palindromes and reported that the genes located in palindrome arms were evolving in concert and underwent extensive gene conversion. The authors compared divergence between paralogous arms of human Y-linked

palindromes to divergence to the orthologous palindromes in chimpanzee. Divergence between orthologous palindrome arms in human and chimpanzee was on average 1.44%, but arm-to-arm divergence within species was only 0.028% in chimpanzees and 0.021% in humans. The authors concluded that the Y-linked palindromes arose before the separation of the two species, and that, in both the human and chimpanzee lineages, the paired arms of the palindromes evolved in concert. Using these divergence estimates they also calculated the rate of gene conversion between palindrome arms to be $1.1X10^{-5}$ per duplicated nucleotide per year, a value four orders of magnitude higher than the estimated Y chromosome mutation rate of $1.6X10^{-9}$ per nucleotide site per year (Rozen *et al.* 2003). Finally these authors argued that the observation that intact testis specific genes tend to be located in palindrome arms whereas non functional copies of these genes seem to be distributed randomly suggests that gene conversion acts to prevent gene loss in a genomic region devoid of heterologous recombination.

As we mentioned above, not only the genes present in palindrome arms of the human MSY are testis specific, but they are also involved in spermatogenesis. The SRY gene is expressed in testis, but mostly during embryogenesis. Also, it is not involved in spermatogenesis. Given this, its location outside human MSY palindromes is not surprising. As described earlier it is well documented that several species of rodents have multiple copies of SRY, but evidence outside of this group is scarce. We have shown in the previous section that not only the SRY gene is duplicated in the rabbit Y chromosome. Here we ask the question, are the two copies of the SRY gene in rabbits evolving in concert through gene conversion?

In the previous section we reported the occurrence of three genotypes for a 7 bp insertion in the 5' UTR of SRY as evidence for the duplication of the SRY gene. Can the observation of these genotypes tell us something about the mode of evolution of this gene?

The fact that this 7 bp insertion is only present in the *O. c. algirus* Y chromosome and not in the *O. c. cuniculus* Y chromosome and the fact that it is also absent from closely related genera (*Lepus* and *Sylvilagus*; Geraldes and Ferrand *in preparation* and Melo-Ferreira personal communication) implies that the presence of the 7 bp motif is the derived state and that its insertion was recent, i.e., after the split of the two subspecies of rabbit. The absence of insertion in both gene copies and the presence of the insertion in only one of the copies can be readily explainable with a

duplicated gene in a non-recombining environment. But the existence of the third state, presence of the insertion in both SRY copies implies either the independent insertion of the 7 bp motif in both copies or the transference of the insertion from one copy to the other by gene conversion (Figure 9). The first hypothesis seems highly unlikely. The mutation rate for size polymorphisms in non repetitive areas is usually much lower than the mutation rate for base substitutions. The likelihood of such a rare event taking place twice independently seems to be extremely low. Support for the hypothesis of gene conversion is much higher given that we know that there is

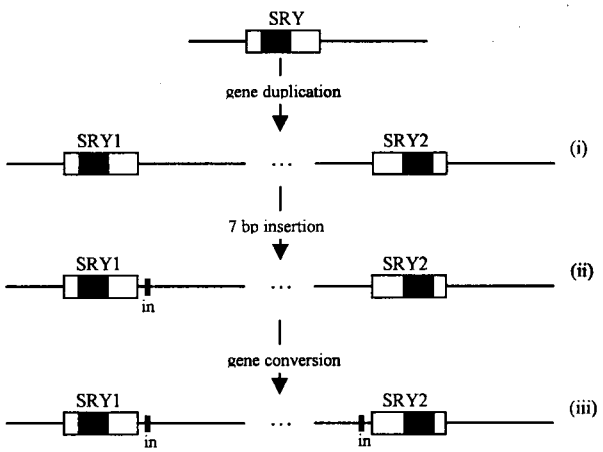


Figure 9- A schematic hypothesis proposed to explain the 7-bp polymorphism at the rabbit SRY locus. The coding region of SRY is represented by a white box, and the HMG domain is represented by a black box. The 7 bp insertion is located downstream from the coding region and is represented by a black square and the letters "in". (i) the two SRY copies resulted from an ancestral gene duplication; (ii) the GAATTAA motif is inserted in the 3' untranslated region of one SRY gene; (iii) this 7-bp motif is transferred to the second SRY copy by a mechanism of gene conversion. Note that subsequently (i) can also be derived from (ii) depending on the converted SRY copy. From Geraldes and Ferrand (2006).

abundant gene conversion between palindrome arms of the human Y chromosome and that this insertion is located in one such palindrome in the rabbit Y chromosome.

Further evidence for the occurrence of gene conversion between palindrome arms of the rabbit Y chromosome comes from the fact that only in very few individuals from our sample we observed base substitution mismatches between the two copies of the SRY. Only one haplotype present in three individuals showed more than one discrepancy between copies. This is surprising given that we found high levels of nucleotide polymorphism in the surveyed region both within subspecies and between subspecies. This indicates that the rate of conversion is sufficiently high to promptly homogenize the two copies of SRY. Furthermore, in the 23 kb repeated region of the BAC clone sequence we only found nine base substitutions between palindrome arms and five size polymorphisms. This corresponds to 99.94% sequence

identity between palindrome arms if both indel and base substitution polymorphisms are considered.

The next obvious question would be: what is the actual rate of gene conversion? In order to calculate the rate of gene conversion we have tried to amplify and sequence the inner boundaries of the palindrome arms in *O. cuniculus*, *Lepus californicus*, *L. castoviejoi* and *Sylvilagus spp.*. We designed one PCR primer within the inner boundary of the palindromes approximately 1 kb before the unique sequence spacer region and two extra primers, one approximately 1 kb after the beginning of the spacer region and one 1 kb before the end of the spacer region. Primers in the spacer region were oriented towards the palindrome arms in order to amplify approximately 2 kb fragments (Figure 10). Cross species amplification of the first boundary was straightforward but the amplification of the second boundary was very problematic and mostly unsuccessful. The sequencing of this particular region also proved to be very difficult because of the abundance of mono and dinucleotide arrays. In the end we were able to sequence almost 1 kb of boundary one in all the species, 1 kb of boundary two in *O. cuniculus*, but only 340 bp of the duplicated region of boundary two from *Lepus castroviejoi* and *Lepus californicus*. This supports the idea

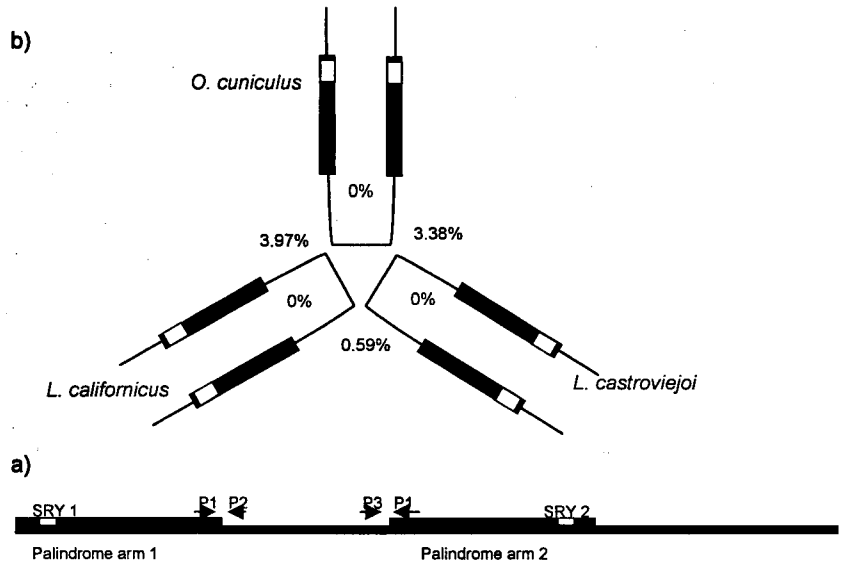


Figure 10- Divergence between orthologue and paralogue Y chromosome palindrome arms in Leporids. a) Experimental design and organization of leporid palindrome arms. Arrows indicate approximate location of primers used in the PCR protocol for the amplification of palindrome arms boundaries. Primer P1 was used in both boundaries and primers P2 and P3 are specific to boundaries one and two. b) Schematic representation of orthologue and paralogue palindrome arms divergence. Estimated paralogue divergence per base pair is indicated within each species palindrome arms and orthologue divergence per base pair is indicated between species palindrome arms.

that the duplication event in this region predates the radiation of the Leporids, but suggests that these regions are probably highly dynamic and change rapidly between species. We used the short sequences obtained in the three species to try to estimate the rate of gene conversion between palindrome arms in *O. cuniculus*. The estimated divergence between palindrome arms was always 0% since no mutations were observed. Between orthologous arms of different species the estimated divergence ranged from 3.97% between *O. cuniculus* and *L. californicus* to only 0.59% between *L. californicus* and *L. Castroviejoi* (Figure 10). This indicates that the rate of gene conversion between palindrome arms in leporids is very high. In the absence of gene conversion we would expect the divergence between palindrome arms of the European rabbit to be at least as high as the divergence between species.

5.2.3 Implications for the study of the mammalian Y-chromosome

Most population genetic studies still lack male specific information. Many still rely solely on mtDNA. By sequencing more than 2 kb of the European rabbit Y chromosome we hope to have facilitated the use of the Y chromosome marker in other Leporids. We have also amplified and sequenced this region in several other species belonging to the genera *Lepus* and *Sylvilagus*, showing that at least in these species amplification of this region is possible. Since most Leporid genera have diverged around the same time 12 MYA we believe that this marker region will also be of use in other genera such as *Brachylagus*, *Romerolagus*, *Bunolagus*, *Caprolagus* and *Pentalagus* among others. Another reason for the lack of use of Y chromosome markers is the lack of nucleotide polymorphism observed in many species. In our study we have found high nucleotide polymorphism levels. Studies in other species (i.e., *Microtus*, *Pan*) also found high levels of nucleotide variability indicating that the idea that selective sweeps usually eliminate nucleotide variability from Y chromosomes is not always true. Finally, the observation that the SRY gene in the European rabbit is a multicopy gene (we have shown that there are at least two copies, but more could be present) evolving through concerted evolution suggests that this could be the case in other species. Recently an exhaustive search for novel genes in the Y chromosome of the Family Carnivora (Murphy *et al.* 2006) showed that the Y chromosome is an actively recruiting chromosome. At least four new multicopy testis specific genes were found. This indicates that the Y chromosome is very variable in

terms of gene content and that testis specific genes tend to be recruited to the Y chromosome. Our results strengthen the idea that the organization of this chromosome is highly volatile between species. In some rodent species the SRY has been shown to exist in multiple copies in the Y chromosome. In other though, no evidence for this has been found. We argue that the occurrence of abundant gene conversion between copies of this gene might make the detection of duplicates harder without either extensive population sequencing or sequencing of large genomic regions.

5.3 References

Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, Soriguer R C, Gaaied A, Hili A and Mounolou J C. 1991. Mitochondrial DNA evolution in Lagomorphs: Origin of systematic heteroplasmy and organization of diversity in european rabbits. *Journal of Molecular Evolution*, 33: 92-102.

Branco M, Ferrand N and Monnerot M. 2000. Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome *b* gene. *Heredity*, 85: 307-317.

Campos P. 2006. Importância da caracterização genética do coelho bravo (*Oryctolagus cuniculus*) para a gestão e conservação das suas populações. (Msc thesis). Universidade do Porto. Porto, Portugal.

Carneiro M. 2005. Análise molecular e evidência de selecção natural no locus K-caseína em populaces selvagens e domésticas do coelho Europeu (*Oryctolagus cuniculus*). (Msc thesis). Universidade do Porto. Porto, Portugal.

Fernandez R, Barragan M J L, Bullejos M, Marchal J A, Martinez S *et al.*. 2002. Mapping the SRY gene in Microtus cabrerae: a vole species with multiple SRY copies in males and females. *Genome*, 45:600-603.

Ferrand N and Branco M. 2006. The evolutionary history of the European rabbit (*Oryctolagus cuniculus*), *in* Phylogeography in Southern European Refugia. Edited by S. Weiss and N. Ferrand. Springer. Amsterdam, The Netherlands. (in press).

Fuller S J, Wilson J C and Mather P B. 1997. Patterns of differentiation among wild rabbit populations *Oryctolagus cuniculus* L. in arid and semi-arid ecosystems of north-eastern Australia. *Molecular Ecology*, 6: 145-153.

Geraldes A, Rogel-Gaillard C and Ferrand N. 2005. High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) *SRY* gene. *Animal Genetics*, 36:349-351.

Geraldes A and Ferrand N. 2006. A 7-bp insertion in the 3' untranslated region suggests the duplication and concerted evolution of the rabbit SRY gene. *Genetics, Selection, Evolution*, 38:313-320.

Geraldes A and Ferrand N. in prep. Sex-specific differences in effective population size and gene flow in subspecies of the European rabbit (*Oryctolagus cuniculus*) inferred from Y chromosome and mtDNA sequence data.

Geraldes A, Nachman M W and Ferrand N. Submitted to Evolution. Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula.

Geraldes A, Ferrand N and Nachman M W. 2006. Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, doi: 10.1534/genetics.105.054106.

Gerard D T and Filatov D A. 2005. Positive and negative selection on mammalian Y chromosomes. *Molecular Biology and Evolution*, 22: 1423-1432.

Gubbay J, Collingnon J, Capel B, Economou B, Munsterberg A *et al.*. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embrionically expressed genes. *Nature*, 346:245-250.

Hammer M F, Blackmer F, Garrigan D, Nachman M W and Wilder J A. 2003. Human population structure and its effects on sampling Y chromosome variation. *Genetics*, 164:1495-1509.

Hayes H, Rogel-Gaillard C, Zijlstra C, de Haan N A, Urien C *et al.*. 2002. establishment of an R-banded rabbit karyotype nomenclature by FISH localization of 23 chromosome-specific genes on both G- and R-banded chromosomes. *Cytogenetic and Genome Research*, 98:199-205.

Hellborg L, Gunduz I and Jaarola M. 2005. Analyses of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, 14: 2025-2031.

Hudson R R. 2000. A new statistic for detecting genetic differentiation. *Genetics*, 155:2011-2014.

Hudson R R, Slatkin M and Maddison W P. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics*, 132:583-589.

Lundrigan B L and Tucker P K. 1997. Evidence for multiple functional copies of the male sex-determining locus, SRY, in African murine rodents. *Journal of Molecular Evolution*, 45:60-65.

Machado C A, Kliman R M, Markert J and Hey J. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*, 19: 472-488.

Matthee C A, van Vuuren B J, Bell D and Robinson T J. 2004. A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the miocene. *Systematic Biology*, 53: 433-447.

Murphy W J, Wilkerson A. J. P, Raudsepp T, Agarwala R, Schäffer A A, Stanyon R and Chowdhary B P. 2006. Novel gene acquisition on carnivore Y chromosomes. *PLOS Genetics*, 2:353-363.

Myers K and Mykytowicz R. 1958. Social behaviour in the wild rabbit. *Nature*, 181: 1515-1516.

Nachman M W. 1998. Y chromosome variation of mice and man. *Molecular Biology and Evolution*, 15:1744-1750.

Nagamine C M. 1994. The testis-determining gene, SRY, exists in multiple copies in old world rodents. *Genetical Research Cambridge*, 64:151-159.

Nei M. 1987. Molecular Evolutionary Genetics. Columbia University Press. New York, US.

Noor M A F, Grams K L, Bertucci L A and Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, 98:12084-12088.

Posada D and Crandall K A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 14:817–818.

Posada D. and Buckley T R. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over Likelihood ratio tests. *Sysematic Biology*, 53:793–808.

Queney G, Ferrand N, Weiss S, Mougel F and Monnerot M. 2001. Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution*, 18: 2169-2178.

Richardson B J, Hayes R A, Wheeler S H and Yardin M R. 2002. Social structures, genetic structures and dispersal strategies in Australian rabbit (*Oryctolagus cuniculus*) populations. *Behavioral Ecology and Sociobiology*, 51:113-121.

Rieseberg L H. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, 16:351-358.

Rieseberg L H, Whitton J and Gardner K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, 152:713-727.

Rogel-Gaillard C, Piumi F, Billaut A, Borgeaux N, Save J C *et al.*. 2001. Construction of a rabbit bacterial artificial chromosome (BAC) library: application to the mapping of the major histocompatibility complex to position 12q.1.1. *Mammalian Genome*, 12:253-255.

Rozen S, Skaletsy H, Marszalek J D, Minx P, Cordum H S *et al.*. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, 423:873-876.

Sinclair A H, Berta P, Palmer M S, Hawkins J R, Griffiths B L *et al.*. 1990. A gene from the human sex-determining region encodes a protein with homology toa conserved DNA-binding motif. *Nature*, 346:240-244.

Skaletsky H, Kuroda-Kawaguchi T, Minx P J, Cordum H S, Hillier L *et al.*. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423:825-837.

Stone A C, Griffiths R C, Zegura S L and Hammer M F. 2002. High levels of Y-chromosome nucleotide diversity in the genus Pan. *Proceedings of the National Academy of Sciences of the United States of America*, 99:43-48.

Swofford D L. 2002. PAUP* ver 2.0.b10. Phylogenetic Analysis Using Parsimony and other methods. Sinauers Associates, Sunderland, Sunderlan MA.

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

Thompson H V and King C M. 1994. *The European rabbit: the history of a successful colonizer.* Oxford Science Publications. Oxford, UK.

Tucker P K and Lundrigan B L. 1991. Rapid evolution of the sex determining locus in Old World mice and rats. *Nature*, 364: 715-717.

Wallner B, Huber S and Achmann R. 2001. Non-invasive PCR sexing of rabbits (*Oryctolagus cuniculus*) and hares (*Lepus europaeus*). *Mammalian Biology*, 66:190-192.

Waterston R H, Lindblad-Toh K, Birney E, Rogers J, Abril J F *et al.*. 2002. Initial sequencing and comparative analyses of the mouse genome. *Nature*, 420:520-562.

White M J D. 1978. *Modes of Speciation*, W. H. Freeman & Co.

Whitfield L S, Lovell-Badge R and Goodfellow P N. 1991. Rapid sequence evolution of the mammalian sex-determining gene SRY. *Nature*, 364:713-715.

Wilder J A, Mobasher Z and Hammer M F. 2004. Genetic evidence for unequal population sizes of Human females and males. *Molecular Biology and Evolution*, 21:2047-2057.

Zenger K R, Richardson B J and Vachot-Griffin A M. 2003. A rapid population expansion retains genetic diversity within European rabbits in Australia. *Molecular Ecology*, 12: 789-794.

6. FINAL REMARKS AND PERSPECTIVES

6.1 Final Remarks

With this work we believe that we have greatly contributed to the knowledge of the evolutionary history of the European rabbit. We feel that we have given valuable insights into the peculiar mode of molecular evolution of the Y chromosome, and the SRY gene in particular. Also, by studying patterns of genetic variation at mitochondrial, Y-linked and X-linked loci we were able to disentangle the male and female contributions to demographic processes. Finally, we think that with the patterns of introgression and divergence described here we have contributed to the still little explored study of speciation in mammals.

By sequencing a BAC clone that maps to the Y chromosome of the European rabbit we showed that the SRY gene is duplicated in the rabbit genome and that it is located in the arms of a palindrome. This organization resembles the organization of most of the euchromatic region of the human MSY where genes that are mostly or exclusively expressed in testis are present in multiple copies and evolve in concert through gene conversion (Skaletsky *et al.* 2003; Rozen *et al.* 2003). In humans the SRY gene is a single copy gene and mostly expressed during embryogenesis in the gonadal primordium. Our observation that in *Oryctolagus cuniculus* this gene is duplicated and that these copies evolve in concert through gene conversion suggests that in other groups the sex determining gene might be protected from degeneration by the same mechanism that prevents the majority of Y-linked genes from degeneration: intrachromosomal recombination. Although we have strong evidence that this duplication predates the Leporid radiation, we hope to demonstrate it beyond reasonable doubt in the near future. We also hope to be able to accurately estimate the rate of gene conversion between palindrome arms of the rabbit Y chromosome.

It has long been realized that the rabbit mitochondrial genome is composed of two highly divergent lineages (Biju-Duval *et al.* 1991) that correspond to the subspecies *Oryctolagus cuniculus cuniculus* in northeastern Iberian Peninsula and most recently colonized areas and *O. c. algirus* in southwestern Iberian Peninsula (Branco *et al.* 2000). Additional evidence for this dichotomy came from studies of protein polymorphism (Ferrand and Branco 2006). Here we performed a detailed analyses of levels and patterns of nucleotide polymorphism at six loci and fully corroborated this view of the rabbit genome. At all six genes two divergent lineages were observed. In regions of inferred high levels of recombination the signal of two

divergent lineages was obscured by the existence of interlineage recombinants. We performed computer simulations of the evolutionary process conditioned on the estimates of polymorphism and recombination of our four X chromosome loci to investigate the likelihood that these lineages could have arisen in a single panmitic population. We found that for three of them the data was incompatible with this scenario and that evolution in allopatry was required to explain the observed patterns of nucleotide variability. This is consistent with the evolution in different glacial refugia in Iberia of the two rabbit subspecies. We dated the divergence of these lineages using maximum likelihood models of nucleotide substitution and found that for most loci (Y-linked SRY gene excluded) this divergence could be placed in the Pliocene/Pleistocene boundary (approximately 2 MYA or more depending on the locus).

Besides the high levels of divergence between rabbit genetic lineages, another interesting aspect of previous genetic studies of rabbit populations was the consistent observation of high levels of genetic polymorphism. In this work we used a coalescent approach to estimate the effective population size of rabbit populations. Estimates of effective population size were very large in relation to other mammals. We found that the effective population size of *algirus* was higher than the effective population size of *cuniculus* irrespectively of the linkage group of the loci analyzed. This means that male and female population sizes of *algirus* populations are higher than that of *cuniculus* populations. Estimates of population size derived from Y chromosome data were always lower than estimates derived from mtDNA. This suggests that the female effective population size of rabbit populations is higher than the male effective population size and that the ratio of female to male can be as high as five to one. This is in agreement with previous behavioral observations that indicate that rabbit social groups present strong hierarchies where a dominant male gets most of the mattings that occur in that social group.

Finally, previous behavioral and genetic data suggested that female rabbits are philopatric while males undergo natal dispersal. Here we used for the first time nucleotide variability (in the same sample of individuals) at the paternally inherited SRY gene and the maternally inherited Cytb gene to directly compare levels of genetic differentiation between populations. We found that rabbit populations are clearly more differentiated for the Cytb gene suggesting that gene flow between populations of the same subspecies of rabbit is mostly male driven. Interestingly we

found that the same does not apply to levels of differentiation between subspecies of rabbit. Here levels of genetic differentiation are very high for both maternally and paternally inherited markers. This suggests that selection may act to prevent the introgression of some genomic regions between subspecies of rabbit and the Y chromosome stands out as one such possible region.

The most striking contrast in levels of introgression between subspecies of rabbit came from the four X-linked loci studied here. The two telomeric genes showed higher levels of nucleotide polymorphism, higher levels of recombination and high levels of introgression between subspecies. The centromeric loci showed the opposite. The comparison of these patterns with the Y and mtDNA data and other data previously published led to the suggestion that in the rabbit genome, areas of low recombination tend to show very low levels of introgression. The observation of lower introgression in areas of low, or no, recombination might simply reflect the effects of speciation genes to linked sites over long distances in such areas. Overall, it seems that these subspecies might be on the early stages of speciation.

Most of our knowledge about speciation comes from studies of fruit flies of the genus *Drosophila*. Studies of speciation are still rare in mammal species. One notable exception is the house mouse (*Mus musculus*). There are several parallels between the rabbit and the house mouse and both have been deeply impacted by human intervention in several ways. Aided by humans, they have colonized most of the world and both have been recently domesticated. This has resulted in the origination of several mouse inbred strains and rabbit domestic breeds. Both have been used for a long time as laboratory models for the study of human diseases, and both genomes are being sequenced (the house mouse genome being in much more advanced stages than the rabbit genome). Of great importance to the subject of this work is the fact that both have described subspecies that hybridize in nature (in *Mus* some authors consider two species instead of subspecies, *Mus domesticus* and *Mus musculus*). In the house mouse several studies have shown that both sex chromosomes show reduced introgression across the hybrid zone (*e.g.*, Tucker *et al.* 1992). A detailed survey of the introgression of the X chromosome showed that a region in the center of the X chromosome shows particularly low levels of introgression (Payseur *et al.* 2004). Independent approaches such as QTL mapping (Storchova *et al.* 2004) and SNP analyses (Harr 2006) identified this same region as being involved in reproductive isolation between species of house mouse. In *Mus* it seems that a few

genomic islands are specially differentiated between species and could harbor genes responsible for reproductive isolation (Harry 2006). In our work we showed that levels of introgression at Y-linked, mtDNA and two centromeric X-linked loci are greatly reduced in the European rabbit. A thorough genome scan for levels of introgression between subspecies of rabbits would help us understand whether in the rabbit reproductive isolation derives from incompatibilities at just a few loci of large effect, or if in the other hand, most of the genome is unable to introgress between subspecies. From our preliminary comparisons of allozyme data with the six loci studied here, it seems that in the rabbit there might be only a few regions involved in reproductive isolation between subspecies.

6.1.2 Perspectives

The completion of this thesis happens in an important transition phase for the study of *Oryctolagus cuniculus*. The first integrated cytogenetic map of the European rabbit has just been completed (Chantry-Darmon 2005), the first genetic map is in progress and the first draft of the rabbit genome was released online in the end of 2005. A joint project between CIBIO (Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto) and IREC (Instituto de Investigación en Recursos Cinegéticos) is now underway that will make possible the study of intersubspecific crossings in the European rabbit. Finally, the first genome wide survey for regions involved in reproductive isolation has just began as a joint effort between CIBIO and the Department of Ecology and Evolutionary Biology at the University of Arizona. We hope that the completion of both these studies will help us understand the nature of reproductive isolation that we have just started to uncover and that these studies, together with those in the house mouse, will shed light into what Darwin named "the mystery of mysteries".

6.3 References

Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, Soriguer R C, Gaaied A, Hili A and Mounolou J C. 1991. Mitochondrial DNA evolution in Lagomorphs: Origin of systematic heteroplasmy and organization of diversity in european rabbits. *Journal of Molecular Evolution*, 33: 92-102.

Branco M, Ferrand N and Monnerot M. 2000. Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome *b* gene. *Heredity*, 85: 307-317.

Chantry-Darmon C. 2005. Construction d'une carte integree genetique et cytogenetique chez le lapin europeen (*Oryctolagus cuniculus*): Application a la primo localization de la mutation rex. PhD thesis. Université de Versailles Saint-Quentin. France.

Ferrand N and Branco M. 2006. The evolutionary history of the European rabbit (*Oryctolagus cuniculus*), *in* Phylogeography in Southern European Refugia. Edited by S. Weiss and N. Ferrand. Springer. Amsterdam, The Netherlands. (in press).

Harr B. 2006. Genomic islands of differentiation between house mouse subspecies. *Genome Research*, DOI/10.1101/gr.5045006.

Payseur B A, Krenz J G and Nachman M W. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two subspecies of house mice. *Evolution*, 58: 2064-2078.

Rozen S, Skaletsy H, Marszalek J D, Minx P, Cordum H S *et al.*. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, 423:873-876.

Skaletsky H, Kuroda-Kawaguchi T, Minx P J, Cordum H S, Hillier L *et al.*. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423:825-837.

Storchova R, Gregorova S, Buckiova D, Kyselova V, Divina P, and Forjeit J. 2004. genetic analyses of X-linked hybrid sterility in the house mouse. *Mammalian Genome*, 15: 515-524.

Tucker P K, Sage R D, Warner G, Wilson A C and Eicher E M. 1992. Abrupt cline for sex chromosomes between two species of mice. *Evolution*, 46:1146-1163.