# Using genomics to dissect the history of divergence and hybridization in the sister broom and Corsican hares (*Lepus* spp.)

João Manuel Meira do Souto

Mestrado em Biodiversidade, Genética e Evolução

Departamento de Biologia Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBIO, Universidade do Porto)

2022

**Orientador**

José Melo-Ferreira, Investigador Principal, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBIO, Universidade do Porto)

**Coorientadores**

Pierre Boursot, Directeur de Recherche, Institut des Sciences de l'Évolution de Montpellier (ISEM, Université de Montpellier)

João Pedro Marques, Investigador de Pós-Doutoramento, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBIO, Universidade do Porto)


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

# Declaração de Honra

Eu, João Manuel Meira do Souto, inscrito no Mestrado em Biodiversidade, Genética e Evolução da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente dissertação de projeto reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta dissertação de projeto, declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente dissertação de projeto quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

João Souto

Porto, 21/11/2022

# Agradecimentos

Este trabalho teve contribuição direta e indireta de várias pessoas às quais quero expressar os meus agradecimentos.

Ao meu orientador, Dr. José Melo-Ferreira, estou eternamente grato por todo o apoio e dedicação. Pelo constante empenho, entusiasmo, disponibilidade e por todos os ensinamentos. Por me incentivares constantemente a fazer mais e melhor. Muito obrigado pela oportunidade, por me confiares este fascinante projeto e por me introduzires ao mundo da Genómica. Não podia ter pedido um melhor orientador!

Ao meu coorientador, Dr. João Pedro Marques, agradeço veemente todo o apoio prestado, empenho e entusiasmo. Desde me ensinares a dar os primeiros passos no mundo da Bioinformática, a me incentivares constantemente a ser crítico dos meus resultados. Obrigado pela tua omnipresença, pelas mil e uma dúvida esclarecidas, e por todos os ensinamentos.

To my co-supervisor Pierre Boursot, thank you for all the help, enthusiasm, support, valuable insights, and constructive criticism provided during this work.

A todos os membros do grupo EVOCHANGE, obrigado por todo o apoio! Pelas constantes reuniões onde pude aprender imenso com os vossos trabalhos e por todas as sugestões e ajudas. Em particular à Inês e Liliana, agraço por toda a ajuda prestada no laboratório e por todos os ensinamentos.

A Fernando Ballesteros, gracias por aportar muestras para este proyecto y por toda la ayuda prestada en el trabajo de campo realizado en Asturias.

Ao Professor Paulo Célio, obrigado por todo o apoio e entusiasmo.

Ao grupo da Invasão a Montpellier, obrigado pela incrível experiência, pelos inúmeras sugestões e conselhos.

Ao grupo das lebres: Sofia, Eugénio e Zé, obrigado por todo o apoio, por estarmos todos no mesmo barco e nos ajudarmos mutuamente. Obrigado por todas as gargalhadas e boa disposição. Zé e Eugénio, obrigado pela incrível companhia na viagem às Astúrias, pelas constantes rizadas e pelos bons momentos, por estarmos uma hora e meia a tentar ver um urso e o máximo que conseguimos ver foi uma pedra.

A todos os amigos agradeço pelo companheirismo e apoio, por serem uma constante fonte de boa disposição e de gargalhadas, e por todas as memórias inesquecíveis. A

vocês António, Bia, Francisco, Inês, Ronne, Gustavo e Zé, obrigado por estarem sempre ao meu lado.

Agradeço também à minha família por todo o apoio incondicional. Por desde sempre acreditarem em mim. Isto não seria possível sem vocês. À minha irmã, mãe, pai e avó, obrigado por todas as brincadeiras e boa disposição, por toda a dedicação, todos os sorrisos e inspiração, por estarem sempre aqui ao meu lado e por serem um exemplo para mim.

# Funding

This work was funded by project HybridChange (PTDC/BIA-EVL/1307/2020) supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia, FCT. Supervisor JMF was supported by FCT (CEEC contract with reference 2021.00150.CEECIND).

# Resumo

A hibridação que leva à troca de variantes genéticas entre espécies (introgressão) é um processo evolutivo ubíquo que pode influenciar fortemente a variação genética das espécies. A caracterização de padrões de introgressão em espécies que hibridam recorrentemente com espécies evolutivamente próximas pode fornecer informações importantes sobre a história das espécies e os processos seletivos e demográficos subjacentes às trocas genéticas. Permite, por exemplo, entender como genomas de entidades divergentes se isolam (especiação), ou o impacto adaptativo de trocas genéticas interespecíficas.

As lebres (género *Lepus*) são um sistema particularmente apropriado para estudar o processo de divergência com fluxo génico e o impacto de hibridação introgressiva ao longo da evolução das espécies. Na Europa, as espécies de lebres passaram por várias mudanças de distribuição geográfica desde o Pleistoceno Superior e por vários períodos de hibridação interespecífica. Notavelmente, no sul da Europa, essas mudanças de distribuição promoveram o contacto entre diferentes linhagens, levando a histórias evolutivas complexas. Neste trabalho investigou-se a história de divergência e hibridação da *L. castroviejoi* (lebre-cantábrica), uma espécie endémica restrita à Cordilheira Cantábrica no norte da Espanha, e a *L. corsicanus* (lebre da Córsega), uma espécie restrita ao sul de Itália e Sicília, e que foi introduzida na Córsega no século XVI. Estudos anteriores sugeriram que estas são espécies irmãs que partilharam um ancestral comum muito recente. Adicionalmente, após a sua separação, poderão ter sofrido eventos de hibridação e introgressão específicos a cada uma das espécies, a escalas temporais que coincidem com hibridação conhecida entre outras espécies de lebre do sul da Europa. Este é assim um modelo apelativo para compreender processos de divergência e subsequente trocas genéticas entre espécies cujo isolamento reprodutivo é ainda incompleto.

Para estudar o processo de divergência entre as espécies irmãs *L. castroviejoi* e *L. corsicanus*, inferir e quantificar eventos de introgressões antigas e recentes, usámos dados de sequenciação de genomas completos de 5 *L. castroviejoi*, 5 *L. corsicanus* e ainda 25 genomas de outras espécies de lebre, potencialmente envolvidas em processos de hibridação e introgressão ou servindo como grupo externo para as análises. Através de análises de estrutura populacional, confirmamos que os genomas de *L. castroviejoi* e *L. corsicanus* são muito semelhantes e têm pouca diferenciação, sugerindo que distribuição alopátrica atual resultou de uma divergência recente.

Modelando o seu processo de divergência, inferimos que essas espécies se separaram durante o Pleistoceno Superior, há cerca de 50 mil anos, a que se seguiu uma redução substancial no tamanho das populações. Apesar da separação recente, *scans* ao longo dos genomas dessas espécies irmãs identificaram picos de divergência, que podem ter resultado de adaptação por alelos alternativos divergentes ou de introgressão de outra fonte, afetando uma mas não a outra espécie. Dada a proximidade genética entre *L. castroviejoi* e *L. corsicanus*, cada uma das espécies pode ser usada como uma representação da população parental da outra para inferir fenómenos evolutivos que ocorreram após sua separação e afetaram apenas uma das espécies. Com foco em *L. castroviejoi*, análises de partilha de variação genética mostram que a espécie possui sinais significativos de introgressão das espécies vizinhas *L. granatensis* (lebre ibérica), *L. europaeus* (lebre europeia) e também de *L. timidus* (lebre da montanha). A história de hibridação foi aprofundada utilizando inferência de ancestralidade local ao longo do genoma. Esta análise sugeriu que cerca de 1% do genoma de *L. castroviejoi* foi afetado por eventos de introgressão, sendo *L. granatensis* o principal contribuinte da sua variação introgredida (0.637%). De facto, examinando a divergência entre *L. castroviejoi* e *L. granatensis* ao longo do genoma, descobrimos que a maioria dos picos de divergência entre *L. castroviejoi* e *L. corsicanus* resulta muito provavelmente de segmentos introgredidos de *L. granatensis* para *L. castroviejoi*. Dado que variantes genéticas introgredidas em *L. castroviejoi* são provenientes de várias espécies, que também se sabe terem hibridado entre si, a inferência de uma certa ancestralidade genética no genoma de *L. castroviejoi* pode resultar de evento de hibridação com uma espécie introgredida (introgressão secundária) e não devido ao contato direto entre as espécies. Analisando as junções de ancestralidade mista nos segmentos genómicos de *L. castroviejoi*, mostrámos que, de facto, parte das contribuições vêm de introgressão secundária, embora não se possa excluir que tenha havido introgressão resultante do contato direto com as três espécies. Para entender se a introgressão de alta frequência de *L. granatensis* poderia afetar genes com um papel funcional relevante, podendo indicar um potencial impacto adaptativo da introgressão, inspecionamos o conteúdo génico dessas regiões genómicas. Encontrámos 247 genes afetados por introgressão de alta frequência (>50%). No entanto, não foi encontrado enriquecimento funcional neste conjunto de genes. Apesar disso, vários genes relacionados com o metabolismo celular foram identificados, os quais merecem investigações futuras para se entender se a introgressão poderá ter sido promovida por adaptação relacionada com as funções dos genes afetados.

Em resumo, este trabalho forneceu pela primeira vez uma caracterização completa da história de divergência entre as espécies irmãs *L. castroviejoi* e *L. corsicanus*, e mostrou que os eventos de hibridação introgressiva que afetaram outras espécies na Península Ibérica após o último máximo glacial impactaram também o património genético de *L. castroviejoi*. Isto permitiu uma melhor compreensão dos processos biogeográficos e evolutivos que guiaram a evolução reticulada das lebres do sul da Europa. Finalmente, este trabalho serve de base a investigações futura dos processos evolutivos que moldaram a mistura genética de *L. castroviejoi*, em particular o equilíbrio entre os processos demográficos e seletivos promotores ou impeditivos de introgressão.

Palavras-chave: Genómica evolutiva, Introgressão, Divergência genética, Lagomorfos, *Lepus castroviejoi*, *Lepus corsicanus*

# Abstract

Hybridization leading to the exchange of genetic variants between species (introgression) is a ubiquitous evolutionary process that may strongly influence the genetic variation of extant species. Characterizing patterns of introgression in species that have recurrently hybridized with neighbour close relatives can provide important insights into the evolutionary history of species and the selective and demographic processes underlying the genetic exchanges. It allows, for example, understanding how genomes from diverging entities become isolated (speciation), or the adaptive impact of interspecific genetic exchanges.

Hares (genus *Lepus*) have emerged as a particularly appropriate system to study the process of divergence with gene flow, and the impact of introgressive hybridization in the long-term course of species evolution. In Europe, hares have experienced several range revolutions since Late Pleistocene and interspecific hybridization at different times and scales. Notably, in Southern Europe these range revolutions promoted the contact between different lineages, leading to complex species' evolutionary histories. In this work, we investigated the history of divergence and genetic admixture of *L. castroviejoi* (the broom hare), an endemic species restricted to the Cantabrian Mountains in Northern Spain, and *L. corsicanus* (the Corsican hare), a species that currently inhabits Southern Italy, Sicily, and that was introduced in Corsica in the 16$^{th}$ century. Previous studies have suggested that these are sister species that shared a very recent common ancestor. In addition, after their split, these species might have undergone independent hybridization events, at timescales that coincide with other hybridization events known to have affected the evolution of neighbouring hare species. This is thus an appealing model to study the process of genetic divergence and subsequent genetic exchanges between species with partial reproductive isolation.

To understand the process of divergence between the sister *L. castroviejoi* and *L. corsicanus*, and infer ancient and recent introgression from different sources we used whole genome sequencing data from *L. castroviejoi* (N=5), *L. corsicanus* (N=5) and from other hare species that may have been involved in hybridization events or that were used as outgroup in the analyses (N=25). Using population structure analyses, we confirm that the genomes of *L. castroviejoi* and *L. corsicanus* are very similar and have shallow differentiation, suggesting that the current allopatric range resulted from a recent split. Modelling their divergence process, we infer that these species split during Late Pleistocene ~50k years ago, which was followed by a substantial decrease in population

size. Despite the recent split, scans of divergence along the genome of these sister species identified peaks of divergence, which could result from local adaptation from alternative diverging alleles, or differential introgression from neighbouring species after the split. In any case, given the close relatedness of *L. castroviejoi* and *L. corsicanus*, each species can be used as a proxy parental population of the other, to infer evolutionary phenomena that occurred after their split and affected one but not the other species. Focusing on *L. castroviejoi*, analyses of shared variation showed that the species holds significant signs of introgression from the neighbouring *L. granatensis* (the Iberian hare) and *L. europaeus* (the European hare) and also from *L. timidus* (the mountain hare). The history of admixture was further investigated using local ancestry inference along the genome, and we estimated that circa 1% of its genome has been affected by introgression events in Iberia, with *L. granatensis* being the major contributor of introgressed variation (0.637%). Indeed, scanning the divergence between *L. castroviejoi* and *L. granatensis* along the genome, we found that most peaks of divergence between *L. castroviejoi* and *L. corsicanus* likely result from introgression from *L. granatensis* to the former. Given that introgression in *L. castroviejoi* comes from multiple source species, which are also known to have admixed with each other, inference of a certain genomic ancestry in the *L. castroviejoi* genome could eventually result from introgression from an already admixed species ("second-hand" introgression) and not from direct contact. Analysing the junctions of mixed ancestry in the genomic tracts of *L. castroviejoi* we show that indeed part of the contributions come from second-hand introgression, though we cannot at this point exclude that introgression also resulted from direct contact with any of the three other hare species. To understand whether high frequency introgression from *L. granatensis* could affect genes collectively with a certain functional role, which could indicate a potential adaptive impact of introgression, we inspected the genic content of such genomic regions. We found 247 genes affected by high frequency introgression (>50%), but no enriched functions were detected in this set. Nevertheless, several genes related to the cell metabolism were identified, which deserve future investigation to understand whether some functional relevance has governed these introgressions.

Altogether, this work provided for the first time a thorough characterization of the history of divergence of the sister *L. castroviejoi* and *L. corsicanus*, and showed that the introgressive hybridization events known to have affected other species in the Iberian Peninsula after the last glacial maximum also impacted the gene pool of *L. castroviejoi*. This adds yet another piece of knowledge to the understanding of the biogeographic and evolutionary processes governing the reticulated evolution of south European hares.

Further, this work sets the ground to investigate the evolutionary processes driving the genetic admixture of *L. castroviejoi*, in particular the balance between demographic and selective processes causing or preventing introgression.

Keywords: Evolutionary Genomics, Introgression, Genetic divergence, Lagomorphs, *Lepus castroviejoi*, *Lepus corsicanus*

# INDEX

# List of Figures

# List of Tables

# List of Appendixes

# Introduction

## 1.1. Speciation, genetic divergence, and gene flow

The ability to analyse complete genomes of natural populations has emphasised the complexity of the processes of evolutionary divergence and exchange of genetic information among diverging evolutionary entities (Sousa & Hey, 2013). Closely related species share large amounts of genetic variation due to common ancestry, in which the shared alleles derived from a recent common ancestor. Nonetheless, shared alleles can also result from gene flow occurring during or at some point in the divergence process (Pinho & Hey, 2010). The genetic exchange between diverging lineages can be current and/or ancient, having a layered effect on the diversification of groups of organisms (Abbott et al., 2013). This allows expanding traditional views of biodiversity as inventories of species with more or less *ad hoc* defined criteria, to the study of what keeps diverging entities apart (and why) – which can be interpreted as speciation, the study of species formation – and how entities we call species are still able to exchange genetic variation to some degree (and why) – what is called introgression. A complete quantification of biodiversity requires understanding both phenomena, and the evolutionary processes that underlie divergence and gene flow.

Which principles should be used to delimit species (the species problem) is one of the most controversial topics in Biology (De Queiroz, 2007; Zachos, 2016). And this problem is not new: "Nor shall I here discuss the various definitions which have been given of the term species. No one definition has as yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species." (Darwin, 1859). Although several concepts have been proposed, perhaps the most widely accepted definition is the biological species concept (BSC) (Dobzhansky, 1937; Mayr, 1942) in which species are defined as “groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups” (Mayr, 1942). Nevertheless, this concept is impractical in allopatric populations and reproductive isolation does not apply to organisms that reproduce asexually (Cronquist, 1978). As such, more technical definitions have been developed (Aldhebiani, 2018; Noor, 2002). A particularly important one was the evolutionary species concept (ESC), proposed with the aim to better delimit biodiversity in the scope of the evolutionary process (Simpson, 1951). The ESC defines a species as “a single lineage of ancestor-descendant populations of organisms which maintains its identity from other such lineages [in space and time] and which has its own

evolutionary tendencies and historical fate" (Wiley, 1981). According to these different views, speciation can be seen as a process where populations diverge and accumulate enough unique characteristics, which can lead to distinct reproductive isolated lineages (De Queiroz, 1998; Wiens, 2004).

Speciation is often classified according to the geographical context, referring to three main modes of speciation: allopatric, parapatric, and sympatric (Coyne & Orr, 2004). Allopatric is often considered the most common mode of speciation, where an ancestral population splits into two geographically isolated ones, due to the formation of an extrinsic barrier, causing a cessation of gene flow between the populations, independent evolution and eventually leading up to reproductively isolated lineages (Coyne & Orr, 2004; Mayr, 1942). Allopatric speciation can be subdivided into (i) dichopatric speciation (traditional allopatric) (Mayr, 1942), in which the populations split without the occurrence of a bottleneck; and (ii) peripatric speciation, in which the ancestral population is divided by a founder effect and one of the new populations is substantially smaller, and due to genetic drift and bottleneck effect, this population acquires new genetic characteristics leading to reproductive isolation (Carson, 1971; Singh, 2012). In the parapatric speciation, two isolated populations have some contact zones (partial barrier), however gene flow between these populations may lead to low fitness offspring and impede assimilation, leading to isolated lineages (Endler, 1977). Although geographical isolation has been considered a major force driving population divergence (Mayr, 1963; Wang et al., 2013; Worsham et al., 2017), in sympatric speciation, a new species evolves from a surviving ancestral species while in the same dispersal area (Bush, 1975). Allopatric and sympatric speciation can be seen as the ends of a continuum of initial levels of gene flow among diverging populations, with the migration rate (m), the proportion of immigrants in a population, being 0 in allopatry and 0.5 in sympatry, while parapatric speciation comprises intermediate values ($0 < m < 0.5$) (Gavrilets, 2004).

Speciation is usually a long process that involves several generations, and thus during this process, the spatial context of these populations can change (Coyne & Orr, 2004). During its evolutionary trajectory, a lineage can experience distinct modes of speciation (Breusing et al., 2020), and different gene exchange rates, with periods in which there is gene flow between populations and others when it is interrupted (Butlin et al., 2008; Cristescu et al., 2012).

Genetic divergence is a mechanism in which two populations can accumulate independent mutations and drift processes through time. Ultimately, the establishment of barriers to gene flow promotes divergence between populations, as drift and mutation

will act independently in each lineage, and, if enough genetic incompatibilities are accumulated, can lead to reproductive isolation (Ferguson, 2002; Kozak et al., 2011). Genetic drift and mutations have a major role in shaping the genetic divergence of populations since mutations generate new variation and genetic drift acts upon it, determining whether a new allele disappears or increases in frequency to fixation (Klopfstein et al., 2006; Millstein, 2016). Over time, genetic drift and mutations will distance the genic pools of diverging lineages by allowing the accumulation of non-adaptive mutations that can facilitate their isolation (Lynch & Walsh, 2007; Yi, 2006). Additionally, evolutionary events such as bottlenecks and founder effects can also promote genetic differentiation, since the new population only has a portion of the ancestral gene pool (Barton & Charlesworth, 1984; Templeton, 2008). Furthermore, genetic divergence can also be induced by natural selection, where genomic segments are favoured and accumulated over time leading to a gene pool distinct from the ancestral one (Ramírez-Valiente et al., 2018; Schreiber & Pfenninger, 2021; Via, 2009).

Moreover, biogeographic dynamics play a major role in shaping the genetic divergence and differentiation of populations, since environmental changes may lead to range shifts and migrations, causing long-term fragmentation of populations, and leading to their isolation (André et al., 2016). These dynamics can also lead to the creation of new peripatric and sympatric zones, between previously independent lineages, and if reproductive isolation is incomplete, to the occurrence of genetic exchanges (i.e. introgression) (Gosden et al., 2011; Quilodrán et al., 2019; Weir & Price, 2011). Additionally, environmental changes can also lead to modifications in the ecosystems and the possible creation of new ecological niches (Laland et al., 2016), inducing new selective pressures which favour the accumulation of adaptive alleles leading to an increase in fitness of those populations and to the divergence from the other lineages (Aggeli et al., 2021; Olson-Manning et al., 2012).

Altogether, the continuous accumulation of new genetic variation may lead to reproductive isolation. Reproductive isolation is often highly polygenic, depending on multiple loci, commonly called “speciation genes”, that underlie different fitness values in the diverging lineages (Dion-Côté & Barbash, 2017; Presgraves, 2010; Wu & Ting, 2004). It is a complex mechanism, and studying patterns of divergence and allele sharing along the genomes of close related diverging lineages can provide insights into the processes that govern the establishment of new species (Kenney & Sweigart, 2016; Muirhead & Presgraves, 2016; Schield et al., 2017).

## 1.2. The impact of introgression on the evolution of species

The modern analysis of whole-genomes data allowed the frequent detection of genetic exchanges between divergent lineages and confirmed a semipermeable view of speciation where introgression is more common than traditionally expected (Harrison & Larson, 2014). These advances have reinforced that the view of evolution as a strictly bifurcating phylogeny was often biased and over simplistic, and shifted the perception of evolution into a more network-like model, in which evolutionary lineages are not completely independent and undergo reticulated relationships (Hallström & Janke, 2010; Mallet et al., 2016). Such scientific progress acknowledged the relevance of introgression in the evolutionary trajectories of species (Aguillon et al., 2022; Suvorov et al., 2022; Wu, 2001).

Traditionally, natural hybridization was considered evolutionarily unimportant (Coyne & Orr, 2004; Darwin, 1859; Fisher, 1930; Mayr, 1963). Indeed, according to the biological concept of species, species are reproductively isolated units, thus interspecific hybridization would be a rare phenomenon as the hybrid offspring would have lower fitness than the parents and backcrossing would lead to less viable genotypes (Coyne & Orr, 2004; Mayr, 1963). However, in the 1930 and 40s, experimental studies in plants evinced that interspecific hybridizations in the botanical field were a common phenomenon (Anderson, 1948; Anderson & Hubricht, 1938; Heiser, 1949). Nevertheless, hybridization in the animal kingdom was still considered rare and its role in animal evolution continued to be neglected (Mayr, 1963).

The advent of genetics allowed understanding that introgressive hybridization is actually a common phenomenon in animals (Arnold, 1997; Seehausen, 2004). However, the evidence was generally based on a limited sampling of the genome, which hampered precise quantifications of genetic exchanges, and hence had limited power in linking evolutionary processes (such as demographic processes and ecological adaptation) with the introgression patterns detected (Allendorf et al., 2010; Twyford & Ennos, 2012). The development of next-generation sequencing (NGS) enabled full genome analyses which set up the emergence of numerous studies detecting patterns of introgression across the genome of a wide taxonomic range (Edelman et al., 2019; Ferreira et al., 2021; Jones et al., 2018; Kamenetzky et al., 2010; Miao et al., 2017; Neafsey et al., 2010; Sheppard et al., 2011; Williams et al., 2020). Indeed, recent research has shown the prevalence of ancient introgression in the genomes of extant species (Moran et al., 2021). For instance, in humans, around 2 to 5% of the genome of some non-African populations are derived from ancient hybridization with Neanderthals and Denisovans (Sankararaman et al.,

2016), and in other taxa, such as swordtail fishes, Italian sparrows, *Heliconius* butterflies, cichlid fishes and sunflowers, over 10% of their genome was derived from ancient admixture (Cui et al., 2013; Hermansen et al., 2011; Martin et al., 2013; Meier et al., 2017). Therefore, these new findings have completed the change of perspective about hybridizations and their role in species evolution, and nowadays introgression is de facto recognized as a common phenomenon and potentially driving force in evolution and adaptation for a wide range of taxa (Adavoudi & Pilot, 2021; Goulet et al., 2017; Moran et al., 2021; Taylor & Larson, 2019).

Introgression does not occur evenly along the genome, and inferring the reasons behind some genomic regions being more permeable to foreign DNA than others is crucial to understand the impacts of introgression in the genomes (de Lafontaine et al., 2015; Harrison & Larson, 2014). Moran et al. (2021) described three emerging principles of hybridization to explain the variance of ancestry in admixed genomes in a background selection against hybridization. A combination of rapid and slower removal of foreign ancestry is expected to stabilize admixed genomes (principle 1). When an admixture event takes place, a quick removal of deleterious ancestry DNA is predicted to occur, leading to a variance in ancestry across the genome (Matute et al., 2020; Schumer et al., 2018), and after this initial fast purging period, populations enter in a slower stage of purging, where selection on individual hybridization-derived haplotypes only subtly shifts genome-wide ancestry proportions (Moran et al., 2021). Additionally, the permeability to foreign alleles also varies across the genome, with some functionally important regions displaying very low proportions of introgression (principle 2) (Wu, 2001). Moreover, the recombination landscape plays a key role in genome stabilization (principle 3), since in regions with low recombination rates, the introgression tracts are longer and more likely to harbour deleterious alleles in the hybrids. Thus, studies have inferred that even when the admixture proportions stabilized along the genome, minor parent ancestry tends to be less prevalent in genomic regions with low recombination rates (Nachman & Payseur, 2012; Seixas et al., 2018; Wu, 2001). Overall, introgressed genomes are a mosaic of ancestry proportions, with regions where minor parent contribution is slight or non-existent and regions where such contribution is substantially more prevalent (Duranton & Pool, 2022; Moran et al., 2021).

Hybridization-derived haplotypes can either be deleterious, neutral, or adaptive, and the prevalence of introgressed segments in a genome can be driven by neutral processes or/and natural selection, or due to selfish genetic elements. Neutral introgression is dependent on genetic drift and demographic processes, whereas adaptive introgression

is shaped by natural selection (Arnold & Martin, 2009; Teixeira & Huber, 2021). Adaptive introgression happens when haplotypes derived from hybridization confer a fitness increase in the gene pool of the recipient population (Burgarella et al., 2019; Moran et al., 2021). While neutral alleles are often lost due to genetic drift, adaptive variants can be maintained by selection and may even reach fixation (Burgarella et al., 2019). The probability of a beneficial allele being introgressed depends on the genomic proximity to a potential deleterious variant and the recombination rate, since it can unlink adaptive haplotypes from harmful variants (Veller et al., 2019). Additionally, the fixation of introgressed haplotypes can also be non-adaptive if involving selfish gene elements (Albrechtova et al., 2012; Crespi & Nosil, 2013), which are genomic segments that have a replication or transmission advantage relative to other genetic elements, but are either neutral or prejudicial to the organism's fitness and reproduction (Werren et al., 1988; Werren & Stouthamer, 2003).

Overall, introgression can be a potential source of allelic novelty, and these introduced new variants distributed genome-wide can provide adaptation even for polygenic phenotypes (Mallet, 2007; Martin et al., 2013) and thus the exchanged genetic variants during hybridization can induce rapid species evolution (Arnold & Kunte, 2017; Baskett & Gomulkiewicz, 2011; Parepa et al., 2014). There have been several cases described of species that have obtained adaptive traits due to introgression, such as abiotic tolerance in sunflowers (Whitney et al., 2010), seasonal coat colour in hares (Giska et al., 2019; Jones et al., 2018), mimicry in *Heliconius* butterflies (Pardo-Diaz et al., 2012), and high latitude adaptation in aspens (Rendón-Anaya et al., 2021). Therefore, the quantification and characterization of introgression events are important to comprehend the mechanisms underlying the persistence of allospecific variants, and consequently better understand the historical dynamics of closely related lineages.

## 1.3. Hares (*Lepus* spp.) as model systems to study drivers of species differentiation and admixture

Hares and jackrabbits (genus *Lepus*) have diverged from the rabbits around 12 MYA, where likely originated in North America and have spread and radiated across Afro-Eurasia most likely in the last 4 – 6 million years (Ferreira et al., 2021; Matthee et al., 2004; Melo-Ferreira et al., 2012; Yamada et al., 2002). Genus *Lepus* consists of 32 extant species native to Africa, Eurasia, and North America and were introduced in

Australia and South America (Smith et al., 2018). Despite being usually associated with grassland habitats, hares can be found in diverse biomes, from savannahs (Abyssinian hare, *L. habessinicus*) to forests (European hare, *L. europaeus*), from wetlands (Indian hare, *L. nigricollis*) to mountains (broom hare, *L. castroviejoi*), from deserts (cape hare, *L. capensis*) to artic biomes (Arctic hare, *L. articus*) (Smith et al., 2018). Hare species have undergone several range shifts, usually due to climatic changes, which induced the contact between closely related species (Ferreira et al., 2021; Melo-Ferreira et al., 2012). As a result, interspecific hybridization is recurrent in hares and various introgression events have been described (Ferreira et al., 2021; Liu et al., 2011). Therefore, hares have emerged as a particularly appropriate system to study the process of divergence with gene flow, and in particular the selective impact of introgressive hybridization in the long-term course of the evolution of species (La Morgia & Venturino, 2017; Seixas et al., 2018).

Interspecific hybridizations in hares often led to the unidirectional introgression of mtDNA of the mountain hare (*L. timidus* Linnaeus, 1758) into other lineages (Alves et al., 2003; Melo-Ferreira et al., 2009; Melo-Ferreira et al., 2007; Yamada et al., 2002). *L. timidus* mtDNA haplotypes are present in more than 10 species distributed both in the New and Old world, including in the geographical distributions (e.g. Iberian and Balkans) where *L. timidus* is currently absent but was present until it went locally extinct at the end of the last glacial period (Alves et al., 2003; Alves et al., 2008a; Melo-Ferreira et al., 2012; Smith et al., 2017).

Several lines of evidence have inferred that hare species in Europe experienced various range revolutions during the last glacial period, and Southern Europe acted as a glacial refugium for some of the species (Acevedo et al., 2015; Lado et al., 2018; Melo-Ferreira et al., 2012; Randi, 2007). Currently, three hare species inhabit the Iberian Peninsula: the broom hare (*L. castroviejoi* Palacios, 1976), the Iberian hare (*L. granatensis* Rosenhauer, 1856), and the European hare (*L. europaeus* Pallas, 1778) (Smith et al., 2018) (Figure 1). Nonetheless, genetic studies, fossil records and ecological niche modelling projections (Acevedo et al., 2015; Lado et al., 2018; Lopez-Martinez, 2008), inferred that at the Late Glacial Maximum the mountain hare (*L. timidus*) was also present in the Northern half of the Iberian Peninsula, while *L. granatensis* was in a refugium in Southwest Iberia (Marques et al., 2017), whereas *L. europaeus* was in a refugium in the Balkans (Stamatis et al., 2009). During the Last Glacial Period, *L. granatensis* presumably expanded north, being favoured by climate change, replacing *L. timidus* in the Northern region of Iberia, which may have contributed to the extinction

of *L. timidus* from the Peninsula (Marques et al., 2017; Melo-Ferreira et al., 2011). Throughout this range replacement event, *L. granatensis* captured *L. timidus* mtDNA as well as some nuclear DNA (nDNA) segments, which witnesses these ancient hybridization events between these species (Melo-Ferreira et al., 2009; Melo-Ferreira et al., 2011; Seixas et al., 2018). Further, it is thought that *L. europaeus* expanded from their Balkan refugium, colonized Central Europe, where contacted with *L. timidus* and exchanged genes, and carried those fragments into Iberia, replacing *L. timidus*, capturing their mtDNA, and colonized Eastern Iberia (Melo-Ferreira et al., 2009; Seixas, 2017). Consequently, in the Iberian Peninsula, a contact zone between *L. europaeus* and *L. granatensis* was formed, resulting in bidirectional introgression (Seixas, 2017). The current Iberian *L. europaeus* population have *L. timidus* mtDNA, but it is yet unclear if it resulted from direct contact with *L. timidus* before invading the Iberian Peninsula, or after by hybridization with *L. granatensis* individuals carrying *L. timidus* mtDNA haplotypes (Melo-Ferreira et al., 2014a; Seixas, 2017; Seixas et al., 2018). Although it has been proposed that *L. timidus* mtDNA haplotypes could have provided environmental adaptive advantages (Melo-Ferreira et al., 2014b; Melo-Ferreira et al., 2007), Seixas et al. (2018) have inferred that both nuclear and mtDNA *L. timidus* introgression patterns in *L. granatensis* can be explained by a range replacement demographic model. The selective advantage of mtDNA introgression in hares is thus yet unclear.

## 1.4. Evolutionary history of the sister broom (*L. castroviejoi*) and Corsican (*L. corsicanus*) hares

The broom hare (*Lepus castroviejoi*) is an endemic species restricted to the Cantabrian Mountains in Northern Spain, and studies have shown it is morphologically and genetically closely related to the Italian hare (*L. corsicanus* De Winton, 1898), a species native to the Apennines and Sicily that was introduced in Corsica in the sixteenth century (Alves et al., 2008a; Mitchell-Jones et al., 1999; Palacios, 1996) (Figure 1). In addition to their genetic similarities, these sister species share ecological niches in their allopatric ranges, which could eventually be used to suggest a conspecific status (Acevedo et al., 2014; Alves & Melo-Ferreira, 2007).

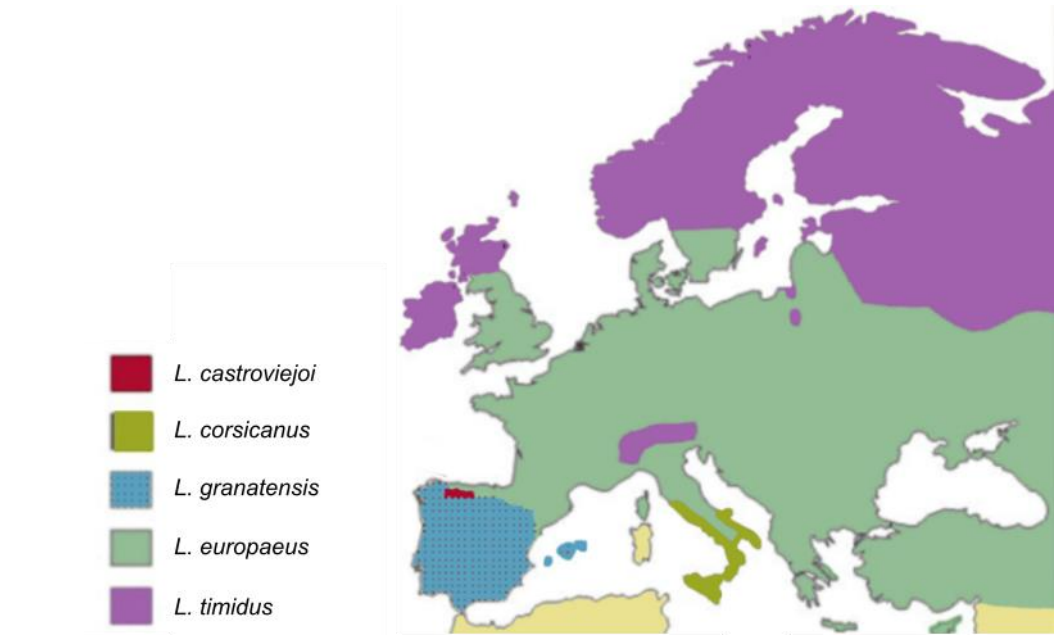



Figure 1 - Geographical distribution of the European hare species according to.Mitchell-Jones et al. (1999).

It has been hypothesized that the common ancestor of *L. castroviejoi* and *L. corsicanus* was more widely distributed in Europe during the Pleistocene (Alves et al., 2008a; Melo-Ferreira et al., 2012). Phylogenetic analyses have suggested that the two species may have split around 120 thousand years ago (Ferreira et al., 2021), likely due to the subsequent climatic changes during the last glacial period, into two allopatric refugia, one in the Iberian Peninsula and another in the Italian Peninsula (Ferreira et al., 2021; Melo-Ferreira et al., 2012). Yet, these studies were based on a limited set of markers and samples, which may limit precise inferences of divergence times. Nevertheless, is still unknown if the divergence of these species was induced by fragmentation, natural selection and environmental adaptation (Alves et al., 2008a). Additionally, current *L. castroviejoi* and *L. corsicanus* populations harbour mtDNA haplotypes that are closely related to *L. timidus*, which discords from phylogenetic inferences based on nuclear DNA that show that the species are not closely related (Alves et al., 2008a; Ferreira et al., 2021; Melo-Ferreira et al., 2012). This mtDNA proximity to *L. timidus* has been shown to be compatible with ancient introgression, and the native haplotypes have likely disappeared from the *L. castroviejoi* and *L. corsicanus* gene pools (Melo-Ferreira et al., 2012; Mengoni et al., 2015; Pietri et al., 2011). Interestingly, *L. castroviejoi* has presumably undergone two introgression events, each one representing a different mtDNA lineage from the mountain hare type: one shared with *L. corsicanus* and thus most likely resulting from introgression during Mid Pleistocene, affecting their common ancestor; and another shared with *L. granatensis* and Iberian *L. europaeus*, which thus

must represent a more recent hybridization event affecting only *L. castroviejoi* (Alves et al., 2008a; Melo-Ferreira et al., 2012). However, it is yet unclear if the presence of this second mtDNA lineage in *L. castroviejoi* results from direct hybridization with *L. timidus* or one of the neighbouring species that carried those mtDNA haplotypes (Melo-Ferreira et al., 2012). Also, it is unknown what is the extent of the nuclear genome affected by these past hybridization events, nor the demographic and selective processes underlying the genetic exchanges (Seixas et al., 2018).

Furthermore, in Corsica, *L. corsicanus* contacts with the also introduced *L. granatensis* and *L. europaeus* (Buglione et al., 2018). Pietri et al. (2011) inferred the genetic diversity of hares in Corsica, and when comparing the mtDNA control region haplotypes with the transferrin nuclear genes, detected *L. corsicanus* x *L. europaeus* hybrids, as well as one *L. corsicanus x L. granatensis* hybrid. Thus, at least in Corsica, *L. corsicanus* has probably hybridized with *L. europaeus*, as well with introduced *L. granatensis*. Moreover, *L. corsicanus* and *L. europaeus* have a contact zone in the Italian peninsula, and although no hybrids were detected until now, the possibility of ancient introgression between these species needs to be properly investigated (Mengoni et al., 2015).

## 1.5. Objectives

Understanding the evolutionary mechanisms of divergence, and how and why species can continue to exchange genetic variation during this process is crucial in evolutionary biology. Thus, studying models where closely related species recently diverged and are still able to hybridize are valuable to comprehend both the process of divergence and introgression.

In this work, we aimed to dissect the process of divergence between *L. castroviejoi* and *L. corsicanus*, as well as the impact and evolutionary processes underlying genetic exchange in the system.

Specifically, we aimed at:

i) Reconstruct the demographic history of divergence between *L. castroviejoi* and *L. corsicanus*.
ii) Infer the impact of introgression in the *L. castroviejoi* and *L. corsicanus* genomes

iii) Clarify the history of genetic exchanges affecting *L. castroviejoi* and *L. corsicanus* after their split. Infer if the genetic contribution was due to direct or secondary introgression.

iv) Incorporate *L. castroviejoi* evolutionary events into the biogeographic history of hares in Iberia.

# 2. Methods

## 2.1. Dataset

### 2.1.1. Sampling and Sequencing

The dataset of this work was composed of 35 individuals from six hare species: *L. castroviejoi* (n=5), *L. corsicanus* (n=5), *L granatensis* (n=10), *L. europaeus* (n=10), *L. timidus* (n=4), *L. americanus* (n=4). Individuals were originally collected at different points of the distribution range of their species, and samples were part of the CIBIO-InBIO biobank (Table 1). We generated new whole genome sequencing data for the *L. castroviejoi* and *L. corsicanus* samples, while for the other hare species we relied on whole genome sequencing data from previous studies (Carneiro et al., 2014; Giska et al., 2019; Seixas, 2017; Seixas et al., 2018). Genomic DNA was extracted using JETquick Tissue DNA Spin Kit (GENOMED) from ear or internal organ tissues that had been preserved in ethanol or RNAlater. Illumina TruSeq DNA v2 genomic libraries were 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) and using inserts of 550 bp for Lcas1 and Lcor1 samples, and inserts of 350bp for the remaining *L. castroviejoi* and *L. corsicanus* samples.

Table 1 - Whole genome Dataset information. (1) Giska et al. (2019); (2) Seixas et al. (2018); (3) Seixas (2017); (4) Carneiro et al. (2014).

| ID | Species | Location | Mitochondrial Lineage | Tissue | Sex | Reference |
|---|---|---|---|---|---|---|
| **Lcas1** | *L. castroviejoi* | Cantabria, Spain | introgressed | ear | F | this work |
| **Lcas2** | *L. castroviejoi* | Alto Sil, León, Spain | introgressed | ear | M | (1) |
| **Lcas3** | *L. castroviejoi* | León, Spain | introgressed | organ | F | this work |
| **Lcas4** | *L. castroviejoi* | Riano, León, Spain | introgressed | ear | F | this work |
| **Lcas5** | *L. castroviejoi* | Cantabria, Spain | introgressed | ear | F | this work |
| **Lcor1** | *L. corsicanus* | Corsica, France | introgressed | organ | F | this work |
| **Lcor2** | *L. corsicanus* | Corsica, France | introgressed | organ | F | this work |
| **Lcor3** | *L. corsicanus* | Corsica, France | introgressed | organ | F | this work |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Lcor4** | *L. corsicanus* | Corsica, France | introgressed | organ | M | this work |
| **Lcor5** | *L. corsicanus* | Corsica, France | introgressed | organ | F | this work |
| **Lgra1** | *L. granatensis* | Alcoutim, Portugal | native | ear | F | (2) |
| **Lgra2** | *L. granatensis* | Peñaflor, Sevilla,Spain | native | ear | F | (2) |
| **Lgra3** | *L. granatensis* | Pancas, Portugal | native | organ (KI) | F | (2) |
| **Lgra4** | *L. granatensis* | Idanha, Castelo Branco, Portugal | native | organ | F | (2) |
| **Lgra5** | *L. granatensis* | Miguelturra, Ciudad Real, Spain | native | organ (KI) | F | (2) |
| **Lgra6** | *L. granatensis* | Valpaços, Portugal | introgressed | organ (KI) | F | (2) |
| **Lgra7** | *L. granatensis* | Algete, Madrid, Spain | introgressed | ear | F | (2) |
| **Lgra8** | *L. granatensis* | Província de Valência, Spain | introgressed | ear | F | (2) |
| **Lgra9** | *L. granatensis* | Monte Allá Detrás,Sauguillo, Spain | introgressed | ear | F | (2) |
| **Lgra10** | *L. granatensis* | Fontellas, Navarra, Spain | introgressed | organ (KI) | F | (2) |
| **Leur1** | *L. europaeus* | Cantabria, Spain | introgressed | ear | F | (3) |
| **Leur2** | *L. europaeus* | Jaca, Spain | introgressed | ear | F | (3) |
| **Leur3** | *L. europaeus* | Villarcayo, Spain | introgressed | ear | F | (3) |
| **Leur4** | *L. europaeus* | Álava, Spain | introgressed | organ | F | (3) |
| **Leur5** | *L. europaeus* | Navarra, Spain | introgressed | organ | F | (3) |
| **Leur6** | *L. europaeus* | Pyrenees, France | native | ear | F | (3) |
| **Leur7** | *L. europaeus* | Ukraine | native | ear | M | (3) |
| **Leur8** | *L. europaeus* | Germany | native | organ | F | (3) |
| **Leur9** | *L. europaeus* | Vienna, Austria | native | organ | F | (3) |
| **Leur10** | *L. europaeus* | Clermont-Ferrand, France | native | organ | F | (3) |
| **Ltim1** | *L. timidus* | Borris-in-Ossory, Ireland | native | organ (KI) | F | (2) |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Ltim2** | *L. timidus* | Captivity (originally from Finland?) | native | organ (KI) | F | (2) |
| **Ltim3** | *L. timidus* | Calfreisen, Egga, Switzerland | native | organ (KI) | F | (2) |
| **Ltim4** | *L. timidus* | Commune Nancy, sur-Cluses, France | native | ear | F | (2) |
| **Lame** | *L. americanus* | Near Lake Inez, Montana – USA | native | organ (HE) | F | (4) |

Abbreviations: HE – Heart; KI – Kidney.

### 2.1.2. Data treatment

Cutadapt version 1.8 (Martin, 2011) was used to filter raw sequence reads by removing the first 5 bp and adapters at the end of reads. Low quality bases (quality < 20 at the end of reads, and 4 consecutive bp with average quality < 30) were removed using Trimmomatic v0.33 (Bolger et al., 2014). Filtered reads were mapped to a *L. timidus* pseudo-reference derived from the European rabbit (*Oryctolagus cuniculus*) generated in (Marques et al., unpublished work) using the BWA-MEM algorithm with default parameters (Li & Durbin, 2009). Read paring information was corrected and mapped reads sorted by coordinates by using Samtools v1.3 (Li et al., 2009), and the further removal of soft clipped bases was performed on NGSutils version 0.5.7 (Breese & Liu, 2013). A read realignment around INDELs was performed to reduce the number of INDELs miscalls using the Genome Analysis Toolkit (GATK v3.2-2) (DePristo et al., 2011; McKenna et al., 2010). The removal of read duplicates was conducted using Picard Markduplicates (http://broadinstitute.github.io/picard/). Bcftools 1.10.2 mpileup (Li, 2011) was used to perform the Multi-sample SNP/genotype calling for each species independently, adopting minimum base and mapping qualities of 20. VCF files file were then merged, INDELs were removed and repetitive regions from the *Oryctolagus cuniculus* genome were extracted from https://genome-euro.ucsc.edu/ and those regions were excluded from our dataset using Bcftools 1.10.2.

The relatedness among samples in the dataset was assessed using the relatedness2 option implemented in vcftools, which is based on the KING method (Danecek et al., 2011). This statistic was calculated using a bi-allelic subset for each species with SNPs subsampled 25kb apart to avoid linked loci.

## 2.2. Population Structure and Evolutionary History

In order to assess genetic variation in the sister *L. castroviejoi* and *L. corsicanus* in the context of the genetic structure of European hares species, the unsupervised principal component analysis (PCA) was performed in PLINK 2.00 (Chang et al., 2015). PCAs were computed using subsets based on bi-allelic SNPs at least 50 kb apart and present in all samples.

Additionally, structure and possible admixture between hare species was investigated by performing an admixture analysis implemented in ADMIXTURE (Alexander et al., 2009). This method uses a Bayesian Markov Chain Monte Carlo model (MCMC) to estimate the ancestry for each specimen based on a SNPs dataset. We used the pruned dataset based on 3,899,363 SNPs containing all European hare species. One run was performed for each number of clusters (k) set from 2 to 7 with 100 bootstrap replicates and 10 cross-validation. Due to lack of computational power, the Admixture runs for K 6 and 7 were based only on data chromosome 20. The most likely number of clusters (K) was determined by considering the cross-validation errors.

Evolutionary relationships between hare species and migration events were inferred using TreeMix v. 1.13 (Pickrell et al., 2012; Pickrell & Pritchard, 2012). TreeMix uses allele frequency differences to quantify drift between populations and to fit a population tree, and then evaluates whether the fit to the data is improved by adding admixtures events. We estimated allele frequencies of the pruned SNPs dataset and subsequently ran the TreeMix model using bootstrapping and accounting for linkage disequilibrium by grouping sites in blocks of 500 single-nucleotide polymorphisms setting the *Lepus americanus* as the root. The best tree topology was inferred following the maximum likelihood approach, and up to 5 migration events were added. The best number of migrations was evaluated according to the standard error values. The inferred maximum-likelihood trees were visualized with the in-built TreeMix R script plotting functions.

## 2.3. Demographic Profiles and History of Divergence

The demographic profiles of *L. castroviejoi* and *L. corsicanus* were reconstructed using a Pairwise Sequentially Markovian Coalescent (PSMC) model (Li & Durbin, 2011). This method examines the variations of heterozygous site densities along the genome to infer the distribution of the most recent common ancestors among genomic regions, from

which it can infer past demography, since the density of coalescence at a given time is inversely proportional to effective population size at that time (Hudson, 1990; Nadachowska-Brzyska et al., 2016). For this analysis, we used one specimen of *L. castroviejoi* and one of *L. corsicanus,* selecting the ones with the highest coverage. Samtools v1.3.1 *mpileup* and *call* modules were used to build the diploid consensus sequences, and only sites with coverage between 6X and twice the average depth and a minimum base and mapping qualities of 20 were called (atomic time intervals were set to 4 + 50*2 + 2 + 4 as Seixas et al. (2018)). Results were calibrated using a mutation rate (μ) of $2.8 \times 10^{-9}$ substitutions/site/generation (Seixas et al., 2018) and by setting the generation time to 2 years (Marboutin & Peroux, 1995). The variance of effective population size ($N_e$) estimates was assessed by 50 bootstraps of randomly sampled segments with replacement.

To better understand the history of divergence between *L. castroviejoi* and *L. corsicanus*, we used a Bayesian demographic inference method, G- PhoCS (Gronau et al., 2011), to estimate effective population sizes of current and ancestral populations, time of split and migration rates. We prepared a dataset with *L. castroviejoi* (n=5) and *L. corsicanus* (n=5) consisting of 2,147 intergenic fragments of 1 kb with a distance between fragments of at least 50kb. Three replicates of a model without gene flow were computed and one run of a model allowing post-split bidirectional gene flow was performed. For all runs, 100,000 generations were discarded as burn-in and 1,000,000 MCMC iterations were run, sampling every 10 iterations. The runs of each model were combined and checked with Tracer v1.7.1 (Rambaut et al., 2014) by examining the effective sample size of each parameter. We converted the scaled demographic parameters obtained from G- PhoCS applying theta=$4N_e$*μ, τ=T*μ/g, M=m/μ, where $N_e$ is the effective population size (in numbers of individuals), g is the average generation time in years, T is the absolute population divergence time in years, μ is the mutation rate in substitutions/site/generation, and m is the probability of migration in a single generation. We assumed a mutation rate of $2.8 \times 10^{-9}$ substitutions/site/generation (Seixas et al., 2018) and a generation length of two years (Marboutin & Peroux, 1995).

### 2.3.1. Detection of localized divergence outliers

To identify divergence outliers, we calculated the genetic distance (Dxy) (Nei, 1987) between *L. castroviejoi* and the other hare species in the dataset. These analyses were based on a subset without indels and with genotypes with a minimum of 4x coverage and a maximum of 60x and a minimum genotype coverage of 10x. Python tools created

by Simon Martin (https://github.com/simonhmartin/genomics_general) were used to filter sites with valid genotypes in at least 75% of the samples (--minCalls), and with at least one valid genotype per population (--minPopCalls). Those python tools were also used to calculate the Dxy values along the genome in 25k windows with a minimum of 250 sites per window. Dxy genome-wide scans and density plots were elaborated using the gwscaR R package. As Dxy is dependent on the mutation rates, which may vary along the genome, these Dxy values were used to calculate variations of relative node depth (RND) (Feder et al., 2005), since this approach mitigates the effect of mutation rate oscillations by using the distance to an outgroup. The RND for each window was calculated by scaling the genetic distance between the sister taxa with their distance to the outgroup to detect outliers of divergence between the sister taxa, using this formula:

$$RND = \frac{Dxy\ L.corsicanus\ L.castroviejoi}{((Dxy\ L.americanus\ L.corsicanus + Dxy\ L.americanus\ L.castroviejoi)/2)}$$

## 2.4. Quantification and characterization of introgression

### 2.4.1. Detection of introgression events

Introgression events that only affected *L. castroviejoi* but not *L. corsicanus* were detected and characterized by using one as the proxy-parental population of the other. First, we identified the genome-wide introgression patterns using *D*-statistics (ABBA-BABA tests) (Durand et al., 2011; Green et al., 2010), which compares the number of shared derived alleles between the sister taxa (P1 and P2) and a possible donor species (P3). In a scenario without gene flow, P1-P3 and P2-P3 should share a similar number of derived alleles. On the other hand, in a scenario with introgression, there will be an excess of shared derived alleles in one of the pairs. We set *L. corsicanus* as P1, *L. castroviejoi* as P2, *L. americanus* as the outgroup (P4), and tested three different donor species (P3): *L. granatensis*, *L. timidus*, *L. europaeus*. For each donor species tested, a SNP subset was made with a minimum quality of 10 and minimum coverage of 4x (3x in the model with *L. europaeus* as donor) and a maximum of 60x per SNP. The ABBA-BABA tests were performed using the Dsuite toolkit (Malinsky et al., 2021) and a z-score with an absolute value of 3 or more was considered to be evidence of significant interspecific gene flow. Additionally, $f_3$ and $f_4$ statistics (Pickrell & Pritchard, 2012; Reich et al., 2009) were calculated using threepop and fourpop from the Treemix package. $F_3$-statistics (target, source 1, source 2) infers if the target population was derived from the Admixture

of sources 1 and 2. $F_4$-statistics assumes the population topology (A,B),(C,D) and evaluates correlation in allele frequency differences between pairs of groups, this way detects whether there was gene flow between the different populations.

Furthermore, introgression signs of *L. granatensis* in *L. castroviejoi* and *L. corsicanus* were also evaluated by inferring the variation of topologies along the genome. Thus, the topology weights of a subset consisting of *L. castroviejoi*, *L corsicanus*, *L. granatensis* and *L. americanus* (outgroup) were performed using TWISST (Martin & Van Belleghem, 2017), since topology weighting is a useful tool to quantify relationships between taxa that are not necessarily monophyletic (Ravinet et al., 2020).

### 2.4.2. Detection of localized introgression

After analyzing the general patterns of introgression, we sought to locate the introgressed regions/genes along the genome. First, we used the $f_{dM}$ statistics which is an *f* statistic that was particularly developed for the inference of introgression in small genomic windows (Malinsky et al., 2015; Malinsky et al., 2018). $F_{dM}$ values vary from -1 to 1 and have similar principles to the ABBA-BABA tests: values between 0 and 1 indicate gene flow between P2 and P3, whereas values between 0 and -1 suggest gene flow between P1 and P3. $F_{dM}$ values were computed for 3 subsets (one for each donor species model), with each subset consisting of SNP filtered genotypes with at least 4 and a maximum of 60 coverage and a minimum of 10 genotype quality, and then valid genotypes in at least 75% of the samples (--minCalls), and with at least one valid genotype per population (--minPopCalls) were called using Python tools created by Simon Martin (https://github.com/simonhmartin/genomics_general). These Python tools were used to calculate the $f_{dM}$ values along the genome in 25k size windows with a minimum of 100 SNPs per window.

Moreover, simulations of Dxy distributions under different models were conducted to estimate the distribution of expected values of Dxy between *L.castroviejoi* and *L. granatensis* under a model with no gene flow, and hence set a threshold under which the Dxy is indicative of introgression. First, demographic parameters (ancient and post-split population sizes, and times of divergence) between *L. castroviejoi* and *L granatensis* were inferred using G-Phocs (Gronau et al., 2011), applying the same procedures described above for the modelling of the *L. castroviejoi* and *L. corsicanus* divergence. Then, the parameters inferred from the *L. castroviejoi L. granatensis* demographic model

were used in msms (Ewing & Hermisson, 2010) to simulate 1,000 fragments of 20 kb under two demographic models:

3) the full demographic model, to assess the reliability of the demographic inference to replicate genome-wide empirical data;

ii) inferred demographic model but without inter-species migration, to assess Dxy expectations under a strict lineage sorting model without gene flow, and set a threshold of Dxy indicative of introgression between *L.castroviejoi* and *L. granatensis*.

For the empirical Dxy distribution, we used the python tools created by Simon Martin (https://github.com/simonhmartin/genomics_general) to calculate the Dxy *L. castroviejoi* *L. granatensis* values for the 1kb intergenic fragments used in the modelling divergence analysis.

Furthermore, we also relied on the previously calculated Dxy genetic distances to identify localized introgression. As *L. corsicanus* and *L. castroviejoi* are sister taxa, is expected for them to be genetically equidistant from *L. granatensis*. Windows where the Dxy *L. castroviejoi-L. granatensis* is substantially lower than that between *L. corsicanus-L. granatensis*, are candidates for introgression between *L. granatensis* and *L. castroviejoi*. A z-score test was performed on the ratio $\frac{Dxy\ L.corsicanus\ L.granatensis}{Dxy\ L.castroviejoi\ L.granatensis}$ to identify windows where the genetic distance between the sister taxa and the *L. granatensis* has an outstanding discrepancy, and identify localized introgression segments from *L. granatensis* along the *L. castroviejoi* (or *L. corsicanus*) genome.

### 2.4.3. Frequency of introgression and junctions between ancestry tracts

To further understand the genetic exchanges that happened after *L. castroviejoi* and *L. corsicanus* split, the ancestry of each position across the *L. castroviejoi* genome was inferred using the Efficient Local Ancestry Inference (ELAI) method (Guan, 2014). Unlike the previously applied methods, ELAI is able to identify introgressed tracts per haplotype in unphased data. This method uses a two-layer hidden Markov model (HMM) to analyse linkage-disequilibrium within and among defined groups and without prior definition of window sizes infers local ancestry of admixed individuals. For each variable position in the genome, the most likely proportions of ancestries are estimated, which can vary from 0 to 1 (0 and 1 indicating homozygous ancestry; 0.5 indicating heterozygous ancestry). ELAI was run using a bi-allelic unphased dataset and by setting two models: 1 – *L. castroviejoi* as the target potentially admixed population and *L. corsicanus*, *L.*

*granatensis*, *L. timidus* and *L. europaeus* as sources; 2 – *L. corsicanus* as the target potentially admixed population and *L. castroviejoi*, *L. granatensis*, *L. timidus* and *L. europaeus* as sources (Appendix 2). For each model, three independent runs were performed using 20 Expectation Maximization (EM) steps and by considering the time of split between *L. castroviejoi* and *L. corsicanus* that was estimated in the G- PhoCS analysis (25,000 generations) as the number of admixture generations. Then the results from the 3 independent runs were averaged. The ancestry for each position was assigned considering the ancestry probability values for each of the four possible ancestries (*L. corsicanus*, *L. granatensis*, *L. europaeus*, *L. timidus*): values higher than 0.7 were considered homozygous; between 0.4 and 0.7 as heterozygous ancestries; values below 0.4 were considered as positions that do not originate from the tested ancestor. For each individual, genomic tracts were formed by merging consecutive positions with the same ancestry.

As previous studies have inferred that *L. granatensis* was impacted by introgression events with *L. europaeus* and *L. timidus* (Seixas, 2017; Seixas et al., 2018), we evaluated if the detected *L. timidus* and *L. europaeus* ancestry tracts across the *L. castroviejoi* genome resulted from a direct contact with these species or were the result of second-hand introgression (hybridization with introgressed *L. granatensis*). In a scenario of second-hand introgression, the segments with *L. timidus* or *L. europaeus* ancestry in the *L. castroviejoi* genomes would be flanked by tracts of *L. granatensis* ancestry. We examined junctions between tracts of different ancestries (junctions between two homozygous SNPs were counted twice). Transitions between SNPs more than 1 kb apart were not considered. This analysis was performed for the ancestries of each *L. castroviejoi* sample separately.

Additionally, the estimation of introgression dates was performed using the introgressed tract lengths detected with ELAI results. Assuming the size tracts are a function of time since the introgression event and depending on the recombination rate, the formula 1/rt, where t is the number of generations since the introgression event and r is the recombination rate per base pair (Liang & Nielsen, 2014; Pool & Nielsen, 2009), was applied to approximately date the admixture events. We used estimates of recombination rate in rabbits ($r = 1.0 \times 10\text{-}8$; Chantry-Darmon et al. 2006) and considered a generation length of 2 years (Marboutin & Peroux, 1995).

### 2.4.4. Evaluation of introgression estimates from different methods

As different approaches applied to estimate introgression were based on distinct principles, their results were contrasted to better understand the overlap of inferences based on different methods. We calculated Dxy and $f_{dM}$ values for each of the introgressed tracts detected with ELAI and grouped them by introgression frequency, to assess the summary statistics compare with ancestry tract inference in detecting high frequency introgression. We also verified the ELAI introgression frequency distribution of the z-score Dxy outliers window to assess the overlap between these two approaches.

### 2.4.5. Functional enrichment analysis for genes introgressed at high frequencies

Introgressed genes favoured by selection tend to reach higher haplotype frequencies (Bay et al., 2019). Thus, to assess the potential adaptive role of *L. granatensis* introgression in *L. castroviejoi*, we were interested in analyzing the introgression segments at high frequency. High frequency segments were based on the ELAI estimates and also on the z-score outliers, since this approach revealed effectiveness in detecting high frequency introgression. The coordinates of *L. castroviejoi* segments with at least 0.5 ingression frequency from *L. granatensis* estimated with ELAI were merged with the outlier windows of the z-score Dxy test, to assemble the coordinates of introgressed segments at high frequency. Then we inferred which genes were within or overlapping those regions and performed functional enrichment analyses using g:Profiler (Raudvere et al., 2019) applying the g:SCS multiple test correction. Only genes within or overlapping windows with more than 250 used sites were considered for the background list of genes. We used the rabbit Gene Ontology (GO) database.

# 3. Results

## 3.1. Sequencing data and relatedness analysis

We sequenced the genomes of 4 *L. castroviejoi* and 5 *L. corsicanus* samples (Table 1). The genomes of 1 *L. castroviejoi*, 10 *L. europaeus*, 10 *L. granatensis* and 4 *L. timidus* previously sequenced by (Giska et al., 2019; Seixas, 2017; Seixas et al., 2018) (Table 1), were also included in this study. *L. castroviejoi* raw coverage ranged between 6-15x and the genotype quality between 18-40, whereas *L. corsicanus* samples had 7-18x raw coverage and 20-49 genotype quality values (Appendix 1).

The relatedness among the hare samples in the dataset was assessed and there was no detection of duplicates or 1$^{st}$ degree samples. Samples Lcas3 Lcas4 were classified as 2$^{nd}$ degree relatives (Appendix 3). The remaining samples were identified as unrelated.

## 3.2. Population Structure and Admixture

Genetic variation structure among European hare species was initially assessed using PCA plots. First, a PCA containing all European species (*L. castroviejoi*, *L. corsicanus*, *L. granatensis*, *L. europaeus*, *L. timidus*) and an outgroup (*L. americanus*) based on 3,899,363 SNPs placed each species in a separate cluster, except for *L. corsicanus* and *L. castroviejoi* which were grouped together (Appendix 4A). PC1 (28.99%) separated *L. castroviejoi* and *L. corsicanus* from the other hare species and PC2 (19.08%) segregated the outgroup. Second, a PCA with only European hare species and based on 2,999,121 SNPs assigned each species to a separate cluster, while grouped together *L. corsicanus* and *L. castroviejoi* (Figure 2A), confirming that the sister *L. castroviejoi* and *L. corsicanus* have low genetic differentiation but are genetically different from the other hare species. To further understand the genetic differentiation between *L. castroviejoi* and *L. corsicanus*, a PCA containing only these sister taxa was performed based on 433,825 SNPs (Figure 2B). Most of the genetic variance in this dataset was found between the two species PC1 (50.74%), while PC2 (8.53%) revealed variation within *L. castroviejoi* (Figure 2B) and PC3 (7.03%) within *L. corsicanus* (Appendix 4D). As PC2 separated Lcas3 and Lcas4 from the other *L. castroviejoi* samples, and as this pair scored high levels of relatedness (Appendix 3), a PCA without Lcas4 was conducted to infer whether

the polarized variation detected intra *L. Castroviejoi* specimens was due to high genetic similarity between Lcas3 Lcas4. The PCA without Lcas4 displayed very similar results to the PCA with Lcas4 (Appendix 4E, F).

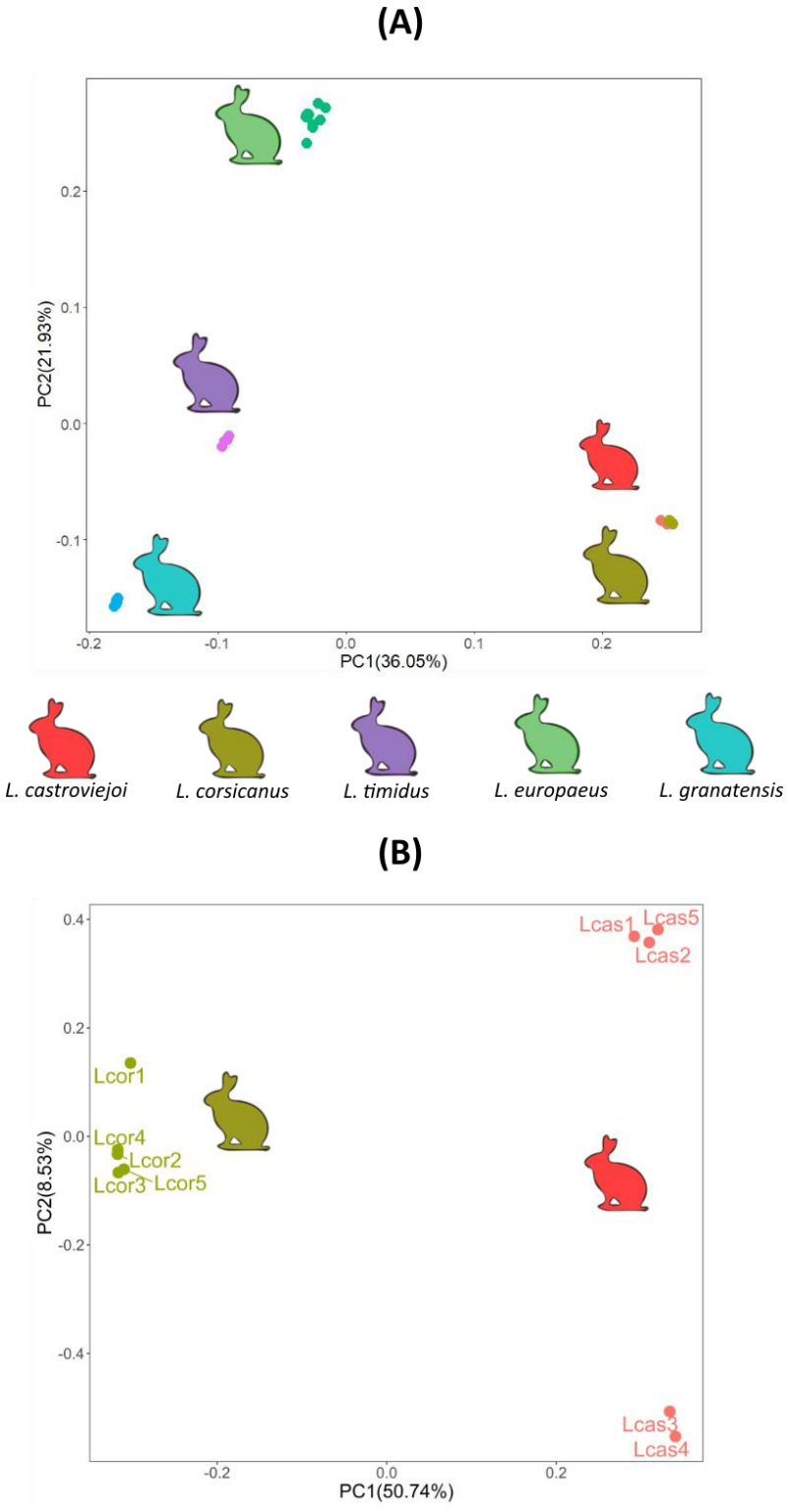



Figure 2 – Principal Component Analysis (PCA) based on SNP data for all European hare species (A) and for *L. castroviejoi* and *L. corsicanus* (B). Only PC1 and PC2 are displayed. Each dot represents a sample.

The admixture analysis for the European species subset (Figure 3) for K=2 grouped *L. castroviejoi* and *L. corsicanus* in one cluster and *L. granatensis* in another, while *L. europaeus* and *L. timidus* had proportions from both clusters. For K=3, *L. corsicanus* and *L. castroviejoi* were assigned to an independent genetic cluster. For the best K according to cross validation assessment (K=4) (Appendix 5) each species was assigned to their own genetic cluster, except for *L. castroviejoi* and *L. corsicanus* which remained grouped together. For the K equal to the number of species (K=5), *L. corsicanus* and *L. castroviejoi* remained in the same genetic cluster, and just split for K=7 (Admixture analysis performed on chromosome 20 – Appendix 6). These results corroborated with the ones obtained in the PCA plots by strongly supporting the genetic similarity between the sister species *L. castroviejoi* and *L. corsicanus*, as well as showing the genetic distinction between these species and the other European hares (Figure 2).

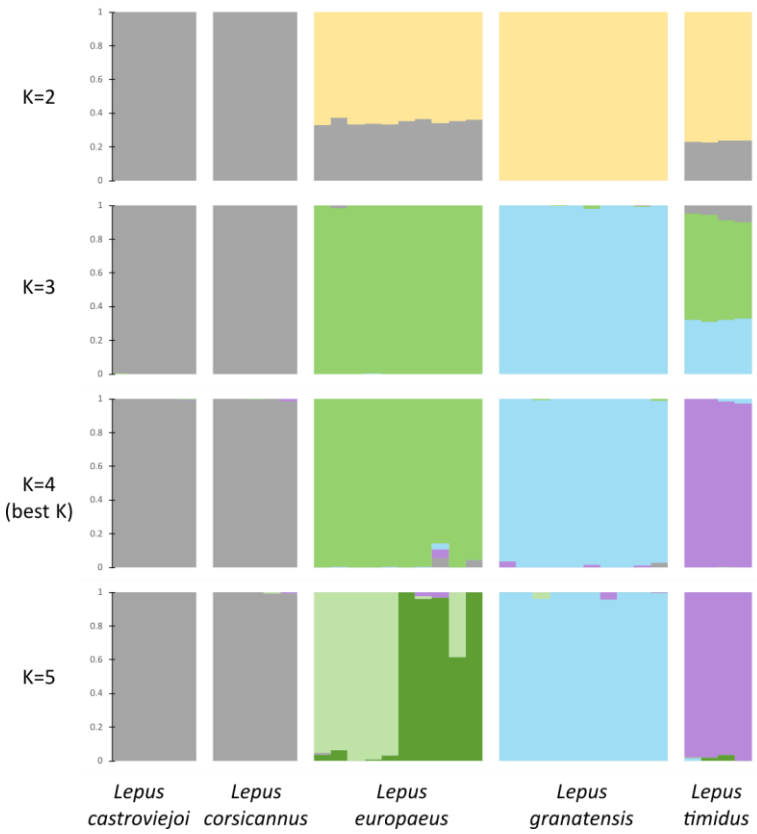



Figure 3 – Admixture analysis for European hare species from K= 2 to K=5, inferred with ADMIXTURE based on ~3 million SNPs filtered by linkage disequilibrium. Each colour represents a distinct genetic cluster.

To further understand the evolutionary relationships among European hare species and whether the inclusion of genetic migration events (introgression) explains better the data,

we performed a species tree using the graph-based model implemented in Treemix, allowing a maximum of five migration evens. The addition of one migration event (Appendix 7B) improves substantially the fit of the model, with the model indicating gene flow between *L. granatensis* and *L. timidus*, which is an introgression event well described in previous studies (Seixas et al., 2018). The addiction of further migration events slightly increased the likelihood of the data, and for m=3 a migration between *L. timidus* and the ancestral of *L. castroviejoi* and *L. corsicanus* was detected, as well as between *L. granatensis* and *L. europaeus*, and between *L. corsicanus* and *L. americanus* (outgroup) (Appendix 7D). The addition of more migration events did not lead to a better model, however it is important to note that for m=4 and m=5 a migration between *L. castroviejoi* and *L. granatensis* was detected (Appendix 7E, F), which can be a sign of an introgression event between these species.

## 3.3. Divergence and Demography

The past population size oscillations of *L. castroviejoi* were inferred using a PSMC model, and there were detected two periods of population growth, 0-40 KYA and 1-2 MYA, and two phases of population decrease, 50-200 KYA and 400-900 KYA (Figure 4). Additionally, a demographic profile of *L. corsicanus* was also inferred, though with a lower resolution. The *L. corsicanus* and *L. castroviejoi* profiles had some discrepancies, with *L. corsicanus* displaying a lower population size, but both species showed similar patterns of population growth and decrease from 700 K to 50 KYA (Appendix 8).

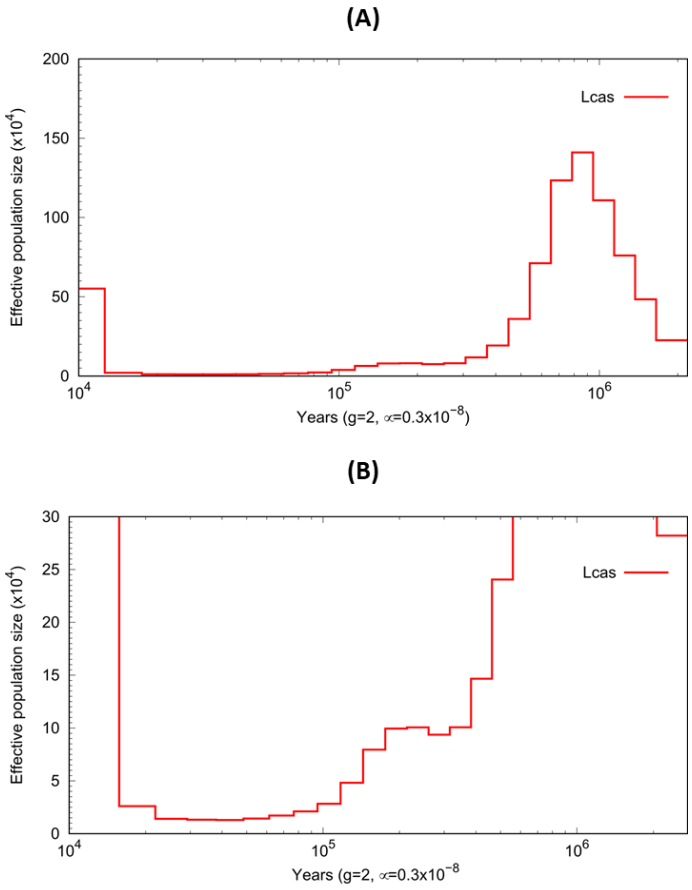



Figure 4 - PSMC inference of *L. castroviejoi* demographic profiles (A) and a zoom-in on the effective population size oscillations under $30x10^4$ (B).

The results from the demographic inference model allowing post-split gene flow did not reveal signs of admixture, and estimated similar divergence times effective population sizes ($N_e$) when compared to the model without gene flow (Appendix 9). Given these results, our analyses were based on the inferences from the model with less parameters, i.e. without post-split gene flow. The time of split between *L. castroviejoi* and *L. corsicanus* was estimated to be circa 47 thousand years ago (kya) (95% Highest Posterior Density (HPD) 30.65kya to 66.49kya) (Figure 5, Appendix 9). The effective population size (Ne) of the ancestral population (before the split) was inferred to be around 158.1k (95% HPD 122.2k to 218.1k) individuals, and the $N_e$ of current populations of *L. castroviejoi* and *L. corsicanus* were estimated around 15.6k (95% HPD 12.5k to 18,7k) and 18.2k (95% HPD 14.7k to 21.9k), respectively.

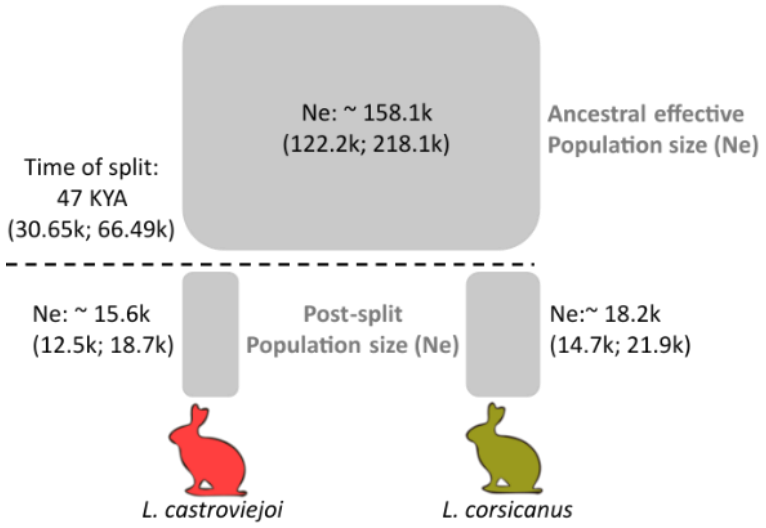



Figure 5 - The history of divergence inferred with G- PhoCS between *L. castroviejoi* and *L. corsicanus* based on a model without post-split gene flow.

### 3.3.1. Peaks of Divergence

The genetic divergence between *L. castroviejoi* and the other European hare species was assessed using pairwise genetic distances (Dxy). First, their overall divergence was investigated by performing density plots using 25kb window-based Dxy values, where the genetic distance between *L. castroviejoi* and *L. granatensis* (median $Dxy_{cas,gra}$ = 0.0085), *L. europaeus* (median $Dxy_{cas,eur}$ = 0.0085), and *L. timidus* (median $Dxy_{cas,tim}$ = 0.0083) were very similar, while the Dxy values between *L. castroviejoi* and *L. corsicanus* were substantially lower (median $Dxy_{cas,cor}$ = 0.0008) (Figure 6A). Subsequently, a genome-wide plot with the Dxy values between *L. castroviejoi* and *L. corsicanus* was conducted and peaks of divergence across the genome were detected (Appendix 10). Additionally, the RND values (a measure of divergence that aims at correcting Dxy for mutation rate variation along the genome) for those windows were also estimated, and several peaks along the genome remained (Figure 6C).

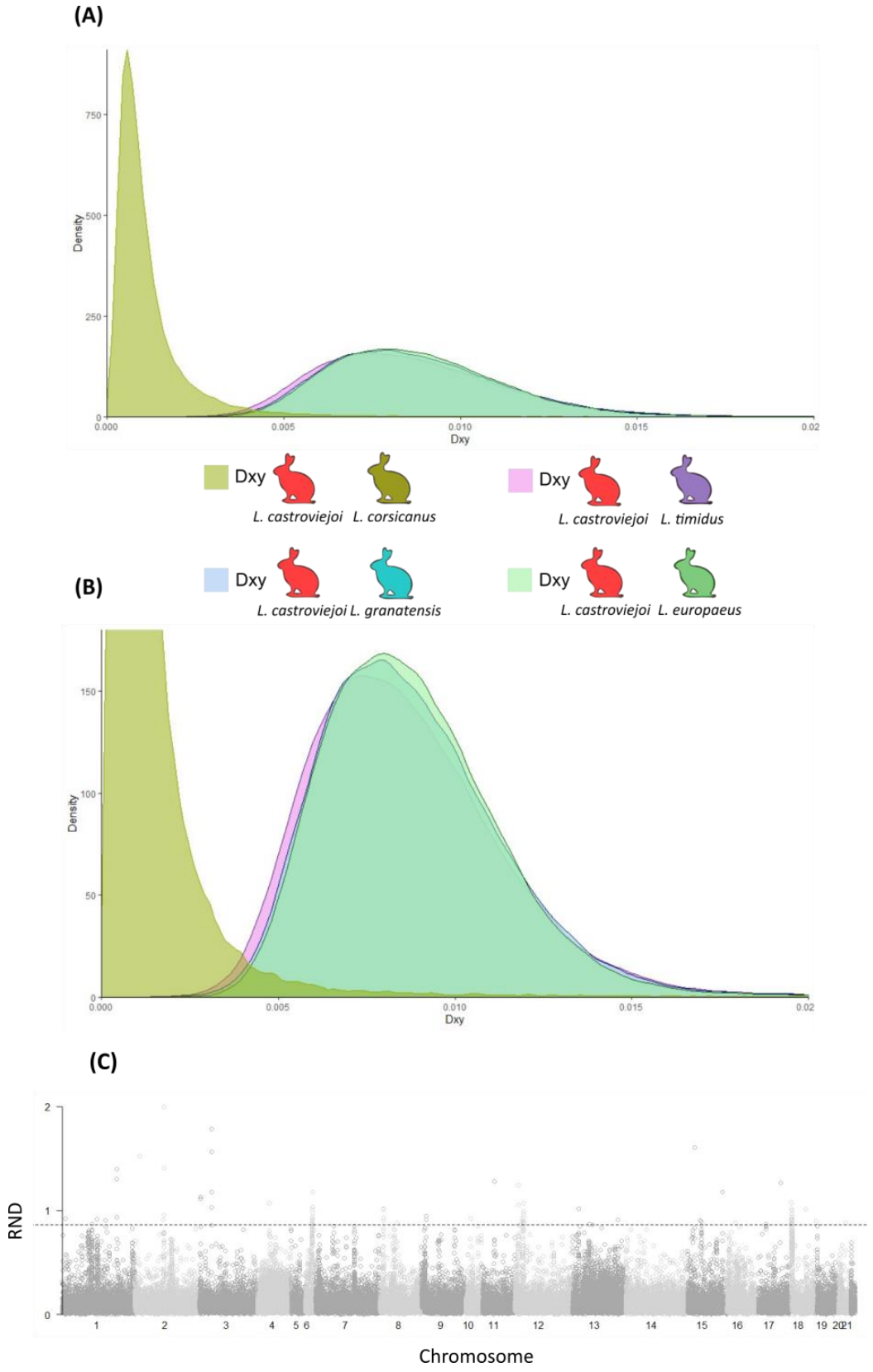



Figure 6 - Dxy distances for 82,478 25kb windows. Density plots for pairwised window based Dxy values (A) and a zoom in on the densities under 180 (B). (C) Genome scan RND distances between *L. castroviejoi* and *L. corsicanus,* using *L. americanus* as the outgroup.

## 3.4. Characterization of Introgression

### 3.4.1. Global Detection of Introgression

To detect and characterize introgression events that affected *L. castroviejoi* but not *L. corsicanus*, we took advantage of their genetic similarity*: L. corsicanus* was used as a proxy of a parental population of *L. castroviejoi*. Having access to a parental population unaffected by the processes increases the power to detect signatures of introgression from other sources (*L. granatensis*, *L. europaeus*, *L. timidus*) in the *L. castroviejoi* genome. Since *L. castroviejoi* and *L. corsicanus* are sister taxa, they are expected to share the same amount of genetic variation with the other hare species, unless an admixture event has occurred after their split, and only affected one of the species. In ABBA-BABA tests (*D*-statistics), significant signs of introgression (z score > 3) between *L. castroviejoi* and the three donor species (P3) (*L. granatensis*, *L. timidus*, *L. europaeus*) were detected, with particular stronger signs of gene flow between *L. castroviejoi* and *L. granatensis* (Figure 7). Furthermore, $f_4$ statistics were also used and similar results were obtained, with the inference of significant gene flow between *L. castroviejoi* and *L. granatensis* and *L. castroviejoi-L. timidus*, but there were no significant signs of admixture between *L. castroviejoi* and *L. europaeus* (Appendix 12). Moreover, $f_3$ statistics were used to assess if one target species (A) could be the result of an admixture event between two source species (B, C). The more negative the $f_3$ value, the more likely the admixture event. In this analysis, all $f_3$ values were positive (Appendix 11). Additionally, the TWISST method was also applied to infer the topology weights of a subset consisting of *L. castroviejoi*, *L corsicanus, L. granatensis,* and *L. americanus,* particularly to quantify the weight of the topologies different from the species tree, since these topologies could be linked to an introgression event. Overall, 95.7% of the windows displayed the species tree topology, placing *L. corsicanus* and *L. castroviejoi* as sister taxa; while 2.24% were assigned to a topology where *L. castroviejoi* and *L. granatensis* were grouped together, whereas 2.06% displayed a topology where *L. corsicanus* was closer to *L. granatensis* (Appendix 13).

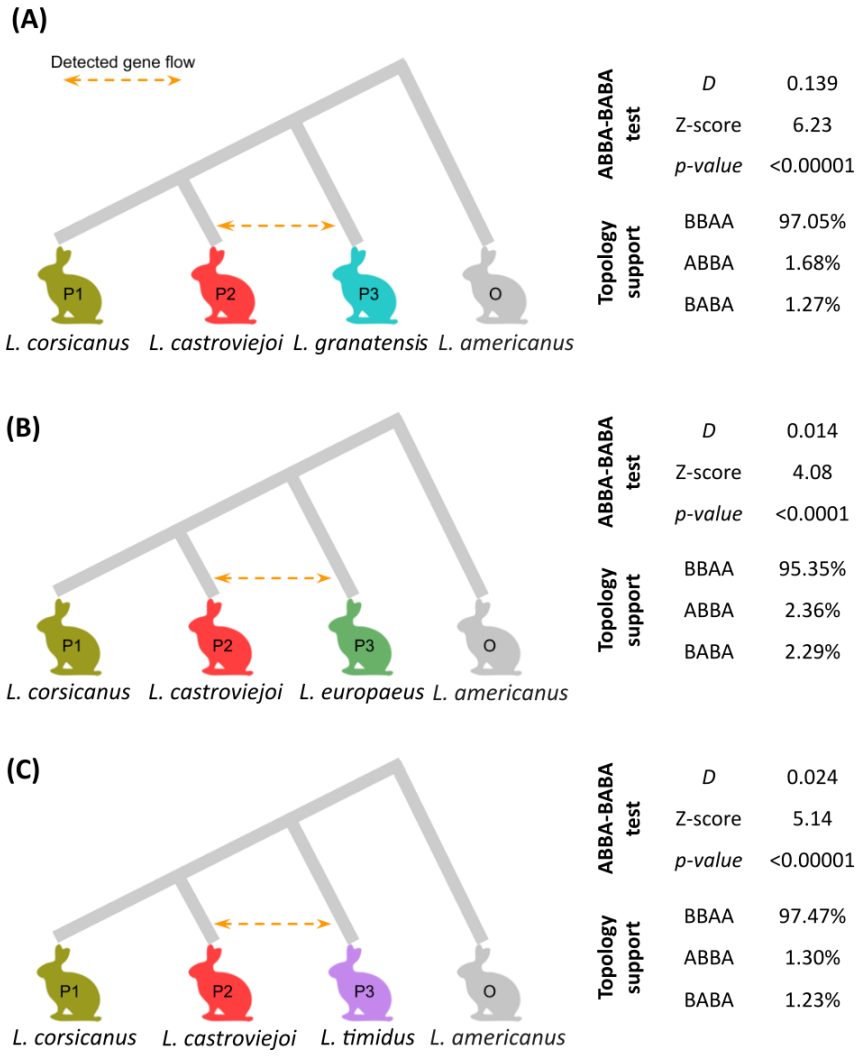



Figure 7 - D statistics (ABBA-BABA tests) for three introgression models, where **P1** – *L. corsicanus*, **P2** – *L. castroviejoi*, **O** (outgroup) – *L. americanus* were fixed, and 3 different P3 were tested: A) **P3** – *L. granatensis*; B) **P3** – *L. europaeus*; C) **P3** – *L. timidus*; Values calculated with Dsuite.

### 3.4.2. Introgression Along the genome

To be able to detect introgression segments along the genome, $f_{dM}$ values were calculated by setting *L. corsicanus* as P1, *L castroviejoi* as P2, *L. americanus* as the outgroup, and testing 3 different donor species (P3): *L. granatensis, L. europaeus, L. timidus*. $f_{dM}$ values for the model with *L. granatensis* as the donor species displayed many positive peaks, which are candidate segments of *L. granatensis* introgression in the *L. castroviejoi* genome (Figure 8). The $f_{dM}$ values for the models with *L. timidus* and *L . europaeus* as the donor species also revealed some positive peaks, which are candidates of *L. timidus* and *L. europaeus* introgression in the *L. castroviejoi* genome (Appendix 14). In contrast, few negative peaks were detected which could represent genetic contribution of these donor species in the *L. corsicanus* genome.

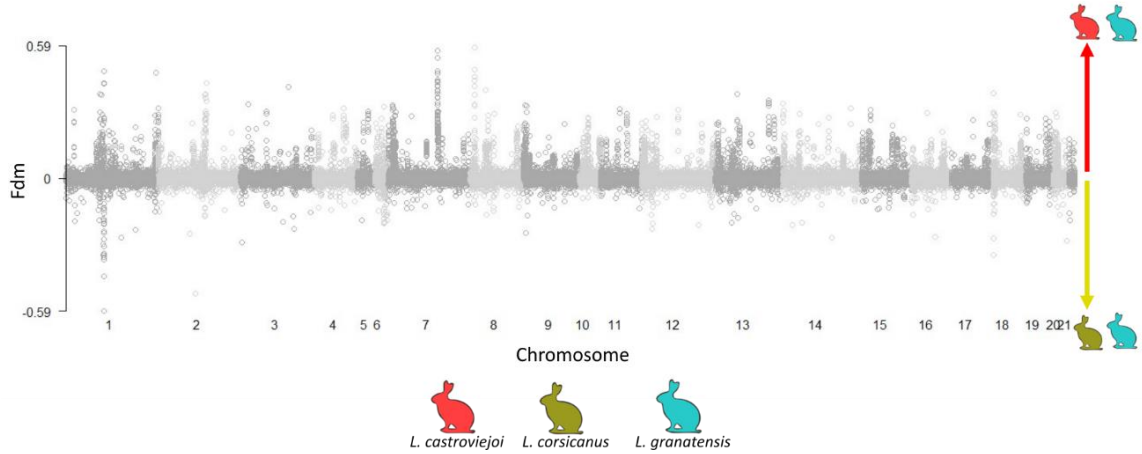



Figure 8 - Genome wide $f_{dM}$ values for the model P1 – *L. corsicanus*, P2 – *L. castroviejoi*, P3 – *L. granatensis*. Positive values suggest gene flow between P2 and P3, while negative values indicate gene flow P1 and P3.

Moreover, we simulated expected Dxy between *L. castroviejoi* and *L granatensis* distributions under i) a full demographic model to assess whether the model is able to recover the empirical distribution, and ii) a model without migration to assess Dxy expectations under a strict lineage sorting model without gene flow, and to define a threshold of introgression for the Dxy estimates (Figure 9A). The empirical Dxy distribution displayed a higher density of close to zero values and peaked in higher Dxy values than the simulated models. The presence or absence of migration did not display a visible effect in the Dxy simulated distributions. The bottom 5% quantile of the model without migration was set as the threshold for introgression candidates for the empirical *Dxy* between *L castroviejoi* and *L. granatensis* (Dxy < 0.00323). The empirical Dxy estimates between *L. castroviejoi* and *L. granatensis* values along the genome were plotted and 139 25kb windows under the threshold of introgression were detected (Figure 9B).

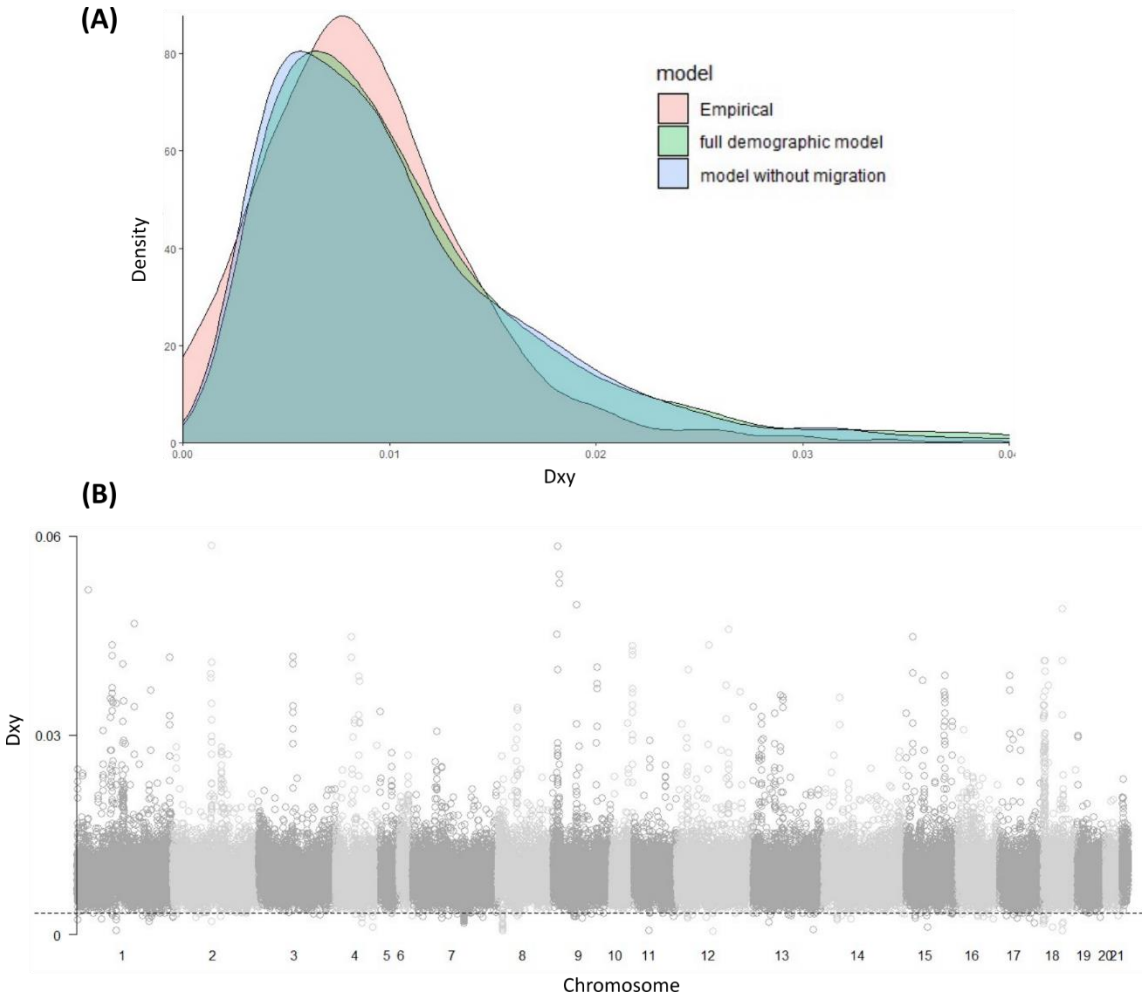



Figure 9 – Dxy L. *castroviejoi L. granatensis* distributions. (A) Dxy distributions for the empirical data and for the data derived from the demographic models simulated in msms. (B) Genome scan for the empirical Dxy *L. castroviejoi L. corsicanus* values based on 82,478 25kb window-based values. The dashed line marks the introgression threshold.

Furthermore, candidates for *L. granatensis* introgression in *L. castroviejoi* were also detected by identifying windows where *L. castroviejoi* was genetically closer to *L. granatensis* than to *L. corsicanus.* This was done by conducting a scatter plot with these two genetic distances as well as a z-score based on the ratio $\frac{Dxy\ L.corsicanus\ L.granatensis}{Dxy\ L.castroviejoi\ L.granatensis}$ (Figure 10). The regression model explained 96.96% of the variability observed and in total 418 outlier windows were detected (z-score > 3 or < -3). 396 of the outliers had z scores > 3, which means in those windows *L. castroviejoi* was genetically substantially closer to *L. granatensis* ($Dxy_{L.castroviejoi\ L.granatensis} < Dxy_{L.\ corsicanus\ L.\ granatensis}$), while 22 outliers had z scores < -3, where in those windows *L. corsicanus* was genetically considerable closer to *L. granatensis* ($Dxy_{L.castroviejoi\ L.granatensis} > Dxy_{L.\ corsicanus\ L.\ granatensis}$).

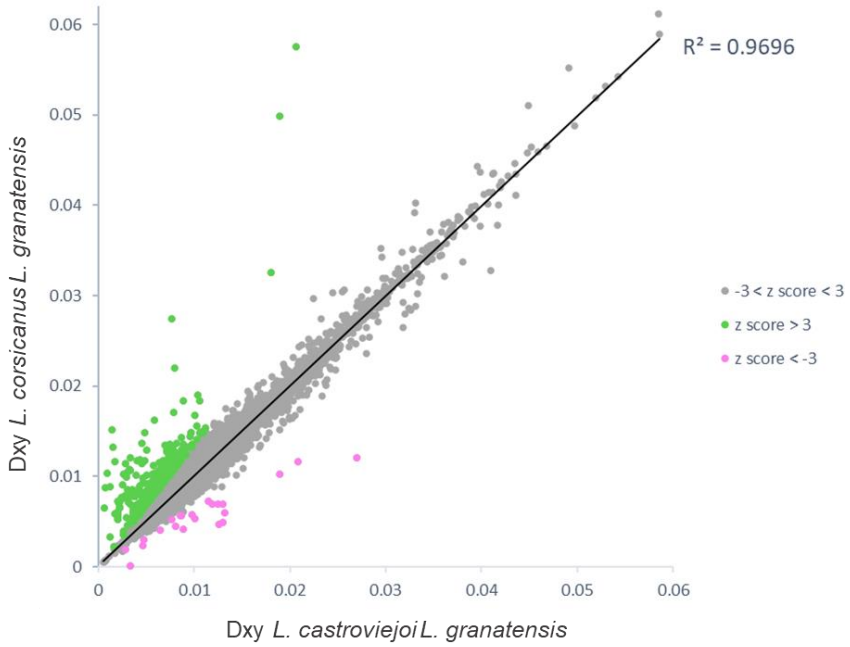



Figure 10 - Scatter plot of Dxy *L. castroviejoi L. granatensis* vs Dxy *L. corsicanus L. granatensis*. The colour of the dots represents the values obtained on a z-score test based on the ratio $\frac{Dxy\ L.corsicanus\ L.granatensis}{Dxy\ L.castroviejoi\ L.granatensis}$.

### 3.4.3. Ancestry inference along the *L. castroviejoi* genome

To better understand the genetic contribution of *L. granatensis, L. timidus* and *L. europaeus* into the genome of *L. castroviejoi*, the ancestry across *L. castroviejoi* genome was inferred by ELAI. Three independent runs were conducted, and their results were merged. Overall, 99.16% of the *L. castroviejoi* genome was attributed to *L. corsicanus* ancestry, while 0.74% was assigned to other ancestries: 0.64% to *L. granatensis*, 0.07% to *L. timidus* and 0.03% to *L. europaeus* (Table 2, Figure 11). These results agree with previous analyses that also detected strong signs of *L. granatensis* introgression as well as smaller genetic contribution of *L. timidus* and *L. europaeus* in *L. castroviejoi* (ABBA-BABA, $f_4$, $f_{dM}$, TWISST).

Table 2 - Elai ancestry proportions with *L. castroviejoi* as target.

| Ind | *L. corsicanus* | *L. granatensis* | *L. europaeus* | *L. timidus* |
|---|---|---|---|---|
| **Lcas1** | 98.823% | 0.895% | 0.041% | 0.097% |
| **Lcas2** | 99.097% | 0.743% | 0.026% | 0.064% |
| **Lcas3** | 99.343% | 0.475% | 0.032% | 0.063% |
| **Lcas4** | 99.349% | 0.465% | 0.035% | 0.063% |
| **Lcas5** | 99.185% | 0.607% | 0.029% | 0.061% |
| **Overall** | **99.160%** | **0.637%** | **0.033%** | **0.070%** |

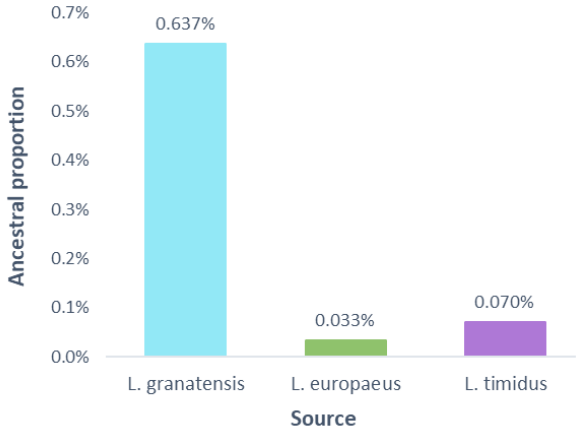



Figure 11 - Overall introgression proportion in the *L. castroviejoi* genome inferred using ELAI.

As *L. granatensis* was identified as the source for most of the introgression tracts in the *L. castroviejoi* genome, we investigated their tract distribution and introgression frequency. *L. granatensis* ancestry varied substantially per sample across the *L. castroviejoi* genome, which was indicative that the *L. granatensis* haplotypes were generally not fixed in *L. castroviejoi* (Figure 12B). Indeed, most *L. granatensis* ancestry tracks were at a lower frequency and only a few were found at high frequencies (Figure 12A).

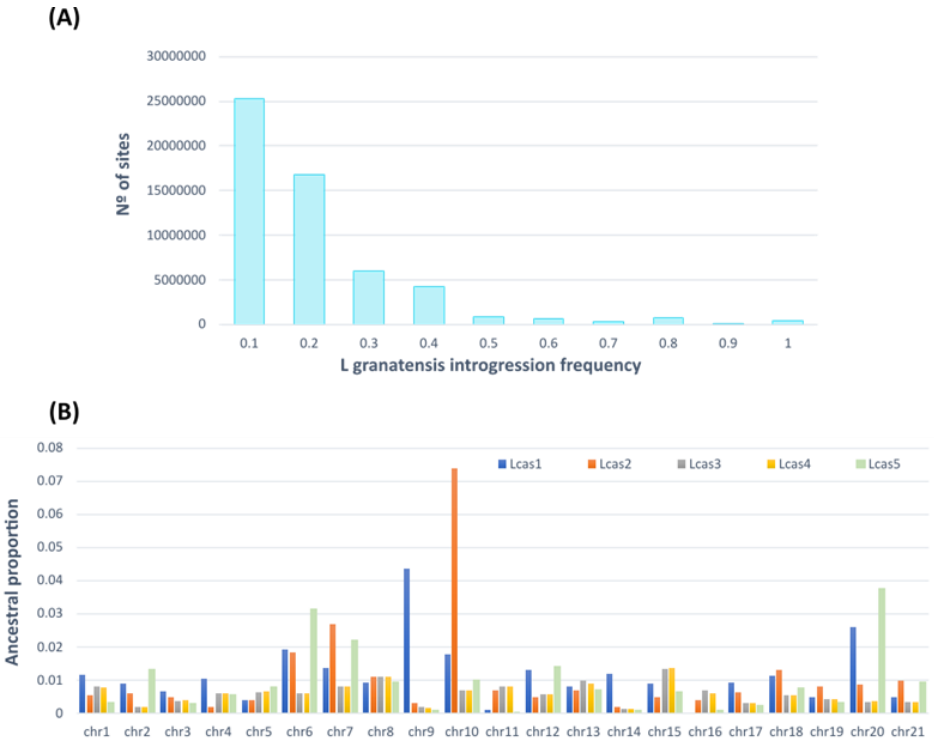



Figure 12 - Info regarding *L. granatensis* ancestry in *L. castroviejoi* genome inferred by ELAI. (A) Distribution of L. granatensis sites per introgression frequency; (B) *L. granatensis* ancestral proportion per sample per chromosome.

### 3.4.4. Junctions between ancestry tracts

As previous studies have detected introgression between *L. granatensis, L. timidus* and *L. europaeus (Seixas et al., 2018)*, we proceeded to assess if the detected genetic contribution of *L. timidus* and *L. europaeus* in *L. castroviejoi* was caused by direct contact between these species, or if it was the result of second-hand introgression (hybridization with introgressed *L. granatensis*). In a scenario of second-hand introgression, the segments with *L. timidus* or *L. europaeus* ancestry would be next to tracts of *L. granatensis* ancestry. The junctions between different ancestries per sample were investigated and 5.9% to 17.3% transitions between *L. timidus* ancestry tracts were found with *L. granatensis* (tim/cor junctions), while 3.6% to 9.1% of the *L. europaeus* junctions were transitions with *L. granatensis* ancestry (eur/cor junctions), which suggests that at least part of the genetic contribution of *L. timidus* and *L. europaeus* in *L. castroviejoi* was due to indirect contact (Figure 13, Appendix 16).

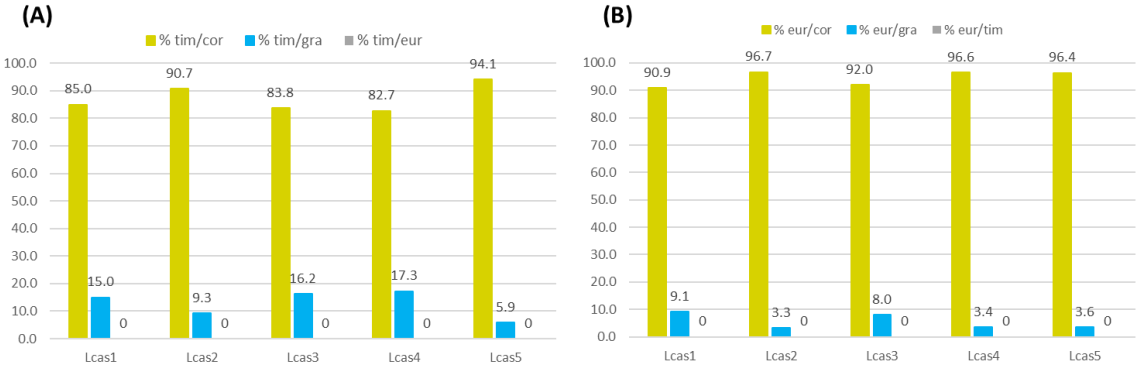



Figure 13 - Percentage of junctions between different ancestries tracts for each sample. (A) *L. timidus* ancestry transitions. (B) *L. europaeus* ancestry transitions. tim/cor: *L. timidus – L. corsicanus* junction; tim/gra: *L. timidus – L. granatensis* junction; tim/eur: *L. timidus – L. europaeus* junction; eur/cor: *L. europaeus – L. corsicanus*; eur/gra: *L. europaeus L. granatensis*.

### 3.4.5. Dating introgression events

The genetic contribution of *L. granatensis*, *L. timidus* and *L.europaeus* in *L. castroviejoi* reveals a complex evolutionary model with several direct contacts and second-hand introgression events. To try to clarify further the timings of these introgression events, we analyzed the tract sizes of the different ancestries to estimate the time of the introgression events. Overall, the tracts median values were relatively small (10.5 – 15.2 kb), with the *L. granatensis* introgression events estimated to have happened circa 10 kya (Table 3).

Table 3 - Number of introgressed tracts, median introgression tract length and the estimated time of introgression in the 5 *L. castroviejoi* individuals, as inferred by ELAI. cor: *L. corsicanus*; gra: *L. granatensis*; tim: *L. timidus*; eur: *L. europaeus*. * proxy of the *L. castroviejoi* parental

| **Parameter** | **ancestry** | | | |
|---|---|---|---|---|
| | **cor*** | **gra** | **tim** | **eur** |
| **nº of tracts** | 3 634 | 1 841 | 292 | 74 |
| **Median length (bp)** | 33 883 | 18 991 | 17 751 | 13 119 |
| **t (years)** | - | 10 531 | 11 267 | 15 246 |

### 3.4.6. Analysis of Introgression signs from different methods

To be able to better interpret the results from the different methods, we analysed the relationship of the inference from the distinct approaches. First, we divided the ELAI ancestry tracts by their introgression frequency and then calculated the $f_{dM}$ and Dxy values for those segments. For the *L. granatensis* introgression tracts, there was a clear correlation between the introgression frequency and the $f_{dM}$ values, and a negative correlation with the genetic distance between *L. castroviejoi* and *L. granatensis* (Figure 14 A, D). For the *L. timidus* tracts, there was also a correlation between the introgression frequency and the $f_{dM}$ values, but there was not a clear correlation with the Dxy values (Figure 14 B, E). For the *L. europaeus* tracts, there was no clear correlation between the introgression frequency and the $f_{dM}$ nor Dxy values, which could be due to the low number of *L. europaeus* introgression segments (Figure 14 C, F).

Furthermore, to further increase the detection power of localized introgression regions, we intersected the *L. castroviejoi L. granatensis* introgression candidate windows from the z-score test with the tracts with *L. granatensis* ancestry inference from ELAI. From the 396 z-score outliers, 267 had *L. granatensis* inferred ancestry, with most windows having an introgression frequency of 0.4 or higher (Figure 15). Overall, the z-score test demonstrated to have strong power to detect medium and high introgression frequency tracts, assuming ELAI as the correct inference.

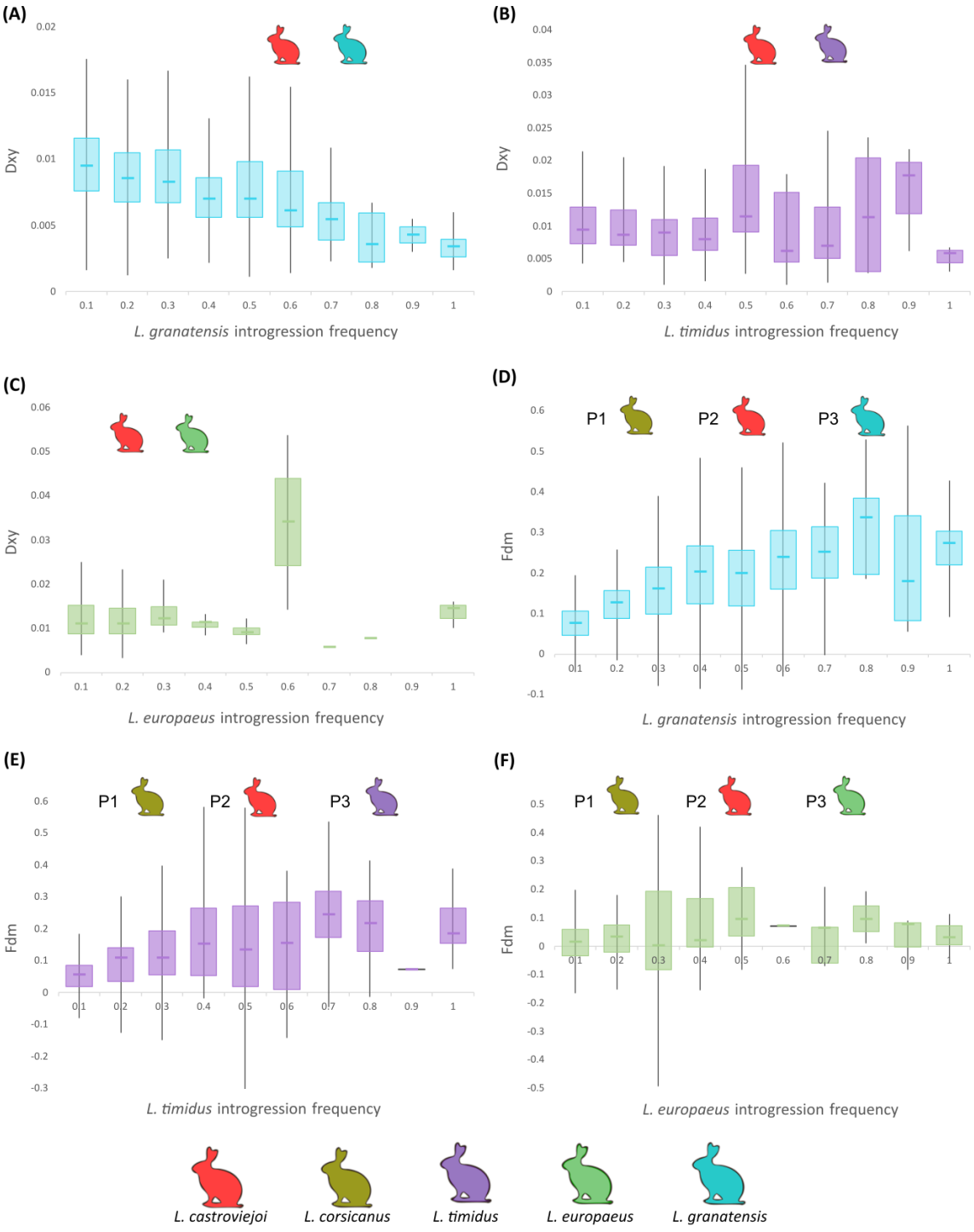



Figure 14 - Boxplots with Dxy distances per ELAI introgression frequencies in *L. castroviejoi*. (A) *L. castroviejoi L. granatensis*; (B) Dxy *L. castroviejoi L. timidus*; (C) Dxy *L. europaeus L. castroviejoi*. Boxplots with $f_{dM}$ values (P1 – *L. corsicanus*; P2 – *L. castroviejoi* per ELAI introgression frequencies in *L. castroviejoi*. (D) P3 – *L. granatensis*; (E) P3 – *L. timidus*; (F) P3 – *L. europaeus*.

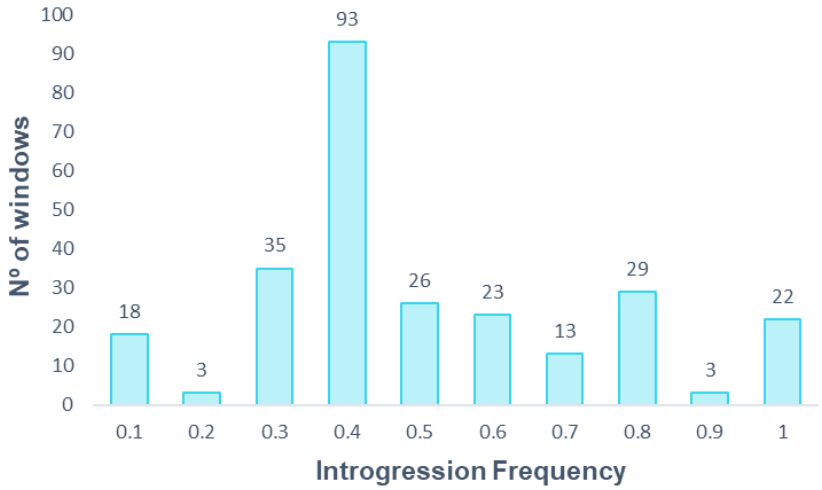
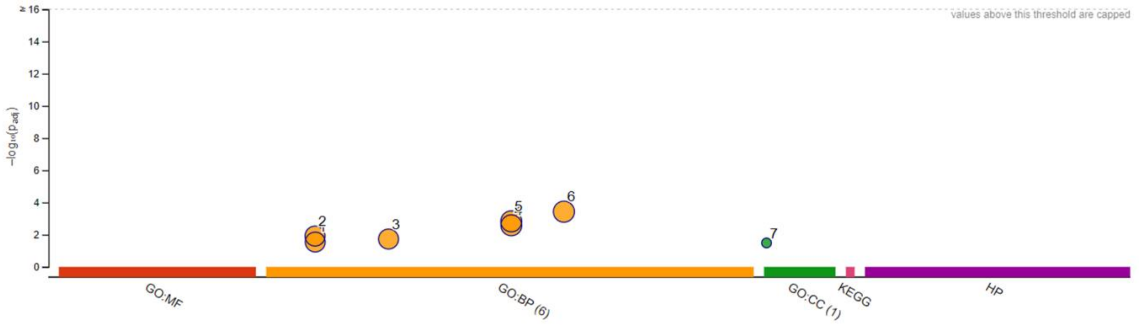



Figure 15 - Elai introgression frequency distribution of Zscore Dxy outliers windows.

### 3.4.7. Functional Impact of Candidate Genes in regions of high frequency introgression

To assess the impact of *L. granatensis* introgressed genes in *L. castroviejoi*, we performed a gene enrichment analysis for all genes within the z-score (>3) Dxy outlier windows or with an ELAI introgression frequency of 0.5. The analysis did not reveal enrichment of a particular function, however several genes linked with cell metabolism (biological regulation, regulation of biological process, regulation of cellular process, cell communication, signalling, signal transduction, voltage-gated sodium channel complex) were detected (Figure 16, Appendix 18).




| ID | Source | Term ID | Term Name | $p_{adj}$ (query_1) |
|---|---|---|---|---|
| 1 | GO:BP | GO:0007165 | signal transduction | $2.872\times10^{-2}$ |
| 2 | GO:BP | GO:0007154 | cell communication | $1.220\times10^{-2}$ |
| 3 | GO:BP | GO:0023052 | signaling | $1.870\times10^{-2}$ |
| 4 | GO:BP | GO:0050794 | regulation of cellular process | $2.621\times10^{-3}$ |
| 5 | GO:BP | GO:0050789 | regulation of biological process | $1.487\times10^{-3}$ |
| 6 | GO:BP | GO:0065007 | biological regulation | $3.737\times10^{-4}$ |
| 7 | GO:CC | GO:0001518 | voltage-gated sodium channel complex | $3.266\times10^{-2}$ |

Figure 16 - g:Profiler Gene Ontology analysis of genes within the z-score outlier windows or showing *L.granatensis* introgression frequencies of at least 50%.

# 4. Discussion

In this work, we used whole genome data to shed light on the evolutionary history of *L. castroviejoi* and *L. corsicanus*. We reconstructed the history of divergence between these sister species, which set up a timeframe for the introgression events affecting *L. castroviejoi* after the split. Genetic exchanges between *L. castroviejoi* and hares currently (*L. granatensis* and *L. europaeus*) or formerly (*L. timidus*) distributed in the Iberian Peninsula were quantified and characterized. Then, the evolutionary events that affected *L. castroviejoi* after the split from *L. corsicanus* were incorporated into the biogeographic history of hares distributed in Iberia. Finally, these results were also interpreted in the scope of the taxonomic classification of *L. castroviejoi* and *L. corsicanus*.

## 4.1. Population history of *L. castroviejoi* and *L. corsicanus*

The analysis of whole-genome data from the sister hare species, *L. castroviejoi* and *L. corsicanus*, and from the neighbouring hares in Europe, allowed providing important novel insights into their history of divergence. Previous insights on this process were limited by the scarce available genomic sampling of *L. castroviejoi* and *L. corsicanus*, and this work increased the power for a more precise understanding of the history of diversification of these species in the frame of the evolution of the genus in Europe. Previous studies relying on mtDNA, a few nuclear markers (Alves et al., 2008a; Alves et al., 2008b; Melo-Ferreira et al., 2012) and on whole exome sequences with a limited species sampling (Ferreira et al., 2021), have suggested that the sister *L. castroviejoi* and *L. corsicanus* are genetically similar, always being grouped together in phylogenetic trees. Our unsupervised population structure analyses (PCA, Figure 2A, Appendix 4C; Admixture analysis, Figure 3) containing all European hare species (*L. castroviejoi, L. corsicanus, L. granatensis, L. europaeus* and *L. timidus*) also consistently clustered these sister taxa together, confirming this close genetic relationship and a recent common evolutionary history. In keeping, genome-wide genetic distance (Dxy) between *L. castroviejoi* and *L corsicanus* was found to be very low (median $Dxy_{cas,cor}$ = 0.0008), when compared to the distances between any of these species and other hare species from Europe (median= ~ 0.0085). Interestingly, Admixture analyses were not able to distinguish *L. castroviejoi* and *L corsicanus* up to a level of population clustering that split

other species into distinct populations (K = 5), suggesting that the sister species may be less differentiated than intraspecific genetic variation in other species. Yet, a zoom in to the differentiation between *L. castroviejoi* and *L. corsicanus* (PCA using only samples from these species, Figure 2B) showed consistent differentiation between the species, which was also detected when increasing the number of K populations in the Admixture analysis to 7 (which for computational limitations was only possible to perform for chromosome 20; see Appendix 6). Globally, these results confirm the genetic similarity of *L. castroviejoi* and *L. corsicanus*, but demonstrate some degree of genetic differentiation between them, compatible with a recent split.

The split between *L. castroviejoi* and *L. corsicanus* had been poorly estimated in previous works due to a lack of intraspecific data for both species, which is needed to improve analyses based on the coalescent (Ferreira et al., 2021; Melo-Ferreira et al., 2012). Here, we relied on whole genome data to provide better insight into the divergence of these sister taxa. Using a Bayesian demographic inference method (G-PhoCS), we estimated the divergence between *L. castroviejoi* and *L. corsicanus* to have very recently occurred during the Late Pleistocene (~47KYA), with the occurrence of a post-split bottleneck in both populations, and no post divergence gene flow (Figure 5). Additionally, past demographic oscillations in *L. castroviejoi* and *L. corsicanus* were also estimated with a PSMC analysis (Appendix 8). Yet, it is important to consider that there were some inconsistencies between *L. castroviejoi* and *L. corsicanus* demographic profiles, which would be expected to coincide for time slices corresponding to their common history pre-split. Particularly, *L. corsicanus* profile had lower resolution, and during 0.8 – 1 MYA time span displayed a different demography from *L. castroviejoi*. From 40 – 800 KYA *L. castroviejoi* and *L. corsicanus* displayed similar demography patterns, with *L. castroviejoi* displaying higher $N_e$ values. After 40/50 KYA they displayed distinct population oscillation patterns which could be interpreted as their time of split. The *L. corsicanus* demographic profile was inferred using a sample from Corsica, as this species was introduced to the island in the sixteenth century, the genome of *L. corsicanus* from Corsica likely reflects a founder effect that may have induced the lower resolution in their demography profile, and the noted discrepancies with *L. castroviejoi* (0.8 – 1 MYA). Moreover, the PSMC analysis uses the genome sequence of a single individual to estimate past demographics across a long period of time (Li & Durbin, 2011), and the inferred population size oscillations can also reflect the divergence between lineages and population structure changes (Bai et al., 2018; Chikhi et al., 2010), whereas G-PhoCS relies on multiple genomes for direct estimation of divergence times and for particular (ancestral and post-split) population size estimates (Gronau et al., 2011). Therefore, these methods

complement each other methodologically and the simultaneous interpretation of their results confers an increase of robustness in reconstructing the history of divergence (Bai et al., 2018; Poelstra et al., 2021). Indeed, despite the low resolution and some discrepancies in the PSMC analyses, their results are consistent with the *L. castroviejoi L. corsicanus* time of split estimated with G-PhoCS. A previous phylogenomics study based on whole exome data and using one *L. castroviejoi* and one *L. corsicanus* individual had estimated a putative time of divergence at 120 KYA (Ferreira et al., 2021). This estimate is likely overestimated given the limited variation present in single individuals of each species, and our estimate of ~50 KYA based on an increased sampling of both species is likely more accurate. Further, by analysing whole genomes, we were able to restrict this analysis to intergenic and thus presumably neutral regions of the genome, which is more appropriate to infer neutral demographic and divergence events, not biased by potential natural selection events that affect coding regions (Johnson et al., 2008; Zhu et al., 2014). Altogether, these analyses confirmed the close genetic similarity of *L. castroviejoi* and *L. corsicanus*, placed their time of split during Late Pleistocene, ~50 KYA, provided a timeframe for evolutionary events and biogeographic scenarios affecting each of the species separately, and allowed using one species as a proxy of the parental population of the other in inferences of these post-split events.

## 4.2. Introgression in the Iberia Peninsula

Understanding the divergence history between *L. castroviejoi* and *L. corsicanus* provided means to infer more precisely evolutionary processes affecting each of the species after the split, in particular those affecting *L. castroviejoi* in the Iberian Peninsula. For this, and given the genetic relatedness of the species, we used *L. corsicanus* as a proxy of the parental population of *L. castroviejoi*, overcoming the difficulty of not having a suitable population of *L. castroviejoi* unaffected by the evolutionary processes we aimed to infer. Using this proxy of the parental we were able to infer the contribution of the other hare species distributed in the Iberian Peninsula (*L. granatensis*, *L. europaeus*) or that used to inhabit the region but are now extinct (*L. timidus*) to the genome of *L. castroviejoi*. It is however important to note that the *L. corsicanus* samples used in this study were from Corsica, where admixture with potentially introduced *L. europaeus* and *L. granatensis* has been detected (Pietri et al., 2011), which may have lead to an underestimation of introgression in our analyses.

Signs of excess of allele sharing between *L. castroviejoi* and the potential three sources were detected in the ABBA-BABA tests (Figure 7), which suggests that after the split from *L. corsicanus*, *L. castroviejoi* has undergone introgression events with *L. granatensis*, *L. europaeus* and *L. timidus*. The $f_4$ statistics also detected significant signs of admixture between *L. castroviejoi* and *L. granatensis*, and between *L. castroviejoi* and *L. timidus*, but not between *L. castroviejoi* and *L. europaeus* (Appendix 12). Indeed, from the 3 models analysed in the ABBA-BABA tests, the one inferring gene flow between *L. castroviejoi* and *L. europaeus* scored the lowest D value (although significant). One possible explanation for the weak introgression signals in the ABBA-BABA tests and $f_4$ statistics is the occurrence of admixture between *L. corsicanus* and *L europaeus*, which has been detected previously in Corsica (Pietri et al., 2011). Given that these methods evaluate the excess of allele sharing, if both P1 (*L. corsicanus*) and P2 (*L. castroviejoi*) have undergone an introgression event with P3 (*L. europaeus*), the admixture signs in this analysis would be masked. The ancestry inference (ELAI) across the *L. castroviejoi* and *L. corsicanus* genomes allowed us to further untangle these admixture events, since this approach infers the ancestry per SNP, quantifying the genetic contribution of foreign sources, assessing the introgression frequency and identifying introgression tracts along the genome. The foreign genetic contribution based on ELAI in *L. castroviejoi* (Figure 11, Table 2) was found to be ten-fold higher than in *L. corsicanus* (Appendix 17), 0.74% and 0.075% respectively, with the ancestry proportion of each source (*L. granatensis, L timidus, L. europaeus*) being higher in *L. castroviejoi*. Nevertheless, *L. europaeus* was the major source of foreign genetic contribution in *L. corsicanus*, 0.024%, while having a contribution of 0.033% in *L. castroviejoi*. The finding of similar contributions of *L. europaeus* in *L. corsicanus* and *L. castroviejoi* genomes may then confirm the masking of *L. europaeus-L. castroviejoi* introgression in the $f_4$ statistics, and the weak signal in the ABBA-BABA test.

Previous studies based on a few markers have detected some signals of introgression between *L. castroviejoi* and *L. granatensis*. Namely, an analysis using a limited sampling detected that the *L. castroviejoi* gene SRY haplotype was also present in the *L. granatensis* populations (Melo-Ferreira et al., 2009), and a coalescent simulation based on 14 nuclear loci detected some degree of nDNA gene flow from *L. granatensis* into *L. castroviejoi* (Melo-Ferreira et al., 2012). Our analyses of whole genomes allowed us to confirm the introgression events between *L. castroviejoi* and *L. granatensis*. In fact, *L. granatensis* was the major source of foreign genetic contribution in the *L. castroviejoi* genome (0.637%) (Figure 11), displayed the strongest signs of introgression in the ABBA-BABA tests (Figure 7A) and $f_4$ statistics (Appendix 12), were detected 139

windows under the estimated Dxy introgression threshold (Figure 9B), and gene flow between *L. granatensis* and *L. castroviejoi* was also observed in the Treemix analysis (m=4 and 5) (Appendix 7).

The simulated Dxy between *L. castroviejoi L. granatensis* values under the full demographic model displayed slightly bigger values than the model without migration (Figure 9A). Since the full demographic model allows post-split gene flow, it would be expected that this gene exchange would lead to lower genetic distances between the two diverging entities than a model without post-split migration. These unexpected distributions were likely caused by the very low post-split migration inferred in the full demographic model (m= 6.2E-16) (Appendix 9B), which likely had almost no impact on the calculation of the demographic parameters. This low post-split from *L. granatensis* to *L. castroviejoi* gene flow is also discordant with the other analyses which detected significant signals of *L. granatensis* to *L. castroviejoi* introgression. As these simulations were based on a reduced number of genomic fragments, they had a lower power to detect introgression than genome-wide methods (such as ELAI, ABBA-BABA tests and $f_4$). Additionally, the model without post-split gene flow included Dxy *L. castroviejoi L. granatensis* values of 0, which indicates that very low Dxy values could be produced in a strict lineage sorting model without post-split migration scenario, thus this simulated Dxy distributions may not have captured well the evolutionary processes and had a limited power in setting a threshold for introgression candidates for the empirical Dxy between *L castroviejoi* and *L. granatensis.*

The *L. Granatensis* introgression into *L. castroviejoi* was found mainly at low frequency, but we did find some instances of higher frequency and even fixed introgression in some genomic tracts (Figure 12B). Although speculative at present, high frequency introgression may result from natural selection. We attempted to understand whether particular functions were enriched in genes contained in these high frequency introgression tracts. The enrichment analyses performed on the *L granatensis* introgressed segments at higher frequency (outliers from the z-score test and tracts with ELAI haplotype frequency ≥ 0.5) (Figure 16, Appendix 18) did not detect a significant enrichment of a particular function, but several genes with roles related to the cell metabolism were identified. Previous studies have inferred that metabolism genes can be linked with high elevation adaptation in other animals (Qiu et al., 2012; Yang et al., 2016), including in American pikas (*Ochotona prínceps*) (Ge et al., 2013). Given that *L. castroviejoi* is distributed in the Cantabrian mountains, and was only found in elevations

above 1,000 meters (Ballesteros & Alves, 2022), thus environmental adaptation to high elevation can be hypothesized, but testing this hypothesis awaits further investigation.

To better understand the timeframe of the genetic exchanges between *L. castroviejoi* and the other hare species, we dated the introgression events from each source by examining the tract sizes of different ancestries and using the median values to date the introgression events (Table 3). The introgression event between *L. castroviejoi* and *L. granatensis* and between *L. castroviejoi* and *L. timidus* was estimated to have occurred 10/11 KYA, while the introgression between *L. castroviejoi* and *L. europaeus* around 15 KYA. However, it is important to note that these size tract analyses relied on a reduced number of tracts, with a dispersed distribution, and therefore the median values may not be properly representing the theoretical tract size distribution. Nonetheless, the results point out that these genetic exchanges most likely occurred at similar times after the last glacial maximum, which also coincides with the inferences of introgression events from *L. timidus* affecting *L. granatensis* (Seixas et al. 2018). The introgression events affecting *L. castroviejoi* are therefore likely ancient, though we cannot exclude continuing introgression with the neighbouring *L. granatensis* and *L. europaeus* as we detected some long introgression tracts from these species. In any, if occurring, ongoing hybridizations seem rare and the analysed *L. castroviejoi* individuals are late generation hybrids.

As previous studies have detected introgression from *L. timidus* into *L. granatensis* and *L. europaeus* and between the latter species (Marques et al., 2017; Seixas et al., 2018), the genetic contribution of *L. europaeus* and *L. timidus* in the *L. castroviejoi* genome could have either resulted from direct contact between these species or from second-hand introgression (hybridization with introgressed *L. granatensis*). The analysis of ancestry tract junctions (Figure 13, Appendix 16) allowed us to infer the origin of those contributions, with the existence of gra/tim and gra/eur junctions in all *L. castroviejoi* samples suggesting that at least part of the genetic contribution of *L. timidus* and *L. europaeus* in *L. castroviejoi* was due to second-hand introgression from *L. granatensis*. Moreover, it is important to note however that segments with eur/cor and tim/cor junctions may have been originally a result of second-hand introgression tracts which were eroded by recombination with native variants. Yet, given these results, we cannot exclude that part of the introgression results from direct contact between *L castroviejoi* and the source species.

Previous works have detected two *L. timidus* mtDNA lineages in the *L. castroviejoi* current populations: (1) the most frequent lineage is shared with *L. corsicanus* and has

most likely resulted from an introgression event between their common ancestor and *L. timidus*; (2) the second lineage is shared with *L. granatensis* and the Iberian *L. europaeus*, which must represent another introgression event (Alves et al., 2008a; Melo-Ferreira et al., 2012). Given the above results, it remains possible that the second mtDNA lineage indicated above introgressed into *L. castroviejoi* from *L. granatensis*, as the *timidus*-like lineage is the most predominant in the northern population do *L. granatensis*.

Altogether, our results provided insights into the admixture events that occurred after the *L. castroviejoi L. corsicanus* split, with a higher focus on the genetic exchanges between *L. castroviejoi* and the other species. Nonetheless, the Treemix analyses also allowed us to have a more general view of both pre and post-split introgression events (Appendix 7). The model with three migration events (m=3) detected a migration between *L. timidus* and the *L. castroviejoi L. corsicanus* ancestor, which may be a validation of the hypothesis of pre-split hybridization and introgression events from *L. timidus*, leading to the capture of the mtDNA lineage. Addictionally, this model also detected signals of migration between *L. corsicanus* and *L. americanus* that might be caused by differences in the sample sizes or related with ancestral events involving species not included in this analysis.

## 4.3. Biogeographic dynamics and introgression

The Pleistocene climatic oscillations between glacial and interglacial periods played a major role in shaping the evolutionary trajectories of hare species. This work provided some insights into the biogeographic history of these sister species and enabled us to reconstruct the time frame of their admixture history. Our results estimated that *L. castroviejoi* and *L. corsicanus* diverged recently, during Late Pleistocene, ~50kya (Figure 5, Appendix 8). Prior to their divergence, the common ancestor of *L. castroviejoi* and *L corsicanus* most likely hybridized with *L. timidus*, which resulted in the shared *L. timidus* mtDNA haplotypes present in the current populations of these sister taxa, which contrasts with the relatively distant relationship between these species and *L. timidus* inferred from nuclear DNA (Alves et al., 2008a; Melo-Ferreira et al., 2012). Previous studies have hypothesized that the common ancestor of *L. castroviejoi* and *L. corsicanus* was more widely distributed in Europe during the Pleistocene, and after their split, these species diverged into two allopatric refugia, one in the Iberian Peninsula and another in the Italian Peninsula (Alves et al., 2008a; Melo-Ferreira et al., 2012). Our demographic inference results detected that the *L. castroviejoi-L. corsicanus* common ancestor had a

substantially higher population size than the post-split populations (Figure 5, Appendix 8). The higher ancestral population size could indicate a larger population, eventually corresponding to a wider distribution, and suggests that the divergence between *L. castroviejoi L. corsicanus* occurred following a peripatric speciation model, where the post-split populations were affected by a bottleneck event. Additionally, a niche modelling analysis (Acevedo et al., 2014) inferred that these sister taxa have similar ecological niches, which suggests that their divergence was not driven by disruptive adaptation to different ecological pressures, but must have resulted from the fragmentation of favourable habitat.

After the split, *L. castroviejoi* and *L.* corsicanus underwent distinct evolutionary events. In this work, we focused more on the *L. castroviejoi* evolutionary trajectory, which was impacted by biogeographic dynamics that occurred in the Iberian Peninsula. In fact, the climate oscillations after the Last Glacial Period favoured the expansion of species well adapted to temperate environments, which promoted range revolutions in hares distributed in Iberia (Lado et al., 2018; Marques et al., 2017; Melo-Ferreira et al., 2009; Seixas et al., 2018). *L. granatensis* expanded from its refugium in Southwest Iberia and replaced *L. timidus* following a south-north invasion, possibly contributing to the extinction of *L. timidus* in the Iberian Peninsula (Lado et al., 2018; Marques et al., 2017; Seixas et al., 2018). *L. europaeus* has been suggested to be a later arrival to western Europe, reaching the Iberian Peninsula after expanding from its previous refugium in the Balkans, and replacing *L. granatensis* in Northeastern Iberia (Seixas et al., 2018; Stamatis et al., 2009). These range replacements were accompanied by introgression (Marques et al., 2017; Seixas et al., 2018). Nevertheless, the effect that these dynamics had on *L. castroviejoi* was mostly ignored in previous studies. Our results detected genetic contribution in *L. castroviejoi* from the three hare species currently or formerly distributed in Iberia: *L. granatensis*, L*. europaeus*, *L. timidus*. Moreover, we detected junctions eur/gra tim/gra (Figure 13, Appendix 16), which suggests that at least part of the *L. europaeus* and *L. timidus* genetic contribution was due to second-hand introgression by *L. granatensis*. Although we cannot discard the possibility of direct contact, our results inferred that *L. granatensis* was the major source of foreign genetic variation in *L. castroviejoi*, and the other potential contacts had a smaller impact on the *L. castroviejoi* genome. Despite the *L. castroviejoi* historical range being poorly understood, it could be a crucial piece to better understanding the drivers of the introgression event between *L. castroviejoi* and *L. granatensis*. A previous niche modelling study (Acevedo et al., 2014) inferred the *L. corsicanus* ecological niche model was able to predict the current full distribution of *L. castroviejoi*, but not the other way

around, which suggested that *L. corsicanus* populations retained the ecological traits of their ancestor. In fact, the *L. corsicanus* niche projections in Iberia estimated a high environmental potential for *L. castroviejoi* in Southern areas of its distribution, which are currently occupied by *L. granatensis*. In cases of introgression, particularly of mtDNA, and considering the range revolutions that occurred in the Iberian Peninsula, range replacement with hybridization has been invoked as a demographic process that promotes asymmetrical introgression (Seixas, 2017; Seixas et al., 2018). With the current data, it is not possible to fully understand these demographic dynamics with regard to *L. castroviejoi*. If *L. granatensis* invaded the *L. castroviejoi* range, the expected asymmetry would be in the opposite direction (introgression from the invaded, *L. castroviejoi,* into the invader, *L. granatensis*). Additionally, a biased assortative mating of females *L. granatensis* with males *L. castroviejoi* could explain the direction of mtDNA introgression and a future study about the Y chromosome could provide more insights into the possible sex-biased effect in this introgression event.

## 4.4. Taxonomic relevance

Our genomic analyses showed that *L. castroviejoi* and *L. corsicanus* are very genetically similar and diverged around 50 KYA. The species are currently classified as distinct (Smith et al., 2018), but there is some controversy around their taxonomic status. The close genetic and ecological niche similarity has fed some suggestions that these species could be classified as varieties of the same species, depending on the species concept used (Acevedo et al., 2014; Alves et al., 2008a).

Whether *L. castroviejoi* and *L. corsicanus* are conspecific or heterospecific can be something dubious and ambiguous. Our history of divergence inference suggest that no gene flow occurred between the species after the split (Appendix 9). These sister taxa are indeed genetically similar, with low genetic divergence, but then there is also enough genetic structure to segregate them (Figure 2B, Appendix 6). Moreover, *L. castroviejoi* and *L. corsicanus* diverged recently during Late Pleistocene (~50 kya) (Figure 5, Appendix 8), but since their split, they have undergone different introgression events impacting them in distinct ways. From their genetic similarity, is plausible to speculate that *L. castroviejoi* and *L. corsicanus* have not diverged enough to develop reproductive incompatibilities. Yet, given their allopatric distribution, we cannot assess whether some degree of reproductive isolation exists. Interpreting *L. castroviejoi* and *L. corsicanus* as different species or as conspecific status would be both valid, depending on the concept

used. Importantly, these sister taxa represent two endemisms, both having a vulnerable conservation status in the IUCN Red List, with *L. castroviejoi* having a small population size and being highly fragmented (Ballesteros & Alves, 2022), with low genetic diversity (Costa, unpublished results), while *L. corsicanus* is in sympatry and competing with *L. europaeus* in mainland Italy, which has likely led to a reduction in their population size (Buglione et al., 2018; Buglione et al., 2020). Thus, these endemisms undergo different threats and their taxonomic classification should also take this into consideration. Overall, our results provide taxonomists with new genetic insights that could be useful for an eventual revaluation of the classification status of *L. castroviejoi* and *L. corsicanus*.

# 5. Conclusions and Future Prospects

In this study, we shed light on the evolutionary history of the sister *L. castroviejoi* and *L. corsicanus*. Our results demonstrate the power of whole genome analyses to dissect the divergence and genetic exchange between closely related lineages. We inferred that *L. castroviejoi* and *L. corsicanus* diverged during Late Pleistocene, and had contact with different genetic entities after the split. Introgression events in Iberia affected up to 1% of the *L. castroviejoi genome*, evidencing genetic contributions from *L. granatensis*, *L. europaeus* and *L. timidus*. Furthermore, we were able to discard the possibility of one wave of introgression, detecting instances of *L. timidus* and *L. europaeus* genetic contribution in *L. castroviejoi* due to second-hand introgression from *L. granatensis*. This work also integrated *L. castroviejoi* introgression events into the biogeographic history of hares in the Iberian Peninsula, and demonstrated how analysing patterns of introgression can be useful to reconstruct the historical dynamics of closely related lineages. It has also reinforced the relevance of interspecific gene flow in the evolutionary trajectory of species and displayed their semipermeable boundaries.

In this work, we inferred the admixture history of *L. castroviejoi* and *L. corsicanus* by using one as the proxy-parental of the other. We detected stronger signals of introgression in *L. castroviejoi*, and thus this study focused more on the characterization of these events. Nevertheless, the allele sharing excess between *L. castroviejoi* and the other species has likely masked introgression signs in *L. corsicanus*. We did detect some foreign genetic contribution in *L. corsicanus*, but the impact of interspecific hybridization in this species' genome should be more deeply assessed. Here, we used *L. corsicanus* samples from Corsica, an island where this species was introduced in the sixteenth century (Scalera & Angelici, 2003). Future work should incorporate *L. corsicanus* samples from its native range (the Apennines and Sicily), as those specimens could have undergone distinct admixture events. Indeed, in mainland Italy, *L corsicanus* distribution is being invaded by *L. europaeus*, with these two species occurring in sympatry in some regions *(Buglione et al., 2018; Buglione et al., 2020; Fulgione et al., 2009)*. Although an exploratory analysis based on the mtDNA control-region, 13 autosomal microsatellites and 9 autosomal SNPs has not detected signs of hybridization between those species in mainland Italy (Mengoni et al., 2015), a deep genomic study still needs to be done in order to discard this hypothesis. As *L. europaeus* is not described in Sicily (Lo Valvo et al., 1997; Mengoni et al., 2015), *L. corsicanus* samples from that island could be used

as a proxy parental of the mainland Italy *L. corsicanus* in analyses to detect introgression of *L. europaeus*. Additionally, the sampling from *L. corsicanus* native range could also clarify ancient introgression events involving this species, since these samples would not be “contaminated” by the contacts that *L. corsicanus* from Corsica had with the other introduced species on the island.

The historical dynamics of Southern European hares are not yet completely resolved. As an example, it is uncertain whether the genetic contribution of *L. timidus* and *L. europaeus* in *L. castroviejoi* was all due to second-hand introgression by *L. granatensis*, or if there were also instances of direct genetic exchanges. Future analysis could tackle this issue by comparing the size of second-hand introgression tracts (tracts with eur/gra and tim/gra junctions) with the size of tracts without *L. granatensis* junctions (eur/cor and tim/cor). To consider the whole second-hand introgression tract size, the *L. europaeus* and *L. timidus* ancestry tracts could be merged with the adjacent *L. granatensis* tracts. If the introgression is only due to second-hand, the distribution of the size tracts with eur/cor tim/cor junctions should be the same as the eur/gra tim/gra and *L. granatensis* adjacent tract distribution. A contribution of first-hand introgression (which is expected to have occurred before the second-hand), should shift the distribution of tracts with eur/cor tim/cor sizes to lower values.

The driving forces underlying the introgression events between *L. castroviejoi* and *L. granatensis* remain unclear. To better assess the ancient dynamics between *L. castroviejoi* and *L. granatensis* introgression, it is important to know the yet poorly understood historical distribution of *L. castroviejoi*. This could be tackled by analysing ancient DNA. Indeed, a previous study analysed 5 hare ancient samples from Southern France and inferred they were *L. granatensis*, demonstrating that this species was anciently found outside of the Iberian Peninsula and shedding light on their historical distribution (Lado et al., 2018). Therefore, a future aDNA study using samples from the surrounding areas of *L. castroviejoi* current distribution could infer whether *L. castroviejoi* was formerly distributed in areas nowadays occupied by *L. granatensis*, thus shedding light on the biogeographic scenario of the introgression between *L. castroviejoi* and *L. granatensis*. Moreover, to further assess the possible biased assortative mating on the *L. granatensis L. castroviejoi* introgression, a more complete analysis of the *L. castroviejoi* Y chromosome could confirm whether this chromosome was also affected by introgression or not, and provide insight on a biased assortative mating scenario. A high-quality male *L. granatensis* reference genome will be assembled in the near future (Melo-Ferreira, personal communication), which will enable future work to assess the Y

chromosomal introgression in *L. castroviejoi*. Furthermore, the possible adaptive role of the introgressed segments in the *L. castroviejoi* genome remains poorly resolved. Although the functional enrichment analysis did not detect enrichment of a particular function, the potential adaptive impact of those introgressed genes cannot yet be discarded. Thus, future work should test if those introgression events were driven by natural selection. The inference of natural selection signatures along the genomes is based on detecting sudden shifts of allele frequencies (Bamshad & Wooding, 2003; Stephan, 2016). Therefore, a sample size of at least 15 individuals per species is recommended to have more precise allele frequencies and consequently accurate inference of selective pressures (Ma et al., 2015). In this work, we only had 5 whole genomes of *L. castroviejoi* and *L. corsicanus*, which made the detection of selective sweeps unreliable. Currently, 12 additional genomes of *L. castroviejoi* and 15 of *L. corsicanus* were sequenced and will be analysed in future works. With this increase in sample size, we will be able to perform genome scans for signatures of selection based on de-correlated composite of multiple signals (DCMS) (Ma et al., 2015) and thus infer if there were cases of adaptive introgression in *L. castroviejoi*.

# 6. References


Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., & Buggs, R. (2013). Hybridization and speciation. *Journal of evolutionary biology*, *26*(2), 229-246.

Acevedo, P., Melo-Ferreira, J., Farelo, L., Beltran-Beck, B., Real, R., Campos, R., & Alves, P. C. (2015). Range dynamics driven by Quaternary climate oscillations explain the distribution of introgressed mt DNA of Lepus timidus origin in hares from the Iberian Peninsula. *Journal of Biogeography*, *42*(9), 1727-1735.

Acevedo, P., Melo-Ferreira, J., Real, R., & Alves, P. C. (2014). Evidence for niche similarities in the allopatric sister species L epus castroviejoi and L epus corsicanus. *Journal of biogeography*, *41*(5), 977-986.

Adavoudi, R., & Pilot, M. (2021). Consequences of hybridization in mammals: A systematic review. *Genes*, *13*(1), 50.

Aggeli, D., Li, Y., & Sherlock, G. (2021). Changes in the distribution of fitness effects and adaptive mutational spectra following a single first step towards adaptation. *Nature communications*, *12*(1), 1-14.

Aguillon, S. M., Dodge, T. O., Preising, G. A., & Schumer, M. (2022). Introgression. *Current Biology*, *32*(16), R865-R868.

Albrechtova, J., Albrecht, T., Baird, S. J., Macholan, M., Rudolfsen, G., Munclinger, P., Tucker, P. K., & Pialek, J. (2012). Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1748), 4803-4810.

Aldhebiani, A. Y. (2018). Species concept and speciation. *Saudi journal of biological sciences*, *25*(3), 437-440.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, *19*(9), 1655-1664.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697-709.

Alves, P., Ferrand, N., Suchentrunk, F., & Harris, D. (2003). Ancient introgression of Lepus timidus mtDNA into L. granatensis and L. europaeus in the Iberian Peninsula. *Molecular phylogenetics and evolution*, *27*(1), 70-80.

Alves, P., Melo-Ferreira, J., Branco, M., Suchentrunk, F., Ferrand, N., & Harris, D. (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*(3), 1191.

Alves, P. C., & Melo-Ferreira, J. (2007). Are Lepus corsicanus and L. castroviejoi conspecific? Evidence from the analysis of nuclear markers. *Conservazione di Lepus corsicanus stato delle conoscenze. IGF Pub, Napoli*, 45-52.

Alves, P. C., 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 B: Biological Sciences*, *363*(1505), 2831-2839.

Anderson, E. (1948). Hybridization of the habitat. *Evolution*, 1-9.

Anderson, E., & Hubricht, L. (1938). Hybridization in Tradescantia. III. The evidence for introgressive hybridization. *American Journal of Botany*, *25*(6), 396-402.

André, T., Salzman, S., Wendt, T., & Specht, C. D. (2016). Speciation dynamics and biogeography of Neotropical spiral gingers (Costaceae). *Molecular phylogenetics and evolution*, *103*, 55-63.

Arnold, M. L. (1997). *Natural hybridization and evolution*. Oxford University Press on Demand.

Arnold, M. L., & Kunte, K. (2017). Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. *Trends in ecology & evolution*, *32*(8), 601-611.

Arnold, M. L., & Martin, N. H. (2009). Adaptation by introgression. *Journal of Biology*, *8*(9), 1-3.

Bai, W. N., Yan, P. C., Zhang, B. W., Woeste, K. E., Lin, K., & Zhang, D. Y. (2018). Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus Juglans (Juglandaceae) revealed by whole-genome sequences. *New Phytologist*, *217*(4), 1726-1736.

Ballesteros, F., & Alves, P. C. (2022). Broom Hare Lepus castroviejoi Palacios, 1977. In *Handbook of the Mammals of Europe* (pp. 1-12). Springer.

Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. *Nature Reviews Genetics*, *4*(2), 99-110.

Barton, N. H., & Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics*, *15*, 133-164.

Baskett, M. L., & Gomulkiewicz, R. (2011). Introgressive hybridization as a mechanism for species rescue. *Theoretical Ecology*, *4*(2), 223-239.

Bay, R. A., Taylor, E. B., & Schluter, D. (2019). Parallel introgression and selection on introduced alleles in a native species. *Molecular ecology*, *28*(11), 2802-2813.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Breese, M. R., & Liu, Y. (2013). NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics*, *29*(4), 494-496.

Breusing, C., Johnson, S. B., Tunnicliffe, V., Clague, D. A., Vrijenhoek, R. C., & Beinart, R. A. (2020). Allopatric and sympatric drivers of speciation in Alviniconcha hydrothermal vent snails. *Molecular biology and evolution*, *37*(12), 3469-3484.

Buglione, M., Maselli, V., Rippa, D., de Filippo, G., Trapanese, M., & Fulgione, D. (2018). A pilot study on the application of DNA metabarcoding for non-invasive diet analysis in the Italian hare. *Mammalian Biology*, *88*, 31-42.

Buglione, M., Petrelli, S., Notomista, T., de Filippo, G., Gregorio, R., & Fulgione, D. (2020). Who is who? High Resolution Melting analysis to discern between hare species using non-invasive sampling. *Conservation Genetics Resources*, *12*(4), 727-732.

Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., Vigouroux, Y., & Berthouly-Salazar, C. (2019). Adaptive introgression: an untapped evolutionary mechanism for crop adaptation. *Frontiers in Plant Science*, *10*, 4.

Bush, G. L. (1975). Modes of animal speciation. *Annual Review of Ecology and Systematics*, 339-364.

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 2997-3007.

Carneiro, M., Rubin, C.-J., Di Palma, F., Albert, F. W., Alföldi, J., Barrio, A. M., Pielberg, G., Rafati, N., Sayyab, S., & Turner-Maier, J. (2014). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *science*, *345*(6200), 1074-1079.

Carson, H. L. (1971). Speciation and the founder principle.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-13015-10047-13748.

Chantry-Darmon, C., Urien, C., De Rochambeau, H., Allain, D., Pena, B., Hayes, H., ... & Rogel-Gaillard, C. (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*(4), 335-341.

Chikhi, L., Sousa, V. C., Luisi, P., Goossens, B., & Beaumont, M. A. (2010). The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, *186*(3), 983-995.

Coyne, J. A., & Orr, H. A. (2004). *Speciation* (Vol. 37). Sinauer Associates Sunderland, MA.

Crespi, B., & Nosil, P. (2013). Conflictual speciation: species formation via genomic conflict. *Trends in ecology & evolution*, *28*(1), 48-57.

Cristescu, M. E., Constantin, A., Bock, D. G., Caceres, C. E., & Crease, T. J. (2012). Speciation with gene flow and the genetics of habitat transitions. *Molecular ecology*, *21*(6), 1411-1422.

Cronquist, A. (1978). Once again, what is a species? Biosystematics in agriculture. Beltsville Symposia in Agr. Res.,

Cui, R., Schumer, M., Kruesi, K., Walter, R., Andolfatto, P., & Rosenthal, G. G. (2013). Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution*, *67*(8), 2166-2179.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*. Books, Incorporated, Pub.

de Lafontaine, G., Prunier, J., Gérardi, S., & Bousquet, J. (2015). Tracking the progression of speciation: variable patterns of introgression across the genome provide insights on the species delimitation between progenitor–derivative spruces (Picea mariana× P. rubens). *Molecular ecology*, *24*(20), 5229-5247.

De Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation. *Endless forms: species and speciation*.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879-886.

De Winton, W. (1898). XXV.—On the hares of Western Europe and North Africa. *Journal of Natural History*, *1*(2), 149-158.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., & Hanna, M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, *43*(5), 491-498.

Dion-Côté, A.-M., & Barbash, D. A. (2017). Beyond speciation genes: an overview of genome stability in evolution and speciation. *Current opinion in genetics & development*, *47*, 17-23.

Dobzhansky, T. (1937). Genetics and the origin of species. New York: Columbia, University Press, 446 p.

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, *28*(8), 2239-2252.

Duranton, M., & Pool, J. E. (2022). Interactions between natural selection and recombination shape the genomic landscape of introgression. *Molecular biology and evolution*, *39*(7), msac122.

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., García-Accinelli, G., Van Belleghem, S. M., Patterson, N., & Neafsey, D. E. (2019). Genomic architecture and introgression shape a butterfly radiation. *science*, *366*(6465), 594-599.

Endler, J. A. (1977). Geographic Variation, Speciation and Clines.(MPB-10), Volume 10. In *Geographic Variation, Speciation and Clines.(MPB-10), Volume 10*. Princeton University Press.

Ewing, G., & Hermisson, J. (2010). MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics*, *26*(16), 2064-2065.

Feder, J. L., Xie, X., Rull, J., Velez, S., Forbes, A., Leung, B., Dambroski, H., Filchak, K. E., & Aluja, M. (2005). Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in Rhagoletis. *Proceedings of the National Academy of Sciences*, *102*(suppl_1), 6573-6580.

Ferguson, J. W. H. (2002). On the use of genetic divergence for identifying species. *Biological Journal of the Linnean Society*, *75*(4), 509-516.

Ferreira, M. S., Jones, M. R., Callahan, C. M., Farelo, L., Tolesa, Z., Suchentrunk, F., Boursot, P., Mills, L. S., Alves, P. C., & Good, J. M. (2021). The legacy of recurrent introgression during the radiation of hares. *Systematic Biology*, *70*(3), 593-607.

Fisher, R. A. (1930). The genetical theory of natural selection. Clarendon. In: Oxford.

Fulgione, D., Maselli, V., Pavarese, G., Rippa, D., & Rastogi, R. K. (2009). Landscape fragmentation and habitat suitability in endangered Italian hare (Lepus corsicanus) and European hare (Lepus europaeus) populations. *European Journal of Wildlife Research*, *55*(4), 385-396.

Gavrilets, S. (2004). *Fitness landscapes and the origin of species (MPB-41)*. Princeton University Press.

Ge, R.-L., Cai, Q., Shen, Y.-Y., San, A., Ma, L., Zhang, Y., Yi, X., Chen, Y., Yang, L., & Huang, Y. (2013). Draft genome sequence of the Tibetan antelope. *Nature communications*, *4*(1), 1-7.

Giska, I., Farelo, L., Pimenta, J., Seixas, F. A., Ferreira, M. S., Marques, J. P., Miranda, I., Letty, J., Jenny, H., & Hackländer, K. (2019). Introgression drives repeated

evolution of winter coat color polymorphism in hares. *Proceedings of the National Academy of Sciences*, *116*(48), 24150-24156.

Gosden, T. P., Stoks, R., & Svensson, E. I. (2011). Range limits, large-scale biogeographic variation, and localized evolutionary dynamics in a polymorphic damselfly. *Biological Journal of the Linnean Society*, *102*(4), 775-785.

Goulet, B. E., Roda, F., & Hopkins, R. (2017). Hybridization in plants: old ideas, new techniques. *Plant Physiology*, *173*(1), 65-78.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., & Fritz, M. H.-Y. (2010). A draft sequence of the Neandertal genome. *science*, *328*(5979), 710-722.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics*, *43*(10), 1031-1034.

Guan, Y. (2014). Detecting structure of haplotypes and local ancestry. *Genetics*, *196*(3), 625-642.

Hallström, B. M., & Janke, A. (2010). Mammalian evolution may not be strictly bifurcating. *Molecular biology and evolution*, *27*(12), 2804-2816.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(S1), 795-809.

Heiser, C. B. (1949). Natural hybridization with particular reference to introgression. *The Botanical Review*, *15*(10), 645-687.

Hermansen, J. S., Sæther, S. A., Elgvin, T. O., Borge, T., Hjelle, E., & SÆTRE, G. P. (2011). Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular ecology*, *20*(18), 3812-3822.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology*, *7*(1), 44.

Johnson, L. A., Chan, L. M., Weese, T. L., Busby, L. D., & McMurry, S. (2008). Nuclear and cpDNA sequences combined provide strong inference of higher phylogenetic relationships in the phlox family (Polemoniaceae). *Molecular phylogenetics and evolution*, *48*(3), 997-1012.

Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J., Jiggins, F. M., Jensen, J. D., Melo-Ferreira, J., & Good, J. M. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *science*, *360*(6395), 1355-1358.

Kamenetzky, L., Asís, R., Bassi, S., de Godoy, F., Bermudez, L., Fernie, A. R., Van Sluys, M.-A., Vrebalov, J., Giovannoni, J. J., & Rossi, M. (2010). Genomic analysis of wild tomato introgressions determining metabolism-and yield-associated traits. *Plant Physiology*, *152*(4), 1772-1786.

Kenney, A. M., & Sweigart, A. L. (2016). Reproductive isolation and introgression between sympatric Mimulus species. *Molecular ecology*, *25*(11), 2499-2517.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular biology and evolution*, *23*(3), 482-490.

Kozak, M., Bocianowski, J., Liersch, A., Tartanus, M., Bartkowiak-Broda, I., Piotto, F. A., & Azevedo, R. A. (2011). Genetic divergence is not the same as phenotypic divergence. *Molecular Breeding*, *28*(2), 277-280.

La Morgia, V., & Venturino, E. (2017). Understanding hybridization and competition processes between hare species: implications for conservation and management on the basis of a mathematical model. *Ecological Modelling*, *364*, 13-24.

Lado, S., Farelo, L., Forest, V., Acevedo, P., Dalén, L., & Melo-Ferreira, J. (2018). Post-glacial range revolutions in South European hares (Lepus spp.): Insights from ancient DNA and ecological niche modelling. *Journal of Biogeography*, *45*(12), 2609-2618.

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary ecology*, *30*(2), 191-202.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987-2993.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493-496.

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*(16), 2078-2079.

Liang, M., & Nielsen, R. (2014). The lengths of admixture tracts. *Genetics*, *197*(3), 953-967.

Liu, J., Yu, L., Arnold, M. L., Wu, C.-H., Wu, S.-F., Lu, X., & Zhang, Y.-P. (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*(1), 1-14.

Lo Valvo, M., Barera, A., & Seminara, S. (1997). Biometria e status della Lepre appenninica (Lepus corsicanus, de Winton 1898) in Sicilia.

Lopez-Martinez, N. (2008). The lagomorph fossil record and the origin of the European rabbit. In *Lagomorph biology* (pp. 27-46). Springer.

Lynch, M., & Walsh, B. (2007). *The origins of genome architecture* (Vol. 98). Sinauer Associates Sunderland, MA.

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., & Simianer, H. (2015). Properties of different selection signature statistics and a new strategy for combining them. *Heredity*, *115*(5), 426-436.

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *science*, *350*(6267), 1493-1498.

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Molecular ecology resources*, *21*(2), 584-595.

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature ecology & evolution*, *2*(12), 1940-1955.

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279-283.

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays*, *38*(2), 140-149.

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

Marques, J. P., Ferreira, M. S., Farelo, L., Callahan, C. M., Hackländer, K., Jenny, H., Montgomery, W. I., Reid, N., Good, J. M., & Alves, P. C. (2017). Mountain hare transcriptome and diagnostic markers as resources to monitor hybridization with European hares. *Scientific data*, *4*(1), 1-11.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), 10-12.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome research*, *23*(11), 1817-1828.

Martin, S. H., & Van Belleghem, S. M. (2017). Exploring evolutionary relationships across the genome using topology weighting. *Genetics*, *206*(1), 429-438.

Matthee, C. A., Van Vuuren, B. J., Bell, D., & Robinson, T. J. (2004). A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Systematic Biology*, *53*(3), 433-447.

Matute, D. R., Comeault, A. A., Earley, E., Serrato-Capuchina, A., Peede, D., Monroy-Eklund, A., Huang, W., Jones, C. D., Mackay, T. F., & Coyne, J. A. (2020). Rapid and predictable evolution of admixed populations between two Drosophila species pairs. *Genetics*, *214*(1), 211-230.

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

Mayr, E. (1963). Animal species and evolution. In *Animal species and evolution*. Harvard University Press.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), 1297-1303.

Meier, J., Marques, D., Mwaiko, S., Wagner, C., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8: 14363. In: Nature Publishing Group.

Melo-Ferreira, J., Alves, P., 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*(12), 2643-2658.

Melo-Ferreira, J., Boursot, P., Carneiro, M., Esteves, P., Farelo, L., & Alves, P. (2012). Recurrent introgression of mitochondrial DNA among hares (Lepus spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology*, *61*(3), 367.

Melo-Ferreira, J., Farelo, L., Freitas, H., Suchentrunk, F., Boursot, P., & Alves, P. C. (2014a). Home-loving boreal hare mitochondria survived several invasions in Iberia: the relative roles of recurrent hybridisation and allele surfing. *Heredity*, *112*(3), 265-273.

Melo-Ferreira, J., Vilela, J., Fonseca, M. M., da Fonseca, R. R., Boursot, P., & Alves, P. C. (2014b). The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome biology and evolution*, *6*(4), 886-896.

Melo-Ferreira, J., Alves, P. C., Rocha, J., Ferrand, N., & Boursot, P. (2011). Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? *Evolution: International Journal of Organic Evolution*, *65*(7), 1956-1968.

Melo-Ferreira, J., Boursot, P., Randi, E., Kryukov, A., Suchentrunk, F., Ferrand, N., & Alves, P. (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*(3), 605-618.

Mengoni, C., Mucci, N., & Randi, E. (2015). Genetic diversity and no evidences of recent hybridization in the endemic Italian hare (Lepus corsicanus). *Conservation genetics*, *16*(2), 477-489.

Miao, B., Wang, Z., & Li, Y. (2017). Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the gray wolf from the Tibetan Plateau. *Molecular biology and evolution*, *34*(3), 734-743.

Millstein, R. L. (2016). Genetic drift.

Mitchell-Jones, A. J., Amori, G., Bogdanowicz, W., Krystufek, B., Reijnders, P., Spitzenberger, F., Stubbe, M., Thissen, J., Vohralik, V., & Zima, J. (1999). *The atlas of European mammals* (Vol. 3). Academic Press London.

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization. *ELife*, *10*, e69016.

Muirhead, C. A., & Presgraves, D. C. (2016). Hybrid incompatibilities, local adaptation, and the genomic distribution of natural introgression between species. *The American Naturalist*, *187*(2), 249-261.

Nachman, M. W., & Payseur, B. A. (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*(1587), 409-421.

Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular ecology*, *25*(5), 1058-1072.

Neafsey, D. E., Barker, B. M., Sharpton, T. J., Stajich, J. E., Park, D. J., Whiston, E., Hung, C.-Y., McMahan, C., White, J., & Sykes, S. (2010). Population genomic sequencing of Coccidioides fungi reveals recent hybridization and transposon control. *Genome research*, *20*(7), 938-946.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia university press.

Noor, M. A. (2002). Is the biological species concept showing its age? *Trends in ecology & evolution*, *17*(4), 153-154.

Olson-Manning, C. F., Wagner, M. R., & Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, *13*(12), 867-877.

Palacios, E. (1996). Systematics of the indigenous hares of Italy traditionally identified as Lepus europaeus Pallas, 1778 (Mammalia: Leporidae). *Bonner Zoologische Beitrage*, *46*, 59-92.

Palacios, F. (1976). Descripción de una nueva especie de liebre (Lepus castroviejoi), endémica de la Cordillera Cantábrica. *Doñana, Acta Vertebrata*, *3*(2), 205-223.

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O., & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in Heliconius butterflies. *PLOS Genetics*, *8*(6), e1002752.

Parepa, M., Fischer, M., Krebs, C., & Bossdorf, O. (2014). Hybridization increases invasive knotweed success. *Evolutionary applications*, *7*(3), 413-420.

Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Mpoloka, S. W., Nakagawa, H., & Naumann, C. (2012). The genetic prehistory of southern Africa. *Nature communications*, *3*(1), 1-6.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics*, *8*(11), e1002967.

Pietri, C., Alves, P. C., & Melo-Ferreira, J. (2011). Hares in Corsica: high prevalence of Lepus corsicanus and hybridization with introduced L. europaeus and L. granatensis. *European journal of wildlife research*, *57*(2), 313-321.

Pinho, C., & Hey, J. (2010). Divergence with gene flow: models and data. *Annual review of ecology, evolution, and systematics*, 215-230.

Poelstra, J. W., Salmona, J., Tiley, G. P., Schüßler, D., Blanco, M. B., Andriambeloson, J. B., Bouchez, O., Campbell, C. R., Etter, P. D., & Hohenlohe, P. A. (2021). Cryptic patterns of speciation in cryptic primates: microendemic mouse lemurs and the multispecies coalescent. *Systematic Biology*, *70*(2), 203-218.

Pool, J. E., & Nielsen, R. (2009). Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics*, *181*(2), 711-719.

Presgraves, D. C. (2010). The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, *11*(3), 175-180.

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., & Larkin, D. M. (2012). The yak genome and adaptation to life at high altitude. *Nature genetics*, *44*(8), 946-949.

Quilodrán, C. S., Nussberger, B., Montoya-Burgos, J. I., & Currat, M. (2019). Hybridization and introgression during density-dependent range expansion: European wildcats as a case study. *Evolution*, *73*(4), 750-761.

Rambaut, A., Suchard, M., Xie, D., & Drummond, A. (2014). Tracer v1. 6. Available from http. *beast. bio. ed. ac. uk/Tracer*.

Ramírez-Valiente, J. A., Deacon, N. J., Etterson, J., Center, A., Sparks, J. P., Sparks, K. L., Longwell, T., Pilz, G., & Cavender-Bares, J. (2018). Natural selection and neutral evolutionary processes contribute to genetic divergence in leaf traits across a precipitation gradient in the tropical oak Quercus oleoides. *Molecular ecology*, *27*(9), 2176-2192.

Randi, E. (2007). Phylogeography of south European mammals. In *Phylogeography of southern European refugia* (pp. 101-126). Springer.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research*, *47*(W1), W191-W198.

Ravinet, M., Kume, M., Ishikawa, A., & Kitano, J. (2021). Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats. *Journal of Evolutionary Biology*, *34*(1), 114-127.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489-494.

Rendón-Anaya, M., Wilson, J., Sveinsson, S., Fedorkov, A., Cottrell, J., Bailey, M. E., Ruņģis, D., Lexer, C., Jansson, S., & Robinson, K. M. (2021). Adaptive introgression facilitates adaptation to high latitudes in European aspen (Populus tremula L.). *Molecular biology and evolution*, *38*(11), 5034-5050.

Sankararaman, S., Mallick, S., Patterson, N., & Reich, D. (2016). The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Current Biology*, *26*(9), 1241-1247.

Scalera, R., & Angelici, F. (2003). Rediscovery of the Apennine hare Lepus corsicanus in Corsica. *Bolletino del Museo Regionale Scienze Naturali Torino*, *20*, 161-166.

Schield, D. R., Adams, R. H., Card, D. C., Perry, B. W., Pasquesi, G. M., Jezkova, T., Portik, D. M., Andrew, A. L., Spencer, C. L., & Sanchez, E. E. (2017). Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. *Ecology and evolution*, *7*(11), 3951-3966.

Schreiber, D., & Pfenninger, M. (2021). Genomic divergence landscape in recurrently hybridizing Chironomus sister taxa suggests stable steady state between mutual gene flow and isolation. *Evolution letters*, *5*(1), 86-100.

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., & Rosenthal, G. G. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *science*, *360*(6389), 656-660.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in ecology & evolution*, *19*(4), 198-207.

Seixas, F. (2017). Genome admixture with massive mitochondrial DNA introgression in hares (Lepus spp.): The relative roles of demography and natural selection. *University of Porto and University of Montpellier, Porto and Montpellier*.

Seixas, F. A., Boursot, P., & Melo-Ferreira, J. (2018). The genomic impact of historical hybridization with massive mitochondrial DNA introgression. *Genome Biology*, *19*(1), 1-20.

Sheppard, S. K., McCarthy, N. D., Jolley, K. A., & Maiden, M. C. (2011). Introgression in the genus Campylobacter: generation and spread of mosaic alleles. *Microbiology*, *157*(Pt 4), 1066.

Simpson, G. G. (1951). The species concept. *Evolution*, *5*(4), 285-298.

Singh, B. (2012). Concepts of species and modes of speciation. *Current Science*, 784-790.

Smith, A. T., Johnston, C. H., Alves, P. C., & Hackländer, K. (2018). *Lagomorphs: pikas, rabbits, and hares of the world*. JHU Press.

Smith, S., Sandoval-Castellanos, E., Lagerholm, V. K., Napierala, H., Sablin, M., Von Seth, J., Fladerer, F. A., Germonpré, M., Wojtal, P., & Miller, R. (2017). Nonreceding hare lines: genetic continuity since the Late Pleistocene in European mountain hares (Lepus timidus). *Biological Journal of the Linnean Society*, *120*(4), 891-908.

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, *14*(6), 404-414.

Stamatis, C., Suchentrunk, F., Moutou, K. A., Giacometti, M., Haerer, G., Djan, M., Vapa, L., Vukovic, M., Tvrtković, N., & Sert, H. (2009). Phylogeography of the brown hare (Lepus europaeus) in Europe: a legacy of south-eastern Mediterranean refugia? *Journal of Biogeography*, *36*(3), 515-528.

Stephan, W. (2016). Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular ecology*, *25*(1), 79-88.

Suvorov, A., Scornavacca, C., Fujimoto, M. S., Bodily, P., Clement, M., Crandall, K. A., Whiting, M. F., Schrider, D. R., & Bybee, S. M. (2022). Deep ancestral introgression shapes evolutionary history of dragonflies and damselflies. *Systematic Biology*, *71*(3), 526-546.

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature ecology & evolution*, *3*(2), 170-177.

Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences*, *118*(10), e2015096118.

Templeton, A. R. (2008). The reality and importance of founder speciation in evolution. *BioEssays*, *30*(5), 470-479.

Twyford, A. D., & Ennos, R. (2012). Next-generation hybridization and introgression. *Heredity*, *108*(3), 179-189.

Veller, C., Edelman, N. B., Muralidhar, P., & Nowak, M. A. (2019). Recombination, variance in genetic relatedness, and selection against introgressed DNA. *BioRxiv*, 846147.

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, *106*(supplement_1), 9939-9946.

Wang, I. J., Glor, R. E., & Losos, J. B. (2013). Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology letters*, *16*(2), 175-182.

Weir, J. T., & Price, T. D. (2011). Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. *The American Naturalist*, *177*(4), 462-469.

Werren, J. H., Nur, U., & Wu, C.-I. (1988). Selfish genetic elements. *Trends in ecology & evolution*, *3*(11), 297-302.

Werren, J. H., & Stouthamer, R. (2003). PSR (paternal sex ratio) chromosomes: the ultimate selfish genetic elements. *Genetica*, *117*(1), 85-101.

Whitney, K. D., Randell, R. A., & Rieseberg, L. H. (2010). Adaptive introgression of abiotic tolerance traits in the sunflower Helianthus annuus. *New Phytologist*, *187*(1), 230-239.

Wiens, J. J. (2004). What is speciation and how should we study it? *The American Naturalist*, *163*(6), 914-923.

Wiley, E. (1981). Remarks on Willis' species concept. *Systematic Biology*, *30*(1), 86-87.

Williams, R. C., Blanco, M. B., Poelstra, J. W., Hunnicutt, K. E., Comeault, A. A., & Yoder, A. D. (2020). Conservation genomic analysis reveals ancient introgression and declining levels of genetic diversity in Madagascar's hibernating dwarf lemurs. *Heredity*, *124*(1), 236-251.

Worsham, M. L., Julius, E. P., Nice, C. C., Diaz, P. H., & Huffman, D. G. (2017). Geographic isolation facilitates the evolution of reproductive isolation and morphological divergence. *Ecology and evolution*, *7*(23), 10278-10288.

Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. *Nature Reviews Genetics*, *5*(2), 114-122.

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of evolutionary biology*, *14*(6), 851-865.

Yamada, F., Takaki, M., & Suzuki, H. (2002). Molecular phylogeny of Japanese Leporidae, the Amami rabbit Pentalagus furnessi, the Japanese hare Lepus brachyurus, and the mountain hare Lepus timidus, inferred from mitochondrial DNA sequences. *Genes & Genetic Systems*, *77*(2), 107-116.

Yang, J., Li, W.-R., Lv, F.-H., He, S.-G., Tian, S.-L., Peng, W.-F., Sun, Y.-W., Zhao, Y.-X., Tu, X.-L., & Zhang, M. (2016). Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Molecular biology and evolution*, *33*(10), 2576-2592.

Yi, S. V. (2006). Non-adaptive evolution of genome complexity. *BioEssays*, *28*(10), 979-982.

Zachos, F. E. (2016). *Species concepts in biology* (Vol. 801). Springer.

Zhu, T., Xu, P.-Z., Liu, J.-P., Peng, S., Mo, X.-C., & Gao, L.-Z. (2014). Phylogenetic relationships and genome divergence among the AA-genome species of the genus Oryza as revealed by 53 nuclear genes and 16 intergenic regions. *Molecular phylogenetics and evolution*, *70*, 348-361.

# 7. APPENDIX

Appendix 1 – Mean depth and genotype quality (GQ) per sample for the final vcf dataset.

| Sample | Mean Depth | Mean GQ |
|---|---|---|
| **Lame** | 27.384 | 75.270 |
| **Lcas1** | 14.198 | 39.555 |
| **Lcas2** | 8.232 | 23.535 |
| **Lcas3** | 6.889 | 21.080 |
| **Lcas4** | 6.035 | 18.391 |
| **Lcas5** | 7.115 | 20.556 |
| **Lcor1** | 17.897 | 48.700 |
| **Lcor2** | 8.910 | 25.496 |
| **Lcor3** | 9.594 | 27.180 |
| **Lcor4** | 7.472 | 20.214 |
| **Lcor5** | 9.248 | 25.815 |
| **Leur1** | 7.032 | 22.863 |
| **Leur10** | 5.542 | 15.713 |
| **Leur2** | 8.201 | 27.097 |
| **Leur3** | 9.065 | 28.803 |
| **Leur4** | 7.835 | 26.189 |
| **Leur5** | 11.083 | 34.632 |
| **Leur6** | 7.514 | 20.063 |
| **Leur7** | 10.867 | 26.576 |
| **Leur8** | 10.715 | 25.678 |
| **Leur9** | 9.269 | 20.155 |
| **Lgra1** | 18.490 | 53.804 |
| **Lgra10** | 20.123 | 58.251 |
| **Lgra2** | 17.596 | 52.610 |
| **Lgra3** | 14.980 | 43.422 |
| **Lgra4** | 18.692 | 53.866 |
| **Lgra5** | 18.884 | 55.586 |
| **Lgra6** | 19.720 | 57.000 |
| **Lgra7** | 19.817 | 60.223 |
| **Lgra8** | 15.312 | 45.963 |
| **Lgra9** | 14.660 | 43.491 |
| **Ltim1** | 22.591 | 67.264 |
| **Ltim2** | 16.746 | 51.877 |
| **Ltim3** | 18.550 | 57.349 |
| **Ltim4** | 21.091 | 63.552 |

Appendix 2 - List of samples and their role in the ELAI models performed.

| Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|
| **Target/Source** | **Species** | **Sample** | **Target/Source** | **Species** | **Sample** |
| Target | *L. castroviejoi* | Lcas1 | Target | *L.corsicanus* | Lcor1 |
| Target | *L. castroviejoi* | Lcas2 | Target | *L.corsicanus* | Lcor2 |
| Target | *L. castroviejoi* | Lcas3 | Target | *L.corsicanus* | Lcor3 |
| Target | *L. castroviejoi* | Lcas4 | Target | *L.corsicanus* | Lcor4 |
| Target | *L. castroviejoi* | Lcas5 | Target | *L.corsicanus* | Lcor5 |
| Source | *L.corsicanus* | Lcor1 | Source | *L. castroviejoi* | Lcas1 |
| Source | *L.corsicanus* | Lcor3 | Source | *L. castroviejoi* | Lcas3 |
| Source | *L.corsicanus* | Lcor5 | Source | *L. castroviejoi* | Lcas5 |
| Source | *L. granatensis* | Lgra1 | Source | *L. granatensis* | Lgra1 |
| Source | *L. granatensis* | Lgra4 | Source | *L. granatensis* | Lgra4 |
| Source | *L. granatensis* | Lgra5 | Source | *L. granatensis* | Lgra5 |
| Source | *L. europaeus* | Leur6 | Source | *L. europaeus* | Leur6 |
| Source | *L. europaeus* | Leur9 | Source | *L. europaeus* | Leur9 |
| Source | *L. europaeus* | Leur10 | Source | *L. europaeus* | Leur10 |
| Source | *L. timidus* | Ltim1 | Source | *L. timidus* | Ltim1 |
| Source | *L. timidus* | Ltim3 | Source | *L. timidus* | Ltim3 |
| Source | *L. timidus* | Ltim4 | Source | *L. timidus* | Ltim4 |

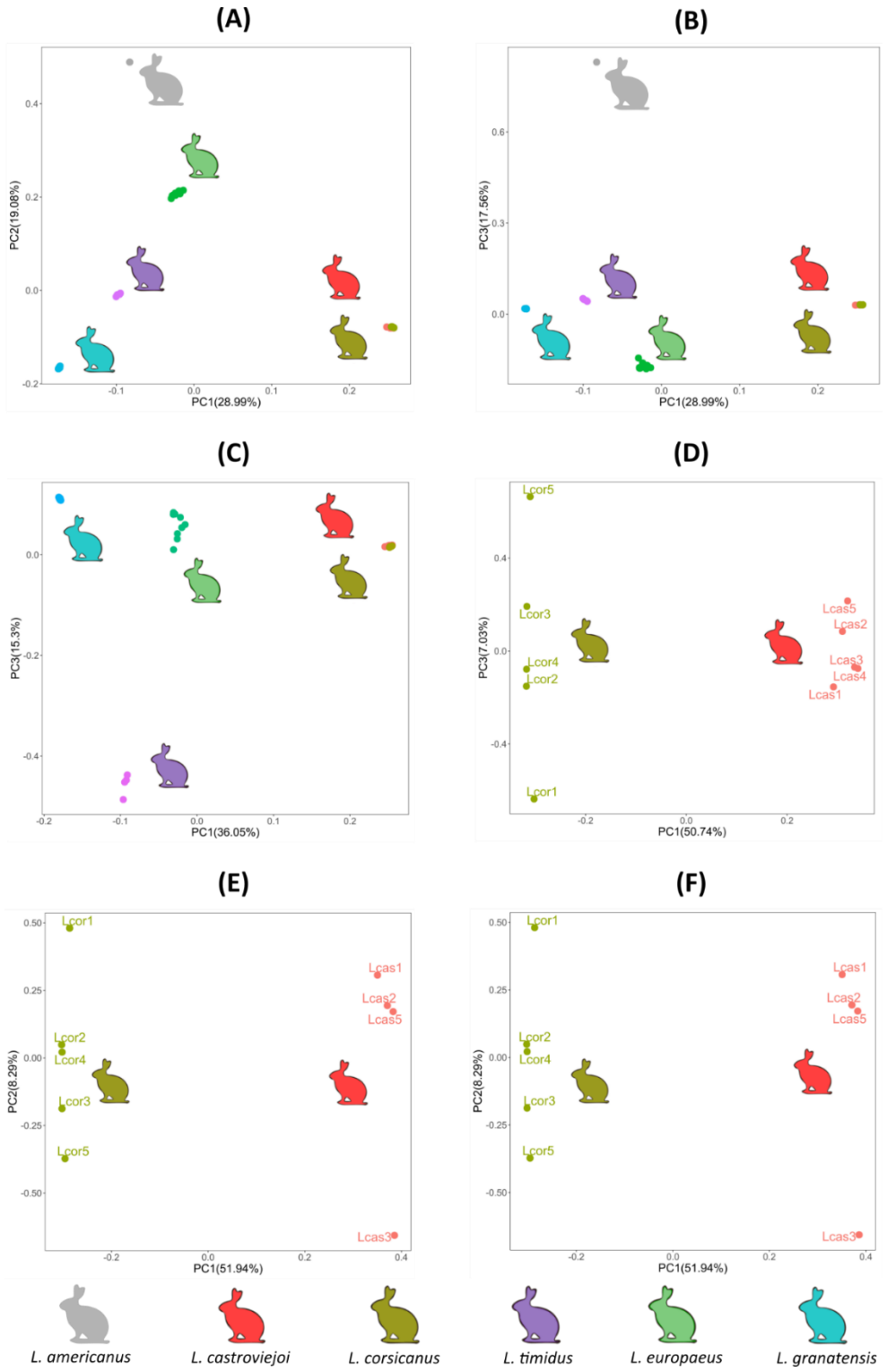
Appendix 3 - Relatedness analysis and info regarding the values.

| IND 1 | IND 2 | RELATEDNESS PHI |
|---|---|---|
| Lcas1 | Lcas2 | -0.0454 |
| Lcas1 | Lcas3 | -0.1038 |
| Lcas1 | Lcas4 | -0.1089 |
| Lcas1 | Lcas5 | -0.0233 |
| Lcas2 | Lcas3 | -0.1490 |
| Lcas2 | Lcas4 | -0.1653 |
| Lcas2 | Lcas5 | -0.0771 |
| **Lcas3** | **Lcas4** | **0.1628** |
| Lcas3 | Lcas5 | -0.1658 |
| Lcas4 | Lcas5 | -0.1733 |
| Lcor1 | Lcor2 | -0.1170 |
| Lcor1 | Lcor3 | -0.2649 |
| Lcor1 | Lcor4 | -0.2013 |
| Lcor1 | Lcor5 | -0.2132 |
| Lcor2 | Lcor3 | -0.1353 |
| Lcor2 | Lcor4 | -0.1265 |
| Lcor2 | Lcor5 | -0.1542 |
| Lcor3 | Lcor4 | -0.2025 |

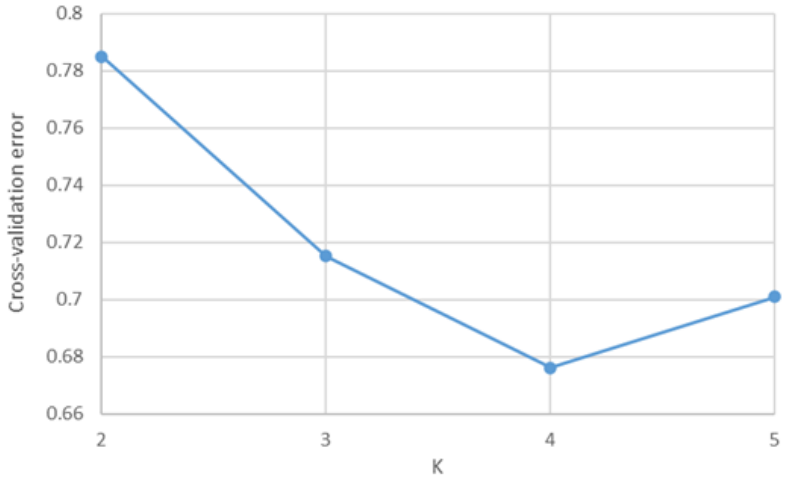
| Values | Info |
|---|---|
| >0.354 | duplicate samples/monozygotic twins |
| 0.177–0.354 | 1st degree relatives |
| 0.0884–0.177 | 2nd degree relatives |
| 0.0442–0.0884 | 3rd degree relatives |
| < 0.0442 | unrelated |

| | | |
|---|---|---|
| Lcor3 | Lcor5 | -0.175566 |
| Lcor4 | Lcor5 | -0.179298 |

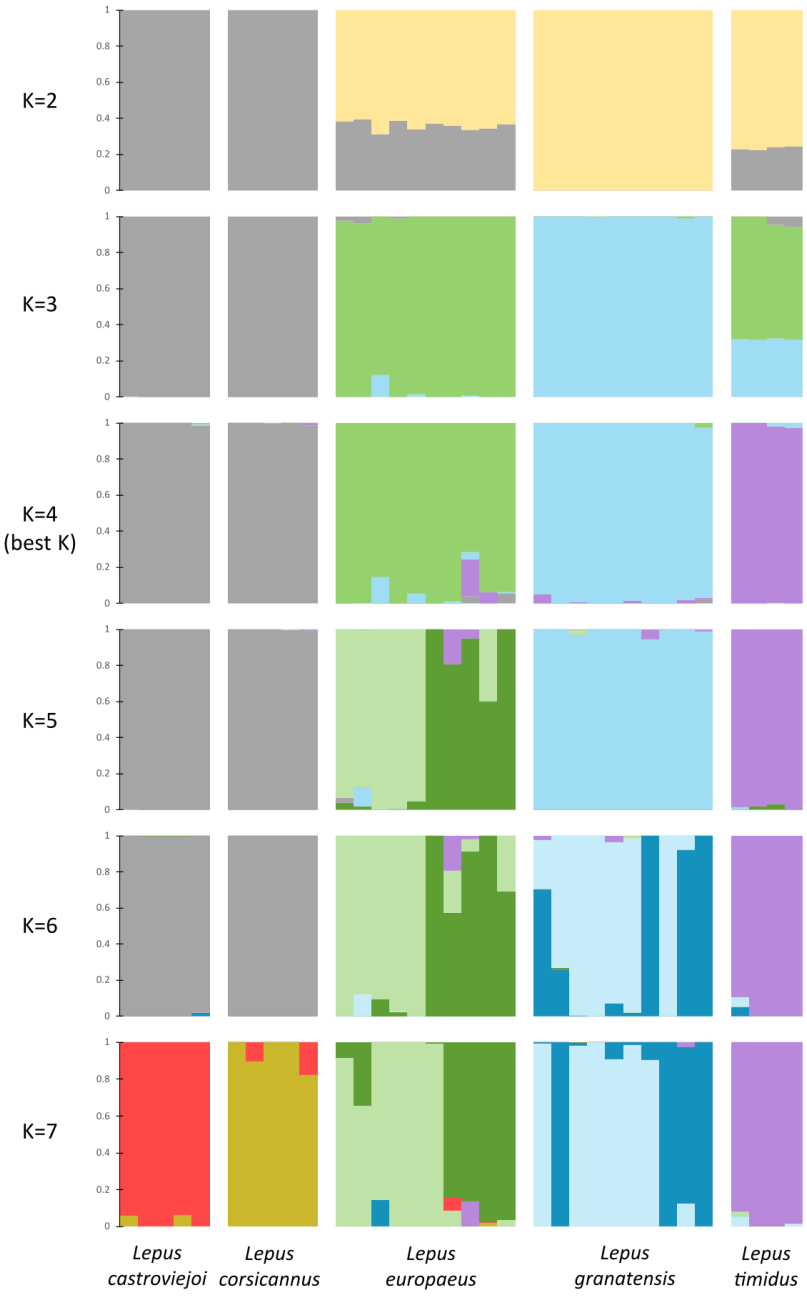
Appendix 4 - Principal Component Analysis (PCA): (A) – European species + outgroup (*L. americanus*), PC1 PC2; (B) – European Species + outgroup (*L. americanus*), PC1 PC3; (C) – European species, PC1 PC3; (D) – *L. castroviejoi* and *L. corsicanus*, PC1 PC3; (E) – *L. castroviejoi* and *L. corsicanus* without Lcas4, PC1 PC2; (F) – *L. castroviejoi* and *L. corsicanus* without Lcas4, PC1 PC3.

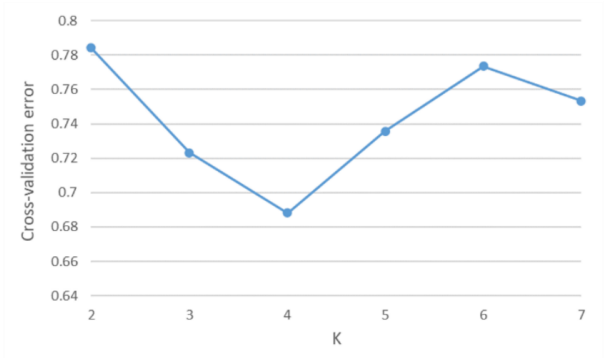


Appendix 5 - Cross-validation errors for the K values calculated in the European hare species Admixture analysis for all chromosomes.

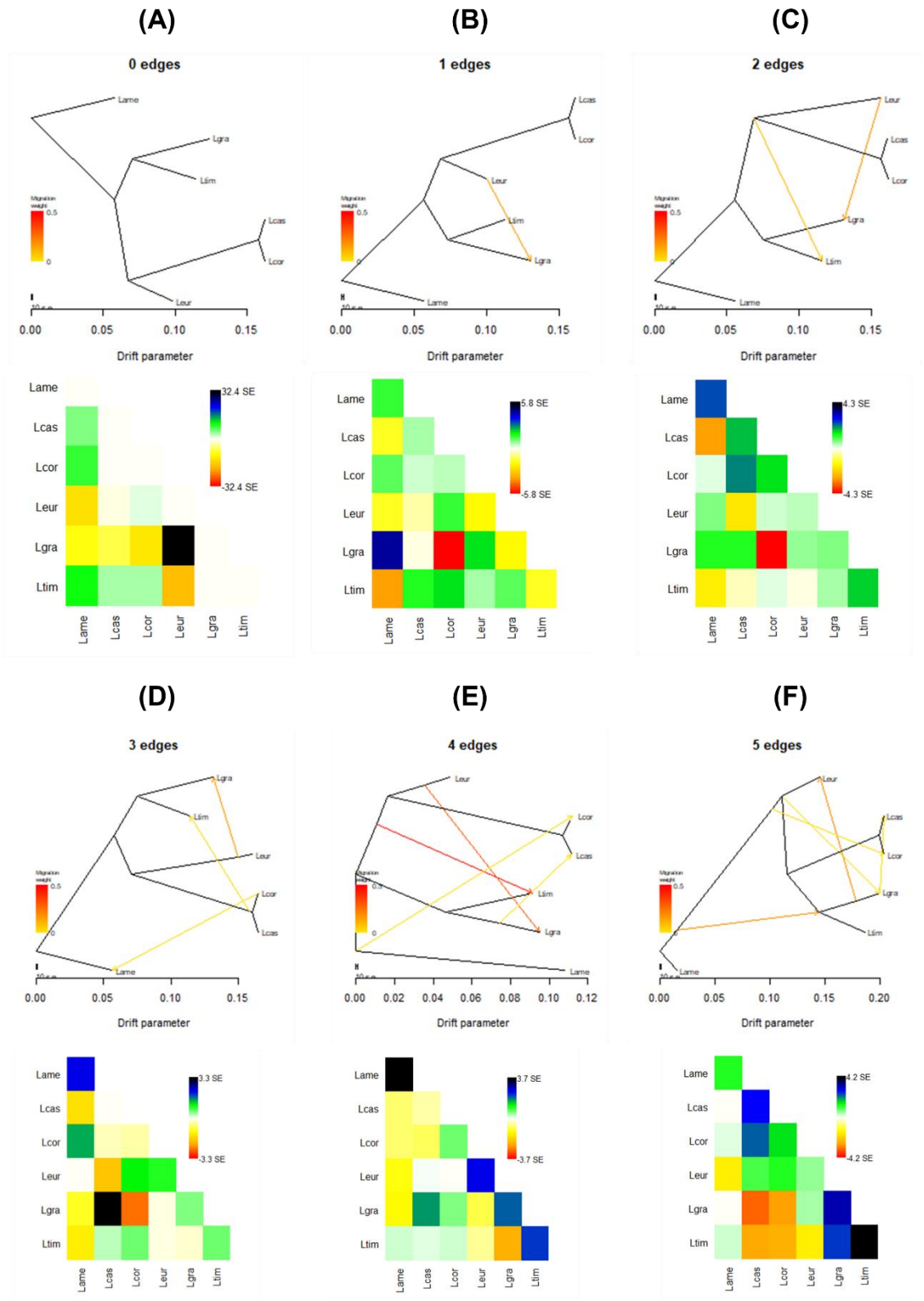


Appendix 6 - Chromosome 20 Admixture analysis for European hare species for K= 2, 3, 4, 5, 6, 7, and their cross-validation errors, inferred with ADMIXTURE.

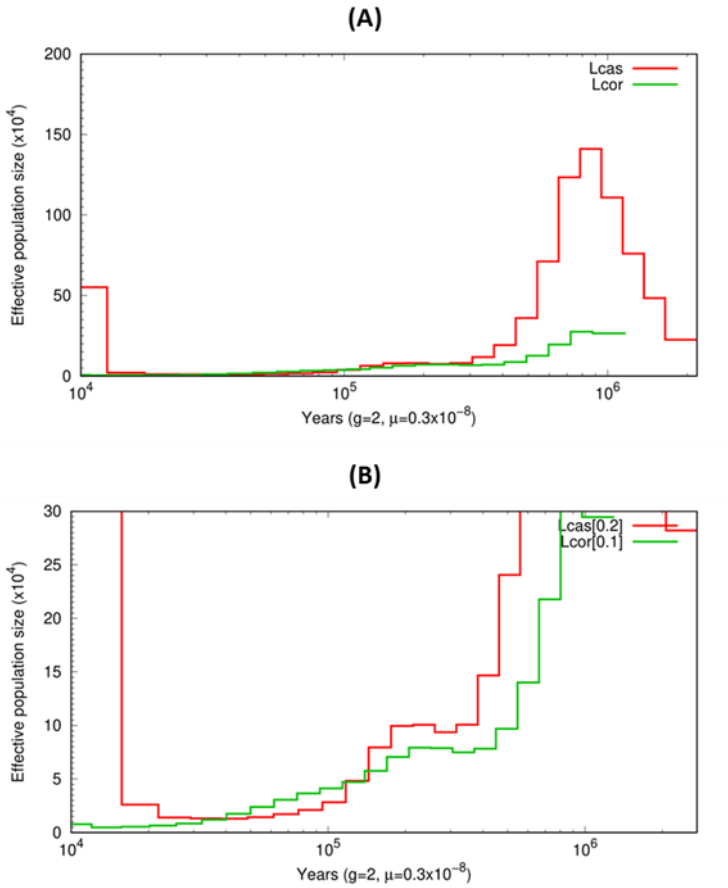


Appendix 7 - Maximum likelihood phylogeny inferred using Treemix using *L. americanus* as the outgroup. (A) – model without migrations; (B) – model with 1 migration event; (C) – model with 2 migration events; (D) – model with 3 migration events; © – model with 4 migration events; (F) model with 5 migration events.



Appendix 8 - PSMC inference of *L. castroviejoi* and *L. corsicanus* demographic profiles (A) and a zoom-in on the effective population size oscillations under $30x10^4$ (B).




Appendix 9 - Demographic parameters inferred with G-PhoCS for the history of divergence between *L. castroviejoi* (lcas) and *L. corsicanus* (lcor) for a model with and without post-split gene flow (A) and for the divergence between *L. castroviejoi* and *L. granatensis* using a model with post-split gene flow (B). Conversions of raw estimates were done by using a generation time of two years and and a mutation rate μ = 2.8 x 10-9 substitutions/site/generation. Mean values of estimated parameters are presented with 95% HPD intervals in parentheses.

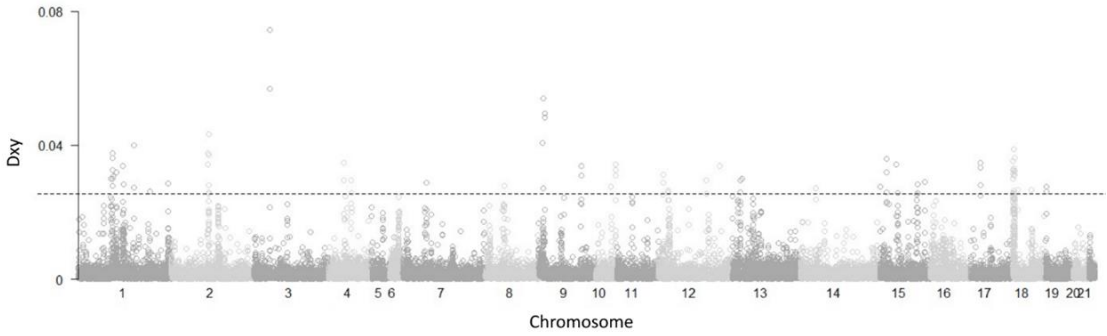
**(A)**

| G-PhoCS parameter | Demographic parameter (95% HPD interval) | |
|---|---|---|
| | **Model without gene flow** | **Model with gene flow** |
| **theta lcor** | 18 232 (14 688 - 21 876) diploid individuals | 17 304 (13 607 – 21 234) diploid individuals |
| **theta lcas** | 15 571 (12 473 – 18 748) diploid individuals | 14 705 (11 558 – 17 989) diploid individuals |
| **theta root** | 158 143 (122 179 – 218 063) diploid individuals | 158 759 (139 022 – 178 671) diploid individuals |
| **tau root** | 47 000 (30 650 – 66 493) generations | 43 214 (35 543 – 50 986) generations |
| **m lcas > lcor** | - | 0 migrants/generation |
| **m lcor > lcas** | - | 0 migrants/generation |

**(B)**

| G-PhoCS parameter | Demographic parameter (95% HPD interval) |
|---|---|
| | model with gene flow |
| **theta lgra** | 326 527 (306 674 – 347 258) diploid individuals |
| **theta lcas** | 52 857 (47 037 – 58 830) diploid individuals |
| **theta root** | 730 982 (664 230 – 801 091) diploid individuals |
| **tau root** | 455 286 (420 071 – 489 936) generations |
| **m lcor > lgra** | 6.2 E-16 (0 – 5.6 E-13) migrants/generation |

Appendix 10 - Genome scan Dxy distances based on 82,478 25kb window-based values. Dashed line marks the top 0.1% values.




Appendix 11 - Admixture $f_3$-statistics where pop A is the result of admixture between pop B and pop C. The more negative the $f_3$ value result, more likely admixture event.

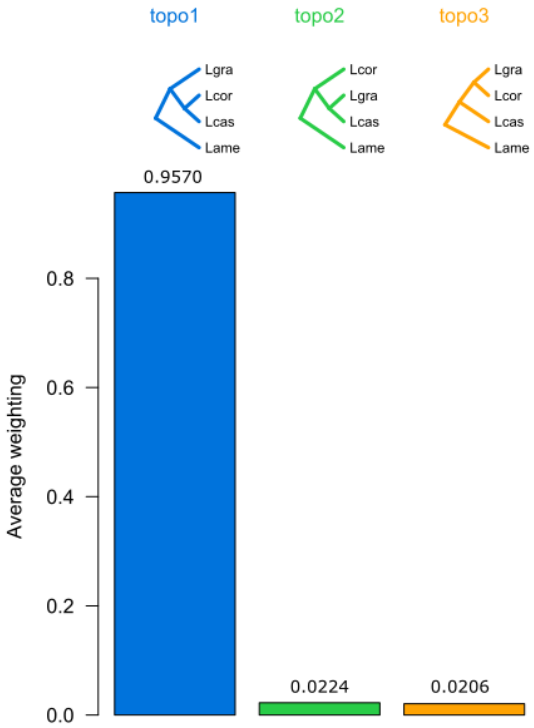
| $f_3$ | | | | | |
|---|---|---|---|---|---|
| **Pop A** | **Pop B** | **Pop C** | **$f_3$ statistics** | **S.E.** | **Z-score** |
| **Lcas** | Lgra | Lcor | 0.00669 | 9.25E-05 | 72.306 |
| **Lcas** | Ltim | Lcor | 0.00745 | 8.38E-05 | 88.841 |
| **Lcor** | Lame | Lcas | 0.00749 | 9.95E-05 | 75.322 |
| **Lcor** | Lcas | Leur | 0.00761 | 9.42E-05 | 80.805 |
| **Lcas** | Lcor | Leur | 0.00770 | 8.93E-05 | 86.132 |
| **Lcas** | Lame | Lcor | 0.00781 | 9.04E-05 | 86.415 |
| **Lcor** | Ltim | Lcas | 0.00786 | 1.10E-04 | 71.202 |
| **Lcor** | Lcas | Lgra | 0.00861 | 1.42E-04 | 60.515 |
| **Lcas** | Ltim | Leur | 0.13248 | 0.000317 | 417.738 |

| | | | | | |
|---|---|---|---|---|---|
| **Lcor** | Ltim | Leur | 0.13264 | 3.09E-04 | 429.140 |
| **Lcor** | Lame | Leur | 0.13282 | 2.80E-04 | 474.798 |
| **Lcas** | Lame | Leur | 0.13303 | 2.89E-04 | 459.690 |
| **Lcas** | Lgra | Leur | 0.14056 | 3.01E-04 | 466.644 |
| **Lcor** | Lgra | Leur | 0.14148 | 2.70E-04 | 523.640 |
| **Lcas** | Lame | Lgra | 0.14763 | 3.33E-04 | 442.876 |
| **Lcor** | Lame | Lgra | 0.14843 | 3.06E-04 | 485.283 |
| **Lcas** | Ltim | Lame | 0.14891 | 3.56E-04 | 418.341 |
| **Lcor** | Ltim | Lame | 0.14895 | 3.43E-04 | 433.891 |
| **Lcas** | Ltim | Lgra | 0.16699 | 4.01E-04 | 416.796 |
| **Lcor** | Ltim | Lgra | 0.16816 | 3.80E-04 | 442.172 |

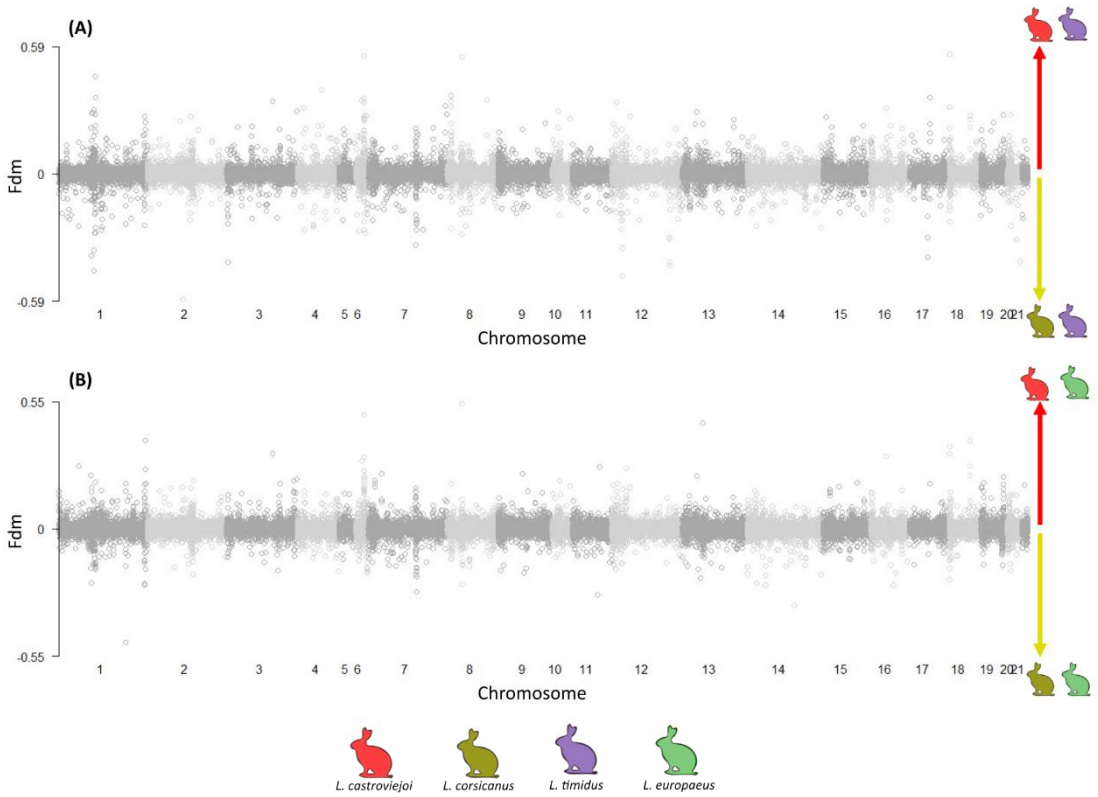
Appendix 12 - $f_4$ statistics based on the topology (A,B)(C,D). Negative values indicates gene flow between C,B or D,A. Positive values imply gene flow between A,C or B,D. Lame – *L. americanus*; Lgra – *L. granatensis*; Lcas – *L. castroviejoi*; Lcor – *L. corsicanus*; Ltim – *L. timidus*; Leur – *L. europaeus*.

| $f_4$ | | | | | | |
|---|---|---|---|---|---|---|
| **Pop A** | **Pop B** | **Pop C** | **Pop D** | **$f_4$ statistics** | **S.E.** | **Z-score** |
| Lame | Lgra | Lcas | Lcor | -0.00112 | 7.86E-05 | -14.29 |
| Ltim | Lgra | Lcas | Lcor | -0.00076 | 6.07E-05 | -12.51 |
| Lame | Leur | Lcas | Lcor | -0.00012 | 4.25E-05 | -2.75 |
| Ltim | Leur | Lcas | Lcor | 0.00025 | 5.06E-05 | 4.89 |
| Ltim | Lame | Lcas | Lcor | 0.00037 | 5.14E-05 | 7.10 |

Appendix 13 - TWISST analysis on dataset consisting on *L. castroviejoi*, *L. corsicanus*, *L. granatensis*, *L. americanus*.




Appendix 14 - Genome wide $f_{dM}$ values for the model P1 – *L. corsicanus*, P2 – *L. castroviejoi*, O – *L. americanus*, and two P3 tested: (A) P3 – *L. timidus* and (B) *L. europaeus*. Positive values suggest geneflow between P2 and P3, while negative values indicate gene flow P1 and P3.



Appendix 15 - Elai ancestry proportions per chromosome, and the number of used SNPs per chromosome.

| Chr | Ancestry | | | | Nº of SNPs |
|---|---|---|---|---|---|
| | *L. corsicanus* | *L. granatensis* | *L. europaeus* | *L. timidus* | |
| chr1 | 0.99053 | 0.00720 | 0.00027 | 0.00180 | **1 869 296** |
| chr2 | 0.99240 | 0.00647 | 0.00000 | 0.00113 | **1 678 349** |
| chr3 | 0.99553 | 0.00447 | 0.00000 | 0.00007 | **1 536 998** |
| chr4 | 0.99293 | 0.00600 | 0.00000 | 0.00100 | **856 761** |
| chr5 | 0.99407 | 0.00580 | 0.00000 | 0.00020 | **382 098** |
| chr6 | 0.96120 | 0.01620 | 0.00293 | 0.01967 | **297 938** |
| chr7 | 0.98387 | 0.01573 | 0.00000 | 0.00060 | **1 632 258** |
| chr8 | 0.98900 | 0.01033 | 0.00000 | 0.00080 | **1 151 088** |
| chr9 | 0.98920 | 0.01027 | 0.00000 | 0.00040 | **1 213 303** |
| chr10 | 0.97580 | 0.02313 | 0.00000 | 0.00113 | **505 446** |
| chr11 | 0.99480 | 0.00487 | 0.00000 | 0.00040 | **848 637** |
| chr12 | 0.98607 | 0.00867 | 0.00387 | 0.00113 | **1 478 759** |
| chr13 | 0.99000 | 0.00820 | 0.00007 | 0.00160 | **1 365 843** |
| chr14 | 0.99647 | 0.00353 | 0.00000 | 0.00000 | **1 528 081** |
| chr15 | 0.99027 | 0.00947 | 0.00000 | 0.00033 | **1 028 676** |
| chr16 | 0.99567 | 0.00360 | 0.00000 | 0.00087 | **845 652** |
| chr17 | 0.99400 | 0.00480 | 0.00000 | 0.00107 | **784 310** |
| chr18 | 0.99133 | 0.00853 | 0.00000 | 0.00000 | **715 408** |
| chr19 | 0.99360 | 0.00500 | 0.00127 | 0.00013 | **566 045** |
| chr20 | 0.98340 | 0.01587 | 0.00060 | 0.00000 | **323 900** |
| chr21 | 0.99373 | 0.00620 | 0.00007 | 0.00000 | **144 657** |

Appendix 16 - Number of *L. europaeus* and *L. timidus* Elai inferred junctions per sample. tim: *L. timidus*; eur: *L. europaeus*; gra: *L. granatensis*; cor: *L. corsicanus*; cas: *L. castroviejoi*.

| sample | ***L.timidus* junctions** | | | | | | |
|---|---|---|---|---|---|---|---|
| | tim/cor | % tim/cor | tim/gra | % tim/gra | tim/eur | % tim/eur | **total** |
| **Lcas1** | 164 | 85.0 | 29 | 15.0 | 0 | 0 | **193** |
| **Lcas2** | 117 | 90.7 | 12 | 9.3 | 0 | 0 | **129** |
| **Lcas3** | 98 | 83.8 | 19 | 16.2 | 0 | 0 | **117** |
| **Lcas4** | 91 | 82.7 | 19 | 17.3 | 0 | 0 | **110** |
| **Lcas5** | 96 | 94.1 | 6 | 5.9 | 0 | 0 | **102** |

| sample | *L.europaeus* junctions | | | | | | |
|---|---|---|---|---|---|---|---|
| | eur/cor | % eur/cor | eur/gra | % eur/gra | eur/tim | % eur/tim | **total** |
| **Lcas1** | 50 | 90.9 | 5 | 9.1 | 0 | 0 | **55** |
| **Lcas2** | 29 | 96.7 | 1 | 3.3 | 0 | 0 | **30** |
| **Lcas3** | 23 | 92.0 | 2 | 8.0 | 0 | 0 | **25** |
| **Lcas4** | 28 | 96.6 | 1 | 3.4 | 0 | 0 | **29** |
| **Lcas5** | 27 | 96.4 | 1 | 3.6 | 0 | 0 | **28** |

Appendix 17 – Elai ancestry proportions with *L. corsicanus* as target.

| ind | *L. castroviejoi* | *L. granatensis* | *L. europaeus* | *L. timidus* |
|---|---|---|---|---|
| **Lcor1** | 99.888% | 0.013% | 0.016% | 0.039% |
| **Lcor2** | 99.903% | 0.015% | 0.015% | 0.033% |
| **Lcor3** | 99.894% | 0.019% | 0.017% | 0.034% |
| **Lcor4** | 99.859% | 0.016% | 0.056% | 0.033% |
| **Lcor5** | 99.896% | 0.020% | 0.017% | 0.030% |
| **Overall** | **99.888%** | **0.017%** | **0.024%** | **0.034%** |

Appendix 18 - List of genes inspected in the Enrichment analysis and their inferred function.

| Chr | Start | End | Gene name | Function |
|---|---|---|---|---|
| 1 | 15441317 | 15579207 | TSTD2 | |
| 1 | 15546196 | 15702310 | TDRD7 | biological regulation, regulation of biological process |
| 1 | 15634894 | 15687161 | ENSOCUG00000029006 | |
| 1 | 72045287 | 72127497 | NAA35 | biological regulation, regulation of biological process, regulation of cellular process |
| 1 | 72131198 | 72165745 | GOLM1 | biological regulation, regulation of biological process |
| 1 | 78359996 | 78439013 | SEMA4D | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 79041952 | 79049783 | ENSOCUG00000022990 | biological regulation |
| 1 | 79074761 | 79146826 | SPIN1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 98635213 | 98636307 | ENSOCUG00000038181 | |
| 1 | 100570515 | 100602524 | ENSOCUG00000038611 | |
| 1 | 147119401 | 147120476 | ENSOCUG00000028163 | |
| 1 | 187613766 | 187614669 | ENSOCUG00000035290 | |
| 1 | 187693360 | 187694262 | ENSOCUG00000030075 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187725519 | 187726448 | ENSOCUG00000024711 | biological regulation, regulation of biological process, regulation of cellular |

| | | | | |
|---|---|---|---|---|
| | | | | process, cell communication, signaling, signal transduction |
| 1 | 187738146 | 187739075 | ENSOCUG00000026679 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187759150 | 187760076 | ENSOCUG00000008122 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187790113 | 187791042 | ENSOCUG00000005268 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187803862 | 187804764 | ENSOCUG00000038037 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187812242 | 187813171 | ENSOCUG00000039133 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187828812 | 187829741 | ENSOCUG00000032013 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187842009 | 187842938 | ENSOCUG00000021660 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187854120 | 187855049 | ENSOCUG00000038724 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187904739 | 187905656 | ENSOCUG00000033859 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 187925966 | 187927503 | ENSOCUG00000024521 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 189063513 | 189064460 | ENSOCUG00000036431 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 194587094 | 194599844 | ENSOCUG00000038329 | |
| 1 | 194597290 | 194599844 | ENSOCUG00000039622 | |
| 1 | 194599865 | 194731670 | KDM2A | biological regulation, regulation of biological process, regulation of cellular process |
| 1 | 194743802 | 194746827 | ENSOCUG00000038516 | |
| 1 | 194754486 | 194756335 | ENSOCUG00000029982 | |
| 1 | 194758430 | 194759825 | ENSOCUG00000006388 | |
| 1 | 194761747 | 194767569 | SSH3 | biological regulation, regulation of biological process, regulation of cellular process |
| 1 | 194770434 | 194835787 | RAD9A | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 1 | 194802505 | 194804152 | ENSOCUG00000004005 | |

| | | | | |
|---|---|---|---|---|
| 1 | 194811735 | 194819288 | CLCF1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 2 | 308930 | 317348 | LYAR | biological regulation, regulation of biological process, regulation of cellular process |
| 2 | 94508442 | 94611216 | ENSOCUG00000028202 | |
| 2 | 98129481 | 98247972 | ENSOCUG00000029292 | |
| 2 | 98223558 | 98454000 | ENSOCUG00000027394 | |
| 2 | 98926095 | 98926409 | ENSOCUG00000025374 | |
| 2 | 98961851 | 98993640 | ENSOCUG00000006530 | |
| 2 | 100020636 | 100144983 | REEP1 | |
| 2 | 100147102 | 100162537 | MRPL35 | |
| 2 | 100167214 | 100212489 | IMMT | biological regulation |
| 2 | 100217116 | 100247027 | ENSOCUG00000010161 | |
| 2 | 100218424 | 100218558 | SNORD94 | |
| 2 | 100247451 | 100326763 | POLR1A | biological regulation, regulation of biological process, regulation of cellular process |
| 2 | 100440078 | 100505486 | ST3GAL5 | |
| 2 | 100543944 | 100553597 | ATOH8 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 3 | 32469752 | 32481040 | ENSOCUG00000035396 | |
| 3 | 32483817 | 32584463 | ENSOCUG00000024432 | |
| 3 | 32505854 | 32510754 | ENSOCUG00000028024 | |
| 3 | 32574069 | 32585184 | ZNF300 | biological regulation, regulation of biological process, regulation of cellular process |
| 3 | 54822318 | 54831487 | ENSOCUG00000031970 | |
| 3 | 54842765 | 54910760 | CPEB4 | biological regulation, regulation of biological process, regulation of cellular process |
| 3 | 146490086 | 146740453 | ADCY8 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 3 | 148923207 | 148936788 | CCN4 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 4 | 9109254 | 9148757 | ENSOCUG00000023941 | |
| 4 | 19488979 | 19532701 | ENSOCUG00000030426 | |
| 4 | 22560026 | 22963461 | TASP1 | biological regulation, regulation of biological process, regulation of cellular process |
| 4 | 24276445 | 24309484 | BTBD3 | |
| 4 | 28349814 | 28509022 | HAO1 | |
| 4 | 37189666 | 37226242 | SP1 | biological regulation, regulation of biological process, regulation of cellular process |
| 4 | 37775877 | 37775984 | MIR196A2 | |
| 4 | 37784456 | 37787223 | HOXC9 | biological regulation, regulation of biological process, regulation of cellular process |

| | | | | |
|---|---|---|---|---|
| 4 | 37793427 | 37795810 | HOXC8 | biological regulation, regulation of biological process, regulation of cellular process |
| 4 | 38080742 | 38101057 | COPZ1 | |
| 4 | 38111631 | 38112818 | ENSOCUG00000010753 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 4 | 38121032 | 38141067 | ZNF385A | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 4 | 38207618 | 38289482 | NCKAP1L | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 4 | 38294892 | 38323526 | PDE1B | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 4 | 38323744 | 38332651 | ENSOCUG00000013555 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 4 | 43108831 | 43197947 | C12orf56 | |
| 4 | 48184819 | 48214405 | CPSF6 | biological regulation, regulation of biological process, regulation of cellular process |
| 4 | 48613716 | 48618906 | FRS2 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 4 | 49389824 | 49465807 | KCNMB4 | biological regulation, cell communication, signaling, |
| 4 | 51397572 | 51830973 | TRHDE | |
| 4 | 60062265 | 60210781 | LIN7A | biological regulation, cell communication, signaling, |
| 4 | 60587600 | 61132121 | PPFIA2 | biological regulation, regulation of biological process, regulation of cellular process |
| 4 | 61629400 | 61741612 | METTL25 | |
| 4 | 62760057 | 62775626 | ENSOCUG00000035502 | |
| 4 | 67196551 | 67242936 | TMTC3 | biological regulation, regulation of biological process, regulation of cellular process |
| 4 | 67493134 | 67575339 | KITLG | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 4 | 75474542 | 75511662 | CDK17 | |
| 4 | 78037409 | 79260279 | ANKS1B | |
| 5 | 13310566 | 13386213 | PSME3IP1 | biological regulation, regulation of biological process, regulation of cellular process |
| 5 | 13386255 | 13439299 | RSPRY1 | |
| 5 | 13449903 | 13471380 | PLLP | |
| 5 | 22005555 | 22036447 | ENSOCUG00000026523 | |
| 5 | 22041880 | 22048898 | CMTM3 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |

| | | | | |
|---|---|---|---|---|
| 5 | 22055909 | 22167329 | DYNC1LI2 | |
| 5 | 22171617 | 22212781 | TERB1 | |
| 5 | 22204313 | 22204506 | U3 | |
| 5 | 22222310 | 22247009 | NAE1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 6 | 1401914 | 1769725 | RBFOX1 | biological regulation, regulation of biological process, regulation of cellular process |
| 6 | 3975325 | 4009191 | ENSOCUG00000034442 | |
| 6 | 5390743 | 5869252 | SNX29 | |
| 6 | 8547405 | 8640042 | VPS35L | |
| 6 | 8644270 | 8663005 | CCP110 | biological regulation, regulation of biological process, regulation of cellular process |
| 6 | 13887348 | 13974482 | USP31 | |
| 7 | 1382865 | 1528527 | PTN | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 7 | 1578626 | 2071619 | DGKI | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 7 | 9252482 | 9334638 | COPG2 | |
| 7 | 11213822 | 11547011 | ENSOCUG00000014032 | |
| 7 | 11630125 | 12818701 | LRGUK | |
| 7 | 14474543 | 14551958 | KLHDC10 | |
| 7 | 14551979 | 14600560 | ZC3HC1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 7 | 15080461 | 15128956 | STRIP2 | biological regulation, regulation of biological process |
| 7 | 67114365 | 68644978 | DPP10 | biological regulation, regulation of biological process, regulation of cellular process |
| 7 | 103860153 | 103987125 | GRB14 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 7 | 104038187 | 104251392 | COBLL1 | |
| 7 | 104321307 | 104385124 | SLC38A11 | |
| 7 | 104351687 | 104351816 | ENSOCUG00000019821 | |
| 7 | 104491906 | 104607488 | SCN3A | voltage-gated sodium channel complex |
| 7 | 104665799 | 104796938 | SCN2A | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction, voltage-gated sodium channel complex |
| 7 | 104824257 | 104850959 | ENSOCUG00000032787 | |
| 7 | 104892048 | 105087752 | CSRNP3 | biological regulation, regulation of biological process, regulation of cellular process |
| 7 | 105115129 | 105115596 | ENSOCUG00000022341 | |
| 7 | 105155110 | 105182854 | GALNT3 | |

| | | | | |
|---|---|---|---|---|
| 7 | 105247455 | 105396969 | TTC21B | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 7 | 105415056 | 105494742 | SCN1A | biological regulation, cell communication, signaling, voltage-gated sodium channel complex |
| 7 | 105535714 | 105535820 | U6 | |
| 8 | 6368815 | 6547738 | KIF21A | |
| 8 | 8903444 | 9118144 | BICD1 | biological regulation, regulation of biological process, regulation of cellular process |
| 8 | 9219782 | 9249780 | RESF1 | |
| 8 | 9272067 | 9272606 | ENSOCUG00000033986 | |
| 8 | 9423421 | 9470014 | AMN1 | |
| 8 | 9454712 | 9454828 | U5 | |
| 8 | 9480156 | 9484835 | ENSOCUG00000023987 | |
| 8 | 9538809 | 9741935 | DENND5B | biological regulation, regulation of biological process |
| 8 | 37314705 | 37492177 | KDM5A | biological regulation, regulation of biological process, regulation of cellular process |
| 8 | 37503424 | 37503678 | ENSOCUG00000023846 | |
| 8 | 37509673 | 37613485 | IQSEC3 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 9 | 1885725 | 2145078 | ENSOCUG00000035357 | |
| 9 | 3603317 | 3925918 | MYRIP | |
| 9 | 3899728 | 3901720 | ENSOCUG00000029704 | |
| 9 | 6755107 | 6847682 | ENSOCUG00000036497 | |
| 9 | 6938543 | 7007704 | KBTBD12 | |
| 9 | 7034415 | 7072608 | SEC61A1 | |
| 9 | 7044984 | 7185439 | RUVBL1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 9 | 87502229 | 87513280 | ATP5F1A | |
| 9 | 87517360 | 87536403 | HAUS1 | |
| 9 | 93286394 | 94114886 | DCC | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling |
| 10 | 15934034 | 16473294 | PDE1C | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 10 | 19832513 | 19882325 | KIAA0895 | |
| 10 | 21116879 | 21121656 | GPR141 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 10 | 21147133 | 21148022 | ENSOCUG00000024486 | |
| 10 | 21176130 | 21219208 | ENSOCUG00000015861 | |
| 10 | 21235086 | 21243552 | SFRP4 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |

| | | | | |
|---|---|---|---|---|
| 10 | 21247907 | 21270997 | EPDR1 | |
| 10 | 21480758 | 21481093 | ENSOCUG00000036137 | |
| 11 | 40285367 | 40739663 | FBXL7 | |
| 11 | 56206706 | 56208844 | ENSOCUG00000029955 | |
| 11 | 56221803 | 56227231 | RAD1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 11 | 56320259 | 56362897 | AGXT2 | biological regulation, regulation of biological process, regulation of cellular process |
| 11 | 56363806 | 56565013 | PRLR | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 12 | 9745147 | 9746083 | ENSOCUG00000022359 | |
| 12 | 23019284 | 23048090 | ENSOCUG00000006769 | |
| 12 | 23039791 | 23041269 | ENSOCUG00000025822 | |
| 12 | 39165077 | 39169010 | ENSOCUG00000035733 | |
| 12 | 39171952 | 39172977 | ENSOCUG00000032803 | |
| 12 | 39173820 | 39175967 | ENSOCUG00000039348 | |
| 12 | 72763182 | 72921498 | ENSOCUG00000032043 | |
| 12 | 75957741 | 75966862 | SRSF12 | biological regulation, regulation of biological process, regulation of cellular process |
| 12 | 126999350 | 127245281 | MAP3K5 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 12 | 127368196 | 127406612 | SLC35D3 | biological regulation, regulation of biological process, regulation of cellular process |
| 13 | 7001249 | 7137678 | CEP350 | |
| 13 | 15260025 | 16352709 | DPP6 | biological regulation, regulation of biological process, regulation of cellular process |
| 13 | 31095033 | 31311165 | ATF6 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 32394249 | 32507941 | ENSOCUG00000022552 | |
| 13 | 32433771 | 32470445 | SLAMF1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 34654583 | 34655527 | ENSOCUG00000001639 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 34670818 | 34678508 | ENSOCUG00000022281 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 34954808 | 35073530 | ENSOCUG00000008307 | |
| 13 | 37158060 | 37335037 | ENSOCUG00000034353 | |
| 13 | 37159192 | 37374813 | ASH1L | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |

| | | | | |
|---|---|---|---|---|
| 13 | 41805373 | 41821930 | ENSOCUG00000022805 | |
| 13 | 47077538 | 47138353 | VTCN1 | biological regulation, regulation of biological process, regulation of cellular process |
| 13 | 47169788 | 47178215 | TRIM45 | |
| 13 | 51967998 | 51969299 | ENSOCUG00000023372 | |
| 13 | 66456505 | 67430660 | DPYD | |
| 13 | 75695873 | 75851300 | KYAT3 | |
| 13 | 75844666 | 75999099 | PKN2 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 77712346 | 77910608 | ENSOCUG00000010672 | |
| 13 | 80727235 | 80949545 | PRKACB | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 87597110 | 87703434 | MIGA1 | |
| 13 | 87714216 | 87784995 | USP33 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 99337246 | 99424659 | DNAI4 | |
| 13 | 101596434 | 101674992 | RAVER2 | |
| 13 | 103843429 | 104078176 | DOCK7 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 113052119 | 113282809 | GLIS1 | biological regulation, regulation of biological process, regulation of cellular process |
| 13 | 113322494 | 113329720 | ENSOCUG00000001341 | biological regulation, regulation of biological process, regulation of cellular process |
| 13 | 113374888 | 113375898 | ENSOCUG00000030578 | |
| 13 | 113475513 | 113539335 | LRP8 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 113549246 | 113556839 | MAGOH | biological regulation, regulation of biological process, regulation of cellular process |
| 13 | 113553616 | 113553722 | U6 | |
| 13 | 113560288 | 113566628 | CZIB | |
| 13 | 113576071 | 113595304 | CPT2 | biological regulation, regulation of biological process |
| 13 | 113579850 | 113579949 | U6 | |
| 13 | 114146858 | 114287036 | TUT4 | |
| 13 | 124058910 | 124223727 | CCDC30 | |
| 13 | 124222520 | 124244815 | ZMYND12 | |
| 13 | 124261082 | 124290797 | RIMKLA | |
| 13 | 129356566 | 129370185 | SRSF10 | biological regulation, regulation of biological process, regulation of cellular process |
| 13 | 129373633 | 129376775 | ENSOCUG00000038691 | biological regulation, regulation of biological process |
| 13 | 131935886 | 132330532 | EIF4G3 | |

| | | | | |
|---|---|---|---|---|
| 13 | 132321361 | 132321521 | U1 | |
| 13 | 135109985 | 135262202 | PUM1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 13 | 135244323 | 135244407 | ENSOCUG00000021668 | |
| 13 | 136462610 | 136587222 | EPB41 | biological regulation, regulation of biological process, regulation of cellular process |
| 13 | 137721511 | 137752045 | FAM76A | |
| 14 | 691101 | 762004 | PIK3R4 | biological regulation, regulation of biological process, regulation of cellular process, cell communication |
| 14 | 737715 | 737842 | 5S_rRNA | |
| 14 | 761497 | 934532 | COL6A6 | |
| 14 | 2341676 | 2648399 | RFTN1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 15 | 4030371 | 4238133 | CSGALNACT1 | |
| 15 | 12609533 | 12611770 | ENSOCUG00000035083 | |
| 15 | 12618216 | 12622501 | ENSOCUG00000031102 | |
| 15 | 18173522 | 18459851 | SLC10A7 | biological regulation |
| 15 | 20734409 | 20789695 | FREM3 | |
| 15 | 20841758 | 20883541 | SMARCA5 | biological regulation, regulation of biological process, regulation of cellular process, , , |
| 15 | 20961381 | 21092482 | GAB1 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 15 | 43060237 | 43171532 | ENSOCUG00000038394 | |
| 17 | 80340533 | 80519441 | ATP10A | biological regulation, regulation of biological process, regulation of cellular process |
| 18 | 3020835 | 3026299 | ZNF25 | biological regulation, regulation of biological process, regulation of cellular process |
| 18 | 3084222 | 3101792 | ENSOCUG00000029543 | biological regulation, regulation of biological process, regulation of cellular process |
| 18 | 3172108 | 3247233 | ENSOCUG00000029443 | biological regulation, regulation of biological process, regulation of cellular process |
| 18 | 6522135 | 6591232 | ZFAND4 | |
| 18 | 7460895 | 8191494 | GRID1 | |
| 18 | 42383100 | 42989156 | ENSOCUG00000015280 | |
| 18 | 61065683 | 61260227 | NRAP | |
| 19 | 4124261 | 4158706 | LRRC75A | |
| 19 | 4160060 | 4190795 | ENSOCUG00000033937 | |
| 19 | 31133263 | 31142229 | ENSOCUG00000011338 | |
| 19 | 43820936 | 43825782 | TMEM106A | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 19 | 43829927 | 43835288 | ENSOCUG00000033652 | |
| 19 | 43837731 | 43837921 | U2 | |

| | | | | |
|---|---|---|---|---|
| 19 | 43838409 | 43864540 | ENSOCUG00000030666 | |
| 19 | 43838431 | 43838537 | U6 | |
| 19 | 43849321 | 43849498 | U2 | |
| 19 | 43990472 | 44044059 | DHX8 | |
| 19 | 51999202 | 52088717 | RGS9 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 19 | 52117642 | 52119972 | ENSOCUG00000036189 | |
| 19 | 52119221 | 52160069 | GNA13 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 19 | 56088008 | 56149875 | ENSOCUG00000031100 | |
| 20 | 8972192 | 9159086 | ENSOCUG00000015859 | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |
| 20 | 10905320 | 10996757 | RHOJ | biological regulation, regulation of biological process, regulation of cellular process, cell communication, signaling, signal transduction |