Genome admixture with massive mitochondrial DNA introgression in hares (Lepus spp.) : the relative roles of demography and natural selection.

Fernando Antonio Pereira da Silva Ferreira Seixas

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Génétique et Génomique

École doctorale GAIA

Unité de recherche ISEM

Genome admixture with massive mitochondrial DNA introgression in hares (*Lepus* spp.): the relative roles of demography and natural selection.

Présentée par Fernando SEIXAS
Le 28 Novembre 2017

Sous la direction de Pierre BOURSOT
et José MELO-FERREIRA

Devant le jury composé de

Nuno FERRAND, Full Professor, CIBIO, UP	Président du Jury
José MELO-FERREIRA, Assistant Professor, CIBIO, UP	Directeur de Thèse
Pierre-André CROCHET, Directeur de Recherche, CEFE, CNRS	Représentant de l'Université de Montpellier
Yves VIGOUROUX, Directeur de Recherche DIADE, IRD	Représentant de l'Université de Montpellier
Laurence DESPRÉS, Professeur, LECA, Université Grenoble Alpes	Examinateur, Rapporteur
Rémy PETIT, Directeur de Recherche, BIOGECO, INRA	Examinateur, Rapporteur

UNIVERSITÉ DE MONTPELLIER

Genome admixture with massive mitochondrial DNA introgression in hares (*Lepus* spp.): the relative roles of demography and natural selection.

Fernando Seixas
Doutoramento em Biodiversidade, Genética e Evolução
Departamento de Biologia
2017

Orientador
José MELO-FERREIRA, Phd, Auxiliary Researcher
CIBIO-InBIO, Laboratório Associado, Universidade do Porto

Coorientador
Pierre BOURSOT, CNRS research director,
Institut des Sciences de l'Évolution, Université de Montpellier

Foreword

In compliance with the no. 2 of article 4 of the General Regulation of Third Cycles of the University of Porto and with the article 31 of the Decree-Law no. 74/2006, of March, with the alteration introduced by the Decree-Law no. 230.2009, of 14 September, the results of already published works were totally used and included in some of the chapters of this dissertation. As these works were performed in collaboration with other authors, the candidate clarifies that, in all these works, participated in obtaining, interpreting, analyzing and discussing the results, as well in the writing of the published forms.

This is a joint doctorate between the University of Porto and University of Montpellier. The Faculdade de Ciências da Universidade do Porto was the home institution of the candidate, and the work was directed by Dr. José Melo-Ferreira, Auxiliary Researcher at Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), and co-directed by Dr. Pierre Boursot, Director of Research at the Institut des Sciences de l'Evolution de Montpellier (ISEM), Université de Montpellier.

This work was financially supported by Fundação para a Ciência e a Tecnologia (FCT) through the attribution of a PhD fellowship to the candidate with reference SFRH/BD/87126/2012, funded by Programa Operacional Potencial Humano from the European Social Fund and Portuguese Ministério da Ciência e Ensino Superior; and the post-doctoral fellowship with reference SFRH/BPD/43264/2008 and Investigador FCT grant with reference IF/00033/2014, funded by Programa Operacional Potencial Humano from the European Social Fund and Portuguese Ministério da Ciência e Ensino Superior to José Melo-Ferreira. Project grants were co-funded by FCT, Fundação para a Ciência e a Tecnologia (Portugese national funds; FCT-ANR/BIA-EVF/0250/2012) and ANR, Agence Nationale de la Recherche (ANR-12-ISV7-0002-01); and by Portuguese national funds through the FCT and Fundo Europeu de Desenvolvimento Regional through COMPETE (PTDC/BIA-EVF/115069/2009 and PTDC/BIA-EVF/111931/2009). Mobility grants were funded by the PESSOA program for transnational cooperation (FCT and Campus France) and Programa de Ações Universitárias Integradas Luso-Francesas (Conférence des Présidents d'Université - CPU - and Conselho de Reitores das Universidades Portuguesas - CRUP). This thesis was conducted in the frame of the International Associated Laboratory (LIA) "Biodiversity and Evolution" supported by InEE (Institut Ecologie et Environnement, CNRS, France) and FCT (Fundação para a Ciência e a Tecnologia, Portugal).


FCT
Fundação
para a Ciência
e a Tecnologia
ANR
UNIÃO EUROPEIA
Fundo Social Europeu
POPH
PROGRAMA OPERACIONAL POTENCIAL HUMANO
COMPETE
QREN
QUADRO
DE REFERÊNCIA
ESTRATÉGICO
NACIONAL
LIA
BIODIVERSITY
AND EVOLUTION
SCIENTIFIC MEETING 2017

Acknowledgements/ Agradecimentos

All the work presented in this dissertation results of the joint effort of many people and institutions to whom I present my sincere thank you.

Ao meu orientador, Dr José Melo-Ferreira, agradeço por todo o apoio que nunca faltou ao longo de todos estes anos. Obrigado por todo o tempo, trabalho e dedicação investidos nesta tese mas também em mim e que me ajudaram a melhorar quer a nível cientifico mas também, e de forma bastante importante, a nível pessoal. Digo-o de boca cheia em conversas e por isso o digo aqui que sinto que aprendi mais contigo neste tempo em que trabalhamos juntos do que nos meus 3 anos de licenciatura. Um honesto e muito sentido obrigado!

To my supervisor at the University of Montpellier, Dr. Pierre Boursot, I thank you first for the way you received me in Montpellier and in your house. I will never forget the first days at your house when we carried all that wood and the very next morning when you took me to my very first mushrooms “hunt”. Also, thank you for your support and for the way you always pushed me to think critically about my work, and all your inspiring insights. Hearing you talk about science is always a pleasure.

Ao Professor Doutor Paulo Célio Alves, obrigado pelo seu constante entusiasmo e o apoio.

À equipa das lebres deixo-vos aqui um muito obrigado por estes anos de convívio e trabalho conjunto. Começo pela Liliana, o elemento que me acompanha na equipa desde que nela me iniciei. Obrigado pelo teu acompanhamento atento e ensinamentos nos meus primeiros tempos no CIBIO. E já depois de me tornar independente no laboratório obrigado pela tua disponibilidade para responderes a todas as questões e mais alguma e em todos os momentos, por estares sempre pronta a discutir comigo qualquer ideia e por nunca me deixares estar fora do acontecimento científico mais recente. À Mafalda, a minha mais recente companheira de congressos, obrigado pela tua boa disposição e por alinhares em todas as minhas brincadeiras, mesmo quando mais ninguém as percebe. Ao João, foi um prazer conviver contigo desde o momento em que entraste para o grupo e ainda mais agora que partilhamos o gabinete. É bom poder ter alguém que também fale a nossa linguagem “paralela” da bioinformática. À Sara, obrigado pela tua simpatia e sorriso constantes, fazes falta cá em Portugal connosco. To Iwona, our most recent acquisition, it’s been a pleasure to discuss with

you every steps of the work our during our lunch time and to share some Eureka moments. A todos um obrigado por aturarem as minhas excentricidades.

Julien, to you I owe a very special thank you and probably much more. You have been the best friend I could ask for in Montpellier. You are one of the most enlightening people I ever met and the way you think about life and fully live it never stops to amaze me. I could feel that I was learning with you in every discussion we had. Thank you for teaching me how to climb, to help me to know more about programming, and for supporting me as a true friend in every moment. And here I cannot obviously forget to thank Laure, for the times both of you received me at your house.

Aos meus amigos de cá, é sempre com alegria que regresso e vos encontro como se fosse ontem. A vocês Susana, Rabaxal, Catarina, Tiago, João, Luís, Vanessa, Filipa também um obrigado.

À minha fantástica família um obrigado muito especial por serem o meu grande alicerce. Quem eu sou hoje e onde cheguei deve-se inteiramente à educação que me deram e inteira dedicação e esforço para que pudesse seguir os meus sonhos.

A ti Ana um obrigado por estares sempre ao meu lado em qualquer circunstância. Obrigado por esperares por mim, pelos teus abraços e mimos que me confortam sempre. Simplesmente obrigado.

Summary

Understanding how and why closely related species continue to exchange genes after they attained partial reproductive isolation provides major insights into the process of species formation and other important evolutionary processes, such as the demographic history of species, interspecific interactions and local adaptation. Introgressive hybridization is a common phenomenon in nature but the causes and consequences of interspecific gene flow are not yet fully understood. In particular, mitochondrial DNA has been widely implicated in cases of introgression, but strongly contrasting explanations for this pattern have been put forth. Two major general hypothesis have emerged. One, supported by theoretical simulations, suggests that during the replacement of a species by an invading one, with hybridization occurring, markers linked to the least dispersing sex, which are often females, tend to introgress farther into the invading population. Given that mitochondria play an important role in producing cellular energy that require the action of co-adapted protein complexes encoded both by the nuclear and mitochondrial DNA, a second hypothesis suggests that introgression of mitochondrial variants is adaptive. These demographic and selective hypotheses can be generalized to the rest of the genome, and be major determinants of introgression.

Hares (*Lepus* spp.) provide an appropriate model to study introgressive hybridization, and in particular of mtDNA. Over 30 species of hares are distributed in the world, and share numerous contacts where they can hybridize. Most cases of mtDNA introgression described so far involve a single species as donor, the mountain hare (*Lepus timidus*), which is widely distributed in northern Eurasia. Introgression occurred in current contacts with the species, but also in regions where *L. timidus* no longer exists but was present during the Pleistocene glaciations, such as southern France or the Iberian Peninsula. These studies questioned whether neutral demography explains multiple cases of mtDNA introgression or a selective advantage of the *timidus* mtDNA would need to be invoked. To tackle this question, we first determined whether such phenomenon extends to other species and geographic regions. We focused on North American hares, and analyzed nuclear and mtDNA sequences of three species widely distributed in the region: the snowshoe hare (*L. americanus),* the black-tailed jackrabbit (*L. californicus*), and the white-tailed jackrabbit (*L. townsendii*). Previous population genetics work had shown that *L. americanus* was composed by three evolutionary units, but one of the mtDNA clades is more closely related with mtDNA haplotypes of *L.*

californicus. Using multilocus coalescent-based approaches, we reconstructed the speciation history of these species and found that the three units in *L. americanus* are deeply divergent but monophyletic. Using coalescent simulations, we determined the distribution of expected mtDNA distances under a strict incomplete lineage-sorting model, and show that the mtDNA proximity to *L. californicus* can only be explained by introgression. This introgression is historical and massive in some populations. We conclude that mtDNA introgression is widespread on hares and not restricted to particular environments or lineages, reinforcing the interest to understand the underlying general mechanisms.

We then focused our research efforts on the hare system of the Iberian Peninsula. Three extant species in Iberia have been massively affected by historical mtDNA introgression from *L. timidus*, when the latter species was present in the region. In the Iberian hare, *L. granatensis*, introgression follows a northwards gradient, from absent in the south to predominant in the north. In Iberian *L. europaeus* and *L. castroviejoi* mtDNA introgression is fixed or nearly fixed. Contrasting with the mtDNA introgression patterns, shallow inspection of the nuclear DNA suggested rare but widespread introgression. Competing demographic and selective hypothesis have been proposed to explain these patterns. This provides a suitable model to understand whether i) mtDNA and nuclear introgression can be reconciled under a single demographic scenario (and potentially explain cases of massive mtDNA introgressions in nature); ii) preferential introgression of nuclear genes with mitochondrial functions occurred (i.e. that nuclear-mitochondrial co-adaptation drove introgression); and iii) genome-wide patterns of historical introgression have been governed by natural selection (acting to prevent or promote introgression). The fact that introgression is repeated in multiple species in northern Iberia provides the power of a replicated experiment to tackle these questions.

The complete genomes of 10 *L. granatensis*, 10 *L. europaeus*, 4 *L. timidus* and 1 *L. americanus* (used as outgroup for some analyses) were sequenced. We used analyses of introgression tracts and geographic distribution of introgression frequencies to infer the complex history of species interactions in northern Iberia. In *L. granatensis*, we found genomic variation that is compatible with a northwards range expansion of the species, and a subtle gradient of increasing frequencies of introgression from *L. timidus* to the north (though nuclear DNA introgression frequencies were low). Geographic explicit coalescent simulations showed that this results from the invasion and replacement of *L. timidus* by *L. granatensis* in northern Iberia, after the last glacial maximum. The northwards increase of introgression tracts also supports this scenario.

Further simulations showed that mtDNA introgression could be reconciled under this single demographic model, if female philopatry and asymmetric introgression are included in the model. This suggests that invasive range replacements are a major determinant of introgression patterns and may account for strong cytonuclear discordances in introgression frequencies.

Given that introgression from *L. timidus* affected multiple species in northern Iberia, we used the sizes of ancestry tracts and transition between tracts of different origin to reconstruct the complex history of hybridization between the species. The first introgression events occurred between *L. granatensis* and *L. timidus* in northern Iberia 7 kY ago. Later, 4kY ago, *L. europaeus* hybridized with *L. timidus* on its way to Iberia where it replaced *L. granatensis* 1 kY ago. Predominance of *timidus*-*europaeus* over *timidus*-*granatensis* transitions found in *L. europaeus* suggests that this species contacted and hybridized directly with *L. timidus*, most likely outside Iberia, and only a small portion of *timidus* introgression was secondarily transmitted through hybridization with *L. granatensis*. However, mtDNA introgression from *L. granatensis* (bearing *timidus* mtDNA) into *L. europaeus* during the invasion of Iberia most likely explains massive mtDNA introgression into Iberian *L. europaeus* (supported by low differentiation of mtDNA between *L. europaeus* and *L. granatensis*). Therefore, *L. timidus* mtDNA represents the ancient distribution of the species in Iberia, as previous ecological modelling had suggested.

No predominant introgression of nuclear genes with mitochondrial functions was found either in *L. granatensis* or in *L. europaeus*, suggesting that this was not a major determinant of general patterns of introgression. However, we did find nuclear genes for which introgression resembles that of mtDNA (one gene, MRPL13, common to both species). These are candidates for cyto-nuclear co-introgression but the possible adaptive relevance of these introgressions must await future studies, including functional assays.

Finally, we found that natural selection strongly determined local genomic patterns of introgression. Depletion of introgression on the X-chromosome and near chromosome centers in *L. granatensis* shows that introgression was prevented by linkage to incompatibility factors, as commonly inferred in other model systems. In addition, we found genes with outlier frequencies of introgression (as determined by simulations in *L. granatensis*), which indicate introgression promoted by natural selection. The inspection of their functions revealed a predominance of genes involved in the immune system and that influence male fertility. Adaptive introgression in immune

system genes may have facilitated adaptation to new pathogenic environments to which *L. timidus* was previously adapted. Male fertility functions invoke a different process, which may be related with compensatory introgression of variants to minimize male deleterious effects of massive mtDNA introgression (this hypothesis is less strong in *L. europaeus*, given the nature of the introgressed genes). Though involving different genes, the similarities of functions of predominantly introgressed genes in *L. granatensis* and *L. europaeus* is noteworthy and may underlie repeated selective processes.

This work shows that broad genomic patterns of introgression, including massive mtDNA introgression, may be strongly determined by the demographic history of the interacting species. However, introgressed variants are an important source of new genetic variation upon which natural selection can act, either promoting or impeding genetic exchanges at local genomic scales.

Sumário

Perceber como e a razão pela qual as espécies continuam a trocar genes depois de atingirem isolamento reprodutivo parcial pode fornecer informações fundamentais sobre o processo de formação das espécies entre outros processos evolutivos relevantes, como a história demográfica das espécies, interações interespecíficas e adaptação local. Apesar da hibridação introgressiva ser um fenómeno comum na natureza, as causas e consequências do fluxo interespecífico de genes não são ainda totalmente claras. Em particular, apesar de o ADN mitocondrial (ADNmt) estar frequentemente implicado em casos de introgressão, as hipóteses avançadas para explicar este fenómeno são bastante dispares, das quais duas se destacam. Uma, que é apoiada por estudos teóricos baseados em simulações, sugere que quando uma espécie invade o território de uma outra espécie residente, a substitui e estas espécies trocam genes, os marcadores ligados ao sexo que menos tende a dispersar (geralmente as fêmeas) são mais propensos a introgredir. Dado que as mitocôndrias desempenham um papel fundamental na produção de energia celular (o que por sua vez requer a ação de complexos de proteínas coadaptadas que são codificados tanto pelo ADN nuclear como pelo ADNmt) uma segunda hipótese sugere que a introgressão dos variantes mitocondriais tem uma função adaptativa. Os processos demográficos e seletivos associados a estas hipóteses podem ser generalizadas para explicar situações de introgressão massiva no resto do genoma sendo fatores determinantes da introgressão.

As lebres (género *Lepus*) são um modelo particularmente adequado para estudar a relevância da introgressão na evolução. Atualmente o género é composto por mais de 30 espécies descritas que ocupam uma grande variedade de habitats. O género é ainda caracterizado por um grande número de casos de fluxo interespecífico de genes. A maioria dos casos descritos envolvem a introgressão do ADNmt da lebre da montanha (*Lepus timidus*), uma espécie boreal atualmente distribuída no Norte da Eurásia e Alpes. Estas situações de introgressão dizem respeito tanto a zonas em que as espécies atualmente contactam, mas também a áreas em que a espécie *L. timidus* não existe nos dias de hoje mas onde esteve presente durante as glaciações do Pleistoceno, como por exemplo o sul de França e a Península Ibérica. Uma questão levantada pelos estudos anteriores que descreveram estes casos de introgressão, foi se processos demográficos neutrais poderiam explicar os vários casos de introgressão de ADNmt observados entre as lebres ou se uma vantagem seletiva do ADNmt da *L. timidus* teria de ser invocada. Para abordar esta questão, neste estudo começamos

por determinar se este fenómeno de introgressão do ADNmt se estende a outras espécies e regiões geográficas. Para tal focamo-nos nas lebres Norte Americanas para as quais analisamos sequências de ADN nuclear e mitocondrial das três espécies com a maior distribuição nesta região: a lebre-americana (*L. americanus*), a lebre-da-Califórnia (*L. californicus*) e a lebre-de-cauda-branca (*L. townsendii*). Trabalhos anteriores utilizando análises de genética populacional mostraram que a *L. americanus* é composta por três unidades evolutivas, mas um dos clados é geneticamente mais próximo ao de *L. californicus*. Usando uma análise de coalescência baseada em múltiplos marcadores moleculares, reconstruimos a história de especiação destas três espécies e confirmamos a existência de três linhagens em *L. americanus* que são monofiléticas mas profundamente divergentes. Usando simulações de coalescência determinamos ainda que a proximidade de um dos clados de *L. americanus* a *L. californicus* no ADNmt não pode ser explicada por coalescência incompleta de linhagens num cenário estrito de divergência só podendo ser explicada por introgressão. Esta introgressão é histórica e massiva em algumas populações. Assim, concluímos que a introgressão de ADNmt é ubíqua nas lebres e não restrita a ambientes ou linhagens particulares, reforçando o interesse em perceber o mecanismo geral na base deste fenómeno.

Neste sentido, focamos então o nosso estudo nas lebres do Velho continente, mais particularmente da Península Ibérica. As três espécies atualmente residentes nesta região foram massivamente afetadas por introgressão histórica do ADNmt de *L. timidus*, que remonta a uma época em que esta espécie estava presente na região. Na lebre-Ibérica, *L. granatensis*, a frequência de introgressão mitocondrial segue um gradiente, desde ausente no Sul a predominante no norte da Península Ibérica. Em populações de lebre-europeia (*L. europaeus*) residentes na Península Ibérica e na lebre-cantábrica (*L. castroviejoi*) o ADNmt de *L. timidus* está fixado ou quase fixado. Em contraste com os padrões de introgressão do ADNmt, uma inspeção superficial do ADN nuclear sugeriu que a introgressão nuclear é rara mas geograficamente dispersa. Tanto hipóteses demográficas como seletivas foram propostas para explicar estes padrões. Neste contexto, o caso das lebres da Península Ibérica apresenta-se como um modelo bastante adequado para tentar perceber: i) se os padrões de introgressão nuclear e mitocondrial podem ser reconciliados sob um modelo demográfico único (e potencialmente explicar casos de introgressão massiva do ADNmt); ii) se ocorreu introgressão preferencial de genes nucleares com funções mitocondriais (ou seja, se a coadaptarão cito-nuclear levou a co-introgressão); e iii) se padrões genómicos de

introgressão histórica foram determinados por seleção natural (a atuar para promover ou prevenir introgressão). O facto da introgressão ocorrer em múltiplas espécies no norte da Península Ibérica dá-nos o poder de uma experiência replicada para abordar estas questões.

Para abordar estas questões sequenciamos os genomas completos de 10 *L. granatensis*, 10 *L. europaeus*, 4 *L. timidus* e 1 *L. americanus* (usada como "outgroup" em algumas análises). A história de interações entre espécies no norte da Península Ibérica foi inferida através da análise de fragmentos de introgressão e da distribuição geográfica das frequências de introgressão. Em *L. granatensis* encontramos padrões de variação genética compatíveis com uma expansão para norte desta espécie, bem como um gradiente subtil das proporções de introgressão (apesar destas serem baixas no genoma nuclear), aumentando de sul para norte. Através de simulações de coalescência geograficamente explícitas demonstramos que estes padrões resultam de uma invasão e substituição de *L. timidus* por *L. granatensis* no norte da Península Ibérica, após a o Último Máximo Glaciar. Além disso, as simulações mostraram que os padrões de introgressão mitocondrial observados podem ser explicados por este mesmo cenário demográfico se se considerar filopatria das fêmeas e assimetrias de introgressão. Estes resultados sugerem que a expansão de uma espécie invasora para o território de uma outra e consequente substituição da espécie residente pode ser um facto determinante dos padrões de introgressão e explicar discordâncias acentuadas das frequências de introgressão entre os genomas nucleares e mitocondrial.

Sendo que a introgressão de *L. timidus* afetou múltiplas espécies no norte da Península Ibérica, usamos o tamanho dos fragmentos de introgressão e a informação baseada na transição entre fragmentos com diferentes origens (junções) para reconstruir a história de contactos e hibridação entre estas espécies. Os primeiros eventos de introgressão ocorreram entre *L. granatensis* e *L. timidus* no norte da Península Ibérica há 7000 anos atrás. Mais tarde durante a sua expansão pela Europa, a *L. europaeus* contactou primeiro com a *L. timidus*, há 4000 anos atrás, e substituiu a *L. granatensis* há 1000 anos atrás já dentro da Península Ibérica. A predominância de junções *timidus-europaeus* em relação a junções *timidus-granatensis* encontradas em *L. europaeus* da Península Ibérica sugere que a introgressão de porções do genoma desta espécie com origem em *L. timidus* resultou do contacto direto entre as duas (provavelmente fora da Península Ibérica) e que só uma porção da introgressão de *L. timidus* foi transmitida secundariamente por contacto com *L. granatensis* (que já estaria afetada por introgressão de *L. timidus*). No entanto, a introgressão massiva de ADNmt

do tipo *timidus* nas *L. europaeus* da Península Ibérica resulta muito provavelmente do contacto com *L. granatensis* (introgredidas com ADNmt do tipo *timidus*) durante a invasão da Península Ibérica pela *L. europaeus*. Esta hipótese é suportada pela baixa diferenciação ao nível do ADNmt entre *L. granatensis* e *L. europaeus*. Deste modo, a introgressão de ADNmt do tipo *timidus* nas espécies da Península Ibérica representa a distribuição histórica de *L. timidus* nesta região, tal como sugerido por estudos de modelação do nicho ecológico desta espécie.

Não encontramos introgressão predominante de genes nucleares com funções mitocondriais nem em *L. granatensis* nem em *L. europaeus* o que sugere que a coevolução cito-nuclear não é um facto determinante dos padrões gerais de introgressão nestas espécies. No entanto, encontramos genes nucleares para os quais os padrões de introgressão são semelhantes aos do ADNmt (um gene, MRPL13, é comum a ambas as espécies). Estes genes correspondem potencialmente a casos de co-introgressão cito-nuclear mas a qualquer relevância adaptativa destas introgressões terá de ser confirmada por estudos futuros incluindo experiências funcionais.

Finalmente, verificamos que a seleção natural determina os padrões locais de introgressão. A depleção da introgressão no cromossoma X e perto do centro dos cromossomas em *L. granatensis* demonstra que a introgressão foi preferencialmente impedida ou dificultada em regiões onde as ligações a fatores de incompatibilidade são mais fortes, algo frequentemente observado noutros sistemas. Além disso, encontramos genes com frequências de introgressão extremas (isto é, não esperadas ou previstas de acordo com as nossas simulações) o que indica que a introgressão foi promovida por seleção natural. A inspeção das funções destes genes demonstrou uma predominância de genes envolvidos no sistema imunitário e que influenciam a fertilidade masculina. A introgressão adaptativa de genes do sistema imunitário poderá ter facilitado a adaptação a novos ambientes patogénicos para os quais *L. timidus* estaria previamente adaptada. Funções relacionadas com a fertilidade masculina invocam um processo diferente, que poderá estar relacionado com a introgressão de variantes compensatórios para minimizar efeitos nocivos nos machos em consequência da introgressão massiva do ADNmt de *timidus* (esta hipótese é menos robusta em *L. europaeus* dada a natureza dos genes introgredidos). Embora diferentes genes estejam envolvidos, as semelhanças nas funções dos genes predominantemente introgredidos em *L. granatensis* e *L. europaeus* é digna de nota e pode ter por base processos seletivos repetidos.

Este trabalho demonstra que padrões gerais de introgressão genómica, incluindo introgressão massiva do ADNmt, podem ser fortemente determinados pela história demográfica das espécies. No entanto, os variantes introgredidos podem também funcionar como uma importante fonte de variação genética sobre a qual a seleção natural pode atuar, tanto promovendo como impedindo o fluxo genético a escalas genómicas e locais.

Extended summary

Closely related taxa in different groups of organisms often show a history of introgressive hybridization. The patterns of introgression are heterogeneous across the genome, since the exchange of genetic material depends on the fitness effects of the regions being exchanged or that of linked regions. For instance, regions involved in species-specific adaptations to their local environments or involved in reproductive isolation are less likely to cross the species barrier. However, if genomic regions are dissociated from these incompatibility regions and are themselves neutral in the recipient species environment or genomic background, then these may more freely introgress. Furthermore, introgression may be promoted in genomic regions that increase the recipient species fitness. Therefore, the genomes of many species pairs may remain semipermeable to gene flow for some time after their initial divergence. The study of patterns of genetic exchanges along the genome can help identify barrier loci, and thus unravel their function and the forces driving speciation, but can also reveal the role of hybridization as a source of novelty with adaptive potential. While these questions have long been a major interest of evolutionary biologists (for instance since early studies of hybrid zones), available data were far from allowing their satisfactory resolution, but the current availability of genomic datasets now allows to tackle them with unprecedented power.

Hares (genus *Lepus*) are a particularly suitable model to study the relevance of introgression in evolution. The genus diversified via rapid radiation, with over 30 described species now occupying a wide range of habitats, and it is characterized by numerous instances of interspecific gene flow. Most of the described cases involve introgression of the mountain hare (*Lepus timidus*) mtDNA, a boreal species currently widely distributed in northern Eurasia and the Alps. The mtDNA of this species introgressed into the four temperate species that inhabit Europe, and has possibly introgressed into at least four other in China. Instances of introgression were described in cases where the species currently contact (e.g. between *L. timidus* and the brown hare (*L. europaeus*) in Sweden, Russia or Alps), but also in cases of past contacts (e.g. involving *L. timidus* and hare species from the Iberian Peninsula). The latter cases result from the distribution of *L. timidus* during Pleistocene glaciations reaching southern Europe, including southern France and the Iberian Peninsula, as attested by paleontological records. Interestingly, in these cases, *timidus* mtDNA introgression can be extreme, reaching very high frequencies in some populations (as is the case of *L.*

granatensis and *L. europaeus* populations in northern Iberia) or leading to complete replacement of the native mitochondria (as in the cases of *L. castroviejoi* and *L. corsicanus*).

The phenomenon of mtDNA introgression appeared frequent in Europe and Asia, and to mostly imply introgression from the arctic lineage into several other species. There was also suspicions of introgression among some Northern American species, so we first wanted to clarify and compare this situation in the New World. Of the more than 30 hare species, nine occur in North America with very distinct habitats and range sizes, and with sometimes overlapping ranges. Among the ones with widest distribution is the snowshoe hare (*L. americanus*), which occupies most of Canada and the Pacific Coast Range and the Rockies in United States. In the South, it is replaced by the black-tailed jackrabbit (*L. californicus*), which occupies the western part of United States and the northern half of Mexico. The white-tailed jackrabbit (*L. townsendii*) in the central part of North America overlaps both with *L. americanus* and *L. californicus*. In fact, *L. californicus* and *L. townsendii* have been suggested to hybridize in the wild but this has never been assessed by genetic studies. A previous work based on microsatellite data, suggested that *L. americanus* comprises three evolutionary units, one occupying the boreal region (Boreal), another the Rocky Mountains (Rockies), and yet another the Pacific Northwest region of USA (PacNW). Interestingly, the PacNW populations make up a mtDNA clade that is more closely related to *L. californicus*, than to the other conspecific clades. Such a pattern is suggestive of mtDNA introgression, but the possibility that this resulted from incomplete lineage sorting (ILS) could not be ruled out. To tackle this question we analyzed a multi-locus dataset including markers with all inheritance patterns (mitochondrial DNA, autosomal, X and Y-linked) sequenced in individuals representative of the three *L. americanus* groups, *L. californicus* and *L. townsendii*.

Using coalescent-based phylogenetic inference methods applied to the nuclear loci we confirm that the three evolutionary units previously inferred in *L. americanus* have genealogical significance, and that particularly the Boreal clade diverged from the other two to the same extent as other bona fide *Lepus* species. Using an Isolation with Migration (IM) model we further show that nuclear gene flow among these species is either null (between *L. californicus* and *L. townsendii*) or very limited (from *L. americanus* to *L. californicus* and *L. townsendii*). In contrast, coalescent simulations of mtDNA divergence using the parameter values inferred from the nuclear DNA markers show that the *L. americanus* PacNW mtDNA haplotypes are more closely related with those from *L. californicus* than expected considering ILS, and thus likely result from introgression.

Notably, *L. californicus* and *L. americanus* PacNW populations do not share haplotypes, suggesting that mitochondrial introgression resulted from ancient hybridization. Since one species is adapted to boreal forest and the other to arid regions, they may have been differently affected by past climatic oscillations, which could have promoted range replacements facilitating hybridization. In contrast to observations in their European counterparts, in these North American species mtDNA introgression occurred from a temperate to a boreal species. However similarly, introgression is geographically limited (Pacific Northwest) but massive, some populations reaching fixation. Interestingly, while *L. americanus* generally changes to a white coat in winter, specimens from the PacNW remain brown year-round, an apparent response to reduced snow-cover in that region. *L. californicus* stays brown also and could have transmitted this property to *L. americanus* together with mtDNA. The answer to this question must await genomic studies describing patterns of introgression genome-wide possibly in association with their functional context.

Introgression, particularly involving mtDNA, is therefore recurrent among hares, often found at high frequencies in the recipient species and it is not restricted to certain lineages and environments. This questions whether a single general mechanism could explain such a replicated evolutionary pattern. Two candidate hypotheses emerge: i) that the dynamics of range replacements between hybridizing species pairs promote mtDNA introgression, and ii) that selective advantages of mtDNA introgression and co-adapted gene complexes promoted massive introgression. We thus directed our work to understand the genomic impact of these ancient hybridization events that resulted in massive mtDNA introgression, using the Iberian system as a model

Three species of hares currently inhabit the Iberian Peninsula. The Iberian hare, *L. granatensis*, occupies most of the Peninsula, being only replaced in the extreme north by *L. castroviejoi*, in the Cantabrian mountains, and by *L. europaeus* from the Cantabrian mountains to the Pyrenean foothills. The range of *L. europaeus* extends towards Central Europe, into Scandinavia, Asia and the Middle East, while the other two species are endemic to Iberia. Populations of all three species in the Iberian Peninsula harbor high frequencies of *timidus* mtDNA. In *L. castroviejoi*, the introgressed *timidus* mtDNA is fixed, and in *L. europaeus* it is almost fixed in its Iberian populations, though not found elsewhere in its range, except where the two species form contact zones. In *L. granatensis*, *timidus* mtDNA follows a south-north gradient, being absent in the southern range of the species but reaching high frequencies in the north.

The repeated introgression of *timidus* mtDNA into the three species resident in the Iberian Peninsula, the massive frequency it reaches in some populations and the fact that it is restricted to the colder northern region of Iberian Peninsula raises the hypothesis that *timidus* mtDNA introgression results from adaptation to cold. In fact, mitochondrial metabolism is involved in thermoregulation, and mtDNA sequence variation has, in several instances, been associated with temperature-related adaptation, and there is some evidence of adaptive mitochondrial protein evolution along the arctic hare branch. However, massive introgression could be an incidental outcome of the process of species replacement, as could have occurred during the drastic post-glacial environmental changes. During expansion of one species into the range occupied by another, drift at the front of invasion can bring rare variants (including introgressed ones) to high frequencies, which can be further propagated by the expansion wave ("allele surfing" on the expansion wave). Previous population genetics data on mtDNA and a handful of nuclear markers had provided evidence in *granatensis* and *europaeus* of past waves of expansion and mitochondrial introgression from *timidus*, along geographic gradients (South-North in *granatensis*, East-West in *europaeus*). However, given the non-recombining nature of mtDNA, demonstrating the adaptive origin of its invasion based on its sole variations is impossible. We reasoned that given the intense collaboration of the mitochondrial and nuclear genomes in many key cellular processes, the two likely co-evolve and thus, whether adaptive or not, massive mtDNA introgression could have affected nuclear encoded genes functionally linked to the mitochondria. We therefore underwent a genome-wide study of genetic exchanges between the Iberian species. This wealth of data also allowed us to reconstruct the history of the interactions between the species, and to test quantitatively the scenarios of species replacements and their ability to explain observed introgression patterns of nuclear and mitochondrial genomes.

To tackle these questions we have sequenced whole genomes of specimens of two species from Iberia. Five *L. granatensis* came from the south, where no mtDNA introgression is observed, and five from the north along the south-north gradient of increasing *timidus* mtDNA introgression. We sequenced five *L. europaeus* from Iberia (where *timidus* mtDNA almost fixed) and five from elsewhere in Europe, from Southern France to Ukraine (not affected by mitochondrial introgression). We also sequenced the genomes of 4 *L. timidus* from the Alps, Ireland and Scandinavia, and one *L. americanus* to use as outgroup.

We inferred local ancestry along the genome of each specimen with the ELAI method, which uses linkage disequilibrium information and a Hidden Markov Model to segment the genome according to inferred ancestry. Based on the taxonomic origins and sizes of inferred introgression tracts, we reconstructed the history and geography of species admixture events, which we could order in time since longer tracts indicate more recent introgression. Previous work based on ecological niche modelling of *L. timidus* distribution at the Last Glacial Maximum predicted that the species was present in the northern half of the Iberian Peninsula, which is confirmed by the fossil record. Other studies suggested that at that time, *L. granatensis* would have been in a southwest Iberian refugium while *L. europaeus* would have been restricted to a refugium in the Balkan area. The expansion of these two species would have only occurred with the warming of the climate, a period more favorable to the two. The first contact to occur was between *granatensis* and *timidus*. The time suggested by average tract sizes, 7 kY ago, is probably underestimated and the size distribution of long identity by state fragments rather suggests 24 kY. Mean introgression tract sizes increase from south to north, indicating that the hybridization wave has progressed in that direction. We also see an increasing gradient of proportion of introgression in the same direction. All these observations sustain the model of invasive replacement of *L. timidus* by *L. granatensis*. The contact between *L. europaeus* and *L. timidus* was estimated to have occurred more recently, 4 kY ago according to average introgression tract lengths. Introgression from *L. timidus* was inferred both inside and outside Iberia, thus suggesting that *L. europaeus* individuals entering the Iberian Peninsula might have already been introgressed. Alternatively, *timidus* introgression in Iberia could have resulted from second-hand hybridization with the already introgressed *L. granatensis*. However, the analysis of allospecific junctions in *L. europaeus* individuals inside Iberia shows a quasi-absence of the *granatensis-timidus* junctions which would be expected to result from second-hand introgression, while *europaeus-timidus* junctions are much more frequent, thus supporting the hypothesis that *timidus* introgression in Iberia resulted from previous contact with *timidus*. Finally, we found introgression from *L. granatensis* into *L. europaeus* individuals from within Iberian Peninsula (it represents up to 7.8% of individuals genomes) while it is much rarer in the opposite direction (0.4% in the contact zone between the two species being almost absent elsewhere). Such asymmetric introgression could have resulted from the range replacement of *L. granatensis* by *L. europaeus* as theory predicts that in such situations, introgression should be more prevalent in the direction of the resident species to the invading one. The contact

between these two species was dated to 1 kY ago based on introgression tract lengths, and could have started in southern France, where we find residual *granatensis* introgression in one individual from the French Pyrenees. In the areas where either *granatensis* or *europaeus* are inferred to have replaced *timidus*, we do not find clear geographical gradients of introgression tract sizes, indicating that the invasion was very rapid. We do however find a clear gradient in *granatensis* outside this area, in southern Iberia, an indication of slower secondary diffusion of introgression tracts from the invasion territory further north.

The biogeographic scenario of species contacts proposed here suggests that *L. timidus* was first replaced in northern Iberia by *L. granatensis*, which was then replaced by *L. europaeus*. The exchanges of mitochondrial DNA in the process allowed that the *timidus* type remained where the species was initially present. In contrast to the mtDNA pattern, nuclear *timidus* introgression in the two other species was found to be geographically widespread and at low frequencies. We thus questioned whether purely demographic processes under a single scenario of invasive species replacements could explain this contrast patterns between the mitochondrial and nuclear genomes. To formally test this we conducted spatially explicit simulations of the demographic and historical context of the interactions between species. For this, we leveraged on the wealth of genetic, ecological and paleo-climatological data gathered by previous studies of *L. granatensis*. More specifically we simulated the range expansion of *L. granatensis* from a south-western refugia at the LGM (20 kY ago) into the territory of *L. timidus* in the northern half of Iberia. The patterns of introgression resulting from the simulations were largely congruent with those observed for the nuclear data: introgression was found at low frequencies and widespread across Iberia. Furthermore, when we considered low intra-species migration rates in the simulations, we were able to recover a south-north gradient of introgression, especially south of the expansion range, as for the empirical data. Importantly, the empirical patterns of *timidus* mtDNA introgression can also be reproduced under this single demographic scenario, if considering its lower effective population size resulting from maternal transmission, and assuming female philopatry and sex-asymmetric introgression between the two species. The contrasting patterns between the nuclear and mitochondrial genomes can thus be explained by a demographic history of range replacement with hybridization, with no need to invoke selection to explain massive mtDNA introgression in *L. granatensis*. These conclusions can readily be extended to the case of *L. europaeus* since the patterns of introgression of the mitochondrial and nuclear genomes are similar to those found in *L. granatensis*:

massive *timidus* mtDNA introgression (captured through hybridization with *L. granatensis*) and limited nuclear introgression (from *L. granatensis*). It thus seems likely that the same demographic process of invasion of *L. europaeus* into part of the territory of *L. granatensis*, associated with female philopatry and asymmetrical hybridization could also result in massive introgression of the *timidus* mtDNA.

Both the biogeographic patterns of introgression and simulations of the demographic history of species contacts and hybridization strongly suggest that *timidus* mtDNA introgression is an accidental by-product of demographic processes. However, the nuclear and mitochondrial genomes are known to interact in fundamental functions for organism fitness (e.g. oxidative phosphorylation – OXPHOS) and the mitochondria depend on many nuclear encoded proteins for their correct functioning and life cycle. Since the two genomes co-evolve we hypothesized that introgression of co-evolving nuclear mitochondrial genes ("mitonuc" genes) should have followed massive mtDNA introgression to rescue incompatibilities resulting from "accidental" mtDNA introgression, notwithstanding the possibility that *timidus* cytonuclear association be intrinsically advantageous under certain environmental conditions. In *L. granatensis*, in addition to ELAI, we also used genetic distance (Relative Node Depth) to specifically detect outliers of high frequency introgression from *L. timidus*. Overall, we did not find evidence for preferential introgression of mitonuc genes as compared to other genes in the nuclear genome of *L. granatensis*. Neither did we find overrepresentation of mitonuc genes among the set of genes following the geographical and frequency patterns of mtDNA introgression. Still, some individual mitonuc genes do co-introgress at high frequencies across Iberia and are thus potential candidates for cytonuclear co-adaptation. Likewise, in *L. europaeus* we found only some of the mitonuc genes to either co-introgress or co-differentiate with *L. timidus* mtDNA. However, only one gene (MRPL13) is common to the two species. This suggests that if co-evolution occurs between the nuclear and mitochondrial genomes in these species it is restricted to very few genes or implied different genes in the two cases.

These analyses however revealed a number of highly introgressed genes not related to mitochondria. In *L. granatensis* we found among them enrichment in genes related to male fertility. Theory predicts that male-harmful mutations can accumulate neutrally on mtDNA because of its maternal transmission. This phenomenon, termed mother's curse, is expected to be counteracted by compensatory mutations at interacting nuclear genes. Some of the massive introgressions of nuclear genes from *timidus* into *granatensis* could thus correspond to such situations. In *L. europaeus* we also found

enrichment of genes massively introgressed in the Iberian Peninsula that affect fertility. However, in this case fertility involvement is not clearly restricted to males and thus a possible association with the mother's curse is not granted.

Finally, although the demographic history of the species seems to explain the global patterns of introgression, the heterogeneity of introgression prevalence across the genome suggests some degree of control by factors of other nature. Because the X is essentially female-transmitted, demographic factors favoring mtDNA introgression should also favor its introgression as compared to autosomes. We however found a clear depletion of introgression of the X (when analyzing *timidus* introgression into both *L. granatensis* and *L. europaeus*, but also *L. granatensis* introgression into *L. europaeus*). We also found significant variations of the prevalence of introgression along *L. granatensis* chromosomes, with an increase from chromosome centers towards chromosome ends. We estimated historical recombination rates along the chromosomes from patterns of linkage disequilibrium, and found they are also positively correlated with distance to chromosome center. The positive relationship between introgression and recombination with chromosome position evidences the existence of incompatibilities spread along the genome which are more effective near chromosome centers where linkage to incompatibility factors is more extensive. These incompatibility factors are also more effective on the X, in line with the general observation of a disproportionate role of the X-chromosome in reproductive isolation (large X-effect).

Our rich dataset also allowed addressing other questions about the evolutionary consequences of hybridization. While as we have seen most of the nuclear introgression tracts occur at low frequencies, some regions in the genomes of both *L. granatensis* and *L. europaeus* show high frequencies of *timidus* introgression. Our demographic simulations suggest that such regions are outliers that cannot result from the pure stochastic effects accounting for the average patterns. In both cases we found among highly introgressed fragments enrichment in genes related with innate immunity. This suggests that the new pathogenic environments encountered by the two species during their Iberian expansion might have imposed strong selective constraints, which promoted adaptive introgression of immune genes allowing the two species to adapt to their newly colonized environments. We note however that different genes were concerned in the two affected species. Several other genes with various functions display such patterns of seemingly adaptive introgression. Only further functional studies could confirm the validity of the hypothesis and reveal the traits subject to selection.

In summary, the ubiquity of mtDNA introgression among numerous hare species is remarkable and in this dissertation we describe it in yet another system of hares in North America. Because mtDNA introgression is so frequent among hares, often involving the same donor, and sometimes massive, we have questioned whether it was determined by natural selection or could be explained by demographic associated with range replacements and hybridization between the involved species. Nuclear genomic patterns of introgression support a major role of demography promoting introgression and our simulations show that the highly discordant nuclear and mitochondrial patterns of introgression can be explained under a demographic scenario. Despite repeated, massive mtDNA introgression was possibly a demographic accident promoted by behavioral traits in association with its peculiar transmission modes, and could have been potentially harmful. However, we find that introgression may have been adaptive for other nuclear genes. Notably, we find evidence of adaptive introgression in genes related with immunity both in *L. granatensis* and *L. europaeus* that could have facilitated adaptation of these two species to their newly colonized habitat in northern Iberia. We thus may be witnessing convergent adaptive introgression. At the same time, introgression along the genomes seems to be restrained by interplay between recombination variations and numerous incompatibility factors, the effect being strongest on the X chromosome. Overall, genomic admixture appears globally impeded by incompatibilities, but locally favored by purely demographic effects, and selective effects, either adaptive or in response to genomic conflicts between the mitochondrial and nuclear genomes.

Sumário detalhado

Nos diferentes grupos de organismos, é frequente observado que vários *taxa* relativamente semelhantes partilham uma história de hibridação introgressiva. Os padrões resultantes da introgressão são heterogéneos ao longo do genoma, dado que a troca de material genético depende diretamente dos efeitos destas regiões, ou de regiões geneticamente ligadas a estas, no "fitness" das espécies afetadas. Por exemplo, regiões implicadas em adaptações específicas das espécies aos seus ambientes locais ou envolvidas no isolamento reprodutivo têm uma menor probabilidade de cruzar a barreira das espécies. No entanto, regiões genómicas que não estejam ligadas a estas regiões de incompatibilidade e que sejam elas mesmas neutras no ambiente ou contexto genómico da espécie recipiente poderão introgredir mais livremente. Além disso, a introgressão pode ser promovida em regiões genómicas que aumentem o "fitness" da espécie introgredida. Deste modo, os genomas de vários pares de espécies podem manter-se semipermeáveis ao fluxo interespecífico de genes por algum tempo após a sua divergência inicial. Dada a existência destes padrões heterogéneos de fluxo genético ao longo do genoma o estudo dos mesmos pode ajudar a identificar regiões do genoma responsáveis por impedir esse mesmo fluxo, a descobrir as funções e os mecanismos associados ao processo de especiação, mas também a esclarecer o papel da hibridação como uma fonte de potencial inovador e adaptativo. Apesar destas questões serem desde há muito tempo de grande interesse para os biólogos evolucionistas (temos como exemplo, os primeiros estudos de zonas híbridas), os dados disponíveis até à data estavam longe de permitir uma resolução satisfatória dos padrões genómicos que permitisse responder a essas mesmas questões. No entanto, os dados genómicos atualmente disponíveis permitem-nos abordar estas questões com um poder sem precedentes.

As lebres (género *Lepus*) são um modelo particularmente adequado para estudar a relevância da hibridação e introgressão na evolução. O género diversificou-se rapidamente sendo atualmente composto por mais de 30 espécies descritas que ocupam uma grande variedade de habitats. O género é ainda caracterizado por um grande número de casos de fluxo-génico interespecífico. A maioria dos casos descritos envolvem a introgressão do ADN mitocondrial (ADNmt) da lebre da montanha (*Lepus timidus*), uma espécie boreal atualmente distribuída no Norte da Eurásia e Alpes. A introgressão do ADNmt desta espécie pode ser atualmente observada em quatro

espécies temperadas que habitam a Europa, e possivelmente noutras quatro na China. Estes casos de introgressão foram descritos tanto em áreas onde as espécies atualmente contactam (p. ex. entre *L. timidus* e a lebre-europeia, *L. europaeus*, na Suécia, Rússia e Alpes), mas também a contactos históricos em áreas onde a *L. timidus* não se encontra nos dias de hoje (p.ex. envolvendo espécies de lebre na Península Ibérica). Estes últimos casos resultam de uma distribuição mais ampla da *L. timidus* durante as glaciações do Pleistoceno, que se estenderia até ao sul da Europa, incluindo o sul de França e a Península Ibérica, como indicam os registos paleontológicos. Curiosamente, nalguns destes casos, a introgressão do ADNmt da *L. timidus* é massiva, atingindo frequências bastante elevadas nalgumas populações (como é o caso de populações de lebre-ibérica, *L. granatensis,* e *L. europaeus* no norte da Península Ibérica) ou tendo resultado na completa substituição da mitocôndria nativa (como no caso da lebre-cantábrica, *L. castroviejoi,* e da lebre-italiana, *L. corsicanus*, no norte da Península Ibérica e na Península Itálica, respetivamente).

Este fenómeno de introgressão do ADNmt é um fenómeno frequente na Europa e Ásia e na maioria dos casos implica introgressão da linhagem ártica noutras espécies. No entanto, também existem suspeitas de introgressão entre algumas espécies de lebre Norte-Americanas e por isso antes de mais quisemos clarificar e comparar esta situação nas lebres do Novo Mundo. Das mais de 30 espécies descritas de lebres, nove ocorrem na América do Norte, em habitats distintos e com áreas de distribuição variáveis, mas que por vezes se sobrepõem. Entre aquelas com a distribuição mais alargada está a lebre-Americana (*L. americanus*), que ocupa uma grande extensão do Canadá, a costa do Pacífico e as Montanhas Rochosas dos Estados Unidos da América. Um trabalho anterior baseado em dados de microssatélites sugeriu que a *L. americanus* é constituída por três unidades evolutivas, uma que ocupa a região Boreal (*Boreal*), outra as Montanhas Rochosas (*Rockies*) e ainda outra que habita a zona do Pacífico Norte dos Estados Unidos (*PacNW*). No Sul, *a L. americanus* é substituída pela lebre-da-Califórnia (*L. californicus*), que ocupa a parte Oeste dos Estados Unidos e a metade Norte do México. Curiosamente, as populações do PacNW de *L. americanus* formam um clado mitocondrial mais próximo do clado de *L. californicus* do que de outros clados de *L. americanus*. Este padrão é sugestivo de introgressão mitocondrial mas não se deve colocar de lado a possibilidade de que resulte de coalescência incompleta de linhagens. A terceira espécie com maior distribuição, a lebre-de-cauda-branca (*L. townsendii*), ocupa a parte central da América do Norte, a sua distribuição sobrepondo-se tanto com a de *L. americanus* no Norte e a de *L. californicus* a sul, com a qual foi sugerido que

hibrida na natureza apesar de tal nunca ter sido confirmado por estudos genéticos. Para entender se estas espécies partilham uma história de introgressão durante a sua evolução analisamos dados de múltiplos *loci*, incluindo marcadores genéticos de todos os compartimentos genéticos (mitocôndria, autossomas e cromossomas X e Y) sequenciados em indivíduos representativos de *L. californicus* e *L. townsendii* e dos 3 grupos de *L. americanus*.

Usando métodos coalescentes de inferência filogenética aplicados aos *loci* nucleares, confirmamos que as três unidades evolutivas anteriormente inferidas em *L. americanus* têm significado genealógico e que particularmente o clado *Boreal* divergiu dos outros dois na mesma extensão que outras espécies de *Lepus bona fide*. Usando um modelo de Isolamento-com-Migração (IM), mostramos ainda que o fluxo de genes nucleares entre essas espécies é nulo (entre *L. californicus* e *L. townsendii*) ou extremamente limitado (de *L. americanus* para *L. californicus* e *L. townsendii*). Em contraste, as simulações coalescentes de divergência do ADNmt, baseadas em valores de parâmetros demográficos inferidos a partir dos marcadores de ADN nuclear, mostram que a proximidade genética entre alguns haplótipos de ADNmt de *L. americanus* e de *L. californicus* é menor do que a esperada considerando coalescência incompleta das linhagens e portanto provavelmente resultam da introgressão. Notavelmente, as populações de *L. californicus* e *L. americanus* do *PacNW* não compartilham haplótipos, sugerindo que a introgressão do ADNmt resultou de eventos históricos de hibridação. Uma vez que uma espécie é adaptada à floresta boreal e outra a regiões áridas, estas podem ter sido afetadas de forma diferente pelas oscilações climáticas, o que poderia ter promovido alterações das suas distribuições, resultando em contactos e daí hibridação. Em contraste com as observações dos seus homólogos Europeus, nestas espécies Norte-Americanas, a introgressão de ADNmt ocorreu de uma espécie temperada para uma espécie boreal. No entanto, e de forma semelhante, a introgressão é geograficamente limitada (noroeste do Pacífico) mas massiva, encontrando-se fixa em algumas populações. Curiosamente, enquanto tipicamente os indivíduos de *L. americanus* geralmente mudam para uma pelagem branca no inverno, indivíduos das populações do *PacNW* permanecem castanhos durante todo o ano, aparentemente em resposta à redução da cobertura de neve nesta região. Os indivíduos de *L. californicus* também permanecem castanhos e podem ter transmitido este fenótipo para as *L. americanus* juntamente com ADNmt. Se foi este o caso ou não, tal questão só poderá ser respondida através de estudos genómicos que descrevam os padrões de

introgressão do genoma em geral, possivelmente em associação com o seu contexto funcional.

A introgressão, particularmente envolvendo o ADNmt, é portanto recorrente entre lebres, muitas vezes encontrada em altas frequências nas espécies afetadas e não se encontra restrita a determinadas linhagens ou ambientes. Esta observação levanta a questão sobre se um único mecanismo geral poderá explicar este padrão evolutivo replicado em diferentes espécies de lebre, sendo que duas hipóteses emergem: i) a dinâmica de alternância das distribuições ocupadas por cada uma das espécies envolvidas em hibridação, com uma espécie a substituir a outra em diferentes períodos do tempo, promove a introgressão de ADNmt; ii) as vantagens seletivas da introgressão de ADNmt e complexos de genes coadaptados promovem a introgressão deste *locus*. Assim, dirigimos o nosso trabalho no sentido de entender o impacto genómico destes eventos de hibridação histórica entre espécies de lebres que resultaram na introgressão massiva do ADNmt, e para tal focamo-nos no sistema Ibérico como modelo.

Atualmente, três espécies de lebre habitam a Península Ibérica. A *L. granatensis*, ocupa grande parte da Península, sendo substituída apenas no extremo norte pela *L. castroviejoi*, nas montanhas Cantábricas e pela *L. europaeus* desde as montanhas Cantábricas aos sopés dos Pirenéus. A distribuição da *L. europaeus* estende-se pela Europa Central, Escandinávia, Ásia e Médio Oriente, enquanto que as outras duas espécies são endémicas da Península Ibérica. Populações de todas as três espécies na Península Ibérica apresentam altas frequências de ADNmt de *L. timidus*. Em *L. castroviejoi*, o tipo mitocondrial de *timidus* encontra-se fixado, e na *L. europaeus* encontra-se quase fixado nas populações Ibéricas, embora não se encontre em mais nenhum outro lugar na sua distribuição, exceto onde a espécie atualmente contacta com *L. timidus.* Em *L. granatensis*, o ADNmt de *timidus* apresenta um gradiente Sul-Norte, sendo a introgressão ausente na parte Sul da Península Ibérica mas atingindo altas frequências no Norte.

A introgressão repetida de ADNmt de *timidus* em três espécies residentes na Península Ibérica, a frequência massiva que atinge em algumas populações e o facto de se encontrar restrita às regiões mais frias da Península Ibérica, levanta a hipótese de que a introgressão do ADNmt de *timidus* resulta de adaptação ao frio. De facto, o metabolismo mitocondrial está envolvido na termorregulação e em vários casos foi associado a adaptações relacionadas com a temperatura. Além disso, existem evidências de evolução adaptativa em proteínas mitocondriais ao longo do ramo das

lebres árticas. No entanto, a introgressão massiva pode também resultar de processos associados à substituição geográfica das espécies. Durante o processo de expansão de uma espécie para o território de uma outra, a deriva genética na frente da "onda de expansão" pode levar variantes raros (incluindo introgredidos) a atingir altas frequências e que podem ser propagados pela "onda de expansão" ("*allele surfing*" na "onda da expansão"). Dados de genética populacional de estudos anteriores relativos ao ADNmt e incluindo um número reduzido de marcadores nucleares evidenciaram tanto em *L. granatensis* como em *L. europaeus* a ocorrência de "ondas de expansão" históricas e de introgressão do ADNmt do tipo *timidus* ao longo de um gradiente geográfico (Sul-Norte em *L. granatensis* e Este-Oeste em *L. europaeus*). No entanto, dada a natureza não recombinante de ADNmt, é impossível demonstrar uma origem adaptativa da sua introgressão tendo apenas por base a variação genética deste marcador. Neste trabalho argumentamos que, dada a intensa colaboração dos genomas mitocondriais e nucleares em muitos processos celulares essenciais os dois genomas muito provavelmente co-evoluíram e portanto, de forma adaptativa ou não, a introgressão mitocondrial massiva poderia ter afetado genes codificados no genoma nuclear e funcionalmente ligados às mitocôndrias. Nesse sentido, realizamos um estudo genómico sobre o fluxo genético entre as espécies Ibéricas. A imensidão dos dados gerados também nos permitiu reconstruir a história das interações entre as espécies e testar quantitativamente os cenários de substituição geográfica de espécies e a sua capacidade em explicar padrões de introgressão observados nos genomas nucleares e mitocondriais.

Para abordar estas questões, sequenciamos genomas completos de indivíduos de duas espécies da Península Ibérica, *L. granatensis* e *L. europaeus*. Cinco *L. granatensis* foram amostradas no Sul da Península Ibérica, onde não se observa introgressão de ADNmt do tipo *timidus*, e cinco no Norte ao longo do gradiente crescente sul-norte da introgressão de ADNmt. Sequenciamos ainda cinco *L. europaeus* da Península Ibérica (onde o ADNmt do tipo *timidus* se encontra quase fixado) e cinco de outras regiões da Europa (não afetados pela introdução de ADNmt), desde o sul de França à Ucrânia. Também sequenciamos os genomas de 4 *L. timidus* com origem nos Alpes, da Irlanda e da Escandinávia, e de um indivíduo de *L. americanus* que usamos como "*outgroup*" em algumas análises.

A ancestralidade local ao longo do genoma de cada indivíduo foi inferida utilizando o método ELAI, que usa informação de desequilíbrio de ligação e um "*Hidden Markov Model*" para segmentar o genoma de acordo com a ancestralidade inferida. Com base na origem taxonómica e tamanho dos fragmentos de introgressão inferidos,

reconstruímos a história e geografia dos eventos de hibridação entre espécies, que pudemos ordenar no tempo dado que fragmentos de introgressão mais longos indicam introgressão mais recente. Estudos anteriores baseados na modelação do nicho ecológico da distribuição de *L. timidus* na altura do Último Máximo Glaciar previram que a espécie estaria presente na metade norte da Península Ibérica, o que é suportado pelo registo fóssil. Outros estudos sugeriram que por essa altura a *L. granatensis* estaria confinada a um refúgio no sudoeste da Península Ibérica enquanto que a *L. europaeus* estaria restrita a um refúgio na zona dos Balcãs. A expansão destas duas espécies só terá começado com o aquecimento do clima, num período mais favorável às duas espécies. De acordo com as nossas inferências o primeiro contacto deu-se entre *L. granatensis* e *L. timidus*. O tempo estimado deste contacto com base no tamanho médio dos fragmentos de introgressão, que data de 7000 anos atrás, é provavelmente uma subestimativa sendo que a distribuição do tamanho de fragmentos mais longos de "Indentity-by-State" partilhados por estas duas espécies sugere antes que a hibridação entre as duas terá ocorrido há cerca de 24'000 anos atrás. O tamanho médio dos fragmentos de introgressão aumenta de sul para norte, indicando que a "onda de expansão" e hibridação progrediu nessa direção. Também observamos um gradiente crescente da proporção de introgressão na mesma direção. Todas estas observações corroboram o modelo de substituição invasiva de *L. timidus* por *L. granatensis*. O contacto entre *L. europaeus* e *L. timidus* foi estimado em 4000 anos atrás de acordo com o tamanho médio dos fragmentos de introgressão. A introgressão de *L. timidus* em *L. europaeus* foi inferida tanto dentro como fora da Península Ibérica sugerindo que os indivíduos de *L. europaeus* que entraram na Península Ibérica poderiam já se encontrarem introgredidos. Em alternativa, a introgressão de *L. timidus* em indivíduos de *L. europaeus* na Península Ibérica pode ter resultado de transmissão indireta através de hibridação com *L. granatensis* já introgredidas. No entanto, a análise de junções heteroespecíficas, isto é de regiões do genoma com diferentes ancestralidades, em indivíduos de *L. europaeus* dentro da Península Ibérica mostra uma quase ausência de junções *granatensis-timidus* que seriam esperadas no caso de introgressão indireta, enquanto que junções *europaeus-timidus* são consideravelmente mais frequentes, suportando a hipótese de que a introgressão de *timidus* na Península Ibérica resultou de contactos prévios entre *L. europaeus* e *L. timidus*. Finalmente, descobrimos introgressão de *L. granatensis* em indivíduos de *L. europaeus* da Península Ibérica (representando até 7.8% dos genomas dos indivíduos de *L. europaeus*) enquanto que é muito mais rara na direção oposta (0.4% na zona de contacto entre as duas espécies,

sendo praticamente ausente em todo o resto da distribuição). Esta introgressão assimétrica pode ter resultado da substituição de *L. granatensis* por *L. europaeus* pois nestas situações, como previsto por estudos teóricos, a introgressão tende a ser predominante na direção da espécie residente para a invasora. O contacto entre estas duas espécies foi datado em 1000 anos atrás baseado no tamanho médio dos fragmentos de introgressão e pode ter começado no sul de França onde encontramos introgressão residual de *L. granatensis* num indivíduo amostrado na parte francesa dos Pirenéus. Nas áreas onde se inferiu que *L. granatensis* ou *L. europaeus* substituíram *L. timidus* não encontramos gradientes geográficos claros dos tamanhos de fragmentos de introgressão indicando que a invasão foi bastante rápida. Ainda assim encontramos um gradiente claro em *L. granatensis* fora desta área, no Sul da Península Ibérica, o que indica difusão secundária e lenta de fragmentos de introgressão da região invadida mais a Norte.

O cenário biogeográfico de contacto entre espécies aqui proposto, sugere que a *L. timidus* foi primeiro substituída por *L. granatensis* no Norte da Península Ibérica, tendo sido depois substituída por *L. europaeus*. As trocas de ADNmt durante este processo permitiram que o tipo *timidus* permanecesse onde a espécie se encontrava inicialmente presente. Em contraste com o padrão do ADNmt, a introgressão nuclear de *timidus* nas duas outras espécies encontra-se espalhada geograficamente e em baixas frequências. Neste sentido questionamo-nos se processos puramente demográficos associados a um cenário de substituição geográfica por espécies invasores poderia explicar estes padrões contrastantes entre genomas nucleares e mitocondriais. De forma a testar formalmente esta hipótese, realizamos simulações espacialmente explícitas da demografia e do contexto histórico das interações entre espécies. Para tal, tiramos vantagem da grande quantidade de dados genéticos, ecológicos e paleoclimatológicos recolhidos em estudos prévios focados em *L. granatensis*. Especificamente, simulamos a expansão de *L. granatensis* a partir de um refúgio no Sudoeste da Península Ibérica há 20'000 anos atrás (Último Máximo Glaciar) para o território de *L. timidus* na metade Norte da Península Ibérica. Os padrões de introgressão resultantes das simulações foram congruentes com os observados para os dados nucleares: encontramos introgressão em baixa frequências e dispersa pela Península Ibérica. Para além disso, quando consideramos baixas taxas de migração intraespecífica nas simulações, conseguimos recuperar um gradiente de introgressão Sul-Norte, especialmente a sul da área de expansão, tal como para os dados empíricos. De salientar, os padrões empíricos de introgressão mitocondrial de *timidus* podem também ser reproduzidos sob

este mesmo cenário demográfico se considerarmos um efetivo populacional mais baixo para o ADNmt (resultante da transmissão materna) e assumindo filopatria das fêmeas e introgressão sexualmente assimétrica entre as duas espécies. Os padrões contrastantes entre os genomas mitocondrial e nuclear podem assim ser explicados por uma história demográfica de substituição com hibridação, sem ser necessário invocar seleção para explicar a introgressão massiva do ADNmt em *L. granatensis*. Estas conclusões podem ser facilmente estendidas para o caso da *L. europaeus* dado que os padrões de introgressão dos genomas mitocondrial e nuclear são semelhantes aos encontrados em *L. granatensis*: introgressão massiva de ADNmt de *timidus* (obtido através de hibridação com *L. granatensis*) e introgressão nuclear limitada (de *L. granatensis*). Parece portanto provável que o mesmo processo demográfico de invasão de *L. europaeus* para parte do território de *L. granatensis*, associado com filopatria das fêmeas e hibridação assimétrica, possa também ter resultado em introgressão massiva de ADNmt de *timidus*.

Tanto os padrões biogeográficos de introgressão como as simulações da história demográfica dos contactos e hibridação entre espécies sugerem fortemente que a introgressão do ADNmt do tipo *timidus* é um subproduto acidental de processos demográficos. No entanto, os genomas nucleares e mitocondriais são conhecidos por interagir em funções fundamentais para o "fitness" dos organismos (por exemplo, fosforilação oxidativa - OXPHOS) e as mitocôndrias dependem de várias proteínas codificadas no genoma nuclear para o seu ciclo de vida e correto funcionamento. Dada a coevolução destes dois genomas colocamos a hipótese de que genes nucleares ("*mitonuc*") que co evoluíssem com genes mitocondriais deveriam ter seguido a introgressão massiva do ADNmt para evitar incompatibilidades resultantes de introgressão acidental do ADNmt, não obstante a possibilidade de associações cito-nucleares de genes de *timidus* poderem também ser intrinsecamente vantajosas sob certas condições ambientais. Em *L. granatensis*, além do método ELAI, usamos também a distância genética ("*Relative Node Depth*") para detetar especificamente regiões do genoma cujas frequências de introgressão fossem "*outliers*" (alta frequência de introgressão). No geral, não encontramos evidências de introgressão preferencial de genes "*mitonuc*" em comparação com outros genes no genoma nuclear de *L. granatensis*, nem entre o conjunto de genes que seguem os padrões geográficos e de frequência da introgressão de ADNmt. Ainda assim, alguns genes "*mitonuc*" individuais apresentam introgressão em altas frequências em toda a Península Ibérica e, portanto, são potenciais candidatos a coadaptarão cito-nuclear. No mesmo sentido, encontramos

em *L. europaeus* apenas alguns genes "*mitonuc*" potencialmente a co-introgredir ou co-diferenciados com ADNmt de *L. timidus*. No entanto, apenas um gene (MRPL13) é comum às duas espécies. Tal sugere que caso exista coevolução entre os genomas nucleares e mitocondriais nestas espécies, esta é restrita a poucos genes ou implicou diferentes genes nos dois casos.

Estas análises revelaram ainda uma série de genes altamente introgredidos não relacionados com a mitocôndria. Em *L. granatensis* encontramos entre estes genes um enriquecimento de genes relacionados com a fertilidade dos machos. Em teoria, mutações nocivas ao sexo masculino podem acumular-se de forma neutral no ADNmt devido à transmissão materna. Este fenómeno, chamado de "maldição materna", pode ser neutralizado por mutações compensatórias em genes nucleares. Algumas das introgressões massivas de *timidus* em genes nucleares de *L. granatensis* podem assim corresponder a tais situações. Em *L. europaeus* também encontramos enriquecimento de genes massivamente introgredidos na Península Ibérica que afetam a fertilidade. No entanto, neste caso, o envolvimento na fertilidade não está claramente restrito aos machos e portanto uma possível associação com a "maldição materna" não será tão óbvia.

Finalmente, embora a história demográfica das espécies pareça explicar os padrões globais de introgressão, a prevalência de heterogeneidade de introgressão em todo o genoma sugere algum grau de controlo por via de outros fatores. Dado que o cromossoma X é essencialmente transmitido pelas fêmeas, os fatores demográficos que favorecem a introgressão de ADNmt deveriam também favorecer a sua introgressão em comparação com os autossomas. Contudo, observamos uma redução clara da introgressão do X (tanto em termos de introgressão de *L. timidus* em *L. granatensis* e *L. europaeus*, como de *L. granatensis* em *L. europaeus*). Também encontramos variações significativas na prevalência de introgressão ao longo dos cromossomas de *L. granatensis*, com um aumento de introgressão desde o centro para as extremidades dos cromossomas. Observamos ainda que a taxa de recombinação histórica ao longo dos cromossomas, estimada a partir dos padrões de desequilíbrio de ligação, se encontra positivamente correlacionados com a distância ao centro do cromossoma. A relação positiva entre introgressão e recombinação com a posição no genoma evidencia a existência de fatores de incompatibilidade espalhados ao longo do genoma que são mais eficazes perto dos centros dos cromossomas onde o desequilíbrio de ligação a fatores de incompatibilidade é mais extenso. Estes fatores de incompatibilidade são

também mais eficazes no cromossoma X, de acordo com a observação geral de um papel desproporcional do cromossoma X no isolamento reprodutivo ("*large X effect*").

Os dados gerados neste estudo permitiram também abordar outras questões sobre as consequências evolutivas da hibridação. Embora tenhamos observado que a maioria da introgressão nuclear ocorre a baixas frequências, algumas regiões dos genomas tanto de *L. granatensis* como de *L. europaeus* mostram altas frequências de introgressão de *timidus*. As nossas simulações demográficas sugerem que tais regiões são "*outliers*" que não podem resultar de efeitos puramente estocásticos. Em ambos os casos, encontramos um enriquecimento de fragmentos altamente introgredidos em genes relacionados com imunidade. Tal sugere que os novos ambientes patogénicos encontrados pelas duas espécies durante a sua expansão para a região norte da Península Ibérica podem ter imposto fortes pressões seletivas que terão promovido a introgressão adaptativa de genes relacionados com a imunidade o que lhes permitiu a adaptação a estes novos ambientes. É, no entanto, importante notar que os genes em questão são diferentes nas duas espécies. Outros genes com funções variadas exibem também padrões de introgressão aparentemente adaptativa. No entanto, apenas através da realização de estudos funcionais se poderá confirmar a hipótese de que a introgressão destes genes teve uma natureza adaptativa e quais os fenótipos sujeitos a seleção.

Em resumo, a ubiquidade da introgressão de ADNmt entre numerosas espécies de lebre é notável e nesta dissertação descrevemos esse mesmo fenómeno num outro sistema de lebres da América do Norte. Como a introgressão de ADNmt é tão frequente entre lebres, muitas vezes envolvendo o mesmo dador, e é por vezes massiva questionamos se esta teria sido determinada por seleção natural ou poderia ser explicada por fatores demográficos associados a situações de substituição da área de distribuição das espécies e hibridação entre elas. Os padrões genómicos de introgressão nuclear revelam um papel importante da demografia no sentido de promover a introgressão. As nossas simulações mostram que os padrões de introgressão nuclear e mitocondrial, ainda que altamente discordantes, podem ser explicados por um único cenário demográfico. Apesar de repetida, a introgressão de ADNmt massiva resultou possivelmente de um acidente demográfico promovido por características comportamentais em associação com o modo de transmissão peculiar deste marcador e pode ter sido potencialmente prejudicial. No entanto, observamos que a introgressão de outros genes nucleares pode ter sido adaptativa. Notavelmente, encontramos evidências de introgressão adaptativa em genes relacionados com a

imunidade tanto em *L. granatensis* como em *L. europaeus* que poderiam ter facilitado a adaptação dessas duas espécies ao seu novo habitat no norte da Península Ibérica. Assim, podemos estar a testemunhar um caso de introgressão adaptativa convergente. Ao mesmo tempo, a introgressão ao longo dos genomas parece ter sido limitada pela interação entre variações dos níveis de recombinação ao longo dos cromossomas e numerosos fatores de incompatibilidade, sendo o efeito mais forte no cromossoma X. Em geral, a introgressão ao nível do genoma parece ser globalmente impedida por incompatibilidades, mas favorecida localmente por efeitos puramente demográficos e efeitos seletivos, tanto adaptativos como em resposta a conflitos genómicos entre os genomas mitocondriais e nucleares.

Résumé détaillé

Les taxons proches de différents groupes d'organismes montrent souvent des histoires d'hybridation introgressive. Les patrons d'introgression sont hétérogènes le long du génome puisque l'échange de matériel génétique dépend des effets des régions échangées ou de celles qui leur sont liées sur la valeur sélective. Par exemple, les régions impliquées dans des adaptations spécifiques à l'environnement local de l'une ou l'autre espèce, ou impliquées dans l'isolement reproductif, sont moins enclines à traverser les barrières spécifiques. Toutefois des régions génomiques libérées de la liaison avec ces incompatibilités et sans influence négative dans le fond génétique étranger pourront introgresser librement, et même être favorisées si elles confèrent un avantage à l'espèce réceptrice. En conséquence les génomes de nombreuses paires d'espèces proches restent semi-perméables aux échanges durant un certain temps après la divergence initiale. L'étude des patrons d'échange le long du génome peut aider à identifier les locus barrière, et la connaissance de leur fonction d'en déduire la nature des moteurs de la spéciation, mais peut aussi révéler le rôle de l'hybridation comme source de potentiel adaptatif. Alors que ces questions ont longtemps suscité l'intérêt en biologie de l'évolution (par exemple depuis les anciennes études de zones hybrides), les données disponibles ont été loin d'en permettre une résolution satisfaisante. Toutefois l'accessibilité récente de jeux de données génomiques permet maintenant d'aborder ces questions avec une puissance sans précédent.

Les lièvres (genre *Lepus*) représentent un modèle de choix pour l'étude de l'importance de l'hybridation en évolution. Le genre s'est diversifié en une rapide radiation produisant plus de 30 espèces décrites occupant une grande diversité d'habitats, et il présente de nombreux cas décrits de flux génétiques interspécifiques. La plupart des cas décrits impliquent l'introgression de l'ADN mitochondrial (ADNmt) du lièvre variable (*L. timidus*), une espèce boréale actuellement largement distribuée en Eurasie du nord et dans les Alpes. L'introgression s'est produite vers quatre espèces d'habitats tempérés européens, et possiblement vers au moins quatre autres espèces de Chine. Des cas d'introgression ont été décrits dans des situations de contact actuel entre les espèces (entre *L. timidus* et le lièvre européen, *L. europaeus*, en Suède, Russie ou dans les Alpes), mais aussi dans des situations de contacts passés (entre *L. timidus* et les espèces actuellement présentes en Ibérie). Ces derniers cas reflètent la présence, documentée par la paléontologie, de *L. timidus* durant le Pléistocène au sud de l'Europe, dont le sud de la France et la péninsule ibérique. Dans ces régions, l'introgression

mitochondriale d'origine *timidus* peut être massive et atteindre des fréquences très élevées dans certaines régions, par exemple au nord de la péninsule en ce qui concerne *L. granatensis* et *L. europaeus*, voire avoir conduit au complet remplacement du génome mitochondrial d'origine, comme dans le cas de *L. castroviejoi* et *L. corsicanus*.

Le phénomène d'introgression mitochondriale apparaissait fréquent en Europe et en Asie, et impliquer uniquement l'introgression depuis une lignée arctique vers plusieurs autres espèces principalement tempérées. Comme il existait aussi des suspicions d'hybridation entre taxons d'Amérique du Nord, nous nous sommes tout d'abord intéressés à la comparaison avec la situation dans le Nouveau Monde. Parmi plus de 30 espèces, neuf sont trouvées en Amérique du Nord, occupant des habitats variés sur des aires de distributions de tailles disparates et parfois chevauchantes. Un de ceux à l'aire la plus vaste est le lièvre d'Amérique (*L. americanus*) qui occupe la plupart du Canada et la frange côtière pacifique des Montagnes Rocheuses aux Etats-Unis. Au sud il est remplacé par le lièvre de Californie (*L. californicus*), qui occupe la partie ouest des Etats-Unis et la moitié nord du Mexique. Le lièvre de Townsend (*L. townsendii*), dans la partie centrale de l'Amérique du Nord, coexiste avec *L. americanus* et *L. californicus*. On a suggéré que *L. californicus* et *L. townsendii* s'hybrident dans la nature, sans que ce soit appuyé par des données génétiques. Une étude précédente basée sur des marqueurs génétiques microsatellites a suggéré la partition de *L. americanus* en trois unités évolutives occupant respectivement la région boréale (« Boreal »), celle des Montagnes Rocheuses (« Rocky ») et finalement la région du pacifique nord-ouest des Etats-Unis (« PacNW »). Curieusement, les populations PacNW constituent un clade mitochondrial plus apparenté à celui typique de *L. californicus* qu'à ceux des conspécifiques d'autres régions. Ceci pourrait suggérer l'introgression, sans qu'il ait toutefois été possible d'exclure le tri incomplet de lignées (« ILS »). Afin de résoudre cette question nous avons analysé un jeu de données multilocus incluant des marqueurs de tout type de mode de transmission (mitochondriaux, autosomaux, liés au X et au Y), séquencés chez des individus représentatifs des trois entités géographiques de *L. americanus*, ainsi que de *L. californicus* et *L. townsendii*.

L'application de méthodes d'inférence phylogénétique basées sur le coalescent aux données nucléaires nous a permis de confirmer la validité généalogique des trois groupes de *L. americanus*, et de révéler la grande divergence du groupe « Boreal » par rapport aux autres, du même ordre de grandeur qu'entre certaines espèces du genre. A l'aide d'ajustements de modèles d'Isolement avec Migration (IM) nous montrons que les échanges génétiques nucléaires entre ces espèces sont soit non détectables (entre *L.*

californicus et *L. townsendii*) soit très limités (de *L. americanus* vers *L. californicus* et *L. townsendii*). Par contre nous montrons, par des simulations de coalescent du génome mitochondrial respectant les modèles démographiques ajustés aux données nucléaires, que les haplotypes mitochondriaux de la population PacNW proviennent d'une introgression depuis *L. californicus*, et non de la persistence de lignées anciennes. Toutefois, *L. californicus* et *L. americanus* PacNW ne partagent aucun haplotype, ce qui indiquerait une hybridation ancienne. L'une de ces espèces étant adaptée à la forêt boréale et l'autre aux régions arides, elles ont dû être différemment affectées par les oscillations climatiques passées, ce qui aurait pu entraîner des remplacements d'aires ayant facilité l'hybridation. Contrairement à ce qui est observé pour les espèces européennes, chez ces espèces américaines l'introgression est observée depuis une espèce tempérée vers une espèce boréale. Toutefois l'introgression est similairement géographiquement limitée (Nord-Ouest pacifique) et massive, atteignant fixation dans certaines populations. Il est intéressant de remarquer que, alors que *L. americanus* mue vers un pelage d'hiver blanc en général, les individus de PacNW restent bruns toute l'année, en réponse apparente à la faible couverture neigeuse de la région. *L. californicus* reste aussi brun toute l'année et pourrait avoir transmis cette faculté à *L. americanus* en même temps que le génome mitochondrial. La réponse à cette question devra attendre des études génomiques décrivant les patrons d'introgression en relation avec de possibles conséquences fonctionnelles.

L'introgression, particulièrement de l'ADNmt, est donc récurrente parmi les lièvres, souvent trouvée à forte fréquence et ne semble pas l'apanage de certaines lignées ou environnements. Ceci soulève la question d'un potentiel mécanisme commun pouvant expliquer une telle répétition du même patron évolutif. Deux hypothèses émergent : i) que la dynamique du remplacement d'aire de distribution promeuve l'introgression ; ii) que l'introgression massive soit promue par un avantage sélectif de l'ADNmt et de gènes nucléaires co-adaptés. Nous avons donc orienté notre travail vers la compréhension de l'impact génomique de ces événements anciens ayant résulté en l'introgression massive d'ADNmt, en prenant le système ibérique comme modèle.

Trois espèces de lièvres sévissent actuellement dans la péninsule ibérique. Le lièvre ibérique*, L. granatensis*, occupe la plupart de la péninsule, n'étant remplacé que pas le lièvre de Castroviejo, *L. castroviejoi*, dans la cordillère cantabrique, et par le lièvre européen, *L. europaeus*, depuis cette cordillère jusqu'aux piémonts pyrénéens. L'aire de *L. europaeus* s'étend toutefois à travers l'Europe centrale jusqu'à la Scandinavie, l'Asie et le Moyen Orient, alors que les deux autres sont endémiques de la péninsule ibérique.

Des populations de ces trois espèces présentent en Ibérie des hautes fréquences d'ADNmt *timidus*. Cela va jusqu'à atteindre la fixation chez *L. castroviejoi*, et la quasi-fixation dans le territoire ibérique de *L. europaeus*, sa présence ailleurs étant limitée à quelques zones de contact actuel avec *L. timidus*. Chez *L. granatensis*, l'ADNmt *timidus* présente un gradient sud-nord, depuis absent au sud jusqu'à de très hautes fréquences au nord.

L'introgression répétée de l'ADNmt *timidus* vers trois espèces de la péninsule ibérique, sa très haute fréquence dans certaines populations, toutes au nord de la péninsule, génère l'hypothèse d'une valeur adaptative de cette introgression, par exemple de résistance au froid. En effet le métabolisme mitochondrial est impliqué dans la thermorégulation, les variations de séquence de l'ADNmt ont pu dans certains cas être mises en relation avec une adaptation à la température, et il a été rapporté des indices d'une divergence fonctionnelle des protéines du génome mitochondrial de la lignée arctique de lièvres. Toutefois, l'introgression massive d'ADNmt pourrait n'être qu'une conséquence fortuite de conditions ayant accompagné le remplacement d'espèces suite aux changements climatiques post-glaciaires drastiques. Durant l'expansion d'une espèce dans le territoire occupé par une autre, la dérive peut amener certains variants rares (y compris issus d'introgression) jusqu'à de très hautes fréquences locales, qui peuvent ensuite être propagées avec la vague d'expansion (« surf » sur la vague d'expansion). De précédentes données de génétique des populations sur l'ADNmt et une poignée de marqueurs nucléaires avaient donné des indices de vagues passées d'expansions géographiques de *granatensis* et *europaeus* ayant accompagné l'introgression mitochondriale de *timidus* le long de gradients géographiques (sud-nord chez *granatensis*, est-ouest chez *europaeus*). Toutefois, étant donné la nature non-recombinante de l'ADNmt, il est impossible de démontrer l'origine adaptative de son invasion sur la seule base de ses variations. Nous remarquons qu'étant donné l'intense collaboration entre les génomes mitochondrial et nucléaire dans plusieurs processus cellulaires clés, les deux génomes co-évoluent et, qu'elle soit adaptative ou pas, l'introgression mitochondriale pourrait avoir affecté des gènes nucléaires fonctionnellement liés à la mitochondrie.

Nous avons ainsi entrepris une étude des échanges génétiques entre espèces de la péninsule ibérique à l'échelle du génome entier. Cette abondance de données nous a aussi permis de reconstituer l'histoire des interactions entre espèces, et de tester quantitativement les scénarios de remplacement d'espèces et leur capacité à expliquer les patrons d'introgression observés pour les génomes mitochondriaux et nucléaires.

Pour aborder cette question nous avons séquencé les génomes complets de spécimens de deux espèces de la péninsule ibérique. Cinq *L. granatensis* provenaient du sud, là où aucune introgression mitochondriale ne prévaut, et cinq autres du nord, le long d'un gradient sud-nord de fréquence croissante d'introgression mitochondriale de *timidus*. Nous avons aussi séquencé cinq *L. europaeus* de la péninsule (où l'ADNmt *timidus* est quasiment fixé) et cinq de diverses autres provenances européennes, du sud de la France à l'Ukraine (régions non affectées par l'introgression mitochondriale). Nous avons également séquencé les génomes de quatre *L. timidus* des Alpes, Irlande et Scandinavie et un *L. americanus* pour servir de groupe externe.

Grâce à la méthode ELAI, qui utilise l'information de déséquilibre de liaison et un modèle de chaîne de Markov cachée, nous avons pu segmenter le génome de chaque individu en fonction des variations d'ancestralité. Sur la base de l'origine taxonomique et la taille des segments d'introgression inférés, nous avons reconstruit l'histoire et la géographie des évènements de mélanges entre les trois espèces, dont nous avons pu déterminer l'ordre chronologique puisque des segments plus longs correspondent à des introgressions plus récentes. Des travaux antérieurs de modélisation de niche écologique avaient prédit la présence de *L. timidus* dans la moitié nord de la péninsule après le dernier maximum glaciaire, ce qui est corroboré par les données fossiles. Il était aussi suggéré qu'en ce temps, *L. granatensis* était confiné dans un refuge au sud-ouest de la péninsule ibérique, tandis que *L. europaeus* était cantonné dans un refuge balkanique. L'expansion de ces deux espèces ne serait intervenue qu'avec le réchauffement climatique post-glaciaire, favorable à toutes deux. Nous démontrons ici que le premier contact s'est produit entre *L. granatensis* et *L. timidus*. L'âge suggéré par la longueur des segments d'introgression, il y a 7 kY, est probablement sous-estimé et la distribution de taille des segments d'identité d'état suggère plutôt 24 kY. La taille des segments d'introgression augmente du sud au nord, indiquant une progression de la vague d'hybridation le long de cette direction. Nous observons également une augmentation de la prévalence de l'introgression le long du même axe et dans la même direction. Toutes ces observations soutiennent le modèle de remplacement invasif de *L. timidus* par *L. granatensis*. Nous estimons que le contact entre *L. europaeus* et *L. timidus* s'est produit plus récemment, il y a 4 kY d'après la taille moyenne des segments d'introgression. L'introgression nucléaire de *L. timidus* prévaut aussi bien en Ibérie qu'en dehors, ce qui suggère que l'introgression avait eu lieu avant l'entrée dans la péninsule. Alternativement, l'introgression *timidus* en Ibérie pourrait s'être faite en seconde main depuis *L. granatensis*, lui-même alors déjà affecté. Toutefois, nous ne détectons chez

europaeus quasiment pas de jonctions *granatensis-timidus* qui seraient caractéristiques d'une telle seconde main, mais quasi-exclusivement des jonctions *europaeus-timidus,* attendues suite à un contact primaire avec *L. timidus* en dehors de la péninsule. Finalement, nous trouvons des nombreux segments d'introgression de *granatensis* vers *europaeus* en Ibérie (représentant jusqu'à 7.8% du génome individuel), mais très peu dans l'autre direction (jusqu'à 0.4%, mais seulement très proche de la zone de contact). Une telle introgression asymétrique pourrait résulter du déplacement d'aire de *L. granatensis* par *L. europaeus*, puisque la théorie prédit dans de telles situations une introgression préférentielle depuis le résident vers l'envahisseur. Sur la base de la taille moyenne des segments d'introgression, nous estimons que le contact entre ces deux espèces s'est produit il y a environ 1 kY, et pourrait avoir été initié dans le sud de la France, où nous trouvons des traces résiduelles d'introgression dans les Pyrénées françaises. Dans les territoires où soit *granatensis*, soit *europaeus* sont supposés avoir remplacé *timidus*, nous ne trouvons pas de gradient géographique de taille des segments d'introgression, une indication que les invasions furent très rapides. Nous trouvons par contre un tel gradient prononcé chez *granatensis* en dehors de cette zone, dans le sud de la péninsule, une indication d'une diffusion plus lente des segments d'introgression depuis le territoire d'invasion plus au nord.

Le scénario biogéographique de contacts interspécifiques proposé ici suggère que *L. timidus* fut d'abord remplacé dans la moitié nord de la péninsule par *L. granatensis*, qui fut ensuite lui-même remplacé par *L. europaeus* dans l'extrême nord. Les échanges mitochondriaux accompagnant ces évènements ont permis au génome mitochondrial *timidus* de rester en place là où était son espèce d'origine. En fort contraste avec le patron mitochondrial, l'introgression nucléaire vers les deux autres espèces est trouvée géographiquement répandue et à faible fréquence moyenne. Nous avons demandé si des patrons tellement contrastés pouvaient résulter des conséquences communes de la seule histoire démographique des contacts entre espèces. Pour tester ceci formellement, nous avons conduit des simulations spatialement explicites du contexte historique et démographique de l'interaction entre les espèces. Nous nous sommes appuyés sur les nombreuses données génétiques, écologiques et paléoclimatiques précédemment recueillies sur *L. granatensis.* Plus spécifiquement, nous avons simulé l'expansion de l'aire de *L. granatensis* depuis un refuge sud-ouest après le dernier maximum glaciaire (20 kY) dans le territoire de *L. timidus*, la moitié nord de la péninsule. Les patrons d'introgression obtenus dans les simulations étaient largement congruents avec ceux observés pour les données nucléaires : une

introgression à basse fréquence et large distribution dans la péninsule. De plus, en considérant des taux de migration intra spécifiques bas, nous reproduisons le gradient de fréquence d'introgression observé en dehors de la zone d'expansion. Finalement le patron empirique d'introgression mitochondriale peut également être reproduit sous ce même scénario démographique, en tenant compte de sa plus faible taille efficace liée à sa transmission femelle, et en supposant la philopatrie des femelles et une asymétrie des croisements entre espèces. Les patrons contrastés d'introgression entre les génomes nucléaire et mitochondrial peuvent ainsi être réconciliés sous un modèle démographique de remplacement d'aire, sans besoin d'invoquer la sélection pour expliquer l'introgression mitochondriale massive. Ces conclusions ont une valeur suffisamment générale pour pouvoir être étendues au cas de *L. europaeus*, étant donné la similitude des patrons d'introgression mitochondriale et nucléaire : introgression massive d'ADNmt (capturé par hybridation avec *L. granatensis*), et introgression nucléaire limitée (depuis *L. granatensis*). Il semble donc vraisemblable qu'un phénomène similaire, d'invasion de *L. europaeus* dans une partie du territoire occupé par *L. granatensis*, associée à une philopatrie des femelles et une hybridation asymétrique ait pu résulter en une introgression massive d'ADNmt *timidus*.

Les patrons biogéographiques d'introgression et les simulations d'histoire démographique de contacts et hybridations entre espèces suggèrent fortement que l'introgression de l'ADNmt *timidus* est un sous-produit accidentel d'un processus démographique. Toutefois, on sait que les génomes nucléaire et mitochondrial interagissent dans des fonctions fondamentales pour la valeur sélective des organismes (par ex. la phosphorylation oxydative, OXPHOS), et les mitochondries dépendent de nombreux gènes nucléaires pour leur fonctionnement correct et leur cycle de vie. Puisque les deux génome co-évoluent, nous avons émis l'hypothèse que l'introgression de gènes nucléaires mitochondriaux (gènes « mitonuc ») co-évoluant aurait pu accompagner l'introgression massive d'ADNmt pour compenser des incompatibilités résultant de l'introgression « accidentelle » de l'ADNmt, sans toutefois exclure la possibilité que certaines combinaisons d'origine *timidus* soient absolument avantageuses dans certaines conditions environnementales. Chez *L. granatensis*, en plus de la méthode ELAI, nous avons utilisé les distances génétiques (« Relative Node Depth, RND ») pour spécifiquement détecter les introgressions à haute fréquence depuis *L. timidus*. Dans l'ensemble, nous ne trouvons pas d'indices d'une introgression préférentielle des gènes mitonuc en comparaison des autres gènes chez *L. granatensis*. Nous ne trouvons pas non plus de sur-représentation des mitonuc parmi ceux aux

patrons d'introgression similaires à ceux de l'ADNmt en fréquence et géographie. Toutefois certains mitonuc individuels co-introgressent à haute fréquence dans la péninsule et représentent donc de potentiels candidats pour la co-adaptation cyto-nucléaire. De même, chez *L. europaeus* nous trouvons certains gènes mitonuc co-introgressés ou co-différenciés avec l'ADNmt *timidus*. Toutefois un seul gène (MRPL13) ressort commun pour les deux espèces. Ceci suggère que si une co-évolution se produit entre les génomes nucléaire et mitochondrial de ces espèces, elle est restreinte à un petit nombre de gènes ou implique des gènes différents dans les deux cas.

Ces analyses ont toutefois révélé un nombre de gènes hautement introgressés mais sans relation avec les fonctions mitochondriales. Chez *L. granatensis* nous trouvons parmi eux un enrichissement en gènes impliqués dans la fertilité mâle. La théorie prédit que des mutations délétères pour les mâles peuvent s'accumuler sur l'ADNmt en raison de sa transmission maternelle. Ce phénomène, baptisé le « mauvais sort des mères », devrait être contrecarré par des mutations compensatoires sur des gènes nucléaires en interaction. Certaines des introgressions massives de gènes nucléaires de *timidus* vers *granatensis* pourraient ainsi correspondre à de telles situations. Chez *L. europaeus* nous trouvons également des gènes massivement introgressés en Ibérie et affectant la fertilité. Toutefois dans ce cas l'implication spécifique dans les fonctions mâles n'est pas avérée.

Finalement, bien que l'histoire démographique des espèces semble pouvoir expliquer les patrons moyens d'introgression, l'hétérogénéité de prévalence de l'introgression le long du génome suggère un degré de contrôle par des facteurs d'une autre nature. Parce que le X est essentiellement transmis par les femelles, les facteurs démographiques favorisant l'introgression mitochondriale devraient aussi favoriser son introgression par rapport aux autosomes. Nous trouvons cependant une déplétion claire d'introgression du X (en analysant l'introgression de *timidus* vers *granatensis* ou *europaeus*, mais aussi de *granatensis* vers *europaeus*). Nous trouvons aussi des variations significatives de la prévalence de l'introgression le long des chromosomes de *L. granatensis*, avec une augmentation depuis le centre vers les extrémités des chromosomes. Nous avons estimé les taux de recombinaison historiques le long des chromosomes à partir des patrons de déséquilibre de liaison, et trouvé qu'ils sont aussi positivement corrélés à la distance au centre des chromosomes. Cette corrélation positive entre recombinaison et introgression atteste de l'existence de nombreuses incompatibilités le long du génome. Nous avons montré que ces incompatibilités

s'expriment plus sur le X, en accord avec une observation générale d'un effet disproportionné du X dans l'isolement reproductif.

Notre riche jeu de données a aussi permis d'aborder d'autres questions en relation avec les conséquences évolutives de l'hybridation. Alors que comme nous l'avons vu la plupart des segments d'introgression se trouvent en faible fréquence, certaines régions des génomes de *L. granatensis* et *L. europaeus* montrent de hautes fréquences d'introgression nucléaire. Nos simulations démographiques suggèrent que ces régions sont des observations aberrantes ne pouvant résulter des purs effets stochastiques rendant compte de la majorité des données. Dans les deux cas nous avons trouvé parmi les gènes hautement introgressés un enrichissement en gènes impliqués dans l'immunité. Ceci suggère que les nouveaux environnements rencontrés par les deux espèces lors de leurs expansions ibériques ont dû imposer des contraintes sélectives qui ont favorisé l'introgression adaptative de gènes immunitaires. Nous avons toutefois constaté que des gènes différents étaient concernés dans les deux espèces. Plusieurs autres gènes aux fonctions variées présentent aussi de tels patrons évocateurs d'introgression adaptative. Seule une caractérisation fonctionnelle pourrait confirmer la validité de l'hypothèse adaptative et révéler les traits ayant donné prise à la sélection.

En résumé, l'ubiquité de l'introgression de l'ADNmt dans plusieurs espèces de lièvre est remarquable et dans cette thèse nous décrivons le phénomène dans un système de plus en Amérique du Nord. Parce que l'introgression de l'ADNmt est si fréquente entre lièvres, impliquant souvent le même donneur, et est souvent massive, nous avons demandé si elle pouvait résulter de causes communes, soit la sélection naturelle, soit la démographie associée à des remplacements d'espèces. Les patrons génomique d'introgression nucléaire soutiennent un rôle majeur de la démographie et nos simulations montrent que les fortes discordances entre patrons mitochondriaux et nucléaire peuvent être réconciliés sous un même scénario démographique. Bien que massive, l'introgression mitochondriale pourrait être un accident démographique favorisé par des traits comportementaux liés au sexe et associés à son mode de transmission particulier, et aurait même pu avoir des conséquences délétères sur les mâles. Toutefois, nous trouvons que l'introgression pourrait avoir été adaptative, par exemple pour des gènes en relation avec l'immunité. En même temps, l'introgression le long du génome semble freinée par l'interaction entre les variations de recombinaison et l'existence de nombreuses incompatibilités, avec un effet particulièrement fort sur le chromosome X. Dans l'ensemble, le mélange génétique semble globalement empêché

par des incompatibilités, mais localement favorisé par des effets purement démographiques, et des effets sélectifs, soit adaptatifs, soit en réponse à des conflits entre génomes nucléaire et mitochondrial

Keywords

- Introgression
- Hares
- Mitochondrial DNA
- Adaptation
- Genomics
- Range replacement

Palavras-chave

- Introgressão
- Lebres
- ADN mitocondrial
- Adaptação
- Genómica
- Substituição da àrea de distribuição

Index of Contents

Index of Figures and Tables

Abbreviations

BI: Bayesian Inference

bp: base pairs

DNA: Deoxyribonucleic acid

EBSP: Extended Bayesian Skyline Plot

ELAI: Efficient Local Ancestry Inference

EM: Expectation Maximization

FDR: False discovery rate

GO: Gene Ontology

HMM: Hidden Markov Model

IBS: Identical by State

ILS: Incomplete Lineage Sorting

IM: Isolation with Migration

INDEL: Insertion/Deletion Polymorphism

Kb: Kilo-base pair

Kya: Thousand Years

MCMC: Markov chain Monte Carlo

ML: Maximum-Likelihood

mtDNA: mitochondrial DNA

Myr: Million Years

PCA: principal component analysis

PCR: Polymerase Chain Reaction

PIRs: Phase informative reads

PSMC: Pairwise Sequentially Markovian Coalescent

RNA: Ribonucleic acid

RND: Relative Node Depth

SNP: Single-nucleotide polymorphism

Chapter 1.
General Introduction

1. The origin of species and the semipermeable view of Speciation
2. Hybridization and Introgression
3. Mitochondrial introgression in animals
4. Hares as a model to study speciation with gene flow
5. Objectives and organization of the thesis
6. References

1. The origin of species and the semipermeable view of Speciation

How species are formed is a fundamental question in evolutionary biology. But what is a species? While this may be easy to answer for distantly related species, the task becomes complex for closely related entities. The main challenge in defining a species comes from the fact that, while species are thought as discrete entities, speciation is a continuous process, gradually leading from a single population to two sister populations that have fixed differences. This question has long led to several and extensive debates and numerous definitions have been proposed (De Queiroz 2007), each of these using its own although many times partially overlapping criteria – morphologic, ecologic, or phylogenetic, among others.

Species definitions greatly vary depending on the field in which they are applied (e.g. Hausdorf 2011). One of the most widely adopted species definition is the Biological Species Concept (BSC; Dobzhansky 1937; Mayr 1942). The BSC defines species as "groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups" (Mayr 1942). Following this definition, speciation can be understood as the attainment of reproductive isolation, i.e. the creation of effective barriers that result in the cessation of gene flow between the newly formed entities, which would otherwise threaten the integrity of the divergent genomes by the homogenization of the gene pools.

Although the BSC species concept ultimately relies on the formation of intrinsic barriers to gene flow, there is the initial underlying assumption that most speciation processes need a period of geographic isolation between populations (allopatric speciation). During such period, the absence of gene flow between the isolated populations would allow for both drift and selection to enhance divergence and thus the building of reproductive isolation. Other geography-based modes of speciation, namely parapatric (divergence of neighbouring populations in results of a transition in the environment that reduces gene flow) and sympatric (species formation from a single randomly mating population without geographic isolation) were considered unlikely as continuous gene flow would prevent the development of reproductive isolation.

Following Mayr, the notion of virtually universal importance of allopatric speciation has remained for many years and considered as the null hypothesis (Coyne and Orr 2004). However, given the duration of the speciation process and that species distributions did not remain static in result of past climatic oscillations, either promoting ranges expansions, contractions, fragmentation or displacement (Hewitt 2011), the maintenance of strict allopatry during the complete process (or any other strict mode of

geographic speciation) is not plausible (Butlin et al. 2008). Alternatively, different stages of the speciation process likely occur in different spatial contexts, divergence being promoted during periods of geographic isolation while hybridization could occur in periods of secondary contact when reproductive isolation is not yet complete, possibly resulting in genetic exchange. Considering these aspects, the study of speciation shifted from focusing on the geographic context of speciation to focusing on the history of gene flow, resulting in the emergence of models of speciation that include gene flow - e.g. Isolation with Migration models (Nielsen & Wakeley 2001; Hey 2005) and Genic View of Speciation (Wu 2001).

Along with this vision that divergence between lineages can be maintained in the face of gene flow it became commonly accepted that, instead of a single cohesive unit (one underlying assumption of the BSC was that the genome evolves as a single cohesive unit involving large sets of strongly co-adapted genes), genomes behave more as mosaics (Wu 2001, others), parts of it being associated with differentiation and isolation, while others are free to be exchanged. From the use of molecular markers, it became clear that the history of divergence of populations could not be depicted from a single random genetic marker, as different markers at times showed discordant patterns (Avise 2004). These different stories result from the fact that different markers may have different modes of inheritance (e.g. nuclear vs organelles) and also from the fact that the genomes are inherited from two parents but during meiosis they are recombined and independently segregated, which results in partially independence of the markers. Furthermore, although markers are expected to achieve reciprocal monophyly with time, the time needed depends on several factors which also vary among markers. These include the stochastic nature of lineage extinction or maintenance due to the random sampling process of variants inherent to reproduction, the effective population size of the genomic region in cause and selection. Also introgression, which can be defined as the "incorporation (usually via hybridization and backcrossing) of alleles from one entity (species) into the gene pool of a second, divergent entity (species)" (Harrison & Larson 2014) can create such discordant phylogenetic patterns among genomic regions that can be more or less prone to introgress. This is supported by early evidence of heterogeneous species boundaries from the study of hybrid zones (Harrison 1990).

It is thus clear that the definition of species is complex and that species boundaries are perhaps best described as semipermeable and variable in time (Figure 1.1).

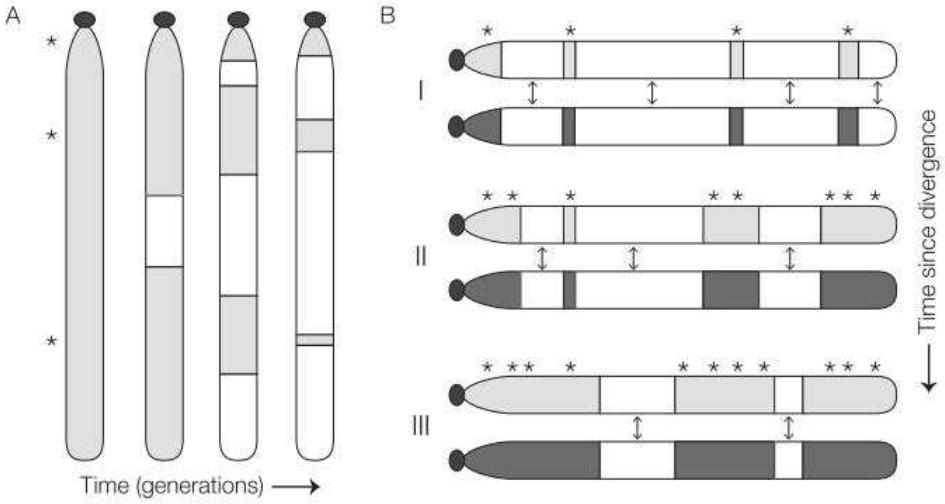


Figure 1.1. Illustration of the semipermeable nature of species boundaries (from Harrison and Larson, 2014). (A) Gene flow following secondary contact continuously homogenizes the genome at neutral loci (white regions) while reproductive barriers (indicated by *) maintain differentiation. With time recombination reduces the extent of linkage-disequilibrium with reproductive barriers and a greater extent of the genome is exchanged. (B) Increasing extension of reproductive isolation along chromosomes with time. In initial steps of divergence only the few loci responsible for reproductive isolation (indicated by *) remain differentiated while other parts of the genome (white) are exchanged. With increasing genetic divergence more loci start contributing to reproductive barriers, thus restricting gene flow in a larger extent of the genome.

2. Hybridization and Introgression

The importance of hybridization and gene flow in evolution has been long appreciated in the study of diversification of plants but only relatively recently became recognized as relevant in animal evolution. This early view was largely influenced by the general acceptance of the Biological Species Concept (BSC), according to which hybridization was a rare phenomenon in animals since F1 hybrids should generally be less viable or sterile, and backcrosses to the parentals normally produce genotypes of inferior viability that are eliminated by natural selection (Mayr 1963). Inter-specific hybridization and introgression of genes between animal taxa were thus seen as rare events that typically lead to evolutionary dead-ends (Mayr, 1963).

However, in the recent decades, the field of evolutionary biology has assisted to a paradigm shift with a growing interest in the study of hybridization and its evolutionary outcomes (see e.g. Arnold 2015). With the advent of the ability to collect molecular genetics data from natural populations, and the accumulation of such data, the notion that gene flow between animal is rare was quickly challenged, as interspecific gene flow became often described, particularly between closely related species (e.g. Good *et al.* 2008; Sequeira *et al.* 2011; Melo-Ferreira *et al.* 2012). For example, by surveying the literature, Mallet (2005) estimated that at least 25% of plant species and 10% of animal species hybridize with at least one closely related species. Likewise, Pinho & Hey (2010) revised the literature estimating gene flow among recently diverged taxa and found that most sister species diverge without or with very little gene flow but that a considerable proportion evolved with a substantial fraction of gene flow (Figure 1.2). Furthermore, they found a decline of the amount of gene flow with time, suggesting that gene flow is more likely to happen when divergence is still recent. Recently, a meta-analysis of 61 population/species pairs at variable divergence states, corroborated this view semi-permeability of the genomes and pervasiveness of genetic exchanges which were observed to continue in species with up to 2% net divergence, independently of their life-history or ecology (Figure 1.3)(Roux *et al.* 2016).

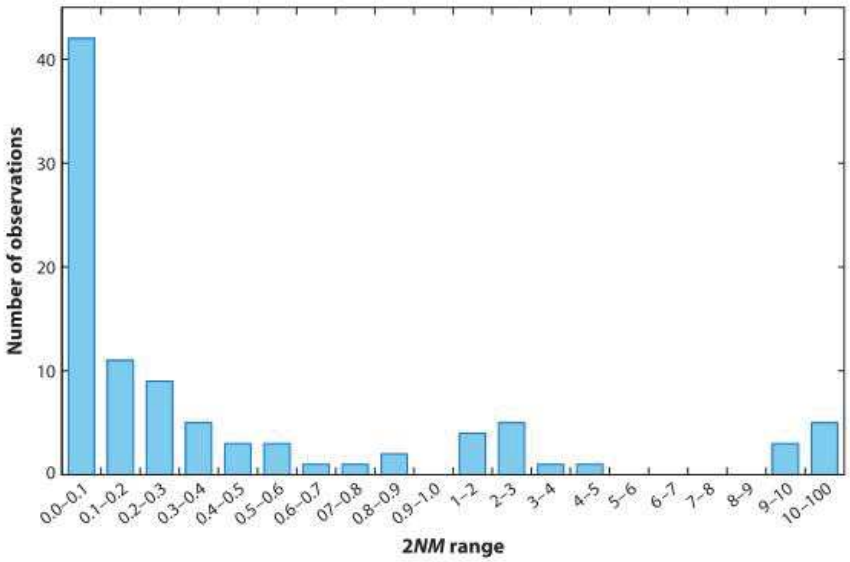


Figure 1.2 The range of gene flow estimates between closely related species (from Pinho and Hey 2010). While most species pairs evolved without or with very little gene flow, in a considerable number of them evolution was accompanied by substantial gene flow.

Interestingly, studies reporting interspecific gene flow also unveiled that patterns of introgression are not homogeneous along the genome (see Harrison & Larson 2014, 2016). Theory predicts that the probability that variation at a locus crosses the species barrier depends on a balance between selection, recombination and dispersal (Barton 1979). Genomic regions responsible for or linked to differential adaptation and/or reproductive isolation are expected to display lower levels of introgression, while neutral unlinked regions should more easily be exchanged (Barton 1979; Barton & Bengtsson 1986; Wu 2001). Furthermore, globally advantageous alleles, either because they increase fitness of individuals in the alternative habitat or interact positively in the foreign genetic background will tend to introgress easily (Barton 1979, 2001). The study of introgression can thus provide valuable information about the past history of species' interactions, the genomic architecture of reproductive barriers, and the adaptive processes that may have been acquired through introgressive hybridization (Abbott *et al.* 2016; Payseur & Rieseberg 2016).

Genomic studies of speciation are now focusing on the differential patterns of divergence across the genomes of species in order to identify the location and number of genomic regions involved in isolation. The study of differential introgression to understand the genetic basis and the architecture of reproductive isolation is particularly promising in naturally hybridizing populations (Nachman & Payseur 2012; Abbott *et al.*

2013; Harrison & Larson 2014, 2016; Payseur & Rieseberg 2016). Compared to laboratory studies of speciation, hybrid zones have the advantage of allowing estimating the fitness of hybrid genotypes under natural conditions along several generations of recombination, which potentially allow fine-scale mapping of genes that contribute to reproductive isolation (Barton and Hewitt 1985; Barton and Gale 1993, Payseur 2010). For instance, by studying both geographic and genomic clines in two recently diverged species of field cricket Larson *et al.* (2014) found several loci showing abrupt clines, which were highly consistent among different geographic regions with different environmental variables, thus suggesting that non-ecological prezygotic barriers (i.e. mate preference, post-mating prezygotic barriers) are likely responsible for maintaining the species boundaries. In another study, Toews *et al.* (2016) show that although genomic differentiation is very low across the genome of golden-winged and blue-winged warblers, six regions associated with feather development or pigmentation exhibit strong differentiation, two of them mapping to the Z-chromosome. Notwithstanding the undeniable potential of population genomic studies to help our understanding of reproductive isolation, interpretation of heterogeneous patterns across the genomes has been the matter of recent debate. While, first studies describing genomic regions of differentiation interpreted these as “genomic islands of speciation” it has been argued that these would be better described as “genomic islands of differentiation” as these could also result from high background selection (or selective sweeps) associated with reduced recombination (Noor and Bennett 2009; Turner and Hahn 2010; Nachman and Payseur 2012; Cruickshank and Hahn 2014)

Documenting the extent and timing of admixture between diverging species can also help clarify the long-standing debate on the role of geographic isolation in speciation. Populations diverging for most of their time in allopatry are expected to show genetic divergence along most of their genome, but if affected by secondary contact this should be characterized by a burst of recent gene flow. On the contrary, in populations diverging mostly in sympatry (or parapatry) genomic divergence is expected to be concentrated in the few regions responsible for the establishment and maintenance of species differentiation and should have a signature of continuous gene flow during speciation. For instance, by sampling both sympatric and allopatric populations and at different stages of the evolutionary process Martin et al. (2013) found evidence that divergence in *Heliconius* species has occurred in the presence of genome-wide admixture which was persistent during long periods of time since the divergence of the species.

Another exciting avenue is the study of introgression as a source of adaptive introgression (Abbott *et al.* 2013). Interspecific gene flow can be source of evolutionary novelty by creating new combinations of alleles not present in any of the parents or by introducing combinations of selectively favoured alleles from one population to another that have already been "tested" by natural selection (Abbott *et al.* 2013; Hedrick 2013). The incorporation of variants already adapted to certain environment in a closely related species can be particularly important during the process of invasion or colonization of new habitats (Rieseberg *et al.* 2007; Hovick & Whitney 2014; but see Rius & Darling 2014). Furthermore, introgression may potentially allow adaptation at faster rates than those possible through de novo mutation (Abbott *et al.* 2013; Hedrick 2013; but see Barton 2013). Adaptive introgression has been long suggested in plants (Anderson 1949) and several examples have been described in a number of plant species (e.g. Louisiana Iris (*Iris*; Martin *et al.* 2006), sunflowers (*Helianthus*; Whitney *et al.* 2010, 2015), ragwort (Kim *et al.* 2008), or poplar (*Populus*; Suarez-Gonzalez *et al.* 2016). Evidence of adaptive introgression in natural animal populations was rare until very recently (Hedrick 2013). However, a few convincing examples started to emerge over the last few years, such as in mice (Song *et al.* 2011; Liu *et al.* 2015), mosquitoes (Clarkson *et al.* 2014; Fontaine *et al.* 2015; Norris *et al.* 2015), butterflies (Pardo-Diaz *et al.* 2012; The Heliconius Genome Consortium 2012; Zhang *et al.* 2016), Darwin finches (Lamichhaney *et al.* 2015) and humans (see Racimo *et al.* 2015). This has been in great part possible by the use of genomic datasets, which have helped uncovering previously undetected cases of introgression, sometimes involving extinct lineages, as is the case in humans (e.g. Reich *et al.* 2010; Green *et al.* 2010; Meyer *et al.* 2012; Prüfer *et al.* 2014).

Interspecific gene flow can also be studied as a mechanism by which new species originate. Hybrid speciation, i.e. the creation of a population that is distinguishable and reproductively isolated from their parents, can either occur without change in chromosome number (homoploid speciation) or with duplication of chromosome number (allopolyploid speciation). While the first is considered rare, allopolyploid speciation plays a major role in the evolution of plants (see Baack & Rieseberg 2007; Abbott *et al.* 2013) and has also been acknowledged in some cases in animals (Nolte *et al.* 2005; Elgvin *et al.* 2011; Hermansen *et al.* 2011).

Finally, interspecific gene flow can be a major concern for conservation. Gene flow contributes to the homogenization of gene pools and can result in the loss of local adaptive variants, outbreeding depression and species collapse (Rius & Darling 2014).

This is particularly relevant since hybridization as a result of human-driven disturbances has been increasing over the years (see Abbott *et al.* 2013).

In sum, with the accumulation of empirical studies using molecular markers it became clear that speciation can continue even with some periods of gene flow, which can occur even between recognized species, and that introgression can have a multitude of potential evolutionary impacts (Abbott et al. 2013).

3. Mitochondrial introgression in animals

One of the most striking patterns in studies that show differential introgression across markers, both by its frequency but also sometimes by its geographical extent, is the pervasiveness of mitochondrial DNA (mtDNA) introgression when compared to nuclear markers (i.e. cytonuclear discordance). In a recent survey, Toews & Brelsford (2012) investigated 126 cases of animal species showing discordant cytonuclear introgression, which was defined as a significant difference in the patterns of introgression between these mitochondrial and nuclear DNA markers (Figure 1.3). This pattern was shown not only to be common, but also to vary in its geographic extent, either being extensive, with complete replacement of the native mtDNA, or more limited, with high frequencies of mtDNA introgression being restricted to a part of the range of the introgressed species.

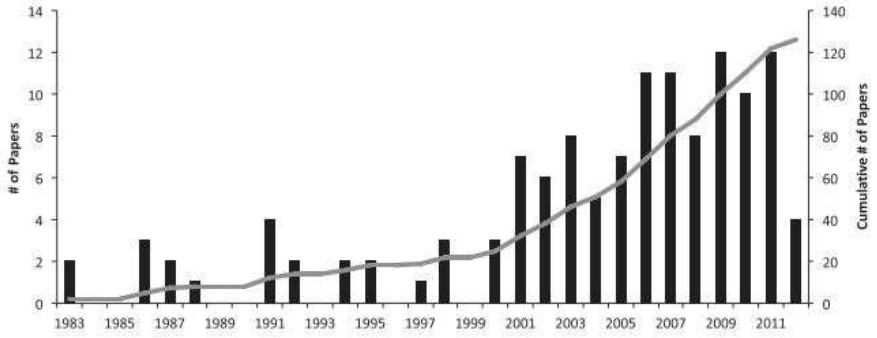


Figure 1.3 Illustration of the increasing number of studies (black bars) reporting mito-nuclear discordance (from Toews and Brelsford, 2012). The cumulative distribution is given by the grey line.

Traditionally, the observation that the mtDNA tends to introgress more than nuclear encoded markers was explained by the neutrality of mtDNA and the fact that it segregates independently from the nuclear genome. Therefore, the mtDNA was thought to be less subject to either direct selection or indirect selection acting over genes that contribute to reproductive isolation as compared to linked nuclear genes (Funk and Omland, 2003 and references therein). However, several lines of evidence contradict this hypothesis. First, nuclear regions not-linked to loci responsible for reproductive isolation should be less constrained to introgress and introgression of these loci could even be enhanced if positively favoured (Barton 2001, 1979; Barton and Bengtsson 1986; Wu 2001). Second, the mitochondrial and nuclear genomes interact in many important physiological functions (e.g. the oxidative phosphorylation - OXPHOS), and mitochondrial DNA transcription, translation and replication depend on nuclear encoded

factors that need to correctly bind to regulatory motifs in the mtDNA to initiate these processes (Smits *et al.* 2010). Given these interdependence these two genomes likely co-evolve (Burton *et al.* 2013; Wolff *et al.* 2014; Sloan *et al.* 2017) and furthermore several examples can be found linking reproductive isolation to cytonuclear incompatibilities (also plastid-nuclear in plants) (see Burton & Barreto 2012; Burton *et al.* 2013). Finally, although mtDNA has been for long considered to be a neutral marker, mitochondrial DNA variation may have significant metabolic and fitness consequences and thus be the target of natural selection (Ballard & Whitlock 2004; Dowling *et al.* 2008; Galtier *et al.* 2009).

Some studies associate massive mtDNA introgression with a selective advantage of the introgressed mitochondria, often related with adaptation to temperature and metabolism. For instance, mtDNA introgression from the arctic char into brook char populations in colder high-altitude lakes as compared to low elevation warmer habitats has been suggested to have occurred as result of adaptation to temperature (Doiron *et al.* 2002). Also in the Eastern Yellow Robin (*Eopsaltria australis*), geographical structured mitochondrial DNA lineages correlate with climate variables, and mtDNA introgression was hypothesized to have been driven by temperature-related adaptation (Morales *et al.* 2015, 2016). In goats, mtDNA introgression is thought to have been promoted by a selective advantage related with adaptation to higher altitude habitats (Ropiquet & Hassanin 2006). In another case, past mtDNA adaptive introgression within a warbler species was associated with different migratory behaviours, mitochondria in different populations showing different metabolic efficiency (Toews *et al.* 2014). However, in the majority of the cases, interpretation of adaptive introgression is speculative or based on indirect evidence and other neutral explanations cannot be discarded (Toews & Brelsford 2012).

It has been suggested that massive mtDNA introgression could be a likely incidental outcome of the process of replacement of a resident species by an invading one, through a purely demographic and drift process. During the range expansion of species, due to the low population density at the front of the invasion wave, the frequency of new and rare variants can increase through a purely demographic and drift process coined as “allele surfing on an expansion wave” (Excoffier & Ray 2008; Currat et al. 2008; Excoffier et al. 2009). When these variants emerge due to hybridization events at the front of invasion a possible outcome is massive introgression from the resident species into the newly colonized territory of the invading species. This demographic process may determine the spread of introgressed nuclear DNA variants, but

introgression is expected to be more prevalent for markers transmitted by the least-dispersing sex (Petit & Excoffier 2009). Thus, in species with male-biased dispersal as is the case of most mammal species, introgression of the maternally transmitted mtDNA tends to be more massive. In addition, the reduced effective size of mtDNA when compared to nuclear DNA may also explain mtDNA biased introgression as it can lead to faster fixation (or loss) of the introgressed alleles due to stronger genetic drift (Takahata and Slatkin, 1984).

Other sex-related asymmetries have also been suggested to promote cytonuclear discordances in patterns of introgression, which involve both incomplete pre-mating and post-mating barriers (see Chan and Levin, 2005). For instance, sex-biased mattings can occur due to frequency-dependent assortative mating. Because of their greater investment in mating, females tend to be choosier than males. When the densities of the two species in contact vary, while the females of the more abundant species will opt to mate with their conspecific males, females of the other species failing to encounter conspecific males will end up mating with a male from the other (see Chan and Levin, 2005 and references therein). Male competition can also lead to sex asymmetries if males of one of the species are able to outcompete the males of the other in mating with females (including heterospecific females).

4. Hares as a model to study speciation with gene flow

Hares and Jackrabbits (genus *Lepus*) belong to order Lagomorpha within which two families are currently recognized Ochotonidae and Leporidae. The former comprises only the genus *Ochotona* (pikas) while the Leporidae include both rabbits (10 genera), and hares and jackrabbits (only genus *Lepus*). Genus *Lepus* is thought to have diverged from rabbits *ca.* 11.8 Mya, likely originating in North America from where it radiated to other continents crossing the Bering Strait ca 5-7 Mya (Matthee et al. 2004; Melo-Ferreira et al. 2012). This genus, which is the most speciose and widespread leporid genus, comprises over 30 species currently distributed all over the world (Alves & Hackländer 2008). This recent and explosive radiation of hares is thought to have resulted from the increasing availability of suitable temperate grasslands 4-6 MYA and the formation of the west Antarctic ice sheet at that period (Matthee et al, 2004; Yamada et al, 2002). Currently, the different hare species can be found in a great variety of habitats, such as tundra and open forest (*L. timidus*), deserts (*L. capensis*), dense boreal forests (*L. americanus*) and open steppe (*L. europaeus*), evidencing the great ecological plasticity of the genus. Furthermore, species distribution ranges very greatly, some species having restricted distributions (e.g. *L. castroviejoi* occurs in the Cantabrian Mountains in Northeast Spain only) while others occur over wide areas (e.g. *L. timidus* occurs in all northern Europe and Asia) (Alves and Hackländer, 2008).

Besides their recent and rapid diversification, hares are also characterized by numerous instances of interspecific gene flow (reviewed in Alves et al, 2008) which make this genus an excellent model to study introgression. Most of the currently described cases seem to involve the mtDNA of the mountain hare (*L. timidus*) which is potentially present in more than 10 species (Alves *et al.* 2008b; Melo-Ferreira *et al.* 2012), both in Eurasia and the North America. While genetic studies in North American hares are still scarce, their European counterparts have been studied in more detail, particularly in respect to hybridization and introgression.

First reports of hybridization and introgression in hares involved the brown hare (*L. europaeus*) and the mountain hare (*L. timidus*). The former was introduced in Sweden in the 1800s and has been since expanding northwards replacing the native *L. timidus* (Thulin 2003). Since its introduction, intermediate forms between the two species were reported and hybridization between the two was confirmed by the analysis of mtDNA variation that showed that the mtDNA of *L. timidus* was present in ca. 10% of *L. europaeus* individuals (Thulin 1997, Thulin and Tegelström, 2002). Interestingly, no introgression was found in the reverse direction in agreement with results from crosses

in captivity in which spontaneous crosses between *L. europaeus* males and *L. timidus* females were observed while the reverse cross could only be performed by artificial insemination (Gustavsson & Sundt 1965). Asymmetric introgression of *L. timidus* mtDNA into *L. europaeus* was also observed in Denmark (Fredsted et al. 2006), and has been suggested to be caused by frequency-dependent hybridization possibly coupled with male competition (Thulin & Tegelström 2002; Thulin et al. 2006a). However, bidirectional mtDNA introgression has been also observed both in Russia (Thulin et al. 2006a) and in the Alps (Suchentrunk et al. 2005, Zachos et al. 2010).

While the previous cases of interspecific gene flow report to areas of present contact of species, instances of introgression were also reported in historical contacts that no longer exist. In the Iberian Peninsula, the mtDNA of *L. timidus*, a species that went locally extinct after the Last Glacial Maximum, can be found in populations of three species inhabiting these region (see Figure 1.4. for distribution of European hare species), namely *L. castroviejoi* (broom hare), *L. granatensis* (Iberian hare) and *L. europaeus*. Mitochondrial DNA introgression into these three species is remarkable by its frequency. In *L. castroviejoi*, which is endemic to the Cantabrian mountains in northern Spain, the native mtDNA is no longer present and instead two mitochondrial lineages are found (Melo-Ferreira et al. 2012). These are likely the result of two distinct introgression events, one during the middle Pleistocene, affecting the common ancestor of *L. castroviejoi* and Italian *L. corsicanus* (a species inhabiting the Italic Peninsula in which the native mtDNA was also lost) and another at the Last Glacial Maximum (LGM) (Alves *et al.* 2008a; Melo-Ferreira *et al.* 2012). Also, in *L. europaeus* populations, which in Iberia are restricted to the Pyrenean foothills, the frequencies of the introgressed lineage reach quasi-fixation. In *L. granatensis*, which is endemic to Iberia and occupies most of the Peninsula, *timidus* mtDNA reaches high frequencies in populations of the northern half of Iberia while it is absent in the south (Melo-Ferreira *et al.* 2005, 2009). In contrast, levels of nuclear gene flow into any of these three species was found to be generally very limited (Melo-Ferreira *et al.* 2009, 2012), although a more thorough investigation of the nuclear genome could potentially uncover other patterns, as is the case for one X-chromosome marker which shows extensive introgression (Melo-Ferreira *et al.* 2011). Also in North America, mtDNA introgression is suspected to occur between the black-tailed jackrabbit (*L. californicus*) and the snowshoe hare (*L. americanus*), since *L. americanus* populations of the Pacific Northwest region harbour mtDNA variants that more closely resemble those found in *L. californicus* as compared to variants found in *L.*

americanus from other areas (Cheng *et al.* 2014). Still, whether such similarity resulted from introgression or incomplete lineage sorting was not tested.

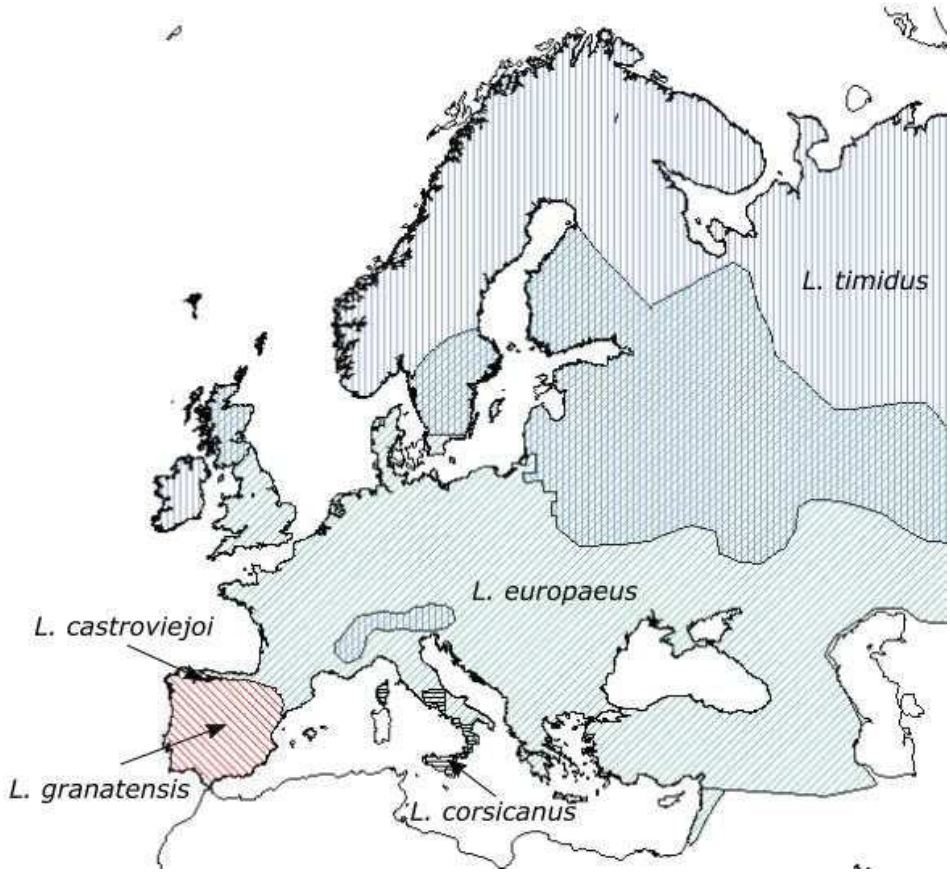


Figure 1.4. Geographic distribution of European hare species according to Mitchell-Jones *et al.* (1999).

The multitude of cases of interspecific and generally unidirectional mtDNA gene flow involving *L. timidus* as the donor species along with the pervasiveness of mtDNA introgression in some cases, has raised questions about its potential adaptive drive (Melo-Ferreira *et al.* 2005, 2009, 2014b). Since mitochondrial metabolism is involved in thermoregulation and mtDNA sequence variation has in several instances been associated with temperature-related adaptation (Ruiz-Pesini *et al.* 2004; Sun *et al.* 2007; Silva *et al.* 2014), mtDNA introgression of the arctic/boreal *L. timidus* may confer a selective advantage related to cold. In Iberian *L. granatensis* this could explain the northward increase in the frequency of introgression of *L. timidus* mtDNA (Melo-Ferreira et al. 2005). In fact, inferences of past demography of the native and introgressed mtDNA

lineages in northern Iberia, where they coexist in the same populations, suggested that the latter out-competed the former, as expected if it was selectively favoured (Melo-Ferreira *et al.* 2011). Also, analyses of complete mtDNA sequences showed evidence for positive selection in genes of the OXPHOS complexes, measured by an increased rate of amino-acid substitutions, particularly affecting *timidus* mtDNA (Melo-Ferreira *et al.* 2014b). However, no effects on the structure and physicochemical properties of the encoded proteins were predicted, suggesting that the focus of selection may lie on complex interactions with nuclear encoded peptides.

However, purely demographic neutral processes could also have been an important determinant of introgression patterns. At the last glacial period, *L. granatensis* has likely started expanded north from a southern refugia and replaced *L. timidus*, in a period in which the changing climate would be presumably favourable for the former but not the latter (Melo-Ferreira et al. 2007; Marques et al. 2017). If this northwards expansion and competitive replacement was accompanied by hybridization this could have led to the gradient of introgression observed in *L. granatensis*. In fact, both the south-north gradient of mtDNA introgression (Melo-Ferreira et al. 2005, 2009) and the phylogeographic structure of mtDNA of *timidus* origin perpendicular to the likely direction of expansion origin (Melo-Ferreira et al. 2011) are congruent with this hypothesis. Also in the Iberian Peninsula, *L. europaeus* is thought to have colonized the region forming a contact zone with *L. granatensis* parallel to the direction of colonization. Studying patterns of differentiation in microsatellite and mtDNA loci, Melo-Ferreira et al. (2013) have shown that *L. europaeus* populations from this region show less mtDNA differentiation with populations of *L. granatensis* across the border of these species than within *L. europaeus* populations along it, although nuclear introgression is quasi-absent between these two species. These results are compatible with an invasion of *L. europaeus* with successive hybridization events with the resident species at the time that could have either been *L. granatensis* (if already introgressed) or *L. timidus*, in each capturing the resident mtDNA thus explaining the lower mtDNA differentiation across the species border (Melo-Ferreira *et al.* 2014a).

With the current knowledge, understanding whether demographic or selective processes (or both) were responsible for massive mtDNA introgression in hares is still a matter of debate, and represents an ideal case-study to tackle this important question. The current availability of genomic datasets now allows us to sample a large number of loci across the genome, which can be viewed as replicates of the same hybridization history giving us power to accurately reconstruct that history and its demographic

context. Furthermore, by scanning the genomes of hybridizing species we can now link genome patterns of interspecific gene flow to the functional and regional contexts which can give us important clues about the role of selection in shaping them. For instance, when analyzed in relation to mtDNA introgression, preferential co-introgression or co-differentiation of nuclear genes interacting with the mitochondria would give convincing evidence of a functional role of mtDNA introgression (adaptive or not).

5. Objectives and organization of the thesis

The general objective of this thesis is to appraise the relative roles of historical demography and natural selection in determining patterns of genomic introgression, using hares as model system. Such an appraisal is fundamental to understand the extent and relevance of adaptive introgression as a source of new adaptations during the evolutionary history of species. Taking advantage of the multiple instances of interspecific gene flow reported in the genus, often involving the mitogenome, an important aim is to investigate the selective causes/consequences of massive mtDNA introgression, and in particular its co-evolution with the nuclear genome. The following main specific objectives were defined:

i) Clarifying the presumable extension of mtDNA vast reticulation from the European hares, where it has been often described, to North American hare species (*L. americanus*, *L. californicus* and *L. townsendii*);

ii) Assessing the importance of historical gene flow between *L. timidus*, *L. granatensis* and *L. europaeus* using a genome-wide perspective;

iii) Formally testing the hypothesis that a range expansion of *L. granatensis* into the historical range of *L. timidus* with hybridization is a major determinant of introgression patterns including of mtDNA;

iv) Assessing the selective consequences (if any) of historical *timidus* mtDNA introgression in the nuclear genome of both *L. granatensis* and *L. europaeus*, either by promoting co-introgression or co-evolution of nuclear encoded genes, particularly those found in or interacting with the mitochondria;

v) Investigating whether historical hybridization events may have led to adaptive introgression, and pinpointing the functional envelope of such adaptation;

vi) Clarifying the complex dynamics of interspecific interactions leading to gene flow among hares (*L. timidus*, *L. granatensis* and *L. europaeus*) in Northern Iberia Peninsula;

vii) Using the repeated introgression events from *L. timidus* into the southern European species to identify common regional selective pressures that may have driven introgression irrespective of the predominant genomic background.

This thesis is organized in four chapters. It includes three scientific manuscripts: one already published in a journal indexed in the Science Citation Index (SCI), one submitted and another in preparation.

The first and current chapter, '*Chapter 1 - General Introduction'*, gives a brief presentation of the subject of this work - the study of speciation with gene flow and the relevance of hybridization and introgression to the evolutionary history of species - followed by a general presentation of the model system used in this work.

In '*Chapter 2 – Promiscuous mitochondrial DNA in hares'* we use a multi-locus dataset to infer the relationships among three North American species and test whether mitochondrial DNA introgression is part of their speciation history, as previously suspected. The work presented in this chapter resulted in one publication in the journal *Molecular Ecology* (journal indexed in the SCI).

Paper I. Melo-Ferreira J, Seixas FA, Cheng E, Mills LS, Alves PC (2014) Molecular Ecology, 23, 4617–4630. **The hidden history of the snowshoe hare, *Lepus americanus*: extensive mitochondrial DNA introgression inferred from multilocus genetic variation.**

The major challenge of the work included in *Paper I* was to infer whether close similarity of mtDNA haplotypes in distinct species can be reconciled under a model with no gene flow with incomplete lineage sorting, or introgression needs to be invoked. We used coalescent-based methods and simulation approaches to show that, like in their European counterparts, gene flow played a role in the mtDNA evolutionary history of North American hares. Also similarly, frequencies of the introgressed mtDNA variants reach high frequencies at a local scale.

In '*Chapter 3 – Genomic perspective of introgression in hares from Iberia'* we use a genomic approach to characterize the patterns of introgression and differentiation genome-wide among three species in northern Iberian Peninsula in order to understand the relative role of selective and demographic processes in the reticulate evolution of these species. Having set that massive mtDNA introgression is a general phenomenon in hares, involving many species and not only the European ones where it had been previously described, we focused on the most intensively researched model to investigate the role of gene flow in the genome: introgression into hares from the Iberian Peninsula. The results are presented in two scientific papers, one submitted to SCI indexed journal and the other is in preparation.

Paper II. Seixas FA, Boursot P, Melo-Ferreira J (2017) **The genomic impact of historical hybridization with massive mitochondrial DNA introgression in the Iberian hare (*Lepus granatensis*).** *Submitted*

Paper III. Seixas FA, Farelo L, Belkir K, Alves PC, Boursot P, Melo-Ferreira J (2017) **Genomic exchanges between three hare species sharing the same mitochondrial genome following massive introgression: the roles of history, adaptation and cytonuclear coevolution.** *In preparation*

In *Paper II* we explore genome-wide patterns of introgression and perform extensive simulations to show that both demography and selection shaped the patterns of *L. timidus* variation present in the current genomes of *L. granatensis*. In *Paper III* we clarify the complex history of replacement and admixture events in northern Iberian Peninsula and find evidence of possible adaptive introgression of *L. timidus* variants into both *L. granatensis* and *L. europaeus* involving genes with similar functions. These studies underline the importance of demographic processes alone promoting introgression which can create, in association with sex-biases, discordant patterns among inheritance compartments but also suggest that selection may play an important role either in allowing species to adapt to newly colonized habitats or in the maintenance of genomic cohesion.

Finally, in 'Chapter 4 – General Discussion' we discuss the most relevant results and both their specific and general implications. Also, the major conclusions as well as implications for future studies and possible lines of research are discussed.

6. References

Abbott RJ, Albach D, Ansell S et al. (2013) Hybridization and speciation. Journal of Evolutionary Biology, 26, 229–246.

Abbott RJ, Barton NH, Good JM (2016) Genomics of hybridization and its evolutionary consequences. Molecular Ecology, 25, 2325–2332.

Alves PC, Hackländer K (2008) Lagomorph species: geographical distribution and conservation status. In: Lagomorph biology: evolution, ecology, and conservation (eds Alves P, Hackländer K, Ferrand N), pp. 395–405. Springer. Berlin, Germany.

Alves PC, Melo-Ferreira J, Branco M et al. (2008a) Evidence for genetic similarity of two allopatric European hares (*Lepus corsicanus* and *L. castroviejoi*) inferred from nuclear DNA sequences. Molecular Phylogenetics and Evolution, 46, 1191–7.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P (2008b) The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, 363, 2831–2839.

Anderson E (1949) Introgressive hybridization. Wiley & Sons, New York.

Arnold ML (2015) Divergence with Genetic Exchange. Oxford University Press, Oxford, UK.

Avise J (2004) Molecular Markers, Natural History, and Evolution. Sinauer & Associates, Sunderland, Massachusetts.

Baack E, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. Current opinion in genetics & development, 513–518.

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

Barton NH (1979) Gene flow past a cline. Heredity, 43, 333–339.

Barton NH, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. Heredity, 57, 357–376.

Barton NH (2001) The role of hybridization in evolution. Molecular Ecology, 10, 551–568.

Barton NH (2013) Does hybridization influence speciation? Journal of Evolutionary Biology, 26, 267–269.

Burton RS, Barreto FS (2012) A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Molecular Ecology, 21, 4942–4957.

Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear Genomic Interactions and Hybrid Breakdown. Annual Review of Ecology, Evolution, and Systematics, 44, 281–302.

Cheng E, Hodges KE, Melo-Ferreira J, Alves PC, Mills LS (2014) Conservation implications of the evolutionary history and genetic diversity hotspots of the snowshoe hare. Molecular ecology, 23, 2929–42.

Clarkson CS, Weetman D, Essandoh J et al. (2014) Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. Nature Communications, 5, 4248.

Dobzhansky T (1937) Genetics and the Origin of Species. 364 pp. 1st ed. Columbia Univiversity Press, New York.

Doiron S, Bernatchez L, Blier PU (2002) A Comparative Mitogenomic Analysis of the Potential Adaptive Value of Arctic Charr mtDNA Introgression in Brook Charr Populations (Salvelinus fontinalis Mitchill). Molecular Biology and Evolution, 19, 1902–1909.

Dowling DK, Friberg U, Lindell J (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. Trends in Ecology and Evolution, 23, 546–554.

Elgvin TO, Hermansen JS, Fijarczyk A et al. (2011) Hybrid speciation in sparrows II: A role for sex chromosomes? Molecular Ecology, 20, 3823–3837.

Fontaine MC, Pease JB, Steele A et al. (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science, 347, 1258524–1258524.

Galtier N, Nabholz B, Glémin S, Hurst G (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular ecology, 18, 4541–50.

Good JM, Hird S, Reid N et al. (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. Molecular ecology, 17, 1313–27.

Green RE, Krause J, Briggs AW et al. (2010) A draft sequence of the Neanderthal genome. Science, 328, 710–22.

Harrison RG (1990) Hybrid zones: windows on evolutionary process (AJ Futuyma D, Ed,). Oxford University Press, New York.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. Journal of Heredity, 105, 795–809.

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Molecular Ecology, 25, 2454–2466.

Hausdorf B (2011) Progress toward a general species concept. Evolution, 65, 923–931.

Hedrick PW (2013) Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology, 22, 4606–4618.

Hermansen JS, Sæther SA, Elgvin TO et al. (2011) Hybrid speciation in sparrows I: Phenotypic intermediacy, genetic admixture and barriers to gene flow. Molecular Ecology, 20, 3812–3822.

Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. Genetica, 139, 617–638.

Hey J (2005) On the number of new world founders: A population genetic portrait of the peopling of the Americas. PLoS Biology, 3, 0965–0975.

Hovick SM, Whitney KD (2014) Hybridisation is associated with increased fecundity and size in invasive taxa: Meta-analytic support for the hybridisation-invasion hypothesis. Ecology Letters, 17, 1464–1477.

Kim M, Cui M-L, Cubas P et al. (2008) Regulatory genes control a key morphological and ecological trait transferred between species. Science, 322, 1116–9.

Lamichhaney S, Berglund J, Almén MS et al. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature, 518, 371–375.

Larson EL, White TA, Ross CL, Harrison RG (2014) Gene flow and the maintenance of species boundaries. Molecular Ecology, 23, 1668–1678.

Liu KJ, Steinberg E, Yozzo A et al. (2015) Interspecific introgressive origin of genomic diversity in the house mouse. Proceedings of the National Academy of Sciences, 112, 196–201.

Mallet J (2005) Hybridization as an invasion of the genome. Trends in Ecology & Evolution, 20, 229–237.

Marques JP, Farelo L, Vilela J et al. (2017) Range expansion underlies historical introgressive hybridization in the Iberian hare. Scientific Reports, 7, 40788.

Martin NH, Bouck AC, Arnold ML (2006) Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. Genetics, 172, 2481–2489.

Mayr E (1942) Systematics and the Origin of Species from the Viewpoint of a Zoologist. Columbia University Press, New York.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P (2009) The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. Molecular Ecology, 18, 2643–58.

Melo-Ferreira J, Alves PC, Rocha J, Ferrand N, Boursot P (2011) Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? Evolution, 65, 1956–68.

Melo-Ferreira J, Boursot P, Carneiro M et al. (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. Systematic Biology, 61, 367–381.

Melo-Ferreira J, Boursot P, Randi E et al. (2007) The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. Molecular ecology, 16, 605–18.

Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. Molecular Ecology, 14, 2459–64.

Melo-Ferreira J, Farelo L, Freitas H et al. (2014a) Home-loving boreal hare mitochondria survived several invasions in Iberia: the relative roles of recurrent hybridisation and allele surfing. Heredity, 112, 265–73.

Melo-Ferreira J, Vilela J, Fonseca MM et al. (2014b) The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. Genome biology and evolution, 6, 886–96.

Meyer M, Kircher M, Gansauge M-T et al. (2012) A high-coverage genome sequence from an archaic Denisovan individual. Science (New York, N.Y.), 338, 222–6.

Mitchell-Jones AJ, Amori G, Bogdanowicz W et al. (1999) The atlas of European mammals. T & AD Poyser Ltd, London.

Morales HE, Pavlova A, Joseph L, Sunnucks P (2015) Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Molecular Ecology, 24, 2820–2837.

Morales HE, Sunnucks P, Joseph L, Pavlova A (2016) Perpendicular axes of incipient speciation generated by mitochondrial introgression. bioRxiv. http://biorxiv.org/lookup/doi/10.1101/072942.

Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 409–421.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics, 158, 885–96.

Nolte AW, Freyhof J, Stemshorn KC, Tautz D (2005) An invasive lineage of sculpins, Cottus sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. Proceedings. Biological sciences / The Royal Society, 272, 2379–2387.

Norris LC, Main BJ, Lee Y et al. (2015) Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proceedings of the National Academy of Sciences, 112, 815–820.

Pardo-Diaz C, Salazar C, Baxter SW et al. (2012) Adaptive introgression across species boundaries in Heliconius butterflies. PLoS Genetics, 8.

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. Molecular Ecology, 25, 2337–2360.

Pinho C, Hey J (2010) Divergence with Gene Flow: Models and Data. Annual Review of Ecology, Evolution, and Systematics, 41, 215–230.

Prüfer K, Racimo F, Patterson N et al. (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. Nature, 505, 43–9.

De Queiroz K (2007) Species concepts and species delimitation. Systematic biology, 56, 879–886.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E (2015) Evidence for archaic adaptive introgression in humans. Nature Reviews Genetics, 16, 359–371.

Reich D, Green RE, Kircher M et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature, 468, 1053–1060.

Rieseberg LH, Kim S-C, Randell RA et al. (2007) Hybridization and the colonization of novel habitats by annual sunflowers. Genetica, 129, 149–165.

Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? Trends in Ecology and Evolution, 29, 233–242.

Ropiquet A, Hassanin A (2006) Hybrid origin of the Pliocene ancestor of wild goats. Molecular Phylogenetics and Evolution, 41, 395–404.

Roux C, Fraïsse C, Romiguier J et al. (2016) Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. PLoS Biology, 14, 1–22.

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of Purifying and Adaptive Selection on Regional Variation in Human mtDNA. Science, 303, 223–226.

Sequeira F, Sodré D, Ferrand N et al. (2011) Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear markers. BMC evolutionary biology, 11, 264.

Silva G, Lima FP, Martel P, Castilho R (2014) Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. Proceedings of the Royal Society B: Biological Sciences, 281, 20141093–20141093.

Sloan DB, Havird JC, Sharbrough J (2017) The on-again, off-again relationship between mitochondrial genomes and species boundaries. Molecular Ecology, 26, 2212–2236.

Smits P, Smeitink J, van den Heuvel L (2010) Mitochondrial Translation and Beyond: Processes Implicated in Combined Oxidative Phosphorylation Deficiencies. Journal of Biomedicine and Biotechnology, 2010, 1–24.

Song Y, Endepols S, Klemann N et al. (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Current Biology, 21, 1296–1301.

Suarez-Gonzalez A, Hefer CA, Christe C et al. (2016) Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). Molecular Ecology, 25, 2427–2442.

Sun C, Kong Q-P, Zhang Y-P (2007) The role of climate in human mitochondrial DNA evolution: a reappraisal. Genomics, 89, 338–342.

The Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature, 487, 94–98.

Thulin CG (2003) The distribution of mountain hares *Lepus timidus* in Europe: A challenge from brown hares *L. europaeus*? Mammal Review, 33, 29–42.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology, 21, 3907–3930.

Toews DPL, Mandic M, Richards JG, Irwin DE (2014) Migration, mitochondria, and the yellow-rumped warbler. Evolution, 68, 241–255.

Toews DPL, Taylor SA, Vallender R et al. (2016) Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. Current Biology, 26, 2313–2318.

Whitney KD, Broman KW, Kane NC et al. (2015) Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. Molecular Ecology, 24, 2194–2211.

Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. New Phytologist, 187, 230–239.

Wolff JN, Ladoukakis ED, Enriquez J a., Dowling DK (2014) Mitonuclear interactions: evolutionary consequences over multiple biological scales. Philosophical Transactions of the Royal Society B: Biological Sciences, 369, 20130443–20130443.

Wu C-I (2001) The genic view of the process of speciation. Journal of Evolutionary Biology, 14, 851–865.

Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR (2016) Genome-wide introgression among distantly related *Heliconius* butterfly species. Genome Biology, 17, 25.

Chapter 2.
Promiscuous mitochondrial DNA in hares

Paper I. Melo-Ferreira J, Seixas FA, Cheng E, Mills LS, Alves PC (2014) Molecular Ecology, 23, 4617–4630. **The hidden history of the snowshoe hare, *Lepus americanus*: extensive mitochondrial DNA introgression inferred from multilocus genetic variation.**

The hidden history of the snowshoe hare, *Lepus americanus*: extensive mitochondrial DNA introgression inferred from multilocus genetic variation

Melo-Ferreira*, J, Seixas, FA[1], Cheng, E, Mills, LS, & Alves, PC

1. Abstract

Hybridization drives the evolutionary trajectory of many species or local populations, and assessing the geographic extent and genetic impact of interspecific gene flow may provide invaluable clues to understand population divergence or the adaptive relevance of admixture. In North America, hares (*Lepus* spp.) are key species for ecosystem dynamics and their evolutionary history may have been affected by hybridization. Here we reconstructed the speciation history of the three most widespread hares in North America - the snowshoe hare (*Lepus americanus*), the white-tailed jackrabbit (*L. townsendii*) and the black-tailed jackrabbit (*L. californicus*) - by analyzing sequence variation at eight nuclear markers and one mitochondrial DNA (mtDNA) locus (6 240 bp; 94 specimens). A multilocus-multispecies coalescent-based phylogeny suggests that *L. americanus* diverged ~2.7 Mya and that *L. californicus* and *L. townsendii* split more recently (~1.2 Mya). Within *L. americanus* a deep history of cryptic divergence (~2.0 Mya) was inferred, which coincides with major speciation events in other North American species. While the isolation-with-migration model suggested that nuclear gene flow was generally rare or absent among species or major genetic groups, coalescent simulations of mtDNA divergence revealed historical mtDNA introgression from *L. californicus* into the Pacific Northwest populations of *L. americanus*. This finding marks a history of past reticulation between these species, which may have affected other parts of the genome and influence the adaptive potential of hares during climate change.

1 These authors contributed equally

2. Introduction

The modern view of interspecific dynamics recognizes that closely related species, even when divergence is irreversible, may exchange genetic material, and that introgressive hybridization plays an important role in shaping the genetic diversity of taxa. Mallet (2005), for example, estimated that 10% of animal species hybridize with at least one other closely related species (see also Pinho and Hey 2010). Understanding patterns of introgression is therefore important to unveil the determinants of major processes of species evolution, such as the genetic nature of population divergence or the generation of adaptive genetic innovation (Abbott et al. 2013; Feder et al. 2012; Seehausen 2004, 2013).

Inferences of introgression have often been based on gene tree polyphyly or paraphyly and incongruence among gene trees (e.g. Bossu & Near 2009; Spinks & Shaffer 2009). However, discordance among markers may arise from the stochasticity of the evolutionary process itself, due to the incomplete sorting of lineages along the divergence of species. Distinguishing these two causes of gene tree discordance is not straightforward, particularly for closely related taxa (Edwards 2009). Nevertheless, several methodological strategies have been created to assess the relative influence of retention of ancestral polymorphism and gene flow in observed patterns of multi-locus genetic variation (e.g. Gerard et al. 2011; Hey 2010; Meng & Kubatko 2009).

Natural hybridization often occurs among species with a rapid and young radiation, and hares (*Lepus* spp.) have emerged as a particularly suitable model to study reticulate evolution (Thulin et al. 2006a, 2006b; Alves et al. 2008; Melo-Ferreira et al. 2009, 2011, 2012; Liu et al. 2011). Even though most instances of introgressive hybridization described among hares relate to areas of present species contact (e.g. between *L. timidus* and *L. europaeus* in Sweden or Russia; Thulin et al. 2006a, 2006b), cases of ancestral introgression between currently allopatric species have also been reported (Alves et al. 2003; Melo-Ferreira et al. 2012). Even though these reticulation events are more pronounced in the mtDNA, they also occur at the nuclear genome, but at different degrees across inheritance pathways and chromosome regions (Melo-Ferreira et al. 2009, 2011, 2012).

Given the widespread nature of genome reticulation and extensive introgression in hares (reviewed by Alves et al. 2008), introgression is expected to have also impacted the evolution of North American species in the U.S. and Canada, with potential consequences to their conservation and adaptive potential. In North America, hares are strong interactors in ecosystem dynamics (Krebs 2011; Lewis et al. 2011; Tyson et al.

2010) and model systems for basic ecological studies ranging from cyclic population dynamics to mechanisms of top-down versus bottom-up population control (Griffin & Mills 2009; Krebs 2011), to the ecology of stress (Boonstra 2013). Also, two of the most widespread hare species in North America (snowshoe hares, *Lepus americanus*, and white-tailed jackrabbits, *L. townsendii*) undergo seasonal coat colour changes, a trait vulnerable to being compromised by climate change, as the number of days of white hares on brown backgrounds increases into the future (Mills et al. 2013; Zimova et al. 2014). Despite these studies, information on the evolutionary history of North American hares is still very scarce. Recently, a comprehensive study by Cheng et al. (2014) based on microsatellites and mitochondrial DNA sequences and covering the entire range of the snowshoe hare suggested that this species is structured in three major evolutionary population clusters with well-defined geographic distributions: Boreal (entire northern and eastern range of the species), Rockies and Pacific Northwest. This pattern of population structure is similar to that inferred for other boreal North American mammals, implying that common phenomena such as climatic oscillations may have shaped the phylogeography of this species. Cheng et al. (2014) also show that the Pacific Northwest population of *L. americanus* possesses an mtDNA lineage that is more closely related to that of the black-tailed jackrabbit, *L. californicus*. This pattern of mtDNA divergence may result from secondary introgression following interspecific hybridization, as often described among species of hares, or from incomplete lineage sorting. However, distinguishing between these competing hypotheses requires reconstructing the speciation history of these species. In addition, Flux (1983) reported that *L. californicus* hybridizes in the wild with the white-tailed jackrabbit but no study of the genetic consequences of this hybridization has been conducted.

In this study, we aim to infer the evolutionary history of the three most widespread North American hare species *L. americanus*, *L. californicus* and *L. townsendii*, by analyzing the sequence variation at nine loci from all inheritance pathways. In addition, we determine the extent and timing of gene introgression in these species, and discuss the potential adaptive importance of hybridization in their evolution.

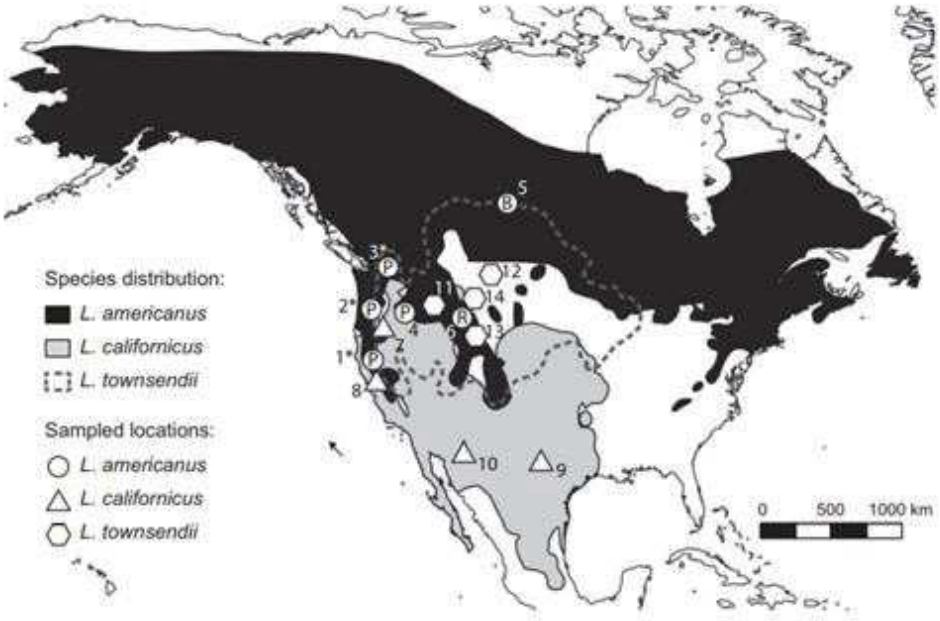


Figure 2.1 Distribution of *L. americanus*, *L. californicus* and *L. townsendii* in North America, and approximate locations of samples used in this study. Letters in *L. americanus* sample locations indicate the microsatellite cluster identified by Cheng et al. (2014): B – Boreal; R – Rockies; P – Pacific Northwest (the localities where the *L. californicus*-like mtDNA was found is indicated by "*"). See Annex I - Table 2.1 for the detailed location of sampling sites (depicted by numbers).

3. Methods

Sampling and data collection

A total of 94 individuals (48 *L. americanus*, 30 *L. californicus* and 16 *L. townsendii*) from 14 sampling locations were used in this study (Table 2.1; Annex I - Tables S2.1 and S2.2), including the three *L. americanus* population clusters described by Cheng et al. (2014) (Figure 2.1). The European rabbit, *Oryctolagus cuniculus*, was used as outgroup for some of the analyses.

Table 2.1 Species and geographic location of the samples collected in this study.

Species	Locality Number	Locality Code[1]	Locality	Sample size
Lepus americanus	1	CA1	California, U.S.A.	8
	2	WA1	Washington, U.S.A.	8
	3	WA4	Washington, U.S.A.	10
	4	OR2	Oregon, U.S.A.	8
	5	SK1	Saskatchewan, Canada	6
	6	WY1	Wyoming, U.S.A.	8
			Total L. americanus	48
Lepus californicus	7	LCA_OR	Oregon, U.S.A.	10
	8	LCA_CA	California, U.S.A.	6
	9	LCA_TE	Texas, U.S.A.	8
	10	LCA_AR	Arizona, U.S.A.	6
			Total L. californicus	30
Lepus townsendii	11	LTO_ID1	Idaho, U.S.A.	8
	12	LTO_MO1	Montana, U.S.A.	2
	13	LTO_WY1	Wyoming, U.S.A.	1
	14	LTO_MO2	Montana, U.S.A (Yellowstone N.P.)	5
			Total *L. townsendii*	16
			Total	94

[1]*Locality codes in L. americanus as in Cheng et al. (2014).*

Total genomic DNA was extracted from muscle and ear tissues using the JETQUICK Tissue DNA Kit (Genomed) following manufacturer's instructions. The sex of the individuals was determined following the PCR approach described by Wallner et al. (2001). Nine loci from all inheritance pathways – five autosomal (SPTBN1, PRKCI, DARC, KITLG, TF), one mitochondrial (CYTB), two X-linked (POLA1, GRIA3) and one Y-linked (SRY) – were amplified by polymerase chain reaction (PCR) (Table 2.2; see Annex I - Table S2.3 for primers and PCR conditions). Purified PCR products were

automatically sequenced (Macrogen Inc, Netherlands) using forward and reverse PCR primers and occasionally internal primers as indicated in Annex I - Table S2.3.

Analysis of sequence data sets

Sequences were visually inspected and aligned using ClustalW (Thompson et al. 1994) as implemented in BioEdit v7.0.5.3 (Hall 1999). Polymorphic tandem repeats and the 5 bp adjacent regions were excluded. Allelic phases were determined using PHASEv2.1.1 (Stephens & Scheet 2005; Stephens et al. 2001). Input files were produced with the online software SeqPHASE (Flot 2010). Haplotypes defined from individuals with heterozygous insertion-deletions, following Flot et al. (2006), were incorporated in the analysis in order to improve phase determination (Stephens et al. 2001). Five replicate runs of 1000 iterations after an initial burn-in of 1000 generations were performed, with a thinning interval of 1, and the run with the best average goodness of fit was retained. Since PHASE has been shown to generate a very low number of false positives (Garrick et al. 2010), the complete dataset including some low-probability calls was kept to avoid biasing levels of diversity and the frequency spectra of mutations. Sequence alignments were reduced to the largest non-recombining blocks using IMgc (Woerner et al. 2007).

Finally, we assessed conformation of the multilocus variation to neutral expectations using the HKA test (Hudson et al. 1987) as implemented in the software HKA (http://genfaculty.rutgers.edu/hey/software#HKA) and using both the rabbit or each of the other hare species as outgroup.

Phylogenetic and Species Delimitation Analysis

To estimate phylogenies of the individual nuclear loci, the European rabbit was used as the outgroup, while for the cytochrome b phylogeny both the European rabbit and the eastern cottontail (*Sylvilagus floridanus*) were used (GenBank Acc. Nrs. in Annex I - Table S2.1). The best-fit model of sequence evolution for each sequenced locus was determined among 88 possible models using jModelTest v0.1.1 (Guindon & Gascuel 2003; Posada 2008) under the Akaike Information Criterion with correction (AICc).
Maximum-Likelihood (ML) and Bayesian Inference (BI) phylogenies were estimated for each nuclear locus using Garli v2.0 (Zwickl 2006) and BEAST v1.7.4 (Drummond et al. 2012) respectively, using European rabbit sequences as outgroup. For Garli, five

replicate runs of 1 million generations were performed using the best-fit mutation model and without fixing the model parameters. For BEAST, three independent runs of 50 million generations were performed, applying the best-fit mutation model or the next-most complex model implemented in the software, a Yule tree prior and an uncorrelated lognormal relaxed clock (Drummond et al. 2006). Runs were examined in Tracer v1.5 (Rambaut & Drummond 2007) and concatenated using LogCombiner, and post-burn-in trees were summarized using TreeAnnotator, part of the BEAST package. For the cytochrome b, both the European rabbit and the eastern cottontail were used as outgroups and similar phylogeny reconstruction analyses were conducted, but running 250 million generations for the BI and 5 million generations and 500 bootstrap replicates for the ML estimate.

Given the stochasticity of the coalescent process, methods that explicitly take into account the possibility of differential lineage sorting across individual loci are expected to perform better in multilocus datasets (Edwards et al. 2007; Kubatko & Degnan 2007). We therefore used the multilocus/multispecies Bayesian inference method *BEAST (Heled & Drummond 2010), as implemented in software BEAST v1.7.4 (Drummond et al. 2012), to infer the phylogeny of the three focal North American *Lepus* species based on the eight nuclear loci. Two strategies of species assignation were used: i) specimens were assigned to the three sampled species, and ii) *L. americanus* specimens were split into three units that correspond to the three population clusters described by Cheng et al. (2014). Given that this method estimates the root of each single-gene tree and uses the multispecies coalescent of the species tree (Heled & Drummond 2010), outgroup sequences were not included. Model choice and post-run examination followed the previously described BEAST analyses but, in this case, three independent *BEAST runs of 500 million generations were performed. The substitution rates of the multiple loci were estimated relative to PRKCI, and the rate for this locus was calibrated using the *Lepus*-*Oryctolagus* uncorrected genetic distance, considering a split time of 11.8 Myr (Matthee et al. 2004).

In order to assess whether the three *L. americanus* population clusters described by Cheng et al. (2014) based on microsatellite data reflect long-term sequence evolution, we performed a Bayesian species delimitation analysis using the nuclear data as implemented in the software BP&P v2.0 (Rannala & Yang 2003; Yang & Rannala 2010). The posterior probability of different possible taxa delimitation models was estimated by collapsing nodes of the species tree considering the three separate population clusters of *L. americanus* (assignation strategy ii described above). Different combinations of

ancestral effective population size (θ) and root age (τ0) priors were used (Yang & Rannala 2010) (see Annex I - Table S2.4). Two runs of 2 500 000 generations were performed. These analyses were also performed randomizing the assignment of the sequences to groups to assess the robustness of inferences.

Isolation-with-Migration Analysis

Given that the multi-species multi-locus phylogeny reconstruction method used here relies on the assumption that no introgression occurred between species (Heled & Drummond 2010), we attempted to quantify gene flow levels regardless of the inferred phylogeny by applying the isolation-with-migration (IM) model implemented in IMa2 (Hey 2010) to pairs of species and/or populations. Three independent runs were performed, varying the parameters' upper bound priors and the starting seeds and using the HKY mutation model (Hasegawa et al. 1985). Significance of gene flow estimates was assessed using Nielsen and Wakeley (2001) approach and also the likelihood-ratio tests of different models implemented in IMa2's L mode. Substitution rates (per generation) were estimated from the *Lepus*-*Oryctolagus* uncorrected genetic distance, considering a split time of 11.8 Myr (Matthee et al. 2004) and a generation time of two years (Marboutin & Peroux 1995).

Demographic analyses

The demographic history of the species was also investigated using the Extended Bayesian Skyline Plot (EBSP) analysis (Heled & Drummond 2008) using software BEAST v1.7.4. The EBSP analysis was performed for each species and for each *L. americanus* cluster (Cheng et al. 2014) separately. Since the *californicus*-like cytochrome b sequences of *L. americanus* from the Pacific Northwest may be the result of introgression they were not included in this analysis. Three independent runs of 200 million generations were performed using the best-fit mutation model selected with jModelTest or the next-most complex model implemented in the program. Tracer v1.5 was used to evaluate the combined runs and EBSPs were plotted using the GraphfromCSV python script provided with BEAST package v1.6.4. The mtDNA substitution rate, estimated from the *Lepus*-*Oryctolagus* average corrected distance and considering a divergence time of 11.8 Myr (Matthee et al. 2004), was used to calibrate the demographic plots.

Coalescent Simulations

We followed a methodology similar to that used by Melo-Ferreira et al. (2012) to understand the contribution of incomplete lineage sorting and introgression to the mtDNA phylogeny. Divergence time and population size estimates obtained between the Pacific Northwest population of *L. americanus* (the one possessing the discordant mtDNA lineage) and *L. californicus* under the IM model were used as input for SimCoal2 v2.1.2 (Laval & Excoffier 2004) to simulate 10 000 cytochrome b datasets mimicking the empirical dataset. Alternatively, the IM parameter values inferred considering *L. americanus* as a single population were also tested. A model where an ancestral haploid population of size NeA/2 splits into two descendant populations of sizes Ne1/2 and Ne2/2, t generations ago, no gene flow occurring between the two descendant populations, was applied. An unequal transition-transversion rate was considered (estimated in jModelTest) and the mtDNA substitution rate per generation was again estimated from the *Lepus*-*Oryctolagus* average corrected distance. The minimum pairwise corrected p-distance between the descendent populations was retained for each replicate. The empirical p-distance was considered to reject the incomplete lineage sorting hypothesis if found to be lower than the 5th percentile of the simulated distribution of minimum distances. This analysis was also performed using the 95% HPD bounds of the IM estimates that maximize incomplete lineage sorting (lower bound of divergence time and upper bounds of effective population sizes).

4. Results

Sequence data and Phylogenetic Inferences

Eight nuclear markers and one mitochondrial DNA locus were sequenced in this study, for a total of 6 184 bp of nuclear DNA and 580 bp of mtDNA (Table 2.2). Limiting the analyses to the largest non-recombining blocks, the nuclear dataset was reduced to 5 660 bp (5 366 bp with the inclusion of *O. cuniculus* as outgroup) (Table 2.2). The HKA test did not detect deviations from neutral expectations (P > 0.05).

Table 2.2 Loci included in this study, length of obtained sequences and inferred mutation models.

			Number of characters				Non-		Mutation
Locus			Total	LNRB[1]	Out.[2]	Variable[3]	coding[4]	Exon[4]	Model[5]
1	SPTBN1	Spectrin, beta, non-erythrocytic 1	636[6]	561	561	26	1-561	-	K80
2	PRKCI	Protein kinase C, iota	436	432	426	36	10-432	1-9	HKY
3	DARC	Duffy blood group, chemokine receptor	783	741	741	26	-	1-741	TPM2uf+Γ
4	KITLG	KIT ligand	552	461	461	23	1-461	-	JC
5	TF	Transferrin	387	316	320	29	1-316	-	JC
6	POLA1	Polymerase, alpha 1, catalytic subunit	813	572	572	34	1-572	-	F81+Γ
7	GRIA3	Glutamate receptor, ionotrophic, AMPA 3	969[6]	969	677	41	1-969	-	TrN
8	SRY	Sex determining region of the Y chromosome	1608	1608	1608	40	1-220; 836-1608	221-835	TIM2
Total nuclear DNA			6184	5660	5366	255			-
9	CYTB	Cytochrome *b*	580	580	580	127	-	580	TPM3uf+Γ
Total			6764	6240	5946	382			-

[1]Largest non-recombining blocks; [2]alignment including outgroup; [3]only ingroup taxa were considered; [4]coordinates of the LNRB alignment; [5]see Posada (2008) for a description of models and references; [6]microsatellites and buffer regions, two in GRIA3 (34 bp; 16 bp) and one in SPTBN1 (19 bp), not considered (see Material and Methods).

The Maximum Likelihood (ML) and Bayesian Inference (BI) phylogenetic reconstructions showed extensive sequence sharing among species (Annex I - Figures S2.1 and S2.2). The multilocus nuclear phylogeny resulting from *BEAST suggests that *L. californicus* and *L. townsendii* are more closely related than either is to *L. americanus*, which is consistent across the replicate runs (Annex I - Figure S2.3). Additionally, the BP&P species delimitation analyses demonstrated high support for the topology considering the three *L. americanus* clusters separately (posterior probability >0.99; Annex I - Table S2.4) but not when sequences were randomly assigned to clusters (not

shown). The *BEAST analysis taking these three *L. americanus* clusters into account recovered a monophyletic *L. americanus* (posterior probability >0.99) (Figure 2.2).

Our divergence estimates suggest that *L. americanus* split from the other two species around 2.7 Mya, while *L. townsendii* and *L. californicus* diverged at 1.2 Mya (Figure 2.2). Within *L. americanus*, the Boreal group was estimated to have diverged 2.0 Mya, while the Rockies and Pacific Northwest groups split much more recently, at about 0.36 Mya (Figure 2.2).

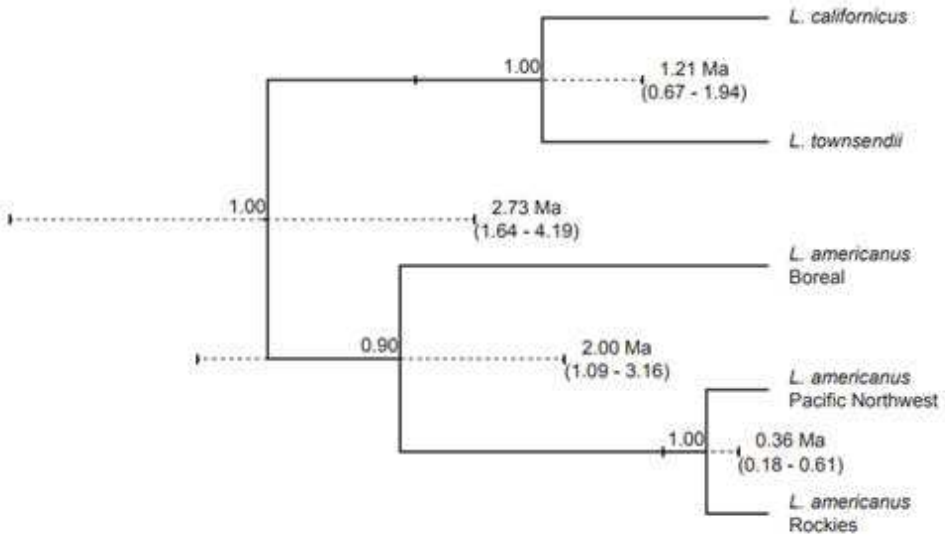


Figure 2.2 Species tree of *L. californicus*, *L. townsendii*, and *L. americanus* inferred with *BEAST, considering the partition of the latter into three discrete populations (Boreal, Rockies and Pacific Northwest). Numbers above branches indicate the posterior probabilities and dashed line the 95% confidence intervals of node ages (mean value and 95% CI are indicated next to the line). The tree was calibrated using a substitution rate of 1.65 x 10-9 substitutions/site/year for the PRKCI fragment.

The mtDNA phylogeny does not conform to that of the nuclear DNA, since *L. americanus* is not recovered as monophyletic (Figure 2.3) given that one group from the Pacific Northwest population cluster (here named PacNW2) is more closely related to *L. californicus*. The remaining *L. americanus* form a distinct clade, classified as Boreal, Rockies, and Pacific Northwest (PacNW1) based on the microsatellite groups of Cheng et al. (2014). The discordant clade (PacNW2) does not include *L. californicus* haplotypes. This result agrees with the observations of Cheng et al. (2014), suggesting either the occurrence of mtDNA introgression from *L. californicus* into *L. americanus* or incomplete mtDNA lineage sorting in the evolution of these species.

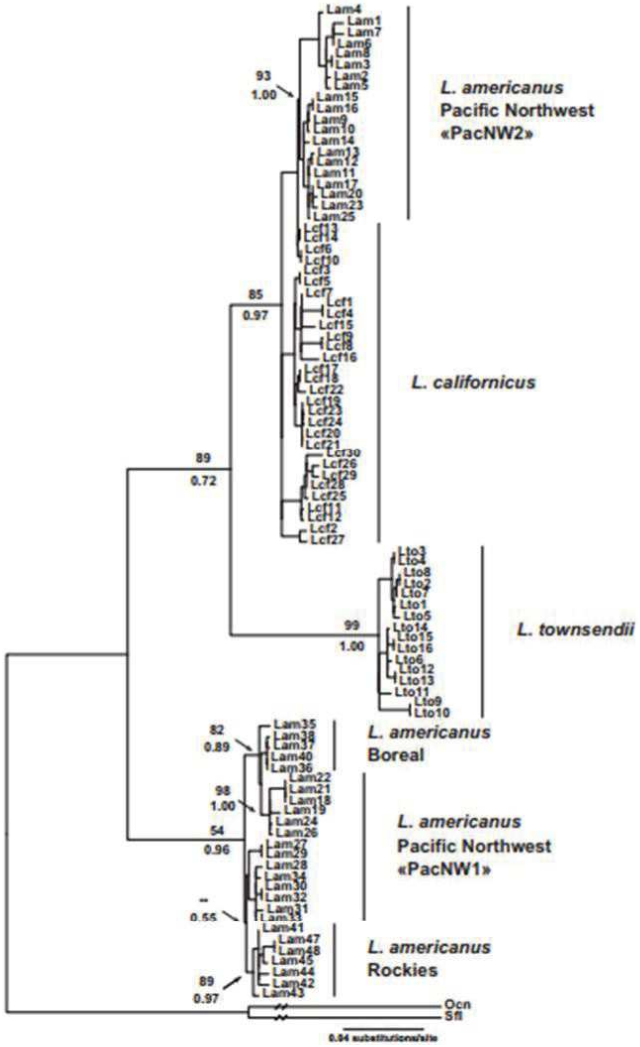


Figure 2.3 Cytochrome b Bayesian inference phylogeny of *L. californicus*, *L. townsendii* and *L. americanus*. Sequences from Oryctolagus cuniculus (Ocn) and Sylvilagus floridanus (Sfl) were used as outgroups. Maximum likelihood bootstrap supports and BI posterior probabilities of the most relevant clades are shown above and below branches, respectively (if bootstrap support was higher than 50% or posterior probability higher than 0.5). See specimen codes and GenBank accession numbers in Annex I - Table S2.1.

Isolation-with-migration and demographic analyses

IMa2 was used to quantify gene flow among species/clusters (see parameter estimates in Table 2.3; Annex I - Tables S2.5 to S2.7). Among species, gene flow was only significant from *L. americanus* into *L. townsendii* and *L. californicus* but at very low levels unlikely to affect phylogenetic reconstruction (Eckert & Carstens 2008). In general, parameter estimates did not differ much when considering *L. americanus* as a single

evolutionary unit (Annex I - Table S2.5). Within *L. americanus*, the three population clusters were found to be remarkably isolated. Nuclear gene flow was only suggested as significant from the Boreal into the Pacific Northwest population and from this to the Rockies population but, again, at very low levels. Estimates of divergence among populations according to the IM model (Table 2.3) were consistent with those inferred with *BEAST (Figure 2.2). *L. californicus* was the species suggested to have the highest effective population size (Ne) and *L. townsendii* the lowest. However, the Pacific Northwest and Rockies populations of *L. americanus* have the lowest Ne across all analyzed populations (Table 2.3).

Table 2.3 ML estimates (95% posterior density intervals in parentheses) of demographic parameters obtained with IMa2 between pairs of populations.

Pop. 1	Pop. 2	N_{e1}[1]	N_{e2}[1]	N_{eA}[1]	t[2]	$2Nm_1$[3]	$2Nm_2$[3]	ABCDD[4]	ABC0D[4]	ABCD0[4]	ABC00[4]
Lam-Bor	Lam-Roc	226 647	38 216	-	2 007 684	-	-	n.s.	n.s.	n.s.	n.s.
		(127 346; 404 666)	(15 347; 83 353)	(0; -)	(986 991; -)	(0.0000; 0.3085)	(0.0000; -)				
Lam-Bor	Lam-PacNW	225 443	77 936	-	2 421 739	-	0.0176*	n.s.	n.s.	*	*
		(128 429; 394 435)	(47 845; 120 605)	(0; -)	(948 474; -)	(0.0000; 0.2761)	(0.0003; 0.1591)				
Lam-Roc	Lam-PacNW	22 111	69 968	196 315	-	0.0904*	-	n.s.	*	n.s.	*
		(7 673; 50 108)	(39 263; 114 190)	(0; -)	-	(0; -)	(0.0000; 0.3504)				
Lam-Bor	Lca	265 404	574 742	207 629	2 972 528	-	-	n.s.	n.s.	n.s.	*
		(152 261; 455 581)	(442 340; 739 641)	-	(1 240 238; -)	(0.0000; 0.6103)	(0.0000; 0.2688)				
Lam-Roc	Lca	60 038	582 686	363 381	3 429 914	-	-	n.s.	n.s.	n.s.	n.s.
		(26637; 117 813)	(453 655; 744 336)	-	(1 452 080; -)	(0.0000; 0.1759)	(0.0000; 0.1036)				
Lam-PacNW	Lca	110 579	597 129	309 217	2 520 919	-	-	n.s.	n.s.	n.s.	n.s.
		(69 968; 164 779)	(465 330; 763 233)	(4 062; 747 946)	(1 537 780; 4 311 946)	(0.0000; 0.0934)	(0.0000; 0.1094)				
Lam-Bor	Lto	320 050	175 612	-	-	-	0.0301*	n.s.	n.s.	*	*
		(179 223; 568 242)	(115 105; 262 275)	-	-	(0.0000; 0.5460)	(0.0000; 0.2029)				
Lam-Roc	Lto	37 470	174 649	-	-	0.0068	0.0263*	n.s.	n.s.	*	*
		(148 89; 78 093)	(114 190; 260 469)	-	-	(0.0000; 0.0629)	(0.0018; 0.1421)				
Lam-PacNW	Lto	78 983	181 871	-	-	-	0.0312*	n.s.	n.s.	*	*
		(47 388; 123 253)	(120 485; 268 534)	-	-	(0.0000; 0.0877)	(0.0000; 0.1687)				
Lca	Lto	641 424	228 813	264 923	1 357 714	0.0033	0.0012	n.s.	n.s.	n.s.	n.s.
		(491 570; 830 998)	(152 984; 334 494)	(91 622; 550 187)	(856 997; 2 166 565)	(0.0000; 0.4184)	(0.0000; 0.2241)				

Lam-Bor: *L. americanus*, Boreal; Lam-Roc: *L. americanus*, Rockies; Lam-PacNW: *L. americanus,* Pacific Northwest; Lca: *L. californicus*; Lto – *L. townsendii*; Missing values correspond to cases where parameters could not be reliably estimated; a substitution rate of 3.45×10^{-9} substitutions/site/generation was estimated. [1]Effective population size of population 1 (N_{e1}), 2 (N_{e2}), and ancestral population (N_{eA}); [2]Time in years since populations 1 and 2 split; [3]Population migration rate into population 1 ($2Nm_1$) and population 2 ($2Nm_2$) (*significant values, $P < 0.05$; Nielsen & Wakeley 2001). [4]Likelihood ratio test of nested models with equal gene flow between populations (ABCDD), no gene flow into population 1 (ABC0D), no gene flow into population 2 (ABCD0), and with no gene flow (ABC00). The test statistic was calculated as follows: ABCDD (2LLR against ABCDE) follows a chi-square distribution with 1 degree of freedom with critical value*$P < 0.05$ at 2LLR > 3.84; ABC0D and ABCD0 (2LLR against ABCDE) and ABC00 (2LLR against ABCDD) follow a chi-square distribution that is 1/2 × chi-square(1) + 1/2 × chi-square(0) with critical value *$P < 0.05$ at 2LLR > 2.70.

The Extended Bayesian Skyline Plot analysis does not suggest any drastic shifts in population size for the species and intraspecific clusters analyzed here, particularly if the large confidence intervals of the inference are taken into account (Figure 2.4).

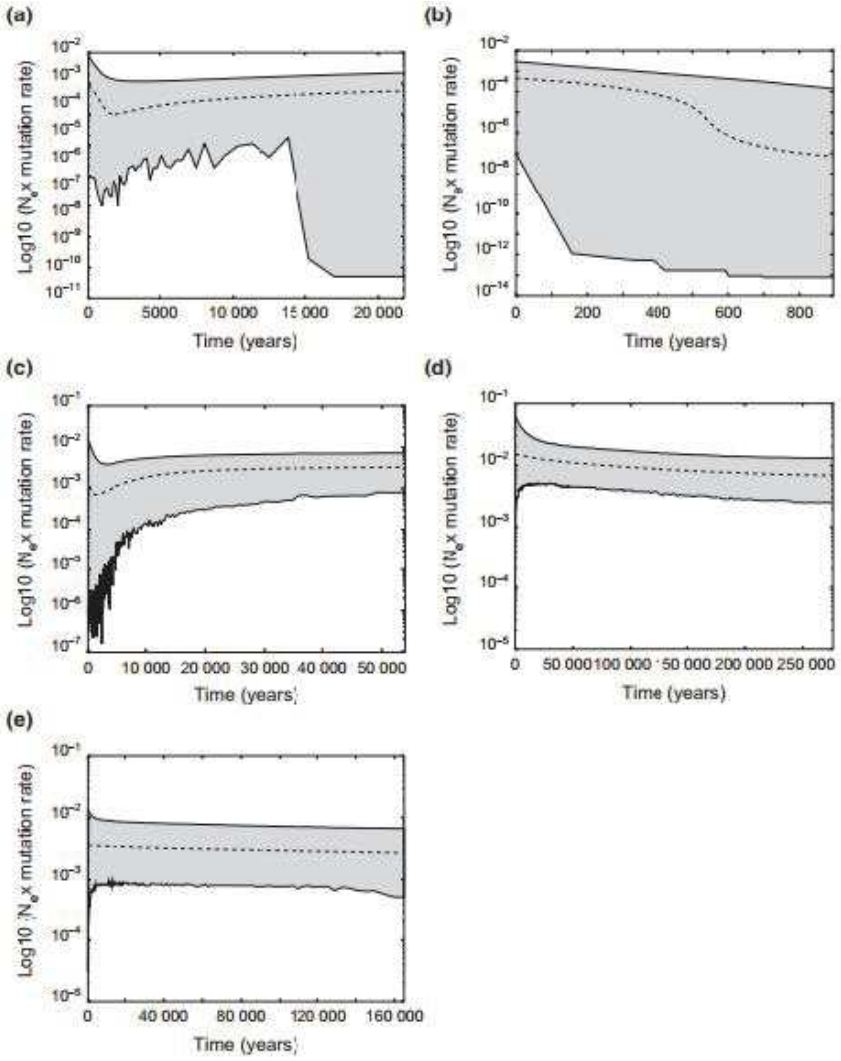


Figure 2.4 Demographic profiles of *L. americanus* Boreal (a), *L. americanus* Rockies (b), *L. americanus* Pacific Northwest (c), *L. californicus* (d) and *L. townsendii* (e), based on Extended Bayesian Skyline Plot analyses. The last 10% of the time points are not shown, except for plot b (see full plots in Annex I - Figure S2.5). Time is in units of years before the present (calibrated using a mtDNA substitution rate of 1.8 x 10-8 substitutions/site/year).

Coalescent Simulations

The effective population sizes and divergence times obtained from IM analysis of the two population clusters possibly involved in the mtDNA introgression events – *L. americanus* Pacific Northwest and *L. californicus* – were used to simulate cytochrome b

datasets under a model with no gene flow. The observed pairwise distances between *L. americanus* PacNW1 cluster and *L. californicus* were found to lie within the range of distances expected under a strict lineage sorting scenario (the same was found for the Boreal, and Rockies groups). On the contrary, all pairwise distances between *L. americanus* PacNW2 and *L. californicus* fell below the 5th percentile of the minimum distances simulated assuming no gene flow, suggesting introgression (Figure 2.5). The same results were obtained maximizing the probability of retention of ancestral polymorphism (Annex I - Figure S2.6). The geographic distribution of mtDNA introgression is shown in Figure 2.1.

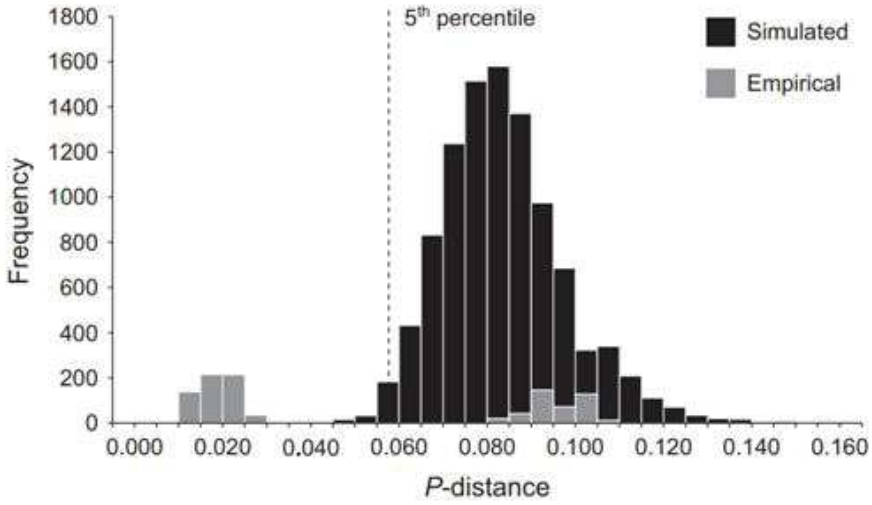


Figure 2.5 Empirical (grey bars) and simulated (black bars) mtDNA distances between *L. californicus* and the Pacific Northwest cluster of *L. americanus*. Simulations were performed under the assumption of no gene flow and a cytochrome b substitution rate of 3.6 x 10-8 substitutions/site/generation. Vertical line indicates the 5th percentile of the distribution of simulated distances.

5. Discussion

Speciation history of North American hares suggests cryptic divergence

Understanding the relative importance of introgression in the evolution of North American hares requires estimating the most relevant parameters of their history of speciation. The *BEAST phylogeny suggest that *L. americanus* diverged from the common ancestor of the three focal species at around 2.7 Mya and that the jackrabbits, *L. californicus* and *L. townsendii*, diverged 1.2 Mya (Figure 2.2). These estimates are generally consistent with those obtained from the IM analyses (2.4-3.1 and 1.4 Mya respectively; Table 2.3) and are more recent than previous estimates based on a molecular supermatrix (4.8 Mya for the stem divergence of *L. americanus*; Matthee et al. 2004) or mtDNA (5.6 Mya for the stem divergence of *L. americanus*; Wu et al. 2005). Interestingly, our analysis suggests that the Pacific Northwest and Rockies populations of *L. americanus* may have diverged ~360 kya, which is consistent with the fragmentation of the western forest of the Pacific coast and Rocky Mountains (see Weir & Schluter 2004); however, the Boreal population diverged from the other two at a deeper evolutionary timescale (2.0 Mya; Figure 2.2). This estimate roughly places the event of fragmentation and divergence in the same period of the split between *L. townsendii* and *L. californicus*, which conforms to the presumed timeframe of speciation events in North American mammals (Arbogast & Kenagy 2001; Demboski & Cook 2001) and birds (Weir & Schluter 2004), and may thus have resulted from common environment-driven fragmentation pressures (Weir & Schluter 2004). The unexpected depth of the snowshoe hare's intraspecific divergence suggests that genetic isolation among groups arose from historic processes and not from recent geographic fragmentation. In addition, the extremely limited levels of gene flow inferred between the three *L. americanus* genetic groups using the IM framework (Table 2.3) suggest that some degree of reproductive isolation may exist. Interestingly, Nagorsen (1985) found no indication of morphological distinctiveness of the Boreal snowshoe hares or conformation to subspecific classifications, and it thus appears that we detected cryptic divergence within *L. americanus*. Although it would be useful to perform coalescent-based analyses with extended sampling of the Boreal group to confirm levels of divergence, we note that Cheng et al. (2014) showed that genetic variation within the Boreal group is homogeneous and thus our sampling may adequately describe genetic variation in that group. Whether or not the inferred divergence and low levels of gene flow justify a taxonomic revision of *L. americanus* must be addressed with an integrative analysis including data from multiple genetic and non-genetic sources.

Demographic history

Although *L. americanus* has the largest distribution among the North American hares, *L. californicus* has the largest effective population size among the three species (Table 2.3), which could reflect different evolutionary histories. Northern *L. americanus* is likely to have been more susceptible to demographic and geographical oscillations due to the repeated advance and retreat of glaciers throughout the Pleistocene (Hewitt 2004). The lower estimated effective population sizes of the Pacific Northwest and Rockies clusters may reflect an increased susceptibility of peripheral populations to demographic fluctuations (Cheng et al. 2014; see Eckert et al. 2008 for a review on the central-marginal hypothesis) (Table 2.3; Figure 2.4). The large distribution of *L. californicus* may have been less affected by climatic oscillations, allowing the species to maintain larger population sizes (Figure 2.4). Our Extended Bayesian Skyline Plot does not suggest changes in population sizes through time for any species or cluster. This may be due to the relatively small sample size in this work for such inferences, particularly in the Boreal snowshoe hare group. Indeed, using only mtDNA but a larger sample size, Cheng et al. (2014) inferred a late Pleistocene demographic expansion of the Boreal group.

Little is known about the population history of *L. townsendii*. Our results suggest that this species has the lowest long-term effective population size among the three studied species (Table 2.3), but no dramatic shift of population size through time were inferred (Figure 2.4). However, fossils suggest that over the past few thousand years this species may have been excluded from some southern regions due to competitive exclusion by *L. californicus* (Lim 1987 and references therein). In addition, *L. townsendii* may have disappeared from some areas due to land use and habitat fragmentation (Berger 2008; but see Gunther et al. 2009). Our analysis suggests that gene flow from *L. americanus* into *L. townsendii* has occurred since the divergence of these species (Table 2.3), although it is difficult to assess whether this corresponds to recent introgression in populations of *L. townsendii*. No evidence of gene flow was found from this species to/from *L. californicus* contrary to the suggestion that these species hybridize in nature (Flux 1983).

Extensive mtDNA introgression from L. californicus into L. americanus

Even though nuclear gene flow among the three North American species seems rare or absent (Table 2.3), previous results of Cheng et al. (2014) suggested that

mitochondrial DNA introgression might have occurred between *L. californicus* and *L. americanus*, considering the sharing of mtDNA linages visible in the mtDNA phylogeny (seen also in this work; Figure 2.3). This contrasts with the monophyly of *L. americanus* that we estimate for nuclear DNA (Figure 2.2). Our coalescent simulations show that the genetic similarity between the PacNW2 mtDNA haplotypes of *L. americanus* and *L. californicus* is incompatible with simple incomplete lineage sorting (contrary to the remaining divergences to *L. californicus*, which are within expectations: PacNW1 DXY = 0.096; Boreal DXY = 0.101; Rockies DXY = 0.098; see Figure 2.5). However, *L. californicus* and *L. americanus* PacNW2 do not share mtDNA haplotypes, which could result from i) ancient introgression, ii) introgression of an extant but unsampled *L. californicus* haplogroup, or iii) introgression from another species not included in this study. We aligned all cytochrome b haplotypes of the three species included in this study and other species available at GenBank to our dataset (Annex I - Figure S2.4). The position of the PacNW2 clade is maintained closer to *L. californicus* in this extended phylogeny suggesting that introgression was likely ancient and of *L. californicus* origin. We estimated that the split between *L. californicus* and PacNW2 mtDNA occurred 470 000 years ago (200 000-906 000 95% HPD), which can thus indicate the time of introgression.

Historical and ongoing gene introgression has been found among other North American mammals (e.g. Chavez et al. 2011; Good et al. 2008), sometimes with massive mtDNA introgression or 'capture' (Good et al. 2008) and little nuclear DNA introgression as found in this work. This may have resulted from the competitive replacement of resident *L. californicus* by invading *L. americanus* during the Pleistocene glaciations, a situation that is expected to lead to introgression into the genome of the invading species (Currat et al. 2008; Excoffier & Ray 2008). These two species have different habitat requirements, *L. americanus* inhabiting for example dense boreal forest and *L. californicus* being distributed in southern open arid regions, and glacial cycles would have differentially shifted these distinct habitats. This competitive replacement model predicts that introgression should prevail for markers transmitted by the least dispersing sex, which is often females in mammals. However, whether this explains massive introgression of mtDNA into *L. americanus* is at this point uncertain. In addition, there is no evidence of sex-biased dispersal in this species (Burton et al. 2002). The asymmetric direction of introgression would also be favored by mechanisms that induce sex-biased matings, such as female choice and frequency-dependent assortative matings, among others (Chan & Levin 2005; Wirtz 1999). Alternatively, mtDNA introgression into *L.*

americanus may have been favored by natural selection. Adaptive introgression of mtDNA has been hypothesized in several species (e.g. Ropiquet & Hassanin 2006; Ruiz-Pesini et al. 2004), including in hares (Melo-Ferreira et al. 2011, 2014), and could explain the pattern observed here if the *L. californicus* mtDNA type is advantageous in the *L. americanus* nuclear background. However, separating the relative contributions of selective and demographic processes to interspecific gene flow is a major challenge and should be the object of future research.

Conclusions and Future Prospects

Our results uncover hidden evolutionary processes in the North American hares: i) deep cryptic divergence exists within *L. americanus*, ii) nuclear gene flow occurred from *L. americanus* into *L. townsendii* and *L. californicus*, and iii) extensive mtDNA introgression occurred from *L. californicus* into the Pacific Northwest populations of *L. americanus*. Introgression is a source of genetic novelty and may set the conditions for adaptation if the introgressed variants underlie favored phenotypes (reviewed by Arnold & Martin 2009). For example, introgression has been shown to enhance abiotic tolerance in sunflowers (Whitney et al. 2010), induce poison resistance in mice (Song et al. 2011), and to generate adaptive wing color variation in butterflies (Pardo-Diaz et al. 2012). It is striking that contrary to the general trend in *L. americanus*, which undergoes seasonal coat color changes from a brown coat in the summer to a white winter coat, part of the Pacific Northwest group retain their summer coat year-round, mimicking the phenotype of *L. californicus*. The dramatic snow pack decrease caused by global warming and the increased tendency of seasonally changing hares to become more mismatched against a snow-free background (Mills et al. 2013) may confer a significant adaptive advantage to the trait present in the Pacific Northwest populations. Although other evolutionary mechanisms can underlie this phenotype, hybridization may have contributed to the retention of the summer coat year-round if introgression affected genomic regions involved in seasonal coat-color change. Although speculative at present, this hypothesis opens new perspectives in the study of the impact of global warming to the survival of boreal species undergoing seasonal coat color change and deserves further investigation.

6. References

Abbott R, Albach D, Ansell S, et al. (2013) Hybridization and speciation. Journal of Evolutionary Biology 26, 229-246.

Alves PC, Ferrand N, Suchentrunk F, Harris DJ (2003) Ancient introgression of *Lepus timidus* mtDNA into *L. granatensis* and *L. europaeus* in the Iberian Peninsula. Molecular Phylogenetics and Evolution 27, 70-80.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P (2008) The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. Philosophical Transactions of the Royal Society B 363, 2831-2839.

Arbogast BS, Kenagy GJ (2001) Comparative phylogeography as an integrative approach to historical biogeography. Journal of Biogeography 28, 819-825.

Arnold ML, Martin NH (2009) Adaptation by introgression. Journal of Biology 8, 82.

Berger J (2008) Undetected species losses, food webs, and ecological baselines: a cautionary tale from the Greater Yellowstone Ecosystem, USA. Oryx 42, 139-142.

Boonstra R (2013) Reality as the leading cause of stress: rethinking the impact of chronic stress in nature. Functional Ecology 27, 11-23.

Bossu CM, Near TJ (2009) Gene Trees Reveal Repeated Instances of Mitochondrial DNA Introgression in Orangethroat Darters (Percidae: Etheostoma). Systematic Biology 58, 114-129.

Burton C, Krebs CJ, Taylor EB (2002) Population genetic structure of the cyclic snowshoe hare (*Lepus americanus*) in southwestern Yukon, Canada. Molecular Ecology 11, 1689-1701.

Chan KM, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution 59, 720-729.

Chavez AS, Saltzberg CJ, Kenagy GJ (2011) Genetic and phenotypic variation across a hybrid zone between ecologically divergent tree squirrels (Tamiasciurus). Molecular Ecology 20, 3350-3366.

Cheng E, Hodges K, Melo-Ferreira J, Alves P, Mills LS (2014) Conservation implications of the evolutionary history and genetic diversity hotspots of the snowshoe hare. Molecular Ecology 23, 2929-2942.

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

Demboski JR, Cook JA (2001) Phylogeography of the dusky shrew, Sorex monticolus (Insectivora, Soricidae): insight into deep and shallow history in northwestern North America. Molecular Ecology 10, 1227-1240.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PloS Biology 4, e88.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29, 1969-1973.

Eckert AJ, Carstens BC (2008) Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Molecular Phylogenetics and Evolution 49, 832-842.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Molecular Ecology 17, 1170-1188.

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

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

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. Trends in Ecology & Evolution 23, 347-351.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. Trends in Genetics 28, 342-350.

Flot JF (2010) seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. Molecular Ecology Resources 10, 162-166.

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

Flux JEC (1983) Introduction to taxonomic problems in hares. Acta Zoologica Fennica 174, 7-10.

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

Gerard D, Gibbs HL, Kubatko L (2011) Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. BMC Evolutionary Biology 11, 291.

Good JM, Hird S, Reid N, et al. (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. Molecular Ecology 17, 1313-1327.

Griffin PC, Mills LS (2009) Sinks without borders: snowshoe hare dynamics in a complex landscape. Oikos 118, 1487-1498.

Guindon S, Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Systematic Biology 52, 696-704.

Gunther KA, Renkin RA, Halfpenny JC, et al. (2009) Presence and Distribution of White-tailed Jackrabbits in Yellowstone National Park. Yellowstone Science 17, 24-32.

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

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22, 160-174.

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

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

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London B-Biological Sciences 359, 183-195.

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

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

Krebs CJ (2011) Of lemmings and snowshoe hares: the ecology of northern Canada. Proceedings of the Royal Society B-Biological Sciences 278, 481-489.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56, 17-24.

Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. Bioinformatics 20, 2485-2487.

Lewis CW, Hodges KE, Koehler GM, Mills LS (2011) Influence of stand and landscape features on snowshoe hare abundance in fragmented forests. Journal of Mammalogy 92, 561-567.

Lim (1987) *Lepus townsendii*. Mammalian Species 288, 1-6.

Liu J, Yu L., Arnold ML, et al. (2011) Reticulate evolution: frequent introgressive hybridization among Chinese hares (genus *Lepus*) revealed by analyses of multiple mitochondrial and nuclear DNA loci. BMC Evolutionary Biology 11, 223.

Mallet J (2005) Hybridization as an invasion of the genome. Trends in Ecology & Evolution 20, 229-237.

Marboutin E, Peroux R (1995) Survival pattern of European hare in a decreasing population. Journal of Applied Ecology 32, 809-816.

Matthee CA, van Vuuren BJ, Bell D, Robinson TJ (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.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P (2009) The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. Molecular Ecology 18, 2643–2658.

Melo-Ferreira J, Alves PC, Rocha J, Ferrand N, Boursot P (2011) Interspecific X-Chromosome and Mitochondrial DNA Introgression in the Iberian hare: Selection or Allele Surfing? Evolution 65, 1956-1968.

Melo-Ferreira J, Boursot P, Carneiro M, et al. (2012) Recurrent Introgression of Mitochondrial DNA Among Hares (*Lepus* spp.) revealed by Species-tree Inference and Coalescent Simulations. Systematic Biology 61, 367-381.

Melo-Ferreira J, Vilela J, Fonseca MM, et al. (2014) The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. Genome Biology and Evolution 6, 886-896.

Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theoretical Population Biology 75, 35-45.

Mills LS, Zimova M, Oyler J, et al. (2013) Camouflage mismatch in seasonal coat color due to decreased snow duration. Proceedings of the National Academy of Sciences of the United States of America 110, 7360-7365.

Nagorsen DW (1985) A morphometric study of geographic variation in the snowshoe hare (*Lepus americanus*). Canadian Journal of Zoology 63, 567-579.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. Genetics 158, 885-896.

Pardo-Diaz C, Salazar C, Baxter SW, et al. (2012) Adaptive introgression across species boundaries in Heliconius butterflies. PLoS Genet 8, e1002752.

Pinho C, Hey J (2010) Divergence with Gene Flow: Models and Data. Annual Review in Ecology, Evolution and Systematics 41, 215–230.

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

Rambaut A, Drummond AJ (2007) Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164, 1645-1656.

Ropiquet A, Hassanin A (2006) Hybrid origin of the Pliocene ancestor of wild goats. Molecular Phylogenetics and Evolution 41, 395-404.

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. Science 303, 223-226.

Seehausen O (2004) Hybridization and adaptive radiation. Trends in Ecology & Evolution 19, 198-207.

Seehausen O (2013) Conditions when hybridization might predispose populations for adaptive radiation. Journal of Evolutionary Biology 26, 279-281.

Song Y, Endepols S, Klemann N, et al. (2011) Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice. Current Biology 21, 1296-1301.

Spinks PQ, Shaffer HB (2009) Conflicting Mitochondrial and Nuclear Phylogenies for the Widely Disjunct Emys (Testudines: Emydidae) Species Complex, and What They Tell Us about Biogeography and Hybridization. Systematic Biology 58, 1-20.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. American Journal of Human Genetics 76, 449-462.

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

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673-4680.

Thulin C-G, Stone J, Tegelström H, Walker CW (2006a) Species assignment and hybrid identification among Scandinavian hares *Lepus europaeus* and *L. timidus*. Wildlife Biology 12, 29-38.

Thulin CG, Fang M, Averianov AO (2006b) Introgression from *Lepus europaeus* to *L. timidus* in Russia revealed by mitochondrial single nucleotide polymorphisms and nuclear microsatellites. Hereditas 143, 68-76.

Tyson R, Haines S, Hodges KE (2010) Modelling the Canada lynx and snowshoe hare population cycle: the role of specialist predators. Theoretical Ecology 3, 97-111.

Wallner B, Huber S, Achmann R (2001) Non-invasive PCR sexing of rabbits (Oryctolagus cuniculus) and hares (*Lepus europaeus*). Mammalian Biology 66, 190-192.

Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. Proceedings of the Royal Society of London B-Biological Sciences 271, 1881-1887.

Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower Helianthus annuus. New Phytologist 187, 230-239.

Wirtz P (1999) Mother species-father species: unidirectional hybridization in animals with female choice. Animal Behaviour 58, 1-12.

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

Wu C, Wu J, Bunch TD, et al. (2005) Molecular phylogenetics and biogeography of *Lepus* in Eastern Asia based on mitochondrial DNA sequences. Molecular Phylogenetics and Evolution 37, 45-61.

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

Zimova M, Mills LS, Lukacs PM, Mitchell MS (2014) Snowshoe hares display limited phenotypic plasticity to mismatch in seasonal camouflage. Proceedings of the Royal Society B-Biological Sciences 281, 20140029.

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, University of Texas.

Chapter 3.

Genomic perspective of introgression in hares from Iberia

Paper II. Seixas FA, Boursot P, Melo-Ferreira J (2017) **The genomic impact of historical hybridization with massive mitochondrial DNA introgression in the Iberian hare (*Lepus granatensis*).** *Submitted*

Paper III. Seixas FA, Farelo L, Belkir K, Alves PC, Boursot P, Melo-Ferreira J (2017) **Genomic exchanges between three hare species sharing the same mitochondrial genome following massive introgression: the roles of history, adaptation and cytonuclear coevolution.** *In preparation*

The genomic impact of historical hybridization with massive mitochondrial DNA introgression

Seixas, FA, Melo-Ferreira, J, Boursot P

1. Abstract

Studying detailed genome-wide patterns of genetic exchanges between species reveals the roles of genome structure, demography and selection in determining speciation, admixture and adaptation. We infer nuclear introgression that accompanied massive mitochondrial DNA introgression from *Lepus timidus* (mountain hare) into *L. granatensis* (Iberian hare) using 13 whole-genome sequences. Identity-by-state and introgression tract length distributions suggest hybridization occurred 7-24 kya ago. Introgression contributes up to 2.44% of individual *granatensis* genomes and underrepresentation on the X-chromosome reflects its disproportionate involvement in hybrid incompatibility. Introgression increases away from chromosome centers, revealing interplay between recombination and hybrid counter-selection at dispersed loci. We find no evidence for enhanced introgression of nuclear genes with mitochondrial functions, which could have indicated cytonuclear co-evolution. Average nuclear introgression occurs at low frequencies all over the species range in Iberia, with a gentle south-north increasing gradient, contrasting with the steep *timidus* mtDNA gradient, from absent to quasi-fixed. Using simulations, we find these geographic and frequency patterns compatible with a demographic model of south-north invasive replacement of *L. timidus* by *L. granatensis* after the last glacial maximum, with asymmetric crossing and male-biased migration. However, a group of genes displays outlying high introgression frequencies that appear selection-driven. Several concern innate immunity, suggesting adaptation to pathogenic environments. Others concern spermatogenesis, and could compensate harmful effects of *timidus* mtDNA on male-specific functions in a *granatensis* background. Although range invasion may determine broad patterns of introgression, nuclear introgression appears impeded by hybrid incompatibilities but enhanced by adaptation to the environment and possibly to mitochondrial introgression.

2. Introduction

Hybridization and genetic introgression between populations with some degree of reproductive isolation, and thus considered different species, is an important evolutionary phenomenon that is widespread in nature (Mallet 2005). Speciation research has come to appreciate that genomes remain permeable to gene flow well after the initiation of the speciation process (see e.g. Harrison and Larson 2014 for a review; Roux et al. 2016). An exciting and vivid line of research in this context is thus to understand the determinants of the amount of gene flow between species, and more interestingly of its variations along the genome, characteristic of the semi-permeability of many species boundaries (Muirhead and Presgraves 2016). Such ideas have been discussed for long and modelled (Barton and Hewitt 1984; Wu 2001), but the advent of the genomic era offers new powerful ways to address them empirically, which hold the promise to understand the genetic origin of reproductive isolation, and the role of natural selection in either preventing or promoting gene flow among species.

Introgression can be a major source of adaptive variation, in addition to standing variation and new mutation (Hedrick 2013; Tigano and Friesen 2016). Introgression of pre-tested genetic combinations may provide important advantages to prosper or invade some habitats (as suggested by e.g. Rieseberg et al. 2007; Quach et al. 2016) although it could also be non-adaptive if involving selfish genetic elements (e.g. Macholán et al. 2008; Albrechtova et al. 2012). Gathering empirical and statistical evidence for such phenomenon is challenging for two reasons. First, one must be able to disentangle the effects of introgression from those of incomplete lineage sorting (i.e. sharing of ancestral variation among daughter populations), which is expected to be pervasive between recently diverged taxa. Second, interpreting a pattern of introgression as adaptive based on its geographic and frequency pattern needs a comparison with a null expectation that depends on the complex and generally unknown historic and demographic conditions that determined the degree of stochasticity of the process of species admixture. For example, during invasion of the range of a species by another, with occasional hybridization, drift at the invasion front (in initially small founding populations) may bring variants introgressed from the resident species into the invading one to high frequencies, which can then propagate further and thus “surf” on the invasion wave (Currat et al. 2008; Excoffier et al. 2009). The occurrence and persistence of introgression occurring under such conditions is favored by high drift and low intraspecific migration rates. These two parameters can vary across genomic regions with different modes of sex-linked transmission if the two sexes have different migration rates. For instance, in species

where females are more philopatric than males, the female-transmitted mitochondrial DNA is expected to be most affected by massive introgression (Currat et al. 2008; Excoffier et al. 2009; Petit and Excoffier 2009).

The vast majority of the reported cases of introgression in animals involve the mitochondrial genome (mtDNA; Toews & Brelsford, 2012), and introgression is often massive (Melo-Ferreira et al. 2005; Good et al. 2008; Sequeira et al. 2011). Explanations for the apparent tendency of mtDNA to cross species boundaries include pure demography/drift, sex-biased interspecific mating, and very often adaptation (reviewed in Toews and Brelsford 2012). Some studies have documented the role nuclear adaptive introgression can have on species evolution and interactions (The Heliconius Genome Consortium 2012; Nadeau et al. 2013; Huerta-Sánchez et al. 2014; Sankararaman et al. 2014; Vernot and Akey 2014; Lamichhaney et al. 2015; Liu et al. 2015), so the question remains open in cases of mtDNA massive introgression. The question can be tackled by genomics in two ways. One consists in evaluating the likelihood of the purely demographic process described above, which requires comparing nuclear and mitochondrial DNA introgression in the framework of the underlying demographic processes. Nuclear introgression in cases of massive mtDNA introgression has rarely been assessed in any detail (see e.g. Good et al 2015), and the likelihood of the purely demographic model was never assessed. Another test of the adaptive nature of mtDNA introgression could however be conceived. The nuclear and mitochondrial genomes closely interact in key cellular functions (e.g. oxidative phosphorylation) and the maintenance of the mitochondria can depend on over 1000 nuclear-encoded genes (see Sloan 2016). Thus, an adaptive role of mtDNA introgression underlies the possibility of cytonuclear co-evolution and adaptation in shaping the patterns of nuclear introgression. In fact, evidence of co-introgression of nuclear genes interacting with the mitochondria has been suggested in a few case studies (e.g. Pritchard & Edmands 2013, Beck et al. 2015). Note though that in several other cases the existence of diverging mtDNA lineages has also been inferred to limit levels of gene flow for the mitochondrial genome and interacting nuclear loci (e.g. Bar-Yaacov et al. 2015; McKenzie et al. 2016; Morales et al. 2016).

In this work, we assess genomic patterns of introgression in a system with massive mtDNA introgression that presumably occurred during a range replacement, providing the opportunity to assess the relative contributions of demographic and selective processes to genetic admixture. In the Iberian Peninsula, the Iberian hare, *Lepus granatensis*, harbors high frequencies of mtDNA from the arctic-boreal mountain

hare, *Lepus timidus* (Melo-Ferreira et al. 2005; Alves et al. 2008). Likewise, the two other hares inhabiting Iberia, *L. europaeus* and *L. castroviejoi*, were also massively affected by mtDNA introgression from *L. timidus* (Melo-Ferreira et al. 2005; 2012). The fossil record suggests that the latter species (currently found only in the northern Palearctic and in some isolated populations, as in Ireland, Scotland, Poland or the Alps) was present in northern Iberia during the Pleistocene but went extinct in the region, presumably after the last glacial maximum (Altuna 1970). Several aspects of mtDNA variation in the Iberian hare appear compatible with a scenario of allele surfing on a wave of expansion of *L. granatensis* into the territory of *L. timidus* in Northern Iberia, with hybridization. These include a south-north gradient of mitochondrial introgression frequency (Melo-Ferreira et al. 2005; 2009) and a perpendicular phylogeographic structure of mtDNA of *timidus* origin (Melo-Ferreira et al. 2011). However, possible signs of competitive replacement of the native mtDNA genome by the alien one (which would be compatible with adaptive introgression) were also proposed (Melo-Ferreira et al. 2007; 2011). Studies of a small number of nuclear markers suggested (i) sporadic low frequency introgression, all over the distribution area, contrary to mtDNA (Melo-Ferreira et al. 2009); (ii) signs of south-north range expansion of *L. granatensis* (Marques et al. 2017); (iii) geographically widespread high frequency introgression of an X chromosome fragment (Melo-Ferreira et al. 2011). These preliminary results draw a contrasted and incomplete picture, leaving open the question of the relative importance of demographic and selective factors in determining introgression in *L. granatensis*, including for mtDNA.

Here we tackle this question by analyzing the complete genome variation of the two species, characterizing the genomic and geographic extent of introgression. We put patterns of nuclear introgression in relation with structural characteristics of the genome and question the relationship between recombination and selective constraints in determining variations of introgression frequencies. We also question whether massive mtDNA introgression was accompanied by substantial introgression of some nuclear mitochondrial genes, and finally look for evidence of adaptive introgression of nuclear genes that could be related to their function.

3. Results

Sampling and Genomic Datasets

We sequenced the genomes of 10 *L. granatensis* specimens sampled over the species distribution range in Iberia, across the reported gradient of mtDNA introgression from *L. timidus*, from absent in the south to frequent in the north (Figure 3.1A). We also sequenced the genomes of three *L. timidus* (two from the Alps and one from Scandinavia, Figure 3.1B) and one *L. americanus* that was used as outgroup for some analyses. Sequencing effort and resulting coverage are shown in Annex II - Table S3.1.

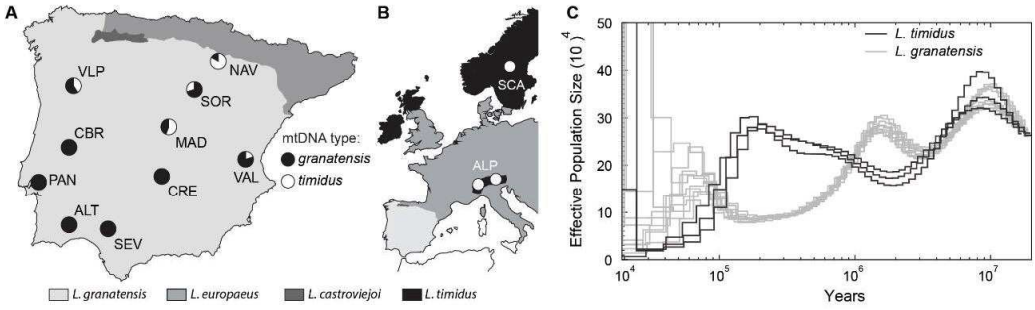


Figure 3.1 Geographic Distribution of hare species in (A) Iberian Peninsula and (B) Western Europe (approximate distributions were based on Mitchel-Jones et al. 1999), and (C) demographic profiles of the boreal (*L. timidus*) and temperate species (*L. granatensis*). (A) Sampling localities and their acronyms for *L. granatensis* (see Annex II - Table S3.1 for a detailed description); pie charts indicate the proportion of *granatensis* and *timidus* type mitochondrial DNA variants estimated from population samples (from Acevedo et al. 2015). (B) Sampling localities and their acronyms for *L. timidus*. (C) Inference of population size changes over time using the PSMC method; the y-axis denotes the scaled effective population size and the x-axis the time in years before present (log-scaled), assuming a substitution rate of 2.8x10-8 substitutions per site per year and a generation time of 2 years.

Using an iterative mapping approach (as in Halligan et al. 2013), we built a hare pseudo-reference genome using the rabbit genome as template. This procedure allowed improving read mapping proportions, averaged across all individuals, from 92.3 to 93.6%. Note that broad synteny between the rabbit and hare karyotypes is expected but some known fusions/fissions exist (Robinson et al. 2002), and were taken into account whenever needed.

Inference and Broad Impact of Nuclear Introgression

We inferred regions of the 10 sequenced *L. granatensis* genomes that were affected by introgression from *L. timidus*. Most methods aimed at detecting local ancestry in admixed populations rely on the observation of presumably pure parental populations (e.g. Price et al. 2009; Liu et al. 2014a; Martin et al. 2014). However previous analyses of *L. granatensis*, although based on a limited number of markers, had suggested that nuclear introgression from *L. timidus* was present all over the range of *L. granatensis* (Melo-Ferreira et al. 2009), so that none of the samples sequenced here could be considered a pure reference. We therefore used a recently developed ancestry inference method, implemented in the ELAI (Efficient Local Ancestry Inference) software (Guan 2014), which can accommodate such situation. This method is not based on an arbitrary segmentation of the genome and is able to infer the boundaries of the introgression tracts in the genome. When one of the parental populations is unobserved, the method is expected to perform properly if the admixed population has a high proportion of ancestry from this unobserved origin. Previous data on a limited number of markers had suggested that this was the case in *L. granatensis* (Melo-Ferreira et al. 2009). However, selection or stochastic processes could have driven high frequency introgression of some particular genomic regions that could remain undetected by this method. We therefore used another method that does not have this limitation, RND (for Relative Node Depth, Feder et al. 2005). For each of the windows in which we segmented the genome, the sequence divergence between statistically phased haplotypes of the focal (here *L. granatensis*) and donor (*L. timidus*) populations was estimated, and standardized by the divergence to the outgroup (*L. americanus*), to control for mutation rate variations across windows. Regions of introgression are expected to produce exceptionally low minimum RND values (RNDmin), independently of the introgression frequency (Rosenzweig et al. 2016). Using the inferences from ELAI, we were able to verify that phasing appeared correct in regions of introgression, where linkage disequilibrium is enhanced, and allowed recovering in-phase parental haplotypes (not shown). Relying on the ELAI inferences, we determined the power and false discovery rate (FDR) of the RND approach depending on the applied threshold (Annex II - Figure S3.1). For each of three window sizes used, we selected the RND threshold that had a predicted FDR of 10%. This resulted in a low estimated power of RND (16.9%, 25.7% and 42.6% for 10kb, 20kb and 50kb RND windows, respectively; Annex II - Figure S3.1).

Introgression was found to occur genome-wide. According to ELAI-based estimates, the proportion of the genome affected by introgression varied between 1.38

and 2.44% among *L. granatensis* specimens. These proportions drop to 0.27-0.79% when based on the RND analysis (0.27-0.40%, 0.42-0.65%, 0.48-0.79% for 10kb, 20kb and 50kb windows respectively), in accordance with the expectations from the power analysis given the FDR rate applied (Annex II - Figure S3.1).

Historic and Geographic Context of Introgressive Hybridization Events

We ran the PSMC method (Li and Durbin 2011) on the whole genome of each individual of both species. This method uses inferred tracts of homozygosity in a diploid genome to reconstruct the rates of coalescence into the past, and infers the evolution of effective size of the supposedly panmictic population in which the two sampled haplotypes have evolved. The results (Figure 3.1C) suggest at least two episodes of population size fluctuation in both species after their divergence (occurring when the two curves merge in the past), that appear synchronized but in opposite directions, i.e. expansion of one species is concomitant with retraction of the other. However, the method is unable to retrieve information on the more recent epoch that would correspond to the last de-glaciation.

We dated the introgression episode using two approaches that use information from tracts of linked variants. The first uses identity by state (IBS) tracts of DNA shared within and between populations to jointly estimate the time and magnitude of introgression along with divergence time and effective population sizes (Harris and Nielsen 2013). The second is based on the expectations of shared haplotype sizes distribution as a function of time since introgression, which should decay due to recombination (Pool and Nielsen 2009; but see Gravel 2012, Liang and Nielsen 2014). Estimates based on both methods suggest that introgression is recent, dating to the end of Pleistocene (58.9 kya, based on IBS tracts for the model with best likelihood; 24.3 kya when considering only IBS tracts larger than 10 kb and thus presumably the most informative about recent migration; Annex II - Table S2), or early Holocene (7 kya, considering the average sizes of introgression tracts, 29’364 bp; Annex II - Figure S3.2). Next, we looked into the geographic partitioning of *L. granatensis* diversity and introgression. A Principal Component Analysis (PCA) performed on the 10 *L. granatensis* revealed geographic differentiation (Annex II - Figure S3.3A), with the first PCA axis being significantly correlated with longitude (Spearman's rank correlation test p-value = 0.0009, $\rho = -0.9$) and latitude (p-value = 0.0159, $\rho = -0.76$) (Annex II - Figure S3.3B). Moreover, we found a correlation between genetic and geographic distances (Annex II -

Figure S3.3C, Mantel test with 9999 permutations p-value = 0.0001, r = 0.85). This could be due to isolation by distance, recent geographic expansion, or a gradient of introgression from *L. timidus*. In order to test the latter possibility we added *L. timidus* into the PCA and a south-north gradient of differentiation was obvious on each of the first two axes (Figure 3.2A). Differentiation along axis 1 could represent a gradient of introgression, as it separates the two species, and differentiation along axis 2 would then be linked to isolation by distance or geographic expansion, as suggested by the correlation of axis 2 values with both longitude (Spearman's rank correlation test p-value << 0.01, ρ = -0.95; Annex II - Figure S3.4A) and latitude (Spearman's rank correlation test p-value = 0.035, ρ = -0.68; Annex II - Figure S3.4A), and by the correlation between genetic (measured as the distance between axis 2 coordinates) and geographic distances (Annex II - Figure S3.4B, Mantel test with 9999 permutations p-value = 0.0001, r = 0.86). This appears confirmed when the analysis is run after replacing *L. timidus* by *L. americanus*, from which no introgression can have occurred: we recover an identical gradient along axis 2, but none along the first axis corresponding to species differentiation (Annex II - Figure S3.5). Furthermore, we ran the same analyses after excluding the genomic regions where introgression was detected with ELAI (Annex II - Figure S3.6). As expected, in this case using *americanus* or *timidus* as outgroup provided similar results: a south-north differentiation axis perpendicular to the species-discriminating axis.

In addition, we found that estimates of genomic proportions of introgression per individual correlate with geography, since they significantly increase with distance to the southernmost sampled point (which is near the inferred origin of a range expansion; Marques et al. 2017), both for ELAI (Spearman's rank correlation test p-value = 0.0009, ρ = 0.90) and RND-based estimates (Spearman's rank correlation test p-value = 0.0027, 0.0035, 0.0045 and ρ = 0.87, 0.85 and 0.84, for 10kb, 20kb and 50kb windows, respectively) (Figure 3.2B). We also found a correlation between mean introgression tract length per individual and geography: the size of introgression tracts significantly increases with distance to the southernmost sampled point (Spearman's rank correlation test p-value = 0.0027 and ρ = 0.87) (Figure 3.2C).

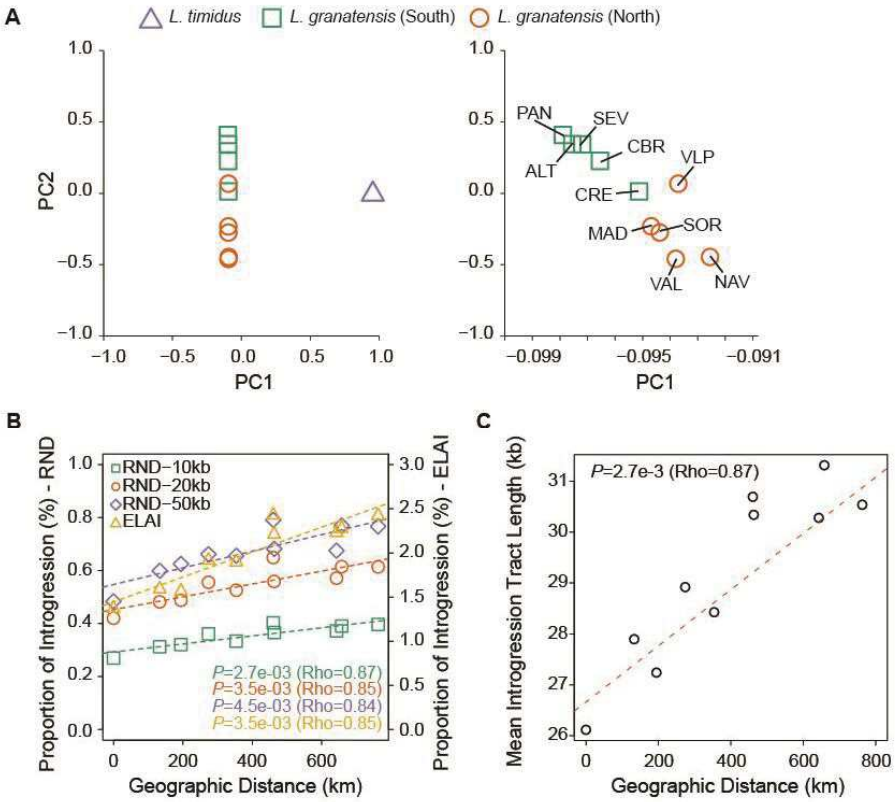


Figure 3.2 Geographic partitioning of *L. granatensis* genetic variation and impact of introgression. (A) PCA summary of genetic variation in *L. granatensis* including one *L. timidus* individual as outgroup, using whole genome data (left), and zoom on *L. granatensis* samples, distinguishing the 5 southernmost and 5 northernmost samples (right). (B) Correlation between individual proportion of introgression and geographical distance (measured in kilometers) to the southernmost sample (Spearman's rank correlation p=0.00 for all methods and window sizes; dashed lines indicate linear regression trendlines). (C) Correlation between individual average introgression tract length and geographical distance (measured in kilometers) to the southernmost individual (Spearman's rank correlation p=0.0027). Dashed lines indicate linear regression trendlines.

Introgression during a range replacement

Patterns of genetic variation in *L. granatensis*, higher impact of introgression towards the north (found here for the nuclear genome and previously for mtDNA), and the northward increase in introgression tract lengths are compatible with introgression occurring during a northward range expansion of the species into the range of *L. timidus* when present in Northern Iberia. However, while mtDNA introgression is strongly structured, being absent in southern Iberia and reaching high frequencies in the North, nuclear DNA introgression is generally rare (most cases affecting a single haplotype)

and present all over the species range. In order to appraise whether these patterns could be generated by a single underlying demographic model, we simulated this process using SPLATCHE2 (Ray et al. 2010). *L. granatensis* was simulated to expand from south-western Iberia 20 kya (Marques et al. 2017), and to replace *L. timidus* where it was likely present in northern Iberia at the LGM (based on ecological niche modelling; Acevedo et al. 2015; Figure 3.3A), using several combinations of parameter values (Table 3.1). The model with lowest carrying capacity (K=1000), highest inter-deme migration (M=0.2) and lowest admixture (A=0.005) resulted in low levels of individual proportion of introgression (mean across individuals = 3.9%), most similar to those observed with ELAI-based inferences (mean across individuals = 2.4%; Table 3.1). Accordingly, the distribution of introgression frequencies shows a skew towards no or low frequency introgression, similarly to the empirical results (Figure 3.3B). Reducing inter-deme migration resulted in a south-north gradient of introgression (Table 3.1).

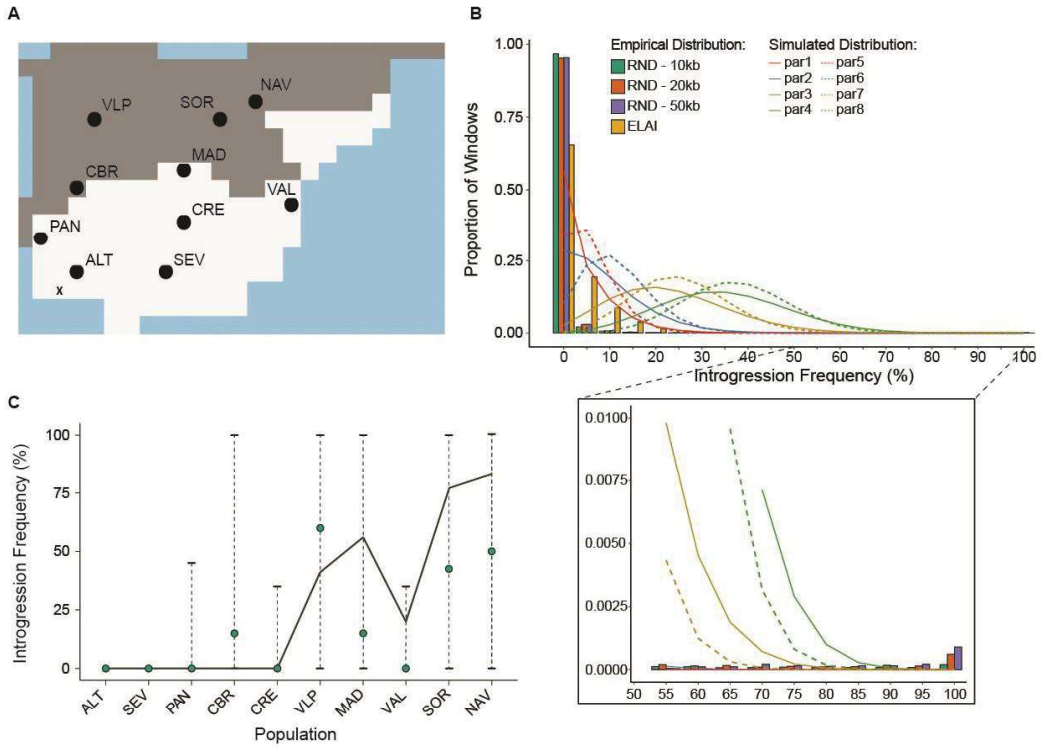


Figure 3.3 Comparison of empirical and simulated introgression frequencies. (A) Simulated landscape of Iberian Peninsula used in SPLATCHE simulations. The dark grey area represents the distribution of *L. timidus* at the last glacial maximum, as determined by ecological niche modelling (probability of presence higher than 0.8 in northern Iberia; Acevedo et al. 2015). Black points indicate demes for which the proportion of introgression was recorded in each of the simulations (the demes corresponding to the geographical locations of the empirical samples – see Annex II - Table S1; location names are given next to the points) and the 'X' marks the origin of *L. granatensis* expansion 20 kya according to

Marques et al (2017). (B) Colored bars illustrate the empirical distribution of nuclear introgression frequencies across genomic windows for the different introgression detection methods used. Solid and dashed lines represent the simulated distributions for different parameter sets (see Table 3.1 for a detailed description of parameter sets of the simulations). (C) Empirical (solid line) and simulated (green dots, median value per population based on 1000 simulations; vertical T lines represent 1.5 x interquartile range (IQR) extensions as an indication of the variance) mitochondrial introgression frequencies in the 10 sampled localities.

Table 3.1 Mean population introgression frequencies based on empirical inference and simulated datasets (using SPLATCHE).

	Parameters					Introgression Frequencies (%)											Max.[a]	Sign.[b]
Set	K_G	K_T	G	M	A	MEAN	ALT	SEV	PAN	CBR	CRE	VLP	MAD	VAL	SOR	NAV		
Empirical																		
ELAI	-	-	-	-	-	2.0	1.3	1.6	1.5	1.9	1.9	2.4	2.2	2.2	2.2	2.4	-	-
RND-10kb	-	-	-	-	-	0.4	0.3	0.3	0.3	0.4	0.3	0.4	0.4	0.4	0.4	0.4	-	-
RND-20kb	-	-	-	-	-	0.5	0.4	0.5	0.5	0.6	0.5	0.6	0.6	0.6	0.6	0.6	-	-
RND-50kb	-	-	-	-	-	0.7	0.5	0.6	0.6	0.7	0.7	0.8	0.7	0.8	0.7	0.8	-	-
Simulated																		
par1	1000	500	0.5	0.2	0.005	3.9	4.0	4.0	3.9	3.9	3.9	4.0	4.0	3.9	4.0	3.9	70	35
par2	1000	500	0.5	0.02	0.005	8.5	7.0	6.7	7.6	8.6	7.6	10.3	8.8	8.2	9.8	10.1	80	45
par3	1000	500	0.5	0.2	0.03	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9	95	70
par4	1000	500	0.5	0.02	0.03	34.9	28.9	28.1	31.1	34.9	31.7	40.9	36.2	34.5	40.7	42.4	95	85
par5	10000	5000	0.5	0.2	0.005	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	50	30
par6	10000	5000	0.5	0.02	0.005	11.3	9.5	9.0	10.2	11.5	10.2	13.5	11.7	10.8	12.9	13.2	65	40
par7	10000	5000	0.5	0.2	0.03	25.2	25.3	25.2	25.2	25.2	25.2	25.2	25.2	25.3	25.2	25.2	80	60
par8	10000	5000	0.5	0.02	0.03	37.4	31.2	30.5	33.4	37.3	34.1	43.2	38.7	37.1	43.2	45.0	95	75

K_G – *L. granatensis* deme carrying capacity; K_T – *L. timidus* deme carrying capacity; G – intrinsic growth rate (same for *L. timidus* and *L. granatensis*); M - migration rates between adjacent demes (same for *L. timidus* and *L. granatensis*); A - bidirectional admixture. Population names are as in Annex II - Table S1 and Figure 3.1. [a]Maximum introgression frequency in percentage. [b]Introgression frequency (in percentage) above which empirical introgression frequencies are significantly higher than expected according to simulations

In order to understand whether the empirical geographic patterns of mtDNA introgression could be recovered under the same model, we repeated the simulations adjusting parameter values to properties of mtDNA transmission. Steep northwards clines of increasing mtDNA introgression were obtained when reducing the effective population size to ¼ of that of the nuclear genome (mimicking female transmission), decreasing inter-deme migration to a minimum (mimicking female philopatry) and setting predominant gene flow in one direction, *L. timidus* into *L. granatensis* (mimicking sex-biased interspecific crosses) (see median of simulated introgression proportions per population in Figure 3.3C). Note however that simulations show that under these conditions geographic patterns can vary substantially from one simulation to the other, but the majority of them shows a higher prevalence of introgression in the north than in the south. Indeed, we note that the difference in mtDNA introgression frequencies between the 5 northernmost and 5 southernmost populations from where our genomic samples come from (55.4%), is within the 95% quantile of the distribution of the same measure obtained from our simulations (Annex II - Figure S3.7). These results suggest nuclear and mtDNA patterns of introgression can be reconciled under a similar demographic model.

Outlier high-frequency introgression

Most introgressions detected by either method occur at low frequencies, with a majority found only in one of the 20 haploid genomes sampled (Figure 3.3B). However, the RND-based method detected a bulk of introgressed fragments at very high frequencies (Figure 3.3B). We questioned whether this could reflect introgression favored by selection, or be a likely stochastic outcome of past demography, hybridization and expansion. Simulations recovering levels of introgression comparable to the empirical values never recovered markers introgressed at frequencies higher than 70% over the 20 sampled haplotypes (par1; Table 3.1). We thus simulated the demographic and coalescent process maximizing the probability of introgression using several parameter combinations (par2-8; Table 3.1). Under these models, we tested the frequency of introgression above which empirical introgression frequencies are significantly higher than expected. We found that for frequencies of introgression >80%, empirical values were always significantly higher than expected, regardless of the simulated parameter set (Table 3.1). In addition, the two extreme conditions (par 4 and par8) led to average frequencies of introgression per specimen ranging from 34.9 to

37.4%, which is ~20-fold higher than inferred for the empirical dataset. These high frequency introgressions are thus clear outliers and were presumably driven by selection.

Taking together the evidence of introgression from all RND windows, we found 139 regions to have outlier introgression frequencies (i.e. >80%) according to our demographic simulations, and they contained 123 genes (Annex II - Table S3). We then inspected the characteristics of these genes. We measured dN/dS between *L. americanus* and *L. timidus* (be it sampled in *timidus* or *granatensis*) and found two genes, "E230025N22Rik" (gene name obtained from the mouse ortholog, as the rabbit gene name is not defined) and HERC6, to have potentially evolved under positive selection (dN/dS > 1). We then performed a Gene Ontology (GO) enrichment analysis of these 123 highly introgressed genes. Applying the Benjamini-Hochberg correction for multiple tests, we found enrichment in several biological processes, including e.g. positive regulation of leukocyte mediated immunity, macroautophagy and spermatogenesis (Annex II - Tables S4, S5). However when applying a more stringent correction method that takes into account the hierarchy of GO term annotation, no significant enrichment was found.

Heterogeneity of Introgression across the Genome

We now study how variations of the distribution of introgression along the genome correlate with those of various structural and functional characteristics of the genome that can affect selection.

We found that average introgression frequency along the X-chromosome is lower than along the autosomes. In fact, the proportion of introgression across individuals according to ELAI was significantly lower on the X chromosome (mean proportion of introgression = 0.24%), when compared to the autosomes (mean proportion of introgression = 2.04%; Mann-Whitney U test p-value << 0.01; Figure 3.4A). Regarding RND-based estimates, we found that mean RND values were significantly lower for the X-chromosome (RND = 0.574, 0.578, 0.577 for 10kb, 20kb and 50kb RND windows, respectively) than the autosomes (RND = 0.578, 0.581, 0.580 for 10kb, 20kb and 50kb RND windows, respectively; Mann-Whitney U test p-value << 0.01), indicating that our detection of introgression was less conservative for the X than the autosomes since we used a common RND threshold. Despite this, the proportion of introgression on the X was also found to be significantly lower (mean proportion of introgression = 0.11% vs

0.34%, 0.16% vs 0.51%, 0.27% vs 0.62% for 10kb, 20kb and 50kb RND windows, respectively; Mann-Whitney U test p-value << 0.01; Figure 3.4A).

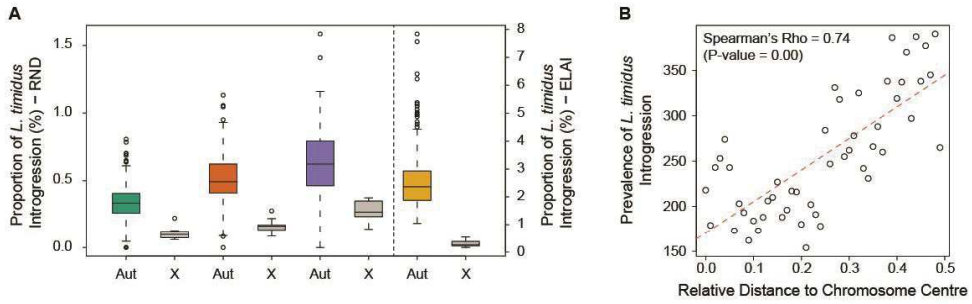


Figure 3.4 Variations of introgression prevalence across the genome. (A) Distribution of the proportion of introgression across individuals for autosomes (colored boxplots; Aut) and X-chromosome (grey boxplots; X) (Mann-Whitney U test p=0.00 for all methods and window sizes). From left to right, the results of the three RND window sizes and of ELAI. Note that different y-axis scales are used for the two methods. (B) Correlation between prevalence of introgression (measured as the number of introgressed ELAI segments overlapping a given position) and relative distance to chromosome center (Spearman's rank correlation p=0.00; dashed line indicates a linear regression trendline).

Furthermore, we found that the impact of introgression is not uniform along the chromosomes. Based on the chromosomal position of informative SNPs, we find that the prevalence of introgression, measured as the number of ELAI introgression segments across all individuals overlapping a given SNP, increases significantly with distance to the chromosome centre (Spearman's rank correlation test p-value << 0.01, ρ = 0.74; Figure 3.4B). Such correlation was not found when considering the distance to the centromere (Spearman's rank correlation test p-value = 0.36, ρ = 0.13; Annex II - Figure S3.8). Using software LDhat (McVean et al. 2002; Auton and McVean 2007) we estimated the population recombination rate (ρ) along the genome and found a significant positive correlation with distance to chromosome center (Spearman's rank correlation test p-value << 0.01, ρ = 0.14) (Annex II - Figure S3.9). We were concerned that this latter correlation could result from LDhat inferring more recombination in regions with more introgression. However the correlation was maintained when LDhat blocks overlapping introgressed fragments (inferred using ELAI) were removed (Spearman's rank correlation test p-value << 0.01, ρ = 0.14) (Annex II - Figure S3.9). We therefore conclude that both introgression and recombination increase from the center to the tips of the chromosomes.

Finally, we tested the relationship between introgression frequency and functional constraint, using the presence of protein coding genes as a proxy. Subsets of genomic windows were repeatedly sampled and the frequencies of introgressed

windows compared among those overlapping genes and those not overlapping. We found that introgressed regions had a significantly higher chance of being found in genic regions, independently of the method used to detect introgression or window size (Wilcoxon rank sum test p-value << 0.001; Annex II - Figure S3.10). This could result from introgression compensating the effects of deleterious recessive mutations segregating in *L. granatensis*. However, when comparing introgressed and non-introgressed genes, we found significant differences neither in dN/dS measured between *L. timidus* and non-introgressed *L. granatensis* (Wilcoxon rank sum test p-value = 0.201), nor for the neutrality index (NI; Wilcoxon rank sum test p-value = 0.609), nor for $\pi N/\pi S$ (*L. granatensis*: Wilcoxon rank sum test p-value = 0.811, *L. timidus*: Wilcoxon rank sum test p-value = 0.171 (Annex II - Table S4). Overall, the genes that introgress do not appear more functionally constrained than those that do not. We however found that the evolutionary rate at both synonymous and non-synonymous sites (the two being correlated; Spearman's rank correlation test p-value << 0.01, $\rho = 0.25$) was significantly higher for non-introgressed genes (Annex II - Table S4). This could suggest a bias towards a better efficiency of introgression detection in regions of low polymorphism and thus less incomplete lineage sorting, which could explain the apparent higher introgression detected in genic regions.

Introgression of nuclear genes with mitochondrial functions

Finally, we ask whether the massive mtDNA introgression in Northern Iberia was accompanied by introgression of some nuclear genes interacting with the mitochondrial genome or its products. Such genes with high frequencies of introgression, paralleling that for mtDNA, would be of particular interest, so we used here the results of the RND test. We examined patterns of variation and introgression at "mitonuc" genes, i.e. nuclear genes the products of which can be found in the mitochondria. Of the 1211 reported such genes (Gu et al. 2011; Calvo et al. 2016), 1178 were covered by at least one RND window passing our threshold of information content (see Methods). Among these, we distinguished a subcategory, which we call "mitonuc-direct", made up of genes with a product known to interact directly with the mitochondrial genome or its products (RNA and proteins). Among the188 genes in this category, 185 overlapped valid RND windows, including all 73 OXPHOS genes.

Among the 3312 genes overlapping introgressed regions, we found 166 mitonuc genes, 23 of which belonged to mitonuc-direct and eight of which were OXPHOS genes

(Annex II - Table S3.7). This does not reflect enrichment either in mitonuc (Pearson's Chi-squared test p-value = 0.554), or mitonuc-direct genes (Pearson's Chi-squared test p-value = 0.385), or OXPHOS genes (Pearson's Chi-squared test p-value = 0.368) (Annex II - Table S3.8).

Introgression frequency of mitonuc genes followed the general genomic pattern, being mostly rare (Annex II - Figure S3.11). However, five mitonuc genes (TYMP, TMLHE, L2HGDH, ATG5 and SDHAF4) and one mitonuc-direct (RARS2) were found introgressed at high frequencies (>80%; Annex II - Table S3.9). We further inspected genes with introgression distribution resembling that of mtDNA (absence of introgression in the 10 southern haploid genomes and at least 20% of introgression in the 10 northern ones). We did not find any enrichment in mitonuc category among such genes. However, 17 mitonuc genes, including three mitonuc-direct, of which two were OXPHOS, showed such a pattern (Annex II - Table S3.10). For these 17 genes, we inspected whether any amino acid replacement between the alleles of *timidus* and *granatensis* origins could suggest a strong functional impact, based on the analysis of sequence conservation at deep evolutionary scales, using SIFT (Kumar et al. 2009). We identified six non-synonymous variants in four of these genes (HEBP1, “ATP5F1”, “HP”, and “RP11-561B11.2”; the latter three names were obtained from the human orthologs as rabbit gene names are not defined), and the introgressed variant was the derived state in all cases (Annex II - Table S3.11). In two of these genes (“HP” and HEBP1), one amino acid change was predicted to potentially influence protein function.

Given that the power of RND to detect introgression is low, we may have missed some important mitonuc genes that are introgressed at high frequencies. We thus relaxed our stringency for introgression detection, but before doing so, re-evaluated power by simulation from the empirical data. We generated artificial introgression by introducing *L. timidus* haplotype fragments of variable sizes, and overlapping mitonuc genes, into the haploid genome of one *L. granatensis* and re-calculated RND, which allowed us to evaluate the power of the method in conditions close to the real data. We found that the power to detect introgression ranged from 2.2 to 10.5% and 11.9 to 30.9%, depending on RND window size, for introgressed fragments of sizes similar to the mode and median size of ELAI introgressed fragments, respectively (Annex II - Table S3.12). We therefore redid the inference of introgression using less conservative RND thresholds that predict a FDR of 30% instead of the previous 10%. In our simulations, this allowed recovering ca. 50% of the artificially introgressed genes with at least one of the RND window sizes, when considering 10kb artificially introgressed fragments (Annex II - Table

S3.13). At this threshold, 8658 genes were found to overlap introgressed regions among which 460 were mitonuc genes, 65 were mitonuc-direct and 25 belong to OXPHOS. Of these, 69 had outlier introgression frequencies (i.e. at least 85%; Annex II - Table S3.9), and 32 mitonuc genes presented introgression distributions resembling that of mtDNA as defined above (Annex II - Table S3.10), of which four are mitonuc-direct (MRPS22, MRPL2, MRPL15, GARS) and one belongs to the OXPHOS ("ATP5F1"; gene name defined from the human ortholog as gene name in rabbit is not defined; Annex II - Table S3.10). However, no amino acid replacements were found between the *granatensis* and *timidus* variants at these genes.

4. Discussion

In this study, we explored genome-wide patterns of historical introgression, and show that both pure demography and natural selection have shaped the genetic contribution of *L. timidus* today embedded in the genome of *L. granatensis*.

Methodological challenges

A limitation of our study results from using the rabbit genome and its annotation as a reference. The creation of a pseudo-reference by iterative mapping allowed improving mapping success of raw reads and we do not expect major biases to result from the mapping process. The major source of potential bias in our results could result from rearrangements between the hare and rabbit genomes, since some of the analyses we performed suppose proper ordering of the sequences along the chromosomes. Although the karyotypes of hares and rabbits appear very similar, and we accounted for the few known exceptions when relevant (chromosomes 1 and 2 of the rabbit are split in hares; Robinson et al. 2002), this approach is blind to potential rearrangements at smaller scales not detectable under the microscope. Even though we may have missed some patterns due to this limitation, it is unlikely that it has created the signals that emerged from our analyses.

Another limitation is of course our power to infer introgression correctly. Although we have used the sophisticated ELAI method based on linkage disequilibrium and genome segmentation by HMM, which is well adapted to our situation (unphased genomes and absence of one of the two reference donor populations), this method is expected under the conditions we used it to recover poorly highly introgressed regions, which are of great interest for our purpose. We thus used the RND method to attempt recovering such regions, but we have determined, both in comparison to the results of ELAI and by simulating introgression, that the method has little power and a high false discovery rate. We were however able to estimate this rate, and to adjust our RND thresholds to keep it reasonably low, at a risk of missing cases of real introgression. An example is a fragment of the X-linked PHKA2 gene that had been previously characterized by classical PCR and Sanger sequencing as a case of high frequency ubiquitous introgression (Melo-Ferreira et al. 2011). We could verify that this introgression was present in our dataset, but was not detected by the methods we used, presumably due to its short length. Again, this lack of power is unlikely to have affected the general patterns that we infer. However, we were interested not only in average

properties of genes or genomic regions, but also in introgression at individual genes with particular functions (mitonuc genes). For this reason, in this context we allowed a relaxation of the criteria retained as evidence for introgression, with increased risk of pointing to false positives. The candidate genes pinpointed by doing so must thus be considered with caution, and should be checked more thoroughly, as described below.

Geographic patterns of nuclear introgression and demographic history

One hypothesis to explain the gradient of mtDNA introgression is the post-glacial expansion of *L. granatensis* from southern Iberia northwards, into a territory then occupied by *L. timidus*, where hybridization occurred. Such event is expected to have left distinctive traces in genomic variation.

First, there should be traces of a demographic expansion of *L. granatensis*, concomitant with a contraction of *L. timidus*. Our PSMC analyses indeed suggested opposite past demographic profiles of the two species, expansion of one of them being contemporaneous with retraction of the other (Figure 3.1C). Given past cycles of climatic oscillations, these opposite demographic profiles appear logical for species adapted to contrasted climatic conditions, one temperate and the other boreal (see also Stewart et al. 2010). The method was unable to recover reliably demographic profiles at the presumed recent time of contact between the two species (LGM) at the origin of the observed introgression. We note however that the demography of Iberian populations of *L. timidus* at that time could not have been inferred since the sampled individuals are not descendants of populations from this region, which are now extinct, but this extinction must have been preceded by local population size collapse. Based on mtDNA variation, extent populations of *L. timidus* have maintained high polymorphism and lack clear geographical differentiation over the species range (Melo-Ferreira et al. 2007; Smith et al. 2017), although the overall population demography appears to have been affected by recent climatic changes (Smith et al. 2017). Our PSMC analysis therefore reveals such species-wide patterns that appear coherent, but cannot of course inform us on the recent local history of the extinct Iberian population of *L. timidus*.

A second prediction of the invasion with replacement model is a gradient of increased introgression in the direction of the expansion. Our PCA analysis indeed revealed a south-north gradient of differentiation in *L. granatensis* along the axis of differentiation with *L. timidus*, presumably resulting from introgression (since it was not found when substituting *L. americanus* to *L. timidus*), and suggesting more introgression

in the North (Figure 3.2A). Our inventory of introgressed genomic regions using both ELAI and RND clearly confirmed this pattern (Figure 3.2B). South-north differentiation was even more marked along a PCA axis perpendicular to that corresponding to differentiation with *timidus* (Figure 3.2A), which presumably results from the south-north expansion, as demonstrated by a previous analysis of Marques et al. (2017) based on 100 SNPs but a much larger and geographically spread sample.

Another prediction of the proposed demographic scenario is the age of the introgression, expected to correspond to the last de-glaciation. We obtained different estimates depending on the method used (IBS tract length distributions or average introgression tract length; 24-7 kya) but they are compatible with hybridization occurring at the end of the last glacial period and possibly persisting towards the Holocene. Independently of the absolute age of the introgression, the invasion model would predict a gradient of introgression age, from most ancient at the initial front of invasion to more recent in more recently invaded territories. This exactly matches the observed gradient of northward increase of average introgression tract sizes, longer tracts reflecting more recent introgression (Figure 3.2C).

By explicitly simulating the proposed demographic model, we were able to reproduce the patterns of introgression observed in our nuclear data (Figure 3.3 and Table 3.1). Introgression frequency distribution was biased toward no or rare introgression, and low proportions of introgression were simulated overall. The overall empirical proportion of introgression was lower than in the simulations, which could be due to the used combination of parameter values in the simulations, or more likely to our lack of power to inventory all introgression tracts. The empirical south-north gradient of increasing proportion of introgression was also recovered when decreasing the intraspecific migration rate. Interestingly, we note that the gradients of introgression proportion have different clines in the northern and southern ranges, a pattern that is equally recovered with the empirical and simulated datasets (Annex II - Figure S3.12B and S3.12C). A similar difference is recovered for mean introgression tract lengths per sequenced specimen, which show a larger range in the south than in the north (Annex II - Figure S3.12A), compatible with faster progression of introgression in the north. This thus goes well with the idea that introgression in the north results from a rapid range replacement, while in the south it results from slower diffusion of introgressed variants due to intraspecific migration, as modelled in our simulations.

Overall, our results are therefore compatible with the invasion-replacement hypothesis and the nuclear and mitochondrial genomes share similar patterns of

increased introgression towards the north. However, levels of nuclear introgression are on average much lower than those found for mtDNA, and the northwards gradient is much shallower (Figure 3.3B). We found that mimicking the haploid nature and maternal transmission of mtDNA, female philopatry and sex-biased introgression, with predominant *timidus* to *granatensis* gene flow, we were able to reproduce the empirical mtDNA introgression patterns (Figure 3.3C). These settings represent commonly invoked causes for the ubiquitous nature of mtDNA introgression. First, the lower effective population size of mtDNA increases the probability of fixation of introgressed variants. Second, lower intra-specific migration resulting from female philopatry decreases the probability that introgressed variants in the invasion front are diluted by migration of native alleles from the parental populations (Currat et al. 2008; Petit and Excoffier 2009). Male hares, as commonly described for many other mammals, tend to disperse farther than females (Bray et al 2007, Avril et al 2011). Female philopatry and male-biased dispersal thus also explain that traces of introgression are found all across the Iberian Peninsula for nuclear DNA, while for mtDNA they remain in the north, where hybridization events took place. Third, sex-biased interspecific crosses, due to male-biased dispersal, frequency dependent assortative mating or other behavioral factors may promote unidirectional introgression of mtDNA (see Chan and Levin 2005). These asymmetries during interspecific crosses have often been invoked in hares (Thulin et al. 2002; 2006).

In a recent study, Bonnet et al. (2017) simulated under a multi-locus framework several demographic and selective scenarios corresponding to verbal hypotheses commonly used to explain cytonuclear discordance in patterns of introgression (including sex related asymmetries, spatial invasion-replacement and selection either promoting mitochondrial DNA introgression or impeding introgression at nuclear loci). They conclude that only positive selection on the introgressed mtDNA could produce massive mtDNA introgression with low levels of nuclear gene flow. The apparent discordance with the present work can nevertheless be explained by two simple factors. First, Bonnet et al. focused on global introgression frequencies, not only at the invasion front. Mitochondrial DNA introgression in *L. granatensis* is massive at the invasion front (the north) but is not even predominant over the species range. Second, asymmetric gene flow was not considered in a scenario of range invasion, and we show here that it is required to reproduce the mtDNA pattern of introgression. Our results thus suggest that, in our system, selection does not need to be invoked to account for cytonuclear differences in introgression prevalence.

Massive mtDNA introgression and cytonuclear co-evolution

This work suggests that the massive and geographically limited mtDNA introgression may result from drift during species replacement. It may seem surprising that selection does not prevent introgression, because the mitochondrial and nuclear genomes extensively interact to control important metabolic functions. They are therefore expected to co-evolve independently in isolated populations and species, a situation that could lead to incompatibilities of heterospecific combinations (see Burton and Barreto 2012; Burton et al. 2013; Levin et al. 2014; Sloan et al. 2017). For instance, cytonuclear incompatibilities result in increased larval mortality rate in male F2 hybrids of Nasonia vitripennis and Nasonia giraulti (Niehuis et al. 2008; Gibson et al. 2013). In addition, Pritchard & Edmands (2013) found evidence of a progressive decrease of cytonuclear mismatch in hybrid swarm replicates of the copepod Tigriopus *californicus*. We therefore explored the hypothesis that mtDNA introgression resulting from allele surfing was not hampered in *L. granatensis* because it was accompanied by co-introgression of interacting nuclear genes. In fact, such co-introgression, if documented, could result either from compensation of accidental mitochondrial DNA introgression, or from adaptive introgression of the co-adapted gene complexes. Beck et al. (2015) found that the complete replacement of Drosophila santomea mitochondrial genome by that of D. yakuba was followed by preferential introgression of cytochrome c oxidase (COX) nuclear encoded genes. Morales et al. (2016) report a case of massive secondary co-introgression of mitochondrial DNA and a set of interacting nuclear genes in the Australian songbird Eopsaltria australis, and provide evidence for a link to climatic adaptation. Overall, this hypothesis would predict rapid co-evolution of the genes involved, driven either by positive selection or to compensate the potential mutation load accumulating in the fast-evolving, low effective size and non-recombining mitochondrial genome (see Burton and Barreto 2012; Levin et al. 2014; Sloan et al. 2017). This has been documented for instance in Nasonia (Werren et al 2010) and Anguilla species (Gagnaire et al 2012), and also in hares (Amoutzias et al. 2016). We found no significant differences of dN/dS between any of the three sets of mitonuc genes (mitonuc, mitonuc-direct and OXPHOS) and background genes in comparisons between the *L. timidus* and *L. granatensis* lineages. We did not either find evidence for mitonuc genes in whatever category to be more subject to introgression than the background. Nor did we find the set of genes with geographic patterns of introgression similar to mitochondrial DNA to be enriched in mitonuc categories. Therefore, we do not detect any general tendency for mitonuc genes to evolve faster or co-introgress more than average.

However, the absence of an overall signal does not preclude that it exists for a few particular genes. By looking at individual patterns of introgression of mitonuc genes, we identified six with high frequency introgression and 17 with a geographic distribution of introgression resembling that of mtDNA. We identified in two genes two amino-acid differences between the native *granatensis* and *timidus* sequences that are predicted to have a strong functional impact, taking into account the conservation levels of the residues at deep evolutionary scales. The concerned genes are HP and HEBP1, which are involved in preventing oxidative stress (Bertaggia et al. 2014) and removal of toxic heme pathway intermediates (Jacob Blackmon et al. 2002), respectively. Relaxing the RND threshold increased the number of candidate genes to 69 that were found introgressed at high frequencies, and 32 resembling the mtDNA pattern. Of the latter, five genes interact directly with the mitochondria in key functions, related with RNA binding (MRPL2 and MRPL15, MRPS22; Marygold et al. 2007), protein biosynthesis (GARS; He et al. 2011) and the ATP synthesis in the oxidative phosphorylation chain ("ATP5F1"; Carbajo et al. 2005). These genes are thus candidates to have been affected by cytonuclear co-evolution during and after the hybridization events, but the functional relevance of these differences must be addressed in future functional assays.

Incompatibilities impede introgression at local genomic scales

We investigated how traces of historical introgression of *L. timidus* origin are distributed along the genome of *L. granatensis* and found non-random patterns of introgression compatible with the existence of genomic incompatibility factors. Modern speciation research has shown that the establishment of genomic incompatibilities between hybridizing species results in heterogeneous patterns of introgression and isolation along the genome (e.g. Muirhead and Presgraves 2016), notwithstanding recent discussion about the nature of genomic islands of speciation vs. differentiation (Noor and Bennett 2009; Turner and Hahn 2010; Nachman and Payseur 2012; Cruickshank and Hahn 2014).

We found that introgression is significantly reduced for the X-chromosome as compared to the autosomes (Figure 3.4A). This is in line with the frequent observation of a disproportionate effect of the X in the establishment of reproductive isolation (large X-effect; Coyne and Orr 1989), resulting in reduced X-linked admixture (Ellegren et al 2012; Martin et al 2013; Fontaine et al 2014; Sankararaman et al 2016). We also found that introgression prevalence increases from the center of the chromosomes towards their

end (Figure 3.4B). A similar trend was reported in cichlid fishes (Gante et al. 2016). This indicates the presence of incompatibility loci along the genome: loci closer to the chromosome end more likely escape from such incompatibilities with a single recombination event, thus facilitating their introgression (Barton and Bengtsson 1986). This effect can be enhanced by an increase of recombination rates towards chromosome ends, as we could verify in hares based on our population genetic estimates of recombination rates (Annex II - Figure S3.9). In any case, this relationship between recombination and introgression suggests that incompatibility factors reduce introgression at linked or partially linked sites, and that there must be a relatively large number of such loci spread in the genome.

Potential adaptive introgression into L. granatensis

In our geographically explicit demographic and coalescent simulations, we were able to reproduce empirical levels and patterns of nuclear and mitochondrial DNA introgression. However, the empirical data displayed a bulk of 123 genes reaching fixation or quasi-fixation for the foreign allele, a result not obtained in the simulations. These remain as significant introgression outliers even when changing simulation parameters to extreme values, to favor gene flow (Figure 3.3A and Table 3.1), suggesting that their introgression was driven by selection. The incorporation of genetic variants previously adapted in a related species can provide an important competitive advantage, particularly for species colonizing new territories (e.g. Rieseberg et al. 2007). As genomics is now widely used to study patterns of admixture across a variety of biological systems, evidence of adaptive introgression has often been suggested, as in plants (see Goulet et al. 2017 for a review) and several animal species, such as humans (see Racimo et al. 2015), mice (Song et al. 2011; Staubach et al. 2012; Liu et al. 2015), Heliconius butterflies (The Heliconius Genome Consortium 2012; Pardo-Diaz et al. 2012; Zhang et al. 2016) and Anopheles mosquitoes (Clarkson et al. 2014; Fontaine et al. 2014; Norris et al. 2015). Proving introgression to be adaptive continues to be a major challenge (Racimo et al 2016), as introgression alone may lead to patterns that can be interpreted as resulting from selection (e.g. extended LD, shift in allele frequencies). However, situations of extreme introgression frequency, especially if shown to be outliers when accounting for the demographic history of the populations offer compelling evidence that introgression may be driven by selection in some instances (Mendez et al. 2012; Sankararaman et al. 2014; Vernot and Akey 2014).

The possible nature of the selection favoring massive introgression can be interrogated by looking at the known functions of the genes concerned. When analysing the functional context of these introgression outliers, we found signals of enrichment of genes involved in the innate immune response. Although the significance of this signal disappears in the more stringent tests accounting for the hierarchical structure of the GO annotation, these genes are worth closer examination. Adaptive introgression of immune-related genes has been inferred in humans (Mendez et al. 2012, Mendez et al. 2013; Dannemann et al. 2016; Deschamps et al. 2016; Nédélec et al. 2016; Quach et al. 2016; Sams et al. 2016), Anopheles (Mancini et al 2015), the Alpine Ibex (Grossen et al. 2014) and house mice (Hasenkamp et al. 2015; Ullrich et al. 2017). These observations suggest that the invasion of new territories with new pathogenic pressures may be facilitated by the incorporation of adapted genetic variants through introgression. Lagomorphs have been widely used as models to understand host-pathogen co-evolution, because viral diseases have recurrently affected them, as witnessed by endogenous viral sequences embedded in their genome (see van der Loo et al. 2009; Pinheiro et al. 2016). Current viral diseases, such as rabbit hemorrhagic disease (RHDV) and myxomatosis (Myxoma Virus) for rabbits, and the European brown hare syndrome (EBHSV) for hares, strongly affect the Iberian populations. Variants of these viruses may change host-specificity and affect other species, as RHDV2 that affects hares (Camarda et al. 2014; Velarde et al. 2016) or EBHSV that affects American rabbits (Sylvilagus) (Lavazza et al. 2015). Interestingly, one of the genes found here introgressed at high frequencies, interleukin 12B (IL12B), has been implicated in the inflammatory process and immune response to RHDV and Myxoma Virus in rabbits (Neves et al. 2015), and to have adaptively introgressed from Neanderthals to modern humans in Europe (Quach et al. 2016). Multiple studies have shown that innate immune system genes have been recurrently affected by positive selection in the evolution of lagomorphs, including in hares (e.g. Lemos de Matos et al. 2011; Lemos De Matos et al. 2014; de Sousa-Pereira et al. 2016).

Another category of genes introgressed at high frequencies is related with spermatogenesis (ALMS1, ARID4B, SPATA6, SLC9C1, KIAA1109, GMCL1 and NEK1). It is interesting to consider this category in a context of interaction with mtDNA. Natural selection cannot act directly against mitochondrial mutations that impair male-specific functions such as spermatogenesis but do not impair female functions, because mtDNA is transmitted maternally. Such male harmful mutations can thus increase in frequency by chance, leading to male-biased mutation load in the mitochondrial genome (termed

"mother's curse"; Gemmell et al. 2004), and could only be counterbalanced by compensatory evolution of interacting nuclear genes (Dowling et al. 2008). The disruption of mtDNA and nuclear DNA co-evolved combinations has been shown to affect male fertility in several cases, e.g. in roosters (Froman 2005), drosophila (Yee et al. 2013), seed beetle (Dowling et al. 2007), but also brown hares, *L. europaeus* (Smith et al. 2010). Therefore, even if *L. timidus* mtDNA hampered spermatogenesis in *L. granatensis* background, this could not have prevented its massive introgression into *L. granatensis*, but could have favored the compensatory introgression and spread of *L. timidus* nuclear alleles restoring the function. Most of the alleles we found highly introgressed occur all over the range of *L. granatensis*, even outside the range of mtDNA introgression, suggesting that they are not harmful in *L. granatensis* context and were thus able to spread south. Since the *L. granatensis* alleles cannot spread north due to the prevalence of *L. timidus* mtDNA, it is logical that the introgression outliers be found at high frequencies all over the range. Only studies evaluating the fertility of males with distinct mitochondrial and nuclear backgrounds could help clarify this question. Note that the phenomenon envisioned here could concern any male-specific trait negatively affected by the alien mitochondrial genome.

Conclusions and future prospects

Speciation research has traditionally paid more attention to processes leading to species divergence and isolation. In this respect, our results are in line with several other studies, i.e. reduced admixture of the X-chromosome as compared to the autosomes. We were able to demonstrate the genome-wide positive relationship between recombination and admixture, without relying on the differentiation proxy often used but potentially misleading (Wolf and Ellegren 2017). Altogether, our results indicate the existence of numerous hybrid incompatibilities along the genome, and especially on the X.

However, we were particularly focused on general evolutionary mechanisms that promote admixture between partially reproductively isolated species. We provide evidence, quantitatively evaluated by simulation, that demographic processes accompanying invasive replacement of one species by the other, with male-biased migration, can determine introgression patterns genome-wide, including strong cytonuclear discordance of admixture levels. This provides an important general neutral

framework under which numerous instances of cytonuclear introgression discordance (revised e.g. in Toews et al. 2012) can be interpreted.

Having set this framework, we could pinpoint outlier genes candidate for selection-driven introgression. Although further phenotypic analyses will be necessary to confirm these candidates, some genes concerned have suggestive functions. For innate immunity genes, adaptation to the environment would be an obvious cause of positive selection. For the other category – spermatogenesis genes – we favor the hypothesis of a role for genetic conflicts, thus having nothing to do with the environment. It is interesting to note that under this hypothesis, mitochondrial massive introgression would be neutral, only driven by demographic and behavioral processes, despite being harmful to males, and thus to the impacted species. Since massive introgression of *L. timidus* mtDNA also occurred into two other species in Iberia (*L. europaeus* and *L. castroviejoi*), it will be possible to conduct a comparative analysis that might reinforce some of the hypotheses put forth here, and especially contribute to evaluating the roles of adaptation and genetic conflicts in driving introgression.

5. Methods

Sampling, genomic DNA Extraction, library construction and sequencing

We performed whole genome sequencing of 10 Iberian hares (*L. granatensis*) and 3 Mountain hares (*L. timidus*), the geographical origins of which are shown on Figure 3.1A and 3.1B, as well as one snowshoe hare (*L. americanus*) – Annex II Table S3.1. Samples were obtained during the hunting season. We used the JETquick Tissue DNA Spin Kit (GENOMED) to extract genomic DNA from ear or internal organ tissues that had been preserved in RNAlater or ethanol. Illumina TruSeq DNA genomic libraries were prepared for the 14 samples and pair-end sequenced (2x100bp) on an Illumina HiSeq 2500 platform at The Genome Analysis Centre (TGAC, Norwich, now Earlham Institute). We also used 30.7 Gb of further sequence data previously generated for the same *L. americanus* individual (Carneiro et al. 2014).

Data Quality Control and Filtering

Raw sequence reads were filtered by removing the first 5 bp using Cutadapt version 1.8 (Martin 2011). Low quality bases were removed using Trimmomatic v0.33 (Bolger et al. 2014) by trimming bases with a quality score lower than 20 at the end of the reads (TRAILING:20) and using a sliding window of 4bp for a minimum average quality of 30 (SLIDINGWINDOW:4:30). Reads shorter than 36 bp were discarded.

Read Mapping, Genotype Calling and Iterative Mapping

Trimmed reads were mapped to the rabbit reference genome available from Ensembl (OryCun2.0, release 80) using the BWA-MEM algorithm with default parameters (Li and Durbin 2009). Samtools v1.3 (Li et al. 2009) "fixmate" and "sort" modules were then used to correct read pairing information and flags and to sort mapped reads by coordinate. Soft clipped bases were further removed using the "removeclipping" module from NGSutils version 0.5.7 (Breese and Liu 2013). Reads were then realigned around INDELs using the Genome Analysis Toolkit (GATK v3.2-2, McKenna et al., 2010; DePristo et al., 2011). Finally, Picard Markduplicates (http://broadinstitute.github.io/picard/) was used to remove read duplicates.

Multi-sample SNP/genotype calling was carried out using the algorithm implemented in Samtools v1.3 for each species independently, requiring minimum base and mapping qualities of 20. Species VCF files were then merged and genotypes filtered

using a minimum site quality (QUAL) of 20, RMS minimum mapping quality (MQ) of 20, minimum individual coverage (FMT/DP) of 8X and maximum overall coverage (DP) of 430X. For variable sites, a minimum genotype quality (FMT/GQ) of 20 was required. All sites failing any of the filtering criteria were coded as missing data. Furthermore, genotypes closer than 10 bp from INDELs were excluded.

In order to improve mapping efficiency, we used the first round of mapping and SNP call to build a pseudo-hare reference genome, by replacing the base in the rabbit reference by that observed in hares at all positions where the latter were fixed for an non-reference state. We used the resulting pseudo-reference to redo the mapping and SNP calling steps. Insertion-deletions were not considered in the processes and original coordinates were thus kept. This iterative mapping procedure has been shown to improve mapping efficiency when using a divergent reference genome (Halligan et al. 2013; Sarver et al. 2017) (5% in this case).

Haplotype Phasing

We used SHAPEITv2.r837 (Delaneau et al. 2012) to perform read-aware phasing, including both *L. granatensis* and *L. timidus* specimens, as we were particularly interested in phasing introgressed regions. Phase informative reads (PIRs), i.e. those that span at least 2 heterozygous sites and thus help local phasing (Delaneau et al. 2013), were extracted from the individual bam files, and phasing was performed using only bi-allelic sites with no more than two individuals with missing information. We ran SHAPEIT for each chromosome using a window size of 0.5Mb (as recommended in the manual) with an MCMC run of 50 main iterations, with 10 burn-in and 10 pruning iterations. We specified an effective population size of 100,000, following the estimates derived in the present paper and by Melo-Ferreira et al. (2012), and a recombination rate of 1 cM/Mb, as inferred for rabbits (Chantry-Darmon et al. 2006).

Estimates of mutation

We estimated mutation rate (μ) based on the sequence divergence between *L. americanus* and rabbit assuming μ=DXY/(2TD+4Ne) (Kimura 1983), where DXY (Nei 1987) is the distance between hares and rabbits averaged across autosomes, TD is the time of divergence (11.8 My, following Matthee et al. 2004), and Ne the ancestral effective population size. We assumed a generation time of 2 years (Marboutin & Peroux

1995) and both a small (10'000) and large (1'000'000) ancestral effective population size value.

Inference of introgression – Efficient Local Ancestry Inference (ELAI)

In order to infer genomic segments of *L. timidus* origin introgressed in *L. granatensis* we first used the Efficient Local Ancestry Inference (ELAI) method (Guan 2014). This method implements a two-layer HMM (hidden Markov model) to infer local ancestry of admixed individuals without prior definition of window sizes, by looking at two layers of linkage-disequilibrium – within and among defined groups. It returns at each variable position in the genome the most likely proportions of ancestries (true values being expected to take values 0, 1 or 2 in two-way admixture). We ran ELAI on the unphased dataset and two population samples: *L. granatensis* defined as the admixed population, and *L. timidus* defined as one of the donors in the admixture. We did not have a pure *L. granatensis* population and therefore let ELAI infer this second ancestry from the data of the admixed population. We set the number of upper-layer groups to 2, representing *L. timidus* and *L. granatensis*, and that of lower-layer clusters to 10 (5 times the number of upper-layer clusters, as recommended). We performed three different Expectation Maximization (EM) runs of 20 steps with mixture generation values of 5,000, 10,000 and 20,000 and different random seeds. ELAI results were averaged over the three independent runs, and sites with a proportion of *L. timidus* ancestry of at least 0.8 were considered as heterozygous for introgression while homozygous for introgression if above 1.8.

Inference of introgression – Relative Node Depth (RND)

In order to infer introgression along the genome, we also analyzed variations of the relative node depth (RND; Feder et al. 2005) because its power to detect introgression does not depend on introgression frequency. This is thus complementary to ELAI, which is not expected to infer properly high frequencies of *L. timidus* ancestry in *L. granatensis* in the absence of a pure *L. granatensis* reference population. Using mvftools (Pease and Rosenzweig 2015) and custom R scripts, we calculated RND from the phased data on non-overlapping windows of 10kb, 20kb or 50kb. We calculated for each *L. granatensis* haplotype its average nucleotide divergence (DXY) (Nei 1987) to all *L. timidus* haplotypes, that we divided by the divergence between *L. timidus* and *L.*

americanus, in order to standardize for potential variations of mutation rates across windows. We then determined the minimum of such values (RNDmin) among haplotypes in each window. Only windows with a minimum of 50 differences between *L. timidus* and *L. americanus* or between *L. timidus* and *L. granatensis* were retained.

Introgression events (whatever their frequency) are expected to produce exceptionally low RNDmin values, but defining thresholds based on empirical distributions can be arbitrary. Therefore, we used ELAI inferences as reference to perform power and false discovery rate (FDR) analyses of the RNDmin method. However, we restricted these analyses to introgression frequencies lower than the highest introgression frequency detected by ELAI (13/20), above which ELAI is expected to perform poorly in our case. RND windows embedded in an ELAI introgression segment were recorded as truly introgressed and as truly non-introgressed if not overlapping an ELAI segment, while RND windows only partially overlapping ELAI segments were not considered. On this basis, we could estimate the FDR and power for the detection of introgression as a function of the RNDmin threshold, and in most downstream analyses, we chose a value predicting a FDR of 10% (Annex II - Figure S3.1), but in some analyses we relaxed this threshold (see results).

Dating Introgression

In order to infer the age of introgression we first used an approach based on Identical by State (IBS) tracts of DNA shared within and between populations (Harris and Nielsen 2013). We used the phased dataset for the 10 *L. granatensis* individuals, and the two *L. timidus* individuals sampled in the Alps, to minimize potential effects of substructure within our geographically widespread *L. timidus* sample (Figure 3.1B). Only sites segregating in this subset were considered. Furthermore, sites with missing genotypes in *L. timidus* or more than 40% missing genotypes in *L. granatensis* were removed. We generated sets of IBS tracts shared within *L. granatensis*, within *L. timidus* and between the species for the 21 autosomes. We excluded un-annotated scaffolds and regions of low SNP density (centromeric regions, regions with more than 10,000 consecutive 'N' bases in the reference genome or regions between SNPs that are 5,000 bp or more apart), in order to avoid erroneously inferring large IBS tracts that span these regions. IBS tracts shared between haplotypes from the same species are informative about the species demographic history while IBS tracts shared between species are informative about their divergence times and the fraction and timing of past genetic

exchanges. We inferred demographic parameters under several demographic models, considering 1 or 4 pulses of introgression, and either constant or variable population size (Annex II - Table S3.2). IBS tract length distributions within species and between species were computed and jointly fit to the observed data. In order to improve computation time and numeric stability, we binned the IBS tract length data by computing the expected abundance of tracts between (3/2)n and (3/2)n+1 bp. We further excluded IBS tracts shorter than 300 bp, following Liu et al. (2014b) .

We also estimated introgression time from the distribution of introgression tract lengths, as inferred with ELAI for the 10 *L. granatensis* genomes, assuming that the distribution is exponential with mean 1/rt, where t is the number of generations since the admixture event and r is the recombination rate per base pair (Pool and Nielsen 2009). We considered a generation time of 2 years and used estimates of recombination rate in rabbits (r = 1.0 x 10-8; Chantry-Darmon et al. 2006).

Long-term demographic profiling of the species

We inferred the long term demographic histories of *L. granatensis* and *L. timidus* with the Pairwise Sequentially Markovian Coalescent (PSMC) method (Li and Durbin 2011), applied to the diploid genome sequence of each individual. Individuals' diploid consensus sequences were generated for each autosome with Samtools v1.3 mpileup, requiring a minimum base and mapping qualities of 20, and coverage between 8 and 50X. Generation time was set to 2 years per generation and the mutation rate (μ) to 3.3 x 10-9 or 2.8 x 10-9 substitutions/site/generation, estimated as described above. The atomic time intervals were set 4+50*2+2+4, meaning the first parameter spans the first 4 atomic intervals, each of the next 50 parameters spans 2 atomic intervals while the last 2 parameters span 2 and 4 atomic intervals respectively.

Principal Component Analysis

We explored population structure within *L. granatensis* with principal component analysis (PCA), as implemented in PLINK 1.9 (Purcell et al. 2007; Chang et al. 2015), and based on a subsample of bi-allelic SNPs at least 50kb apart and without missing genotypes. The PCA analysis was performed either on *L. granatensis* alone, or together with a *L. timidus* or a *L. americanus* individual in order to see the effects of introgression into *L. granatensis* (expected to have occurred from the former but not the latter) on

intraspecific structure. For the same purpose, we also performed PCA on subsets of SNPs outside introgressed regions, as inferred by ELAI.

Geographic distribution of introgression proportions

We estimated the proportion of introgression from *L. timidus* into the genome of each sequenced *L. granatensis* specimen. This was defined as the fraction of the genome length within introgressed segments inferred by ELAI, while for RND this was the average across the genome of the introgression value (0, 1 or 2) of each unit of observation (windows). We then tested the correlation between the level of introgression and the geographic distance to the southernmost sample locality, using the Spearman's rank correlation test.

Spatially explicit coalescent simulations of demographic expansion and introgression

Using the spatially-explicit coalescent simulator SPLATCHE2 (Ray et al. 2010), we simulated the presumed history of the interaction between *L. timidus* and *L. granatensis*. The Iberian Peninsula was subdivided in demes of 50x50 km, and *L. granatensis* was simulated to expand from a deme located in southwest Portugal (as suggested by Marques et al. 2017) 20,000 years ago, progressively replacing the resident *L. timidus* in the northern half of Iberia. The range of *L. timidus* in the Northern demes was determined by a probability of presence higher than 0.8 at the last glacial maximum, as determined by ecological niche modelling (Acevedo et al. 2015). All simulations were performed using a density-independent competition model (model 6) in two layers (as used in Currat et al. 2008), corresponding to the two species, and implied the complete replacement of *L. granatensis* by *L. timidus* at the time of sampling. Admixture between layers was allowed in co-occupied demes. As in Currat et al. (2008), the intrinsic growth rate was set to a fixed value (0.5 in this work) and different carrying capacities, migration rates and admixture rates values were tested, totaling 8 combinations of parameter values. Two values of deme carrying capacity (K) of *L. granatensis* were considered, K=1000 and K=10000. The first corresponds to an inferred effective population size of ~100,000 (this work and Melo-Ferreira et al. 2012) divided by the ~200 demes where the species exists in Iberia. The second value of K used increases 10 times the estimates of effective population size, to evaluate the influence of this parameter on proportions of introgression. During the replacement, the carrying

capacity of *L. timidus* was considered half of that for *L. granatensis*. Two migration rates between adjacent demes were tested – M=0.02 and M=0.2 – and bidirectional admixture at two distinct rates was assumed – gamma=0.005 and gamma=0.03. Larger carrying capacities and admixture rates, and lower migration rates were expected to result in higher levels of introgression (see Currat et al. 2008). We simulated 100 replicates of genomic introgression (demographic and coalescent simulations) per set of parameter values, each corresponding to 51,247 independent markers, mimicking the total number of 50k windows used for the RND-based estimates. We recorded for each simulated marker the proportion of introgression in each of 10 *L. granatensis* simulated individuals, located in the demes corresponding to the geographical locations of the empirical samples.

We have also simulated the same demographic scenario but adjusting parameters to represent commonly invoked causes of massive mtDNA introgression: (i) carrying capacity (K) was reduced to ¼ of that of the nuclear genome (250 and 125 for *L. granatensis* and *L. timidus*, respectively); (ii) inter-deme migration was set to the minimum (M=0.005); and (iii) gene flow was set to be predominant from *L. timidus* into *L. granatensis* (A=0.025 from *L. timidus* to *L. granatensis* and 0.001 in the other direction). An intrinsic growth rate of 0.5 was maintained. We simulated 10000 replicates each with only one marker per simulation.

Gene Ontology (GO) enrichment analyses

We tested for functional enrichment of genes with high introgression frequencies using the g:Profiler R package (Reimand et al. 2016, 2007). Categories with less than 5 genes were excluded and either the Benjamini-Hochberg correction for multiple testing or the Set Counts and Sizes (g:SCS) was applied. While the first assumes GO terms to be independent, the latter takes into consideration the non-independence of GO terms due to the hierarchical nature of the GO annotation (Reimand et al. 2007). Only genes within or overlapping RND windows with more than 50 informative sites were considered for the background list of genes. We used both the rabbit GO term annotation and the more complete mouse one. For the latter, only one-to-one rabbit to mouse orthologous genes were considered. Finally, we summarized GO terms using REVIGO (Supek et al. 2011).

Relationship between chromosome position and introgression.

We tested the correlation of introgression and recombination with position along the chromosomes, expressed either by the relative distance to the centromere or to the chromosome centre. The population-scaled recombination rate coefficient ρ was estimated along the *L. granatensis* genome using the reversible-jump MCMC algorithm interval implemented in LDhat v2.2 (McVean et al. 2002; Auton and McVean 2007). The method fits a uniform recombination rate over a region from patterns of linkage disequilibrium across genotypes. We selected only variable sites without missing information with VCFtools (0.1.15, Danecek et al, 2011) to create LDhat input files. We calculated ρ along the chromosomes in segments of up to 2000 variable sites, as recommended for the method. The interval algorithm was run for 1,000,000 iterations, sampling every 5,000 iterations, discarding the first 10% as burn-in. We specified a block penalty of 5 in all analyses. We then attributed to each SNP the ρ value of the LDhat fragment in which it was included. Introgression prevalence at a given SNP position in the genome was measured as the introgression frequency of the RND window the SNP belonged to, or as the number of ELAI introgressed fragments across individuals overlapping that SNP. To ensure independence, we subsampled SNPs that were at least 50kb apart. The relative distance of a SNP to either the centromere or the chromosome center was calculated by dividing the distance to this reference point (in base pairs) by the length of the chromosome arm or chromosome length, respectively. We excluded rabbit chromosomes 1 and 2 from these analyses given their known structural differences between rabbit and hare (both are split in hares; Robinson et al. 2002). SNPs were grouped by bins of distance, and the prevalence of introgression per bin was calculated as the sum of introgression frequencies across the SNPs, while ρ values per bin were measured as the average of values across SNPs composing the bin. Because ρ estimates could have been affected by introgression, the correlation was also evaluated after excluding SNPs within introgressed fragments, as assessed by ELAI. The correlations were tested with Spearman's rank correlation test.

Introgression in genic and non-genic regions

In order to assess whether genic and non-genic regions were differentially affected by introgression, windows of 10kb, 20kb and 50kb (same windows as for the RND tests of introgression) were annotated as genic if overlapping a protein-coding coding gene annotation and non-genic otherwise. Each window was then classified as

introgressed if having at least one introgressed haplotype as defined by the RND analysis or non-introgressed otherwise. Regarding ELAI-based inferences of introgression, windows were considered introgressed if overlapping an ELAI introgression fragment. In both cases, a bootstrap distribution of proportions of introgression in genic and non-genic regions was obtained from 10,000 replicates of 100 randomly sampled genic and 100 non-genic windows (to avoid non-independence of adjacent windows). We used a Wilcoxon rank sum test to compare introgression prevalence between genic and non-genic windows.

Analyses of nuclear genes with mitochondrial functions

We generated a list of nuclear genes with mitochondrial functions (mitonuc genes) by combining two public databases: InterMitoBase (Gu et al. 2011) and MitoCarta2.0 (Calvo et al. 2016). These databases provide lists of human annotated genes encoding proteins that are present in the mitochondria ("mitonuc" genes). We identified rabbit orthologous genes using the Ensembl Biomart query tool (Kinsella et al. 2011). Of the 708 human annotated nuclear genes in InterMitoBase, 615 were found annotated in the rabbit, while 1030 genes from the 1147 nuclear genes from Mitocarta2.0 were annotated in the rabbit genome. The union of the two databases resulted in 1210 mitonuc genes annotated in rabbit. We further added 1 OXPHOS gene (NDUFA4L2), which was missing from both databases. We also defined a subset of this list by retaining only genes with protein products reported to directly interact with the mitochondrial DNA, or mitochondrially encoded RNAs and proteins (Sloan, Fields, & Havird, 2015). This list, which we call "mitonuc-direct", contained 179 genes among which 154 had an ortholog in the rabbit annotation. This led to a dataset of 188 mitonuc-direct rabbit annotated genes.

Because we were particularly interested in detecting introgression of mitonuc genes, we assessed the power of our introgression detection methods specifically for these genes. Using our real data, we artificially introgressed portions of these genes from *L. timidus* into *L. granatensis*. Several introgression fragment sizes were used – 5Kb, 10Kb, 15Kb, 20Kb, 25Kb and 30Kb – in order to mimic the bulk of introgression tracts inferred with ELAI (mode: 10kb, median: 19kb and mean: 29kb). Introgression fragments smaller than gene size were entirely contained in the gene, and those longer than the gene included the whole gene. The artificially introgressed sequences were taken from a *L. timidus* haplotype and replaced the orthologous sequence in a *L. granatensis*

haplotype. Because we were interested in detecting mitonuc genes introgressed at high frequencies (similarly to mtDNA), we only tested the power of the RND method, presumed to be more powerful than ELAI in such situations. We calculated RND as previously described, in non-overlapping windows of 10kb, 20kb and 50kb, between the artificially introgressed *L. granatensis* haploid genomes and that of *L. timidus* (excluding the one used as source of the introgressed fragment). We expressed the power to detect introgression for each window size as the proportion of mitonuc genes with at least one overlapping RND window where introgression was correctly inferred, for a given RND value. We further summarized introgression detection power as the proportion of mitonuc genes overlapped by at least one introgressed RND window of any given size (10kb, 20kb or 50kb).

From the sets of mitonuc genes, we verified those showing a geographic introgression pattern that could be similar to mtDNA: i) absence of introgression in southern individuals (no mtDNA introgression is found in the south; Melo-Ferreira et al. 2005; Alves et al. 2008); ii) at least 2 introgressed haplotypes in the 5 northernmost samples. At least two introgressed haplotypes would be expected in the north, given the frequencies of mtDNA introgression in the northern populations (see Acevedo et al, 2015 and Annex II - Figure S3.13). For each gene, the window with the highest total frequency of introgression was retained.

In order to examine the potential functional impact of amino-acid differences observed in mitonuc genes, we used the Aligned Sequences tool implemented in SIFT v1.03 (Kumar et al. 2009), available at http://sift.jcvi.org. This method assumes that amino acid changes occurring in a given linage at positions otherwise conserved at a deeper phylogenetic scale likely affect protein function. All available one2one orthologs (all metazoans) for the candidate genes with amino acid changes between introgressed and non-introgressed *L. granatensis* were downloaded from ENSEMBL. The translated protein sequences were aligned with ClustalW v2.0 (Larkin et al. 2007) and the impact of the nonsynonymous mutations between introgressed and non-introgressed *L. granatensis* was assessed. Functional changes were assumed for normalized probabilities of tolerated change ≤0.05.

Gene variation statistics

We produced alignments of all rabbit annotated genes (19280) between the phased genomes of our samples of both hare species. For each gene, we obtained the exon

coordinates of the largest transcript from the Ensembl Biomart query tool. We excluded from the alignments sites with more than two alleles. Alignments including SNPs with allele frequencies markedly deviating from Hardy-Weinberg proportions in either species (exact test p-value < 0.01; using Plink 1.9) were discarded, since this could indicate the presence of paralogs. Sequences with more than 50% missing data were removed from the alignments. Furthermore, haplotypes in *L. granatensis* inferred to be of *L. timidus* origin were excluded from the *L. granatensis* alignment. Sites with less than four haplotypes with information in at least one of the species were masked with Ns. Finally, alignments with less than 100 codons or with premature stop codons were removed. We estimated dN and dS (Jukes-Cantor) (rates of non-synonymous and synonymous substitutions, respectively) between all inter-species pairwise comparisons, using the Bioperl DNAStatistics module (available in http://search.cpan.org/dist/BioPerl/Bio/Align/DNAStatistics.pm). For each gene, dN/dS was calculated as the average of dN/dS pairwise estimates. We also calculated the neutrality index between the two species, as in the PopGenome package implemented in R (Pfeifer et al. 2014). Finally, we estimated πS (per-site synonymous diversity) and πN (per-site non-synonymous diversity) in each species, using the Bioperl DNAStatistics module. Calculations of dN/dS were also performed between *L. americanus* and *L. timidus* haplotypes (either found in *timidus* or *granatensis*).

We also compared dN/dS between mitonuc genes (and subcategories, mitonuc-direct and OXPHOS) and all others. The comparisons were between *timidus* and *granatensis* only (excluding introgressed haplotypes).

6. References

Acevedo P, Melo-Ferreira J, Farelo L, Beltran-Beck B, Real R, Campos R, Alves PC. 2015. Range dynamics driven by Quaternary climate oscillations explain the distribution of introgressed mtDNA of *Lepus timidus* origin in hares from the Iberian Peninsula. J Biogeogr 42: 1727–1735.

Albrechtova J, Albrecht T, Baird SJE, Macholan M, Rudolfsen G, Munclinger P, Tucker PK, Pialek J. 2012. Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse. Proc R Soc B Biol Sci 279: 4803–4810.

Altuna J. 1970. Hallazgo de una liebre artica (*Lepus timidus*) en el yacimiento prehistorico de Urtiga (Guipuzcoa). Munibe 22: 165–168.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P. 2008. The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. Philos Trans R Soc Lond B Biol Sci 363: 2831–2839.

Amoutzias GD, Giannoulis T, Moutou KA, Psarra AMG, Stamatis C, Tsipourlianos A, Mamuris Z. 2016. SNP identification through transcriptome analysis of the european brown hare (*Lepus europaeus*): Cellular energetics and mother's curse. PLoS One 11: 1–17.

Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. Genome Res 17: 1219–1227.

Avril A, Léonard Y, Letty J, Péroux R, Guitton JS, Pontier D. 2011. Natal dispersal of European hare in a high-density population. Mamm Biol 76: 148–156.

Bar-Yaacov D, Hadjivasiliou Z, Levin L, Barshad G, Zarivach R, Bouskila A, Mishmar D. 2015. Mitochondrial Involvement in Vertebrate Speciation? The Case of Mito-nuclear Genetic Divergence in Chameleons. Genome Biol Evol 7: 3322–3336.

Barton NH, Bengtsson BO. 1986. The barrier to genetic exchange between hybridising populations. Heredity 57: 357–376.

Barton NH, Hewitt G. 1985. Analysis of Hybrid Zones. Annu Rev Ecol Evol Syst 16: 113–148.

Beck EA, Thompson AC, Sharbrough J, Brud E, Llopart A. 2015. Gene flow between Drosophila yakuba and Drosophila santomea in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression. Evolution 69: 1973–1986.

Bertaggia E, Scabia G, Dalise S, Lo Verso F, Santini F, Vitti P, Chisari C, Sandri M, Maffei M. 2014. Haptoglobin Is Required to Prevent Oxidative Stress and Muscle Atrophy. PLoS One 9: e100745.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Bonnet T, Leblois R, Rousset F, Crochet P-A. 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution 1–45. http://doi.wiley.com/10.1111/evo.13296.

Bray Y, Devillard S, Marboutin E, Mauvy B, Péroux R. 2007. Natal dispersal of European hare in France. J Zool 273: 426–434.

Breese MR, Liu Y. 2013. NGSUtils: A software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics 29: 494–496.

Burton RS, Barreto FS. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Mol Ecol 21: 4942–4957.

Burton RS, Pereira RJ, Barreto FS. 2013. Cytonuclear Genomic Interactions and Hybrid Breakdown. Annu Rev Ecol Evol Syst 44: 281–302.

Calvo SE, Clauser KR, Mootha VK. 2016. MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res 44: D1251–D1257.

Camarda A, Pugliese N, Cavadini P, Circella E, Capucci L, Caroli A, Legretto M, Mallia E, Lavazza A. 2014. Detection of the new emerging rabbit haemorrhagic disease type 2 virus (RHDV2) in Sicily from rabbit (Oryctolagus cuniculus) and Italian hare (*Lepus corsicanus*). Res Vet Sci 97: 642–645.

Carbajo RJ, Kellas FA, Runswick MJ, Montgomery MG, Walker JE, Neuhaus D. 2005. Structure of the F1-binding Domain of the Stator of Bovine F1Fo-ATPase and How it Binds an α-Subunit. J Mol Biol 351: 824–838.

Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alfoldi J, Barrio AM, Pielberg G, Rafati N, Sayyab S, Turner-Maier J, et al. 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science 345: 1074–1079.

Chan KMA, Levin SA. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution 59: 720–9.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7.

Chantry-Darmon C, Urien C, De Rochambeau H, Allain D, Pena B, Hayes H, Grohs C, Cribiu EP, Deretz-Picoulet S, Larzul C, et al. 2006. A first-generation microsatellite-based integrated genetic and cytogenetic map for the European rabbit (Oryctolagus cuniculus) and localization of angora and albino. Anim Genet 37: 335–341.

Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, Field SG, Webster M, Antão T, MacInnis B, et al. 2014. Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. Nat Commun 5: 4248.

Coyne JA, Orr AH. 1989. Two rules of speciation. In Speciation and its Consequences (ed. S. Associates), pp. 180–207, Sunderland, MA.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol 23: 3133–3157.

Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. Evolution 62: 1908–20.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

Dannemann M, Andrés AM, Kelso J. 2016. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. Am J Hum Genet 98: 22–33.

de Sousa-Pereira P, Abrantes J, Baldauf H-M, Keppler OT, Esteves PJ. 2016. Evolutionary study of leporid CD4 reveals a hotspot of genetic variability within the D2 domain. Immunogenetics 68:477–482.

Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. 2013. Haplotype estimation using sequencing reads. Am J Hum Genet 93: 687–696.

Delaneau O, Marchini J, Zagury J-F. 2012. A linear complexity phasing method for thousands of genomes. Nat Methods 9: 179–81.

DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491–8.

Deschamps M, Laval G, Fagny M, Itan Y, Abel L, Casanova J, Patin E, Quintana-Murci L. 2016. Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. Am J Hum Genet 98: 5–21.

Dowling DK, Friberg U, Lindell J. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. Trends Ecol Evol 23: 546–554.

Dowling DK, Nowostawski AL, Arnqvist G. 2007. Effects of cytoplasmic genes on sperm viability and sperm morphology in a seed beetle: Implications for sperm competition theory? J Evol Biol 20: 358–368.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, et al. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature 491: 756–760.

Excoffier L, Foll M, Petit RJ. 2009. Genetic Consequences of Range Expansions. Annu Rev Ecol Evol Syst 40: 481–501.

Feder JL, Xie X, Rull J, Velez S, Forbes A, Leung B, Dambroski H, Filchak KE, Aluja M. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in Rhagoletis. Proc Natl Acad Sci U S A 102 Suppl: 6573–80.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov I V, Jiang X, Hall AB, Catteruccia F, Kakani E, et al. 2014. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347: 1258524–1258524.

Froman DP. 2005. Sperm Mobility: Phenotype in Roosters (Gallus domesticus) Determined by Mitochondrial Function. Biol Reprod 72: 562–567.

Gagnaire PA, Normandeau E, Bernatchez L. 2012. Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. Mol Biol Evol 29: 2909–2919.

Gante HF, Matschiner M, Malmstrøm M, Jakobsen KS, Jentoft S, Salzburger W. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. Mol Ecol 25: 6143–6161.

Gemmell NJ, Metcalf VJ, Allendorf FW. 2004. Mother's curse: The effect of mtDNA on individual fitness and population viability. Trends Ecol Evol 19: 238–244.

Gibson JD, Niehuis O, Peirson BRE, Cash EI, Gadau J. 2013. Genetic and developmental basis of F2 hybrid breakdown in Nasonia parasitoid wasps. Evolution 67: 2124–32.

Good JM, Hird S, Reid N, Demboski JR, Steppan SJ, Martin-Nims TR, Sullivan J. 2008. Ancient hybridization and mitochondrial capture between two species of chipmunks. Mol Ecol 17: 1313–27.

Good JM, Vanderpool D, Keeble S, Bi K. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. Evolution 69: 1961–1972.

Goulet BE, Roda F, Hopkins R. 2017. Hybridization in Plants: Old Ideas, New Techniques. Plant Physiol 173: 65–78.

Gravel S. 2012. Population Genetics Models of Local Ancestry. Genetics 191: 607–619.

Grossen C, Keller L, Biebach I, Croll D. 2014. Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex ed. M.H. Schierup. PLoS Genet 10: e1004438.

Gu Z, Li J, Gao S, Gong M, Wang J, Xu H, Zhang C, Wang J. 2011. InterMitoBase: an annotated database and analysis platform of protein-protein interactions for human mitochondria. BMC Genomics 12: 335.

Guan Y. 2014. Detecting structure of haplotypes and local ancestry. Genetics 196: 625–42.

Halligan DL, Kousathanas A, Ness RW, Harr B, Eöry L, Keane TM, Adams DJ, Keightley PD. 2013. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. PLoS Genet 9: e1003995.

Harris K, Nielsen R. 2013. Inferring Demographic History from a Spectrum of Shared Haplotype Lengths. PLoS Genet 9: 1–18.

Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. J Hered 105: 795–809.

Hasenkamp N, Solomon T, Tautz D. 2015. Selective sweeps versus introgression - population genetic dynamics of the murine leukemia virus receptor Xpr1 in wild populations of the house mouse (Mus musculus). BMC Evol Biol 15: 248.

He W, Zhang H-M, Chong YE, Guo M, Marshall AG, Yang X-L. 2011. Dispersed disease-causing neomorphic mutations on a single protein promote the same localized conformational opening. Proc Natl Acad Sci 108: 12307–12312.

Hedrick PW. 2013. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol 22: 4606–4618.

Huerta-Sánchez E, DeGiorgio M, Pagani L, Tarekegn A, Ekong R, Antao T, Cardona A, Montgomery HE, Cavalleri GL, Robbins P A., et al. 2013. Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. Mol Biol Evol 30: 1877–1888.

Jacob Blackmon B, Dailey TA, Lianchun X, Dailey HA. 2002. Characterization of a human and mouse tetrapyrrole-binding protein. Arch Biochem Biophys 407: 196–201.

Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database 2011: bar030.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518: 371–375.

Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.

Lavazza A, Cavadini P, Barbieri I, Tizzani P, Pinheiro A, Abrantes J, Esteves PJ, Grilli G, Gioia E, Zanoni M, et al. 2015. Field and experimental data indicate that the eastern cottontail (Sylvilagus floridanus) is susceptible to infection with European brown hare

syndrome (EBHS) virus and not with rabbit haemorrhagic disease (RHD) virus. Vet Res 46: 1–10.

Lemos De Matos A, McFadden G, Esteves PJ. 2014. Evolution of viral sensing RIG-I-like receptor genes in Leporidae genera *Oryctolagus*, *Sylvilagus*, and *Lepus*. Immunogenetics 66: 43–52.

Lemos de Matos A, van der Loo W, Areal H, Lanning DK, Esteves PJ. 2011. Study of Sylvilagus rabbit TRIM5α species-specific domain: how ancient endoviruses could have shaped the antiviral repertoire in Lagomorpha. BMC Evol Biol 11: 294.

Levin L, Blumberg A, Barshad G, Mishmar D. 2014. Mito-nuclear co-evolution: The positive and negative sides of functional ancient mutations. Front Genet 5: 1–11.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–60.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature 475: 493–6.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Liang M, Nielsen R. 2014. The Lengths of Admixture Tracts. Genetics 197: 953–967.

Liu KJ, Dai J, Truong K, Song Y, Kohn MH, Nakhleh L. 2014a. An HMM-Based Comparative Genomic Framework for Detecting Introgression in Eukaryotes. PLoS Comput Biol 10: e1003649.

Liu KJ, Steinberg E, Yozzo A, Song Y, Kohn MH, Nakhleh L. 2015. Interspecific introgressive origin of genomic diversity in the house mouse. Proc Natl Acad Sci 112: 196–201.

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C, et al. 2014b. Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. Cell 157: 785–794.

Macholán M, Baird SJ, Munclinger P, Dufková P, Bímová B, Piálek J. 2008. Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? BMC Evol Biol 8: 271

Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol Evol 20: 229–237.

Mancini E, Spinaci MI, Gordicho V, Caputo B, Pombi M, Vicente JL, Dinis J, Rodrigues A, Petrarca V, Weetman D, et al. 2015. Adaptive potential of hybridization among malaria vectors: Introgression at the immune locus TEP1 between Anopheles coluzzii and A. gambiae in “Far-West” Africa. PLoS One 10: 1–13.

Marques JP, Farelo L, Vilela J, Vanderpool D, Alves PC, Good JM, Boursot P, Melo-Ferreira J. 2017. Range expansion underlies historical introgressive hybridization in the Iberian hare. Sci Rep 7: 40788.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Res 23: 1817–1828.

Martin SH, Davey JW, Jiggins CD. 2014. Evaluating the Use of ABBA-BABA Statistics to Locate Introgressed Loci. Mol Biol Evol 32: 244–257.

Marygold SJ, Roote J, Reuter G, Lambertsson A, Ashburner M, Millburn GH, Harrison PM, Yu Z, Kenmochi N, Kaufman TC, et al. 2007. The ribosomal protein genes and Minute loci of Drosophila melanogaster. Genome Biol 8: R216.

Matthee C, Van Vuuren B, Bell D, Robinson T. 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.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297–1303.

McKenzie JL, Dhillon RS, Schulte PM. 2016. Steep, coincident, and concordant clines in mitochondrial and nuclear-encoded genes in a hybrid zone between subspecies of Atlantic killifish, Fundulus heteroclitus. Ecol Evol 6: 5771–5787.

McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160: 1231–1241.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P. 2009. The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. Mol Ecol 18: 2643–58.

Melo-Ferreira J, Alves PC, Rocha J, Ferrand N, Boursot P. 2011. Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? Evolution 65: 1956–68.

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

Melo-Ferreira J, Boursot P, Randi E, Kryukov A, Suchentrunk F, Ferrand N, Alves PC. 2007. The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. Mol Ecol 16: 605–18.

Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC. 2005. Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. Mol Ecol 14: 2459–64.

Mendez FL, Watkins JC, Hammer MF. 2012. A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am J Hum Genet 91: 265–74.

Mendez FL, Watkins JC, Hammer MF. 2013. Neandertal Origin of Genetic Variation at the Cluster of OAS Immunity Genes. Mol Biol Evol 30: 798–801. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst004.

Mitchell-Jones AJ, Amori G, Bogdanowicz W, Krystufek B, Reijnders PJH, Spitzenberger F, Stubbe M, Thissen JBM, Vohralık V, Zima J. 1999. The atlas of European mammals. T & AD Poyser Ltd, London.

Morales HE, Pavlova A, Amos N, Major R, Bragg J, Kilian A, Greening C, Sunnucks P. 2016. Mitochondrial-nuclear interactions maintain a deep mitochondrial split in the face of nuclear gene flow. bioRxiv.

Muirhead CA, Presgraves DC. 2016. Hybrid Incompatibilities, Local Adaptation, and the Genomic Distribution of Natural Introgression between Species. Am Nat 187: 249–261.

Nachman MW, Payseur BA. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos Trans R Soc B Biol Sci 367: 409–421.

Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, Baxter SW, Blaxter ML, Mallet J, Jiggins CD. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Mol Ecol 22: 814–826.

Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, Grenier J-C, Freiman A, Sams AJ, Hebert S, et al. 2016. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell 167: 657–669.e21.

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

Neves F, Abrantes J, Almeida T, de Matos AL, Costa PP, Esteves PJ. 2015. Genetic characterization of interleukins (IL-1α, IL-1β, IL-2, IL-4, IL-8, IL-10, IL-12A, IL-12B, IL-15 and IL-18) with relevant biological roles in lagomorphs. Innate Immun 21: 787–801.

Niehuis O, Judson AK, Gadau J. 2008. Cytonuclear Genic Incompatibilities Cause Increased Mortality in Male F 2 Hybrids of Nasonia giraulti and N. vitripennis. Genetics 178: 413–426.

Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 104: 418–418.

Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, Lanzaro GC. 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc Natl Acad Sci 112: 815–820.

Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, McMillan WO, Jiggins CD. 2012. Adaptive introgression across species boundaries in Heliconius butterflies. PLoS Genet 8.

Pease JB, Rosenzweig BK. 2015. Encoding data using biological principles: The multisample variant format for phylogenomics and population genomics. IEEE/ACM Trans Comput Biol Bioinforma PP: 1–14.

Petit RJ, Excoffier L. 2009. Gene flow and species delimitation. Trends Ecol Evol 24: 386–393.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An efficient swiss army knife for population genomic analyses in R. Mol Biol Evol 31: 1929–1936.

Pinheiro A, Neves F, Lemos de Matos A, Abrantes J, van der Loo W, Mage R, Esteves PJ. 2016. An overview of the lagomorph immune system and its genetic diversity. Immunogenetics 68: 83–107.

Pool JE, Nielsen R. 2009. Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics 181: 711–719.

Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S. 2009. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet 5: e1000519.

Pritchard VL, Edmands S. 2013. The genomic trajectory of hybrid swarms: Outcomes of repeated crosses between populations of Tigriopus *californicus*. Evolution (N Y) 67: 774–791.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.

Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, Laval G, Patin E, Harmant C, Lopez M, et al. 2016. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell 167: 643–656.e17.

Racimo F, Marnetto D, Huerta-Sánchez E. 2017. Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. Mol Biol Evol 34: 296–317.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. Nat Rev Genet 16: 359–371.

Ray N, Currat M, Foll M, Excoffier L. 2010. SPLATCHE2: A spatially explicit simulation framework for complex demography, genetic admixture and recombination. Bioinformatics 26: 2993–2994.

Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. 2016. g:Profiler—a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res 44: W83–W89.

Reimand J, Kull M, Peterson H, Hansen J, Vilo J. 2007. G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res 35: 193–200.

Rieseberg LH, Kim S-C, Randell RA, Whitney KD, Gross BL, Lexer C, Clay K. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. Genetica 129: 149–165.

Robinson TJ, Yang F, Harrison WR. 2002. Chromosome painting refines the history of genome evolution in hares and rabbits (order Lagomorpha). Cytogenet Genome Res 96: 223–7.

Rosenzweig BK, Pease JB, Besansky NJ, Hahn MW. 2016. Powerful methods for detecting introgressed regions from population genomic data. Mol Ecol 25: 2387–2397.

Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. PLoS Biol 14: 1–22.

Sams AJ, Dumaine A, Nédélec Y, Yotova V, Alfieri C, Tanner JE, Messer PW, Barreiro LB. 2016. Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. Genome Biol 17: 246.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507: 354–7.

Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol 26: 1241–1247.

Sarver BAJ, Keeble S, Cosart T, Tucker PK, Dean MD, Good JM. 2017. Phylogenomic Insights into Mouse Evolution Using a Pseudoreference Approach. Genome Biol Evol 9: 726–739.

Sequeira F, Sodré D, Ferrand N, Bernardi J a R, Sampaio I, Schneider H, Vallinoto M. 2011. Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads Rhinella marina and R. schneideri inferred from mtDNA and nuclear markers. BMC Evol Biol 11: 264.

Sloan DB, Fields PD, Havird JC. 2015. Mitonuclear linkage disequilibrium in human populations. Proc R Soc B Biol Sci 282: 20151704.

Sloan DB, Havird JC, Sharbrough J. 2016. The on-again, off-again relationship between mitochondrial genomes and species boundaries. Mol Ecol 26: 2212–2236.

Smith S, Sandoval-Castellanos E, Lagerholm VK, Napierala H, Sablin M, Von Seth J, Fladerer FA, Germonpré M, Wojtal P, Miller R, et al. 2017. Nonreceding hare lines: genetic continuity since the Late Pleistocene in European mountain hares (*Lepus timidus*). Biol J Linn Soc 120: 891–908.

Smith S, Turbill C, Suchentrunk F. 2010. Introducing mother's curse: Low male fertility associated with an imported mtDNA haplotype in a captive colony of brown hares. Mol Ecol 19: 36–43.

Song Y, Endepols S, Klemann N, Richter D, Matuschka FR, Shih CH, Nachman MW, Kohn MH. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Curr Biol 21: 1296–1301.

Staubach F, Lorenc A, Messer PW, Tang K, Petrov D a., Tautz D. 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (Mus musculus). PLoS Genet 8: e1002891.

Stewart JR, Lister AM, Barnes I, Dalen L. 2010. Refugia revisited: individualistic responses of species in space and time. Proc R Soc B Biol Sci 277: 661–671.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6: e21800.

The Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94–98.

Thulin C, Tegelström H, Tegelstrm H. 2002. Biased geographical distribution of mitochondrial DNA that passed the species barrier from mountain hares to brown hares (genus *Lepus*): an effect of genetic incompatibility and mating behaviour? J Zool 258: 299–306.

Thulin C, Stone J, Tegelström H, Walker CW. 2006. Species assignment and hybrid identification among Scandinavian hares *Lepus europaeus* and *L. timidus*. Wildlife Biol 12: 29–38.

Tigano A, Friesen VL. 2016. Genomics of local adaptation with gene flow. Mol Ecol 25: 2144–2164.

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol 21: 3907–3930.

Turner TL, Hahn MW. 2010. Genomic islands of speciation or genomic islands and speciation? Mol Ecol 19: 848–850.

Ullrich KK, Linnenbrink M, Tautz D. 2017. Introgression patterns between house mouse subspecies and species reveal genomic windows of frequent exchange. http://dx.doi.org/10.1101/168328.

van der Loo W, Abrantes J, Esteves PJ. 2009. Sharing of Endogenous Lentiviral Gene Fragments among Leporid Lineages Separated for More than 12 Million Years. J Virol 83: 2386–2388.

Velarde R, Cavadini P, Neimanis A, Cabezón O, Chiari M, Gaffuri A, Lavín S, Grilli G, Gavier-Widén D, Lavazza A, et al. 2016. Spillover Events of Infection of Brown Hares (*Lepus europaeus*) with Rabbit Haemorrhagic Disease Type 2 Virus (RHDV2) Caused Sporadic Cases of an European Brown Hare Syndrome-Like Disease in Italy and Spain. Transbound Emerg Dis 1–12.

Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343: 1017–21.

Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Beukeboom LW, Desplan C, Elsik CG, Grimmelikhuijzen CJP, et al. 2010. Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species. Science 327: 343–348.

Wolf JBW, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. Nat Rev Genet 18: 87–100.

Wu C-I. 2001. The genic view of the process of speciation. J Evol Biol 14: 851–865.

Yee WKW, Sutton KL, Dowling DK. 2013. In vivo male fertility is affected by naturally occurring mitochondrial haplotypes. Curr Biol 23: R55–R56.

Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR. 2016. Genome-wide introgression among distantly related Heliconius butterfly species. Genome Biol 17: 25.

Genomic exchanges between three hare species sharing the same mitochondrial genome following massive introgression: the roles of history, adaptation and cytonuclear coevolution

Seixas FA, Farelo L, Belkir K, Alves PC, Boursot P, Melo-Ferreira J

1. Abstract

Introgression can be an important source of adaptive variation, and situations of repeated introgression into multiple species are powerful models to detect and understand such introgressions driven by adaptation. In the Iberian Peninsula, northern populations of the brown hare (*L. europaeus*) and the Iberian hare (*L. granatensis*) have been strongly affected by mitochondrial DNA introgression from the locally extinct mountain hare (*L. timidus*), and a genomic study in *L. granatensis* suggested adaptive introgression of several nuclear genes. Here we examine genome-wide introgression in *L. europaeus*, with particular interest into common patterns of introgression in *L. granatensis*, which can provide further evidence of introgression of locally adapted genes. Based on whole genome sequences of 10 newly sequenced *L. europaeus*, 10 previously sequenced *L. granatensis* and four *L. timidus* (1 newly sequenced), we first investigate the complex history of range replacements of these three species using transitions between genomic regions with different ancestries. We infer that the sequence of interspecific contacts first involved the replacement of *L. timidus* by *L. granatensis*, which was then replaced by *L. europaeus* in northern Iberia. Repeated introgression of the *timidus* mitochondrial DNA allowed it to remain in place, depicting the historical distribution of the species in Iberia. Range replacement of *L. granatensis* by *L. europaeus* and allele surfing of the introgressed variants may explain massive *timidus* mtDNA introgression into *L. europaeus* as inferred in *L. granatensis*. We find evidence of massive *timidus* introgression in several nuclear genes in *L. europaeus* but only few are common to *L. granatensis*. However, many of these are involved in similar functions, including genes related with immunity. Our results thus suggest common determinants of introgression in the two Iberian species, which may have facilitated adaptation to a common environment.

2. Introduction

Closely related species can continue to exchange genes long after their initial divergence resulting in semipermeable genomes (Harrison & Larson 2016). The intensity of gene flow varies across genomes is sometimes massive, raising questions about the adaptive nature of introgression. Adaptive introgression is the transfer of genes from one species to another resulting in increased fitness of the introgressed individuals (Anderson 1949). Such introgression has the potential of outpacing the rates of *de novo* mutation and standing variation in providing adaptive variation (Hedrick 2013; Abbott et al. 2013; but see Barton 2013). Also importantly, introgression facilitates the transfer of combinations of allelic combinations (either at the same gene or in different genes – "cassette-like" variation) that have been previously tested by natural selection in the donor species (Rieseberg 2011; Abbott et al. 2013; Suarez-Gonzalez et al. 2016).

There is a growing list of studies suggesting that adaptive introgression is an important evolutionary mechanism in animals, including in Drosophila (Clarkson et al 2014; Fontaine et al. 2015), mice (Song et al. 2011; Liu et al. 2015; Hasenkamp et al. 2015; Ullrich et al. 2017), *Heliconius* (Pardo-Diaz et al. 2012; The Heliconius Genome Consortium 2012; Zhang et al. 2016), Darwin finches (Lamichhaney et al. 2015) and humans (Racimo et al. 2015 for a review). Interpretations of adaptive introgression in these studies are based on extended linkage-disequilibrium (LD) and frequency of the introgressed variants, sometimes in relation with geography. For instance, genomic regions with unusually high frequencies of introgressed alleles are suggestive of adaptive introgression, particularly if these are numerous and uniquely shared between the source population and the population subject to introgression (Racimo et al. 2017). Functional studies evaluating the adaptive fitness effects of the introgressed variants in the recipient backgrounds could provide further and direct evidence of adaptive introgression. However, such studies are difficult to perform in many of the taxa. Few exceptions exist, including the demonstration of resistance to rodenticide acquired through introgression in the house mouse (Song et al. 2011) and adaptive introgression of drought tolerance in *Helianthus* (Whitney et al. 2010). Alternatively, comparative studies of multiple species involved in similar situations of admixture, either sharing the same donor species or inhabiting similar habitats and preferentially both, can give us important information about the selective nature of introgression.

In hares (*Lepus* spp), numerous instances of interspecific introgression have been described, often involving the mountain hare, *L. timidus* (Alves et al. 2008b). In the Iberian Peninsula, populations of three hare species are affected by historical introgression from *L timidus* (Melo-Ferreira et al. 2005), which was present in the region until the Last Glacial Maximum (Altuna 1970). In all three species the mitogenome of *L. timidus* thrives at high frequencies (Melo-Ferreira et al. 2005; Alves et al. 2008b): it is almost fixed in the Iberian range of the brown hare (*L. europaeus*); is absent in the south but very frequent in northern Iberian populations of the Iberian hare (*L. granatensis*); and in the broom hare (*L. castroviejoi*), which is restricted to the Cantabrian Mountains, the native mitogenome can no longer be found as it has been replaced by that of *L. timidus (Alves et al. 2008a; Melo-Ferreira et al. 2012)*. The repeated and massive frequency of *timidus* mtDNA introgression raises questions about its adaptive nature but could also have resulted from neutral demographic processes (Melo-Ferreira et al. 2011). A previous study conducted in *L. granatensis* tested the latter hypothesis to show that massive and geographically restricted mtDNA introgression in this species could be explained by a demographic history of range expansion of *L. granatensis* from a south-west Iberian refugium followed by range replacement and hybridization with *L. timidus* in the north (Seixas et al. submitted). Still, an investigation of introgression patterns along the genome of this species allowed detecting genes with outlier frequencies of introgression, which could have resulted from selection. Among these we find genes potentially co-evolving with the mtDNA (mitonuc genes), which could have resulted from selective pressures to maintain co-adapted cyto-nuclear complexes (see Beck et al. 2015; Pritchard and Edmands 2013; Morales et al. 2016). Other outliers include genes related with: i) spermatogenesis, which could have introgressed as compensation for massive introgression of potentially male harmful *timidus* mitochondria in a *granatensis* nuclear background, and ii) with immune response that if advantageous in a new pathogenic environment could have been adaptively introgressed. A comparative analysis of the patterns of introgression including other species affected by *timidus* introgression could further inform us about the role of natural selection.

In this study, we take advantage of the repeated genetic admixture in northern Iberia to tackle this question, focusing in particular in *L. europaeus*. In Iberian Peninsula *L. europaeus* is restricted to the northernmost part, but it extends its

distribution towards central Europe. This species is thought to have first colonized Central Europe from a refugium in south/central Balkans during the late glacial or early Holocene and later entered Iberian Peninsula (Stamatis et al. 2009). MtDNA introgression from *L. timidus* into *L. europaeus* is not observed outside Iberia except for areas where the two species currently contact (Suchentrunk et al. 2005; Thulin et al. 2006a; Thulin et al. 2006b). In the Iberian Peninsula, *timidus* mtDNA introgression into *L. europaeus* was suggested to result from the invasion of this territory with repeated hybridization along the invasion front with the resident species that harboured *L. timidus* mtDNA at the time (Melo-Ferreira et al 2014a). It is however not clear whether *L. europaeus* replaced *L. timidus* or *L. granatensis* bearing *timidus* mtDNA in northern Iberia.

We analyse the complete genomes of two of the introgressed species in Iberia (*L. europaeus* and *L. granatensis*) and of the donor species (*L. timidus*). We first reconstruct the history of hybridization in northern Iberia by investigating the patterns of introgression between these three species and investigate whether a neutral model could explain massive mtDNA introgression as in *L. granatensis*. Then, we characterize nuclear introgression to, taking advantage of the binary situation of mtDNA introgression between *L. europaeus* populations inside (almost fixed) and outside of Iberia (almost absent), investigate whether massive *timidus* mtDNA introgression in Iberian *L. europaeus* was followed by co-introgression or co-differentiation of some nuclear genes associated with the mitochondria. Furthermore, *timidus* introgression into populations of both *L. europaeus* and *L. granatensis* in Northern Iberian Peninsula could have facilitated the colonization of this region by these two species. The joint analysis of introgression into the two species provides a powerful test to understand how the environment promotes introgression. We thus look for genes undergoing apparent introgression sweeps in Iberian *L. europaeus* and compare to the results obtained previously for *L. granatensis*.

3. Methods

Sampling and Sequencing

Whole genome sequencing data was newly generated for one *L. timidus* sampled in Ireland and 10 *L. europaeus* individuals from Europe with emphasis in the Iberian Peninsula (see detailed geographical origins in Figure 3.5 and Annex III – Table S3.14). These data were put together with whole genome sequencing data of three other *L. timidus* individuals from Scandinavia and Alps, 10 *L. granatensis* from across the Iberian Peninsula and one *L. americanus* (data from Seixas et al. submitted and Carneiro et al. 2014). Genomic DNA was extracted from ear or internal organ tissues, preserved in ethanol or RNA later, using the JETquick Tissue DNA Spin Kit (GENOMED) and treated with RNAse, proteinase K and phosphate buffered saline (PBS) to remove RNA and protein contaminants, following manufacturers' instructions. Illumina TruSeq DNA libraries were prepared for the 10 *L. europaeus* samples and the sequencing of the libraries was performed on the Illumina HiSeq 1500 platform at the NEWGEN sequencing platform at the Research Centre in Biodiversity and Genetic Resources (CIBIO, Vairão, Portugal), generating paired-end sequence data (2x100-125 bp; see Annex III – Table S3.14 for details). Preparation of the Illumina TruSeq DNA library and overlapping paired-end sequencing (2x100bp) of the *L. timidus* individual was performed in The Genome Analysis Centre (TGAC, Norwich, now Earlham Institute).

Data Filtering, Mapping and SNP call

Data quality control and filtering was performed as in (Seixas *et al.* submitted). Filtered reads were mapped to a *Lepus* pseudo-reference generated in (Seixas *et al.* submitted) using the BWA-MEM algorithm implemented in the Burrows-Wheeler Aligner (Li & Durbin 2009), with default parameters. Prior to SNP calling, Samtools v1.3 (Li et al> 2009) 'fixmate' and 'sort' modules were used to correct read pairing information and flags and to sort mapped reads by coordinate, respectively. We further removed soft clipped bases using the 'removeclipping' module from NGSutils version 0.5.7 (Breese & Liu 2013). In order to reduce the number of miscalls of INDELs we've performed realignment of reads around INDELs using the Genome Analysis Toolkit (GATK v3.2-2, McKenna et al., 2010; DePristo et al., 2011). Multi-sample SNP/genotype calling was carried out using the algorithm implemented in Samtools v1.3 for each species independently, requiring a minimum base quality of 20 and minimum mapping quality of

20. Species VCF files were filtered by species using the following criteria. Individual genotypes for variable and invariable sites were retained only for minimum quality (QUAL) of 20, RMS mapping quality (MQ) of 20, individual coverage (FMT/DP) of eight (except in *L. europaeus* in which it was set to 6 due to their lower coverage) and not exceeding 45X (*L. americanus*), 45X (*L. timidus*), 36X (*L. granatensis*), and 24X (*L. europaeus*). Furthermore, sites with species overall coverage exceeding 120X (*L. timidus*), 270X (*L. granatensis*), and 180X (*L. europaeus*) we excluded. For variable sites, a minimum genotype quality (FMT/GQ) of 20 was required. All genotypes failing these parameters were coded as missing data. Furthermore, genotypes distancing less than 10 bp from INDELs were excluded. VCF files were then merged.

Principal Component Analyses

To explore the population structure within *L. europaeus* we performed a principal component analysis (PCA), as implemented in PLINK 1.9 (Purcell et al. 2007; Chang et al. 2015), using a subset of autosomal bi-allelic SNPs with no missing genotypes and 50,000 bp apart, to guarantee their independence (filtering processed in PLINK 1.9). The PCA analysis was repeated using one *L. granatensis*, *L. timidus* or *L. americanus* together with *L. europaeus* in order to interpret intraspecific variation in the axis of interspecific divergence.

Global Detection of Introgression

In order to assess whether gene flow occurred between *L. europaeus* and either *L. granatensis* or *L. timidus* we used the D-statistic (commonly known as ABBA-BABA test; (Green et al. 2010; Durand et al. 2011), implemented in the POPSTATS program (available in https://github.com/pontussk/popstats). This method counts two phylogenetic patterns, ABBA and BABA (A denotes the ancestral and B the derived variants), in the tree (((P1,P2),D),O), where O is the outgroup, D is the putative donor species, and P1 and P2 are two populations of the target species. Under a neutral coalescent model with no gene flow, the two phylogenetic patterns have equal chances to occur. Differential introgression from D into either P1 or P2 is expected to produce a significant increase of either BABA or ABBA counts, respectively. Only autosomal bi-allelic SNPs with no missing genotypes and polymorphic both in P1+P2 and D+O were

considered. The significance of the D-statistic was determined based on Z-scores by performing a jackknife approach of 5 Mb blocks.

Detection of Introgression tracts

In order to infer genomic segments of either *L. timidus* or *L. granatensis* origin introgressed in *L. europaeus* we used the Efficient Local Ancestry Inference (ELAI) method (Guan 2014). By implementing a two-layer HMM (hidden Markov model), that looks at linkage-disequilibrium within and among defined groups, this method infers local ancestry of admixed individuals without prior definition of window sizes. It returns at each variable position in the genome the most likely proportions of ancestries which vary from 0 to 2 (true values being expected to take values of 0 and 2 for homozygous ancestry, or 1 for heterozygous ancestry). We ran ELAI on our unphased dataset considering only bi-allelic sites with no more than 25% of individuals with missing information per population. The number of upper-layer groups, representing *L. timidus*, *L. granatensis and L. europaeus*, was set to 3 and that of lower-layer groups to 10. Three independent ELAI runs with 20 Expectation Maximization (EM) steps were performed, considering different mixture generation values (5,000, 10000, 20000) and different random seeds. The results were averaged over the three independent runs. Finally, for all downstream analyses, we defined each SNP state as from *L. timidus* or *L. granatensis* ancestry state, as non-introgressed (if less than 0.5), heterozygous for introgression (if between 0.5 and 1.5) or homozygous for introgression (if greater than 1.5), unless stated otherwise.

Ancestry tract Junctions – Historical Recombination Points of Introgression

In order to trace the origin of *L. timidus* into *L. europaeus*, we identified and quantified the different types of junctions, characterised by which pair of parental genomes they join in our samples. For the set of SNPs for which ancestry was inferred with ELAI, we defined the ancestry state for each of the three possible ancestries (*L. timidus*, *L. granatensis* and *L. europaeus*): 0 if ancestry below or equal to 0.2, 1 if ancestry between 0.9 and 1.1, and 2 if ancestry equal or above 1.8. Any values outside these ranges was considered of ambiguous ancestry and the SNPs discarded. The reason to use this more stringent criteria to define ancestries was to avoid counting artefact junctions which could result from increased uncertainty in transitions between states. The combination of the three possible ancestry states for each SNP allowed

determining its genotype: native homozygous (homozygous *L. europaeus* ancestry), homozygous introgressed (from either *L. timidus* or *L. granatensis* ancestry) and heterozygous introgressed (with three possible combinations of ancestry: *L. timidus-L. europaeus* (tim-eur), *L. timidus-L. granatensis* (tim-gra) *and L. europaeus-L. granatensis* (eur-gra)). Junctions were considered as transitions between different genotypes. Transitions between SNPs more than 1 kb apart were not considered. The number of *L. europaeus-L. timidus*, *L. europaeus-L. granatensis* and *L. timidus-L. granatensis* junctions were counted, transitions between two homozygous states being counted twice. This analysis was performed for each of the three ELAI replicates independently in order to avoid the increased uncertainty of ancestry estimates in regions of transition resulting from averaging ancestries among replicates.

The date of Introgression

The distribution of introgressed segment lengths as inferred with ELAI was used in order to estimate the dates of introgression, assuming that the distribution is exponential with mean 1/*rt*, where *t* is the number of generations since the admixture event and *r* is the recombination rate per base pair (Pool and Nielsen 2009). We considered a generation time of 2 years and used estimates of recombination rate in rabbits (r = 1.0 x 10-8; Chantry-Darmon et al. 2006).

Geographically structured introgression and differentiation between Iberian and non-Iberian L. europaeus

In order to inspect whether we find genes in the nuclear genome of *L. europaeus* co-introgressing with *L. timidus* mtDNA or with evidence of being swept in Iberian Peninsula possibly as a result of local adaptation, we searched for SNPs with differential *L. timidus* introgression between Iberian and non-Iberian *L. europaeus.* Regions of differential introgression were determined by grouping SNPs less than 10 kb apart, and were discarded if smaller than 500 bp or with a density lower than 5 and higher than 50 SNPs/kb.

To measure differentiation across the genomes of Iberian and non-Iberian *L. europaeus* we calculated Weir and Cockerham's F*st* in VCFtools (Danecek et al. 2011). F*st* values were calculated per SNP and averaged within windows of 200 SNPs with

steps of 20 SNPs, corresponding to an average physical window of *ca.* 39 kb. Only bi-allelic sites with no more than two individuals with missing information were considered. We further removed singleton sites as these can result from technical errors (e.g. sequencing and PCR errors) and bias the estimate. F*st* windows overlapping centromeric regions were discarded. We considered as outliers the windows with F*st* values equal or above the 99.9th percentile of the empirical distribution.

Gene Ontology (GO) enrichment

Statistical analyses of overrepresentation of functions for genes overlapping or within a 10 kb range of regions with differential introgression between Iberian and non-Iberian *L. europaeus* were performed in G:Profiler (Reimand et al. 2007, 2016). Functional categories with less than 5 genes were not considered and either the Benjamini-Hochberg correction for multiple testing or the Set Counts and Sizes (g:SCS) correction was applied. The first assumes GO terms as independent, while the latter takes into consideration the non-independence of GO terms due to the hierarchical nature of the GO annotation (Reimand et al. 2007). The background list of genes were defined as genes within 10 kb distance to SNPs used in ELAI analysis. The same analysis was performed for genes overlapping outlier regions of differentiation between Iberian and non-Iberian *L. europaeus*. Only genes overlapping sampled F_{ST} windows were considered for the background list of genes.

4. Results

Genomic Sampling and Sequencing

We sequenced the genomes of 10 *L. europaeus* individuals across the species European range. Specimens were sampled from within northern Iberian Peninsula, in the Pyrenean foothills where *L. timidus* mitochondrial DNA introgression is frequent, and from Central and Eastern European populations devoid of known *L. timidus* mtDNA introgression (see Figure 3.5 and Annex III – Table S3.14). We further sequenced the genome of one *L. timidus* individual from Ireland. The genomes of three more *L. timidus* specimens (from the and Alps and Scandinavia), 10 *L. granatensis* across the species range in Iberia and 1 *L. americanus* previously sequenced by (Seixas *et al*. submitted) were also included in this study. *L. europaeus* individual samples raw coverage ranged between 13-19X while the sequencing effort for the *L. timidus* individual resulted in 37X raw coverage.

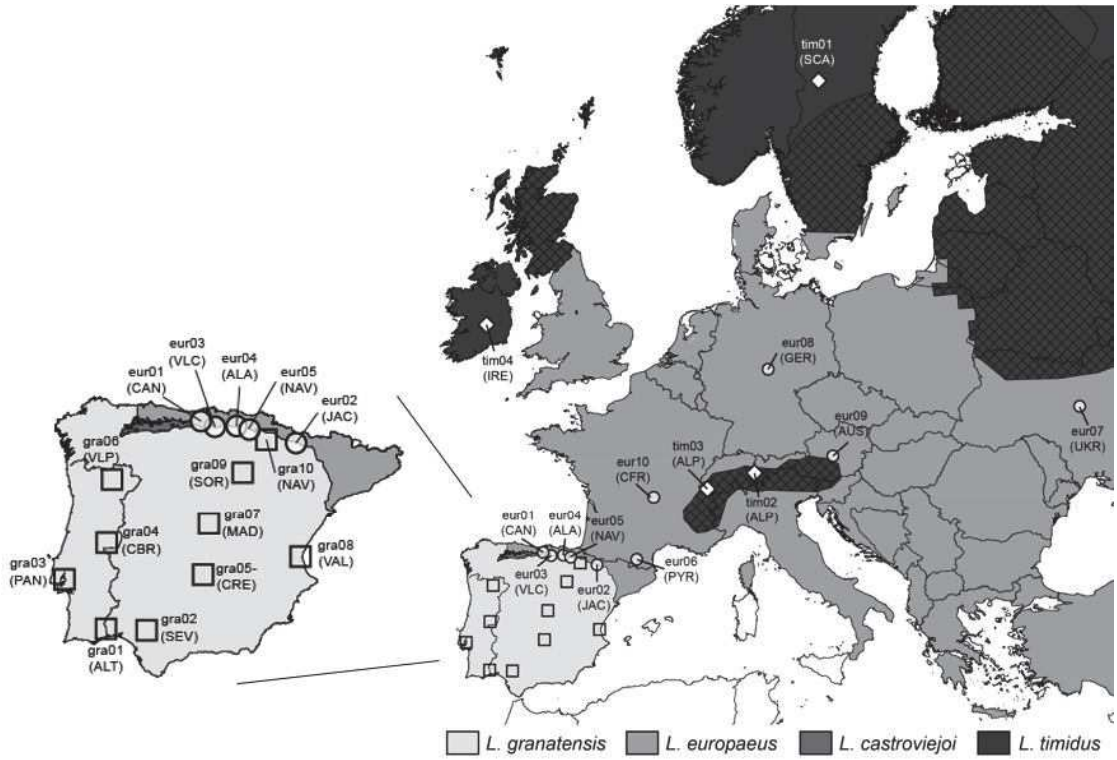


Figure 3.5 Species distribution ranges and sampling locations of the individuals used in this study. Circles - *L. europaeus*; Squares - *L. granatensis*; Diamonds - *L. timidus*.

Genetic Structure and Admixture in L. europaeus

We first analysed the geographic partitioning of *L. europaeus* diversity and introgression. A Principal Component Analysis (PCA) including all *L. europaeus* individuals revealed two main groups based on the first axis of variation: one that includes only the Ukrainian individual and another grouping all the remaining individuals (Annex III – Figure S3.14). This likely reflects the two main lineages known to exist in *L. europaeus*, the Anatolian to which the Ukrainian individual likely belongs, the others belonging to the European lineage (Stamatis et al. 2009), in keeping with its mtDNA lineage. The second axis of differentiation shows a gradient of differentiation within individuals from the European lineage, likely reflecting Isolation by Distance of individuals within the Central Europe/Balkans lineage. Since one a priori possibility is that Iberian *L. europaeus* individuals replaced and hybridized either *L. timidus* or *L. granatensis* (or both) when invading Iberia, we performed a PCA analysis with all *L. europaeus* individuals together with 1 specimen of each of the other species sampled in this study. When considering *L. europaeus* together with the *L. granatensis* individual (southernmost individual thus presumably the least affected by introgression of any source) we found differentiation between individuals from within Iberia Peninsula and individuals from outside Iberia along axis 1 (Annex III – Figure S3.15A). This pattern could result from introgression from *L. granatensis*. On the contrary, when analysing *L. europaeus* together with 1 *L. timidus* individual (from the Alps) differentiation along axis 1 shows that the individual from Germany and in particular that from Ukraine are closer to *L. timidus* again suggestive of introgression particularly affecting these two individuals (Annex III – Figure S3.15B). When *L. europaeus* individuals were analysed together with the *L. americanus* individual, from which no introgression occurred, we find no differentiation along axis 1 (Annex III – Figure S3.15C) thus supporting the hypothesis that introgression is driving differentiation within *L. europaeus* along the axis of species differentiation when analysed together with *L. granatensis* or *L. timidus*.

In order to confirm this interpretation of the PCA, we used the *D*-statistic to detect introgression genome-wide. The analysis using *L. timidus* as the donor and comparing Iberian and non-Iberian *L. europaeus* populations indicates significant introgression into the latter (D = -0.024, Z-score = -6.1; Annex III – Table S3.15). The analysis considering all *L. europaeus* pairwise comparisons suggested this signal is mostly driven by introgression into two individuals (from Ukraine and Germany; Annex III – Table S3.16). When considering *L. granatensis* as donor, introgression mostly affects the Iberian *L.*

europaeus population (*D* = 0.285, Z-score = 31.2, Annex III – Table S3.15). The individual-based analysis shows that all *L. europaeus* specimens from Iberia are significantly affected by *L. granatensis* introgression (Annex III – Table S3.16).

This method relies on the imbalance between ABBA and BABA patterns and thus reveals differential introgression between the two focal populations, but is not able to reveal introgression into both contrasted populations. Therefore, we used ELAI to infer introgression tracts in *L. europaeus,* of either *L. timidus* or *L. granatensis* origin. This method uses patterns of linkage disequilibrium among and between populations to infer local ancestry along the genomes of members of an admixed individual. Here the *L. europaeus* sample was specified as the admixed population. The *granatensis* and *timidus* samples were given as parental populations, but we did not have a pure *europaeus* population to represent the third source of admixture. We therefore let ELAI infer this third source from the data on the admixed population (Annex III – Figure S3. 16A). This is expected to work well if the admixed population is mainly derived from this uncharacterised source, which is likely the case here. On Figure 3.6A we show the estimated overall contributions of the two other species to the genomes of the *europaeus* specimens. Introgression from *granatensis* is substantial in Iberian individuals (5.3-7.9%), and also detectable at a lower level in the sample from the French Pyrenees (0.6%), but absent in all other non-Iberian samples. *L. timidus* contribution was found in all *europaeus* samples at low frequencies (0.6-1.8%), except for the samples from Ukraine (11.7%) and Germany (4.9%).

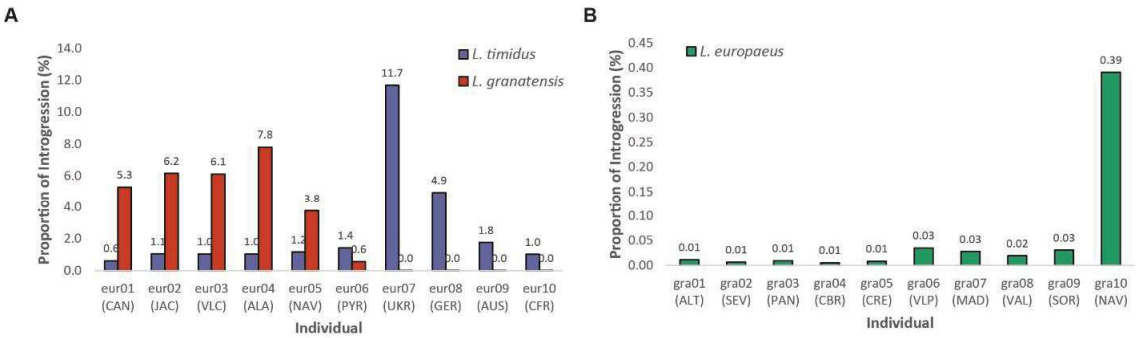


Figure 3.6 Individuals proportions of introgression in (A) *L. europaeus* from either *L. timidus* (blue bars) and *L. granatensis* (red bars) introgression (ELAI setting in Annex III - Figure S3.16A); and (B) in *L. granatensis* from *L. europaeus* (green bars; ELAI setting in Annex III - Figure S3.16B). Proportion of introgression is measured as the proportion of the genome within introgressed segments inferred by ELAI.

Time and origin of Introgression in Northern Iberian Peninsula

The patterns of admixture described above reveal a complex history of three-way hybridization in northern Iberia. We tried to clarify this history of admixture by inspecting the characteristics of the introgression tracts. In particular, we are interested to determine whether the *timidus* contribution in Iberian *europaeus* was acquired before the latter species reached Iberia, directly from *timidus*, or rather in Iberia, then possibly through a *granatensis* intermediate (since *granatensis* is known to be introgressed by *timidus*; Melo-Ferreira et al. 2005, 2009; Seixas et al. submitted). Hereafter for simplicity we will refer to introgression tracts by using abbreviations: tim2eur and gra2eur will designate introgression tracts from *timidus* or *granatensis*, respectively, into *europaeus*.

Table 3.2 Number of introgressed tracts, mean introgression tract length and estimated time of introgression for both *L. timidus* and *L. granatensis* introgression into the 10 *L. europaeus* individuals, as inferred by ELAI (ELAI setting in Annex III - Figure S3.16A).

Ind.	Nb. Introgressed Tracts		Mean Introgression Tract Size		Time Introgression (ya)	
	L. granatensis	*L. timidus*	*L. granatensis*	*L. timidus*	*L. granatensis*	*L. timidus*
Iberian Peninsula						
eur01 (CAN)	1651	499	123035	48822	1626	4096
eur02 (JAC)	1205	685	151111	46680	1324	4284
eur03 (VLC)	1052	620	176777	46499	1131	4301
eur04 (ALA)	1041	623	193381	47195	1034	4238
eur05 (NAV)	480	648	172364	53621	1160	3730
Non-Iberian						
eur06 (PYR)	130	972	145271	44253	1377	4519
eur07 (UKR)	0	3520	-	53554	-	3735
eur08 (GER)	2	2103	-	54994	-	3637
eur09 (AUS)	4	1258	-	38453	-	5201
eur10 (CFR)	3	573	-	51586	-	3877

Table 3.2 shows the number and average sizes of the introgression tracts. The gra2eur tracts are relatively numerous (over 1'000) in most Iberian samples but rarer in one of them (480) and even rarer in the French Pyrenees (130), but absent in all other

samples. In all cases the tracts are relatively long (123-193 kb). The presence of gra2eur in the French Pyrenees could indicate either that *granatensis* used to thrive in this region, and was replaced by *europaeus*, or that gra2eur recently leaked out of Iberia. In comparison, tim2eur are found in all samples, are shorter (46-54 kb) and generally less numerous (below 1'000), except for the Ukrainian and German samples (over 3'000 and 2'000, respectively). For both sources of introgression, there is no clear relationship between the number of tracts and their sizes. Shorter tracts indicate that tim2eur occurred before gra2eur, and we attempted to date these events based on the average tract sizes (Pool and Nielsen 2009; but see Gravel 2012, Liang and Nielsen 2014). We estimate that tim2eur in Iberia result from hybridization 4'000 years ago (Table 3.2). The estimates outside Iberia, although similar, are a bit more heterogeneous, which might correspond to different events given the geographic spread of the samples. Gra2eur is estimated to be substantially more recent (1'000-1'600 years).

Given this temporal frame, the *timidus* contribution in *europaeus* must have been essentially acquired directly from *timidus* rather than being second-hand from *granatensis*. We confirmed this by counting the different types of junctions between segments of different origins. First-hand *timidus* introgression would generate tim-eur junctions, while second-hand would generate many tim-gra junctions in the affected *europaeus* genomes. As expected, we find a majority of tim-eur junctions and very few tim-gra junctions (Figure 3.7). However the latter could still indicate rare instances of second-hand introgression.

Figure 3.7 Counts of junctions between different ancestry states inferred for each of the three ELAI independent runs (REP 1-3). G/E: *L. granatensis-L. europaeus* junction; t/E: *L. timidus-L. europaeus* junction; T/G: *L. timidus-L. granatensis* junction.

Finally, given *L. granatensis* and *L. europaeus* contact in northern Iberia, it is possible that gene flow also occurred from *L. europaeus* into *L. granatensis*. We thus inferred the ancestry of *L. granatensis* by performing an ELAI analysis considering all *L. granatensis* individuals as the admixed population and *L. timidus* and non-Iberian *L. europaeus* as two potential parental populations (Annex III – Figure S3.16B). We specified a third unsampled parental population (thus representing *L. granatensis*) and allowed ELAI to infer its variation from the admixed population. We found that *L. europaeus* introgression in *L. granatensis* was generally absent or rare, with a mean proportion of introgression among all individuals of 0.055%. The most affected individual was the one closest to the contact zone between *L. granatensis* and *L. europaeus,* in Navarra (proportion of introgression 0.39%; Figure 3.6B). Melo-Ferreira et al. (2013) had described, on the basis of microsatellite markers, limited exchanges between the two species close to the contact zone, but not further away. We conclude that nuclear introgression from *europaeus* to *granatensis* is hardly detectable, except very close to the contact zone.

Candidate genes for Cytonuclear Co-evolution and Local Adaptation in Northern Iberia

The mtDNA of *L. timidus* is quasi-fixed while in Iberian *L. europaeus* but absent outside Iberia (Melo-Ferreira et al. 2009). This contrast offers a suitable design to study the genomic impact and correlates of massive mitochondrial introgression. First, we looked for regions of the genome for which introgression from *L. timidus* is frequent in Iberia while rare or absent outside. We found that most of the *L. timidus* introgression in *L. europaeus* occurs at low frequencies, with the majority occurring in only one individual (Annex III - Figure S3.17), and in Iberia the maximum frequency of introgression is 70% (7 haplotypes out of 10 sampled, Annex III – Figure S3.18). We extracted genomic regions showing at least 50% of introgressed alleles in Iberia and less or equal to 20% outside Iberia, which would mimic the mtDNA structure of introgression. We found 40 such regions, harbouring 33 genes Annex III – Table S3.17). Among these genes two have known functions in the mitochondria (mitonuc genes) – BDH1 and MRPL13. The GO enrichment analyses on this set of genes suggests an enrichment in several biological functions including response to mitochondrial depolarization and macromitophagy (Annex III - Table S3.18).

The approach taken above has one potential limitation, because genomic regions with high *timidus* introgression frequencies in *granatensis* (shown to exist by Seixas et al. submitted) could have been misclassified in the ELAI ancestry deconvolution using *granatensis* and *timidus* as parental populations. Therefore some tim2eur tracts could have been missed if lying in a region of high frequency tim2gra. We thus ran ELAI using only Iberian *L. europaeus* as focal population and *L. timidus* and non-Iberian *L. europaeus* as parental populations (Annex III - Figure S3.16C). This design should correctly recover tim2eur with high frequencies in Iberia and low outside Iberia, whatever their status in *granatensis*. Using this approach, tim2eur in Iberia reached frequencies of 90% (Annex III – Figure S3.19). Again we looked at regions with introgression frequencies of at least 50% and found 130 such regions harbouring 85 genes (Annex III – Table S3.19). Within this set of genes we found 4 mitonuc genes (SLC25A30, BDH1, UQCRC2, ATP5L/ATP5L2*; *genes without available name in rabbit and for which the mouse orthologue gene name is given), the latter two belonging to the OXPHOS complexes III and V, respectively. We performed a GO analysis of these 85 genes and found an enrichment in chemokine activity involving four C-C motif chemokine ligand genes (CCL14, CCL15, CCL6* and CCL9; *genes without available name in rabbit and for which the mouse orthologue gene name is given) (Annex III – Table S3.20). However, performing the same analyses leaving only one gene of this cluster, no significant enrichment was found in chemokine activity. Instead, we found an enrichment in meiotic telomere clustering (Annex III - Table S3.20). Interestingly, the two genes producing the signal of enrichment in this category (TERB1 and RAD21L1) are involved in meiosis.

Finally we inspected regions of the genome highly differentiated between Iberian and non-Iberian *L. europaeus*, by calculating *Fst* between these two populations. The average level of genetic differentiation was low (*Fst* = 0.034), facilitating the detection of regions of high genetic differentiation, which could be suggestive of positive selection. We found 103 outlier regions harbouring 91 genes (Annex III – Table S3.21), four of which were mitonuc: GSTK1, SLC25A21, APOOL and MRPL22. The latter (MRPL22) interacts directly with the mitochondria or its products. The GO enrichment analysis of these genes suggested an enrichment in genes involved in Golgi to plasma membrane protein transport (GOPC, SPTBN1 and BLZF1) and in trace-amine receptor activity (TAAR5, ENSOCUG00000024372, ENSOCUG00000026295; cluster of genes in chromosome 12) (Annex III - Table S3.22).

5. Discussion

Cases of massive introgression in nature raise questions about the role of genetic admixture in local adaptation. Comparative studies can give valuable clues about underlying deterministic forces, if patterns of introgression are common to multiple species affected by introgression. In this study, we take advantage of a situation of historical genomic introgression from a boreal species into several species in northern Iberian Peninsula to infer repeated patterns of introgression across species.

The history of colonization and hybridization in northern Iberia

The Northern Iberian Peninsula has been the stage of a complex history of species replacements and admixture, as attested by the presence of *L. timidus* mtDNA in the three species from this region (Melo-Ferreira et al. 2005). Our results and those of Seixas et al. (submitted) now allow a reconstruction of the time and geographic frames of the contacts between the three species. Seixas et al. (submitted) had inferred that *granatensis* replaced *timidus* in the Iberian Peninsula through a south-north invasion wave during which *timidus* mtDNA was captured and reached high frequencies in the northern half of the Peninsula. Using the same method as we used here, based on average tim2gra tract lengths, they estimated the age of hybridization around 7'000 years ago. Based on tim2eur found in Iberia, we estimate the age of the contact to about 4'000 years ago. Unexpectedly, however, we found extensive tim2eur tracts outside Iberia. Although the date of hybridization there is similar to that inferred in Iberia, given the geographical spread of the samples it cannot be considered to represent a single event in biogeographic terms. However, it shows that hybridization outside Iberia occurred pervasively and that the tim2eur tracts in Iberia could have been imported from non-Iberian populations. It seems to be the case since we show that Iberian *europaeus* populations hybridized more recently with *granatensis* (1'000 years ago based on gra2eur tracts) than with *timidus.* Furthermore, we found little evidence that the tim2eur tracts in Iberia are second-hand from *granatensis* given the quasi-absence of gra2tim junctions in Iberian *europaeus*. We can therefore infer a complete scenario of historical species interactions in the region and their genomic consequences. Note that the dates we inferred are probably rough and substantially underestimated because they rely on the mean of the distribution of tract sizes, and our ability to detect short tracts is low, so that the empirical distribution is likely biased towards larger values. This is obvious when inspecting the shape of the empirical distributions, which as compared to the expected

exponentials present a deficit in short sizes (not shown). However, we consider that the order of events that is inferred is robust since the level of age underestimation is expected to increase with age (i.e. with a larger expected proportion of small tracts).

Initially (after the last glacial maximum) *timidus* was present in southern Europe, at least as far south as the northern half of Iberia, in agreement with ecological niche modelling projection into past climate (Acevedo et al. 2015). *L. granatensis* was present in a refugium in SW Iberia, from where it expanded north with climate warming (Marques et al. 2017), replacing *timidus* in the northern half of Iberia and capturing its mtDNA as well as a number of genomic fragments that then spread back south through male migration (Seixas et al. submitted). This process resulted in the extinction of *L. timidus* in Iberia. In the meantime, *europaeus* invaded Western Europe from its eastern refuge, contacting with *timidus* on its way to Iberia, and capturing *timidus* genomic fragments. It imported these fragments into Iberia (as attested by our analysis of the junctions), where it met *granatensis* that it replaced in extreme Northern Iberia, capturing large *granatensis* genomic fragments. Note that *granatensis* could have been present outside Iberia in Southern France at that time, since our sample from the French Pyrenees also shows gra2eur tracts. These tracts could result from secondary leaking out of Iberia, but an on-going analysis of ancient DNA (aDNA) from bones dating back 5.5-7 ky collected in Southern France revealed nuclear genotypes diagnostic of *granatensis*, but mtDNA of *timidus* origin (Melo-Ferreira, unpublished). This shows that the northwards expansion of *L. granatensis* has likely reached Southern France, with individuals that carried *timidus* mtDNA.

It is however not clear from our results whether Iberian *europaeus* captured *timidus* mtDNA from *timidus* before invading Iberia or after, from *granatensis*. The results of Melo-Ferreira et al. (2014a), clearly plead in favour of the second hypothesis, since these authors found less differentiation for mtDNA of *timidus* origin across than along the present contact zone between *europaeus* and *granatensis.* Such pattern demonstrates that the phylogeographic structure of *timidus* mtDNA in Iberian *europaeus* is a remnant of that which existed in *granatensis* at the time of invasion of *europaeus*, witnessed by the proxy of its present structure in *granatensis*. A compatible alternative would be that *timidus* mtDNA was brought into Iberia in the first place by *europaeus,* and then leaked south into *granatensis*, thus explaining both the lack of differentiation across the contact zone mentioned above, and the southward decreasing frequency gradient in *granatensis*. This would however mean that *timidus* mtDNA introgression into *granatensis* occurred after and independently from nuclear introgression in that direction.

This cannot be formally rejected from the available data, but would reveal a selective advantage of *timidus* mtDNA introgression into *granatensis*, since we found that nuclear genome leaking in that direction is hardly detectable. One can however wonder why *timidus* mtDNA would have "waited" so long to selectively sweep into *granatensis* when it had ample previous opportunities to do so during the replacement of *timidus* by *granatensis* in Iberia.

Massive mitochondrial DNA introgression: selective sweep or demographic accident?

Seixas et al. (submitted) have concluded that massive *timidus* mtDNA introgression into *granatensis* could be attributed to a "demographic accident", namely allele surfing on the wave of expansion of *granatensis* into the territory of *europaeus*. They showed that the contrast between geographically limited and massive mtDNA introgression and geographically widespread but low frequency nuclear introgression could be accounted for by the lower effective population size of mtDNA, and by supposing female philopatry and asymmetrical hybridization on the invasion front. Among the patterns sustaining this scenario, they found a gradient of increasing prevalence of nuclear introgression and of increasing sizes of introgression tracts in the direction of expansion (south-north). According to our reconstructed scenario, *L. europaeus* got its *timidus* mtDNA from *granatensis* and while invading the *granatensis* territory, in a presumably east to west direction given *europaeus* distribution in Northern Iberia. We however did not detect a gradient of nuclear gra2eur prevalence or tract lengths along that direction in *europaeus*. The study of Melo-Ferreira et al. (2014a) had demonstrated that repeated hybridization with *granatensis* had occurred along the expansion front of *europaeus* in Iberia, so such a gradient could have been expected.

There can be several explanations to why this gradient is not observed. Male mediated gene flow could be high enough to have homogenized the effects of introgression for the nuclear genome, despite the persistence of a structure for mtDNA, preserved by female philopatry (Melo-Ferreira et al. 2014a). Indeed using microsatellites, Melo-Ferreira et al. (2014a) found little differentiation along the east-west direction in Iberian *europaeus*. Alternatively, the colonization process may have been so rapid that the gradient is not visible (i.e. the time difference between the initial and final contacts are so close that no perceptible difference of tract lengths results). In fact, the gradient reported by Seixas et al. in *granatensis* was only significant in the south, outside of the presumed zone of invasion, and was inferred to result mostly from diffusion of

introgression from the invasion zone into the aboriginal territory. We do not have enough samples from an equivalent territory in *europaeus* (Southern France) to compare. Additionally, it is likely that the connectivity of the invasion territory with the aboriginal territory is not as good in *europaeus,* with the Pyrenees standing in-between, as in *granatensis* where there is no major barrier. Such connections participate in establishing gradients by diluting introgression through fuelling of pure parental genome (Currat et al. 2008; Excoffier et al. 2009). Also, we note that our Iberian *europaeus* samples are geographically relatively close to each other and far from covering the whole area of presumed past interaction between the species, a situation that is not favourable to detect a gradient. And finally, we cannot exclude that invasion of *europaeus* into Iberia took several routes, for instance one along the Mediterranean coast and another along the Atlantic coast, a situation that would result in two gradients in opposing directions, even more difficult to detect.

Although we have not conducted geographically explicit demo-genetic simulations as Seixas et al. (submitted) did for *granatensis*, it seems likely that the same process explains the similar results in the two cases. Therefore massive mtDNA introgression (of *timidus* origin but through a *granatensis* smuggler) and limited nuclear introgression (from *granatensis*) in Iberian *europaeus* presumably also result from the stochastic outcome of invasion with replacement of *L. europaeus* into part of the *L. granatensis* territory, in a context of female philopatry and asymmetrical hybridization.

Adaptive nuclear introgression and cytonuclear co-evolution.

We have tried to identify genomic regions with a pattern of introgression or differentiation similar to those of mitochondrial DNA in *europaeus*, i.e. strong differentiation in Iberia as compared to outside Iberia, or high frequency introgression from *timidus* specifically in Iberia (also resulting in high differentiation). Nuclear genetic differentiation between these two geographic regions is on average limited (F_{ST} = 0.034), and so is the average tim2eur introgression frequency in Iberia (0.92%). Thus, nuclear genes with patterns similar to those of mtDNA would stand as outliers, suggesting that these patterns result from selection favouring either differentiation or preferential introgression in Iberia. We could anticipate two possible sources of selection: adaptation to the environment of the newly colonised area in Iberia, and selection linked to the high prevalence of the alien mitochondrial genome. Such links would result from epistatic interactions between the nuclear and mitochondrial genomes, and some of these outliers

could have evolved from standing variation or by co-introgression of genes involved in independent cyto-nuclear coevolution in the two species. Evidence of cyto-nuclear coevolution have been reported, but they generally result in reproductive isolation rather than in promoting introgression (Bar-Yaacov et al. 2015; McKenzie et al. 2016; Sharbrough et al. 2017). However cases of co-introgression of interacting mitochondrial and nuclear genes have been reported (Pritchard & Edmands 2013; Beck et al. 2015), and in some cases their adaptive nature could be suspected (Beck et al. 2015). Note that if mitochondrial introgression merely resulted from a "demographic accident", as we believe is the case here, co-introgression or differentiation of nuclear mitochondrial genes would not necessarily result from adaptation to the new environment, but could just reflect adaptation to the alien mitochondrial genome.

Inferring selection from outlying patterns of differentiation or introgression poses some difficulties because safely deciding of the outlying status requires some knowledge of the expected patterns and their variance in the absence of selection, which may vary in particular according to population history. To help identify such outliers of introgression in *granatensis*, Seixas et al. (submitted) had conducted extensive geographically explicit demo-genetic simulations taking into account the ecological characteristics of the species and paleo-climatological data. In this study, given the lack of such a null model to define outliers, we relied on the analysis of the function of the genes leveraging on the power of the comparison between the two studies, since finding common outliers could greatly strengthen their validity.

We used various approaches to identify genes either highly introgressed in Iberian *L. europaeus* but not outside Iberia, or strongly differentiated between the two regions, because they were expected to contain those potentially linked to mitochondrial introgression. In the first approach we took (using scheme A of the ELAI design, which is not the most powerful to detect high frequency tim introgression in eur when it co-occurs in gra; Annex III - Figure S3.16A) we found the list of candidate genes to be enriched for terms 'macromitophagy' and 'response to mitochondrial depolarization', driven by a single gene – AMBRA1. Macromitophagy is the process by which damaged or excessive mitochondria are eliminated and thus works as a mechanism to protect the cell from damaged mitochondria and energy homeostasis (Goldman et al. 2010). Interestingly, Melo-Ferreira et al. (2014) inferred a putative cloverleaf structure specific to the control region of arctic mitochondrial lineages (including *L. timidus* and *timidus*-

like mtDNA lineages in Iberian hares). Such structures may influence the efficiency of mtDNA replication and transcription, and may have impacts on the regulation of mtDNA content in cells and thus the efficiency of mitochondrial function (see Melo-Ferreira et al. 2014b and references therein). AMBRA1 could have co-introgressed with mtDNA if it was involved in this process, although this is extremely speculative. In the set of candidate genes of this first approach, we also found an enrichment in the term 'protein ubiquitination' (Annex III – Table S3.18), a mechanism that in the mitochondria is associated with mitochondrial protein quality control (see Taylor and Reuter 2011). Whether there are possible interactions with the mitochondrial genome or its products is however not clear.

When putting together all our attempts to detect candidate genes for cytonuclear coevolution, we found nine genes that are functionally linked with the mitochondria. These included two solute carrier genes (SLC25A21 and SLC25A30), two mitochondrial ribosomal proteins (MRPL13 and MRPL22) as well as BDH1, GSTK1, APOOL, UQCRC2 and ATP5L/L2. Five have products that can be found in the mitochondria (SLC25A21, SLC25A30BDH1, BDH1, GSTK1 and APOOL), two directly interact with the mitochondrial genome or its products (MRPL13 and MRPL22) and the two others are part of the OXPHOS (UQCRC2 and ATP5L/L2) path. These are thus potential candidates to be involved in cytonuclear co-evolution. However, only one of them (MRPL13, involved in mitochondrial translation) was found as candidate for co-introgression in the similar study in *granatensis* (Seixas et al. submitted), but a SNP analysis of this gene over the range of *L. granatensis* did not reveal any association between variation at this gene and mtDNA introgression prevalence (Marques et al. 2017).

Introgression may have been favoured not only as a response to mtDNA introgression, but also to adaptation, particularly in the context of colonisation of new territories. It is therefore interesting to examine the nature of the highly introgressed genes prevailing in the new territory. Scheme C of the ELAI (Annex III – Figure S3.16C) design appears particularly suited since it is expected to detect introgression even if they happened in both *europaeus* and *granatensis*. We found 85 candidate genes using this scheme and their functional analyses revealed an enrichment in the GO term 'chemokine activity'. This partly concerned a cluster of genes in chromosome 19 including CCL5, CCL6*, CCL9* and CCL14 (*CCL6 and CCL9 gene names obtained from mouse

annotation by orthology with the rabbit). Because these genes are clustered in the genome, the significance of the enrichment can be questioned, but it remains that these genes have the introgression pattern we seek, and interestingly chemokines play a crucial role in immune and inflammatory responses (Charo & Ransohoff 2006). Previous studies have shown they evolve under strong purifying selection in three Lagomorph genera, including *Lepus* (de Matos et al. 2014). A detailed inspection to the list of strongly introgressed genes revealed other genes involved in immune response (IL16, FUT8, TNFSF13B). Interestingly, Seixas et al. (submitted) reported enrichment in innate immunity genes among highly introgressed genes from *timidus* into *granatensis*. However, the genes involved were different.

Another functional category we found enriched among strongly introgressed genes was 'meiotic telomere clustering', which included two genes: RAD21L1 and TERB1. Both are involved in the meiotic process, RAD21L1 being required during the initial steps of prophase I in male meiosis and TERB1 involved in attachment of telomeres to membranes also during prophase I. Furthermore, these genes seem to be associated with phenotypes related with infertility in both sexes in mice. For instance male mice lacking of RAD21L1 are infertile while females are fertile but develop sterility with age (Herrán et al. 2011). The disruption of TERB1 results in complete infertility in both sexes (Shibuya et al. 2014). Seixas et al. (submitted) reported an enrichment of genes related with male fertility, but in the case involved in spermatogenesis. This led them to suggest a possible relation with mother's curse processes, where demography-driven massive mtDNA introgression affects male fertility, but cannot be purged by selection because mtDNA is exclusively female-transmitted. Compensatory introgression of nuclear variants would have re-established male fitness. Here however, the phenotypes of the genes in question seem to affect both sexes in mice, and thus a possible association with the mother's curse is not obvious. It cannot however be discarded.

Conclusions and future prospects

Our analysis provided new insights into the history of species contacts and admixture among hares in northern Iberian Peninsula. It revealed that following the replacement of *L. timidus* by *L. granatensis* in this area, this species was then replaced by *L. europaeus* and that their contact likely started in southern France. Early on its way

to Iberian Peninsula however, *L. europaeus* must have hybridized with *L. timidus* since we find evidence of admixture with this species that cannot be accounted by indirect transfer from *L. granatensis*. We could not test the scenario of range replacement with hybridization into Iberia, which could have helped explaining the massive mtDNA introgression in this area, as suggested in *L. granatensis* based on geographic gradients of introgression. However, our geographically limited sampling may have hindered our ability to recover such gradients, which could have also been subtle if the invasion was fast, or erased due to high male dispersal or if the invasion took several routes.

Finally, our analysis of common patterns of massive *L. timidus* introgression in the genomes of *L. europaeus* and *L. granatensis* did not show strong overlap of introgressed genes. However, the functions of some of these genes are remarkably similar. These include genes involved in immunity, the introgression of which may have facilitated the adaptation of the invading species to new pathogenic environments in northern Iberian Peninsula. We may thus be witnessing a remarkable case of convergent adaptive introgression mechanisms. Even if speculative at present, this is an exciting hypothesis that should guide future research.

6. References

Abbott RJ, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Altuna J (1970) Hallazgo de una liebre artica (Lepus *timidus*) en el yacimiento prehistorico de Urtiga (Guipuzcoa). *Munibe*, **22**, 165–168.

Alves PC, Melo-Ferreira J, Branco M *et al.* (2008a) Evidence for genetic similarity of two allopatric European hares (*Lepus corsicanus* and *L. castroviejoi*) inferred from nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, **46**, 1191–7.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P (2008b) The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus Lepus. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 2831–2839.

Anderson E (1949) *Introgressive hybridization*. Wiley & Sons, New York.

Arnold ML (2015) *Divergence with Genetic Exchange*. Oxford University Press, Oxford, UK.

Bar-Yaacov D, Hadjivasiliou Z, Levin L *et al.* (2015) Mitochondrial Involvement in Vertebrate Speciation? The Case of Mito-nuclear Genetic Divergence in Chameleons. *Genome Biology and Evolution*, **7**, 3322–3336.

Barton NH (2013) Does hybridization influence speciation? *Journal of Evolutionary Biology*, **26**, 267–269.

Beck EA, Thompson AC, Sharbrough J, Brud E, Llopart A (2015) Gene flow between Drosophila yakuba and Drosophila santomea in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression. *Evolution*, **69**, 1973–1986.

Breese MR, Liu Y (2013) NGSUtils: A software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics*, **29**, 494–496.

Carneiro M, Rubin C-J, Di Palma F *et al.* (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*, **345**, 1074–1079.

Chantry-Darmon C, Urien C, De Rochambeau H *et al.* (2006) A first-generation

microsatellite-based integrated genetic and cytogenetic map for the European rabbit (Oryctolagus cuniculus) and localization of angora and albino. *Animal Genetics*, **37**, 335–341.

Charo IF, Ransohoff RM (2006) The Many Roles of Chemokines and Chemokine Receptors in Inflammation. *New England Journal of Medicine*, **354**, 610–621.

Clarkson CS, Weetman D, Essandoh J *et al.* (2014) Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. *Nature Communications*, **5**, 4248.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–20.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, **43**, 491–8.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution*, **28**, 2239–2252.

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Fontaine MC, Pease JB, Steele A *et al.* (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, **347**, 1258524–1258524.

Goldman SJ, Taylor R, Zhang Y, Jin S (2010) Autophagy and the degradation of mitochondria. *Mitochondrion*, **10**, 309–315.

Green RE, Krause J, Briggs AW *et al.* (2010) A draft sequence of the Neandertal genome. *Science*, **328**, 710–22.

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, **25**, 2454–2466.

Hasenkamp N, Solomon T, Tautz D (2015) Selective sweeps versus introgression -

population genetic dynamics of the murine leukemia virus receptor Xpr1 in wild populations of the house mouse (Mus musculus). *BMC Evolutionary Biology*, **15**, 248.

Hedrick PW (2013) Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606–4618.

Herrán Y, Gutiérrez-Caballero C, Sánchez-Martín M *et al.* (2011) The cohesin subunit RAD21L functions in meiotic synapsis and exhibits sexual dimorphism in fertility. *The EMBO journal*, **30**, 3091–3105.

Lamichhaney S, Berglund J, Almén MS *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371–375.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **25**, 1754–60.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Liu KJ, Steinberg E, Yozzo A *et al.* (2015) Interspecific introgressive origin of genomic diversity in the house mouse. *Proceedings of the National Academy of Sciences*, **112**, 196–201.

Marques JP, Farelo L, Vilela J *et al.* (2017) Range expansion underlies historical introgressive hybridization in the Iberian hare. *Scientific Reports*, **7**, 40788.

McKenna A, Hanna M, Banks E *et al.* (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

McKenzie JL, Dhillon RS, Schulte PM (2016) Steep, coincident, and concordant clines in mitochondrial and nuclear-encoded genes in a hybrid zone between subspecies of Atlantic killifish, Fundulus heteroclitus. *Ecology and Evolution*, **6**, 5771–5787.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P (2009) The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. *Molecular Ecology*, **18**, 2643–58.

Melo-Ferreira J, Alves PC, Rocha J, Ferrand N, Boursot P (2011) Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? *Evolution*, **65**, 1956–68.

Melo-Ferreira J, Boursot P, Carneiro M *et al.* (2012) Recurrent introgression of mitochondrial DNA among hares (lepus spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology*, **61**, 367–381.

Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (Lepus *timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, **14**, 2459–64.

Melo-Ferreira J, Farelo L, Freitas H *et al.* (2014a) Home-loving boreal hare mitochondria survived several invasions in Iberia: the relative roles of recurrent hybridisation and allele surfing. *Heredity*, **112**, 265–73.

Melo-Ferreira J, Vilela J, Fonseca MM *et al.* (2014b) The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome biology and evolution*, **6**, 886–96.

Morales HE, Pavlova A, Amos N *et al.* (2016) Mitochondrial-nuclear interactions maintain a deep mitochondrial split in the face of nuclear gene flow. *bioRxiv*.

Pardo-Diaz C, Salazar C, Baxter SW *et al.* (2012) Adaptive introgression across species boundaries in Heliconius butterflies. *PLoS Genetics*, **8**.

Pritchard VL, Edmands S (2013) The genomic trajectory of hybrid swarms: outcomes of repeated crosses between populations of Tigriopus *californicus*. *Evolution; international journal of organic evolution*, **67**, 774–91.

Racimo F, Marnetto D, Huerta-Sánchez E (2017) Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. *Molecular biology and evolution*, **34**, 296–317.

Reimand J, Arak T, Adler P *et al.* (2016) g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research*, **44**, W83–W89.

Reimand J, Kull M, Peterson H, Hansen J, Vilo J (2007) G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids*

Research, **35**, 193–200.

Seixas FA, Boursot P, Melo-Ferreira J (2017). The genomic impact of historical hybridization with massive mitochondrial DNA introgression. *Submitted*

Sharbrough J, Havird JC, Noe GR, Warren JM, Sloan DB (2017) *The Mitonuclear Dimension of Neanderthal and Denisovan Ancestry in Modern Human Genomes*.

Shibuya H, Ishiguro K, Watanabe Y (2014) The TRF1-binding protein TERB1 promotes chromosome movement and telomere rigidity in meiosis. *Nature Cell Biology*, **16**, 145–56.

Song Y, Endepols S, Klemann N *et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, **21**, 1296–1301.

Stamatis C, Suchentrunk F, Moutou KA *et al.* (2009) Phylogeography of the brown hare (Lepus *europaeus*) in Europe: A legacy of south-eastern Mediterranean refugia? *Journal of Biogeography*, **36**, 515–528.

Suchentrunk F, Mamuris Z, Stamatis C (2005) Introgressive hybridization in wild living mountain hares (*L. timidus* varronis) and brown hares (*L. europaeus*) and morphological consequences. *Mammalian Biology*, **70**, 39–40.

The Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Thulin C, Fang M, Averianov AO (2006a) Introgression from Lepus *europaeus* to *L. timidus* in Russia revealed by mitochondrial single nucleotide polymorphisms and nuclear microsatellites. *Hereditas*, **143**, 68–76.

Thulin C, Stone J, Tegelström H, Walker CW (2006b) Species assignment and hybrid identification among Scandinavian hares Lepus *europaeus* and *L. timidus*. *Wildlife Biology*, **12**, 29–38.

Ullrich KK, Linnenbrink M, Tautz D (2017) Introgression patterns between house mouse subspecies and species reveal genomic windows of frequent exchange. , **1**.

Whitney KD, Broman KW, Kane NC *et al.* (2015) Quantitative trait locus mapping

identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. *Molecular Ecology*, **24**, 2194–2211.

Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR (2016) Genome-wide introgression among distantly related Heliconius butterfly species. *Genome Biology*, **17**, 25.

Chapter 4.

General discussion

1. The relevance of introgression in the evolutionary history of species
2. Cytonuclear discordance – Causes and Consequences
3. Adaptive introgression
4. Interspecific incompatibilities – variable recombination and the large X-effect
5. Conclusions
6. Final considerations, undergoing work and future directions
7. References

1. The relevance of introgression in the evolutionary history of species

One of the most surprising findings since the implementation of molecular evolutionary genetics is that a considerable number of species hybridize (Mallet 2005) and continue to exchange genetic material for some time after their initial divergence (Pinho and Hey 2010, Roux et al. 2016). This is particularly true for many closely related species pairs, whose genomes are semi-permeable to gene flow with different regions of the genome showing different permeability to gene exchange (Roux et al. 2016; Harrison and Larson 2016). This raises a number of questions of general interest for evolutionary biology that go from the determination of maintenance of reproductive isolation, to the factors promoting differential gene flow along the genomes and notably the significance of introgression for the adaptive ability of species.

Ubiquitous mtDNA introgression?

The very first step towards a better understanding of the relevance of introgression in the evolutionary history of species is to detect and describe the patterns of interspecific gene flow. One pattern recurrently revealed by such studies is the remarkable tendency for the mtDNA being the introgressed marker (see e.g. Toews and Brelsfrod 2012). In this framework, one remarkable example of ubiquitous mitochondrial introgression are hares (*Lepus ssp.*) in which the mitogenome of the arctic species *L. timidus* is found in four southern European hares (*L. granatensis*, *L. europaeus*, *L. castroviejoi* and *L. corsicanus*) and may have introgressed in four other species in Asia. Using coalescent simulations of mtDNA divergence Melo-Ferreira et al. (2012) confirmed that the presence of the *timidus* mitochondrial DNA in European hares resulted from past hybridization and gene-flow with *L. timidus*. Still, mtDNA introgression was also suspected in other species (Alves et al. 2008) including Northern American hares (Cheng et al. 2014). We thus asked whether this pervasiveness nature of mtDNA introgression could be found in other systems or was rather a regional phenomenon, or restricted to particular species. Using simulations of mitochondrial DNA divergence we found in *Paper I* evidence of extensive mtDNA and geographically restricted introgression from *L. californicus* into a group of *L. americanus* populations from the Pacific Northwest region. Interestingly and contrarily to the general observation in Europe, mtDNA introgression occurs from a temperate species to a boreal one. MtDNA introgression is now also found

in other systems that do not involve *L. timidus*, in Asia between *L. capensis* and *L. yarkandensis* (Wu et al. 2011) and Africa between *L. capensis* and *L. europaeus* (Lado et al., unpublished) and possibly three hares from Ethiopia (Tolesa et al. 2017). These results indicate that mtDNA introgression occurs independently of the lineages involved or the local environment and rather a rule than an exception within the genus *Lepus*.

Extent of nuclear introgression.

In order to properly portray and evaluate the frequency and importance of introgression during animal speciation we must have a genome-wide perspective of this phenomenon. For some time this was only possible in a few model organisms for which genomic resources were available, but with the advent of Next Generation Sequencing (NGS) technology we are now able to easily examine the genetic variation across the genomes of many individuals in virtually any species in a cost-effective way and thus appraise the extent of gene flow between hybridizing species.

To understand the importance of historical gene flow in hares, in *Papers II* and *III* we examined the extent of nuclear gene flow between *L. timidus*, *L. granatensis* and *L. europaeus* from whole genome sequences. We show that between 1.3-2.4% of the sampled *L. granatensis* genomes have ancestry in *L. timidus* (*Paper I*), while the genomes of *L. europaeus* harbor proportions of *L. timidus* ancestry that go from 0.6% to 11.7% in some individuals (*Paper II*). We have also found evidence of introgression from *L. granatensis* to *L. europaeus* (0.5-7.8%), while in the opposite direction introgression is much less prevalent (0.01-0.39%) and predominant in the current contact zone (*Paper II*).

These results show historical hybridization with massive mtDNA introgression involved non-negligible interspecific nuclear gene flow, though in much lower frequencies. Also, it suggests that the view of introgression as an important phenomenon in the genus is not only the result of a bias of analysis mostly focused on the mtDNA which can be misleading in some cases (Good et al. 2015). Our results come in line with several others now showing that species can continue to exchange a considerable part of their genomes for some time after initial divergence (Fontaine et al. 2015; Poelstra et al. 2014; Zhang et al. 2016) as is the case of our own species (Green et al. 2010; Reich et al. 2010; Meyer et al. 2012; Prüfer et al. 2014), further supporting the modern view of speciation which accepts the semipermeability of genomes.

2. Cytonuclear discordance – Causes and Consequences

As Toews and Brelsford (2012) suggested it is important to go beyond the description of patterns of cytonuclear discordance and start testing hypotheses regarding the processes that may lead to these patterns (e.g. natural selection, demography). This was one of the major objectives of this thesis and to tackle it we focused on the cases of *L. timidus* introgression into two hares in the Iberian Peninsula, *L. granatensis* and *L. europaeus*, for which these questions have long been studied. In both cases, previous studies have shown massive levels of mtDNA introgression with little nuclear introgression for the very few markers analysed, the results regarding the drivers of discordance have been inconclusive so far. Furthermore, we have investigated the consequences of massive mtDNA introgression on the nuclear genomes of these species.

Testing the null hypothesis – Range replacement and sex-biases explain massive and geographically structured mtDNA introgression with little nuclear introgression

It is now well established that hybridization between species pairs often results in patterns of geographic discordance of introgression between mitochondrial and nuclear loci (cytonuclear discordance), the most common form of cytonuclear discordance being asymmetric mtDNA introgression (Toews and Brelsford 2012). In the majority of cases, the higher prevalence of mtDNA introgression is suggested to result from adaptive mtDNA introgression, demographic disparities and sex-biased asymmetries (in many cases authors presented more than one explanation for discordant patterns in a given system). However in most of these studies, these hypotheses are only based on patterns of biogeographic discordance and thus remain to be tested (Toews and Brelsford 2012). This was one of the major objectives of the thesis: to evaluate the plausibility of demographic range expansions with hybridization to create such cytonuclear discordant patterns.

We show the plausibility of this hypothesis by first demonstrating that geographic patterns of introgression conform to a range expansion of *L. granatensis* from a SW Iberian refugium followed by hybridization with *L. timidus* in the north (*Paper II*). The south-north differentiation in the genetic variation within *L. granatensis* supports the

south-north range expansion. Furthermore, both the proportion of *timidus* introgression and the mean size of introgression tracts increase towards the north, which is compatible with more recent hybridization towards the former range of *timidus* in northern Iberia and the drift of introgressed variants at the front of the invasion to high frequencies as expected in a situation of range replacement. Interestingly, we do not find clear geographical gradients of introgression tract sizes when focusing only in the north, the area where *L. granatensis* is inferred to have replaced *timidus*. The same also applies to *L. europaeus* in which we find a lack of a geographic gradient of introgression in Iberia (Figure 4.1). These results seem to indicate that both invasions were very rapid. However, in *L. granatensis* there is a clear gradient in southern Iberia, suggesting that introgression in this region results from secondary diffusion of introgression tracts from the invasion territory further north.

In *Paper II* we have formally tested the hypothesis that a range invasion of *L. granatensis* into the territory of *L. timidus*, with hybridization between the two species, resulted in both massive and geographically restricted mtDNA and rare and geographically widespread nuclear introgression observed in *L. granatensis*. According to our geographically-explicit demographic simulations we find that indeed under this single demographic scenario the two very distinct patterns can be recovered. However, to be able to reproduce the patterns observed for the mtDNA it was further necessary to consider the reduced effective size of this marker (due to its haploid nature and maternal inheritance) and two assumptions commonly invoked for the ubiquitous nature of mtDNA introgression. The first was female philopatry, migration being male-driven. Although we do not have information directly for *L. granatensis*, telemetry studies in *L. europaeus* have described that males tend to migrate more than females (Bray et al 2007, Avril et al 2011), as reported for many other mammals. It would be important to conduct similar studies in *L. granatensis* to confirm this expectation. The second, involved predominant gene-flow from *L. timidus* to *L. granatensis*, which could result from male-biased dispersal, frequency assortative mating or other behavioral factors. Again, we lack such information regarding the interactions between these two species as there are no current contact zones between the two. However, such asymmetries have been often invoked in contact zones of *L. timidus* and *L. europaeus* (Thulin et al. 2002; 2006).

An important aspect of our results is its potential broader significance. In the studies surveyed by Toews and Brelsford (2012), in 40% of the times demography was invoked to explain cytonuclear discordances while sex-biases were used as an

explanation in 53%. Only in 17% of the cases both demography and sex-biases are proposed (and not necessarily together). Here, we show that in our system the two must be considered together to produce massive cytonuclear discordance. The neutral scenario that we propose could possibly be a major mechanism that explains several instances of cytonuclear discordance described in other study systems, particularly in situations in which range expansions are suspected and when mtDNA introgression is massive and geographically restricted. In fact, in our simulations of mtDNA introgression we never see complete or nearly complete introgression (maximum is 68% and not extending to all populations; not shown). Accordingly, Bonnet et al. (2017) have recently shown that in extreme cases of cytonuclear discordance, i.e. when all individuals or nearly all individuals of one species are introgressed for the mtDNA of another but with little nuclear introgression, the massive discordance is best explained by positive selection on mitochondria. Without selection, other processes such as demography and sex-biases only rarely can result in massive cytonuclear discordance. It thus seems that demographic processes together with sex-biases could explain geographically restricted cytonuclear discordances but that positive selection is needed to create complete or nearly complete mitochondrial DNA replacements when nuclear introgression is limited. These hypotheses must await confirmation from future studies in several other systems, hares included, which could use the framework we provide here to test them.

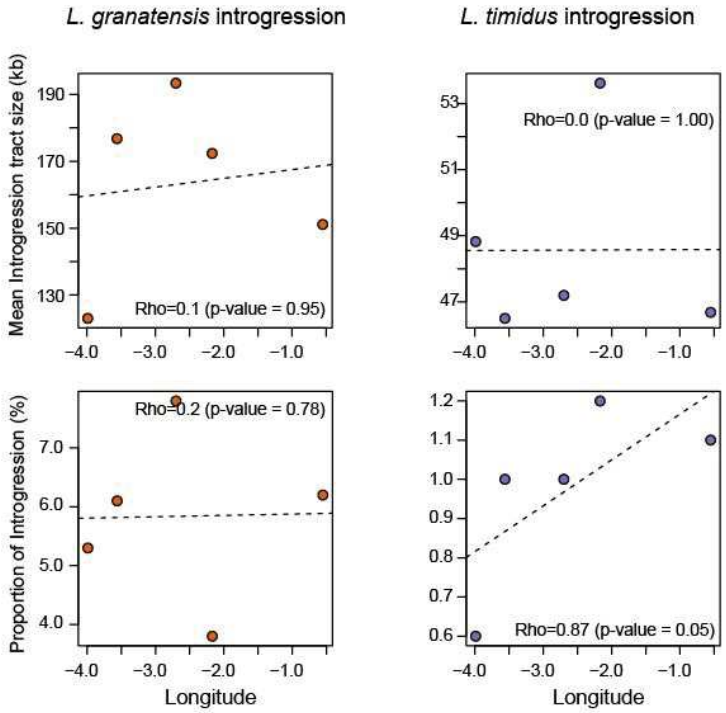


Figure 4.1. Correlation between introgression in *L. europaeus* and geography. For each of the 5 Iberian *L. europaeus* samples, longitude (x axis) is plotted against different characteristics of introgression: mean introgression tract size (top) and observed proportion of the genome introgressed (bottom). Introgression from *L. granatensis* is considered in the left panels, and from *L. timidus* in the right panels. Correlations were tested with Spearman's rank correlation test. Dashed lines represent linear regression trendlines.

Cytonuclear co-evolution in hares?

Situations of massive mitochondrial DNA introgression raise another important question: what is its effect on the nuclear genome and how does the nuclear genome respond? This question results mainly from the fact that these two genomes largely interact in key cellular functions (Burton et al. 2013; Sloan et al. 2017; Wolff et al. 2014) and likely co-evolve Burton et al. 2013; Sloan et al. 2017; Wolff et al. 2014). After a period in which the two genomes co-evolved separately in two sister species, the formation of heterospecific combinations resulting from secondary contact may lead to incompatibilities due to the disruption of co-adapted combinations of mitochondrial and nuclear alleles (see Burton and Barreto 2012; Burton et al. 2013; Levin et al. 2014; Sloan et al. 2016). In fact, cytonuclear incompatibilities are now suspected to play a disproportionate role in the creation of species boundaries (Burton and Barreto 2012; Hill 2016). However and surprisingly, massive mtDNA introgression is observed in many taxa

which may suggest that cytonuclear co-adaptation is not pervasive and that mtDNA of one species can readily function in the nuclear backgrounds of other species (Burton and Barreto 2012). Simple explanations for such phenomenon involve lack of sufficient time or of relevant variation for co-adaptation to develop. Alternatively, it could be that co-introgression of nuclear co-adapted genes (mitonuc genes) could mitigate the deleterious effects of disrupting the cytonuclear combinations (see Burton and Barreto, 2012; Sloan et al., 2016), an hypothesis that is been only tested in less than a handful of cases (Pritchard and Edmands 2013; Beck et al. 2015).

Despite massive *L. timidus* introgression is found both in *L. granatensis* and *L. europaeus*, we found no evidence of preferential co-introgression of mitonuc genes in *L. granatensis* (*Paper II*) and although in *L. europaeus* this could not be properly tested due to the low number of introgressed genes, only a few mitonuc genes were found to co-introgress or co-differentiate with *timidus* mtDNA (*Paper III*). These results suggest that, at least in our system, co-evolution does not seem to be a major mechanism, either promoting or impeding gene flow. This was confirmed in *Paper II*, where we failed to find increased dN/dS ratios in mitonuc genes compared to the remainder of genes in the genome among the species pairs analysed (*L. granatensis* – *L. timidus*, *L. granatensis* – *L. americanus* or *L. timidus* – *L. americanus*) as it would have been expected if the two genomes largely co-evolve (Burton and Barreto 2012; Sloan et al. 2017). Whether the same is true regarding *L. europaeus* still needs to be assessed. Given the rapid and recent radiation of hares, the absence of generalized cytonuclear co-evolution could be interpreted as resulting of lack of time for significant co-evolution to have evolved. Still, it is interesting to note that within *L. europaeus* co-evolution of nuclear and mitochondrial genes of the OXPHOS complexes has been suggested to have resulted in a strong barrier to gene flow between two recently diverged mitochondrial lineages (Amoutzias et al. 2016; Giannoulis et al. 2017) and thus time does not seem to be necessarily a limitation in all cases.

Although we do not find a general pattern for cytonuclear co-introgression, we cannot exclude that co-evolution of the two genomes could be restricted to a few genes. In both *L. granatensis* and *L. europaeus* we find a number of such candidate genes, either co-introgressed at high frequencies with mtDNA (*Paper II* and *III*) or co-differentiated with it in the case of *L. europaeus* (*Paper III*). However, since in both studies the limited sample size is prone to false positives, candidates will need to be confirmed with population level studies.

3. Adaptive introgression

Along with new mutations and standing variation, interspecific gene flow (introgression) can also work as a valuable, and possibly faster and more effective, source of adaptive variation (Abbott et al. 2013; Hedrick 2013). By bypassing the normal waiting time needed for new mutations to arise, introgression provides a faster mean of adaptation by the immediate introduction of genetic novelty in populations (e.g. Grant & Grant, 1994), although this can also be achieved from standing variation or from new mutations provided large effective population sizes exist (see Abbott et al. 2013). Furthermore, introgression has the potential of introducing complex combinations of alleles, involving one or several genes, which have been already previously tested in nature. However, and although introgression is now appreciated as a common phenomenon in nature, one major challenge in the study of introgression continues to be the demonstration of its adaptive significance (Abbott et al. 2013; Hedrick 2013). Providing convincing evidence of adaptive introgression can be challenging as it requires i) identifying both the genes and traits involved, ii) documenting that the variants present in one species at these genes resulted from introgression from another species, and also iii) showing the adaptive significance or fitness effects of the introgressed variants in the recipient species (Rieseberg 2011).

Although providing all this information can be difficult in most cases, a growing number of studies has suggested adaptive introgression at some loci (e.g. Zhang et al. 2016; Liu et al. 2015; Lamichhaney et al. 2015; Norris et al. 2015; also see Racimo et al. 2015 for a review of studies in humans). Many of these studies generally follow the rationale of first detecting introgression regions and then using common statistics to test for positive selection as a potential indicator of its adaptive advantage (but also the opposite, thus detecting introgression in regions previously suggested to be under positive selection; see Racimo et al. 2016, 2017). These statistics generally rely on the observation of linkage disequilibrium (as selection is expected to increase linkage disequilibrium) or shifts in allelic frequencies (e.g. high frequency of a variant of the donor species also present in the recipient populations but not in other related populations). One caveat of such approaches is that introgression itself can change the distribution of allele frequencies and also affects haplotype structure, thus confounding traditional tests for detecting positive selection that rely in these patterns. In particular, the demographic context of species admixture can have particular influence on the frequency of

introgressed variants as it has been shown that during the range replacement of the resident by the invading species, introgressed variants can increase in frequency by drift alone (Currat et al. 2008; Excoffier et al. 2009). A more appropriate setting would thus be to confront situations of extreme frequencies of introgression with null expectations based on simulations of the process of admixture which also account for the demographic (Mendez et al. 2012; Vernot and Akey 2014; Sankararaman et al. 2014) and geographic context of this process as we did in *Paper II*. Based on this approach, which integrated genetic, ecological and paleo-climatological data, we were able to pinpoint several regions of the genome with outlier frequencies of introgression and thus potentially adaptively introgressed, notably involved in genes related with immune processes and spermatogenesis.

Another important source of evidence for adaptive introgression can come from comparative studies of multiple species involved in the admixture phenomenon. In such studies, showing introgression in the same genes or categories of genes may provide strong evidence for an underlying role of deterministic processes driving introgression. This is the case in northern Iberia, where there is evidence for admixture involving *L. europaeus*, *L. granatensis* and *L. timidus*. This region was colonized in different times by these three species, with *L. timidus* replacing *L. granatensis,* which was later replaced by *L. europaeus*. It is thus possible that the same local selective pressures promoted introgression in these species. Interestingly, several regions of the genomes of the latter two were shown to harbor high frequencies of *L. timidus* variation and although not the same genes were involved, their functions were remarkably similar.

Among these genes is worthy to note those involved in immune response. This is an interesting result in light of recent studies that mark a tendency showing an association of adaptive introgression with immune related genes (Sams et al. 2016; Quach et al. 2016; Dannemann et al. 2016; Ullrich et al. 2017; Hasenkamp et al. 2015; Grossen et al. 2014). This class of genes is known for being under strong natural selection (either purifying, positive or balancing selection; see e.g. Quintana-Murci and Clark 2013). The incorporation of haplotypes from closely related species may thus be a particularly efficient source of viable and potentially adaptive diversity in immune related genes. For instance in humans, despite evidence for widespread negative selection against Neanderthal ancestry in genic regions (Sankararaman et al. 2014) multiple studies have shown that several innate immunity genes present higher Neanderthal ancestry than the remainder of the coding genome and has been suggested as an important factor for

adaptation to new pathogens when facing novel environments (Sams et al. 2016; Dannemann et al. 2016; Quach et al. 2016). Also, in situations of evolutionary arms-race, involving rapid cycles of adaptation, variants conferring resistance to novel pathogens can be acquired from other populations providing a prompt response (e.g. Hasenkamp et al. 2015).

Although immune-related genes are obvious candidates for adaptive introgression, evidence for the functional significance of introgression in our study is indirect without further functional studies that assess the fitness effects of the introgressed variants. This is common to most studies suggesting adaptive introgression, the few exceptions including adaptive introgression of rodenticide resistance in the house mouse (Song et al. 2011) and tolerance to drought in *Helianthus annuus* (Whitney et al. 2010). In other cases there is strong evidence for adaptive introgression, since an association can be established between genotype and phenotypes known to confer an advantage on particular populations. This is the case in Tibetans in which the EPAS1 gene has been inferred as being introgressed from Denisovans (or a close population; Huerta-Sánchez et al. 2014). Since variation in this gene has also been significantly associated with haemoglobin levels and is likely linked with adaptation to high-altitude hypoxia, it is likely that the introgressed variants have undergone positive selection but this remains to be tested. Another similar situation is that of *Heliconius* butterflies, which display Müllerian mimetic wing color patterns that work as a warning to predators of their toxicity. The genes in two genomic regions that control for wing color pattern were found to be exchanged by closely related species (The Heliconius Genome Consortium 2012; Pardo-Diaz et al. 2012) but again the adaptive nature of introgression in these genes remains hypothetical.

In sum, the increasing body of literature now suggesting an adaptive nature of interspecific gene flow suggest that, as has long been suggested for plants (Anderson 1949), hybridization and introgression may be in fact an important evolutionary force also in animals. Still, future studies integrating evidence of genes and traits associated with adaptation, the introgressive origin of these genes and fitness advantage of the introgressed variants in their recipient background will be needed to appraise the total extent and relative contribution of adaptive introgression in evolution.

4. Interspecific incompatibilities – variable recombination and the large X-effect

One major prospect of the use of genomic datasets is the ability to describe genome-wide patterns of differentiation and gene flow, which allows identifying regions with unusual patterns. The study of heterogeneous genomic patterns can give important clues about the formation and maintenance of reproductive barriers (Payseur and Rieseberg 2016; Harrison and Larson 2016; Abbott et al. 2016). However, these patterns should be interpreted with caution since other factors rather than reproductive isolation (e.g. natural selection) can account for heterogeneous genomic divergence (Nachman & Payseur 2012; Cruickshank & Hahn 2014).

The ubiquitous mtDNA introgression, that in this study we show to extend to yet another system (*Paper I*), and the evidence of considerable nuclear gene flow among the three European hares considered in here (*Papers II* and *III*) suggest that there are no clear obstacles to hybridization among hares. Such result may not be surprising given the rapid radiation of the genus (Matthee et al. 2004) and the lack of major morphological or karyological barriers to gene flow (Angermann 1983; Robinson et al. 1983; Flux and Angermann 1990). Still some strong barriers must come into play to impede global genomic reticulation and in *Paper II* we gave the first steps towards a better understanding of these. First we found an increase of introgression in chromosome ends where introgressed loci are more likely to escape incompatibility loci with single recombination event (Barton and Bengtsson 1986), the effect being enhanced by increased recombination rates in these regions (*Paper II*). Second, we found reduced introgression on the X chromosome both in *L. granatensis* (*Paper II*) but also *L. europaeus* either considering *L. timidus* or *L. granatensis* introgression into this species (Figure 4.2). These results suggest the existence of numerous incompatibility loci across the genomes and that these are particularly clustered in the X chromosome indicating inviability linked to the heterogametic sex as an obvious candidate as an important factor for speciation in hares.

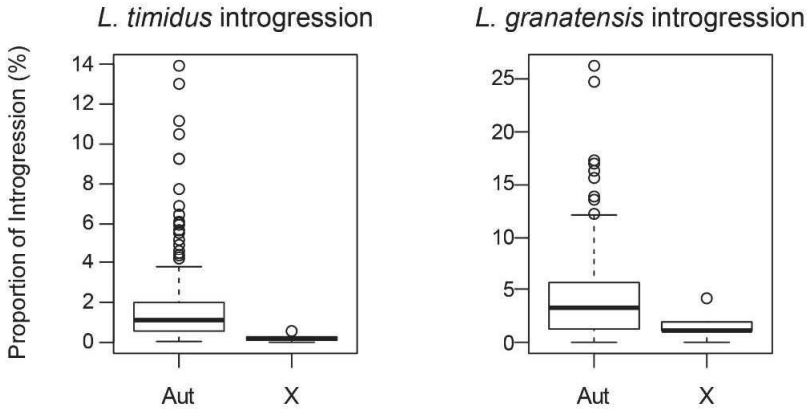


Figure 4.2. Distribution of the proportion of introgression across individuals for autosomes (Aut) and X-chromosome (X) (Mann-Whitney U test $p<0.05$ for both *L. timidus* (left) and *L. granatensis* (right) introgression).

These results are in line with two major patterns emerging in the speciation literature in studies using genomic datasets. The first is that despite outlier regions of differentiation can be found across the genome, these tend to be overrepresented in regions of reduced recombination (Brandvain *et al.* 2014; Janoušek *et al.* 2015), as for instance centromeres (e.g. Carneiro *et al.* 2014) and inversions (e.g. Lohse *et al.* 2015; Roesti *et al.* 2015). Several models posit that such regions may limit gene-flow by harbouring a disproportionate number of linked genomic incompatibilities (reviewed in Faria and Navarro 2010). The second, is that sex-linked loci tend to show reduced gene flow in comparison to autosomes (Payseur et al. 2004; Carneiro et al. 2014; Fontaine et al. 2015; Sankararaman et al. 2016; Martin et al. 2013; Macholán et al. 2007), a pattern that is in line with studies of controlled crosses in the laboratory that show a higher density of incompatibility loci on the X chromosome (Masly and Presgraves 2007). These are in line with the idea that when sterility or reduced fertility occurs in hybrids from interspecific crosses it generally affected the heterogametic sex (Haldane's rule; Haldane 1922) but also of a disproportionate effect of the X chromosome on hybrid sterility and inviability relative to autosomes (large X-effect; Coyne and Orr 1989).

5. Conclusions

The work presented in this thesis contributes to a better knowledge about the relevance of interspecific gene-flow in the history of species and to our understanding of the evolutionary causes and consequences of species admixture. The major conclusions of this work are:

i) Massive mitochondrial DNA introgression is a ubiquitous phenomenon in hares, and is not limited to certain lineages or environments. Introgression is however not restricted to the mitochondria, being also relevant in the nuclear genome;

ii) Northern Iberian Peninsula has been the stage of multiples waves of invasion and range replacements. More precisely, the biogeographic scenario we propose includes the range replacement of *L. timidus* by *L. granatensis* after the Last Glacial Maximum, which was then more recently replaced by *L. europaeus*. Despite the successive invasions, *L. timidus* mitochondrial DNA remained as a mark of the historical range of the species;

iii) Demographic replacements with hybridization in combination with behavioral traits and different transmission modes can result in massive discordant cytonuclear patterns of introgression. In particular, we show that the combination of such phenomena is sufficient to explain massive and geographically structured mtDNA introgression even when nuclear introgression is widespread and rare, without the need to invoke positive selection favouring the introgressed mtDNA. This could thus be the general mechanism behind cytonuclear discordance which is observed in several species;

iv) Although the two Iberian species analysed in this study were affected by massive mtDNA introgression, this was not followed by general co-introgression or co-differentiation of nuclear genes functionally linked to it, suggesting that cytonuclear co-evolution is not a major phenomenon in your model system. However, massive mtDNA introgression may still have had functional consequences, potentially linked to male fertility, as we find genes related with spermatogenesis introgressed at high frequencies;

v) While global nuclear introgression is a by-product of demographic processes, resulting in rare and geographic widespread introgression, we find evidence of adaptive introgression in both *L. granatensis* and *L. europaeus*. Notably, in both cases this seems to involve genes related with immunity, which could indicate that introgression of these genes was important for the successful colonization of new pathogenic environments in northern Iberian Peninsula by these two species. Interestingly, introgression of this

category of genes is increasingly being reported in other systems, suggesting that interspecific gene flow may be a major source of adaptive variance in these genes;

vi) Introgression was counter-selected in result of numerous genetic incompatibilities scattered across the genome in interplay with recombination. Incompatibility loci are particularly dense in the X chromosome suggesting a disproportionate role of this sexual chromosome in reproductive isolation (large X-effect).

6. Final considerations, undergoing work and future directions

The research associated with this thesis shed light on several aspects regarding the nature of the processes leading to introgression and also its outcomes. However, it also raised several questions that would be interesting to address in future studies, some of which are highlighted here.

In this thesis we were able to show that the joint action of demographic processes during range replacements and sex-biased behaviours have likely led to massive *timidus* mtDNA introgression in *L. granatensis*, and could be a major general mechanism generating cytonuclear discordance in the many studies in which this is observed. The framework that we set in this study can now be applied to these other systems including hares to appraise the generality of the mechanism we propose here. In this sense, the North American system could be an interesting system to start with, namely focusing on the interactions between *L. americanus* and *L. californicus*. Similarly, to the *L. granatensis*-*L. timidus* system, we show massive mtDNA introgression, following a north-south gradient, between two species with very different ecological requirements (one is temperate and the other boreal) and thus likely with contrasting demographic histories in result of climatic oscillations.

We found that cytonuclear co-evolution does not play a major role in determining patterns of introgression in *L. granatensis*, but still both in this species and in *L. europaeus* we find a few nuclear genes functionally linked to the mitochondria to be potential candidates for co-introgression or co-differentiation with the introgressed mtDNA. However, given the limited sampling of individuals in both these studies our approach is prone to false-positives, especially in the case of *L. granatensis* where we impose a binary pattern to test for co-introgression with the mtDNA, while *timidus* mtDNA introgression in this species rather follows a gradient. A population-approach screening the geographic structure at these genes will thus be needed to confirm their status as candidates for cytonuclear co-evolution. Furthermore, in the two studies we found the little congruence of candidate mitonuc genes, which could have partially resulted from the fact that we used different methodologies to detect introgression. While in *L. europaeus* we relied on ELAI to detect introgression, in *L. granatensis* we used the less powerful RND approach to define candidates (although we tried to overcome this problem by relaxing the false discovery rate and thus increase our power). Perhaps an alternative to screen for mitonuc genes repeatedly co-introgressing with *timidus* mtDNA

would be to use phylogenetic-based approaches including the four European species massively affected by *L. timidus* mtDNA introgression and look for mitogenes showing phylogenetic discordances that can be evocative of introgression.

Another factor that may have hindered our ability to correctly detect local patterns of introgression in some cases is the use of the relatively distant rabbit reference genome. The use of a distant genome can hinder mapping success and mapping/genotype quality thus reducing the final amount of usable information but can also result in convergence of the called genotypes towards the reference and an underestimation of divergence. To overcome this problem we created a pseudo-reference by iterative mapping. Such an approach has been shown to be successful in correcting such mapping biases (Sarver et al. 2017), and was here applied quite successfully. Still, this approach does not take into account rearrangements between the rabbit and hare genomes. Although the hare and rabbit genomes show high synteny and we accounted for major the known chromosomal rearrangements (chromosomes 1 and 2 of the rabbit is split in two in hares; Robinson et al. 2002) other smaller scale rearrangements could have biased our analysis at the local level. Given initial uncertainties about the efficiency of using the rabbit genome as reference, we produced in parallel sequence data to *de novo* assemble a hare genome. We have performed the *de novo* assembly following the standard recipe of ALLPATHS-LG, using overlapping 180 bp paired-end reads and mate-pairs with different insert sizes (2.5 kb, 4.5kb and 8kb). The resulting assembled genome was evaluated by calculating simple statistics (Genome size, number of contigs, N50) and examining the presence of core eukaryotic genes. We further validated our assembly by mapping the reads into the scaffolds in order to assess its internal consistency and break scaffolds when this was not met. Next, we used a SSPACE to re-scaffold our assembly as this software is specifically designed for the purpose. This led to a final assembled genome of 2.7 Gb with an N50 of 420'320 bp, the largest scaffold having 3'358'433 bp. Finally, we also produced gene annotation for our *de novo* assembly based on *ab-initio* predictions, transcriptome evidence and homology based approaches. However, the produced de novo assembly is too fragmented, as attested by the large number of scaffolds (33018), and thus not adequate to properly apply genome scans as used in Papers II and III. Still, the produced genome is an important resource for any genomic study in hares, particularly for analyses at local genomic scales where small rearrangements relative to the rabbit genome can affect them.

Although initially we were looking for nuclear genes following the *timidus* mtDNA introgression in search of signals of indirect adaptive introgression of *timidus* mtDNA, instead we found that massive *timidus* mtDNA introgression might have had negative effects in the recipient species, as in *L. granatensis* we find massive introgression of some genes related with spermatogenesis. This hypothesis is based on the premise that *timidus* mtDNA introgression would carry male-harmful mutations when in a *L. granatensis* background, and that it could have been compensated by introgression of *timidus* variants in order to re-establish male fertility. While this is only speculative for the moment, it would be interesting to evaluate the effects of hetero-specific combinations of mitochondria and nuclear backgrounds, on male fertility. This could be tested by performing experimental crosses for instance between *L. timidus* and *L. europaeus* individuals with pure nuclear background and measuring male breeding success. Such studies have already been successfully performed in *L. europaeus* and show the existence of the mother's curse between populations with divergent mitochondrial lineages.

Finally, in this study and in regard to the Iberian hares we have mostly focused our analysis into the patterns of introgression of *timidus* origin and with especial attention to Northern Iberian Peninsula. We found cases evidence suggestive of adaptive introgression, which in some cases could potentially be linked to adaptation to the local pathogenic environment. However, *timidus* introgression was also found outside Iberia and thus the analysis of *timidus* introgression over *L. europaeus* distribution could help uncover other cases of adaptive introgression. Likewise, interspecific gene flow was also shown to occur in both directions between *L. europaeus* and *L. granatensis* and thus give us the opportunity to explore other potential cases of adaptive introgression not involving *timidus*. The data collected in this study further allows us to inspect the combined landscape of introgression in different species and of several origins. The study of the common patterns of introgression can help us understand for instance which regions of the genome are more prone to repeatedly introgress and thus benefit most from the introduction of genetic variation from interspecific gene-flow (see e.g. Ullrich et al. 2017). In the other sense, this combined landscape of introgression can shed further light into genomic architecture of reproductive isolation for instance by looking for genomic regions that never introgress (see e.g. Sankararaman et al. 2016).

In sum, our study gives major insights into the mechanisms leading to reticulate evolution but also opens many exciting avenues for the study of these same

mechanisms, from the involvement of demographic processes and behavioural traits promoting massive discordance of patterns of introgression among genetic compartments, the role of genomic conflicts further shaping introgression in result of accidental introgression, to the relevance of adaptive introgression in species evolution.

7. References

Abbott RJ, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and speciation. *J Evol Biol* **26**: 229–246.

Abbott RJ, Barton NH, Good JM. 2016. Genomics of hybridization and its evolutionary consequences. *Mol Ecol* **25**: 2325–2332.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P. 2008. The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus Lepus. *Philos Trans R Soc Lond B Biol Sci* **363**: 2831–2839.

Amoutzias GD, Giannoulis T, Moutou KA, Psarra AMG, Stamatis C, Tsipourlianos A, Mamuris Z. 2016. SNP identification through transcriptome analysis of the European Brown hare (*Lepus europaeus*): Cellular energetics and mother's curse. *PLoS One* **11**: 1–17.

Angermann R. 1983. The taxonomy of Old World Lepus. *Acta Zool Fenn* **174**: 17–21.

Barton NH, Bengtsson BO. 1986. The barrier to genetic exchange between hybridising populations. *Heredity* **57**: 357–376.

Beck EA, Thompson AC, Sharbrough J, Brud E, Llopart A. 2015. Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression. *Evolution* **69**: 1973–1986.

Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. 2014. Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet* **10**: e1004410.

Burton RS, Barreto FS. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol* **21**: 4942–4957.

Burton RS, Pereira RJ, Barreto FS. 2013. Cytonuclear Genomic Interactions and Hybrid Breakdown. *Annu Rev Ecol Evol Syst* **44**: 281–302.

Carneiro M, Albert FW, Afonso S, Pereira RJ, Burbano H, Campos R, Melo-Ferreira J, Blanco-Aguiar J a., Villafuerte R, Nachman MW, et al. 2014. The Genomic Architecture of Population Divergence between Subspecies of the European Rabbit ed. J.L. Feder. *PLoS Genet* **10**: e1003519.

Cheng E, Hodges KE, Melo-Ferreira J, Alves PC, Mills LS. 2014. Conservation implications of the evolutionary history and genetic diversity hotspots of the snowshoe hare. *Mol Ecol* **23**: 2929–42.

Coyne JA, Orr AH. 1989. Two rules of speciation. In *Speciation and its Consequences* (ed. S. Associates), pp. 180–207, Sunderland, MA.

Dannemann M, Andrés AM, Kelso J. 2016. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am J Hum Genet* **98**: 22–33.

Faria R, Navarro A. 2010. Chromosomal speciation revisited: Rearranging theory with pieces of evidence. *Trends Ecol Evol* **25**: 660–669.

Flux J, Angermann R. 1990. The Hares and Jackrabbits. In *Rabbits, Hares and Pikas: Status Survey and Conservation Action Plan* (eds. J.A. Chapman and J. Flux), pp. 61–94, Gland, Switzerland: International Union for Conservation of Nature and Natural Resources.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov I V, Jiang X, Hall AB, Catteruccia F, Kakani E, et al. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**: 1258524–1258524.

Giannoulis T, Stamatis C, Tsipourlianos A. 2017. Mitogenomic analysis in European brown hare (*Lepus europaeus*) proposes genetic and functional differentiation between the distinct lineages. *Mitochondrial DNA Part A*. http://dx.doi.org/10.1080/24701394.2016.1278540

Good JM, Vanderpool D, Keeble S, Bi K. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution* **69**: 1961–1972.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. 2010. A draft sequence of the Neandertal genome. *Science* **328**: 710–22.

Grossen C, Keller L, Biebach I, Croll D. 2014. Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex. *PLoS Genet* **10**: e1004438.

Haldane J. 1922. Sex ratio and unisexual sterility in hybrid animals. *J Genet* **12**: 101–109.

Harrison RG, Larson EL. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Mol Ecol* **25**: 2454–2466.

Hasenkamp N, Solomon T, Tautz D. 2015. Selective sweeps versus introgression - population genetic dynamics of the murine leukemia virus receptor Xpr1 in wild populations of the house mouse (*Mus musculus*). *BMC Evol Biol* **15**: 248.

Hill GE. 2016. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol Evol* **6**: 5831–5842.

Huerta-Sánchez E, Jin X, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, Ni P, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*.

Janoušek V, Munclinger P, Wang L, Teeter KC, Tucker PK. 2015. Functional Organization of the Genome May Shape the Species Boundary in the House Mouse. *Mol Biol Evol* **32**: 1208–1220.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**: 371–375.

Liu KJ, Steinberg E, Yozzo A, Song Y, Kohn MH, Nakhleh L. 2015. Interspecific introgressive origin of genomic diversity in the house mouse. *Proc Natl Acad Sci* **112**: 196–201.

Lohse K, Clarke M, Ritchie MG, Etges WJ. 2015. Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution (N Y)* **69**: 1178–1190.

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

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res* **23**: 1817–1828.

Masly JP, Presgraves DC. 2007. High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS Biol* **5**: 1890–1898.

Matthee C, Van Vuuren B, Bell D, Robinson T. 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.

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

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**: 222–6.

Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, Lanzaro GC. 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc Natl Acad Sci* **112**: 815–820.

Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, McMillan WO, Jiggins CD. 2012. Adaptive Introgression across Species Boundaries in *Heliconius* Butterflies ed. M. R. Kronforst. *PLoS Genet* **8**: e1002752.

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

Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Mol Ecol* **25**: 2337–2360.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**: 1410–1414.

Pritchard VL, Edmands S. 2013. The genomic trajectory of hybrid swarms: outcomes of repeated crosses between populations of *Tigriopus californicus*. *Evolution* **67**: 774–91.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**: 43–9.

Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, Laval G, Patin E, Harmant C, Lopez M, et al. 2016. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**: 643–656.e17.

Quintana-Murci L, Clark AG. 2013. Population genetic tools for dissecting innate immunity in humans. *Nat Rev Immunol* **13**: 280–93.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet* **16**: 359–371.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**: 1053–1060.

Rieseberg LH. 2011. Adaptive introgression: The seeds of resistance. *Curr Biol* **21**: R581–R583.

Robinson TJ, Elder FF, Chapman J a. 1983. Karyotypic conservatism in the genus *Lepus* (order Lagomorpha). *Can J Genet Cytol* **25**: 540–544.

Robinson TJ, Yang F, Harrison WR. 2002. Chromosome painting refines the history of genome evolution in hares and rabbits (order Lagomorpha). *Cytogenet Genome Res* **96**: 223–7.

Roesti M, Kueng B, Moser D, Berner D. 2015. The genomics of ecological vicariance in threespine stickleback fish. *Nat Commun* **6**: 8767.

Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biol* **14**: 1–22.

Sams AJ, Dumaine A, Nédélec Y, Yotova V, Alfieri C, Tanner JE, Messer PW, Barreiro LB. 2016. Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. *Genome Biol* **17**: 246.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**: 354–7.

Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Curr Biol* **26**: 1241–1247.

Sarver BAJ, Keeble S, Cosart T, Tucker PK, Dean MD, Good JM. 2017. Phylogenomic Insights into Mouse Evolution Using a Pseudoreference Approach. *Genome Biol Evol* **9**: 726–739.

Sloan DB, Havird JC, Sharbrough J. 2017. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol Ecol* **26**: 2212–2236.

The Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**: 94–98.

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* **21**: 3907–3930.

Tolesa Z, Bekele E, Tesfaye K, Ben Slimen H, Valqui J, Getahun A, Hartl GB, Suchentrunk F. 2017. Mitochondrial and nuclear DNA reveals reticulate evolution in hares (*Lepus* spp., Lagomorpha, Mammalia) from Ethiopia ed. T.-Y. Chiang. *PLoS One* **12**: e0180137.

Ullrich KK, Linnenbrink M, Tautz D. 2017. Introgression patterns between house mouse subspecies and species reveal genomic windows of frequent exchange. **1**.

Wolff JN, Ladoukakis ED, Enriquez J a., Dowling DK. 2014. Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Philos Trans R Soc B Biol Sci* **369**: 20130443–20130443.

Wu Y, Xia L, Zhang Q, Yang Q, Meng X. 2011. Bidirectional introgressive hybridization between *Lepus capensis* and *Lepus yarkandensis*. *Mol Phylogenet Evol* **59**: 545–555.

Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR. 2016. Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biol* **17**: 25.

Annexes

Annex I. Supplementary material from paper I in Chapter 2. Promiscuous mitochondrial DNA in hares

Annex II. Supplementary material from paper II in Chapter 3. Genomic perspective of introgression in hares from Iberia

Annex III. Supplementary material from paper III in Chapter 3. Genomic perspective of introgression in hares from Iberia

Annex I. Supplementary material from paper I in Chapter 2. Promiscuous mitochondrial DNA in hares

List of Supplementary Figures and Tables

Figure S2.1 Individual phylogenies of nuclear loci generated from the outputs of BEAST.

Figure S2.2 Individual phylogenies of nuclear loci inferred using Garli

Figure S2.3 Species tree of *L. californicus*, *L. townsendii*, and *L. americanus* inferred with *BEAST.

Figure S2.4 Cytochrome b phylogenies of all North American hare species and one sequence representative of each non-North American hare species available in GenBank.

Figure S2.5 Demographic profiles of *L. americanus* population clusters based on Extended Bayesian Skyline Plot analyses.

Figure S2.6 Empirical and simulated mtDNA distances between *L. californicus* and the PacNW2 group of *L. americanus* (simulations used highest 95% HPD estimate of effective population sizes and the lowest 95% HPD estimate of divergence time inferred using IMa2).

Table S2.1 Detailed information of sequences obtained per individual with GenBank accession numbers.

Table S2.2 Geographic coordinates of sampling sites.

Table S2.3 Analysed loci, PCR conditions and primers.

Table S2.4 Posterior probabilities for models of taxa delimitation estimated using BP&P.

Table S2.5 ML estimates of demographic parameters obtained with IMa2 between pairs of species.

Table S2.6 ML estimates of demographic parameters obtained with IMa2 among the three species.

Table S2.7 ML estimates of demographic parameters obtained with IMa2 among the three *L. americanus* groups.

SPTBN1

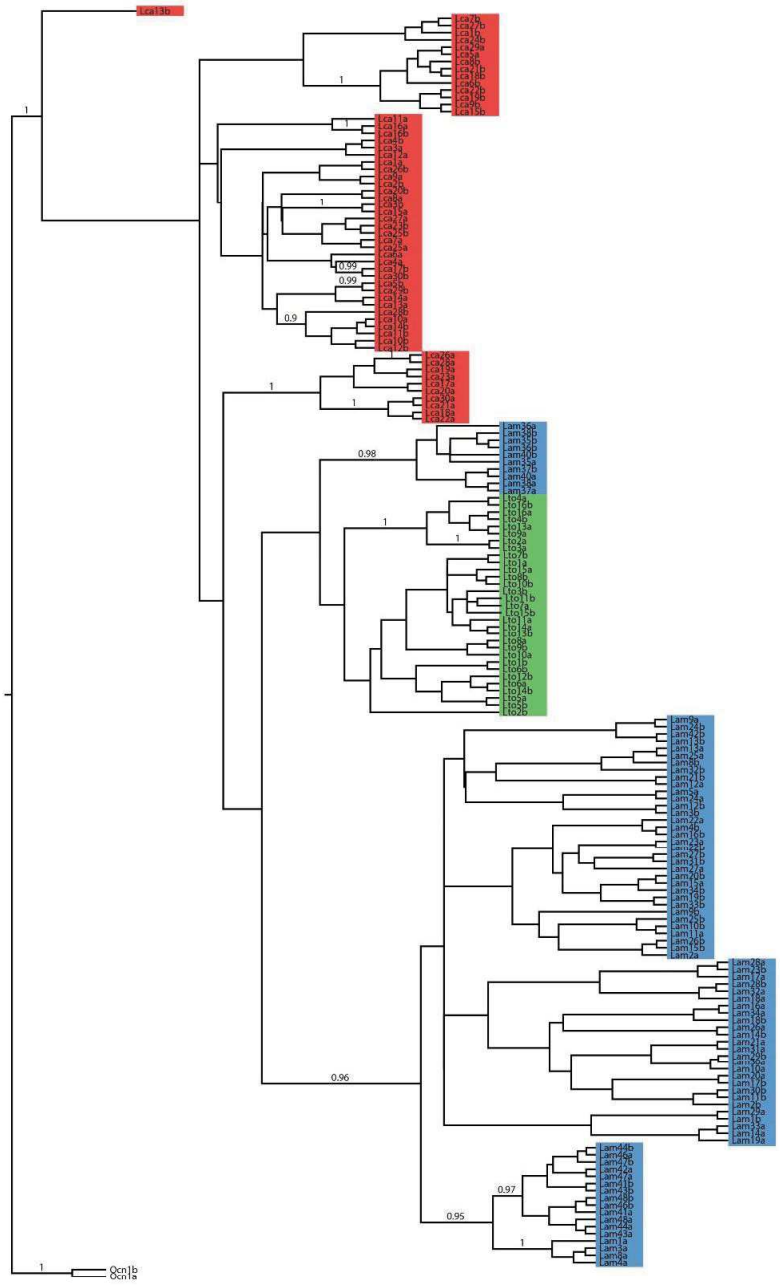


Figure S2.1 Individual phylogenies of nuclear loci generated from the outputs of BEAST (Drummond and Rambaut, 2007) (numbers close to nodes indicate the posterior probabilities if higher than 0.9). Coloured shades indicate the species: blue - *L. americanus*; red - *L. californicus*; green - *L. townsendii*. Codes of sequenced specimens are those shown in Table S2.1.

PRKCI

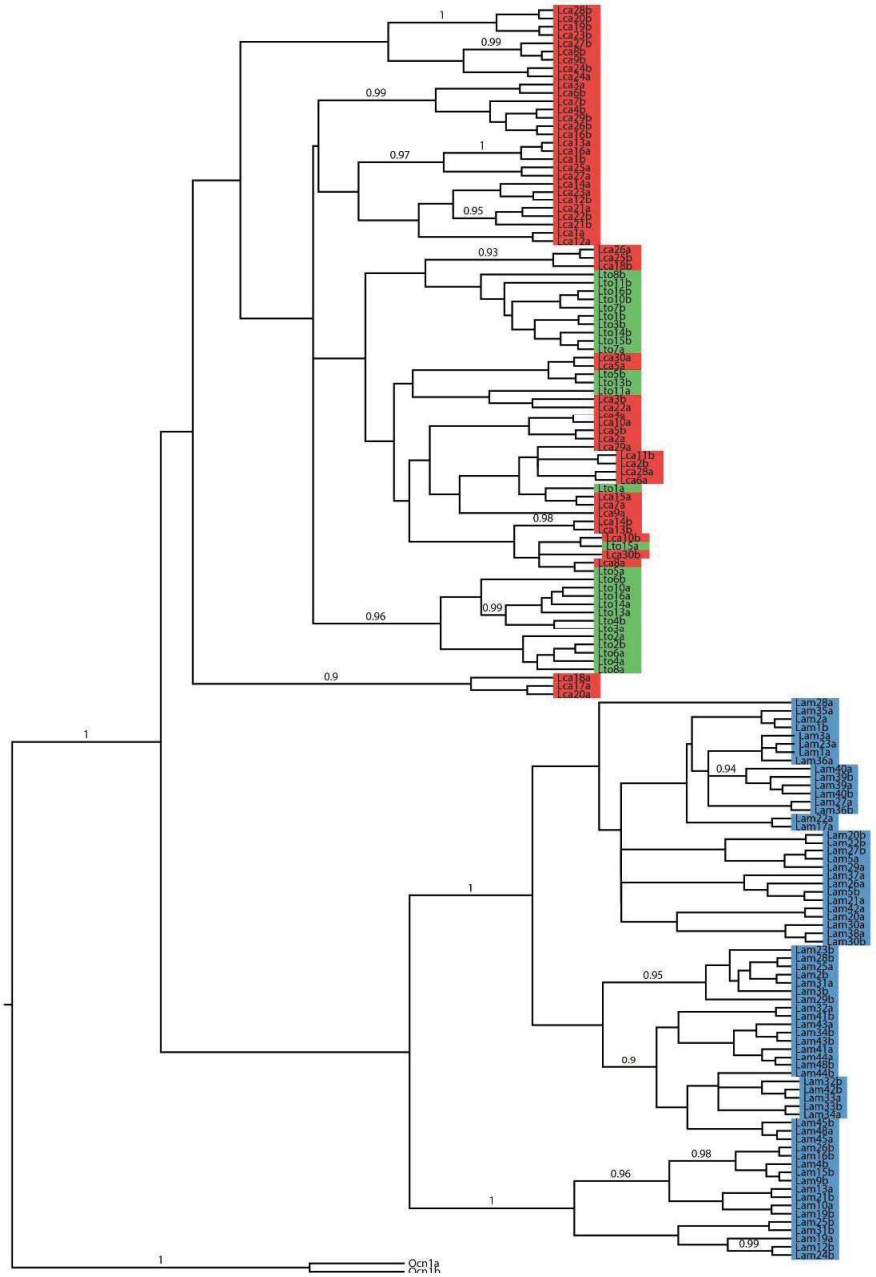


Figure S2.1 (cont'd)

DARC

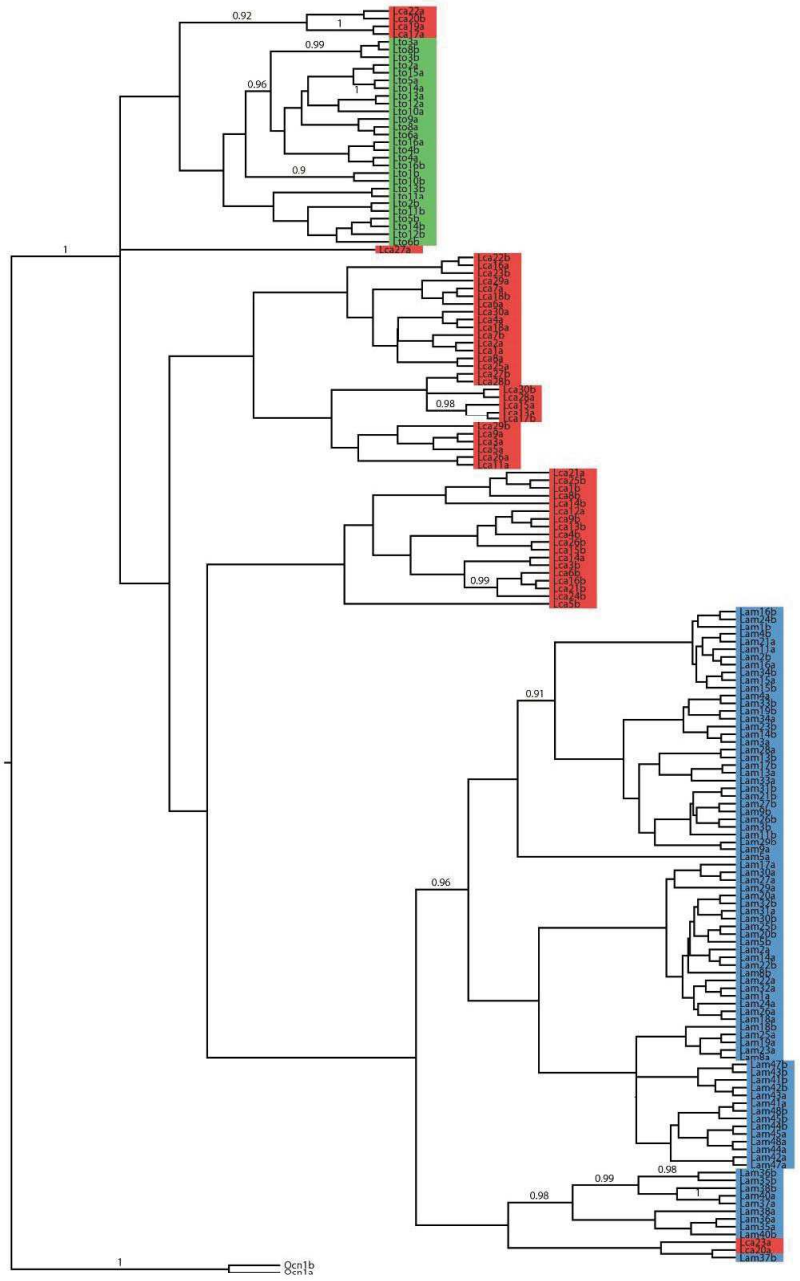


Figure S2.1 (cont'd)

KITLG

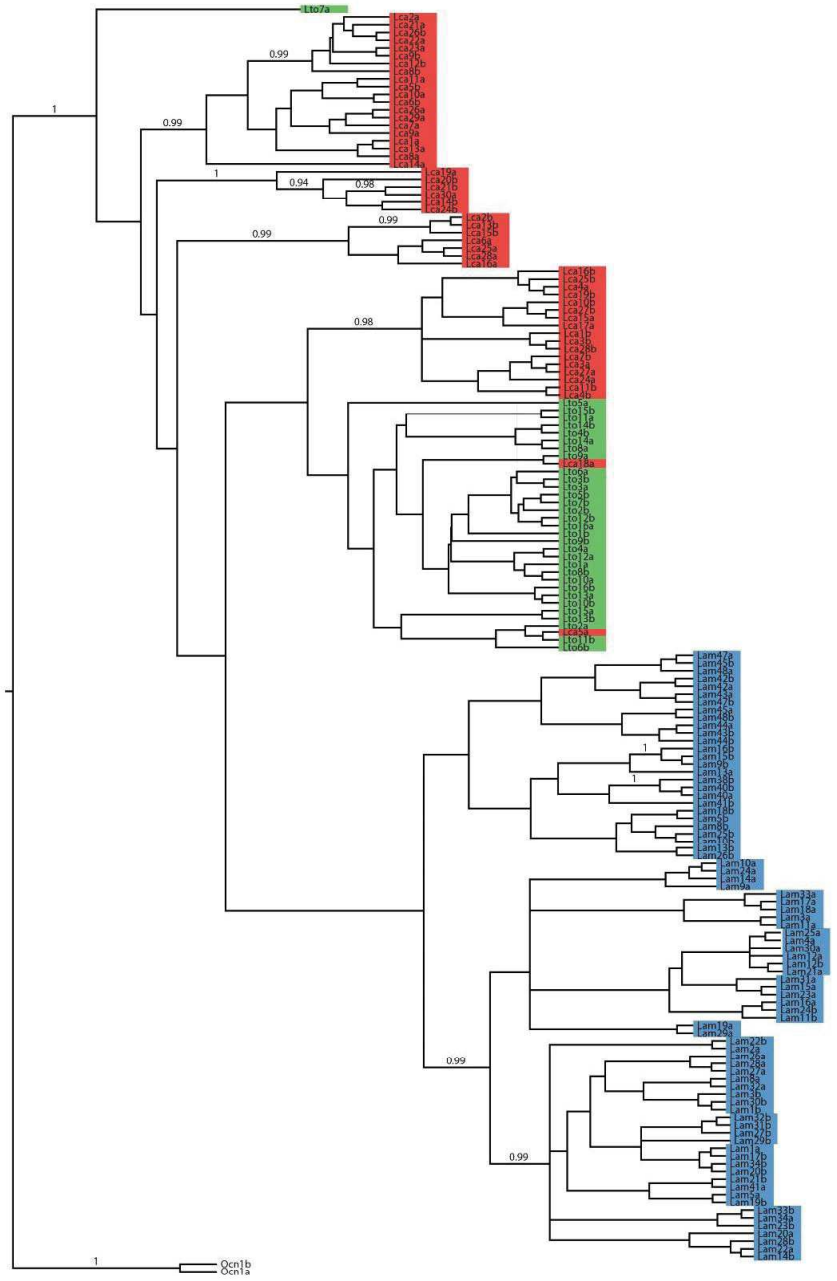


Figure S2.1 (cont'd)

TF

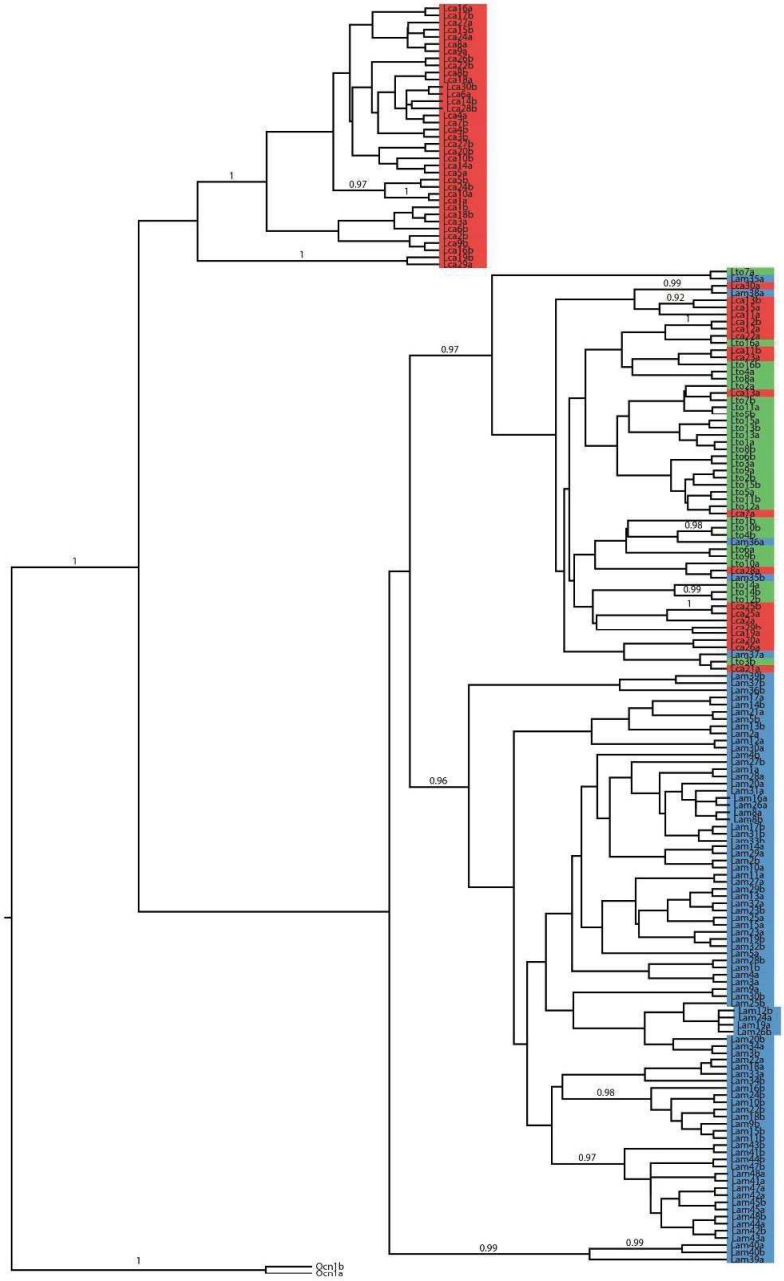


Figure S2.1 (cont'd)

POLA1

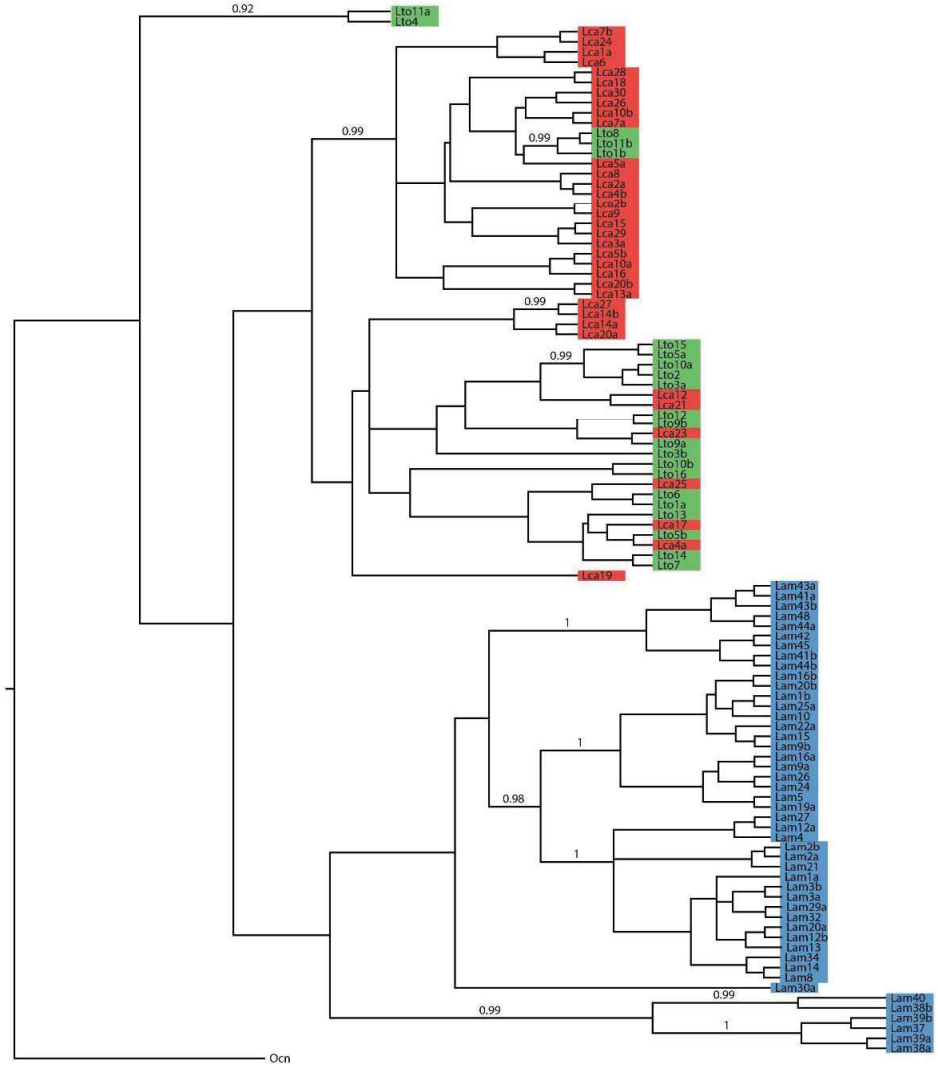


Figure S2.1 (cont'd)

GRIA3

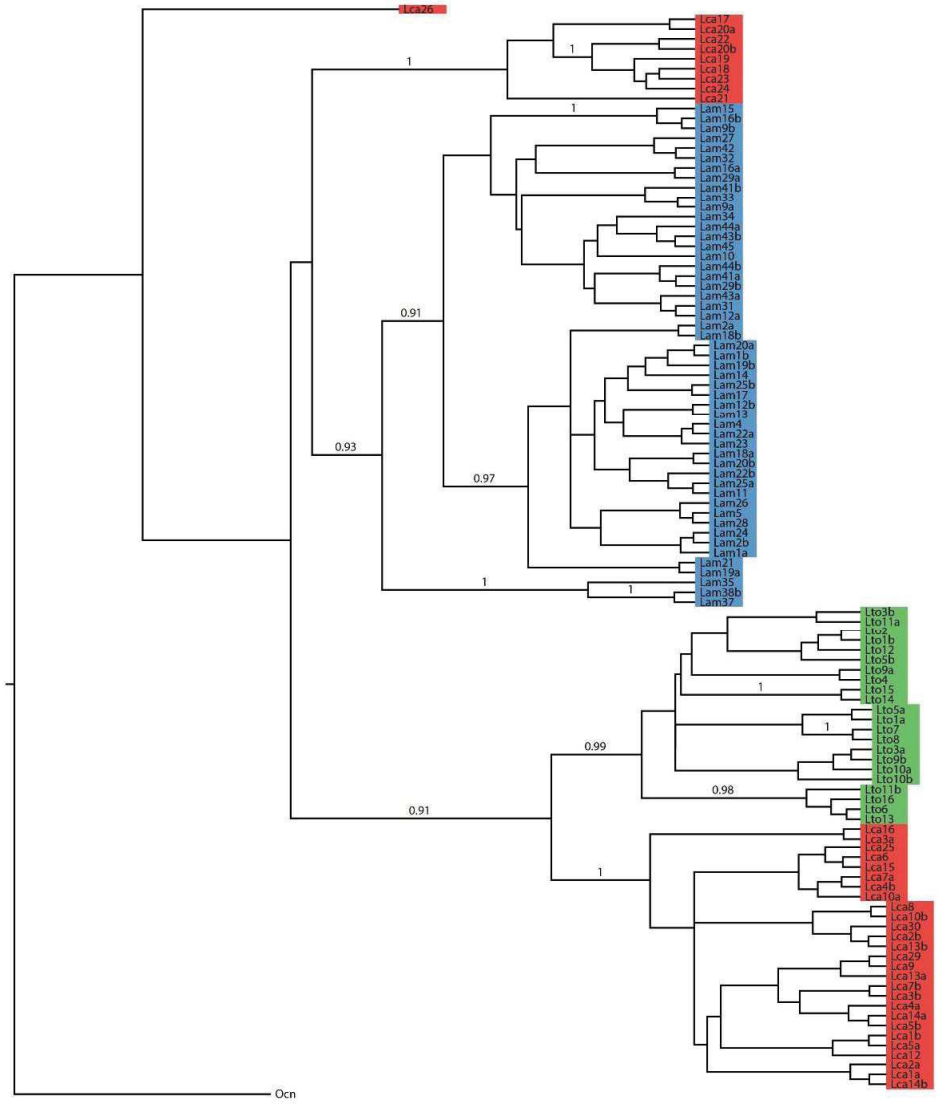


Figure S2.1 (cont'd)

SRY

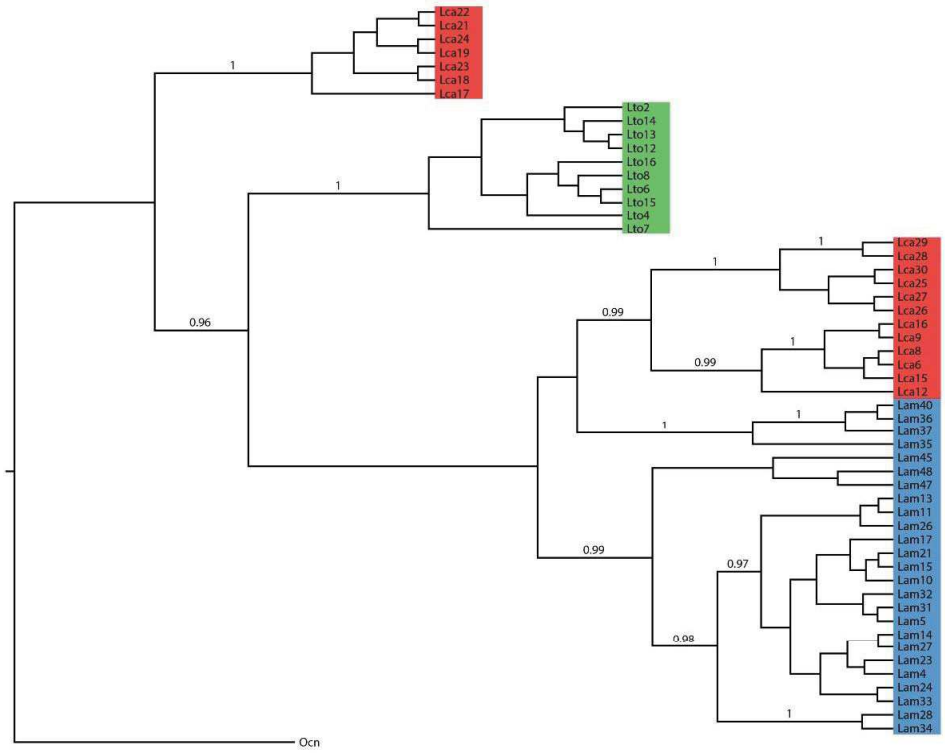


Figure S2.1 (cont'd)

SPTBN1

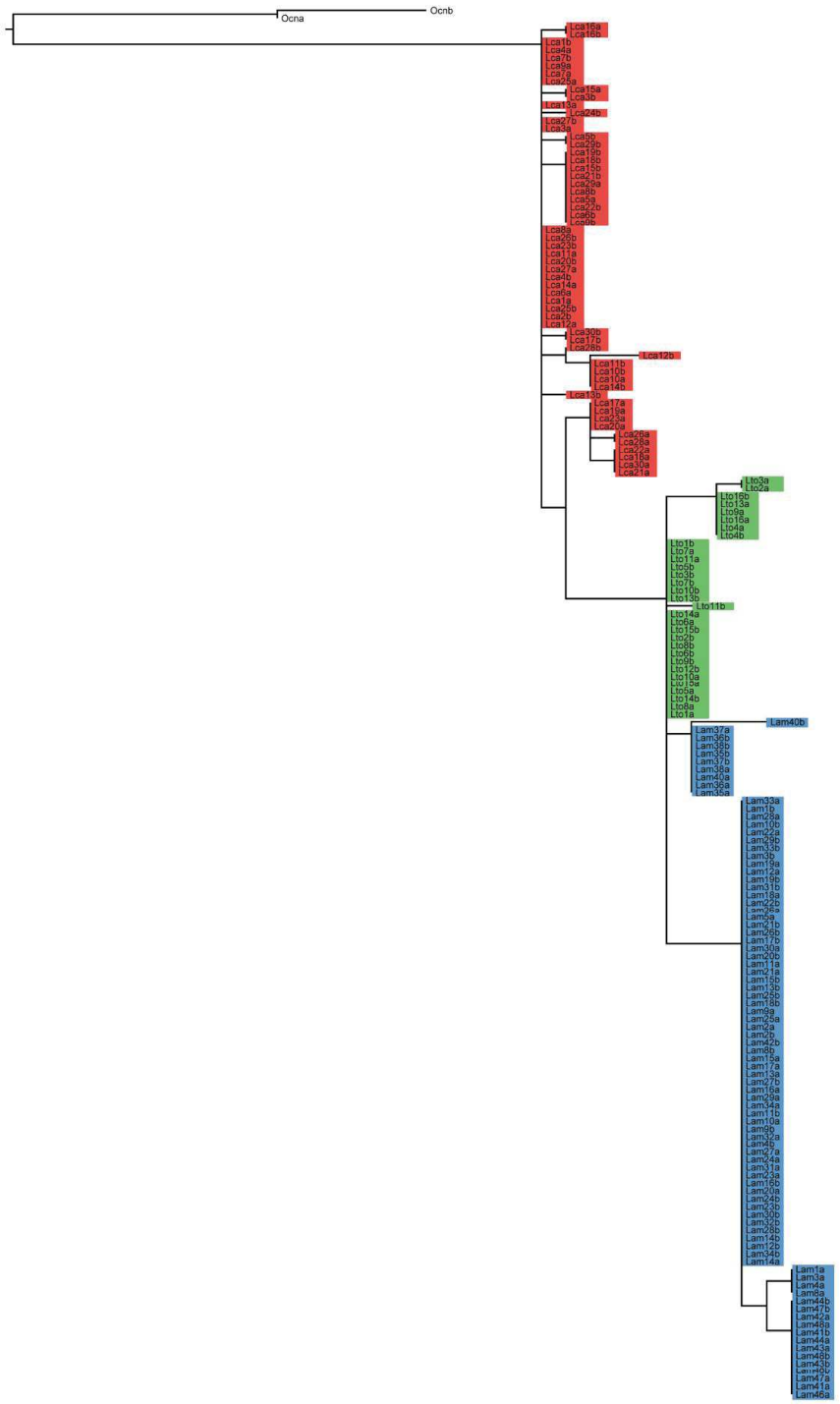


Figure 2.2 Individual phylogenies of nuclear loci inferred using Garli v1.0. Coloured shades indicate the species: blue - *L. americanus*; red - *L. californicus*; green - *L. townsendii*. Codes of sequenced specimens are those shown in Table S2.1.

PRKCI

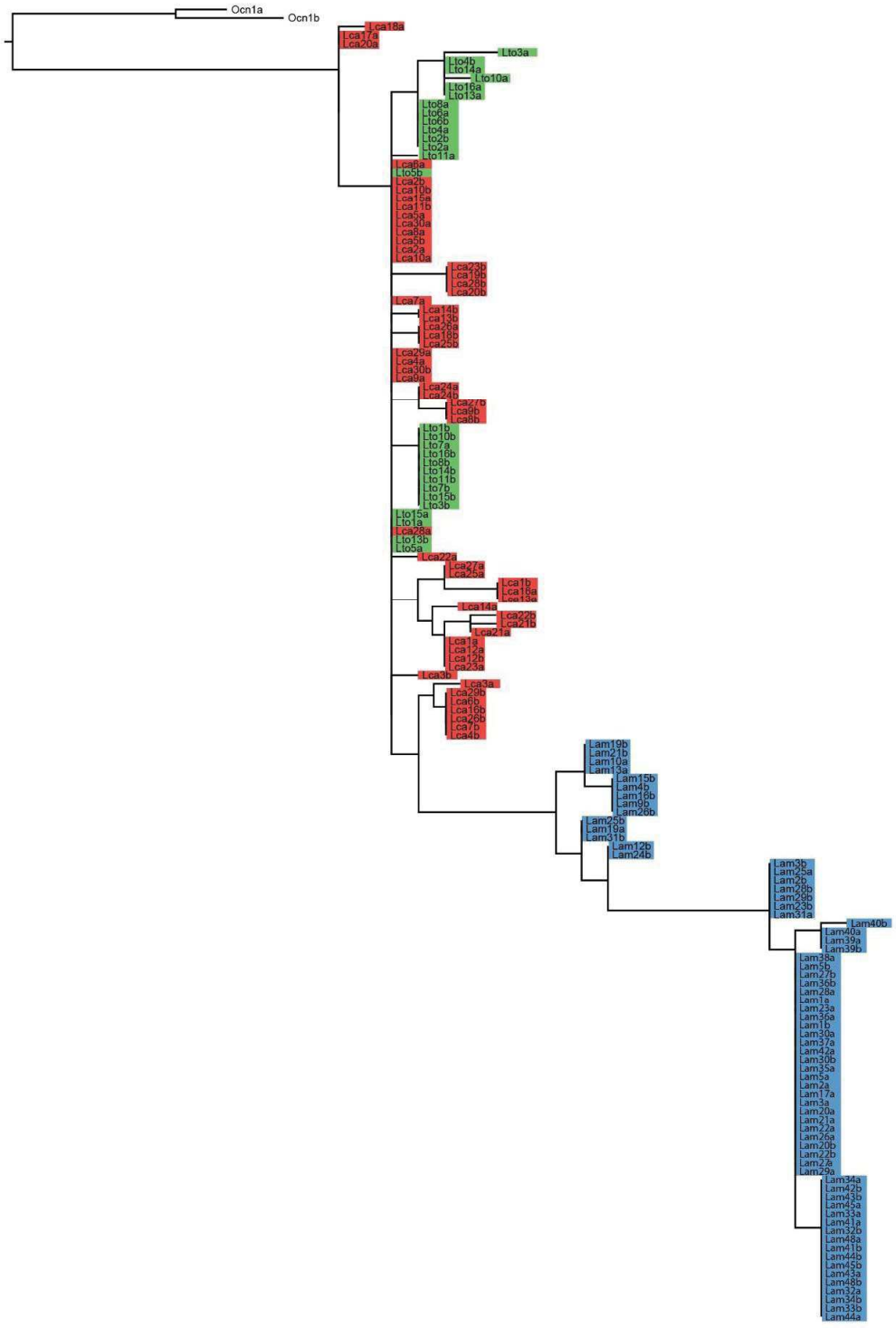


Figure S2.2 (cont'd)

DARC

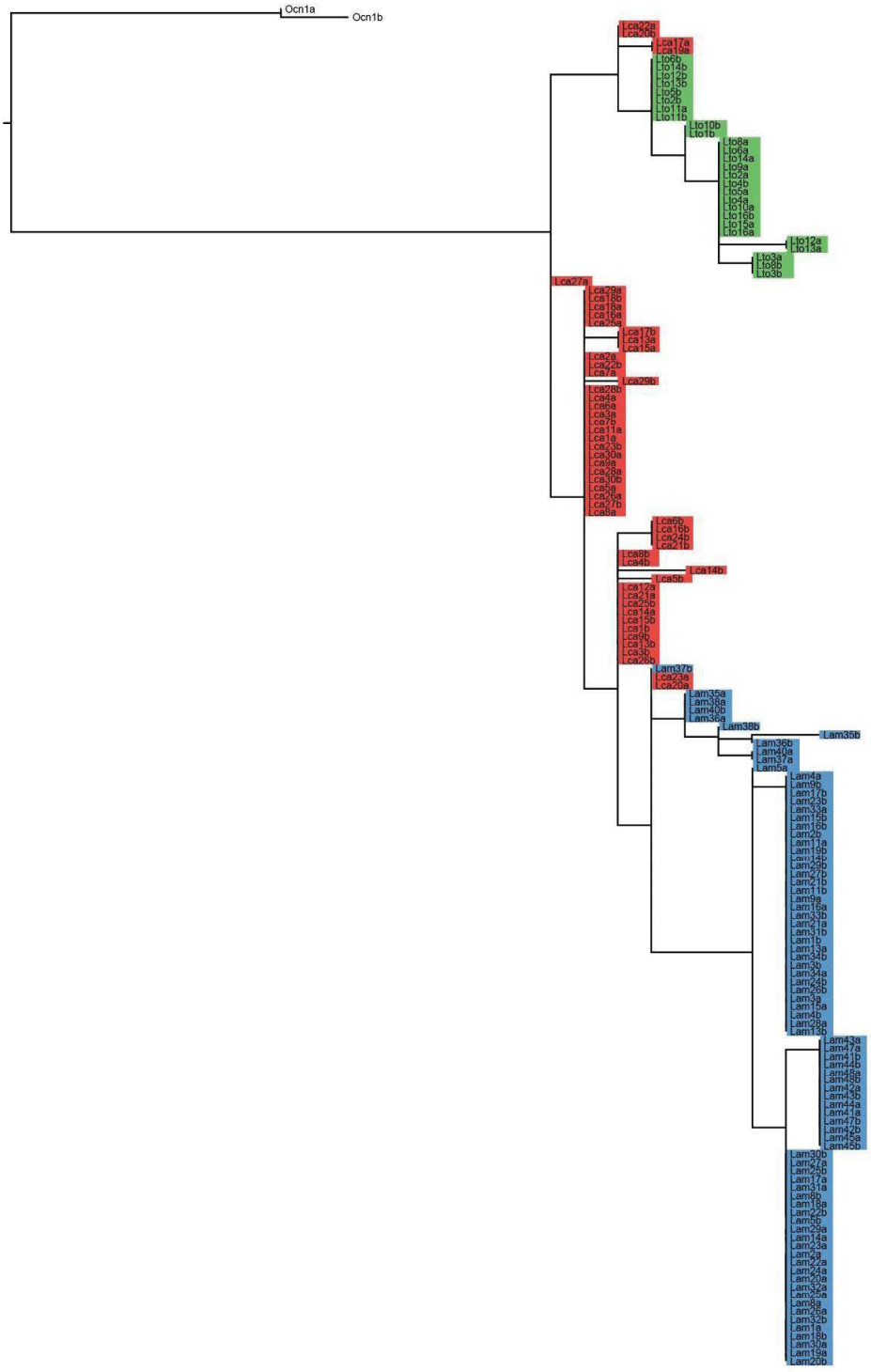


Figure S2.2 (cont'd)

KITLG

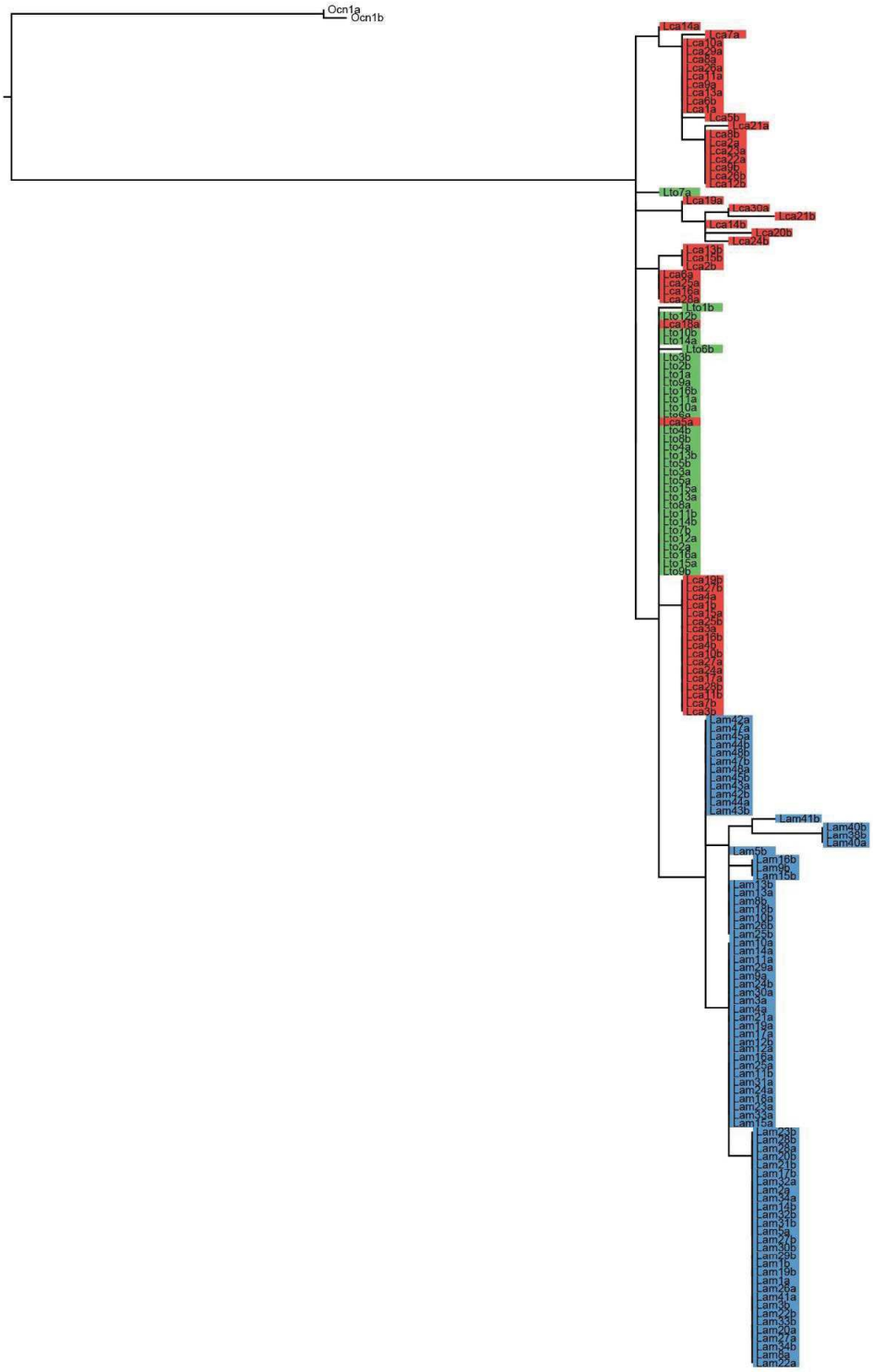


Figure S2.2 (cont'd)

TF

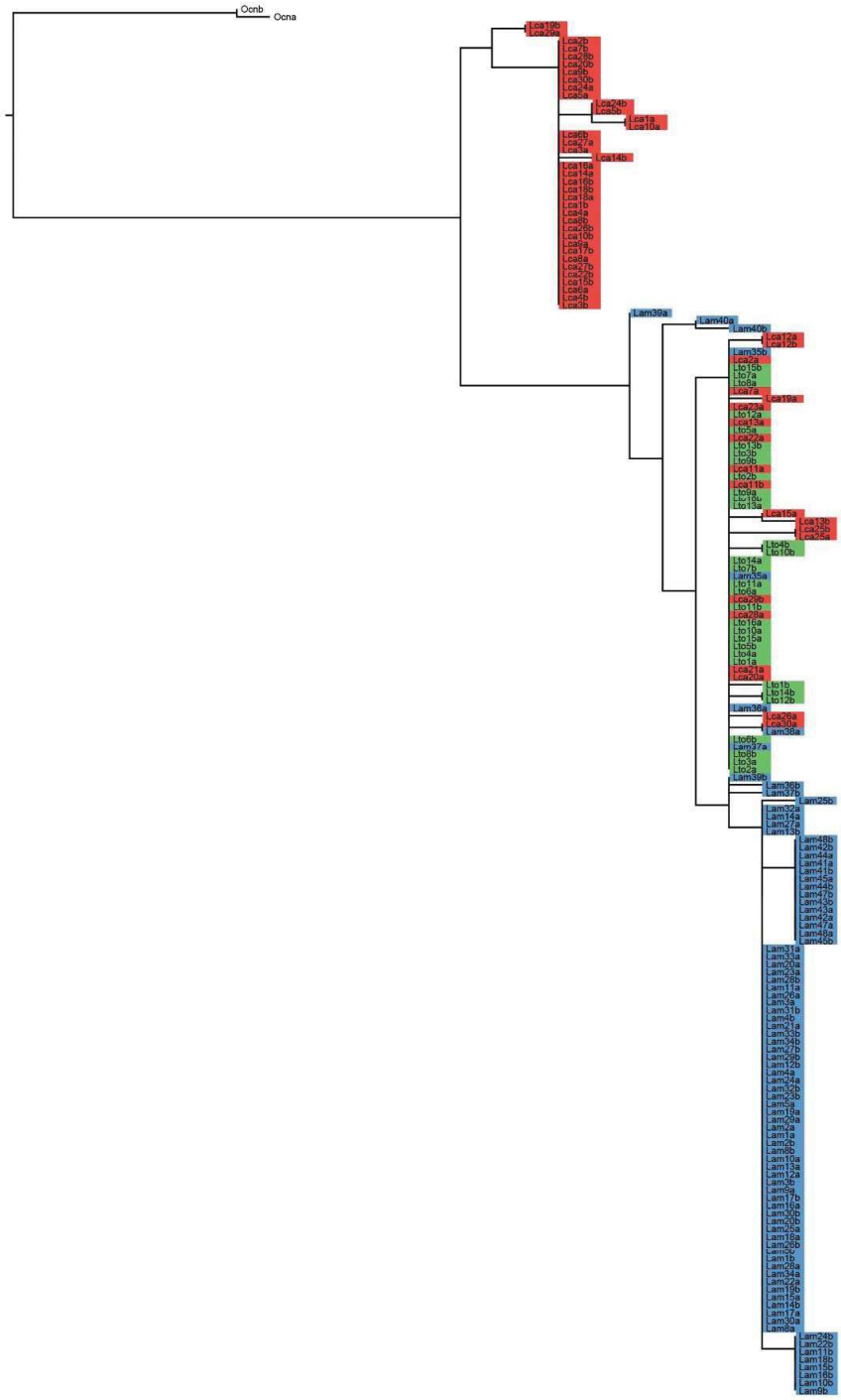


Figure S2.2 (cont'd)

POLA1

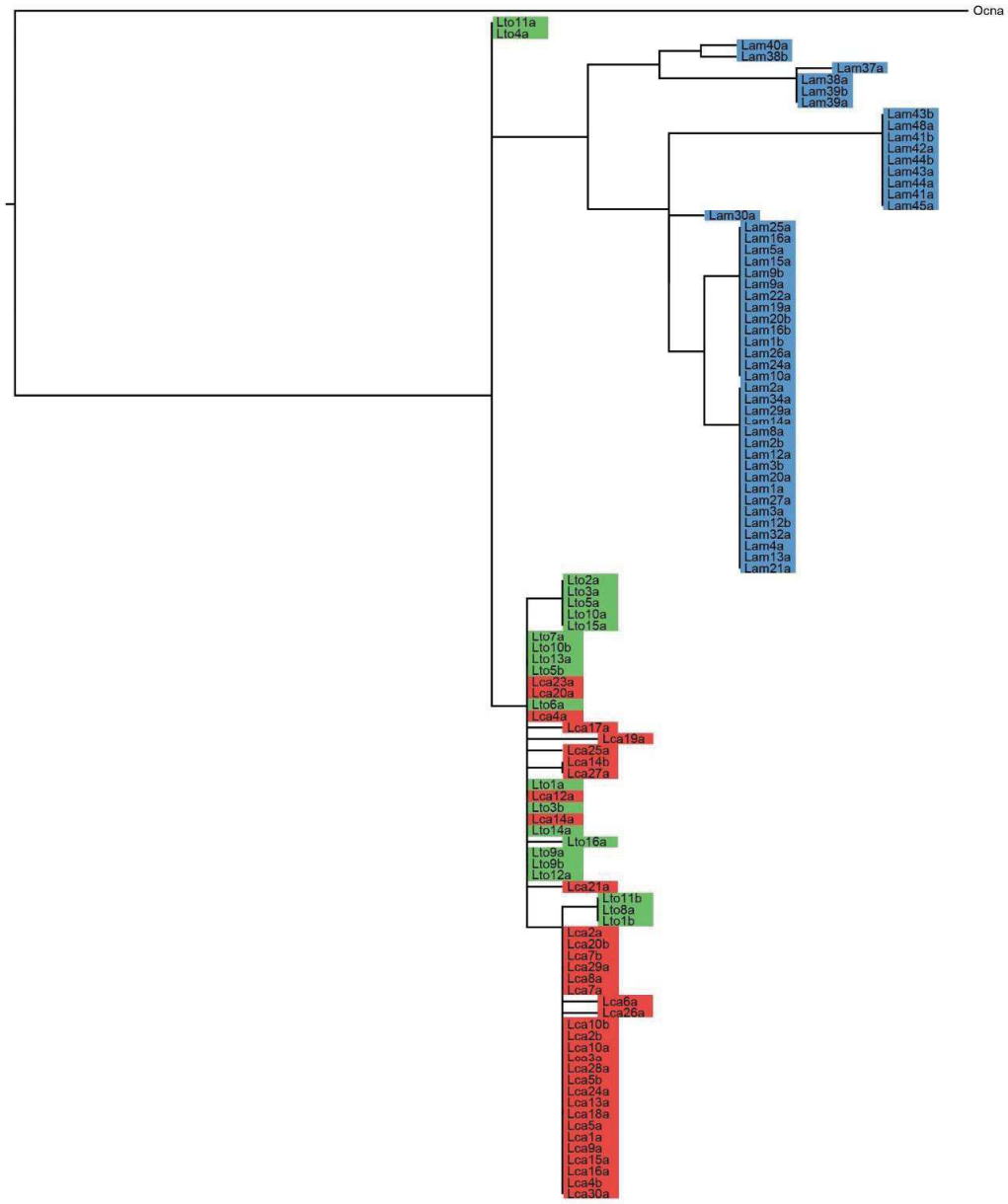


Figure S2.2 (cont'd)

GRIA3

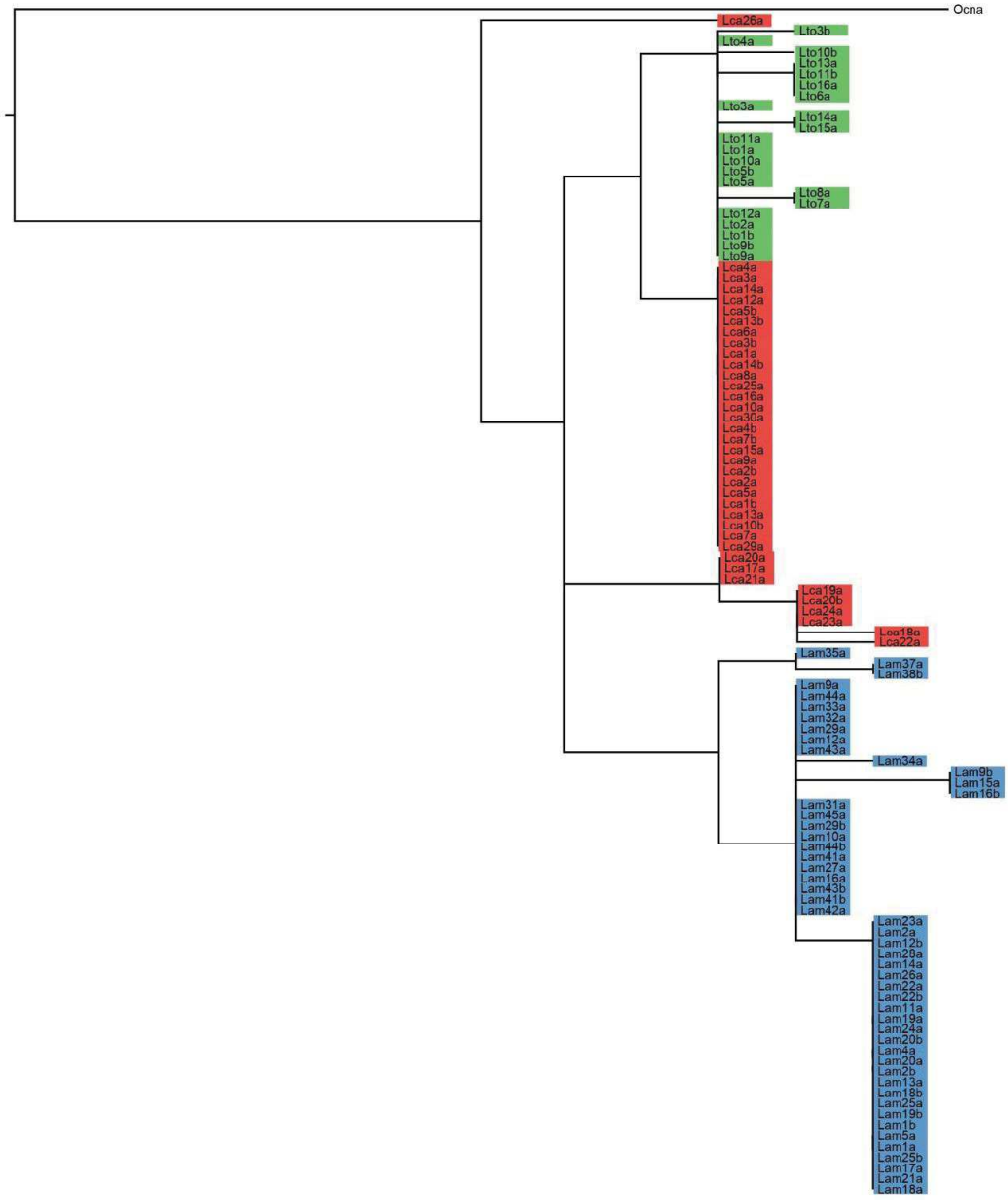


Figure S2.2 (cont'd)

SRY

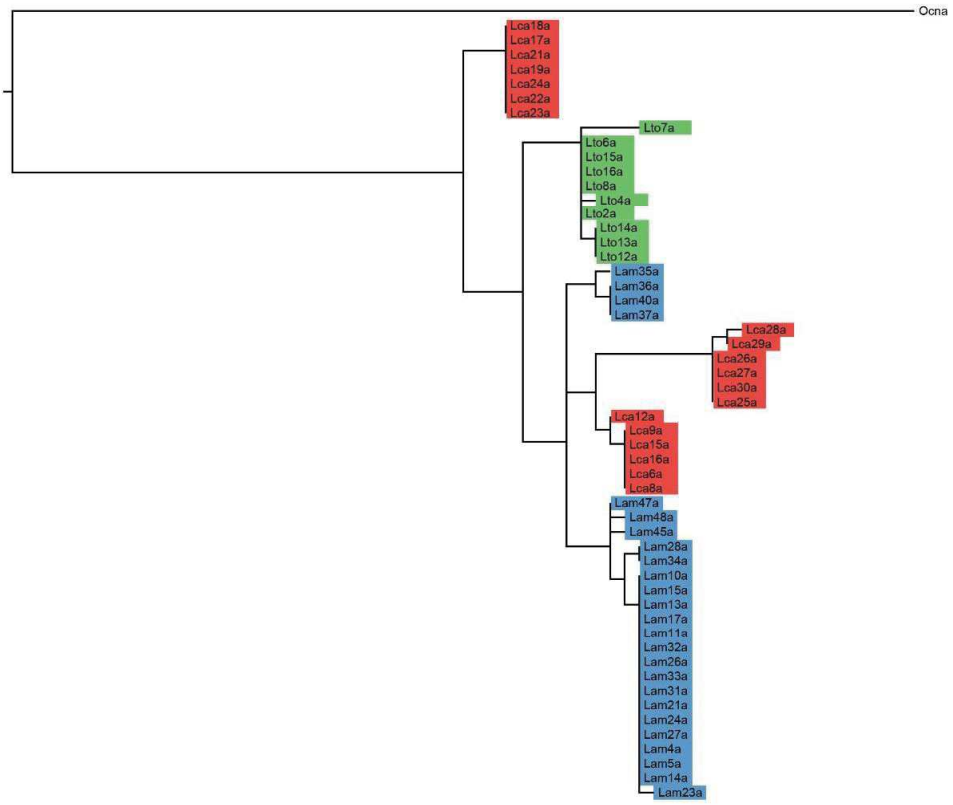


Figure S2.2 (cont'd)

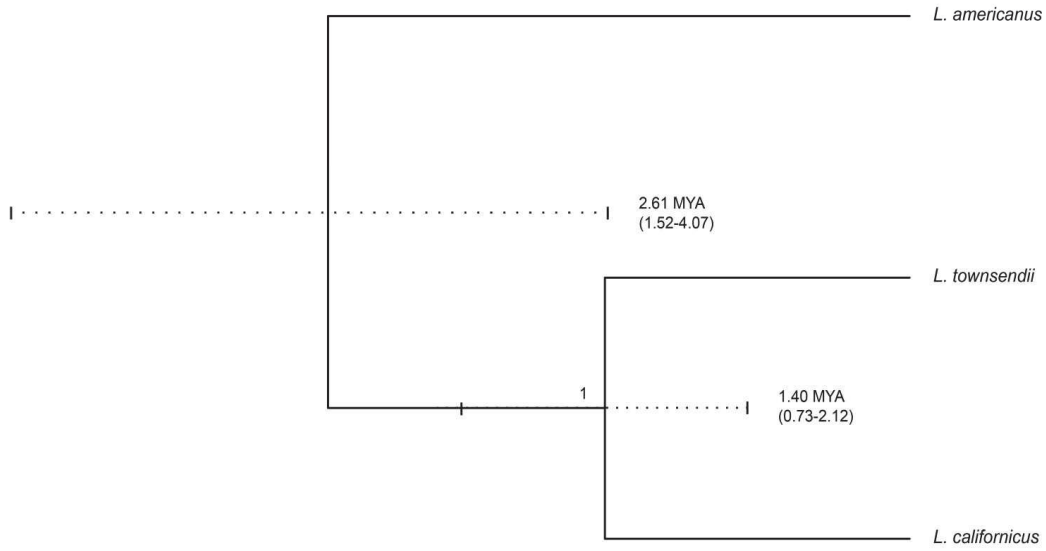
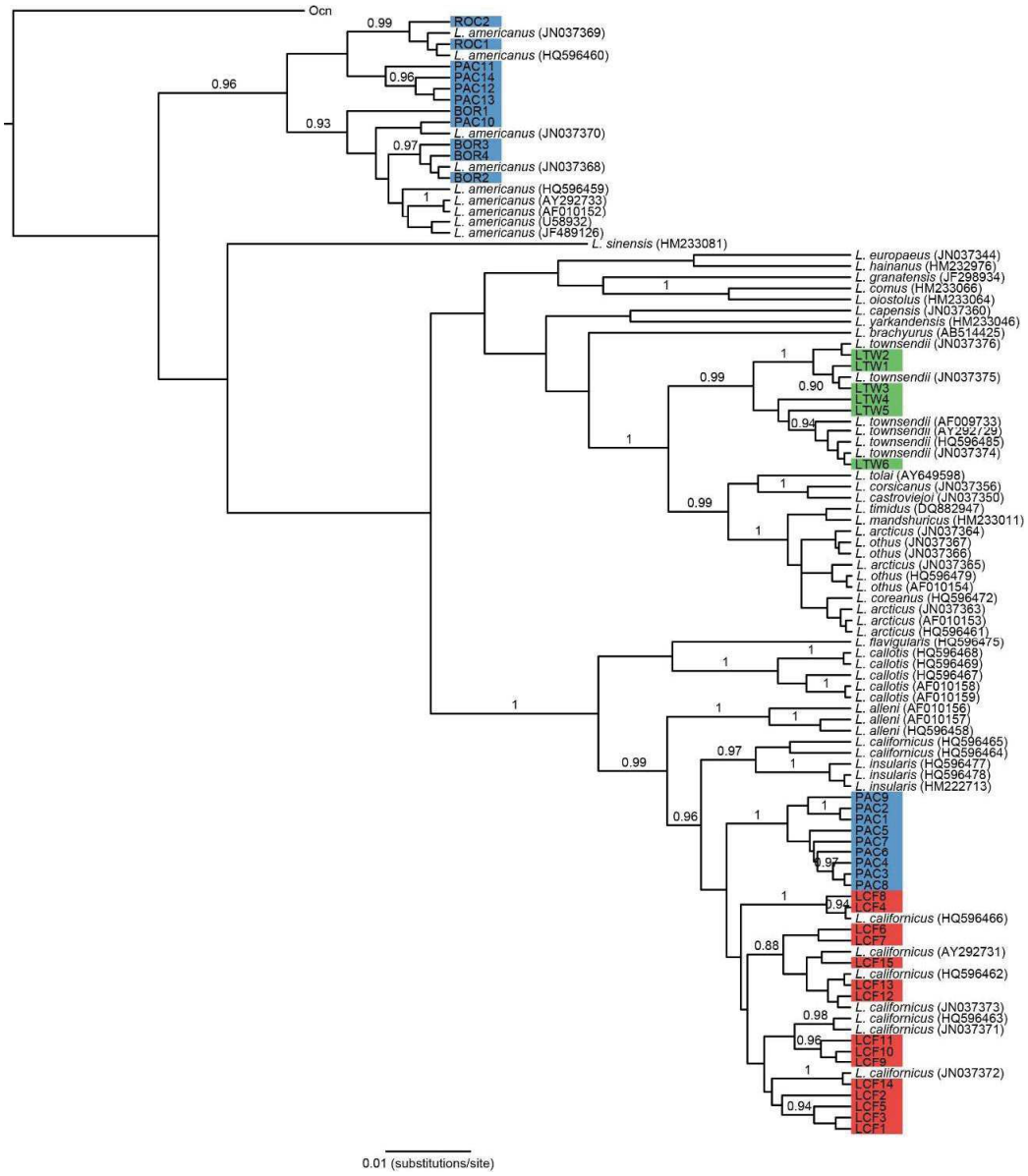


Figure S2.3 Species tree of *L. californicus*, *L. townsendii*, and *L. americanus* inferred with *BEAST. Numbers above branches indicate the posterior probabilities and dashed line the 95% confidence intervals of node ages (mean value and 95% CI are indicated next to the line).

a)


Ocn
ROC2
L. americanus (JN037369)
ROC1
L. americanus (HQ596460)
PAC11
PAC14
PAC12
PAC13
BOR1
PAC10
L. americanus (JN037370)
BOR3
BOR4
L. americanus (JN037368)
BOR2
L. americanus (HQ596459)
L. americanus (AY292733)
L. americanus (AF010152)
L. americanus (U58932)
L. americanus (JF489126)
L. sinensis (HM233081)
L. europaeus (JN037344)
L. hainanus (HM232976)
L. granatensis (JF298934)
L. comus (HM233066)
L. oiostolus (HM233064)
L. capensis (JN037360)
L. yarkandensis (HM233046)
L. brachyurus (AB514425)
L. townsendii (JN037376)
LTW2
LTW1
L. townsendii (JN037375)
LTW3
LTW4
LTW5
L. townsendii (AF009733)
L. townsendii (AY292729)
L. townsendii (HQ596495)
L. townsendii (JN037374)
LTW6
L. tolai (AY649598)
L. corsicanus (JN037356)
L. castroviejoi (JN037350)
L. timidus (DQ882947)
L. mandshuricus (HM233011)
L. arcticus (JN037364)
L. othus (JN037367)
L. othus (JN037366)
L. arcticus (JN037365)
L. othus (HQ596479)
L. othus (AF010154)
L. coreanus (HQ596472)
L. arcticus (JN037363)
L. arcticus (AF010153)
L. arcticus (HQ596461)
L. flavigularis (HQ596475)
L. callotis (HQ596468)
L. callotis (HQ596469)
L. callotis (HQ596467)
L. callotis (AF010158)
L. callotis (AF010159)
L. alleni (AF010156)
L. alleni (AF010157)
L. alleni (HQ596458)
L. californicus (HQ596465)
L. californicus (HQ596464)
L. insularis (HQ596477)
L. insularis (HQ596478)
L. insularis (HM222713)
PAC9
PAC2
PAC1
PAC5
PAC7
PAC6
PAC4
PAC3
PAC8
LCF8
LCF4
L. californicus (HQ596466)
LCF6
LCF7
L. californicus (AY292731)
LCF15
L. californicus (HQ596462)
LCF13
LCF12
L. californicus (JN037373)
L. californicus (HQ596463)
L. californicus (JN037371)
LCF11
LCF10
LCF9
L. californicus (JN037372)
LCF14
LCF2
LCF5
LCF3
LCF1
0.01 (substitutions/site)

b)

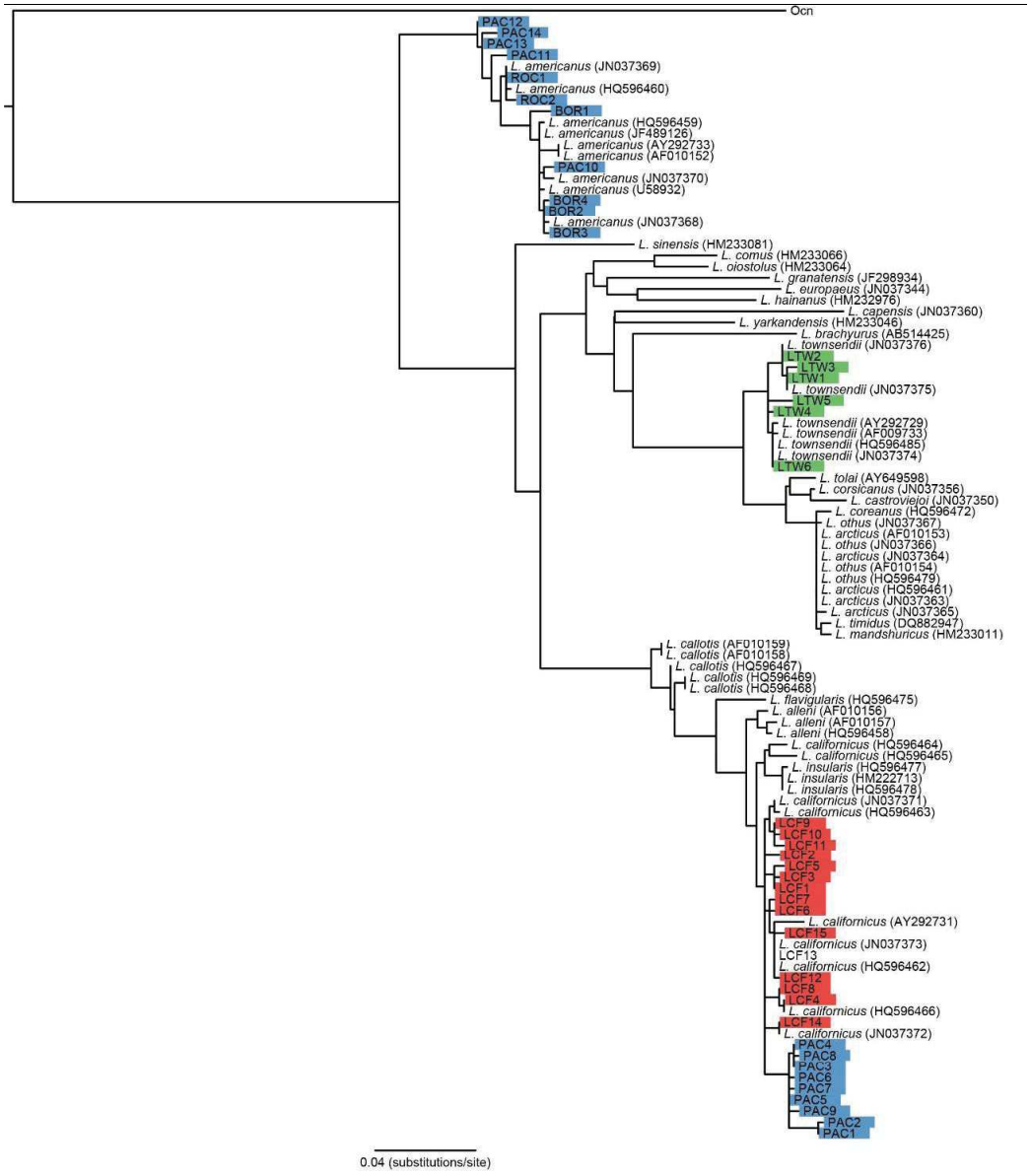


Figure S2.4 Cytochrome b Bayesian Inference (a) and Maximum Likelihood (b) phylogenies of all North American hare species and one sequence representative of each non North American hare species available in GenBank. GenBank sequences are represented on the tree by the name of the species they are reported to belong to and their respective accession number within brackets. Only cytochrome b haplotypes of sequences produced in this study were included (codes are as in Sup. Table 1) and coloured shades indicate the species to which they belong: blue - *L. americanus*; red - *L. californicus*; green - *L. townsendii*. Posterior probabilities greater than 0.9 are given above the branches.

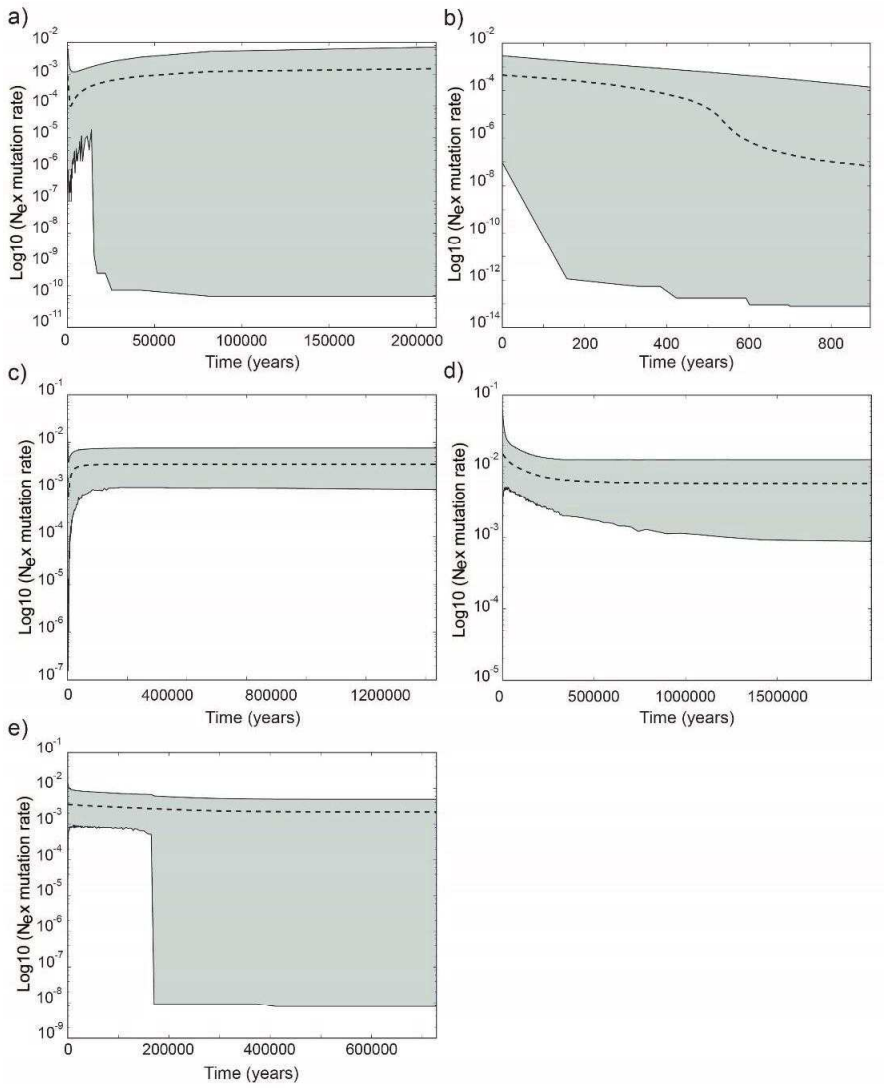


Figure S2.5 Demographic profiles of *L. americanus* Boreal (a), *L. americanus* Rockies (b), *L. americanus* Pacific Northwest (c) population cluster, *L. californicus* (d), and *L. townsendii* (e), based on Extended Bayesian Skyline Plot analyses. Time is in units of years before the present.

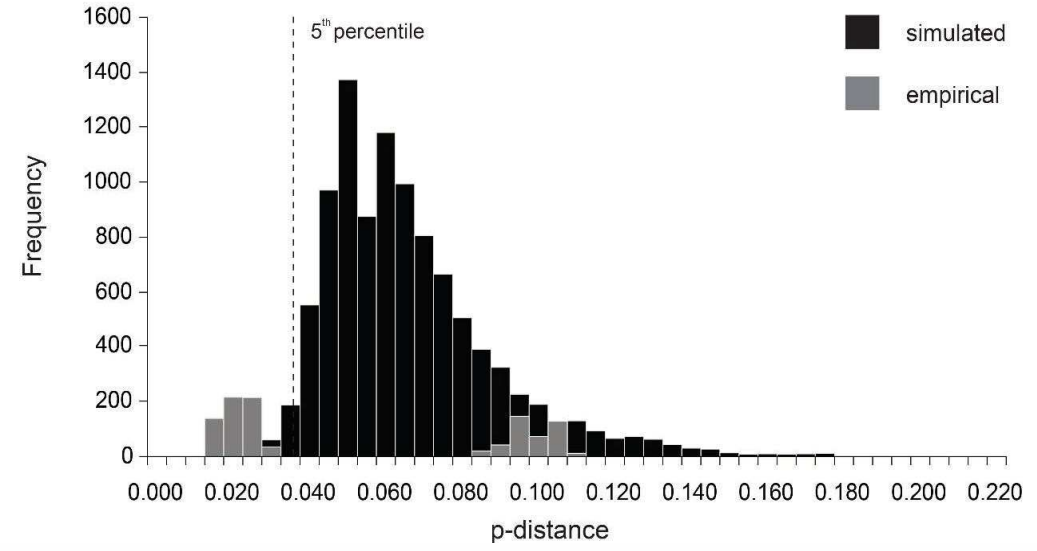


Figure S2.6 Empirical (grey bars) and simulated (black bars) mtDNA distances between *L. californicus* and the PacNW2 group of *L. americanus*. Simulations were performed under the assumption of no gene flow and using the highest 95% HPD estimates of current and ancestral population sizes and the lowest 95% HPD estimate of divergence time obtained under the IM model.

Table S2.1 Detailed information of sequences obtained per individual with GenBank accession numbers (new sequences in bold).

Species	Sample Locations		Specimens				Autossomes					X chr		Y chr	mtDNA
	State/Province (Country)	Pop. Code	Nr.	Code	Original Lab Codes	Sex[a]	SPTBN1	PRKCI	DARC	KITLG	TF	POLA1	GRIA3	SRY[b]	CYTB[c]
Snowshoe Hare	California,	CA1	1	Lam1	*CHENG9	F	**KM260760**	**KM260850**	**KM260937**	**KM261027**	**KM261117**	**KM261208**	**KM261298**		**KM261435** (PAC1)
- *L. americanus* -	(U.S.A.)		2	Lam2	*CHEYNE1	F	**KM260761**	**KM260851**	**KM260938**	**KM261028**	**KM261118**	**KM261209**	**KM261299**		KF781408 (PAC1)
			3	Lam3	*CHENG11	F	**KM260762**	**KM260852**	**KM260939**	**KM261029**	**KM261119**	**KM261210**	-		**KM261436** (PAC1)
			4	Lam4	*CHENG12	M	**KM260763**	**KM260853**	**KM260940**	**KM261030**	**KM261120**	**KM261211**	**KM261300**	**KM261380**	**KM261437** (PAC2)
			5	Lam5	*CHENG13	M	**KM260764**	**KM260854**	**KM260941**	**KM261031**	**KM261121**	**KM261212**	**KM261301**	**KM261381**	KF781404 (PAC1)
			6	Lam6	*SIMONS3	M	-	-	-	-	-	-	-	-	**KM261438** (PAC1)
			7	Lam7	*SIMONS4	F	-	-	-	-	-	-	-		**KM261439** (PAC1)
			8	Lam8	*SIMONS5	M	**KM260765**	-	**KM260942**	**KM261032**	**KM261122**	**KM261213**	-	-	KF781423 (PAC1)
	Washington,	WA1	9	Lam9	*CHENG14	F	**KM260766**	**KM260855**	**KM260943**	**KM261033**	**KM261123**	**KM261214**	**KM261302**		**KM261440** (PAC3)
	(U.S.A.)		10	Lam10	*CHENG15	M	**KM260767**	**KM260856**	**KM260944**	**KM261034**	**KM261124**	**KM261215**	**KM261303**	**KM261382**	**KM261441** (PAC4)
			11	Lam11	*CHENG19	M	**KM260768**	**KM260857**	**KM260945**	**KM261035**	**KM261125**	**KM261216**	**KM261304**	**KM261383**	**KM261442** (PAC5)
			12	Lam12	*MACCRAC77	F	**KM260769**	**KM260858**	**KM260946**	**KM261036**	**KM261126**	**KM261217**	**KM261305**		**KM261443** (PAC6)
			13	Lam13	*MACCRAC80	M	**KM260770**	**KM260859**	**KM260947**	**KM261037**	**KM261127**	**KM261218**	**KM261306**	**KM261384**	**KM261444** (PAC6)
			14	Lam14	*MACCRAC82	M	**KM260771**	**KM260860**	**KM260948**	**KM261038**	**KM261128**	**KM261219**	**KM261307**	**KM261385**	**KM261445** (PAC7)
			15	Lam15	*STRAUSER51	M	**KM260772**	**KM260861**	**KM260949**	**KM261039**	**KM261129**	**KM261220**	**KM261308**	**KM261386**	**KM261446** (PAC8)
			16	Lam16	*STRAUSER52	F	**KM260773**	**KM260862**	**KM260950**	**KM261040**	**KM261130**	**KM261221**	**KM261309**		**KM261447** (PAC3)
	Washington,	WA4	17	Lam17	*MACCRAC73	M	**KM260774**	**KM260863**	**KM260951**	**KM261041**	**KM261131**	**KM261222**	**KM261310**	**KM261387**	**KM261448** (PAC9)
	(U.S.A.)		18	Lam18	*MACCRAC83	F	**KM260775**	**KM260864**	**KM260952**	**KM261042**	**KM261132**	**KM261223**	**KM261311**		**KM261449** (PAC1)
			19	Lam19	*MACCRAC85	F	**KM260776**	**KM260865**	**KM260953**	**KM261043**	**KM261133**	**KM261224**	**KM261312**		**KM261450** (PAC1)
			20	Lam20	*MACCRAC94	F	**KM260777**	**KM260866**	**KM260954**	**KM261044**	**KM261134**	**KM261225**	**KM261313**		**KM261451** (PAC9)
			21	Lam21	*MACCRAC102	M	**KM260778**	**KM260867**	**KM260955**	**KM261045**	**KM261135**	**KM261226**	**KM261314**	**KM261388**	**KM261452** (PAC1)
			22	Lam22	*MACCRAC105	F	**KM260779**	**KM260868**	**KM260956**	**KM261046**	**KM261136**	**KM261227**	**KM261315**		**KM261453** (PAC1)
			23	Lam23	*STRAUSER34	M	**KM260780**	**KM260869**	**KM260957**	**KM261047**	**KM261137**	**KM261228**	**KM261316**	**KM261389**	**KM261454** (PAC9)
			24	Lam24	*STRAUSER35	M	**KM260781**	**KM260870**	**KM260958**	**KM261048**	**KM261138**	**KM261229**	**KM261317**	**KM261390**	**KM261455** (PAC1)
			25	Lam25	*STRAUSER36	F	**KM260782**	**KM260871**	**KM260959**	**KM261049**	**KM261139**	**KM261230**	**KM261318**		**KM261456** (PAC9)
			26	Lam26	*STRAUSER40	M	**KM260783**	**KM260872**	**KM260960**	**KM261050**	**KM261140**	**KM261231**	**KM261319**	**KM261391**	**KM261457** (PAC1)
	Oregon	OR2	27	Lam27	*STRAUSER85	M	**KM260784**	**KM260873**	**KM260961**	**KM261051**	**KM261141**	**KM261232**	**KM261320**	**KM261392**	**KM261458** (PAC1)
	(U.S.A.)		28	Lam28	*STRAUSER88	M	**KM260785**	**KM260874**	**KM260962**	**KM261052**	**KM261142**	**KM261233**	**KM261321**	**KM261393**	**KM261459** (PAC1)
			29	Lam29	*STRAUSER89	F	**KM260786**	**KM260875**	**KM260963**	**KM261053**	**KM261143**	**KM261234**	**KM261322**		**KM261460** (PAC1)
			30	Lam30	*STRAUSER92	F	**KM260787**	**KM260876**	**KM260964**	**KM261054**	**KM261144**	**KM261235**	-		**KM261461** (PAC1)
			31	Lam31	*STRAUSER94	M	**KM260788**	**KM260877**	**KM260965**	**KM261055**	**KM261145**	**KM261236**	**KM261323**	**KM261394**	**KM261462** (PAC1)
			32	Lam32	*STRAUSER95	M	**KM260789**	**KM260878**	**KM260966**	**KM261056**	**KM261146**	**KM261237**	**KM261324**	**KM261395**	**KM261463** (PAC1)
			33	Lam33	*STRAUSER97	M	**KM260790**	**KM260879**	**KM260967**	**KM261057**	**KM261147**	**KM261238**	**KM261325**	**KM261396**	**KM261464** (PAC1)
			34	Lam34	*STRAUSER101	M	**KM260791**	**KM260880**	**KM260968**	**KM261058**	**KM261148**	**KM261239**	**KM261326**	**KM261397**	**KM261465** (PAC1)

	Saskatchewan,	SK1	35	Lam35	*GORDON1	M	**KM260792**	**KM260881**	**KM260969**	**KM261059**	**KM261149**	**KM261240**	**KM261327**	**KM261398**	**KM261466** (BOR1)
	Canada		36	Lam36	*GORDON25	M	**KM260793**	**KM260882**	**KM260970**	**KM261060**	**KM261150**	**KM261241**	**KM261328**	**KM261399**	**KM261467** (BOR2)
			37	Lam37	*GORDON7	M	**KM260794**	**KM260883**	**KM260971**	**KM261061**	**KM261151**	**KM261242**	**KM261329**	**KM261400**	**KM261468** (BOR3)

Table S2.1 (cont'd)

			38	Lam38	*WEBER1	F	**KM260795**	**KM260884**	**KM260972**	**KM261062**	**KM261152**	**KM261243**	**KM261330**		**KM261469** (BOR4)
			39	Lam39	*WEBER2	F	-	**KM260885**	-	-	**KM261153**	**KM261244**	-		-
			40	Lam40	*WEBER3	M	**KM260796**	**KM260886**	**KM260973**	**KM261063**	**KM261154**	**KM261245**	**KM261331**	**KM261401**	**KM261470** (BOR2)
	Wyoming,	WY1	41	Lam41	*MILLS_R1405	F	**KM260797**	**KM260888**	**KM260974**	**KM261064**	**KM261155**	**KM261246**	**KM261332**		**KM261471** (ROC1)
	(U.S.A.)		42	Lam42	*MILLS_R1573	M	**KM260798**	**KM260887**	**KM260975**	**KM261065**	**KM261156**	**KM261247**	**KM261333**	**KM261402**	**KM261472** (ROC1)
			43	Lam43	*MILLS_R1676	F	**KM260799**	**KM260890**	**KM260976**	**KM261066**	**KM261157**	**KM261248**	**KM261334**		**KM261473** (ROC2)
			44	Lam44	*MILLS_R1791	F	**KM260800**	**KM260889**	**KM260977**	**KM261067**	**KM261158**	**KM261249**	**KM261335**		KF781413 (ROC1)
			45	Lam45	*NBERG10	M	-	**KM260891**	**KM260978**	**KM261068**	**KM261159**	**KM261250**	**KM261336**	**KM261403**	**KM261474** (ROC1)
			46	Lam46	*NBERG29	M	**KM260801**	-	-	-	-	-	-	-	-
			47	Lam47	*NBERG33	M	**KM260802**	-	**KM260979**	**KM261069**	**KM261160**	-	-	**KM261404**	**KM261475** (ROC1)
			48	Lam48	*NBERG6	M	**KM260803**	**KM260892**	**KM260980**	**KM261070**	**KM261161**	**KM261251**	-	**KM261405**	KF781358 (ROC1)
Black-tailed	Oregon,	LCA_OR	49	Lca1	*BURKE2	F	**KM260804**	**KM260893**	**KM260981**	**KM261071**	**KM261162**	**KM261252**	**KM261337**		**KM261476** (LCF1)
Jackrabbit	(U.S.A.)		50	Lca2	*BURKE3	F	**KM260805**	**KM260894**	**KM260982**	**KM261072**	**KM261163**	**KM261253**	**KM261338**		**KM261477** (LCF2)
- *L. californicus* -			51	Lca3	*BURKE5	F	**KM260806**	**KM260895**	**KM260983**	**KM261073**	**KM261164**	**KM261254**	**KM261339**		**KM261478** (LCF1)
			52	Lca4	*BURKE6	F	**KM260807**	**KM260896**	**KM260984**	**KM261074**	**KM261166**	**KM261255**	**KM261340**		**KM261479** (LCF3)
			53	Lca5	*BURKE7	F	**KM260808**	**KM260897**	**KM260985**	**KM261075**	**KM261165**	**KM261256**	**KM261341**		**KM261480** (LCF1)
			54	Lca6	*HENNINGS1	M	**KM260809**	**KM260898**	**KM260986**	**KM261076**	**KM261167**	**KM261257**	**KM261342**	**KM261406**	**KM261481** (LCF4)
			55	Lca7	*HENNINGS2	F	**KM260810**	**KM260899**	**KM260987**	**KM261077**	**KM261168**	**KM261258**	**KM261343**		**KM261482** (LCF1)
			56	Lca8	*HENNINGS3	M	**KM260811**	**KM260900**	**KM260988**	**KM261078**	**KM261169**	**KM261259**	**KM261344**	**KM261407**	**KM261483** (LCF5)
			57	Lca9	*HENNINGS4	M	**KM260812**	**KM260901**	**KM260989**	**KM261079**	**KM261170**	**KM261260**	**KM261345**	**KM261408**	**KM261484** (LCF5)
			58	Lca10	*HENNINGS5	F	**KM260813**	**KM260902**	**KM260990**	**KM261080**	**KM261171**	**KM261261**	**KM261346**		**KM261485** (LCF4)
	California,	LCA_CA	59	Lca11	*BURKE11	M	**KM260814**	**KM260903**	**KM260991**	**KM261081**	**KM261172**	**KM261262**	-	-	**KM261486** (LCF6)
	(U.S.A.)		60	Lca12	*BAUER1	M	**KM260815**	**KM260904**	**KM260992**	**KM261082**	**KM261173**	**KM261263**	**KM261347**	**KM261409**	**KM261487** (LCF7)
			61	Lca13	*BAUER2	F	**KM260816**	**KM260905**	**KM260993**	**KM261083**	**KM261174**	**KM261264**	**KM261348**		**KM261488** (LCF8)
			62	Lca14	*BAUER7	F	**KM260817**	**KM260906**	**KM260994**	**KM261084**	**KM261175**	**KM261265**	**KM261349**		**KM261489** (LCF8)
			63	Lca15	*BAUER9	M	**KM260818**	**KM260907**	**KM260995**	**KM261085**	**KM261176**	**KM261266**	**KM261350**	**KM261410**	**KM261490** (LCF1)
			64	Lca16	*BAUER14	M	**KM260819**	**KM260908**	**KM260996**	**KM261086**	**KM261177**	**KM261267**	**KM261351**	**KM261411**	**KM261491** (LCF1)
	Texas,	LCA_TE	65	Lca17	*DOWLER1	M	**KM260820**	**KM260909**	**KM260997**	**KM261087**	**KM261178**	**KM261268**	**KM261352**	**KM261412**	**KM261492** (LCF9)
	(U.S.A.)		66	Lca18	*DOWLER2	M	**KM260821**	**KM260910**	**KM260998**	**KM261088**	**KM261179**	**KM261269**	**KM261353**	**KM261413**	**KM261493** (LCF9)
			67	Lca19	*DOWLER5	M	**KM260822**	**KM260911**	**KM260999**	**KM261089**	**KM261180**	**KM261270**	**KM261354**	**KM261414**	**KM261494** (LCF1)
			68	Lca20	*DOWLER7	F	**KM260823**	**KM260912**	**KM261000**	**KM261090**	**KM261181**	**KM261271**	**KM261355**		**KM261495** (LCF9)
			69	Lca21	*DOWLER9	M	**KM260824**	**KM260913**	**KM261001**	**KM261091**	**KM261182**	**KM261272**	**KM261356**	**KM261415**	**KM261496** (LCF9)
			70	Lca22	*DOWLER10	M	**KM260825**	**KM260914**	**KM261002**	**KM261092**	**KM261183**	**KM261273**	**KM261357**	**KM261416**	**KM261497** (LCF1)
			71	Lca23	*DOWLER11	M	**KM260826**	**KM260915**	**KM261003**	**KM261093**	**KM261184**	**KM261274**	**KM261358**	**KM261417**	**KM261498** (LCF9)

			72	Lca24	*DOWLER15	M	**KM260827**	**KM260916**	**KM261004**	**KM261094**	**KM261185**	**KM261275**	**KM261359**	**KM261418**	**KM261499** (LCF9)
	Arizona,	LCA_AR	73	Lca25	LCF.TUC.2055	M	**KM260828**	**KM260917**	**KM261005**	**KM261095**	**KM261186**	**KM261276**	**KM261360**	**KM261419**	**KM261500** (LCF1)
	(U.S.A.)		74	Lca26	LCF.TUC.2056	M	**KM260829**	**KM260918**	**KM261006**	**KM261096**	**KM261187**	**KM261277**	**KM261361**	**KM261420**	**KM261501** (LCF1)
			75	Lca27	LCF.TUC.2057	M	**KM260830**	**KM260919**	**KM261007**	**KM261097**	**KM261188**	**KM261278**	-	**KM261421**	**KM261502** (LCF1)
			76	Lca28	LCF.TUC.2058	M	**KM260831**	**KM260920**	**KM261008**	**KM261098**	**KM261189**	**KM261279**	-	**KM261422**	**KM261503** (LCF1)
			77	Lca29	LCF.TUC.2059	M	**KM260832**	**KM260921**	**KM261009**	**KM261099**	**KM261190**	**KM261280**	**KM261362**	**KM261423**	**KM261504** (LCF1)
			78	Lca30	LCF.TUC.2060	M	**KM260833**	**KM260922**	**KM261010**	**KM261100**	**KM261191**	**KM261281**	**KM261363**	**KM261424**	**KM261505** (LCF1)

Table S2.1 (cont'd)

White-tailed	Idaho,	LTO_ID1	79	Lto1	LTW.2206	F	**KM260834**	**KM260923**	**KM261011**	**KM261101**	**KM261192**	**KM261282**	**KM261364**		**KM261506** (LTW1)
Jackrabbit	(U.S.A.)		80	Lto2	LTW.2208	M	**KM260835**	**KM260924**	**KM261012**	**KM261102**	**KM261193**	**KM261283**	**KM261365**	**KM261425**	**KM261507** (LTW1)
- L. townsendii -			81	Lto3	LTW.2211	F	**KM260836**	**KM260925**	**KM261013**	**KM261103**	**KM261194**	**KM261284**	**KM261366**		**KM261508** (LTW2)
			82	Lto4	LTW.2212	M	**KM260837**	**KM260926**	**KM261014**	**KM261104**	**KM261195**	**KM261285**	**KM261367**	**KM261426**	**KM261509** (LTW2)
			83	Lto5	LTW.2213	F	**KM260838**	**KM260927**	**KM261015**	**KM261105**	**KM261196**	**KM261286**	**KM261368**		**KM261510** (LTW3)
			84	Lto6	LTW.2214	M	**KM260839**	**KM260928**	**KM261016**	**KM261106**	**KM261197**	**KM261287**	**KM261369**	**KM261427**	**KM261511** (LTW4)
			85	Lto7	LTW.2215	M	**KM260840**	**KM260929**	**KM261017**	**KM261107**	**KM261198**	**KM261288**	**KM261370**	**KM261428**	**KM261512** (LTW1)
			86	Lto8	LTW.2217	M	**KM260841**	**KM260930**	**KM261018**	**KM261108**	**KM261199**	**KM261289**	**KM261371**	**KM261429**	**KM261513** (LTW1)
	Montana,	LTO_MO1	87	Lto9	LTW.MTA.2570	F	**KM260842**	-	**KM261019**	**KM261109**	**KM261200**	**KM261290**	**KM261372**		**KM261514** (LTW5)
	(U.S.A.)		88	Lto10	LTW.MTA.2571	F	**KM260843**	**KM260931**	**KM261020**	**KM261110**	**KM261201**	**KM261291**	**KM261373**		**KM261515** (LTW5)
	Wyoming,	LTO_WY1	89	Lto11	LTW.WYO.2572	F	**KM260844**	**KM260932**	**KM261021**	**KM261111**	**KM261202**	**KM261292**	**KM261374**		**KM261516** (LTW6)
	(U.S.A.)														
	Montana,	LTO_MO2	90	Lto12	LTW.WYO.2575	M	**KM260845**	-	**KM261022**	**KM261112**	**KM261203**	**KM261293**	**KM261375**	**KM261430**	**KM261517** (LTW4)
	(U.S.A.)		91	Lto13	LTW.WYO.2576	M	**KM260846**	**KM260933**	**KM261023**	**KM261113**	**KM261204**	**KM261294**	**KM261376**	**KM261431**	**KM261518** (LTW4)
			92	Lto14	LTW.WYO.2577	M	**KM260847**	**KM260934**	**KM261024**	**KM261114**	**KM261205**	**KM261295**	**KM261377**	**KM261432**	**KM261519** (LTW4)
			93	Lto15	LTW.WYO.2578	M	**KM260848**	**KM260935**	**KM261025**	**KM261115**	**KM261206**	**KM261296**	**KM261378**	**KM261433**	**KM261520** (LTW4)
			94	Lto16	LTW.2085	M	**KM260849**	**KM260936**	**KM261026**	**KM261116**	**KM261207**	**KM261297**	**KM261379**	**KM261434**	**KM261521** (LTW4)
Wild Rabbit			95	Ocn1			JN037052	JN037024	JN036940	JN036996	JN037078	HM028509	HM028196	AY785433	AJ001588
- O.cuniculus -															
Eastern cottontail			96	Sfl											AY292724
- S. floridanus -															

Notes:

[a]F: female; M: male;

[b]only information regarding males is shown;

[c]Cytochrome *b* haplotype names are within brackets (see Sup. Fig. 4);

*sample codes as used in Cheng et al (2014);

"-" denotes missing sequences.

Table S2.2 Geographic coordinates of sampling sites.

Sample Locations	Specimens		Coordinates (WGS84)	
Pop. Code	**Nr.**	**Code**	**Lon.**	**Lat.**
Snowshoe Hare (*L. americanus*)				
CA1	1	Lam1	41.571722	-121.715064
	2	Lam2	41.569627	-121.824790
	3	Lam3	41.557523	-121.832900
	4	Lam4	41.482549	-121.771698
	5	Lam5	41.528527	-121.910745
	6	Lam6	41.284587	-121.879545
	7	Lam7	41.423940	-121.964380
	8	Lam8	41.270580	-122.200240
WA1	9	Lam9	45.365927	-121.570805
	10	Lam10	45.364283	-121.569646
	11	Lam11	45.458621	-121.656680
	12	Lam12	45.632373	-122.122864
	13	Lam13	45.687980	-122.055111
	14	Lam14	45.631773	-122.126851
	15	Lam15	45.371210	-121.568578
	16	Lam16	45.365171	-121.568653
WA4	17	Lam17	47.733598	-120.810582
	18	Lam18	47.751741	-120.792396
	19	Lam19	47.366743	-120.832570
	20	Lam20	47.758922	-120.844936
	21	Lam21	47.731739	-120.801377
	22	Lam22	47.749236	-120.791301
	23	Lam23	47.797118	-121.276684
	24	Lam24	47.798073	-121.274969
	25	Lam25	47.796683	-121.275309
	26	Lam26	47.799824	-121.284941
OR2	27	Lam27	44.551404	-118.361107
	28	Lam28	44.560983	-118.367866
	29	Lam29	44.563736	-118.372489
	30	Lam30	44.552151	-118.360356
	31	Lam31	44.552927	-118.366267
	32	Lam32	44.562954	-118.371639
	33	Lam33	44.559028	-118.366498
	34	Lam34	44.559191	-118.367194
SK1	35	Lam35	53.750000	-104.150000
	36	Lam36	53.750000	-104.150000
	37	Lam37	53.750000	-104.150000
	38	Lam38	53.777530	-106.915510
	39	Lam39	53.777530	-106.915510
	40	Lam40	53.800650	-106.920320
WY1	41	Lam41	44.147605	-110.673523
	42	Lam42	44.625319	-110.852888
	43	Lam43	44.147605	-110.673523
	44	Lam44	44.147605	-110.673523
	45	Lam45	43.761417	-110.507404
	46	Lam46	43.761417	-110.507404
	47	Lam47	43.761417	-110.507404
	48	Lam48	43.761417	-110.507404

Sample Locations	Specimens		Coordinates (WGS84)	
Pop. Code	**Nr.**	**Code**	**Lon.**	**Lat.**
Black-tailed Jackrabbit (*L. californicus*)				
LCA_OR	49	Lca1	43.372083	-121.074000
	50	Lca2	43.372083	-121.074000
	51	Lca3	43.372083	-121.074000
	52	Lca4	43.268360	-120.783810
	53	Lca5	43.257870	-120.636950
	54	Lca6	42.724412	-120.685600
	55	Lca7	42.724412	-120.685600
	56	Lca8	42.719928	-120.589900
	57	Lca9	42.723173	-120.637400
	58	Lca10	43.130588	-120.919640
LCA_CA	59	Lca11	38.376460	-121.362070
	60	Lca12	40.912994	-121.712122
	61	Lca13	40.912994	-121.712122
	62	Lca14	40.912994	-121.712122
	63	Lca15	40.365414	-120.431314
	64	Lca16	40.365414	-120.431314
LCA_TE	65	Lca17	31.257230	-100.683170
	66	Lca18	31.257230	-100.683170
	67	Lca19	31.257230	-100.683170
	68	Lca20	31.257230	-100.683170
	69	Lca21	31.257230	-100.683170
	70	Lca22	31.257230	-100.683170
	71	Lca23	31.257230	-100.683170
	72	Lca24	31.257230	-100.683170
LCA_AR	73	Lca25	31.898333	-110.547222
	74	Lca26	31.898333	-110.547222
	75	Lca27	31.898333	-110.547222
	76	Lca28	31.898333	-110.547222
	77	Lca29	31.898333	-110.547222
	78	Lca30	31.898333	-110.547222
White-tailed Jackrabbit (*L. townsendii*)				
LTO_ID1	79	Lto1*	45.036391	-113.923045
	80	Lto2*	45.036391	-113.923045
	81	Lto3*	45.036391	-113.923045
	82	Lto4*	45.036391	-113.923045
	83	Lto5*	45.036391	-113.923045
	84	Lto6*	45.036391	-113.923045
	85	Lto7*	45.036391	-113.923045
	86	Lto8*	45.036391	-113.923045
LTO_MO1	87	Lto9	47.700145	-107.167736
	88	Lto10	47.700145	-107.167736
LTO_WY1	89	Lto11	42.560090	-109.503270
	90	Lto12	45.032834	-110.728905
LTO_MO2	91	Lto13	44.991048	-110.694458
	92	Lto14	44.978354	-110.692784
	93	Lto15	44.986834	-110.693521
	94	Lto16	45.034116	-110.732068

*Approximate coordinates.

Table S2.3 Analysed loci, PCR conditions and primers.

Loci		PCR conditions		PCR primers		
Nr.	Symbol	AT[a]	E[b]	Forward(F)/Reverse(R) - (5'-3')		Reference
Autosomes						
1	SPTBN1	65	45"	F	*TGATAGCAGAACTCCATGTGG	Matthee et al. 2004
				R	CTCTGCCCAGAAGTTTGCAAC	Matthee et al. 2004
2	PRKCI	58	45"	F	*AAACAGATCGCATTTATGCAAT	Matthee et al. 2004
				R	TGTCTGTACCCAGTCAATATC	Matthee et al. 2004
3	KITLG	56	45"	F	*AAATATCAGTCTTGAATCTTAC	Matthee et al. 2004
				R	TTTTAGATGAATTACAGTGTCC	Matthee et al. 2004
4	TF	56	45"	F	*GCCTTTGTCAAGCAAGAGACC	Matthee et al. 2004
				R	CACAGCAGCTCATACTGATCC	Matthee et al. 2004
5	DARC	56	45"	F	*CTCTCAGTTGACCCAAATTC	Melo-Ferreira et al. 2009
				R	GCCTTTAATTCAGGTTGACG	Melo-Ferreira et al. 2009
X chromosome						
6	POLA1	57[c]	1'30"	F	*GGTATTTCTGTTTGGCAAGGTTTG	Carneiro et al. 2010
				R	*CTTGGACTTGAATTTCATGATTC	Carneiro et al. 2010
7	GRIA3	57[c]	1'30"	F	*CTCAGATCAGCAAATCAGCAATG	Carneiro et al. 2010
				R	*CATAGGCTAAGTCTACACAATAG	Carneiro et al. 2010
Y chromosome						
8	SRY	56[c]	1'30"	F	*CATGCTTTGAGGCAAATGAATAAC	This work
				R	*TTTTGAACCTTGAACTTGGCATC	This work
				F	*CTGTTGCAGCATGCTTTGAG	Melo-Ferreira et al. 2009
				R	*GATTTGACGAATGCCAAGTGTTTC	Melo-Ferreira et al. 2009
Mitochondrial						
9	CYTB	50"	30"	F	*AGCCTGATGAAACTTTGGCTC	Alves et al. 2003
				R	GGATTTTATTCTCGACTAAGC	Alves et al. 2003
				R	CCGAGAAGGTCAGGAGAGAA	Melo-Ferreira et al. 2005
				R	GTTGGCAGGGGTGTAGTTGT	This work

[a]Annealing temperature;

[b]Extension step length;

[c]PCR amplification of this locus required an initial touchdown phase with a decrease of the annealing temperature of 0.5ºC per cicle starting at 64ºC;

*Sequencing primer.

Table S2.4 Posterior probabilities for models of taxa delimitation estimated using BP&P using different combinations of ancestral effective population size (θ) and root age (τ_0) priors.

		Model[2]				
θ prior[1]	τ_0 prior[1]	1:2:3:4:5	1:2:3:(4+5)	1:2:(3+4):5	1:2:(3+5):4	1:2:(3+4+5)
(2.0, 2000)	(2.0, 2000)	1.0	0.0	0.0	0.0	0.0
(2.0, 2000)	(1.0, 10)	1.0	0.0	0.0	0.0	0.0
(0.02, 20)	(0.02, 20)	1.0	0.0	0.0	0.0	0.0
(0.02, 20)	(0.1, 1.0)	1.0	0.0	0.0	0.0	0.0
(1.0, 10)	(2.0, 2000)	1.0	0.0	0.0	0.0	0.0
(1.0, 10)	(1.0, 10)	1.0	0.0	0.0	0.0	0.0
(0.1, 1.0)	(0.02, 20)	1.0	0.0	0.0	0.0	0.0
(0.1, 1.0)	(0.1, 1.0)	1.0	0.0	0.0	0.0	0.0

[1]Gamma distribution of the priors, considering small ancestral effective population sizes or shallow divergence (2.0, 2000), (0.02, 20), and large effective population sizes or deep divergence (1.0, 10), (0.1, 1.0); [2]Models of taxa delimitation: 1 – *L. californicus*; 2 – *L. townsendii*; 3 – Boreal, *L. americanus*; 4 – Pacific Northwest, *L. americanus*; 5 – Rockies, *L. americanus*. Posterior probabilities were identical across independent replicate runs.

Table S2.5 ML estimates (95% posterior density intervals in parentheses) of demographic parameters obtained with IMa2 between pairs of species (population clusters of *L. americanus* not considered).

Pop. 1	Pop. 2	N_{e1}[1]	N_{e2}[1]	N_{eA}[1]	t[2]	$2Nm_1$[3]	$2Nm_2$[3]
Lam	Lca	376140	587982	181149	3153556	0.0566*	0.0031
		(288274, 484468)	(456784, 750474)	(0, 628905)	(2026942, 4492012)	(0.0000, 0.2240)	(0.0000, 0.1676)
Lam	Lto	319929	204861	-	2447737	0.0246	0.0226*
		(243377, 417305)	(138781, 298264)	-	-	(0.0000, 0.1794)	(0.0000, 0.1964)
Lca	Lto	641424	228813	264923	1357714	0.0033	0.0012
		(491570, 830998)	(152984, 334494)	(91622, 550187)	(856997, 2166565)	(0.0000, 0.4184)	(0.0000, 0.2241)

Lam: *L. americanus*, Lca: *L. californicus*; Lto – *L. townsendii*; Missing values correspond to cases where parameters could not be reliably estimated; [1]Effective population size of population 1 (N_{e1}), 2 (N_{e2}), and the ancestral population (N_{eA}); [2]Time in years since species 1 and 2 split (calibrated using a rabbit-hare divergence of 11.8 My; Matthee *et al.* 2004); [3]Population migration rate into population 1 ($2Nm_1$) and population 2 ($2Nm_2$) (*significant values, $P < 0.05$; Nielsen & Wakeley 2001).

Table S2.6 ML estimates (95% posterior density intervals in parentheses) of demographic parameters obtained with IMa2 among the three species (topology is Pop 1;(Pop 2;Pop 3)).

Pop. 1	Pop. 2	Pop. 3	N_{e1}[1]	N_{e2}[1]	N_{e3}[1]	N_{eA1}[1]	N_{eA2}[1]	t1[2]	t2[2]
Lam	Lca	Lto	356520	601463	226647	-	-	1452080	-
			(272987, 461839)	(479894, 601463)	(149012, 336660)	-	-	(931623, 2288856)	-
Lca	Lam	Lto	558372	360613	205342	-	193667	2725058	3090966
			(425610, 722670)	(275756, 466293)	(134086, 306449)	-	-	(1704365, 3668717)	-
Lto	Lam	Lca	199083	357002	572816	-	189093	2378407	3264292
			(129512, 295616)	(272145, 462682)	(436443, 746984)	-	-	(1454006, 3495392)	-

Pop. 1	Pop. 2	Pop. 3	$2Nm_{1(2)}$[3]	$2Nm_{2(1)}$[3]	$2Nm_{1(3)}$[3]	$2Nm_{3(1)}$[3]	$2Nm_{2(3)}$[3]	$2Nm_{3(2)}$[3]	$2Nm_{1(4)}$[3]	$2Nm_{4(1)}$[3]
Lam	Lca	Lto	0.0244	-	-	-	-	0.0240	-	-
			(0.0000, 0.1679)	-	-	-	-	(0.0000, 0.2776)	-	-
Lca	Lam	Lto	-	0.0234	-	-	-	-	-	-
			-	(0.0000, 0.1719)	-	-	-	-	-	-
Lto	Lam	Lca	-	-	-	0.0922	0.0278	-	-	-
			-	-	-	(0.0000, 0.7177)	(0.0000, 0.1802)	-	-	-

Lam: *L. americanus*, Lca: *L. californicus*; Lto – *L. townsendii*; Missing values correspond to cases where parameters could not be reliably estimated; [1]Effective population size of population 1 (N_{e1}), 2 (N_{e2}), 3 (N_{e3}), and the ancestral populations of populations 2 and 3(N_{eA1}) and of all populations (N_{eA2}); [2]Time in years since the first split between species 1 and the ancestral of 2 and 3 (t2), and since 2 and 3 split (t1) (calibrated using a rabbit-hare divergence of 11.8 Mya; Matthee *et al.* 2004); [3]Population migration rate from population Y into population X ($2Nm_{X(Y)}$) (*significant values, $P < 0.05$; Nielsen & Wakeley 2001).

Table S2.7 ML estimates (95% posterior density intervals in parentheses) of demographic parameters obtained with IMa2 among the three *L. americanus* groups (topology is Pop 1;(Pop 2;Pop 3)).

Pop. 1	Pop. 2	Pop. 3	N_{e1}[1]	N_{e2}[1]	N_{e3}[1]	N_{eA1}[1]	N_{eA2}[1]	$t1$[2]	$t2$[2]
Lam-Bor	Lam-Roc	Lam-PacNW	207148	-	-	-	-	1916207	2840608
			(118704, 359770)	(30236, 97941)	(30236, 101564)	-	-	-	-
Lam-Roc	Lam-Bor	Lam-PacNW	45582	228813	65442	-	-	2104939	7110186
			(23026, 83497)	(129512, 406712)	-	-	-		
Lam-PacNW	Lam-Bor	Lam-Roc	64552	228813	46485	-	-	1968204	-
			(33847, 105175)	(127707, 409360)	(23026, 84412)	-	-	(949630, 3852634)	-

Pop. 1	Pop. 2	Pop. 3	$2Nm_{1(2)}$[3]	$2Nm_{2(1)}$[3]	$2Nm_{1(3)}$[3]	$2Nm_{3(1)}$[3]	$2Nm_{2(3)}$[3]	$2Nm_{3(2)}$[3]	$2Nm_{1(A)}$[3]	$2Nm_{A(1)}$[3]
Lam-Bor	Lam-Roc	Lam-PacNW	-	-	-	-	-	-	-	-
			-	-	-	-	-	-	-	-
Lam-Roc	Lam-Bor	Lam-PacNW	0.0061	-	0.6974*	-	-		-	-
			-	-	-	-	-	-	-	-
Lam-PacNW	Lam-Bor	Lam-Roc	-	-	-	-	-	-	-	-
			-	-	-	-	-	-	-	-

Lam-Bor: *L. americanus*, Boreal; Lam-Roc: *L. americanus*, Rockies; Lam-PacNW: *L. americanus,* Pacific Northwest; Missing values correspond to cases where parameters could not be reliably estimated; [1]Effective population size of population 1 (N_{e1}), 2 (N_{e2}), 3 (N_{e3}), and the ancestral populations of populations 2 and 3(N_{eA1}) and of all populations (N_{eA2}); [2]Time in years since the first split between species 1 and the ancestral of 2 and 3 (t2), and since 2 and 3 split (t1) (calibrated using a rabbit-hare divergence of 11.8 Mya; Matthee *et al.* 2004); [3]Population migration rate from population Y into population X ($2Nm_{X(Y)}$) (*significant values, $P < 0.05$; Nielsen & Wakeley 2001)

Annex II. Supplementary material from paper II in Chapter 3. Genomic perspective of introgression in hares from Iberia

List of Supplementary Figures and Tables

Figure S3.1 Relative Node Depth (RND) Power and False Discovery Rate (FDR) for inferring introgression detected with the ELAI method.

Figure S3.2 Introgression tract-size distribution.

Figure S3.3 Geographic partitioning of *L. granatensis* genetic variation using whole genome data. (A) *L. granatensis* Principal Component Analysis (PCA) analysis from genotype data.

Figure S3.4 Geographic partitioning of *L. granatensis* genetic variation including one *L. timidus* individual as outgroup, using whole genome data.

Figure S3.5 PCA summary of genetic variation in *L. granatensis* including one *L. americanus* individual as outgroup, using whole genome data.

Figure S3.6 PCA summary of genetic variation in *L. granatensis* including outgroups and excluding introgressed segments: (A) including *L. timidus* in the analysis; (B) including *L. americanus* instead

Figure S3.7 Distribution of differential levels of average introgression between the 5 northern and the 5 southern individuals across the 1000 simulations of mitochondrial introgression.

Figure S3.8 Correlation between prevalence of introgression and relative distance to centromere (Spearman's rank correlation p=0.36).

Figure S3.9 Correlation between the population recombination rate (*rho*) and distance to chromosome centre.

Figure S3.10 Functional context of introgression.

Figure S3.11 Introgression Frequency Distribution a mitonuc (lines) and background (bars) genes.

Figure S3.12 Correlation between introgression and geography.

Figure S3.13 Expected introgression frequency distribution in a sample of 10 *L. granatensis* individuals with the same geographic origin as the 10 samples used in this study, supposing introgression frequencies of these population were as previously estimated for mtDNA in large samples (Acevedo et al. 2015).

Table S3.1 Sampling localities, tissue used for genomic DNA extraction and conservation method, mitochondrial DNA lineage and raw sequencing coverage of specimens sequenced in this study.

Table S3.2 Demographic Inferences from IBS tracts.

Table S3.3 List of nuclear genes introgressed with outlier introgression frequencies (>= 85%).

Table S3.4 Gene Ontology Functional enrichment analyses of genes overlapping with outlier frequencies of introgression (>= 85%).

Table S3.5 Summary of GO functional categories significantly enriched in the set of genes with outlier introgression frequencies (at least 85%).

Table S3.6 Estimates of polymorphism and divergence patterns.

Table S3.7 List of introgressed "mitonuc" genes.

Table S3.8 Test of enrichment of mitonuc genes within the sets of i) introgressed genes and ii) genes with geographic patterns of introgression similar to the mitochondrial DNA (introgression frequency = 0% in the south and >= 20 % in the north).

Table S3.9 List of mitonuc genes with outlier frequencies of introgression (introgression frequency >= 85%).

Table S3.10 List of mitonuc genes with geographic patterns similar to the mtDNA (introgression frequency = 0% in the south and >= 20 % in the north).

Table S3.11 Nonsynonymous mutations detected within three mitonuc genes candidates to have co-introgressed with mitochondrial DNA and their potential functional impact inferred using SIFT.

Table S3.12 RND power to detect introgression at artificially introgressed mitonuc genes using the RND threshold defined at 10% FDR.

Table S3.13 RND power to detect introgression at artificially introgressed mitonuc genes using different RND thresholds based on different FDRs.

References

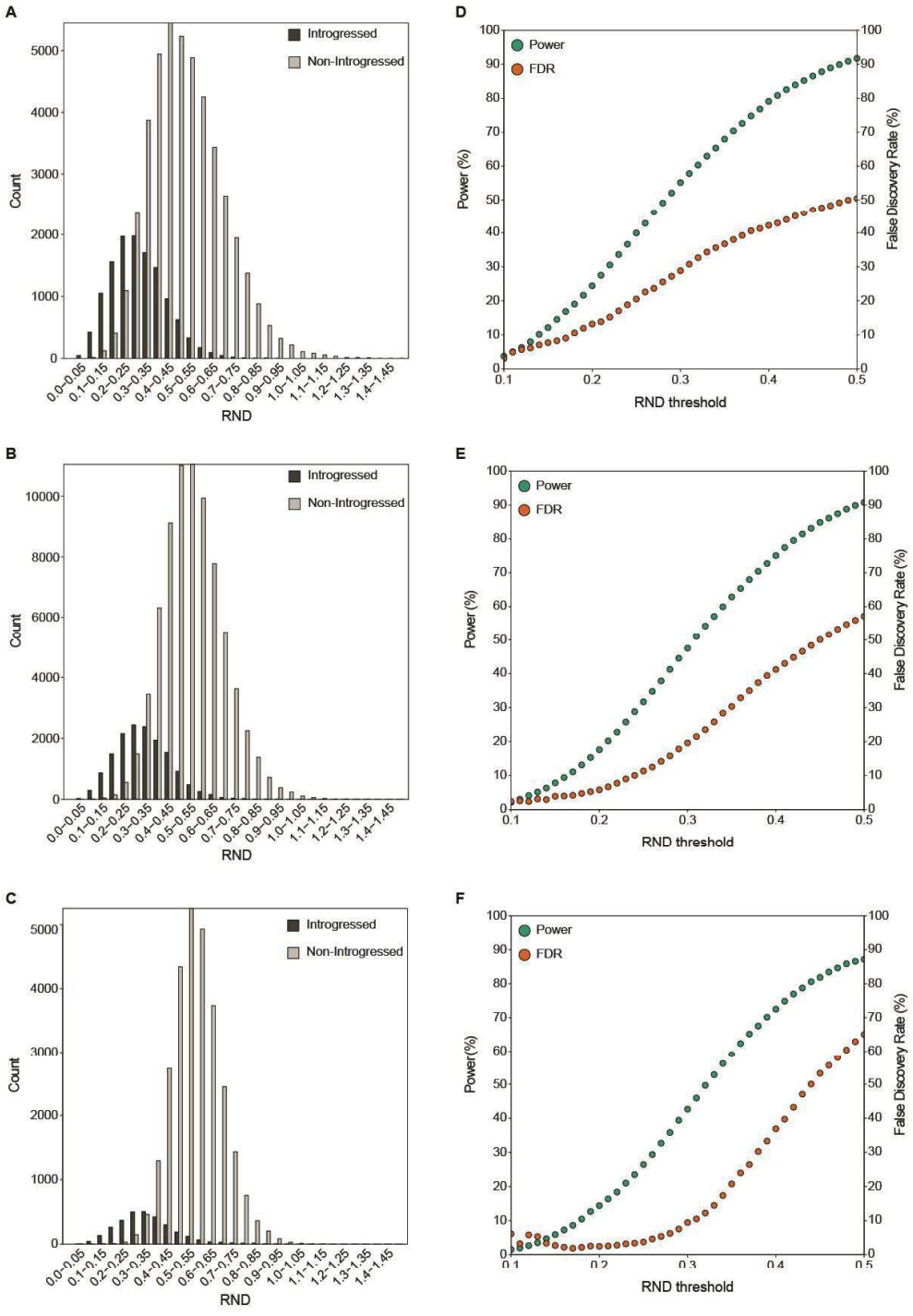


Figure S3.1 Relative Node Depth (RND) Power and False Discovery Rate (FDR) for inferring introgression detected with the ELAI method. (A-C) Distribution of RND minimum values across all individuals for windows completely within ELAI introgression fragments (black) and windows not overlapping such fragments (grey), for RND window sizes of (A) 10kb, (B) 20kb and (C) 50kb. (D-F) Estimates of Power (green) and FDR (red) assuming different RND minimum thresholds to define RND windows as introgressed, for RND window sizes of (D) 10kb, (E) 20kb and (F) 50kb.

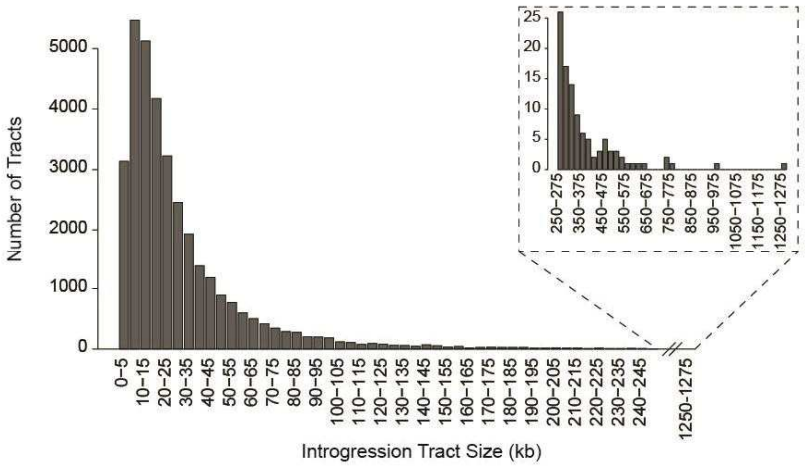


Figure S3.2 Introgression tract-size distribution. Introgression tracts inferred by ELAI across all individuals were grouped in 5 kb bins. Mean tract size is 29364 bp.

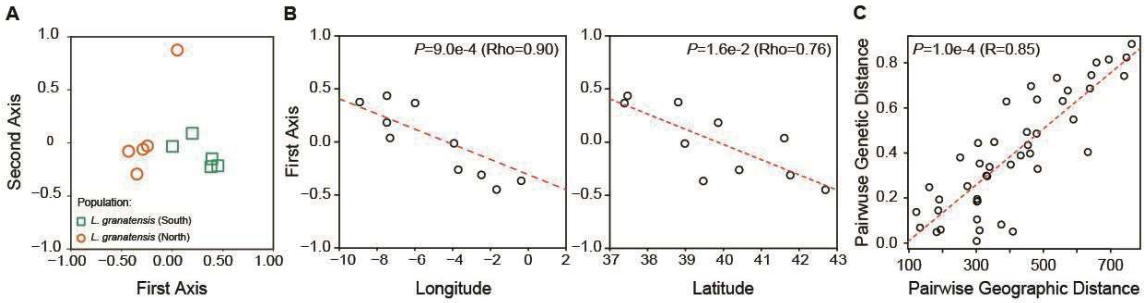


Figure S3.3 Geographic partitioning of *L. granatensis* genetic variation using whole genome data. (A) *L. granatensis* Principal Component Analysis (PCA) analysis from genotype data. Different symbols indicate the 5 southernmost and 5 northernmost samples. (B) Correlation between coordinates on the first PCA axis of genetic variation in *L. granatensis* and geographical coordinates of sample localities (Spearman's rank correlation p=0.000 and p=0.016, for correlation with Longitude and Latitude respectively; dashed line indicates a linear regression trendline). (C) Correlation between genetic differentiation (measured as the absolute difference of pairwise values of the first *L. granatensis* PCA axis) and geographic distance (measured in kilometers) among pairs of individuals (Mantel test with 9999 permutations p=0.0001; dashed line indicates a linear regression trendline).

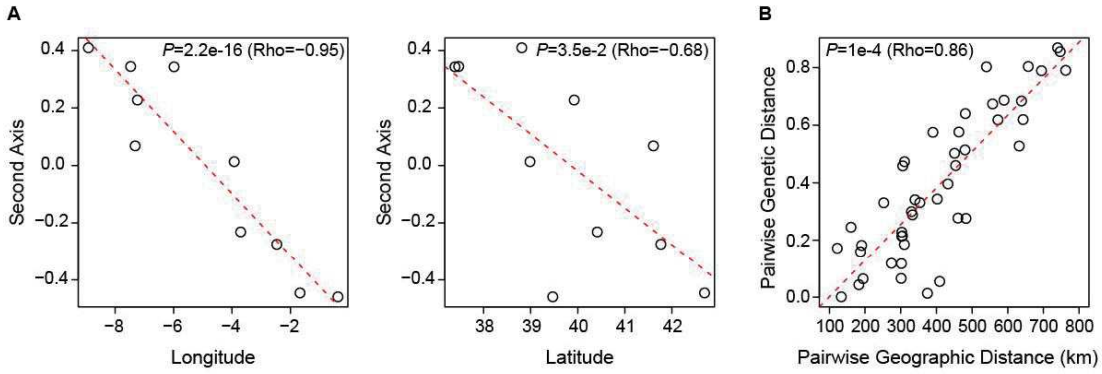


Figure S3.4 Geographic partitioning of *L. granatensis* genetic variation including one *L. timidus* individual as outgroup, using whole genome data. The results of the PCA can be seen on Fig. 2A. (A) Correlation between the second PCA axis of genetic variation and geographical coordinates of sample localities (Spearman's rank correlation p=0.000 and p=0.035, for correlation with Longitude and Latitude respectively; dashed line indicates a linear regression trendline). (B) Correlation between genetic differentiation (measured as the absolute difference of pairwise values of the second PCA axis) and geographic distance (measured in kilometers) among pairs of individuals (Mantel test with 9999 permutations p=0.0001; dashed line indicates a linear regression trendline).

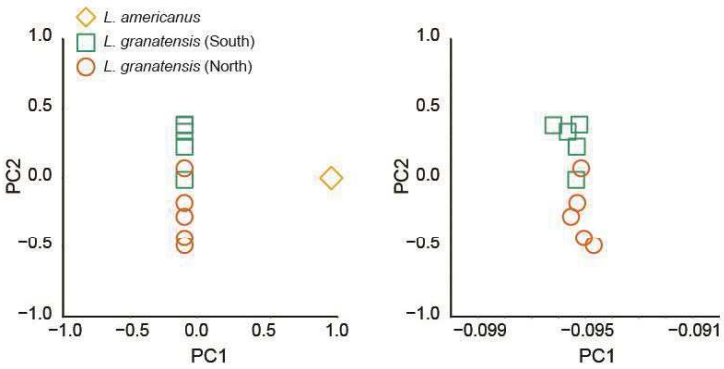


Figure S3.5 PCA summary of genetic variation in *L. granatensis* including one *L. americanus* individual as outgroup, using whole genome data. Zoom in *L. granatensis* (right part of the graph) shows lack of differentiation between the 5 southernmost and the 5 northernmost samples along the axis of interspecific differentiation, contrary to what was found when using *L. timidus* as outgroup.

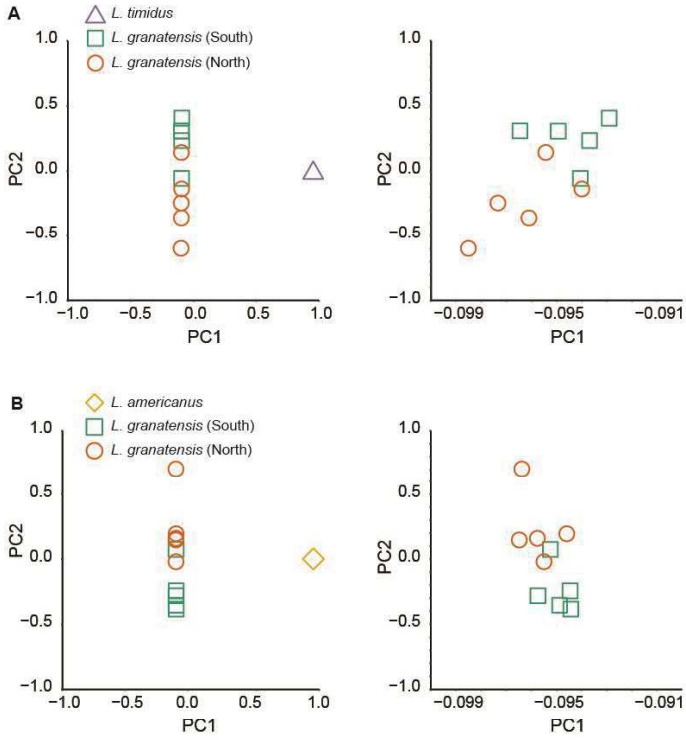


Figure S3.6 PCA summary of genetic variation in *L. granatensis* including outgroups and excluding introgressed segments: (A) including *L. timidus* in the analysis; (B) including *L. americanus* instead. Plots on the left show all samples, those on the right a zoom in *L. granatensis* samples only.

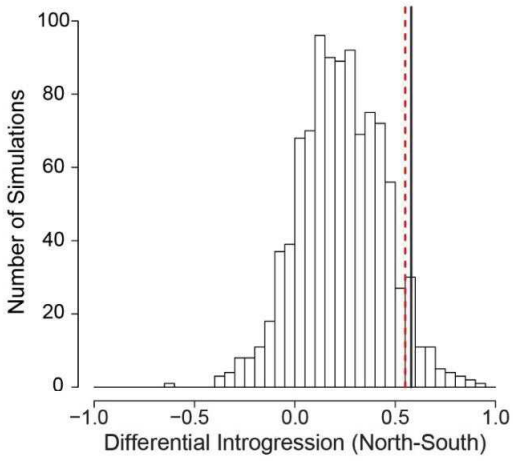


Figure S3.7 Distribution of differential levels of average introgression between the 5 northern and the 5 southern individuals across the 1000 simulations of mitochondrial introgression. The vertical red dashed line indicates the empirical difference while the solid black line represents the 95% percentile value of the simulated distribution.

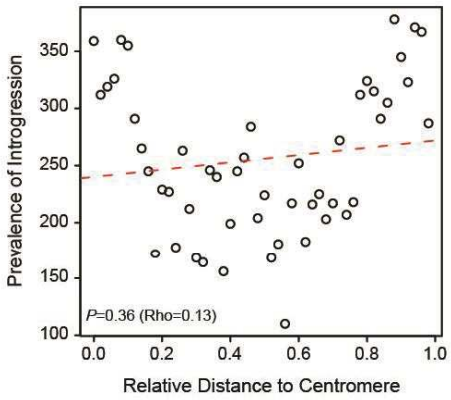


Figure S3.8 Correlation between prevalence of introgression and relative distance to centromere (Spearman's rank correlation p=0.36). Dashed lines indicates a linear regression trendline. Only a subset of SNPs, at least 50kb apart from each other to avoid dependence, was considered.

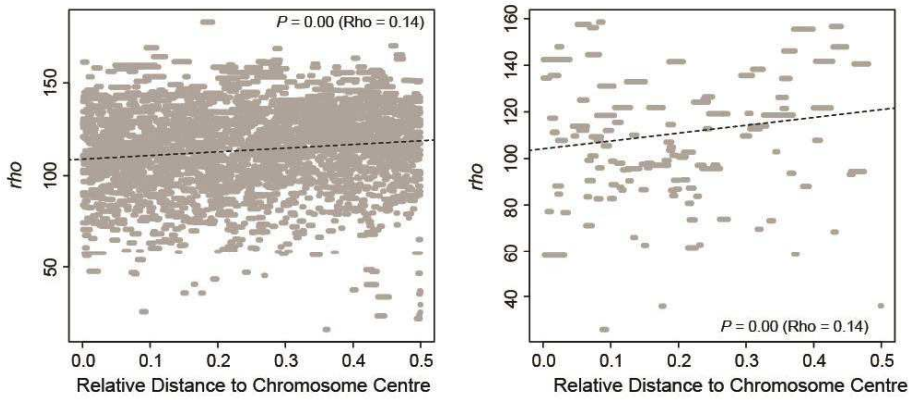


Figure S3.9 Correlation between the population recombination rate (rho) and distance to chromosome centre. Only a subset of SNPs, at least 50kb apart from each other to avoid dependence, was considered. In (A) all SNPs within the defined subset were considered independently of whether lying within or outside an ELAI introgression segment. In (B) SNPs lying within ELAI segments of introgression were further discarded from the subset of independent SNPs in order to remove the effects of introgression. Dashed lines indicate linear regression trendlines.

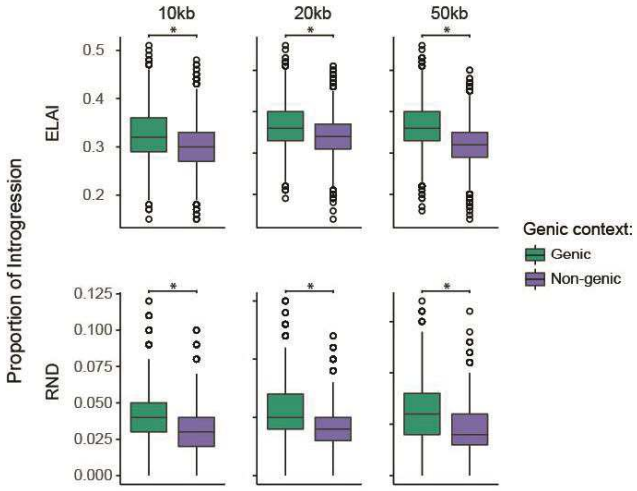


Figure S3.10 Functional context of introgression. The distributions of the proportions of introgressed windows in genic (green) and non-genic regions (blue) across 10000 replicates (y-axis) for the different methods are presented. For each of the window sizes used (10kb, 20kb and 50kb) the genic or non-genic state was defined if overlapping or not a protein-coding coding gene. For RND, windows were defined as introgressed if at least one haplotype was found as introgressed. For ELAI, the same windows as defined for RND were used, a window being considered as introgressed if overlapping an ELAI introgressed segment. Statistical differences between distributions were assessed using a Wilcoxon rank sum test. Significant differences (P << 0.01) are indicated by an asterisk.

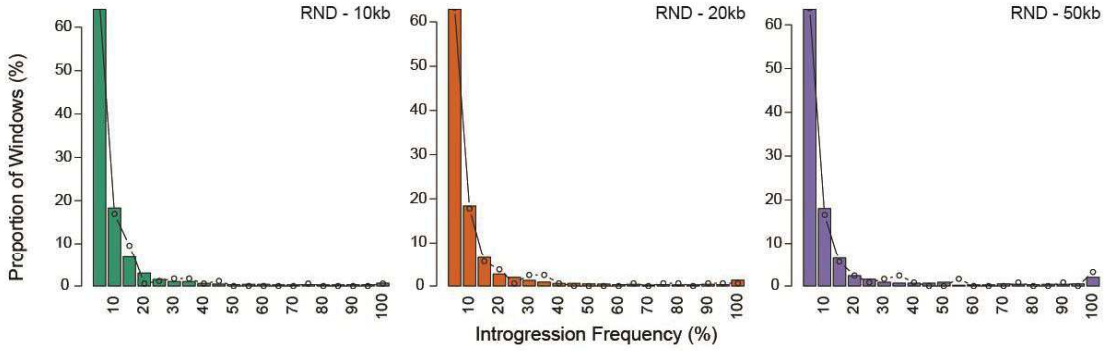


Figure S3.11 Introgression Frequency Distribution a mitonuc (lines) and background (bars) genes. The frequency of windows (y-axis) at each introgression frequency (x-axis) was estimated only considering windows with at least one haplotype introgressed (introgression frequency >= 5%).

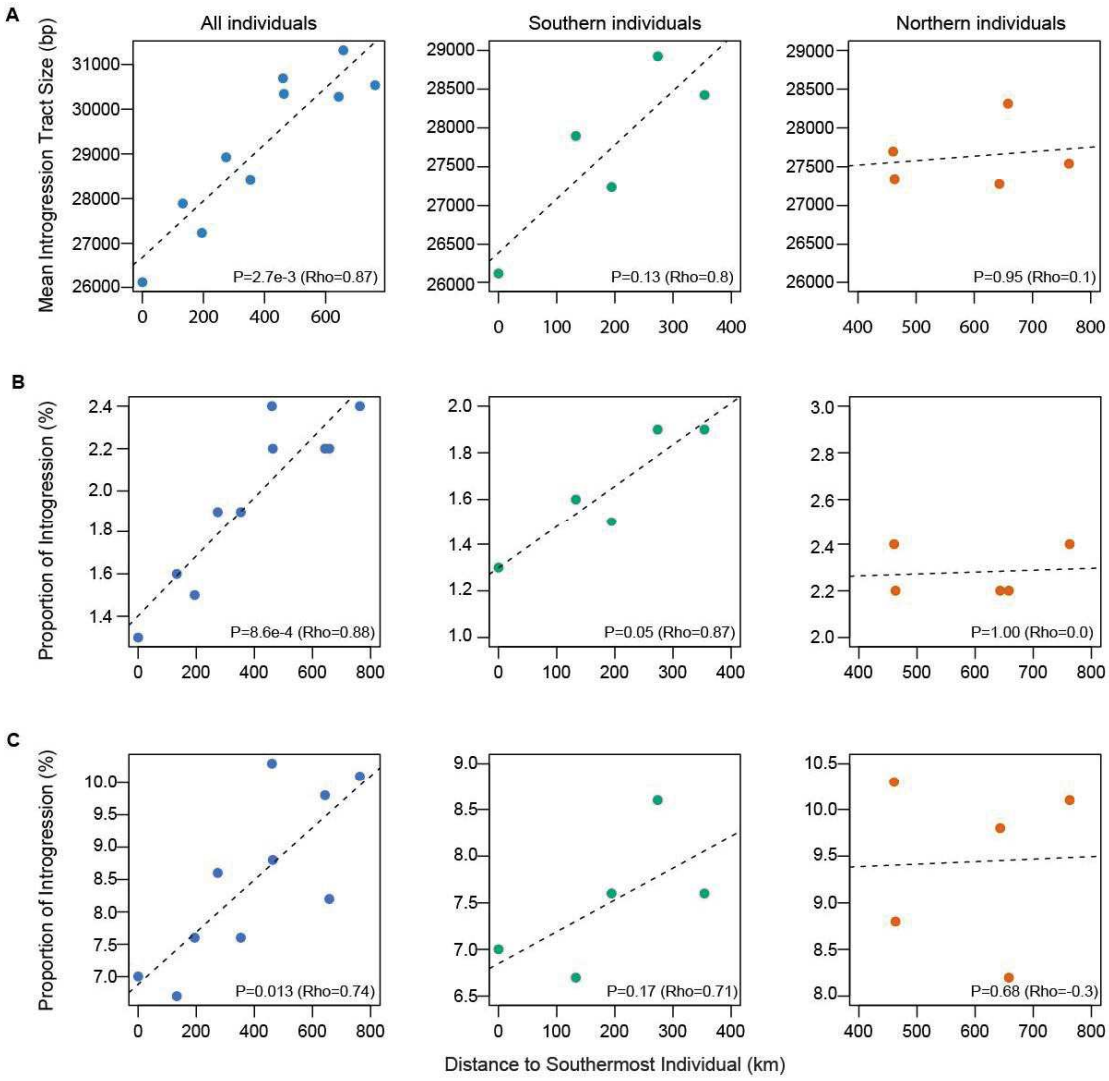


Figure S3.12 Correlation between introgression and geography. For each of the 10 samples, distance to the southernmost sample (x axis) is plotted against different characteristics of introgression. (A) mean introgression tract size, (B) observed proportion of the genome introgressed and (C) simulated proportion of genome introgressed for simulation parameter set par2. In the left panels, all samples are considered, and in the two rightmost panels, the 5 southern (centre panels) and 5 northern samples (right panels) are considered separately. Correlations were tested with Spearman's rank correlation test. Dashed lines represent linear regression trendlines.

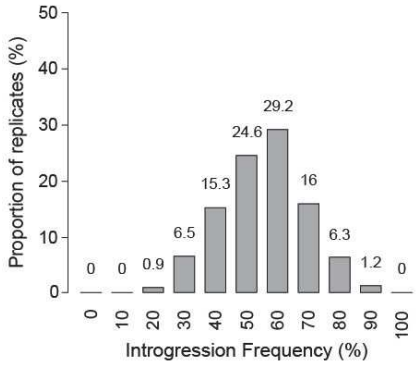


Figure S3.13 Expected introgression frequency distribution in a sample of 10 *L. granatensis* individuals with the same geographic origin as the 10 samples used in this study, supposing introgression frequencies of these population were as previously estimated for mtDNA in large samples (Acevedo et al. 2015). We simulated sampling of two haplotypes per population, with a probability of being introgressed equal to the empirical mtDNA introgression frequency of the population, and calculated introgression frequency over the 10 populations. The final distribution was built from a sample of 10´000 replicates.

Table S3.1 Sampling localities, tissue used for genomic DNA extraction and conservation method, mitochondrial DNA lineage and raw sequencing coverage of specimens sequenced in this study.

Species	Individual code	Population Code	Locality	Latitude	Longitude	Tissue	Conservation	Year	mtDNA haplotype[a]	Raw Coverage (X)	Accession Number
Lepus granatensis											
	LGR.3085	ALT	Alcoutim, Portugal	37.469978	-7.473078	Ear	Ethanol	2012	nat	26.9	submitted
	LGR.1163	SEV	Seville, Spain	37.389092	-5.984459	Ear	Ethanol	2012	nat	25.5	submitted
	LGR.2018	PAN	Pancas, Portugal	38.809101	-8.918929	Kidney	RNAlater	2009	nat	22.8	submitted
	LGR.147	CBR	Castelo Branco, Portugal	39.924751	-7.241590	Organ	Frozen	-	nat	26.2	submitted
	LGR.2553	CRE	Ciudad Real, Spain	38.984829	-3.927378	Kidney	RNAlater	2011	nat	25.6	submitted
	LGR.2786	VLP	Valpaços, Portugal	41.608715	-7.310906	Kidney	RNAlater	2012	iC	27.6	submitted
	LGR.1028	MAD	Madrid, Spain	40.416775	-3.70379	Ear	Ethanol	2002	iA	28.7	submitted
	LGR.1294	VAL	Valencia, Spain	39.469910	-0.376288	Ear	Ethanol	2002	iB	23.2	submitted
	LGR.1184	SOR	Soria, Spain	41.764431	-2.463772	Ear	Ethanol	2003	iB	23.3	submitted
	LGR.2544	NAV	Navarra, Spain	42.695393	-1.676069	Kidney	RNAlater	2001	iA	27.7	submitted
Lepus timidus											
	LTM.2012	SCA	Scandinavia	-	-	Kidney	RNAlater	2009	-	23.2	submitted
	LTM.3121	ALP	Switzerland, Alps	46.841560	9.594860	Kidney	RNAlater	2012	-	25.1	submitted
	LTM.3109	ALP	France, Alps	46.043150	6.579070	Ear	Ethanol	2012	-	28.5	submitted
Lepus americanus											
	LAM.2013	MON	Montana, USA	47.040180	-113.554680	Ovarian	RNAlater	2009	-	27.6 (37.4[b])	submitted SRX265626

[a]mtDNA lineages were inferred based on a fragment of the D-loop control region and follow the notation of Melo-Ferreira et al. (2011);

[b]Including data from Carneiro et al. (2014).

Table S3.2 Demographic Inferences from IBS tracts. Four demographic models were tested, all based on a simple model of divergence from an ancestral population. Other demographic events included: 1P - one admixture pulse from *L. timidus* into *L. granatensis*; 1PC - one admixture pulse with size change in both populations at time of admixture; 4P - four admixture pulses; 4PC - four admixture pulses with size change in both populations at time of the last admixture (in coalescent time).

Model parameters	IBS tracts > 300 bp				IBS tracts > 10 kb			
	1P	**1PC**	**4P**	**4PC**	**1P**	**1PC**	**4P**	**4PC**
Negative log likelihood	14,871,204	14,838,966	5,581,353	5,336,008	55,528	55,379	55,434	55,276
Divergence time (kya)	463.6	464.0	451.1	479.1	518.0	532.2	531.1	743.4
Time of ancient bottleneck	-	-	-	-	-	-	-	-
Time of most ancient gene flow (kya)	-	-	51.1	79.1	-	-	21.6	24.6
Time of most recent gene flow (kya)	63.6	64.0	50.8	58.9	2.0	4.9	21.3	24.3
L. granatensis population size (ancient)	-	43559	-	44236	-	18789	-	10000
L. granatensis population size (current)	40080	39460	35185	35735	37881	256872	37950	38007
L. timidus population size (ancient)	-	43691	-	50000	-	66744	-	10000
L. timidus population size (current)	38222	37888	56020	22829	45496	37762	45283	45743
Ancestral population size	10000	10000	10000	10000	373491	653271	737827	416232

Table S3.3 List of nuclear genes introgressed with outlier introgression frequencies (>= 85%).

Introgression Frequency	Ensembl Gene ID	Associated Gene Name	Description
20	ENSOCUG00000000064	DNAJB6	DnaJ heat shock protein family (Hsp40) member B6
20	ENSOCUG00000000323	NUP188	nucleoporin 188
20	ENSOCUG00000001270	VAMP7	vesicle associated membrane protein 7
20	ENSOCUG00000001402	FNTA	-
20	ENSOCUG00000002022	RAB3GAP1	RAB3 GTPase activating protein catalytic subunit 1
20	ENSOCUG00000003708	-	-
20	ENSOCUG00000004871	RARS2	arginyl-tRNA synthetase 2, mitochondrial
20	ENSOCUG00000005079	-	-
20	ENSOCUG00000005804	RBL1	RB transcriptional corepressor like 1
20	ENSOCUG00000005948	DMXL2	Dmx like 2
20	ENSOCUG00000005958	ATRNL1	attractin like 1
20	ENSOCUG00000006052	SOS2	SOS Ras/Rho guanine nucleotide exchange factor 2
20	ENSOCUG00000006224	ALMS1	ALMS1, centrosome and basal body associated protein
20	ENSOCUG00000006605	ARID4B	AT-rich interaction domain 4B
20	ENSOCUG00000007181	-	-
20	ENSOCUG00000008572	RBPMS	RNA binding protein with multiple splicing
20	ENSOCUG00000008669	TMLHE	trimethyllysine hydroxylase, epsilon
20	ENSOCUG00000008704	SPATA6	spermatogenesis associated 6
20	ENSOCUG00000009937	SLC9C1	solute carrier family 9 member C1
20	ENSOCUG00000010045	ZNF638	zinc finger protein 638
20	ENSOCUG00000011118	-	-
20	ENSOCUG00000011569	GLRA2	glycine receptor alpha 2
20	ENSOCUG00000011572	-	-
20	ENSOCUG00000011774	FBXL5	F-box and leucine rich repeat protein 5
20	ENSOCUG00000011808	STARD13	StAR related lipid transfer domain containing 13
20	ENSOCUG00000012104	ADAM7	ADAM metallopeptidase domain 7
20	ENSOCUG00000012315	RNF168	ring finger protein 168
20	ENSOCUG00000013358	-	-
20	ENSOCUG00000013380	ADAM9	ADAM metallopeptidase domain 9
20	ENSOCUG00000013397	-	-
20	ENSOCUG00000013788	-	-
20	ENSOCUG00000014811	KIAA1109	KIAA1109
20	ENSOCUG00000014825	-	-
20	ENSOCUG00000015577	C3orf20	chromosome 3 open reading frame 20
20	ENSOCUG00000015808	GMCL1	germ cell-less, spermatogenesis associated 1
20	ENSOCUG00000016322	IL12B	interleukin 12B
20	ENSOCUG00000016382	ZNF106	zinc finger protein 106
20	ENSOCUG00000016532	ATP13A3	ATPase 13A3
20	ENSOCUG00000016756	NEK1	NIMA related kinase 1
20	ENSOCUG00000017435	RIF1	replication timing regulatory factor 1
20	ENSOCUG00000017527	LRRIQ1	leucine rich repeats and IQ motif containing 1
20	ENSOCUG00000017627	UGGT2	UDP-glucose glycoprotein glucosyltransferase 2
20	ENSOCUG00000017859	RBM46	RNA binding motif protein 46

20	ENSOCUG00000018195	-	-
20	ENSOCUG00000019104	-	-
20	ENSOCUG00000020014	-	-
20	ENSOCUG00000020052	-	-
20	ENSOCUG00000020561	-	-
20	ENSOCUG00000021795	-	-
20	ENSOCUG00000021956	-	-
20	ENSOCUG00000022218	SPRY3	sprouty RTK signaling antagonist 3
20	ENSOCUG00000022616	-	-
20	ENSOCUG00000024224	-	-
20	ENSOCUG00000024280	TMEM5	transmembrane protein 5
20	ENSOCUG00000024354	-	-
20	ENSOCUG00000025062	SDHAF4	succinate dehydrogenase complex assembly factor 4
20	ENSOCUG00000025461	-	-
20	ENSOCUG00000027756	-	-
20	ENSOCUG00000027897	-	-
20	ENSOCUG00000027911	-	-
20	ENSOCUG00000028219	FASLG	Fas ligand
20	ENSOCUG00000028725	-	-
20	ENSOCUG00000028980	-	-
20	ENSOCUG00000029249	LOC100340453	-
20	ENSOCUG00000029319	-	-
20	ENSOCUG00000029499	CWC27	CWC27 spliceosome associated protein homolog
20	ENSOCUG00000029585	ZNF584	zinc finger protein 584
19	ENSOCUG00000000397	KALRN	-
19	ENSOCUG00000002212	TIPIN	TIMELESS interacting protein
19	ENSOCUG00000007258	STRN	striatin
19	ENSOCUG00000007754	CAMSAP2	calmodulin regulated spectrin associated protein family member 2
19	ENSOCUG00000008782	MAP2K1	dual specificity mitogen-activated protein kinase kinase 1
19	ENSOCUG00000010614	ATAD2	ATPase family, AAA domain containing 2
19	ENSOCUG00000012757	NUP205	nucleoporin 205
19	ENSOCUG00000013917	C4H12orf55	-
19	ENSOCUG00000015796	ANXA4	annexin A4
19	ENSOCUG00000015849	-	-
19	ENSOCUG00000016222	TXLNB	taxilin beta
19	ENSOCUG00000017001	KANSL1L	KAT8 regulatory NSL complex subunit 1 like
19	ENSOCUG00000017266	FAM168B	family with sequence similarity 168 member B
19	ENSOCUG00000023033	-	-
19	ENSOCUG00000025652	LOC100355582	-
19	ENSOCUG00000026715	ERLIN1	ER lipid raft associated 1
18	ENSOCUG00000000334	TOX4	TOX high mobility group box family member 4
18	ENSOCUG00000000336	METTL3	methyltransferase like 3
18	ENSOCUG00000001397	LMF2	lipase maturation factor 2
18	ENSOCUG00000001401	NCAPH2	non-SMC condensin II complex subunit H2
18	ENSOCUG00000001408	TYMP	thymidine phosphorylase

18	ENSOCUG00000002156	PATL1	PAT1 homolog 1, processing body mRNA decay factor
18	ENSOCUG00000004218	CD14	CD14 molecule
18	ENSOCUG00000004236	TMCO6	transmembrane and coiled-coil domains 6
18	ENSOCUG00000004657	NR2C1	nuclear receptor subfamily 2 group C member 1
18	ENSOCUG00000006543	FAM169A	family with sequence similarity 169 member A
18	ENSOCUG00000006794	OPTN	optineurin
18	ENSOCUG00000007815	LOC100337847	-
18	ENSOCUG00000011422	L2HGDH	L-2-hydroxyglutarate dehydrogenase
18	ENSOCUG00000012252	DPP6	dipeptidyl peptidase like 6
18	ENSOCUG00000012915	SIK2	salt inducible kinase 2
18	ENSOCUG00000021766	-	-
18	ENSOCUG00000023678	-	-
18	ENSOCUG00000024805	COL28A1	collagen type XXVIII alpha 1 chain
18	ENSOCUG00000025155	-	-
18	ENSOCUG00000026117	-	-
18	ENSOCUG00000026317	-	-
18	ENSOCUG00000026983	-	-
18	ENSOCUG00000027464	SALL2	spalt like transcription factor 2
18	ENSOCUG00000027602	-	-
18	ENSOCUG00000028610	-	-
18	ENSOCUG00000029177	LOC100338099	-
18	ENSOCUG00000029651	LOC100351290	-
17	ENSOCUG00000001172	CCDC138	coiled-coil domain containing 138
17	ENSOCUG00000002694	KPNA1	karyopherin subunit alpha 1
17	ENSOCUG00000003372	ART3	ADP-ribosyltransferase 3
17	ENSOCUG00000004519	ERCC8	ERCC excision repair 8, CSA ubiquitin ligase complex subunit
17	ENSOCUG00000006203	DUSP27	dual specificity phosphatase 27 (putative)
17	ENSOCUG00000008904	RNASEH2B	ribonuclease H2 subunit B
17	ENSOCUG00000011270	CD1B	T-cell surface glycoprotein CD1b precursor
17	ENSOCUG00000011880	SPIN1	-
17	ENSOCUG00000014615	MSH6	mutS homolog 6
17	ENSOCUG00000014631	FBXO11	F-box protein 11
17	ENSOCUG00000014849	CACNB2	voltage-dependent L-type calcium channel subunit beta-2
17	ENSOCUG00000025553	-	-
17	ENSOCUG00000029524	FAM227B	family with sequence similarity 227 member B

Table S3.4 Gene Ontology Functional enrichment analyses of genes overlapping with outlier frequencies of introgression (>= 85%). Evidence of RND 10kb, 20kb and 50kb window size analyses were combined. Rabbit and Mouse Ensemble IDs (only one2one orthologues were considered) were used in this analysis. Corrected p-values were obtained by using the Benjamini-Hochberg multiple testing correction algorithm.

Species	GO Term	Term Description	corrected p-value	Nb. Genes in Term	Nb. Genes in Query	Intersection Term/Query
Rabbit						
	GO:0048172	regulation of short-term neuronal synaptic plasticity	0.045	6	123	1
	GO:0061734	parkin-mediated mitophagy in response to mitochondrial depolarization	0.038	5	123	1
	GO:0016236	Macroautophagy	0.049	107	123	3
	GO:0007033	vacuole organization	0.043	101	123	3
	GO:0035196	production of miRNAs involved in gene silencing by miRNA	0.012	22	123	2
	GO:0051547	regulation of keratinocyte migration	0.045	6	123	1
	GO:0000076	DNA replication checkpoint	0.045	6	123	1
	GO:0045739	positive regulation of DNA repair	0.022	30	123	2
	GO:2000618	regulation of histone H4-K16 acetylation	0.038	5	123	1
	GO:0051095	regulation of helicase activity	0.038	5	123	1
	GO:0006283	transcription-coupled nucleotide-excision repair	0.045	6	123	1
	GO:0051608	histamine transport	0.038	5	123	1
	GO:0086027	AV node cell to bundle of His cell signaling	0.045	6	123	1
	GO:0018410	C-terminal protein amino acid modification	0.045	6	123	1
	GO:0051135	positive regulation of NK T cell activation	0.038	5	123	1
	GO:0080009	mRNA methylation	0.045	6	123	1
	GO:0071361	cellular response to ethanol	0.045	6	123	1
	GO:0050829	defense response to Gram-negative bacterium	0.014	24	123	2
	GO:0072594	establishment of protein localization to organelle	0.045	337	123	6
	GO:0043001	Golgi to plasma membrane protein transport	0.008	18	123	2
	GO:0002705	positive regulation of leukocyte mediated immunity	0.015	68	123	3
	GO:0008333	endosome to lysosome transport	0.024	31	123	2
	GO:0001956	positive regulation of neurotransmitter secretion	0.038	5	123	1

	GO:0060440	trachea formation	0.045	6	123	1
	GO:0042976	activation of Janus kinase activity	0.045	6	123	1
	GO:0097527	necroptotic signaling pathway	0.038	5	123	1
	GO:0009411	response to UV	0.003	81	123	4
	GO:0090051	negative regulation of cell migration involved in sprouting angiogenesis	0.045	6	123	1
	GO:0034454	microtubule anchoring at centrosome	0.045	6	123	1
	GO:0042267	natural killer cell mediated cytotoxicity	0.024	31	123	2
	GO:0032819	positive regulation of natural killer cell proliferation	0.045	6	123	1
	GO:0002285	lymphocyte activation involved in immune response	0.044	102	123	3
	GO:0050765	negative regulation of phagocytosis	0.038	5	123	1
	GO:0030101	natural killer cell activation	0.028	34	123	2
	GO:0034773	histone H4-K20 trimethylation	0.038	5	123	1
	GO:0009437	carnitine metabolic process	0.045	6	123	1
	GO:0000045	autophagosome assembly	0.049	46	123	2
	GO:0071500	cellular response to nitrosative stress	0.045	6	123	1
	GO:0070257	positive regulation of mucus secretion	0.038	5	123	1
	GO:0034393	positive regulation of smooth muscle cell apoptotic process	0.038	5	123	1
	GO:0032760	positive regulation of tumor necrosis factor production	0.029	35	123	2
	GO:0048840	otolith development	0.045	6	123	1
	GO:0070339	response to bacterial lipopeptide	0.038	5	123	1
	GO:0046666	retinal cell programmed cell death	0.038	5	123	1
	GO:0051957	positive regulation of amino acid transport	0.038	5	123	1
	GO:0001866	NK T cell proliferation	0.045	6	123	1
	GO:0060324	face development	0.036	39	123	2
	GO:0008380	RNA splicing	0.040	167	123	4
Mouse						
	GO:0007528	neuromuscular junction development	0.003	37	88	3
	GO:0046578	regulation of Ras protein signal transduction	0.050	177	88	4
	GO:0002449	lymphocyte mediated immunity	0.029	150	88	4
	GO:0007283	spermatogenesis	0.013	349	88	7

GO:0002705	positive regulation of leukocyte mediated immunity	0.025	77	88	3
GO:0002230	positive regulation of defense response to virus by host	0.009	16	88	2
GO:0002697	regulation of immune effector process	0.024	219	88	5
GO:0009411	response to UV	0.001	97	88	5
GO:0002204	somatic recombination of immunoglobulin genes involved in immune response	0.042	36	88	2
GO:1902580	single-organism cellular localization	0.001	744	88	13
GO:0002323	natural killer cell activation involved in immune response	0.006	13	88	2
GO:0071359	cellular response to dsRNA	0.004	41	88	3
GO:0043984	histone H4-K16 acetylation	0.011	18	88	2
GO:0050829	defense response to Gram-negative bacterium	0.034	32	88	2
GO:0045739	positive regulation of DNA repair	0.040	35	88	2
GO:2000001	regulation of DNA damage checkpoint	0.008	15	88	2
GO:0050775	positive regulation of dendrite morphogenesis	0.021	25	88	2

Table S3.5 Summary of GO functional categories significantly enriched in the set of genes with outlier introgression frequencies (at least 85%). Redundant terms were removed using REVIGO.

Species	Representative Term
Rabbit	
	positive regulation of leukocyte mediated immunity
	response to UV
	Golgi to plasma membrane protein transport
	mRNA methylation
	macroautophagy
	face development
	carnitine metabolism
Mouse	
	response to UV
	single-organism cellular localization
	neuromuscular junction development
	spermatogenesis

Table S3.6 Estimates of polymorphism and divergence patterns. †Statistical differences between introgressed and non-introgressed genes distributions were appraised using the Wilcoxon rank sum test. gra - *L. granatensis*; tim - *L. timidus*. NI- *Neutrality index*.

	All Genes			**Introgressed Genes**			**Non-Introgressed Genes**			**P-value†**		
	πS(%)	**πN(%)**	**πN/πS**	**πS(%)**	**πN(%)**	**πN/πS**	**πS(%)**	**πN(%)**	**πN/πS**	**πS**	**πN**	**πN/πS**
gra												
mean	0.44	0.08	0.26	0.42	0.07	0.23	0.45	0.08	0.26	0.496	0.671	0.811
sd	0.41	0.14	0.91	0.36	0.11	0.41	0.42	0.15	0.97			
tim												
mean	0.66	0.12	0.20	0.53	0.08	0.18	0.69	0.12	0.20	2.2E-16	4.6E-07	0.171
sd	0.67	0.24	0.30	0.46	0.16	0.27	0.70	0.25	0.31			
	dS	**dN**	**dN/dS**	**dS(%)**	**dN(%)**	**dN/dS**	**dS(%)**	**dN(%)**	**dN/dS**	**dS**	**dN**	**dN/dS**
tim-gra												
mean	0.1837	0.0113	0.0018	0.1702	0.0100	0.0015	0.1837	0.0113	0.0018	1.5E-13	2.2E-05	0.201
sd	0.2369	0.0071	0.0024	0.2129	0.0063	0.0019	0.2369	0.0071	0.0024			
tim-gra			**NI**			**NI**			**NI**			**NI**
mean			0.99			0.90			1.01			0.609
sd			2.05			1.60			2.13			

Table S3.7 List of introgressed "mitonuc" genes.

Mitonuc Category	Gene Ensembl ID	Gene Name	Gene Description
OXPHOS	ENSOCUG00000006359	NDUFS3	NADH:ubiquinone oxidoreductase core subunit S3
OXPHOS	ENSOCUG00000007809	-	-
OXPHOS	ENSOCUG00000008113	UQCRB	ubiquinol-cytochrome c reductase binding protein
OXPHOS	ENSOCUG00000009899	NDUFV2	NADH:ubiquinone oxidoreductase core subunit V2
OXPHOS	ENSOCUG00000013818	COX7A2	cytochrome c oxidase subunit 7A2
OXPHOS	ENSOCUG00000014894	-	-
OXPHOS	ENSOCUG00000022925	-	-
OXPHOS	ENSOCUG00000025074	NDUFB9	NADH:ubiquinone oxidoreductase subunit B9
mitonuc-direct	ENSOCUG00000000867	MRPS28	mitochondrial ribosomal protein S28
mitonuc-direct	ENSOCUG00000001119	MRPS22	mitochondrial ribosomal protein S22
mitonuc-direct	ENSOCUG00000001383	MRPL9	mitochondrial ribosomal protein L9
mitonuc-direct	ENSOCUG00000001964	LARS2	leucyl-tRNA synthetase 2, mitochondrial
mitonuc-direct	ENSOCUG00000002951	MRPL3	mitochondrial ribosomal protein L3
mitonuc-direct	ENSOCUG00000003348	ATP5S	ATP synthase, H+ transporting, mitochondrial Fo complex subunit s (factor B)
mitonuc-direct	ENSOCUG00000004871	RARS2	arginyl-tRNA synthetase 2, mitochondrial
mitonuc-direct	ENSOCUG00000007961	MRPS27	mitochondrial ribosomal protein S27
mitonuc-direct	ENSOCUG00000013542	MRPL44	mitochondrial ribosomal protein L44
mitonuc-direct	ENSOCUG00000016189	MRPL13	mitochondrial ribosomal protein L13
mitonuc-direct	ENSOCUG00000016214	MRPS35	mitochondrial ribosomal protein S35
mitonuc-direct	ENSOCUG00000017469	NARS2	asparaginyl-tRNA synthetase 2, mitochondrial (putative)
mitonuc-direct	ENSOCUG00000021496	-	-
mitonuc-direct	ENSOCUG00000023231	-	-
mitonuc-direct	ENSOCUG00000025086	-	-
mitonuc-direct	ENSOCUG00000025605	-	-
mitonuc	ENSOCUG00000000482	NUDT6	nudix hydrolase 6
mitonuc	ENSOCUG00000000525	-	-

mitonuc	ENSOCUG00000000584	ARG2	arginase 2
mitonuc	ENSOCUG00000000749	HTATIP2	HIV-1 Tat interactive protein 2
mitonuc	ENSOCUG00000001086	RMDN1	regulator of microtubule dynamics 1
mitonuc	ENSOCUG00000001121	ATIC	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase
mitonuc	ENSOCUG00000001387	TDRKH	tudor and KH domain containing
mitonuc	ENSOCUG00000001408	TYMP	thymidine phosphorylase
mitonuc	ENSOCUG00000001434	SPTLC2	serine palmitoyltransferase long chain base subunit 2
mitonuc	ENSOCUG00000001496	STARD7	StAR related lipid transfer domain containing 7
mitonuc	ENSOCUG00000001620	SND1	staphylococcal nuclease and tudor domain containing 1
mitonuc	ENSOCUG00000001622	-	-
mitonuc	ENSOCUG00000001696	DNAJC15	DnaJ heat shock protein family (Hsp40) member C15
mitonuc	ENSOCUG00000001972	TXNDC12	thioredoxin domain containing 12
mitonuc	ENSOCUG00000002085	-	-
mitonuc	ENSOCUG00000002179	STOML1	stomatin like 1
mitonuc	ENSOCUG00000002380	NIT1	nitrilase 1
mitonuc	ENSOCUG00000002403	PPOX	protoporphyrinogen oxidase
mitonuc	ENSOCUG00000002594	-	ADP/ATP translocase 1
mitonuc	ENSOCUG00000002683	TUFM	Tu translation elongation factor, mitochondrial
mitonuc	ENSOCUG00000002690	ACSF2	acyl-CoA synthetase family member 2
mitonuc	ENSOCUG00000003029	PEX11B	peroxisomal biogenesis factor 11 beta
mitonuc	ENSOCUG00000003191	SUGCT	succinyl-CoA:glutarate-CoA transferase
mitonuc	ENSOCUG00000003666	DGUOK	deoxyguanosine kinase
mitonuc	ENSOCUG00000003830	-	-
mitonuc	ENSOCUG00000003862	DUT	deoxyuridine triphosphatase
mitonuc	ENSOCUG00000003871	GOT2	glutamic-oxaloacetic transaminase 2
mitonuc	ENSOCUG00000003883	BRCA1	BRCA1, DNA repair associated
mitonuc	ENSOCUG00000004145	ATP7B	ATPase copper transporting beta
mitonuc	ENSOCUG00000004217	STX17	syntaxin 17
mitonuc	ENSOCUG00000004274	REXO2	RNA exonuclease 2

mitonuc	ENSOCUG00000004276	MTERF3	mitochondrial transcription termination factor 3
mitonuc	ENSOCUG00000004330	NRDC	nardilysin convertase
mitonuc	ENSOCUG00000004525	NDUFAF2	NADH:ubiquinone oxidoreductase complex assembly factor 2
mitonuc	ENSOCUG00000005249	DDAH1	dimethylarginine dimethylaminohydrolase 1
mitonuc	ENSOCUG00000005330	ACADSB	acyl-CoA dehydrogenase, short/branched chain
mitonuc	ENSOCUG00000005347	PIF1	PIF1 5'-to-3' DNA helicase
mitonuc	ENSOCUG00000005403	SLC30A6	solute carrier family 30 member 6
mitonuc	ENSOCUG00000005862	AFG1L	AFG1 like ATPase
mitonuc	ENSOCUG00000005913	CHCHD7	coiled-coil-helix-coiled-coil-helix domain containing 7
mitonuc	ENSOCUG00000005938	-	glutathione peroxidase 1
mitonuc	ENSOCUG00000006446	CBR4	carbonyl reductase 4
mitonuc	ENSOCUG00000006799	COQ6	coenzyme Q6, monooxygenase
mitonuc	ENSOCUG00000006904	DCAKD	dephospho-CoA kinase domain containing
mitonuc	ENSOCUG00000006962	ALDH1L2	aldehyde dehydrogenase 1 family member L2
mitonuc	ENSOCUG00000007245	SIRT5	sirtuin 5
mitonuc	ENSOCUG00000007261	-	-
mitonuc	ENSOCUG00000007293	CLPX	caseinolytic mitochondrial matrix peptidase chaperone subunit
mitonuc	ENSOCUG00000007430	AUH	AU RNA binding methylglutaconyl-CoA hydratase
mitonuc	ENSOCUG00000007708	GADD45GIP1	GADD45G interacting protein 1
mitonuc	ENSOCUG00000007842	KYAT3	kynurenine aminotransferase 3
mitonuc	ENSOCUG00000007892	FBXL4	F-box and leucine rich repeat protein 4
mitonuc	ENSOCUG00000007893	GPAM	glycerol-3-phosphate acyltransferase, mitochondrial
mitonuc	ENSOCUG00000008004	PDHX	pyruvate dehydrogenase complex component X
mitonuc	ENSOCUG00000008144	CHCHD6	coiled-coil-helix-coiled-coil-helix domain containing 6
mitonuc	ENSOCUG00000008162	GUCY2C	guanylate cyclase 2C
mitonuc	ENSOCUG00000008364	PTPN4	protein tyrosine phosphatase, non-receptor type 4
mitonuc	ENSOCUG00000008393	OSBPL1A	oxysterol binding protein like 1A
mitonuc	ENSOCUG00000008468	MTO1	mitochondrial tRNA translation optimization 1
mitonuc	ENSOCUG00000008669	TMLHE	trimethyllysine hydroxylase, epsilon
mitonuc	ENSOCUG00000008867	FASTKD5	FAST kinase domains 5

mitonuc	ENSOCUG00000008974	-	-
mitonuc	ENSOCUG00000009058	-	-
mitonuc	ENSOCUG00000009096	LIAS	lipoic acid synthetase
mitonuc	ENSOCUG00000009162	NADK2	NAD kinase 2, mitochondrial
mitonuc	ENSOCUG00000009296	DHTKD1	dehydrogenase E1 and transketolase domain containing 1
mitonuc	ENSOCUG00000009328	CRLS1	cardiolipin synthase 1
mitonuc	ENSOCUG00000009420	NLN	neurolysin
mitonuc	ENSOCUG00000009541	NFU1	NFU1 iron-sulfur cluster scaffold
mitonuc	ENSOCUG00000009630	SLC25A17	solute carrier family 25 member 17
mitonuc	ENSOCUG00000009642	XPNPEP3	X-prolyl aminopeptidase 3
mitonuc	ENSOCUG00000009815	PNPLA8	patatin like phospholipase domain containing 8
mitonuc	ENSOCUG00000009953	SLC25A13	solute carrier family 25 member 13
mitonuc	ENSOCUG00000010029	GRSF1	G-rich RNA sequence binding factor 1
mitonuc	ENSOCUG00000010544	CROT	carnitine O-octanoyltransferase
mitonuc	ENSOCUG00000010814	ME1	malic enzyme 1
mitonuc	ENSOCUG00000011052	PITRM1	pitrilysin metallopeptidase 1
mitonuc	ENSOCUG00000011422	L2HGDH	L-2-hydroxyglutarate dehydrogenase
mitonuc	ENSOCUG00000011557	-	-
mitonuc	ENSOCUG00000011677	NSUN3	NOP2/Sun RNA methyltransferase family member 3
mitonuc	ENSOCUG00000011782	MIPEP	mitochondrial intermediate peptidase
mitonuc	ENSOCUG00000011936	FHIT	fragile histidine triad
mitonuc	ENSOCUG00000012174	NRF1	nuclear respiratory factor 1
mitonuc	ENSOCUG00000012197	EHHADH	enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase
mitonuc	ENSOCUG00000012376	SUCLG2	succinate-CoA ligase GDP-forming beta subunit
mitonuc	ENSOCUG00000012393	MCCC1	methylcrotonoyl-CoA carboxylase 1
mitonuc	ENSOCUG00000012433	ME2	malic enzyme 2
mitonuc	ENSOCUG00000012449	HCLS1	hematopoietic cell-specific Lyn substrate 1
mitonuc	ENSOCUG00000012747	PRKAR2B	protein kinase cAMP-dependent type II regulatory subunit beta
mitonuc	ENSOCUG00000012793	CMC1	C-X9-C motif containing 1
mitonuc	ENSOCUG00000013212	ACLY	ATP citrate lyase

mitonuc	ENSOCUG00000013310	AMT	aminomethyltransferase
mitonuc	ENSOCUG00000013358	ATG5	autophagy related 5
mitonuc	ENSOCUG00000013720	CLYBL	citrate lyase beta like
mitonuc	ENSOCUG00000013727	-	-
mitonuc	ENSOCUG00000014248	CYB5R3	cytochrome b5 reductase 3
mitonuc	ENSOCUG00000014405	TFAM	transcription factor A, mitochondrial
mitonuc	ENSOCUG00000014453	ACACA	acetyl-CoA carboxylase alpha
mitonuc	ENSOCUG00000014871	COQ5	coenzyme Q5, methyltransferase
mitonuc	ENSOCUG00000014909	VWA8	von Willebrand factor A domain containing 8
mitonuc	ENSOCUG00000015008	SLC25A40	solute carrier family 25 member 40
mitonuc	ENSOCUG00000015071	ALDH1L1	aldehyde dehydrogenase 1 family member L1
mitonuc	ENSOCUG00000015262	TEFM	transcription elongation factor, mitochondrial
mitonuc	ENSOCUG00000015344	HEBP1	heme binding protein 1
mitonuc	ENSOCUG00000015621	GPT2	glutamic--pyruvic transaminase 2
mitonuc	ENSOCUG00000015832	OPA1	OPA1, mitochondrial dynamin like GTPase
mitonuc	ENSOCUG00000015868	OXNAD1	oxidoreductase NAD binding domain containing 1
mitonuc	ENSOCUG00000015870	CEP89	centrosomal protein 89
mitonuc	ENSOCUG00000016168	ABCD2	ATP binding cassette subfamily D member 2
mitonuc	ENSOCUG00000016359	ALKBH3	alkB homolog 3, alpha-ketoglutaratedependent dioxygenase
mitonuc	ENSOCUG00000016368	-	-
mitonuc	ENSOCUG00000016540	ACAA2	acetyl-CoA acyltransferase 2
mitonuc	ENSOCUG00000016663	-	-
mitonuc	ENSOCUG00000016875	GATB	glutamyl-tRNA amidotransferase subunit B
mitonuc	ENSOCUG00000017039	RPUSD4	RNA pseudouridylate synthase domain containing 4
mitonuc	ENSOCUG00000017050	-	-
mitonuc	ENSOCUG00000017062	ROMO1	reactive oxygen species modulator 1
mitonuc	ENSOCUG00000017169	-	-
mitonuc	ENSOCUG00000017356	CLPB	ClpB homolog, mitochondrial AAA ATPase chaperonin
mitonuc	ENSOCUG00000017437	ADCK1	aarF domain containing kinase 1
mitonuc	ENSOCUG00000017543	-	-

mitonuc	ENSOCUG00000017713	NUBPL	nucleotide binding protein like
mitonuc	ENSOCUG00000017828	MPG	N-methylpurine DNA glycosylase
mitonuc	ENSOCUG00000020964	OCIAD1	OCIA domain containing 1
mitonuc	ENSOCUG00000021229	-	-
mitonuc	ENSOCUG00000021709	-	-
mitonuc	ENSOCUG00000022237	PRELID3A	PRELI domain containing 3A
mitonuc	ENSOCUG00000022653	PISD	phosphatidylserine decarboxylase
mitonuc	ENSOCUG00000022723	ACAD11	acyl-CoA dehydrogenase family member 11
mitonuc	ENSOCUG00000023109	-	-
mitonuc	ENSOCUG00000024056	PPWD1	peptidylprolyl isomerase domain and WD repeat containing 1
mitonuc	ENSOCUG00000024190	-	-
mitonuc	ENSOCUG00000024318	VDAC3	voltage dependent anion channel 3
mitonuc	ENSOCUG00000024892	NLRX1	NLR family member X1
mitonuc	ENSOCUG00000025062	SDHAF4	succinate dehydrogenase complex assembly factor 4
mitonuc	ENSOCUG00000025221	FAM210A	family with sequence similarity 210 member A
mitonuc	ENSOCUG00000025844	HINT3	histidine triad nucleotide binding protein 3
mitonuc	ENSOCUG00000027099	RPL34	ribosomal protein L34
mitonuc	ENSOCUG00000027290	-	-
mitonuc	ENSOCUG00000027386	-	-
mitonuc	ENSOCUG00000029502	MICU2	mitochondrial calcium uptake 2
mitonuc	ENSOCUG00000029729	-	-

Table S3.8 Test of enrichment of mitonuc genes within the sets of i) introgressed genes and ii) genes with geographic patterns of introgression similar to the mitochondrial DNA (introgression frequency = 0% in the south and >= 20 % in the north). Two different RND thresholds were used (10% FDR and 30 %FDR). [a]At least one category does not have enough elements to correctly perform a Fisher Exact Test.

FDR	Geographic set	Mitonuc set	Introgressed Genes		All Genes		P-value
			Mitonuc Category	All	Mitonuc Category	All	
10%							
	All introgression	Mitonuc	166	3312	1178	22553	0.554
		Mitonuc-direct	23	3312	185	22553	0.385
		OXPHOS	8	3312	73	22553	0.368
	Mitochondrial-like	Mitonuc	17	274	1178	22553	0.463
		Mitonuc-direct	3	274	185	22553	0.612[a]
		OXPHOS	2	274	73	22553	0.234[a]
30%							
	All introgression	Mitonuc	460	8658	1178	22553	0.633
		Mitonuc-direct	65	8658	185	22553	0.361
		OXPHOS	25	8658	73	22553	0.466
	Mitochondrial-like	Mitonuc	32	548	1178	22553	0.512
		Mitonuc-direct	5	548	185	22553	0.809[a]
		OXPHOS	1	548	73	22553	0.556[a]

Table S3.9 List of mitonuc genes with outlier frequencies of introgression (introgression frequency >= 85%). Two different RND thresholds were used (10% FDR and 30 %FDR).

FDR	Mitonuc Category	Gene Ensembl ID	Gene Name	Gene Description
10%				
	mitonuc-direct	ENSOCUG00000004871	RARS2	arginyl-tRNA synthetase 2, mitochondrial
	mitonuc	ENSOCUG00000025062	SDHAF4	succinate dehydrogenase complex assembly factor 4
	mitonuc	ENSOCUG00000001408	TYMP	thymidine phosphorylase
	mitonuc	ENSOCUG00000008669	TMLHE	trimethyllysine hydroxylase, epsilon
	mitonuc	ENSOCUG00000011422	L2HGDH	L-2-hydroxyglutarate dehydrogenase
	mitonuc	ENSOCUG00000013358	ATG5	autophagy related 5
30%				
	OXPHOS	ENSOCUG00000006359	NDUFS3	NADH:ubiquinone oxidoreductase core subunit S3
	OXPHOS	ENSOCUG00000007600	-	-
	OXPHOS	ENSOCUG00000009606	COX6B1	cytochrome c oxidase subunit 6B1
	OXPHOS	ENSOCUG00000009899	NDUFV2	NADH:ubiquinone oxidoreductase core subunit V2
	OXPHOS	ENSOCUG00000012164	ATP5G3	ATP synthase, H+ transporting, mitochondrial Fo complex subunit C3 (subunit 9)
	mitonuc-direct	ENSOCUG00000004871	RARS2	arginyl-tRNA synthetase 2, mitochondrial
	mitonuc-direct	ENSOCUG00000005261	-	-
	mitonuc-direct	ENSOCUG00000007961	MRPS27	mitochondrial ribosomal protein S27
	mitonuc-direct	ENSOCUG00000017798	MARS2	methionyl-tRNA synthetase 2, mitochondrial
	mitonuc-direct	ENSOCUG00000021496	-	-
	mitonuc	ENSOCUG00000000628	MUL1	mitochondrial E3 ubiquitin protein ligase 1
	mitonuc	ENSOCUG00000000823	PDE12	phosphodiesterase 12
	mitonuc	ENSOCUG00000001408	TYMP	thymidine phosphorylase
	mitonuc	ENSOCUG00000002448	GLS	glutaminase
	mitonuc	ENSOCUG00000002594	-	ADP/ATP translocase 1
	mitonuc	ENSOCUG00000002763	KIF1B	kinesin family member 1B
	mitonuc	ENSOCUG00000002827	IDE	insulin degrading enzyme

mitonuc	ENSOCUG00000003191	SUGCT	succinyl-CoA:glutarate-CoA transferase
mitonuc	ENSOCUG00000003666	DGUOK	deoxyguanosine kinase
mitonuc	ENSOCUG00000004217	STX17	syntaxin 17
mitonuc	ENSOCUG00000004330	NRDC	nardilysin convertase
mitonuc	ENSOCUG00000004525	NDUFAF2	NADH:ubiquinone oxidoreductase complex assembly factor 2
mitonuc	ENSOCUG00000004820	ABCD3	ATP binding cassette subfamily D member 3
mitonuc	ENSOCUG00000005249	DDAH1	dimethylarginine dimethylaminohydrolase 1
mitonuc	ENSOCUG00000005349	GFM2	G elongation factor mitochondrial 2
mitonuc	ENSOCUG00000005403	SLC30A6	solute carrier family 30 member 6
mitonuc	ENSOCUG00000005515	HIBCH	3-hydroxyisobutyryl-CoA hydrolase
mitonuc	ENSOCUG00000005571	CA5B	carbonic anhydrase 5B
mitonuc	ENSOCUG00000005719	5-Mar	membrane associated ring-CH-type finger 5
mitonuc	ENSOCUG00000005927	ACSM5	acyl-CoA synthetase medium-chain family member 5
mitonuc	ENSOCUG00000005977	FAM185A	family with sequence similarity 185 member A
mitonuc	ENSOCUG00000007245	SIRT5	sirtuin 5
mitonuc	ENSOCUG00000007418	PDK3	pyruvate dehydrogenase kinase 3
mitonuc	ENSOCUG00000007842	KYAT3	kynurenine aminotransferase 3
mitonuc	ENSOCUG00000008144	CHCHD6	coiled-coil-helix-coiled-coil-helix domain containing 6
mitonuc	ENSOCUG00000008669	TMLHE	trimethyllysine hydroxylase, epsilon
mitonuc	ENSOCUG00000009420	NLN	neurolysin
mitonuc	ENSOCUG00000009429	AGPAT5	1-acylglycerol-3-phosphate O-acyltransferase 5
mitonuc	ENSOCUG00000009500	TIMMDC1	translocase of inner mitochondrial membrane domain containing 1
mitonuc	ENSOCUG00000010012	AK4	adenylate kinase 4
mitonuc	ENSOCUG00000010200	PMPCB	peptidase, mitochondrial processing beta subunit
mitonuc	ENSOCUG00000010405	METTL5	methyltransferase like 5
mitonuc	ENSOCUG00000010593	UQCC1	ubiquinol-cytochrome c reductase complex assembly factor 1
mitonuc	ENSOCUG00000010814	ME1	malic enzyme 1

mitonuc	ENSOCUG00000011422	L2HGDH	L-2-hydroxyglutarate dehydrogenase
mitonuc	ENSOCUG00000011557	-	-
mitonuc	ENSOCUG00000012376	SUCLG2	succinate-CoA ligase GDP-forming beta subunit
mitonuc	ENSOCUG00000013358	ATG5	autophagy related 5
mitonuc	ENSOCUG00000013402	AKAP10	A-kinase anchoring protein 10
mitonuc	ENSOCUG00000013662	ACADM	acyl-CoA dehydrogenase, C-4 to C-12 straight chain
mitonuc	ENSOCUG00000013751	MCUB	mitochondrial calcium uniporter dominant negative beta subunit
mitonuc	ENSOCUG00000014103	TMEM11	transmembrane protein 11
mitonuc	ENSOCUG00000014871	COQ5	coenzyme Q5, methyltransferase
mitonuc	ENSOCUG00000015076	FXN	frataxin
mitonuc	ENSOCUG00000015408	ACYP2	acylphosphatase 2
mitonuc	ENSOCUG00000015635	PSMA6	proteasome subunit alpha 6
mitonuc	ENSOCUG00000016663	-	-
mitonuc	ENSOCUG00000016808	HSDL2	hydroxysteroid dehydrogenase like 2
mitonuc	ENSOCUG00000017050	-	-
mitonuc	ENSOCUG00000017062	ROMO1	reactive oxygen species modulator 1
mitonuc	ENSOCUG00000017713	NUBPL	nucleotide binding protein like
mitonuc	ENSOCUG00000022017	DIABLO	diablo IAP-binding mitochondrial protein
mitonuc	ENSOCUG00000022393	-	-
mitonuc	ENSOCUG00000022642	TOMM6	translocase of outer mitochondrial membrane 6
mitonuc	ENSOCUG00000022723	ACAD11	acyl-CoA dehydrogenase family member 11
mitonuc	ENSOCUG00000022792	-	-
mitonuc	ENSOCUG00000025062	SDHAF4	succinate dehydrogenase complex assembly factor 4
mitonuc	ENSOCUG00000027700	YBEY	ybeY metallopeptidase (putative)
mitonuc	ENSOCUG00000029502	MICU2	mitochondrial calcium uptake 2

Table S3.10 List of mitonuc genes with geographic patterns similar to the mtDNA (introgression frequency = 0% in the south and >= 20 % in the north). Two different RND thresholds were used (10% FDR and 30 %FDR).

FDR	Mitonuc Category	Gene Ensembl ID	Gene Name	Gene Description
10%				
	OXPHOS	ENSOCUG00000009899	NDUFV2	NADH:ubiquinone oxidoreductase core subunit V2
	OXPHOS	ENSOCUG00000007809	-	-
	mitonuc-direct	ENSOCUG00000016189	MRPL13	mitochondrial ribosomal protein L13
	mitonuc	ENSOCUG00000002683	TUFM	Tu translation elongation factor, mitochondrial
	mitonuc	ENSOCUG00000006446	CBR4	carbonyl reductase 4
	mitonuc	ENSOCUG00000006904	DCAKD	dephospho-CoA kinase domain containing
	mitonuc	ENSOCUG00000007893	GPAM	glycerol-3-phosphate acyltransferase, mitochondrial
	mitonuc	ENSOCUG00000008974	-	-
	mitonuc	ENSOCUG00000009058	-	-
	mitonuc	ENSOCUG00000011052	PITRM1	pitrilysin metallopeptidase 1
	mitonuc	ENSOCUG00000012393	MCCC1	methylcrotonoyl-CoA carboxylase 1
	mitonuc	ENSOCUG00000012449	HCLS1	hematopoietic cell-specific Lyn substrate 1
	mitonuc	ENSOCUG00000013720	CLYBL	citrate lyase beta like
	mitonuc	ENSOCUG00000014405	TFAM	transcription factor A, mitochondrial
	mitonuc	ENSOCUG00000015344	HEBP1	heme binding protein 1
	mitonuc	ENSOCUG00000016663	-	-
	mitonuc	ENSOCUG00000027099	RPL34	ribosomal protein L34
30%				
	OXPHOS	ENSOCUG00000007809	-	-
	mitonuc-direct	ENSOCUG00000001119	MRPS22	mitochondrial ribosomal protein S22
	mitonuc-direct	ENSOCUG00000001586	MRPL2	mitochondrial ribosomal protein L2
	mitonuc-direct	ENSOCUG00000001949	MRPL15	mitochondrial ribosomal protein L15
	mitonuc-direct	ENSOCUG00000005295	GARS	glycyl-tRNA synthetase
	mitonuc	ENSOCUG00000000563	QRFPR	pyroglutamylated RFamide peptide receptor
	mitonuc	ENSOCUG00000001170	RCC1L	RCC1 like
	mitonuc	ENSOCUG00000001915	-	-

mitonuc	ENSOCUG00000002151	SDR39U1	short chain dehydrogenase/reductase family 39U member 1
mitonuc	ENSOCUG00000002683	TUFM	Tu translation elongation factor, mitochondrial
mitonuc	ENSOCUG00000003969	ABHD10	abhydrolase domain containing 10
mitonuc	ENSOCUG00000006446	CBR4	carbonyl reductase 4
mitonuc	ENSOCUG00000006904	DCAKD	dephospho-CoA kinase domain containing
mitonuc	ENSOCUG00000007360	PTS	6-pyruvoyltetrahydropterin synthase
mitonuc	ENSOCUG00000007893	GPAM	glycerol-3-phosphate acyltransferase, mitochondrial
mitonuc	ENSOCUG00000007949	OPA3	OPA3, outer mitochondrial membrane lipid metabolism regulator
mitonuc	ENSOCUG00000008974	-	-
mitonuc	ENSOCUG00000009058	-	-
mitonuc	ENSOCUG00000009369	VDAC1	voltage dependent anion channel 1
mitonuc	ENSOCUG00000012645	MECR	mitochondrial trans-2-enoyl-CoA reductase
mitonuc	ENSOCUG00000012924	ACOX1	acyl-CoA oxidase 1
mitonuc	ENSOCUG00000013456	OMA1	OMA1 zinc metallopeptidase
mitonuc	ENSOCUG00000014667	DLD	dihydrolipoamide dehydrogenase
mitonuc	ENSOCUG00000015344	HEBP1	heme binding protein 1
mitonuc	ENSOCUG00000016929	APBB1	amyloid beta precursor protein binding family B member 1
mitonuc	ENSOCUG00000017234	GSR	glutathione-disulfide reductase
mitonuc	ENSOCUG00000021541	SHMT1	Serine hydroxymethyltransferase, cytosolic
mitonuc	ENSOCUG00000022444	-	-
mitonuc	ENSOCUG00000023754	PNKD	paroxysmal nonkinesigenic dyskinesia
mitonuc	ENSOCUG00000027099	RPL34	ribosomal protein L34
mitonuc	ENSOCUG00000027418	-	-
mitonuc	ENSOCUG00000029156	PRELID2	PRELI domain containing 2

Table S3.11 Nonsynonymous mutations detected within three mitonuc genes candidates to have co-introgressed with mitochondrial DNA and their potential functional impact inferred using SIFT. ‡Gene names were obtained from the ortholog gene in the mouse when gene names not available in the rabbit annotation.

Chromosome	Position Genome (bp)	Gene	Ensembl ID	Mitonuc category	Reference base	Alternative base	Amino acid and mRNA details	Nb. Species Position	Effect	Score
5	27499631	HP‡	ENSOCUG00000009058	mitonuc	A	C	Tyr(Y)>Ser(S)	25	POTENTIALLY TOLERATED	0.20
5	27499719	HP‡	ENSOCUG00000009058	mitonuc	A	G	Thr(T)>Ala(A)	25	POTENTIALLY TOLERATED	0.20
5	27499743	HP‡	ENSOCUG00000009058	mitonuc	T	C	Phe(F)>Leu(L)	25	POTENTIALLY FUNCTIONAL	0.02
8	29806739	HEBP1	ENSOCUG00000015344	mitonuc	G	C	Glu(E)>Gln(Q)	44	POTENTIALLY FUNCTIONAL	0.00
17	54857177	RP11-561B11.2‡	ENSOCUG00000016663	mitonuc	G	T	Ala(A)>Ser(S)	46	POTENTIALLY TOLERATED	0.68

Table S3.12 RND power to detect introgression at artificially introgressed mitonuc genes using the RND threshold defined at 10% FDR. Power is defined as the proportion of artificially introgressed mitonuc genes detected as introgressed by at least one RND window overlapping the gene.

RND window size	Size of artificially introgressed fragment					
	5kb	**10kb**	**15kb**	**20kb**	**25kb**	**30kb**
10kb	0.83%	10.51%	22.44%	30.94%	37.45%	42.29%
20kb	0.85%	4.48%	12.01%	24.20%	34.74%	43.83%
50kb	0.78%	2.24%	3.36%	6.98%	11.87%	18.49%

Table S3.13 RND power to detect introgression at artificially introgressed mitonuc genes using different RND thresholds based on different FDRs. Power is defined as the proportion of artificially introgressed mitonuc genes detected as introgressed by at least one RND window overlapping the gene of any size (10kb, 20kb or 50kb).

Size of artificially introgressed fragment	Power of detection of Introgression (%)				
	10% FPR	**20% FPR**	**30% FPR**	**40% FPR**	**50% FPR**
5 kb	2.00%	9.77%	21.95%	36.73%	63.94%
10 kb	15.10%	33.61%	48.87%	63.30%	82.57%
15 kb	28.86%	52.46%	68.06%	79.23%	91.24%
20 kb	41.03%	65.47%	78.73%	86.32%	94.16%
25 kb	51.29%	74.31%	84.57%	90.24%	95.83%
30 kb	57.55%	78.82%	87.07%	91.99%	96.66%

References

Acevedo P, Melo-Ferreira J, Farelo L, Beltran-Beck B, Real R, Campos R, Alves PC. 2015. Range dynamics driven by Quaternary climate oscillations explain the distribution of introgressed mtDNA of *Lepus timidus* origin in hares from the Iberian Peninsula. *J Biogeogr* **42**: 1727–1735. http://doi.wiley.com/10.1111/jbi.12556.

Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alfoldi J, Barrio AM, Pielberg G, Rafati N, Sayyab S, Turner-Maier J, et al. 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science (80-)* **345**: 1074–1079. http://www.sciencemag.org/content/345/6200/1074.full.html.

Melo-Ferreira J, Alves P, Rocha J, Ferrand N, Boursot P. 2011. Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? *Evolution* **65**: 1956–68. http://www.ncbi.nlm.nih.gov/pubmed/21729051 (Accessed August 15, 2012).

Annex III. Supplementary material from paper III in Chapter 3. Genomic perspective of introgression in hares from Iberia

List of Supplementary Figures and Tables

Table S3.19 List of genes with *L. timidus* introgression at frequencies of at least 50% in Iberian Peninsula *L. europaeus* as inferred by the ELAI analysis considering Iberian *L. europaeus* as focal population and only *L. timidus* and non-Iberian *L. europaeus* as parental populations.

Table S3.20 List of GO enriched terms for the set genes in regions of introgression frequency of at least 50% in Iberian Peninsula *L. europaeus* as inferred by the ELAI analysis considering Iberian *L. europaeus* as focal population.

Table S3.21 List of genes in outlier FST windows (99.9% percentile) between Iberian and non-Iberian *L. europaeus*.

Table S3.22 List of GO enriched terms for the set genes in outlier FST windows (99.9% percentile) between Iberian and non-Iberian *L. europaeus*.

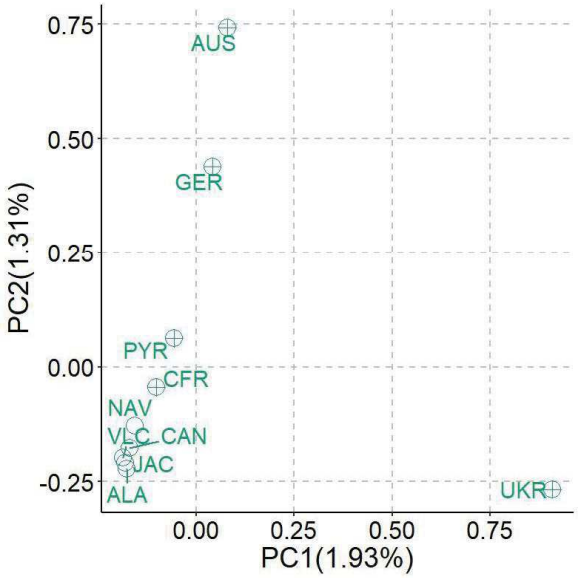


Figure S3.14 Principal component analyses with all *L. europaeus* individuals. Eigenvalues for each of the principal components are given within parenthesis.

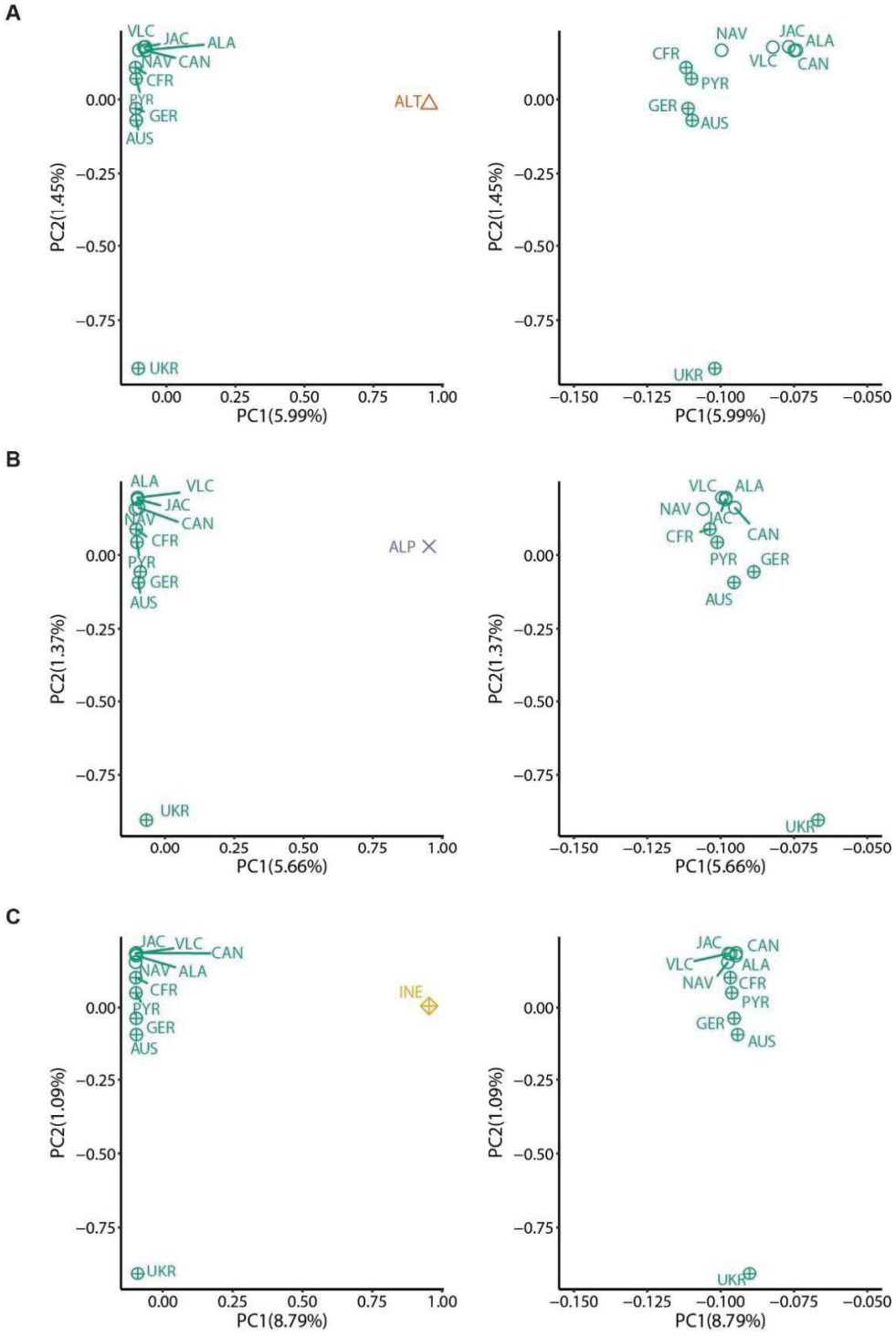


Figure S3.15 PCA summary of genetic variation in *L. europaeus* including outgroups: (A) including *L. granatensis*; (B) including *L. timidus*; (C) including *L. americanus*. Plots on the left show all samples, those on the right a zoom in *L. europaeus* samples only. Eigenvalues for each of the principal components are given within parenthesis.

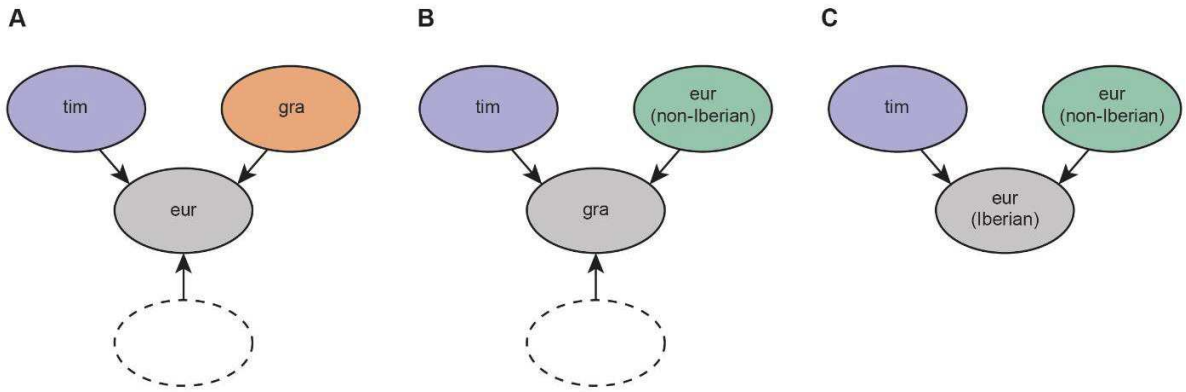


Figure S3.16 Schematic representation of ELAI analysis using three different settings. (A) ELAI inference of the origin of admixture in the population of *L. europaeus* (including all individuals sampled in this study; grey circle) from three sources: *L. timidus* (blue circle), *L. granatensis* (red circle) and *L. europaeus* (dashed white circle). Note that given the lack of a pure parental *L. europaeus* population we let ELAI infer this third source from the data of the admixed population. (B) ELAI inference of the origin of admixture in the population of Iberian *L. granatensis* (grey circle) from three sources: *L. timidus* (blue circle), non-Iberian *L. europaeus* (green circle) and *L. granatensis* (dashed white circle). Again, given the lack of a pure parental *L. granatensis* population we let ELAI infer this third source from the data of the admixed population. (C) ELAI inference of the origin of admixture in the Iberian population of *L. europaeus* (grey circle) from two sources: *L. timidus* (blue circle) and non-Iberian *L. europaeus* (green circle).

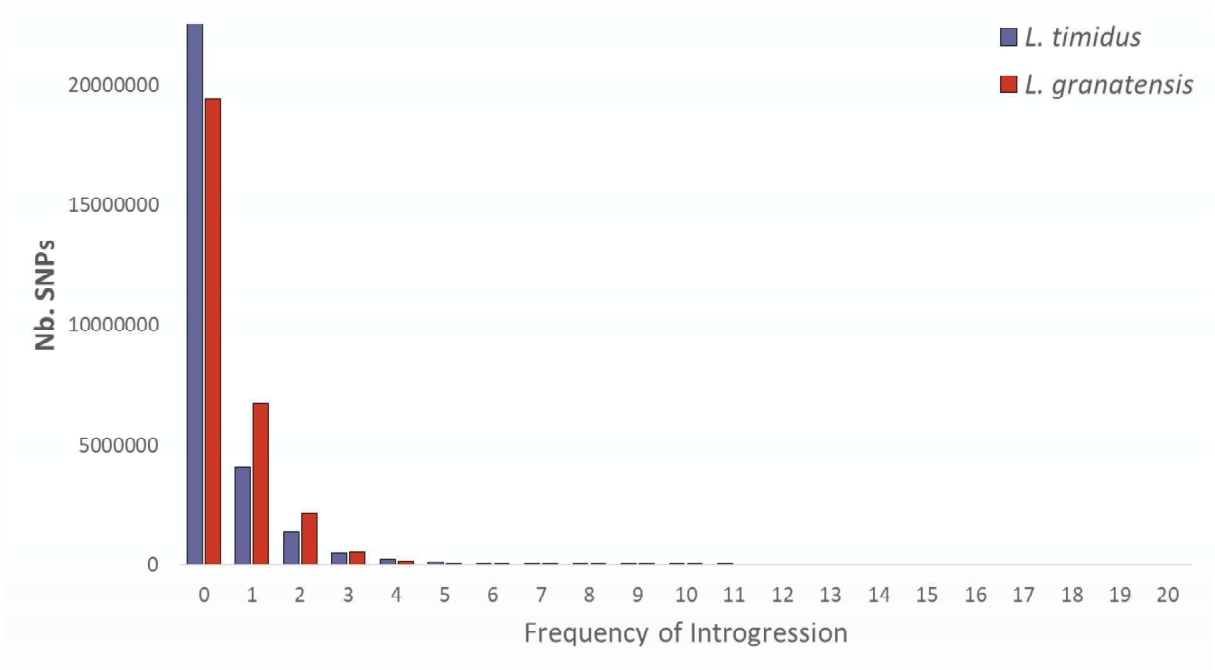


Figure S3.17 Frequency of introgression by SNP of *L. timidus* (blue) and *L. granatensis* (red) origin into the 10 *L. europaeus* individuals, as inferred by ELAI.

		Iberian Peninsula										
		0	1	2	3	4	5	6	7	8	9	10
Europe	0	2E+07	6E+05	2E+05	60494	24833	4324	1548	0	0	0	0
	1	3E+06	3E+05	1E+05	44988	26038	6490	414	134	0	0	0
	2	9E+05	1E+05	61946	26268	14048	2754	2415	0	0	0	0
	3	2E+05	57287	25642	11716	6450	1732	550	0	0	0	0
	4	54035	13840	5784	3484	998	94	0	0	0	0	0
	5	12552	4473	1010	64	39	763	28	0	0	0	0
	6	2202	823	70	4	0	0	0	0	0	0	0
	7	191	266	0	0	8	0	0	0	0	0	0
	8	327	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	0

Figure S3.18 2D frequency of introgression by SNP of *L. timidus* origin into the 5 Iberian and 5 non-Iberian *L. europaeus* individuals, as inferred by ELAI.

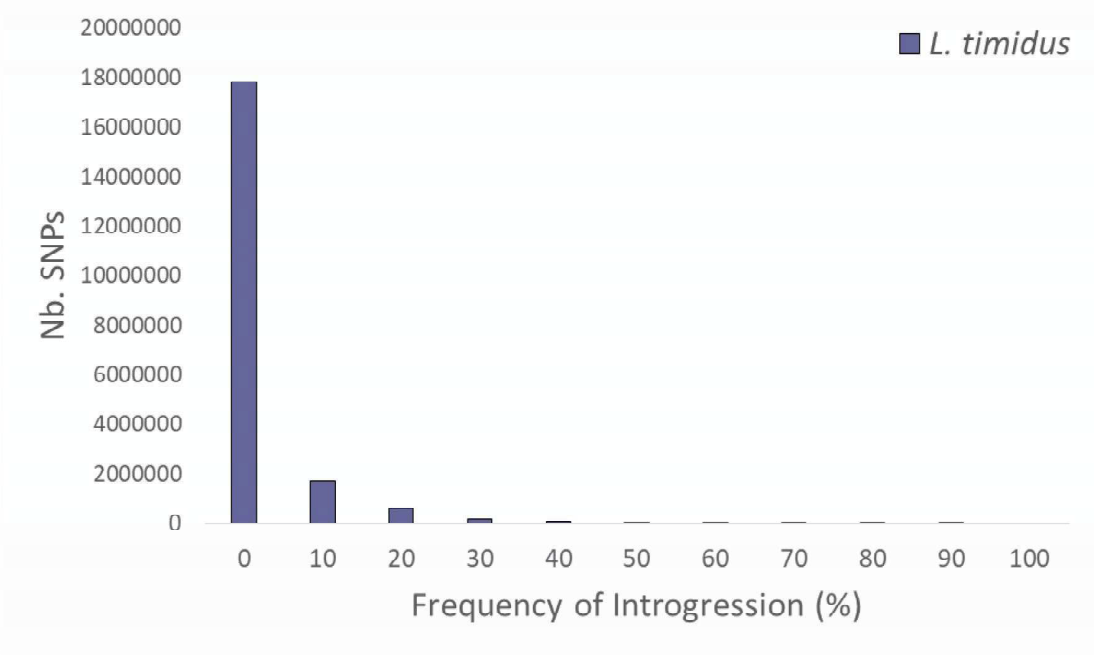


Figure S3.19 Frequency of introgression by SNP of *L. timidus* (blue) origin into the 5 Iberian *L. europaeus* individuals, as inferred by ELAI.

Table S3.14 Sampling localities, tissue used for genomic DNA extraction, mitochondrial DNA type and raw sequencing coverage of specimens sequenced in this study.

Species	Individual code	Population Code	Locality	Lat	Lon	Tissue	mtDNA type[a]	Raw Coverage (X)	Reference
Lepus *europaeus*									
	eur01	CAN	Cantabria, Spain	43.182890	-3.986640	ear	tim	13.4	this work
	eur02	JAC	Jaca, Spain	42.570060	-0.547060	ear	tim	16.3	this work
	eur03	VLC	Villarcayo, Spain	42.944830	-3.561030	ear	tim	14.1	this work
	eur04	ALA	Alava, Spain	42.910000	-2.698390	organ	tim	13.2	this work
	eur05	NAV	Navarra, Spain	42.895640	-2.170810	organ	tim	18.9	this work
	eur06	PYR	French Pyrenees, France	42.516670	2.016670	ear	eur	14.1	this work
	eur07	UKR	Ukraine	-	-	ear	eur	16.9	this work
	eur08	GER	Germany	-	-	organ	eur	15.5	this work
	eur09	AUS	Vienna, Austria	-	-	organ	eur	18.6	this work
	eur10	CFR	Clermont-Ferrand, France	45.777220	3.087060	organ	eur	11.5	this work
Lepus granatensis									
	gra01	ALT	Alcoutim, Portugal	37.469978	-7.473078	Ear	gra	26.9	Seixas et al. submitted
	gra02	SEV	Seville, Spain	37.389092	-5.984459	Ear	gra	25.5	Seixas et al. submitted
	gra03	PAN	Pancas, Portugal	38.809101	-8.918929	Kidney	gra	22.8	Seixas et al. submitted
	gra04	CBR	Castelo Branco, Portugal	39.924751	-7.241590	Organ	gra	26.2	Seixas et al. submitted
	gra05	CRE	Ciudad Real, Spain	38.984829	-3.927378	Kidney	gra	25.6	Seixas et al. submitted
	gra06	VLP	Valpaços, Portugal	41.608715	-7.310906	Kidney	tim	27.6	Seixas et al. submitted
	gra07	MAD	Madrid, Spain	40.416775	-3.70379	Ear	tim	28.7	Seixas et al. submitted
	gra08	VAL	Valencia, Spain	39.469910	-0.376288	Ear	tim	23.2	Seixas et al. submitted
	gra09	SOR	Soria, Spain	41.764431	-2.463772	Ear	tim	23.3	Seixas et al. submitted

gra10	NAV	Navarra, Spain	42.695393	-1.676069	Kidney	tim	27.7	Seixas et al. submitted
Lepus timidus								
tim01	SCA	Scandinavia	-	-	Kidney	-	23.2	Seixas et al. submitted
tim02	ALP	Switzerland, Alps	46.841560	9.594860	Kidney	-	25.1	Seixas et al. submitted
tim03	ALP	France, Alps	46.043150	6.579070	Ear	-	28.5	Seixas et al. submitted
tim04	IRE	Borris-in-Ossory, Ireland	-	-	Kidney	-	37.7	this work
Lepus americanus								
ame01	MON	Montana, USA	47.040180	-113.554680	Ovarian	-	37.4	Seixas et al. submitted; Carneiro et al. 2014

[a]mtDNA type: eur - *L. europaeus*; gra - *L. granatensis*; tim - *L. timidus*.

Table S3.15 Results of the *D*-statistic calculated between Iberian and non-Iberian *L. europaeus* populations (focal) and using either *L. timidus* or *L. granatensis* as the donor population. Negative values of the D-statistic indicate introgression into the non-Iberian population while positive values indicate introgression into the Iberian population. Standard error was estimated using a weighted block jackknife approach with blocks of 5 Mb.

popA (outgroup)	popB (donor)	popX (D<0)	popY (D>0)	D(A, B; X, Y)	SE	Z-score (D/SE)	Nb. Informative sites	Nb. Jackknife blocks
L. americanus	*L. timidus*	*L. europaeus* (non-Iberian)	*L. europaeus* (Iberia)	-0.024	0.004	-6.064*	947514	432
L. americanus	*L. granatensis*	*L. europaeus* (non-Iberian)	*L. europaeus* (Iberia)	0.285	0.009	31.196*	1263788	432

**D* significantly different from zero after converting Z-cores to a two-tailed P-value and using α = 0.01 as a cutoff for significance.

Table S3.16 - *D*-statistic between pairwise comparisons of *L. europaeus* individuals as focal and with *L. timidus* (top) or *L. granatensis* (bottom) as donor populations. Negative values of the D-statistic indicate introgression into individuals from population X while positive values indicate introgression into individuals of population Y. Underscored values indicate comparisons for which *D* was significantly different from zero after converting Z-cores to a two-tailed P-value and using α = 0.01 as a cutoff for significance.

***timidus* donor**		**Population Y (D>0)**									
	Individuals	eur01 (CAN)	eur02 (JAC)	eur03 (VLC)	eur04 (ALA)	eur05 (NAV)	eur06 (PYR)	eur07 (UKR)	eur08 (GER)	eur09 (AUS)	eur10 (CFR)
Population X (D<0)	eur01 (CAN)	-	-	-	-	-	-	-	-	-	-
	eur02 (JAC)	-0.013	-	-	-	-	-	-	-	-	-
	eur03 (VLC)	0.008	0.021	-	-	-	-	-	-	-	-
	eur04 (ALA)	0.000	0.013	-0.008	-	-	-	-	-	-	-
	eur05 (NAV)	0.021	0.036	0.014	0.023	-	-	-	-	-	-
	eur06 (PYR)	0.025	0.039	0.019	0.027	0.006	-	-	-	-	-
	eur07 (UKR)	-0.111	-0.104	-0.121	-0.115	-0.136	-0.142	-	-	-	-
	eur08 (GER)	-0.071	-0.061	-0.081	-0.073	-0.097	-0.102	0.053	-	-	-
	eur09 (AUS)	0.032	0.045	0.027	0.034	0.015	0.010	0.150	0.110	-	-
	eur10 (CFR)	0.069	0.078	0.056	0.064	0.044	0.036	0.175	0.141	0.023	-

***granatensis* donor**		**Population Y (D>0)**									
	Individuals	eur01 (CAN)	eur02 (JAC)	eur03 (VLC)	eur04 (ALA)	eur05 (NAV)	eur06 (PYR)	eur07 (UKR)	eur08 (GER)	eur09 (AUS)	eur10 (CFR)
Population X (D<0)	eur01 (CAN)	-	-	-	-	-	-	-	-	-	-
	eur02 (JAC)	-0.016	-	-	-	-	-	-	-	-	-
	eur03 (VLC)	0.021	0.038	-	-	-	-	-	-	-	-
	eur04 (ALA)	-0.007	0.010	-0.030	-	-	-	-	-	-	-
	eur05 (NAV)	0.183	0.208	0.173	0.201	-	-	-	-	-	-
	eur06 (PYR)	0.279	0.308	0.275	0.301	0.123	-	-	-	-	-
	eur07 (UKR)	0.258	0.280	0.251	0.272	0.119	0.019	-	-	-	-
	eur08 (GER)	0.290	0.317	0.286	0.310	0.141	0.026	0.003	-	-	-
	eur09 (AUS)	0.312	0.336	0.308	0.330	0.167	0.054	0.028	0.028	-	-
	eur10 (CFR)	0.382	0.378	0.349	0.373	0.192	0.055	0.024	0.024	-0.007	-

Table S3.17 List of genes with *L. timidus* introgression at frequencies of at least 50% in Iberian Peninsula *L. europaeus* and not greater than 20% in the non-Iberian *L. europaeus* as inferred by the ELAI analysis considering all *L. europaeus* individuals as focal population.

Chromosome	Gene start (bp)	Gene end (bp)	Ensembl ID	Gene Name	Description
1	185708242	185920855	ENSOCUG00000016117	AMBRA1	autophagy and beclin 1 regulator 1
1	185709794	185716156	ENSOCUG00000024752	-	-
1	185716977	185718413	ENSOCUG00000021064	-	-
3	136227979	136273841	ENSOCUG00000016189	MRPL13	mitochondrial ribosomal protein L13
3	136273885	136364423	ENSOCUG00000016195	MTBP	MDM2 binding protein
3	136372159	136676563	ENSOCUG00000016211	SNTB1	syntrophin beta 1
7	164083316	164128824	ENSOCUG00000005829	MOGAT1	monoacylglycerol O-acyltransferase 1
7	164069167	164069263	ENSOCUG00000020962	-	-
10	42978137	43100880	ENSOCUG00000012016	SEMA3D	semaphorin 3D
12	151424148	151424387	ENSOCUG00000025750	-	-
12	151412157	151412502	ENSOCUG00000026726	-	-
13	26987889	26999994	ENSOCUG00000008446	UCK2	uridine-cytidine kinase 2
13	31095227	31311165	ENSOCUG00000007930	ATF6	activating transcription factor 6
13	31321012	31329121	ENSOCUG00000007924	DUSP12	dual specificity phosphatase 12
13	95981970	96177185	ENSOCUG00000001980	-	-
13	133680611	133842737	ENSOCUG00000000264	UBR4	ubiquitin protein ligase E3 component n-recognin 4
13	31317949	31318082	ENSOCUG00000018742	-	-
14	91324582	91356602	ENSOCUG00000010327	BDH1	3-hydroxybutyrate dehydrogenase 1
16	67517847	67541186	ENSOCUG00000000005	MDM4	MDM4, p53 regulator
X	58424321	58587973	ENSOCUG00000016955	TMEM164	transmembrane protein 164
AAGW02080735	25889	36474	ENSOCUG00000023676	-	-
GL018703	910314	1631064	ENSOCUG00000017426	GRM7	glutamate metabotropic receptor 7
GL018704	4179058	4793829	ENSOCUG00000007194	CSMD2	CUB and Sushi multiple domains 2

GL018709	4159443	4365170	ENSOCUG00000017773	TNKS	tankyrase
GL018716	2281015	2537427	ENSOCUG00000015023	-	-
GL018716	1851532	1851695	ENSOCUG00000019588	-	-
GL018718	1534013	2520890	ENSOCUG00000001573	PTPRT	protein tyrosine phosphatase, receptor type T
GL018753	152336	181782	ENSOCUG00000014169	ASB7	ankyrin repeat and SOCS box containing 7
GL018770	9788	22645	ENSOCUG00000023975	-	-
GL018770	29588	34299	ENSOCUG00000024982	-	-
GL018801	661701	673463	ENSOCUG00000004892	-	-
GL018885	300368	301279	ENSOCUG00000017075	ORYCUNV1R1598	Oryctolagus cuniculus vomeronasal 1 receptor oryCunV1R1598 (ORYCUNV1R1598), mRNA
GL018930	125577	261350	ENSOCUG00000012906	SMOC2	SPARC related modular calcium binding 2

Table S3.18 List of GO enriched terms for the set genes in regions of introgression frequency of at least 50% in Iberian Peninsula *L. europaeus* and not greater than 20% in the non-Iberian *L. europaeus* as inferred by the ELAI analysis considering all *L. europaeus* individuals as focal population.

Ontology	GO code	P-value	GO description
BP			
	GO:0046463	4.58E-02	acylglycerol biosynthetic process
	GO:0007196	1.85E-02	adenylate cyclase-inhibiting G-protein coupled glutamate receptor signaling pathway
	GO:0045839	2.25E-03	negative regulation of mitotic nuclear division
	GO:0046339	2.77E-02	diacylglycerol metabolic process
	GO:0045023	1.54E-02	G0 to G1 transition
	GO:0034502	8.52E-03	protein localization to chromosome
	GO:0098780	4.28E-02	response to mitochondrial depolarisation
	GO:0009173	1.54E-02	pyrimidine ribonucleoside monophosphate metabolic process
	GO:0007205	3.98E-02	protein kinase C-activating G-protein coupled receptor signaling pathway
	GO:0043173	2.16E-02	nucleotide salvage
	GO:1904355	3.37E-02	positive regulation of telomere capping
	GO:0034091	2.16E-02	regulation of maintenance of sister chromatid cohesion
	GO:1900087	4.58E-02	positive regulation of G1/S transition of mitotic cell cycle
	GO:0035518	3.68E-02	histone H2A monoubiquitination
	GO:0043552	4.58E-02	positive regulation of phosphatidylinositol 3-kinase activity
	GO:0030815	2.77E-02	negative regulation of cAMP metabolic process
	GO:0036003	3.68E-02	positive regulation of transcription from RNA polymerase II promoter in response to stress
	GO:0000423	4.28E-02	macromitophagy
	GO:0016567	1.57E-02	protein ubiquitination
CC			
	GO:0005614	3.98E-02	interstitial matrix

MF

GO:0004143	2.77E-02	diacylglycerol kinase activity
GO:0010851	1.85E-02	cyclase regulator activity
GO:0045295	3.07E-02	gamma-catenin binding
GO:0019206	1.85E-02	nucleoside kinase activity
GO:0004721	5.00E-02	phosphoprotein phosphatase activity
GO:0003951	3.98E-02	NAD+ kinase activity
GO:0045294	2.46E-02	alpha-catenin binding
GO:0008270	4.80E-02	zinc ion binding
GO:0035497	2.77E-02	cAMP response element binding

Table S3.19 List of genes with *L. timidus* introgression at frequencies of at least 50% in Iberian Peninsula *L. europaeus* as inferred by the ELAI analysis considering Iberian *L. europaeus* as focal population and only *L. timidus* and non-Iberian *L. europaeus* as parental populations.

Chromosome	Gene start (bp)	Gene end (bp)	Ensembl ID	Gene Name	Description
3	102379944	102774820	ENSOCUG00000029386	CNBD1	cyclic nucleotide binding domain containing 1
3	109343610	109373029	ENSOCUG00000010331	FAM92A	family with sequence similarity 92 member A
3	109374574	109377481	ENSOCUG00000010353	RBM12B	RNA binding motif protein 12B
3	109385936	109445106	ENSOCUG00000010390	TMEM67	transmembrane protein 67
3	136273885	136364423	ENSOCUG00000016195	MTBP	MDM2 binding protein
3	136372159	136676563	ENSOCUG00000016211	SNTB1	syntrophin beta 1
3	109336145	109338151	ENSOCUG00000024826	-	-
4	8577541	8622454	ENSOCUG00000016747	-	-
4	8653983	8661756	ENSOCUG00000025098	C20orf202	chromosome 20 open reading frame 202
4	8672171	8707135	ENSOCUG00000003130	RAD21L1	RAD21 cohesin complex component like 1
4	8719686	8753923	ENSOCUG00000013195	SNPH	syntaphilin
4	79595275	79648458	ENSOCUG00000015995	SLC17A8	solute carrier family 17 member 8
4	8636138	8636423	ENSOCUG00000028996	-	-
5	22139416	22167617	ENSOCUG00000006143	DYNC1LI2	dynein cytoplasmic 1 light intermediate chain 2
5	22171278	22219216	ENSOCUG00000000569	TERB1	telomere repeat binding bouquet formation protein 1
5	22204312	22204522	ENSOCUG00000028866	-	-
6	12718088	12751332	ENSOCUG00000003891	-	-
6	12750865	12767101	ENSOCUG00000003902	PDZD9	PDZ domain containing 9
6	12774572	12835114	ENSOCUG00000029509	C16orf52	chromosome 16 open reading frame 52
6	14533586	14885558	ENSOCUG00000013916	PRKCB	protein kinase C beta type isoform II
6	12744568	12744697	ENSOCUG00000020870	-	-
6	12744907	12745074	ENSOCUG00000028431	-	-

8	13154735	13210330	ENSOCUG00000008344	TM7SF3	transmembrane 7 superfamily member 3
8	13200127	13224570	ENSOCUG00000008342	FGFR1OP2	FGFR1 oncogene partner 2
8	13241744	13277235	ENSOCUG00000008328	INTS13	integrator complex subunit 13
8	49695149	49715031	ENSOCUG00000008419	SLC25A30	solute carrier family 25 member 30
8	49843112	49974880	ENSOCUG00000003537	GTF2F2	general transcription factor IIF subunit 2
8	109997646	109998659	ENSOCUG00000015739	ABHD13	abhydrolase domain containing 13
8	110029766	110063343	ENSOCUG00000012271	TNFSF13B	TNF superfamily member 13b
10	21833327	22020788	ENSOCUG00000000676	VPS41	VPS41, HOPS complex subunit
10	44317155	44562791	ENSOCUG00000010520	SEMA3E	semaphorin 3E
12	142269249	142620735	ENSOCUG00000023878	SYNE1	spectrin repeat containing nuclear envelope protein 1
13	5891451	6011914	ENSOCUG00000005161	RALGPS2	Ral GEF with PH domain and SH3 binding motif 2
13	5947850	5969699	ENSOCUG00000005167	ANGPTL1	angiopoietin like 1
13	26987889	26999994	ENSOCUG00000008446	UCK2	uridine-cytidine kinase 2
13	27115123	27175196	ENSOCUG00000008441	TMCO1	transmembrane and coiled-coil domains 1
13	31095227	31311165	ENSOCUG00000007930	ATF6	activating transcription factor 6
13	31321012	31329121	ENSOCUG00000007924	DUSP12	dual specificity phosphatase 12
13	31355697	31360643	ENSOCUG00000007914	FCRLB	Fc receptor like B
13	80013899	80103166	ENSOCUG00000024789	LPAR3	lysophosphatidic acid receptor 3
13	133680611	133842737	ENSOCUG00000000264	UBR4	ubiquitin protein ligase E3 component n-recognin 4
13	5840645	5840720	ENSOCUG00000021139	-	-
13	31317949	31318082	ENSOCUG00000018742	-	-
14	89279682	89426464	ENSOCUG00000015801	ATP13A4	ATPase 13A4
14	91324582	91356602	ENSOCUG00000010327	BDH1	3-hydroxybutyrate dehydrogenase 1
16	36735239	36782855	ENSOCUG00000005645	TRIM67	tripartite motif containing 67
16	67517847	67541186	ENSOCUG00000000005	MDM4	MDM4, p53 regulator
17	76553682	76586741	ENSOCUG00000003358	ACTR10	actin-related protein 10 homolog

19	24904373	24921322	ENSOCUG00000002842	HEATR9	HEAT repeat containing 9
19	24922478	24932917	ENSOCUG00000002848	CCL5	C-C motif chemokine ligand 5
19	24961758	24964697	ENSOCUG00000002849	CCL14	C-C motif chemokine ligand 14
19	24970288	24977839	ENSOCUG00000029281	-	-
19	24980659	24987202	ENSOCUG00000000686	-	-
19	49023108	49117761	ENSOCUG00000012112	TEX2	testis expressed 2
20	6418615	6659515	ENSOCUG00000004667	FUT8	fucosyltransferase 8
20	32947712	32948001	ENSOCUG00000012114	-	-
20	6682135	6682241	ENSOCUG00000020950	-	-
X	58424321	58587973	ENSOCUG00000016955	TMEM164	transmembrane protein 164
AAGW02082356	111	7682	ENSOCUG00000023579	-	-
AAGW02082356	12602	14216	ENSOCUG00000022221	-	-
AAGW02082456	1551	13076	ENSOCUG00000027331	-	-
GL018704	4179058	4793829	ENSOCUG00000007194	CSMD2	CUB and Sushi multiple domains 2
GL018715	937998	1232057	ENSOCUG00000004407	-	-
GL018716	1851532	1851695	ENSOCUG00000019588	-	-
GL018725	1443600	1498348	ENSOCUG00000017615	SLC13A3	solute carrier family 13 member 3
GL018728	2482081	2640926	ENSOCUG00000002964	AGTPBP1	ATP/GTP binding protein 1
GL018737	100294	510852	ENSOCUG00000002672	ADAMTSL3	ADAMTS like 3
GL018737	1518058	1683475	ENSOCUG00000010384	IL16	interleukin 16
GL018737	1637547	1638296	ENSOCUG00000022959	-	-
GL018746	755979	814644	ENSOCUG00000009907	IREB2	iron responsive element binding protein 2
GL018753	1000851	1137948	ENSOCUG00000006434	LRRC28	leucine rich repeat containing 28
GL018753	1139165	1233439	ENSOCUG00000006429	TTC23	tetratricopeptide repeat domain 23
GL018754	762078	832090	ENSOCUG00000000100	-	-
GL018770	1957	7765	ENSOCUG00000021842	-	-
GL018770	9788	22645	ENSOCUG00000023975	-	-

GL018770	29588	34299	ENSOCUG00000024982	-	-
GL018770	105257	106348	ENSOCUG00000027867	-	-
GL018770	383108	407803	ENSOCUG00000022395	-	-
GL018770	1225692	1226857	ENSOCUG00000007188	-	-
GL018780	345126	451719	ENSOCUG00000007071	ZRANB1	zinc finger RANBP2-type containing 1
GL018782	434593	463433	ENSOCUG00000002397	UBE4A	ubiquitination factor E4A
GL018782	465651	474793	ENSOCUG00000029455	-	-
GL018801	661701	673463	ENSOCUG00000004892	-	-
GL018868	9385	70351	ENSOCUG00000022005	ICE1	interactor of little elongation complex ELL subunit 1
GL019253	52485	53285	ENSOCUG00000025774	-	-

Table S3.20 List of GO enriched terms for the set genes in regions of introgression frequency of at least 50% in Iberian Peninsula *L. europaeus* as inferred by the ELAI analysis considering Iberian *L. europaeus* as focal population.

Ontology	GO code	P-value	GO description
MF			
-	GO:0008009	0.05	chemokine activity
*BP**			
-	GO:0045141	0.05	meiotic telomere clustering
-			

*keeping only one chemokine gene

Table S3.21 List of genes in outlier FST windows (99.9% percentile) between Iberian and non-Iberian *L. europaeus*.

Chromosome	Gene start (bp)	Gene end (bp)	Ensembl ID	Gene Name	Description
1	156820308	156950700	ENSOCUG00000003080	INSC	inscuteable homolog (Drosophila)
1	104608911	104645120	ENSOCUG00000012930	PPP2R1B	protein phosphatase 2 scaffold subunit Abeta
1	104476339	104590490	ENSOCUG00000012966	-	-
2	82686719	82793559	ENSOCUG00000003608	DAAM1	dishevelled associated activator of morphogenesis 1
2	158389352	158658476	ENSOCUG00000004316	-	-
2	131494285	131640710	ENSOCUG00000005008	SPTBN1	spectrin beta, non-erythrocytic 1
2	152968663	153407758	ENSOCUG00000009598	LTBP1	latent transforming growth factor beta binding protein 1
2	37666226	37749814	ENSOCUG00000012352	DCUN1D4	defective in cullin neddylation 1 domain containing 4
2	169765514	169912749	ENSOCUG00000015795	LDAH	lipid droplet associated hydrolase
2	5807716	5930877	ENSOCUG00000016551	C1QTNF7	C1q and TNF related 7
3	106086826	106087319	ENSOCUG00000001379	-	-
3	109343610	109373029	ENSOCUG00000010331	FAM92A	family with sequence similarity 92 member A
3	109374574	109377481	ENSOCUG00000010353	RBM12B	RNA binding motif protein 12B
3	109385936	109445106	ENSOCUG00000010390	TMEM67	transmembrane protein 67
3	36258249	36304851	ENSOCUG00000011180	GEMIN5	gem nuclear organelle associated protein 5
3	36306185	36340892	ENSOCUG00000011189	MRPL22	mitochondrial ribosomal protein L22
3	97298413	97321456	ENSOCUG00000014535	IMPA1	inositol monophosphatase 1
3	97336812	97338128	ENSOCUG00000017530	SLC10A5	solute carrier family 10 member 5
3	97344555	97361180	ENSOCUG00000017532	ZFAND1	zinc finger AN1-type containing 1
3	36354511	36354762	ENSOCUG00000024002	-	-
3	109336145	109338151	ENSOCUG00000024826	-	-
4	25612276	25788826	ENSOCUG00000002878	SLX4IP	SLX4 interacting protein

7	3003229	3024229	ENSOCUG00000001661	PDIA4	protein disulfide isomerase family A member 4
7	8814083	8814918	ENSOCUG00000002030	TMEM139	transmembrane protein 139
7	8794797	8812804	ENSOCUG00000002158	CASP2	caspase 2
7	38467943	38533983	ENSOCUG00000005195	RSBN1L	round spermatid basic protein 1 like
7	21978280	22440132	ENSOCUG00000007425	CADPS2	calcium dependent secretion activator 2
7	2872600	3027194	ENSOCUG00000008416	ZNF398	zinc finger protein 398
7	8755732	8788889	ENSOCUG00000011147	CLCN1	chloride voltage-gated channel 1
7	55487700	56233251	ENSOCUG00000016546	CNTNAP5	contactin associated protein like 5
7	72797849	73025870	ENSOCUG00000017366	R3HDM1	R3H domain containing 1
7	72954492	72954573	ENSOCUG00000018487	-	-
7	3028374	3028485	ENSOCUG00000018632	-	-
7	3031347	3031448	ENSOCUG00000018974	-	-
7	76787042	76787284	ENSOCUG00000021532	-	-
7	21975922	21977049	ENSOCUG00000024828	RNF133	ring finger protein 133
7	8825809	8855988	ENSOCUG00000029426	GSTK1	glutathione S-transferase kappa 1
8	68573629	68730823	ENSOCUG00000000774	PCDH9	protocadherin 9
9	42469361	42469470	ENSOCUG00000021387	5S_rRNA	5S ribosomal RNA
11	81327343	81634734	ENSOCUG00000001525	ADAMTS6	ADAM metallopeptidase with thrombospondin type 1 motif 6
12	123624576	123889314	ENSOCUG00000000635	EYA4	EYA transcriptional coactivator and phosphatase 4
12	80433019	80622164	ENSOCUG00000000672	EPHA7	EPH receptor A7
12	106520031	106564747	ENSOCUG00000002627	GOPC	golgi associated PDZ and coiled-coil motif containing
12	106622959	106651983	ENSOCUG00000011553	NUS1	NUS1 dehydrodolichyl diphosphate synthase subunit
12	122856450	122856830	ENSOCUG00000013335	-	-
12	98387435	98575755	ENSOCUG00000014528	CDK19	cyclin dependent kinase 19
12	122862128	122863138	ENSOCUG00000017661	TAAR5	trace amine associated receptor 5
12	35172602	35275786	ENSOCUG00000022531	-	-

12	106577021	106606698	ENSOCUG00000024042	-	-
12	122871275	122875771	ENSOCUG00000024372	-	-
12	122880816	122881847	ENSOCUG00000026295	-	-
13	23534600	23560296	ENSOCUG00000002795	FMO4	dimethylaniline monooxygenase
13	23602459	23630041	ENSOCUG00000008982	BLZF1	basic leucine zipper nuclear factor 1
13	32379017	32393383	ENSOCUG00000013551	CD48	CD48 molecule
13	32319849	32338970	ENSOCUG00000013559	SLAMF7	SLAM family member 7
14	122637852	122708360	ENSOCUG00000012874	-	-
15	73520194	73520676	ENSOCUG00000023590	-	-
16	56817993	57649567	ENSOCUG00000014710	USH2A	usherin
16	20132168	20462378	ENSOCUG00000017317	CDC42BPA	CDC42 binding protein kinase alpha
17	37364573	37470179	ENSOCUG00000003488	AQR	aquarius intron-binding spliceosomal factor
17	56307183	56534369	ENSOCUG00000014333	SLC25A21	solute carrier family 25 member 21
17	37353737	37355818	ENSOCUG00000024670	ZNF770	zinc finger protein 770
18	31335757	31990614	ENSOCUG00000000268	PCDH15	protocadherin related 15
19	25217906	25272383	ENSOCUG00000003682	HNF1B	HNF1 homeobox B
20	24195498	24359811	ENSOCUG00000001374	TSHR	thyroid stimulating hormone receptor
X	75931336	75970114	ENSOCUG00000003369	APOOL	apolipoprotein O like
X	16918695	16998407	ENSOCUG00000007836	TAB3	TGF-beta activated kinase 1 and MAP3K7 binding protein 3
X	75973254	75986688	ENSOCUG00000010085	-	-
X	75846416	75849351	ENSOCUG00000021447	-	-
X	75833970	75834413	ENSOCUG00000023135	-	-
X	75691868	75708334	ENSOCUG00000027774	-	-
GL018700	1672768	1673516	ENSOCUG00000006218	-	-
GL018700	6808359	6809328	ENSOCUG00000022716	-	-
GL018705	3366033	3426434	ENSOCUG00000006781	KPNA3	karyopherin subunit alpha 3

GL018705	5355122	5546728	ENSOCUG00000010151	WDFY2	WD repeat and FYVE domain containing 2
GL018705	3368395	3368452	ENSOCUG00000028621	-	-
GL018728	1063449	1072376	ENSOCUG00000010364	-	-
GL018728	1089242	1091107	ENSOCUG00000013511	RMI1	RecQ mediated genome instability 1
GL018731	1024919	1072973	ENSOCUG00000009436	CNGA2	cyclic nucleotide gated channel alpha 2
GL018731	714121	733055	ENSOCUG00000009667	GABRE	gamma-aminobutyric acid type A receptor epsilon subunit
GL018731	641498	643924	ENSOCUG00000009992	MAGEA10	MAGE family member A10
GL018731	785566	786627	ENSOCUG00000016680	-	-
GL018731	727736	727815	ENSOCUG00000022740	-	-
GL018731	726699	726783	ENSOCUG00000023918	-	-
GL018734	2451183	2593562	ENSOCUG00000012263	ENTHD1	ENTH domain containing 1
GL018735	587287	636672	ENSOCUG00000015683	LNX2	ligand of numb-protein X 2
GL018735	685385	687368	ENSOCUG00000015706	POLR1D	RNA polymerase I subunit D
GL018748	1308978	1354445	ENSOCUG00000001479	MAGED1	MAGE family member D1
GL018748	1207064	1208845	ENSOCUG00000014736	-	-
GL018748	1337365	1337496	ENSOCUG00000019505	-	-
GL018748	1320192	1320303	ENSOCUG00000019704	5S_rRNA	5S ribosomal RNA

Table S3.22 List of GO enriched terms for the set genes in outlier FST windows (99.9% percentile) between Iberian and non-Iberian *L. europaeus*.

Correction Method	Ontology	GO code	P-value	GO description
Benjamini-Hochberg	*BP*			
		GO:0043001	0.05	Golgi to plasma membrane protein transport
	MF			
		GO:0001594	0.03	trace-amine receptor activity
SCS	*MF*			
-		GO:0001594	0.03	trace-amine receptor activity