Using genomic tools to understand species differentiation and admixture in hares and mice

João Pedro Nogueira Marques

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

En génétique et génomique

École doctorale GAIA

Unité de recherche ISEM (Institut des Sciences de l'Evolution Montpellier)

En partenariat international avec Faculdade de Ciências da Universidade do Porto, Portugal

Using genomic tools to understand species differentiation and admixture in hares and mice

Utilisation des outils génomiques pour comprendre la différenciation et le mélange des espèces chez les lièvres et les souris

Présentée par João Pedro MARQUES

Le 19 avril 2022

Sous la direction de Pierre BOURSOT et José MELO-FERREIRA

Devant le jury composé de

Stéphane BOISSINOT, Professeur, NYU Abu Dhabi	Présidente du jury et Rapporteur
Lounès CHIKHI, Directeur de recherche, CNRS	Rapporteur
Pierre BOURSOT, Directeur de recherche, ISEM - CNRS	Directeur de thèse
Paulo Célio ALVES, Maître de conférences (Equivalent), BIOPOLIS et Universidade do Porto	Membre du jury
Pierre-André CROCHET, Directeur de recherche, CEFE - CNRS	Membre du jury
Catarina PINHO, chargé de recherche, CIBIO-BIOPOLIS et Universidade do Porto	Membre du jury
Vítor SOUSA, Maître de conférences (Equivalent), CE3C et Universidade de Lisboa	Membre du jury

UNIVERSITÉ DE MONTPELLIER

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

PhD

PhD

3.º CICLO

FCUP
UM
CIBIO
ISEM
2022

U.PORTO

Using genomic tools to understand species differentiation and admixture in hares and mice

João Pedro Nogueira Marques

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

UNIVERSITÉ
DE MONTPELLIER

Using genomic tools to understand species differentiation and admixture in hares and mice

João Pedro Nogueira Marques

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto e
Université Montpellier
Biodiversidade, Genética e Evolução
2022

PhD

PhD

Foreword

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

The candidate was financially supported by Fundação para a Ciência e a Tecnologia (FCT) through the attribution of a PhD fellowship (SFRH/BD/115089/2016), funded by POPH-QREN funds from the European Social Fund. Support was also obtained from the research project HybridChange (reference PTDC/BIA-EVL/1307/2020) funded by national funds through FCT. Additional support was obtained from the Portugal–France bilateral cooperation programs PESSOA (FCT in Portugal; Ministère de l'Europe et des Affaires etrangères (MEAE) and Ministère de l'Education nationale, de l'Enseignement sup erieur et de la Recherche (MESRI) in France), and Cotutelles/Cotutelas PESSOA (FCT through Conselho de Reitores das Universidades Portuguesas in Portugal; MEAE and MESRI through Conference des Présidents d'Université in France). Support was also obtained from French CNRS-InEE 5institut Ecologie et Environnement) through the LIA (Laboratoire International Associé), then IRP (International Research Project) "Biodiversity and Evolution". Candidate training was supported by the two doctoral schools: GAIA - UM and BioDiv – FCUP, and additionally by Universitat Leipzig to attend the "6th Programming for evolutionary biologists" in 2017 and Evomics to attend the "2018 Workshop on population and speciation genomics" at Cesky Krumlov. Whole genome sequencing data was collected in the frame of the HybridAdapt project jointly funded by FCT (reference FCT-ANR/BIA-EVF/0250/2012) and by the French Agence Nationale de la Recherche (reference ANR-12-ISV7–0002-01). Advisor José Melo-Ferreira was supported by FCT via IF (IF/00033/2014) and CEEC (CEECIND/00372/2018) contracts.
Additional support is listed in each respective chapter.

Acknowledgements

I barely remember my life before being a PhD student! The (long) years during which I developed the work present in this dissertation were not only a period of great professional development but also of tremendous change and personal growth. The journey was tortuous, with immense obstacles and this work would not have been possible without the support and joint effort of several people (mentioned or not along this document) and to all of them I offer my sincere “Obrigado! Thank You! Merci!”.

Ao meu orientador e Amigo, José Melo-Ferreira. Obrigado por todo o apoio profissional, mas também pessoal que me deste. Durante estes últimos anos foste a âncora que me manteve agarrado à ciência, mesmo quando as coisas não corriam como planeávamos ou quando entrava nos meus exageros. Nas nossas reuniões infindáveis, mas também nas futeboladas ou nos festejos de mais uma conquista. Que continuemos a colaborar e a produzir boa ciência, mas também ótimos momentos para lá do “trabalho”.

A toi Pierre, pour tout ton soutien. Pour l'enthousiasme avec lequel tu parles de sujets scientifiques, pour les enseignements incroyables, pour tout l'accompagnement à Montpellier. Merci pour ta patience à corriger mes fautes d’orthographe et de prononciation. Merci aussi pour ces moments où j'ai quitté nos réunions avec un sentiment d'incompétence totale, ils étaient et seront indispensables pour être meilleur (personne et chercheur)!

To both of you, who were always with me along the way, I genuinely thank you for being exceptional persons and scientists. I felt lucky to have been guided by two excellent role models. And I aspire one day to think as sharply and clearly as you, to have my own team, develop my own projects and guide others as you do. Our (never-ending) meetings, discussions and your feedback were crucial for the progress of my work and to the improvement of my scientific knowledge. All your enthusiasm, recognitions, supportive words but also criticisms were essential to persuade me to continue and to finish this immense chapter of my (our) life. Thank you and I am sure that we will keep working together! This is just the beginning!

Je souhaiterais également remercier les membres de mon comité de thèse, Pierre-Alexandre Gagnaire, Camille Roux et Mathieu Joron. Merci d'avoir accepté de faire partie de mon jury et pour vos conseils et les discussions qui m'ont permis d'avoir de nouvelles idées.

Ao Professor Paulo Célio, ao Rui Faria, à Catarina Pinho, ao Bruno Louro e ao Marco Campinho porque me permitiram chegar onde estou hoje e porque acreditaram em mim quando ainda nada de relevante tinha feito. Ao Paulo um agradecimento especial por ter visto para além de um elemento para jogar à bola, por ter acreditado mais do que ninguém (inclusivé eu próprio) que podia um dia fazer este doutoramento e claro por me ter apresentado os meus orientadores. Acrecento ainda um muito obrigado à Graciela Sotelo, Pedro Esteves, Miguel Carneiro, Filipe Castro, Martim Melo e à Teresa pelos inúmeros bons momentos que tivemos e que foram muito para lá das discussões científicas, mas que são tão ou mais importantes!

To past and current members of the EVOCHANGE group, a very big thank you for being such great colleagues and for dealing with my criticism and obsession with technical details. Our discussions and your insights were always important for my work, and I have learned so much from you. Thank you for the feedback, support, suggestions and patience along the last years. I sincerely miss seeing you all in person every day and having beers at Teresa's bar after a group achievement! Hope this time gets back soon. Merci aussi à Paul Saunders, Pierre Paulus et Guylène Soula. Vous avez rendu mon séjour à Montpellier très facile et plus agréable!

I also thank the institutions or entities that contributed to the work developed in this thesis. Those include, CIBIO, Universidade do Porto, BIODIV, ISEM, Université Montpelllier, GAIA, CNRS, FCT, CampusFrance, CRUP, LIA, BioPolis, Universität Leipzig, Evomics and many others that due to exhaustion I forgot to mention.

À professora Natália Ferreira e ao fantastico grupo de alunos da escola secundária de Rio Tinto, ao Luís, ao João (Rocha e Saravia), ao Miguel, à Ana, à Cristina, à Catarina e ao Bernardo um especial obrigado por me obrigarem a ser melhor comunicador de ciência e me relembrarem que de nada vale o que possa descobrir se não o disser aos outros.

To the wonderful people I met in Leipzig and Czesky Krumlov with whom I learned so much and I had so much fun. I miss you all, and I keep hoping we can further collaborate in the near future! To the Bad Bromance team and particularly to Cerca a special thank you! It was really a pleasure to work and have fun with all of you!

On a (even) more personal remark, I would also like to thank the friends that were with me along the way.

To the guys of “Barracões”, d’O Clube and BIODIV a special thank you! You maid this journey much funnier and easier! Most of you are way more than colleagues and I miss you all!

Aos amigos de sempre e para sempre, o meu MUITO obrigado por estarem sempre lá! Nos bons, mas especialmente nos maus momentos! Por me permitirem descontrair e divertir, mas especialmente por me relembrarem que existe imenso mundo para além destas páginas! Por serem a familia que escolhi (ou que me escolheu)!

A ti que demasiado cedo nos deixaste e que desmoronaste o meu mundo como nunca ninguém o tinha feito. Sempre foste um exemplo de perseverança, de luta por mais e melhor, e sempre me ensentivaste a ser uma pessoa melhor! Sei que estarás sempre comigo porque assim o vou garantir! E enquanto depender de mim não serás esquecido, mas sim relembrado!

Agradeço também à minha família, que são a razão da minha existência e o meu alicerce. Em particular, obrigada aos meus pais, irmão, avós, padrinhos, tios e primos que durante os últimos anos tiveram de conviver com as minhas ausências e por vezes “alheamentos” por estar a divagar nos meus pensamentos.

Finalmente, a ti, e à nossa Inês! Simplesmente OBRIGADO…vocês são TUDO e espero vos poder dar agora a atenção devida!

Sem vocês todos nada disto seria possível!

Summary

Speciation is undoubtedly an important process in generating biodiversity over evolutionary timescales. It is therefore important to understand the underlying evolutionary mechanisms leading to speciation, but also the potential limits of this somewhat arbitrary classification in describing fully the patrimony of genetic variation available for evolution to occur.

Although species are described as discrete entities, the process that leads to their formation is continuous and gradual, and there is increasing evidence that genetic exchanges can continue to occur along the process, and/or even after it is well advanced, i.e. after hybrids have reduced fertility or viability. Therefore, understanding the evolution of the diversification of life needs, on the one hand, to unravel the mechanisms leading to reproductive isolation, i.e. determining what part of the genetic divergence between species is contributing to making them reproductively isolated, and what evolutionary forces have led to such divergence: mutation and drift linked to demography/biogeography, as well as selection, be it positive or negative, adaptive or non-adaptive (the latter caused by genetic conflicts resulting from selfish genetic elements). On the other hand, we also need to understand the possible consequences of these apparently frequent genetic exchanges on the evolution of species: to what extent do they participate in adaptation (and potentially to speciation), or on the contrary constitute an evolutionary burden?

In this context, situations of hybridization/admixture between diverged populations, that we would tend to call species, offer interesting natural (or artificial) laboratories to address the two aspects. Modern genomic techniques and statistical developments hold promise of the possibility to infer the genomic landscape of admixture (i.e. to infer the ancestry of every position in the genome of an admixed population), and to infer the evolutionary forces that have determined these genomic patterns (i.e. inferring the influence and intensity of positive or negative selection on local patterns). Major advances in this direction have come from human population genetics, because of the enormous amount of highest quality genomic data available in very large samples. However, studies on other species cannot generally access the full power of the

developed methods, and must rely on more indirect and less powerful inferences. In all cases however, going from patterns to processes is a complex task for several reasons: first, the landscapes of differentiation and admixture are influenced by stochastic and/or poorly documented processes (respectively mutation and drift, and past demography/biogeography). Second, the different positions in the genome do not evolve independently from one another, because of genetic linkage causing interference between the different evolutionary forces that operate at different sites (positive or negative selection of various intensities). Furthermore, the degree of linkage varies with the amount of recombination, which is variable along the genome, as can be mutation rates as well. Finally, such genomic scans are generally operated independently of any link between phenotype and genotype. Although they may suggest candidate genomic regions for selectively driven reduced or enhanced introgression, the origin of selection can at best sometimes be suggested through the function of the genes concerned. However, since phenotypic scan is most often out of reach, there is a promise that the functions of the genomic regions concerned be suggestive of the biological processes and traits subject to the selective pressures inferred. Additionally, comparative studies on replicates or on different species can also be suggestive of categories of functions that appear to be often involved in reproductive isolation or adaptive introgression.

The present thesis has brought contributions of various nature to the issues mentioned above, on two different biological models with documented hybridization between closely related taxa: hares (*Lepus* spp.) and mice (*Mus musculus*).

First, this work has contributed to the development of the genomic resources available to study hare population genomics, by providing the first *de novo* assembly of a hare genome (for the mountain hare, *Lepus timidus*), and assessing its utility as compared to the rabbit assembly, previously available. We have also generated the first mountain hare transcriptome, and the most complete among the currently available *Lepus* transcriptomes. In combination with published data on the European brown hare (*L. europaeus*), we pinpointed candidate fixed differences between the two species that can be used to build genotyping tools to monitor gene exchange in contact zones.

Second, we have contributed to the understanding of the documented massive introgression of the mitochondrial genome from the mountain hare to the Iberian hare (*L. granatensis*), by reconstructing the post-glacial demographic dynamics of the latter species using Single Nucleotide Polymorphism data. We demonstrated that this introgression occurred at the favor of the invasive replacement of the donor species by

the recipient one during the last deglaciation, thus showing the importance of demographic and biogeographic history in driving introgression.

Third, studying the house mouse (*Mus musculus*), we analyzed whole genome sequences from several populations in Iran and demonstrated the existence in Central Iran of a population resulting from extensive past admixture of two subspecies (*M. m. domesticus* and *M. m. musculus*) that, in contrast, form in Europe a tension zone attesting reproductive isolation . Studying the variations of levels of admixture along the genome in this hybrid Iranian population has stzrted to shed light on the genomic architecture and functional origin of reproductive isolation.. Our analyses also suggest a selective advantage of non-*domesticus* Y chromosome in this context of admixture.

Fourth, we discover in NW Iran geographic region where mice are predominantly of *domesticus* ancestry, although admixed with *musculus*. However, a *musculus*-like Y chromosome is fixed in this region. We searched for genes of the X chromosome and the autosomes showing similar massive introgression and found an enrichment on male fertility associated genes. Furthermore, we tested the potential link of Y introgression with an arms-race between ampliconic regions on the X and Y chromosomesknown to antagonistically affect the sex-ratio in a dosage-dependent manner.. We found the correlation between copy numbers of Y and X ampliconic families (*Sly/Slx*) expected to result from such a conflict. We show that the *musculus* Y invasion seems to have occurred in all regions of subspieces admixture but not outside, suggesting that this genomic conflict is not the only cause of this massive introgression..

Overall, the biological models presented in this thesis promise to be important case studies that may constitute key elements for the clarification of the determinants and consequences of admixture and introgression between species.

Keywords

Genomic resources, Speciation Genomics, Introgression, Genomic incompatibilities, Sex chromosomes

Sumário

A especiação é um dos mais importantes processos que leva à geração de biodiversidade ao longo do tempo evolutivo. É, portanto, um tema importante em biologia evolutiva compreender quais os mecanismos evolutivos por detrás da formação das espécies, mas também perceber os potenciais limites da classificação simplista das espécies para descrever de uma forma completa o património genético disponível para que a evolução se desenrole.

Apesar da espécie ser geralmente descrita como entidade discreta, os processos evolutivos que levam à sua formação são contínuos e graduais, e há evidência científica substancial de que trocas genéticas podem continuar a ocorrer ao longo de todo o processo de especiação, mesmo quando este se encontra em fases bastante avançadas, i.e. mesmo quando os híbridos apresentam reduzida fertilidade e viabilidade. Deste modo, compreender o processo evolutivo que leva à formação e diversificação das espécies requer por um lado desvendar os mecanismos que levam ao isolamento reprodutivo, ou seja, que parte da divergência genética entre as espécies contribuiu para torná-las isoladas reprodutivamente e quais as forças evolutivas que originaram essa mesma divergência: mutação e deriva genética ligadas à demografia/biogeografia, bem como a seleção, seja ela positiva ou negativa, adaptativa ou não adaptativa (neste último caso causada por conflitos genéticos decorrentes de elementos genéticos egoístas). Por outro lado, requer também perceber as possíveis consequências das aparentemente frequentes trocas genéticas entre as espécies durante a sua evolução: de que forma serão estas fundamentais para a sua adaptação (e potencialmente especiação), ou se pelo contrário constituem um fardo evolutivo.

Neste contexto, fenómenos de hibridação/mistura genética entre populações divergentes, às quais habitualmente chamamos espécies, oferecem um interessante laboratório natural (ou nalguns casos artificial) para o estudo destes dois aspetos. O desenvolvimento de modernas tecnologias de sequenciação genómica e os recentes avanços metodológicos que permitem maximizar a sua análise, prometem permitir inferir a paisagem genómica da mistura genética entre espécies, i.e. inferir a ancestralidade ao longo do genoma de populações resultantes da mistura de duas ou mais entidades divergentes) e identificar as forças evolutivas responsáveis pelo padrão inferido (i.e.

inferir a influência e intensidade da seleção, positiva ou negativa, nos padrões genómicos locais de variação). Os maiores avanços científicos nesta direção resultam do estudo da genética populacional humana devido à quantidade e qualidade de dados genómicos em amostras populacionais grandes. Contudo, estudos noutras espécies não conseguem ainda na generalidade aceder ao potencial completo dos métodos desenvolvidos, e dependem de inferências indiretas e com menor poder. Em qualquer caso, passar da descrição de padrões para a inferência de processos é uma tarefa complexa por diversas razões: em primeiro lugar porque os padrões de diferenciação e mistura ao longo do genoma são influenciados por processos estocásticos ou raramente bem documentados (p.e. mutação, deriva genética, demografia/biogeografia passada). Em segundo lugar, as posições ao longo do genoma não evoluem de forma completamente independente, uma vez que o desequilíbrio de ligação causa interferência entre as diferentes forças evolutivas que atuam em diferentes regiões do genoma (seleção positiva ou negativa com diferentes intensidades). A situação é ainda mais complexa porque o grau de desequilíbrio de ligação varia com a taxa de recombinação, que também é variável ao longo do genoma, assim como as taxas de mutação. Por último, a maioria dos estudos genómicos é realizado sem que haja o conhecimento da ligação entre o genótipo e o fenótipo. Ou seja, ainda que os resultados destes estudos possam sugerir regiões genómicas candidatas a promover ou impedir introgressão, a origem da seleção é muitas vezes apenas sugerida pela função dos genes envolvidos. No entanto, uma vez que análises funcionais estão muitas vezes fora do alcance, existe a expectativa de que as funções das regiões genómicas em questão sejam, em alguns casos, claramente sugestivas dos processos biológicos e dos fenótipos sujeitos às pressões seletivas. Neste sentido, estudos comparativos entre réplicas do mesmo processo ou em diferentes espécies poderão permitir identificar categorias de funções que poderão estar frequentemente envolvidas no isolamento reprodutivo ou introgressão adaptativa.

A presente tese trouxe contribuições de diferentes naturezas sobre as questões levantadas em cima, em dois modelos biológicos com evidências bem documentadas de hibridação entre taxa evolutivamente próximos: lebres (*Lepus* spp.) e ratos (*Mus musculus*). Em primeiro lugar, este trabalho contribuiu para o desenvolvimento de recursos genómicos para o estudo populacional em lebres, através da produção do primeiro genoma de referência reconstruído *de novo* de uma espécie de lebre (a lebre variável, *Lepus timidus*). Foi realizada uma avaliação da sua utilidade para estudos nas lebres em comparação com a utilização do genoma de coelho que era, até então, o

genoma referência mais próxima disponível. Foi também produzido o primeiro transcriptoma da lebre variável, sendo à data o transcriptoma mais completo de uma espécie de lebre. Em combinação com dados de lebre europeia (*L. europaeus*) previamente publicados, identificámos diferenças genómicas fixadas entre as duas espécies que poderão ser utilizadas na produção de ferramentas de monitorização de hibridação em zonas de contacto.

Em segundo lugar contribuímos para o entendimento de uma introgressão comprovada e massiva do genoma mitocondrial da lebre variável para a lebre Ibérica (*L. granatensis*), reconstruindo as dinâmicas demográficas pós-glaciares desta última, através da análise de polimorfismos nucleotídicos simples ao longo do genoma. Este trabalho demonstrou que esta introgressão ocorreu a favor do sentido geográfico da invasão e substituição da espécie dadora pela recetora durante a deglaciação, mostrando assim a importância da demografia e história biogeográfica na promoção de introgressão entre espécies.

Em terceiro lugar, estudando o rato doméstico, analisámos dados de genomas completos e demonstrámos a existência de uma população resultante da mistura ancestral de duas subespécies que estão geneticamente isoladas no presente. Estudar o mosaico de ancestralidade ao longo do genoma da população híbrida do Centro do Irão permitirá estabelecer potenciais candidatos ligados ao isolamento reprodutivo. As nossas análises sugerem ainda uma desvantagem seletiva do cromossoma Y de *domesticus* em cenários que envolvem mistura de linhagens.

Em quarto lugar, apesar do envolvimento geralmente pronunciado dos cromossomas sexuais no isolamento reprodutivo entre espécies, usando dados de genomas completos demonstrámos um caso de extensa introgressão do cromossoma Y de uma subespécie de ratinho noutra. Este modelo foi então usado para identificar regiões do cromossoma X e dos autossomas que potencialmente coevoluíram com o cromosoma Y introgredido, tendo encontrado um enriquecimento em genes ligados à fertilidade dos machos. Testámos ainda a potencial coevolução entre regiões amplicónicas dos cromossomas X e Y que se sabe poderem manipular a proporção dos sexos na descendência, abordando assim o papel de conflitos genéticos na promoção de introgressão. E verificamos uma correlação entre o número de cópias de uma família de elementos amplicónicos (*Sly/Slx*) cuja interação se sabe estar associada ao controlo da transmissão dos cromossomas sexuais.

Em suma, os modelos biológicos apresentados na presente tese providenciam importantes casos de estudo que poderão constituir elementos-chave para a clarificação do papel de seleção natural e da hibridação na especiação.

Palavras-chave

Recursos genéticos, Genómica da Especiação, Introgressão, Incompatibilidades genómicas, Cromossomas sexuais

Résumé

Une des questions fondamentales et largement débattue en biologie évolutive est celle des déterminants de la spéciation, c'est à dire de l'acquisition de l'isolement reproductif entre deux entités issues d'un ancêtre commun. L'avènement de la génomique des populations promet la possibilité de reconstituer finement et sur l'ensemble du génome l'histoire et les modalités de la divergence génétique, et donc de trouver ces déterminants. Elle a toutefois aussi largement démontré que les espèces ne sont pas des compartiments étanches, et que les échanges secondaires sont très fréquents, ce qui pose la question de leurs causes et conséquences évolutives, une autre question très importante pour comprendre l'évolution du vivant. Que ce soit sous l'angle de la spéciation ou de l'introgression, l'information la plus recherchée est celle de la détection de l'action de la sélection naturelle. Pour ce faire, il faut être capable de faire des prédictions d'un attendu de patrons de variation génétique en l'absence de sélection. Or l'attendu doit être déterminé en utilisant les mêmes données. D'où le difficile exercice de pouvoir comprendre la divergence en présence de mélange, et de comprendre l'histoire neutre en présence de sélection. Nous avons appliqué ici les méthodes de la génomique des populations pour progresser dans ces voies sur deux modèles biologiques : les lièvres en raison de l'occurrence spectaculaire d'introgressions mitochondriales massives et répétées entre espèces, et la souris domestique comme modèle de spéciation incipiente.

Nous avons contribué à la compréhension des circonstances ayant abouti à un phénomène remarquable, qui est l'introgression massive du génome mitochondrial entre espèces de lièvres dans la péninsule ibérique. On avait décrit dans l'espèce endémique de la péninsule ibérique *Lepus granatensis* un gradient de fréquence du génome mitochondrial provenant d'une espèce boréale (*L. timidus*), qui n'est plus présente dans la péninsule, mais l'était jusqu'au début du dernier réchauffement climatique post-glaciaire. Cette introgression est absente dans le sud de la péninsule, et augmente à partir du milieu, jusqu'à atteindre la quasi-fixation dans le nord. Nous avons typé une batterie de marqueurs génétiques (de type SNP, polymorphismes de nucléotides uniques) sur un échantillon de populations de *L. granatensis* réparties sur l'ensemble de la péninsule. L'analyse statistique des variations spatiales de fréquences alléliques

montre un accord avec un modèle d'expansion géographique passée, depuis le sud vers le nord. Nous avons donc émis l'hypothèse que cette expansion géographique de *L. granatensis* s'est faite lors du réchauffement post-glaciaire, et a conduit à l'envahissement du territoire jusqu'alors occupé par *L. timidus,* dans la moitié nord de la péninsule, région dont le climat était à l'époque encore favorable à cette espèce boréale, mais est devenu progressivement plus favorable à l'espèce tempérée *L. granatensis*. Des hybridations répétées durant la progression de ce front d'invasion auraient permis l'introgression génétique depuis *L. timidus*, et son augmentation de fréquence vers le nord. Le processus se serait arrêté après l'invasion complète de la péninsule, et l'extinction de *L. timidus* de ce territoire. Les données génétiques disponibles suggéraient toutefois qu'une telle introgression massive était limitée au génome mitochondrial. Nous avons émis l'hypothèse que ceci était dû à la transmission maternelle du génome mitochondrial, au rôle prépondérant des mâles dans le processus de colonisation et d'échanges génétiques, et à la philopatrie des femelles. Ce modèle neutre, ne faisant pas intervenir la sélection naturelle mais seulement des processus démographiques, a servi d'hypothèse nulle dans des études ultérieures de l'introgression entre ces deux espèces à l'échelle génomique.

Nous avons contribué au développement futur de la génomique sur les lièvres en produisant le premier assemblage *de novo* de génome pour ce genre (pour l'espèce *Lepus timidus*), et nous avons pu évaluer sa qualité et son utilité en comparaison au génome de référence du lapin qui était disponible. Nous avons d'autre part produit un transcriptome de bonne qualité pour *L. timidus*, et défini des marqueurs qui seront utiles pour l'étude de son hybridation naturelle fréquente avec *L. europaeus*, qui pose des questions de conservation et d'adaptation aux changements climatiques.

Nous avons également contribué à la compréhension de l'histoire de la différentiation entre sous-espèces de la souris domestique (*Mus musculus*), et des conséquences de l'hybridation secondaire. La souris domestique était connue pour être structurée en trois sous-espèces génétiquement différenciées (*M. m. domesticus*, *M. m. musculus* et *M. m. castaneus*), à distributions parapatriques (c'est à dire disjointes mais adjacentes) sur le continent eurasiatique. Les deux sous-espèces *domesticus* et *musculus* sont originaires du Proche ou Moyen-Orient, et ont colonisé l'Europe récemment, en association avec leur commensalisme avec l'espèce humaine, suivant

deux corridors géographiques distincts. Il y a quelques milliers d'années, leurs aires de distribution se sont rencontrées en Europe. L'étude génétique détaillée de cette zone de contact a montré qu'une zone de tension s'était formée, c'est à dire une zone géographiquement étroite de transition, maintenue par l'équilibre entre la migration et la contre-sélection des hybrides, opérant à de nombreux locus répartis dans le génome (avec toutefois un rôle prépondérant du chromosome X). Il s'agit donc d'une situation de spéciation en cours, selon un modèle dit de spéciation allopatrique (c'est à dire suite à une divergence en isolement génétique). Afin de mieux comprendre les contextes géographiques et temporels de la mise en place de cet isolement reproductif depuis la divergence initiale (qui était estimée à quelques centaines de milliers d'année), nous avons étudié des populations du berceau géographique de ces sous-espèces, en Iran à partir de séquençage de génomes complets représentant plusieurs populations de cette région et de sa périphérie (incluant les trois sous-espèces précédemment décrites). Nous avons découvert plusieurs phénomènes remarquables.

Nous avons tout d'abord découvert au centre de l'Iran une population (CEI) qui, dans une analyse descriptive sans *a priori* de sa composition génétique (analyse en composante principale des génotypes à de nombreuses positions nucléotidiques indépendants dans le génome), apparaît clairement distincte des trois sous-espèces connues. En combinant les résultats de l'analyse de la distribution des allèles dérivés entre populations (f-statistiques), l'analyse des corrélations de fréquences alléliques entre populations (« admixture graphs »), et l'inférence de la distribution temporelle des taux de coalescences croisées entre populations (à partir de paires d'haplotypes de chromosomes X de populations différentes et la méthode PSMC), nous proposons que la population CEI résulte du mélange secondaire d'environ 40% provenant de *domesticus* d'Iran, et 60% d'une population rattachée à la branche évolutive qui mène à *musculus*. Nous confirmons cette interprétation et ces proportions en appliquant une méthode de reconstitution des variations d'ancestralité le long du génome basée sur l'analyse du déséquilibre de liaison, plutôt que les fréquences alléliques (méthode ELAI). En intégrant ces résultats avec ceux obtenus sur les marqueurs dont la transmission est liée au sexe (génome mitochondrial, chromosomes sexuels), et leurs discordances quant à l'ancestralité, nous proposons un modèle phylogéographique de différenciation et re-mélange de ces entités depuis leurs divergences initiales, envisageant divers scénarios de remplacement invasif pour expliquer le mélange conduisant à CEI. Nous ne pouvons réconcilier toutes ces données dans un tel modèle sans supposer que le

chromosome Y de la population CEI (proche de celui de *musculus*) s'est maintenu dans cette population hybride parce qu'il présentait un avantage sélectif. Nous proposons sur la base de ces résultats que le couple *domesticus-musculus* présente des caractéristiques d'une espèce en anneau (« ring species » en anglais), où l'isolement reproductif est moins prononcé entre les populations proches du berceau de différenciation qu'entre celles issues de colonisations indépendantes d'aires secondaires disjointes.

Nous avons ensuite étudié la nature génétique des peuplements du nord-ouest de l'Iran, aussi à partir de génomes complets d'échantillons distribués dans cette zone. L'analyse descriptive basée sur l'ACP (Analyse en Composantes Principales) des génotypes suggère deux entités géographiquement séparées mais adjacentes dans cette région, caractéristiques de *musculus* (au nord, population CAU) et *domesticus* (plus au sud, population NWI). L'analyse suggère toutefois un certain degré d'échange entre ces deux entités, ainsi qu'une contribution de la population CEI nouvellement découverte, ce qui est confirmé par l'analyse de la distribution des allèles dérivés (f3-satistics). Les données mitochondriales sont concordantes avec cette interprétation, du point de vue de la phylogéographie et de l'origine des contributions. Cette région géographique apparaît donc comme une région de re-mélange à trois voies entre les trois entités génétiques géographiquement adjacentes (*domesticus*, *musculus* et CEI). Sur la base des déséquilibres de liaison (méthode ELAI), nous reconstituons l'ancestralité le long du génome et estimons pour cette population *domesticus* (NWI) une contribution de quelques pourcents (<9%) chacun pour *musculus* et CEI. Nous avons noté que les méthodes classiquement utilisées pour modéliser les divergences avec échange à partir des fréquences alléliques, qu'elles soient basées sur des modèles simplificateurs (ADMIXTURE) ou sur des heuristiques particulières (TreeMix) échouaient à rendre compte de l'histoire de différenciation et re-mélange dans une situation de cette complexité, où les échanges se sont produits entre toutes les entités, et à diverses périodes.

Nous montrons enfin que la distribution de deux lignées très divergentes de chromosome Y, caractéristiques de *domesticus* et *musculus* en dehors de la région étudiée ici, n'est pas concordante avec celle de la moyenne du génome dans cette zone géographique. En effet, la lignée Y *musculus* est fixée dans la population à majorité

domesticus. Nous avons recherché d'autres régions génomiques présentant un patron d'introgression similairement extrême, et avons trouvé un excès de gènes impliqués dans la fertilité mâle. Nous avons aussi caractérisé les variations de nombre de copies de régions ampliconiques (en utilisant une méthode basée sur l'analyse des k-mers) dans ces populations, et trouvé une famille du chromosome X (contenant le gène *Slx*) qui suit le patron connu du chromosome Y (pour la famille ampliconique contenant le gène *Sly*), c'est à dire un nombre de copies plus faible chez *domesticus*, et plus élevé chez *musculus* (y compris dans la population SWI, *domesticus* mais possédant un chromosome Y *musculus*). Ces deux familles ampliconiques sont connues pour agir de manière antagoniste et dosage-dépendante sur le sex-ratio de la descendance. Nous discutons la possibilité que ce conflit soit à l'origine de la corrélation des nombres de copies X et Y, et de l'envahissement du chromosome Y *musculus* dans toutes les régions de re-mélange, phénomène qui semble s'être produit également dans la population CEI, comme nous l'avons vu plus haut, mais aussi dans toutes les autres régions de contact ou re-mélange entre sous-espèces connues en dehors d'Iran. Sur la base d'indices connus sur les effets du chromosome Y *musculus* et du nombre de copies ampliconiques sur la fertilité, et de nos propres données d'estimation comparée du nombre de copies dans les populations étudiées, nous proposons qu'un avantage intrinsèque du chromosome Y *musculus* en situation d'hybridation est une explication plus vraisemblable de son succès, et que les propriétés de distorsion de transmission représentent plutôt un coût, lié à une baisse de fertilité associée.

Nos travaux ouvrent de nombreuses perspectives dans le domaine de la génomique de la spéciation et de l'hybridation. Nous fournissons de nouveaux outils pour la génomique d'un groupe d'espèces non modèles, les lièvres. Nous fournissons un cadre historique pour les conditions d'interaction et d'hybridation entre espèces de lièvres dans la péninsule ibérique, qui contribuera à améliorer l'analyse fine de l'introgression entre ces deux espèces, et la détermination de son contrôle par la sélection, en s'affranchissant des effets confondants de la démographie. En ce qui concerne la souris domestique, la population hybride CEI représente une situation précieuse pour progresser dans la compréhension des déterminants de l'isolement reproductif entre sous-espèces, en recherchant plus finement dans les régions génomiques qui ne participent pas au re-mélange en Iran, et celles qui ont particulièrement divergé lors de la colonisation des deux branches de l'espèce en anneau. De même, la comparaison des modalités d'échanges génétiques entre les

situations de contact en Iran et en Europe sera du plus grand intérêt. Nous contribuons aussi à la construction d'un modèle de plus en plus complet et plausible de l'histoire (très complexe) de la différenciation de ces sous-espèces. La qualité d'un tel modèle est essentielle pour pouvoir interpréter les patrons de variation et inférer l'influence de la sélection sur la divergence et le re-mélange. Enfin, nous émettons une hypothèse intéressante concernant l'évolution du chromosome Y, combinant ses effets potentiels sur la fertilité mâle, et ses contraintes liées au conflit avec le chromosome X.

Mots clés

Ressources génétiques, Génomique de spéciation, Introgression, Incompatibilités génomiques, Chromosomes sexuels

Publications

Published:

Chapter II

Publication I - Marques, Joao P, Mafalda S Ferreira, Liliana Farelo, Colin M Callahan, Klaus Hacklander, Hannes Jenny, W Ian Montgomery, et al. 2017. 'Mountain Hare Transcriptome and Diagnostic Markers as Resources to Monitor Hybridization with European Hares.' *Scientific Data* 4 (December): 170178–170178. https://doi.org/10.1038/sdata.2017.178.

Publication II - Marques, João P., Fernando A. Seixas, Liliana Farelo, Colin M. Callahan, Jeffrey M. Good, W. Ian Montgomery, Neil Reid, Paulo C. Alves, Pierre Boursot, and José Melo-Ferreira. 2020. 'An Annotated Draft Genome of the Mountain Hare (Lepus Timidus)'. *Genome Biology and Evolution* 12 (1): 3656–62. https://doi.org/10.1093/gbe/evz273.

Chapter III

Publication III - Marques, João P., Liliana Farelo, Joana Vilela, Dan Vanderpool, Paulo C. Alves, Jeffrey M. Good, Pierre Boursot, and José Melo-Ferreira. 2017. 'Range Expansion Underlies Historical Introgressive Hybridization in the Iberian Hare'. *Scientific Reports* 7 (1): 40788. https://doi.org/10.1038/srep40788.

In preparation:

Chapter III

Publication IV - João Pedro Marques, José Melo-Ferreira, and Pierre Boursot. 2022. Extensive admixture in the cradle of differentiation and speciation of house mice.

Publication V - João Pedro Marques, José Melo-Ferreira, and Pierre Boursot. 2022. A successful Y chromosome lineage in regions of admixture between house mouse subspecies.

FCUP and U. Montpellier

Using genomic tools to understand species differentiation and admixture in hares and mice

Index

Index of Figures

Publication 1

Publication 2

Publication 3

Publication 4

Publication 5

Index of Tables

Publication 1

Publication 2

Publication 3

Publication 4

Publication 5

Abbreviations

μ – mutation rate
AFG – Population from Afghanistan
BDMIs – Bateson-Dobzhansky-Muller hybrid incompatibilities
bp – base pairs
CAS – *M. m. castaneus*
CAU – Population from caucasus
CDS – coding sequence
CEI – central iranian population
Chr – Chromosome
CNV – copy number variation
crb-blast – Conditional Reciprocal Best BLAST
CYP – *Mus cypriacus*
CZ – Population from Czechia
DOM – *M. m. domesticus*
Dxy – genetic divergence
EHZ – European hybrid zone
ELAI – Efficient Local Ancestry Inference
FCT – Fundação para a Ciência e a Tecnologia
FDR – False discovery rate
FR – Population from France
Fst – Fixation index
GER – Population from Germany
GO – Gene Ontology
GRA – *Lepus granatensis*
HMM – hidden Markov model
hPSMC – X-hybrid Pairwise sequentially Markovian coalescent
ISEM – Institut des Sciences de l'Evolution Montpellier
kb – kilobase pairs
KEGG – Kyoto Encyclopedia of Genes and Genome
LD – *Linkage disequilibrium*
LEP – re-scaffolded reference
MAF – minor allele frequency
mtDNA – mitochondrial DNA
MUS – *M. m. musculus*
MWI – Population from Zagroos mountains
Myr – millions of years
Ne – effective population size
NI – Population from North India
NWI – Population from northwest Iran
OCN – *Oryctolagus cuniculus*
ORF – open reading frame
PAK – Population from Pakistan
PC – principal component
PCA – principal component analysis
PSE – pseudo-hare reference
PSMC – Pairwise Sequentially Markovian Coalescent
RIN – RNA Integrity Number
Scaf – Scaffold
SE – Standard error
SNP – single nucleotide polymorphism
SWI – Population from southwest Iran
TIM – *Lepus timidus*
TMRCA – time to the most recent common ancestor
θ – population mutation rate
Π (pi) – nucleotide diversity

Chapter I

General Introduction

Understanding how species originate, differentiate, and admix is one of the major goals in evolutionary biology. In 1942, in his seminal work, Mayr considered that the formation of many new species is what leads to evolutionary diversity: "Without speciation, there would be no diversification of the organic world, no adaptive radiation, and very little evolutionary progress." (Mayr 1942). Since then, the concept itself has evolved and speciation can be described as the evolutionary process by which new species can evolve from an ancestral one, through the continuous accumulation of reproductive barriers which eventually leads to complete isolation (Nosil, Funk, and Ortiz-Barrientos 2009; Nosil, Harmon, and Seehausen 2009; de Queiroz 1998). Even though many species concepts exist and great controversy among specialists arises when discussing criteria for species definition, the establishment of reproductive isolation (i.e. the biological species concept; (Mayr 1942) remains among the most widely considered when studying speciation (Butlin and Stankowski 2020). Speciation can be understood as a continuous process that can proceed even with some levels of gene flow between the diverging taxa, either as a result of continuous migration during the process (isolation with migration), or after a period of allopatric divergence (secondary contact). Incomplete reproductive isolation between the diverging populations allows these genetic exchanges between the species, in a process called “introgression”. Introgression has been shown not only to continue during speciation but also to contribute to diversification and adaptation throughout the tree of life (e.g. mammals (Giska et al. 2019; Liu et al. 2015), fishes (Meier et al., 2017; Svardal et al., 2020), or plants (Whitney, Randell, and Rieseberg 2010)).

1.1. Genetic models of speciation

The possibility of hybridization and genetic exchanges between species depends on numerous factors, such as reproductive barriers, time since divergence, system-specific differences, among others. Reproductive barriers are expected to be prevalent between diverging species, leading to low frequency of hybrids, impeding introgression. A similar influence is expected for the time of divergence, with experimental studies suggesting a strong negative correlation between hybridization and genetic distance (Coyne and Orr 1997; Price and Bouvier 2002). This seems to be also applicable to natural populations, as can be seen in the *Heliconius* system, where races that abundantly hybridize in nature have mtDNA sequence differences of less than 2%, races that occasionally hybridize are mostly 2-6% divergent and no hybrids were found between races with more than 10% divergence (Mallet and Joron 1999), or in the hare system where admixture proportions decrease with genetic divergence (Ferreira et al. 2021). These findings suggest a gradual and progressive development of reproductive isolation with time of divergence (Roux et al. 2016). Such relationship is compatible with a progressive accumulation of incompatibility factors along the genome, and the establishment of increasingly larger blocks of genomic isolation, until reproductive isolation is complete. Importantly it establishes the gene as the central unit of speciation (the genic view of speciation (Wu 2001)), contrary to models where the genome behaves as a single cohesive unit.

Sex chromosomes are especially important to the establishment of reproductive barriers. The differential inheritance in males and females, the reduced recombination and the faster evolution of gene repertoire and gene expression create a fertile environment for the establishment of incompatibilities across lineages. The role of sex chromosomes in speciation is well exemplified in the “two rules of speciation”: first when hybrid sterility is limited to one sex, it is almost always the heterogametic sex – Haldane’s Rule(Haldane 1922). Second, hybrid dysfunction differentially maps to the X or Z chromosome – large X-effect (Coyne and Orr 1989). Another important factor associated with sex dimorphism is the importance of behavioural decisions to mate, which greatly prevent hybridization between sympatric animals (e.g. revised in birds (Price and Bouvier 2002) and *Drosophila* (Coyne and Orr 1997)).

Bateson-Dobzhansky-Muller hybrid incompatibilities (BDMIs) is one of the best studied models of speciation (originally descripted in (Dobzhansky and Gould 1982; Bateson 1909; Muller 1942) and revised later by Maheshwari and Barbash 2011 and Unckless and Orr 2009). Decades of empirical and theoretical work have demonstrated that negative interaction of divergent genomic regions in hybrid genomes, or

incompatible interactions between mutations that are derived in each of the parental species, are a central mechanism underlying reproductive isolation once species are formed, and a common cause of inviability and infertility in hybrids (Moran et al. 2021). This model predicts that incompatibilities can arise if there is a deleterious interaction between variants at two loci in a hybrid context, for combinations that have never occurred within species (explaining why these derived variants have not been purged by selection upon emergence within species). Assuming two loci, locus A and locus B, where one mutation in locus A gets fixed in one species because it is not deleterious, and another mutation in locus B also gets fixed in another species, when these new variants at two loci are combined in a hybrid then they can be selected against. Interspecific crosses in the laboratory may allow identifying QTLs for reduced hybrid fitness, but the precise identification of interacting genes determining a given incompatibilty is rarely possible. Consequently, very few examples of BDMIs have been clearly described (Case et al. 2016; P. Christie and Macnair 1987; Nosil and Schluter 2011; D. C. Presgraves 2010; Sweigart and Flagel 2015; Vyskočilová et al. 2005; Wright et al. 2013).

Ecological or environmental selection is another potentially important promoter of speciation. Although the general understanding is that environmental selection inevitably leads to a change in the composition of genes within a population, little is often known about the genetic architecture of these ecological adaptations and consequently how these environmentally selected alleles can be decoupled. Admixed individuals may express intermediate phenotypes and selection will disfavour them (Delmore and Irwin 2014; Gow, Peichel, and Taylor 2007; Melo et al. 2014; Rundle and Nosil 2005; Scordato et al. 2020; Soria-Carrasco et al. 2014), yet sometimes these admixed lineages can be favoured if they occur in intermediate ecological niches (Hessenauer et al. 2020; Loren H. Rieseberg, Archer, and Wayne 1999; Schluter and McPeek 2000; Selz and Seehausen 2019; Stelkens and Seehausen 2009). Due to its functional connection and complex architecture, ecological selection is predicted to bias ancestry around ecological functionally genes, at least in the case of traits underlain by single or few loci.

Finally, the case of polygenic traits is much more complex and the outcomes are harder to predict, as recombination tends to decouple these association of parental alleles. Hybrid phenotypes fall sometimes outside of the phenotypic optima of either parental species, reducing their fitness (Fraïsse et al. 2016; Simon, Bierne, and Welch 2018). This phenomenon is known as segregation load: when parental alleles are mixed into distinct backgrounds, hybrids can evidence a larger variance in a trait in comparison

with what is observed in its parentals (N. H. Barton and Hewitt 1989; Slatkin and Lande 1994).

In summary, hybrid lineages can only be established if the genetic combinations from distinct parental species (with implications e.g. in protein-protein interactions, or gene regulatory mechanisms) allow survival and reproduction across generations. The challenge is then disentangling the factors impacting genome evolution in each case and understanding how genome stabilization can be achieved after hybridization between two diverged genomes. Historically, hybrid incompatibilities have been seen as the major cause of hybrid fitness reduction and therefore conditioning its establishment, but recent work has suggested that modelling hybrid fitness as a function of admixture and heterozygosity provides more general fit to observed data on hybrid fitness (Simon, Bierne, and Welch 2018).

1.2. Speciation and hybridization

Ronald Fischer stated that “the grossest blunder in sexual preference which we can conceive of an animal making, would be to mate with a species different from its own and with which the hybrids are either infertile or, through the mixture of instincts and other attributes appropriate to different courses of life, at so serious a disadvantage as to leave no descendants” (Fisher 1930). Indeed, the debate on the contribution of hybridization during speciation has been almost restricted to plants and its importance for animal speciation was neglected (Arnold 1992; Coyne and Orr 2004; Mayr 1963). However, we now understand that hybridization is a pervasive phenomenon across all living world, and the sequencing technology advances have greatly contributed to this clarification (reviewed in Moran et al. 2021). This technological revolution has allowed moving from the analysis of a small number of loci to full genomic landscapes, with the correspondent impact on the way we perceive the complexity of the evolutionary process.

Over the last two decades, genomic data have increasingly provided evidence that introgressive hybridization, or merely introgression, is a critically important source of genetic variation. We now know that introgression can be pervasive, leave signatures along profound timescales of evolution, affect genomes at distinct scales, and that genetic variation resulting from hybridization, when present, can vary from a modest genomic contribution (<10%) (e.g. humans (Harris and Nielsen 2016), hares (Giska et al. 2019; M. R. Jones et al. 2018; Seixas, Boursot, and Melo-Ferreira 2018), mice (Song

et al. 2011)), to very large genome proportions, such as hybrid species where balanced parental contributions are found (e.g. house sparrow (Elgvin et al. 2017), *Anopheles* mosquitoes (Fontaine et al. 2015), tiger swallowtail butterflies (Kunte et al. 2011) or wild sunflowers (Rieseberg 2003)).

It has been estimated that on average around 10% of animal and 25% of plant species hybridize with at least one other species (Mallet, Besansky, and Hahn 2016). However, in radiations, where species are more closely related and diversification occurs over short time spans, hybridization can be exceptionally prevalent, with introgression events occurring among many of the diverging lineages and layered throughout time, as for example in the well document case of ducks (Wang et al. 2019), cichlids (Svardal et al. 2020) and hares (Ferreira et al. 2021). Recent studies have also demonstrated the occurrence of both ancient and recent introgression on the genomes of extant species (e.g. sunflowers (Rieseberg 2003), sparrows (Hermansen et al. 2011) or *Heliconius* butterflies (Mallet et al. 2007)) but also on extinct emblematic species such as the mammoths (e.g. Enk et al., 2016, 2011; van der Valk et al., 2021). Although recent introgression among species is more easily detected (due to its larger genomic segments not yet broken by recombination), the signatures of ancient introgression can persist for thousands or even millions of years after the initial divergence (Payseur and Rieseberg 2016).

Over the last decade, genetic exchanges were not only established as frequent but seen as important adaptive forces in several studies (reviewed in Taylor and Larson, 2019). Although most genetic exchanges that prevail between closely related species are expected to be neutral, introgression can occasionally introduce adaptive variants. Indeed, “adaptive introgression” has now been documented across a wide range of taxa, linked, for example, to pesticide resistance in mice (Song et al. 2011) and mosquitoes (Norris et al. 2015), mimicry in *Heliconius* butterflies (Consortium 2012), abiotic tolerance in wild sunflowers (Whitney, Randell, and Rieseberg 2010), coat colour in hares (Ferreira et al. 2017; Jones et al. 2018; Giska et al. 2019; Ferreira et al. 2020; Jones et al. 2020) and has also been proposed in modern humans, with for example modern Tibetan populations showing local adaptation to high-altitude caused by Denisovan introgression (Huerta-Sánchez et al. 2014).

The introgression of beneficial (adaptive) or neutral alleles overcomes background genome-wide selection against hybrids (Arnegard et al. 2014; K. Christie and Strauss 2018; Orr 1995; Svedin et al. 2008). Such forces that impede introgression can result from single or combined selective mechanisms, as ecological selection, hybridization load or Bateson-Dobzhansky-Muller hybrid incompatibilities. The selection against

foreign genetic ancestry can thus be particularly informative about the mechanisms leading to genomic isolation and ultimately to speciation.

1.3. Past demography and introgression

While patterns of abundant introgression at particular loci may be indicative of adaptive processes underlying the genetic exchange, certain demographic processes may also promote introgression, which may be confounded with natural selection. The relative roles of population history and neutral demography in promoting introgression are an important topic of debate.

The evolutionary history of several organisms may have resulted from range expansions and population replacements, particularly during the glacial cycles oscillations on the result of drastic climate changes (e.g. Duvernell et al., 2019; Langdon et al., 2020; Nielsen et al., 2017; Seixas et al., 2018). Theoretical modelling, supported with an increasingly number of empirical models, shows that drift in the front of range expansions can lead to increases in frequency of rare alleles, a phenomenon coined « allele surfing » (Klopfstein et al., 2006). In situation of range invasions and replacements between hybridizing species, introgressed alleles at the leading edge of population expansion (wave front) may reach very high frequencies in the newly occupied areas through the combination of founder effects, genetic drift, and demographic expansion (Excoffier and Ray 2008). In an invasive range expansion scenario, a local introgressed allele at the wave front, where population density is especially low, has less chance to be lost by drift when the invader population is rapidly growing (Excoffier et al. 2009; Excoffier and Ray 2008). The allele surfing may result in a clinal distribution of allele frequencies along the axis of expansion (Excoffier et al 2009). The asymmetrical pattern of introgression that occurs almost exclusively from the local to the invading species in a range expansion scenario, was firstly noticed by Currat et al. (2008). The pattern described above was found in several published studies documenting introgression during range expansions (reviewed in Quilodrán et al., 2020).

1.4. The differentiation landscape and its potential determinants

The advent of next generation sequencing technologies has allowed us moving from studies based on few markers to genome-wide patterns. The innovative genomic perspective has been helping to turn towards a more integrative understanding of the

genomic architecture of speciation and how it may mediate or impede further divergence. With the implicit or explicit promise of solving "the mystery of mysteries" (Darwin and Kebler 1859) evolutionary biologists went through a journey to find the "genomic islands of speciation". The 'island' metaphorical concept (Turner and Hahn 2010) represents any genomic element (from a single nucleotide to an entire chromosome) which exhibits significantly greater differentiation than expected by neutral evolution alone (sea level in the metaphor). The pervasive occurrence of interspecific gene flow led researchers to interpret these genomic regions of high differentiation as those impeding gene flow and thus involved in reproductive barriers and speciation (genomic islands of divergence) (T. L. Turner and Hahn 2010). However, an alternative interpretation is that such islands are regions where selection (mostly purifying or background selection, possibly positive selection) is most intense because recombination is locally reduced, increasing linkage between sites under selection and neutral sites (genomic islands of differentiation) (Cruickshank and Hahn 2014; Nachman and Payseur 2012). It could be that either or both hypotheses and models apply, depending on whether divergence occurred in sympatry or allopatry, and on the intensity of gene flow between species, thus depending on the biology and history of the species considered (Nosil and Feder 2012). This is one vivid area of debate in evolutionary biology and the study of speciation and hybridization should be extended to a broader variety of organisms with different life-histories, recombination landscapes, divergence histories and level of interspecific genetic exchanges to draw more general conclusions.

Predicting how the genome will evolve after hybridization implies understanding the sources of selection that act on admixed genomes. This allows not only to study the mechanisms underlying species admixture but also those preventing it, shedding light on both the process of introgression and species formation (Nosil and Feder 2012). The landscape of admixture along the genome is expected to depend on the complex interplay between migration,selection (be it positive or negative), and recombination (Gavrilets 2003; Kirkpatrick and Ravigné 2002).

Genomics landscapes/scans of divergence or differentiation can hold information about the complex process of speciation. But going from patterns to processes requires understanding the relative forces determining such landcapes. The antagonism between the different forces can create very heterogenous genomic landscapes, which makes the inference of populations' history fundamental to correctly interpret those genomics islands and the forces responsible for their formation, size (small or large blocks), number (few or multiple loci), dispersion (concentrated or widely spread along the genome) and genomic content (genic or non-genic) (Nosil and Feder 2012).

1.5. Ancestry landscape and its interpretation

Many speciation genomic studies have widely shown that hybridization is a common phenomenon in nature. These assumptions are mainly supported by the findings of hybrid ancestries along many species genomes (reviewed in Payseur and Rieseberg, 2016; Schumer et al., 2018, 2015). An emerging and salient pattern is the strong relationship between variations of recombination and ancestry along the genome (Duranton and Pool 2021; Martin et al. 2019; Schumer et al. 2018). Theory predicts a positive correlation between recombination rate and introgression when the latter is deleterious (for instance driven by incompatibilities), such as shown in humans with Neanderthal introgression (Juric, Aeschbacher, and Coop 2016; Sankararaman et al. 2014; Schumer et al. 2018), between subspecies of house mice (Janoušek et al. 2015b), in hybrid lineages of swordtail fishes (Schumer et al. 2018), and within populations of *Heliconius* butterfly (Martin et al. 2019). The justification is that neutral introgressed alleles can more easily recombine away from deleterious ones and persist in the genome in highly recombining regions (N. Barton and Bengtsson 1986). Whereas in regions of low recombination, where linkage disequilibrium is higher, neutral alleles tend to be removed with the deleterious ones, reducing the introgression levels (Charlesworth, Morgan, and Charlesworth 1993). However, negative correlations between recombination and foreign ancestry have been recently found in natural populations of *Drosophila melanogaster* (Corbett-Detig and Nielsen 2017; Pool 2015). This could be attributed to the effect of positive selection (Duranton and Pool 2021): introgressed alleles favoured by selection will tend to bring along them larger linkage blocks, predominantly in the low recombining regions (Corbett-Detig and Nielsen 2017).

As previously stated, recombination is responsible for the shortening of introgressed tracts over time as it tends to break long ancestry tracts. This role of recombination on shaping hybrid genomes is particularly prominent in the early generations after hybridization, although it can still be effective after the stabilization of admixture proportions (Harris and Nielsen 2016; Nachman and Payseur 2012; Veller et al. 2019). Another consistent outcome from recent studies suggests that genomic regions enriched in coding or conserved elements tend to be particularly resistant to gene flow between species (Brandvain et al. 2014; Calfee et al. 2021; Martin et al. 2013; Masly and Presgraves 2007; Maxwell et al. 2018; Sankararaman et al. 2014; Teeter et al. 2008). This observation suggests that selection is particularly effective at generating barriers against introgression on functional elements (Sankararaman et al. 2014).

1.6. Genomic conflicts and speciation/introgression.

Genomic conflicts can be seen as a negative epistasis between two genomic elements (DNA sequences or their products), where an increase in the fitness of one decreases the fitness of the co-evolver. This notion can be illustrated by the relation predator-pray, where a fitness gain in the prey (e.g. ability to evade a predator) results in a lower fitness of the predator. The fitness incentive to the species that is 'losing' the evolutionary race can *per se* spur a tit-for-tat adaptation that may 'change the game' and increase the predator fitness and consequently reduce the fitness of the prey (reviewed in McLaughlin and Malik, 2017). This concept of perpetual antagonistic coevolution was described as "the Red Queen's Hypothesis" (Valen 1973). This intrinsic susceptibility to conflict creates a selfishness pervasive effect that shapes fundamental aspects at all levels of biology, from individuals to inter or intragenomic elements. At the intragenomic scale ('genome' meaning all inherited genetic material), selfish elements can affect pairs of antagonistically interacting loci and readily generate BDMIs, which are a major cause of reproductive isolation. These elements under conflict can directly distort their own transmission acting directly on gametes, interfering with transmission, or replicating disproportionally across the genome. Such directly distorting elements are thus expected to reach fixation, unless counter selected (Crespi and Nosil 2013; McLaughlin and Malik 2017). Alternatively, selfish elements can interfere with the patterns of inheritance, either on biparentally inherited elements (nuclear genes) or mono-parentally inherited (cytoplasmatic or organelle genes).

Although genomic conflicts could generally promote speciation, they could also prevent or constrain the process, especially in cases of population divergence with gene flow. One example is a hypothetical case where two driver alleles arise in two different populations connected through gene flow. The alleles will tend to move between populations, and because one is often more advantageous than the other, will tend to fix in both populations (Crespi and Nosil 2013). The role of intragenomic conflicts in speciation has been a matter of debate across times and the conceptual framework of "conflictual speciation" (Crespi and Nosil 2013) that unifies a significant collection of intragenomic conflicts that have been implicated in speciation. The range of mechanisms that could cause reproductive isolation via intragenomic conflict is vast and include: meiotic drive (or simply drive), where an allele - usually a selfish genetic element - biases its transmission into gametes, away from Mendelian expected ratios and often in detriment of the rest of the genome (Presgraves 2009); imprinting, where genes in diploid organisms express preferentially or exclusively the copy inherited from one of the parents

(this is the case in *Peromyscus* and *Mus,* where dosage imbalances of imprinted genes and epigenetic dysregulation have been associated with abnormal placenta development in interspecific hybrids (Shi et al. 2005; Vrana 2007)); cytonuclear incompatibilities (also known as 'mother's curse') produced by the conflict between maternally inherited organelles and biparentally-inherited nuclear 'restorer' genes; chromosomal conflicts induced by chromosomes with differing sex-linked transmissions (autosomes, X, and Y chromosomes) can be involved in evolutionary conflicts around the determination of sexually differentiated characters (Frank and Crespi 2011). Contrarily to other causes of large X effects, the chromosomal conflict theory predicts epistatic interactions between sex and autosomal chromosomes (Chase et al. 2005; Crespi 2008). For example, interactions of X-linked from one parental species with autosomal or Y-linked genes of the other parent are known to impact the genus *Mus* (see Good et al. 2010; White et al. 2011); transposable elements which are very abundant, rapidly evolving genomic elements that can copy and move to new locations within the genome. These elements were proposed as major drivers of speciation due to remarkable differences among lineages, ability to cause deleterious mutations and mediate genome reorganization as well as recurrent 'domestication' with adaptation purposes (Böhne et al. 2008; Brown and O'Neill 2010; Burt and Trivers 2006; Jurka, Bao, and Kojima 2011; Rebollo et al. 2010; Zeh, Zeh, and Ishida 2009).

Overall, conflicts involving the antagonistic co-evolution of different parts of the genome could either lead to homogenisation or isolation. It could be said that grossly, the outcome should depend on whether the arms race has followed similar paths in the two species or not, and on the relative pace of the races in the two species. When different paths were followed, incompatibility appears a more likely outcome, the result resembling the classical BDMI model. If similar paths were followed, invasion would appear more likely if the arms race had proceeded at different paces in the two species (in which case the faster evolving system should win).

1.7. Study systems

1.7.1. The impact of hybridization on the evolution of hares: current knowledge and future challenges

Hares (*Lepus spp.*) belong to the Leporidae family, which also includes several rabbit genera (Smith 2018), that together with Pikas constitute the Lagomorph order. These small to medium size mammals have likely originated in North America at ~12 MYA and

have dispersed and diversified across the world over the last 4-6 million years (Ferreira et al. 2021; Matthee et al. 2004; Melo-Ferreira et al. 2012b; Yamada et al. 2002). Hare species are currently naturally present all over the Northern Hemisphere and Africa (Ferreira et al. 2021; Matthee et al. 2004). Although generally associated with grasslands, they are very widespread and adapted to very distinct ecosystems, from deserts (Cape hare, *L. capensis*) to Artic biomes (artic hare, *L. articus*), from temperate (European hare, *L. europaeus*) to cold climates (mountain hare, *L. timidus*), from sea level to high elevation (Ethiopian highland hare, *L. starcki*) (Smith 2018). Species generally have parapatric distributions, and the range shifts due to climate oscillations have promoted opportunities for natural hybridization between lineages (Ferreira et al. 2021; Matthee et al. 2004; Melo-Ferreira et al. 2012b). Hybridization has been described across several hare species pairs (e.g. Alves et al. 2008; Giska et al. 2019; Jones et al. 2018; Levänen et al. 2018; Melo-Ferreira et al. 2012b; 2005; Seixas et al. 2018; Thulin, Jaarola, and Tegelström 1997), in current but also in ancient contacts, and in some cases resulted in adaptive introgression (Giska et al. 2019; Jones et al. 2018), suggesting that introgressed variation can be pervasive in the gene pool of species, and questioning its role in adaptive evolution.

Over the last decades, hares have been established as an exciting model to study hybridization between closely related species. Introgressive hybridization is an important and widespread evolutionary process, but the relative roles of neutral demography and natural selection in promoting introgression are difficult to assess and an important matter of debate. Hares from Southern Europe provide an appropriate system to study this question. Numerous and sequential range invasions with hybridization between species occurred since the Pleistocene, and the genetic variation detected in modern populations witness these events (e.g. Melo-Ferreira et al., 2014, 2014; Seixas et al., 2018).

The Iberian Peninsula is currently occupied by three hare species: the endemic Iberian hare (*L. granatensis*) that occurs in most of the territory, the broom hare (*L. castroviejoi*) currently restricted to the Cantabrian Mountains and a result of a recent fragmentation (Acevedo et al. 2014; Alves et al. 2008; Estonba et al. 2006), and the widely distributed European hare (*L. europaeus*) that likely invaded north-eastern Spain after the last glacial maximum (Alves et al. 2008; Melo-Ferreira et al. 2014b; 2005; Sanz-Martín et al. 2014). All these species show high frequencies of mitochondrial DNA (mtDNA) haplotypes introgressed from *L. timidus*, as a result of hybridization at the end of the last glacial period before the species retreated north (Melo-Ferreira et al. 2005; Seixas, Boursot, and Melo-Ferreira 2018). The mountain hare, *Lepus timidus*, is a boreal

species currently distributed in the northern Palearctic and in some isolated populations such as Ireland, Scotland, Poland and the Alps, but fossils from southern France or northern Iberian Peninsula show that it inhabited southern Europe during the Pleistocene (Altuna 1970). In the broom hare, introgression has led to the complete replacement of its native mitogenome and that of its sister-taxa Italian hare (*L. corsicanus*). Iberian populations of European hares harbour very high frequencies of mtDNA of *timidus* origin, which appears almost fixed, and the result from the Iberian colonization and its replacement of a resident species (still uncertain if the mountain hare or the Iberian hare) (Paulo C. Alves et al. 2008; Melo-Ferreira et al. 2005). In the Iberian hare, mtDNA introgression is strongly structured, being absent in the south and increasing in frequency towards the north. This pattern appears compatible with a northwards expansion after the last glacial maximum, replacing and hybridizing with the mountain hare, and spreading the traces of introgression to the north (Melo-Ferreira et al. 2014b; Melo-Ferreira et al. 2007; Seixas, Boursot, and Melo-Ferreira 2018). Although the northwards invasion hypothesis requires population genetics validation, this scenario makes it a potentially outstanding model to understand the impact of purely demographic processes on genome landscape (Seixas, Boursot, and Melo-Ferreira 2018).

The Iberian system promises to undisclose many open questions provided the right questions are asked and the right tools used to address them. Despite the great amount of genomic data produced in the last decade, it is still mandatory complement them with system-specific genomic resources.

1.7.2. The house mouse (*Mus musculus*) as a model to study hybridization and speciation

The house mouse (*Mus musculus*) has long been a scientific “swiss army knife”. The use of house mice in science can be at least tracked back to the 17th century in the early days of modern science, when William Harvey used them in his comparative anatomy experiences (Macholán et al. 2012). At the beginning of 20th century, Abbie E. C. Lathrop, a methodical breeder but also self-made scientist, recognized the animals' potential for genetic research and help establish the standard mouse model (Steensma, Kyle, and Shampo 2010). In 1965 Sturtevant suggested that Mendel had recognized the same potential several decades before. Sturtevant implied that Mendel has originally worked his ‘laws’ in mice, but had to suppress his findings fearing antagonizing his ecclesiastical superiors (mice were considered voluptuous and libidinous creatures) (Macholán et al.

2012). Since then, “lab mice” have been used to make great scientific and medical advances, from cancer drugs and HIV antiretrovirals to the yearly flu vaccine. There are over 450 laboratory strains and their variants have already been characterized and represent most of all laboratory animals (Beck et al. 2000).

House mice also provide a valuable model for evolutionary genetics and speciation, particularly to shed light on the genetic basis of reproductive isolation and on the impacts of hybridization to speciation (Dod et al. 1993; Forejt and Iványi 1974; Janoušek et al. 2012; Larson et al. 2018; Mihola et al. 2009; Phifer-Rixey, Harr, and Hey 2020; Phifer-Rixey, Bomhoff, and Nachman 2014; L. M. Turner and Harr 2014; Vanlerberghe et al. 1986; White et al. 2011). *Mus musculus* forms a species complex genetically structured in three well accepted and several not yet fully described evolutionary units (Guénet and Bonhomme 2003; Hardouin et al. 2015; Phifer-Rixey and Nachman 2015). The well-established subspecies, thought to have originated in the northern part of the Indian Subcontinent (Boursot et al. 1993; Din et al. 1996), are currently distributed parapatrically. *Mus musculus domesticus* is distributed from southwest Iran to the Middle east, around the Mediterranean, in Western Europe (including Norway (E. P. Jones et al. 2010)) and recently expanded its range to the Americas and Oceania transported by humans. *Mus musculus musculus* is present from Eastern Europe to Northern Asia. And *Mus musculus castaneus* is distributed from Central to Southeast Asia. Japanese house mice have been given a specific name, *Mus musculus molossinus*, but shown to be a hybrid between *musculus* and *castaneus* (Hiromichi Yonekawa et al. 1981). All the entities above have been well characterised from their distributions at the periphery of the Eurasian Continent. The situation in more central regions (Iran and its surroundings) is less well understood and additional taxonomic denominations are used in the recent literature. For instance *Mus musculus gentilulus* is described in the south-eastern coast of Arabia to Oman (Duplantier et al. 2002; Harrison 1970; Prager, Orrego, and Sage 1998) based on its specific mitochondrial lineage, and two other names used to represent variation found in Iran: *M. m. isatissus* (Hamid et al. 2017; Hardouin et al. 2015) and *M. m. bactrianus* (Hardouin et al. 2015; Schwarz and Schwarz 1943).

The history of house mouse differentiation is mainly documented either for old or very recent periods (reviewed in Boursot et al., 1993). The major subspecies are believed to have diverged at roughly the same time ~130,000-420,000 years ago, (similar to the split of chimpanzees and humans in numbers of generations and relative molecular divergence) (Geraldes et al. 2008, 2011; Phifer-Rixey, Harr, and Hey 2020), spread across the world through diverse habitats and come into secondary contact much more

recently (e.g. Cucchi et al., 2005; Duvaux et al., 2011). Interestingly, they have all become commensal with humans between the Mesolithic and Neolithic (Cucchi, Auffray, and Vigne 2012; Weissbrod et al. 2017) and consequently have spread in the context of developing human activities in Europe and Asia within the past few thousand years, followed by a colonization of the rest of the world in the wake of transcontinental transport within the past few hundred years (Macholán et al. 2012). The evolutionary history is thus complicated by their relatively recent divergence, ongoing gene flow, and complex demographic histories.

Each subspecies meets and hybridizes with the other two sub-species where their ranges come into contact (e.g. Boursot et al., 1993; Duvaux et al., 2011; Fujiwara et al., 2021; Phifer-Rixey and Nachman, 2015; Teeter et al., 2008), creating hybrid zones. The European hybrid zone (EHZ) between *musculus* and *domesticus* is a narrow 2500 km long contact zone, that extends from the Jutland Peninsula in Denmark (Raufaste et al. 2005) to the Black Sea in Bulgaria (Macholán, Kryštufek, and Vohralík 2003) and is the best studied (Janoušek et al. 2015a; Macholán et al. 2008a; Martincová et al. 2019; Teeter et al. 2008). The East Asian contact between *musculus* and *castaneus* is a broader hybrid zone (Boursot et al. 1993), where the prevalence of hybridization was recognized to be more frequent than in the EHZ (suggested by Boursot *et al* (Boursot et al. 1993) as recently confirmed (H Suzuki et al. 2013; Hitoshi Suzuki et al. 2015; Fujiwara et al. 2021). Finally, the Iranian plateau, often considered the species cradle of differentiation, is the place in the world with the highest apparent diversity of mouse lineages within the same geographical area (Hardouin et al. 2015). This region however remains poorly studied. In 2015, *Hardouin* and collaborators (Hardouin et al. 2015) analysed a battery of microsatellite loci, using extensive sampling (963 individuals from 47 populations), and produced the most comprehensive genetic study done in the region. The results confirmed the presence of the three well described subspecies (*musculus* in the North, *domesticus* in the west and *castaneus* in the east). Furthermore, were also describe two putative new distinct genetic groups in Central and Southeast Iran (roughly corresponding to the proposed *M. m. isatissus* (Hamid et al., 2017) and *M. m. bactrianus* (Schwarz and Schwarz, 1943). Populations in and around the region likely represent a long history of divergence with gene flow, which presumably provides multiple replicates of the process, allowing to capture the sorting of genomic incompatibilities across different pairs of subspecies.

The combination of laboratory crosses with the study of hybrid zones (particularly the EHZ), has revealed that the degree of reproductive isolation differs in pairwise comparisons among subspecies. Lab crosses between *domesticus* and *musculus* result in significant reductions of male fertility (e.g. Good et al., 2008; White et al., 2012), although studies in the EHZ reveal introgression across the contact zone, inclusively and particularly striking a massive unidirectional *musculus* Y chromosome invasion into the *domesticus* range, a situation that contradicts the Haldane's rule (Macholán et al. 2008a; Martincová et al. 2019). In contrast, crosses between *castaneus* and *domesticus* have not led to a significant infertility until the F2 generation (even here the degree of infertility is not as severe) (White et al. 2012). Regarding *castaneus* and *musculus,* there are no published studies documenting reduced fertility in lab crosses, which is in line with the use of a cross between *castaneus* and *musculus* for a recombination mapping study in which infertility was not observed (Dumont and Payseur 2011) and the description of *molossinus*, as a hybrid between *musculus* and *castaneus* (Takada et al. 2013; H Yonekawa et al. 1988).

The unparalleled amount of genomic resources (reference-genomes, recombination maps, hybridization crosses, SNP databases, etc), in combination with its recent divergence with ongoing gene flow and complex demographic histories, makes the house mouse a powerful system to understand the genetic mechanisms generating barriers among the diverging subspecies during the early stages of speciation. In particular, the Iranian plateau and its numerous natural contact zones appears to be an important territory to understand the origin of the radiation.

1.8. Objectives

The major goal of this thesis was to contribute to the understanding of major processes driving relevant evolutionary mechanisms in hares and mice, most notably related with those underlying admixture between species and the establishment of genomic barriers to gene flow (and that may be involved in speciation). Further, since genomic resources for hares were yet scarce, we aimed at generating new resources that can be used in the future to deepen evolutionary studies in the genus. To investigate the consequences of hybridization in the species formation, we took advantage of published datasets (Harr et al. 2016; Thybert et al. 2018) and extensive population genomics data (whole genome sequences of population samples) newly collected for this work. It concerns several recently diverged species of hares (*genus Lepus*) and mice (*genus Mus*). Inside each genus they represent various degrees of divergence, of gene

flow and ecological differentiation, and thus constitute a potential appropriate framework to explore comparative evolutionary mechanisms in two distinct systems.

On the one hand, this thesis contributes to generate resources that can be used in the future to study the genomics of speciation and hybridization. On the other hand, it also contributes to understand processes governing the isolation and admixture of diverging genomes. We explore the landscape of genomic divergence and differentiation between several pairs of recently diverged taxa with varying degrees of divergence, hybridization, and ecological differentiation. We address the role of natural selection in genome divergence and reproductive isolation leading to speciation, but also in admixture particularly by introgression.

More specifically, this thesis had the following objectives:

i. Increase the genomic resources available for hares and mice. This comprises the production of the first annotated genome assembly of a hare species, as well as the first *L. timidus* transcriptome assembly and a set of diagnostic SNPs between *L. timidus* and *L. europaeus* that can be easily used as an important hybridization monitoring tool. Additionally, it includes the generation of large amounts of whole genome sequences of house mouse population samples that can play an important role to help increasing the knowledge about the system and its evolutionary history.

ii. Understand the relative roles of neutral demography and natural selection as introgression promoting processes.

iii. Clarify the evolutionary history of the house mouse populations from the Iranian plateau – the cradle of diversification of the species – and particularly clarify the genomic makeup and evolutionary origin of the central Iranian population.

iv. Contribute to understand the role of sex chromosomes in introgression and isolation processes, and its impacts on whole genome admixture landscape among species.

1.9. Thesis organization

This thesis is organized in four chapters. In chapter I, I provide a general introduction to the main themes of this work, such as the current understanding of the genomic process of speciation, the relevance of hybridization and genetic exchanges between diverging populations, and the power to study these processes with the advent of next generation sequencing. In chapter II, I present two articles where genomic resources for

hares were developed, which can be used for future evolutionary studies in the system. The first paper describes the assembly of the first mountain hare (*Lepus timidus*) transcriptome and the discovery of a set of diagnostic SNPs that can be used as a tool for studying species admixture, particularly in the numerous contact zones where the mountain hare tends to be replaced by the invading European hare (*Lepus europaeus*). This study has been published at the SCI indexed journal *Scientific Data* (João P. Marques, Ferreira, et al. 2017). The second paper in Chapter II, describes a major genomic resource for hares: the first annotated draft genome of a hare species. This represents a paradigmatic change, as current genomic studies on the system are constrained by the use of the high-quality European rabbit (*Oryctolagus cuniculus*) reference genome. The new hare reference genome is therefore a new resource to uncover and explore hare-specific variation. This work was published in the SCI indexed journal *Genome Biology and Evolution* (João P. Marques et al. 2020).

In Chapter III, I present three evolutionary studies on hares and mice. In the first publication of this chapter, we investigate the role of range expansion and species replacement as the major force promoting hybridization and mitochondrial DNA introgression from the arctic/boreal mountain hare *(L. timidus)* into the Iberian hare (*L. granatensis*). This publication also resulted in the production of the Iberian hare most complete and annotated transcriptome. This work was published in the SCI indexed journal *Scientific Reports* (João P. Marques, Farelo, et al. 2017). In the second manuscript of Chapter III, we use whole genome sequences of various populations to demonstrate that mice from Central Iran that result from the ancient admixture of two presently reproductively isolated subspecies. The manuscript is currently under preparation for submission to a peer reviewed journal. Finally, in the last manuscript of this thesis, also based on whole genome sequences, we characterise admixture patterns in NW Iran, in particular those of the sex-chromosomes that appear discordant. The manuscript is currently under preparation for submission to a peer reviewed journal.

Chapter IV recapitulates the main conclusions of this thesis as well as a personal view about the future directions to the field in the transforming times that we live with the power given by the new sequencing technologies. A final brief note is given about the responsibility of each scientist as a science communicator in an Era marked by climatic changes causing species extinction, which needs to be addressed not only by the scientists but also by the general society.

1.10. References

Acevedo, Pelayo, Alberto Jiménez-Valverde, José Melo-Ferreira, Raimundo Real, and Paulo Célio Alves. 2012a. 'Parapatric Species and the Implications for Climate Change Studies: A Case Study on Hares in Europe'. *Global Change Biology* 18 (5): 1509–19. https://doi.org/10.1111/j.1365-2486.2012.02655.x.

———. 2012b. 'Parapatric Species and the Implications for Climate Change Studies: A Case Study on Hares in Europe'. *Global Change Biology* 18 (5): 1509–19. https://doi.org/10.1111/j.1365-2486.2012.02655.x.

Acevedo, Pelayo, José Melo-Ferreira, Raimundo Real, and Paulo C. Alves. 2014. 'Evidence for Niche Similarities in the Allopatric Sister Species Lepus Castroviejoi and Lepus Corsicanus'. *Journal of Biogeography* 41 (5): 977–86. https://doi.org/10.1111/jbi.12270.

Altuna, Echave. 1970. 'Hallazgo de una liebre ártica (Lepus timidus L.) en el yacimiento prehistórico de Urtiaga (Guipúzcoa)', 4.

Alves, Paulo C., José Melo-Ferreira, Hélder Freitas, and Pierre Boursot. 2008. 'The Ubiquitous Mountain Hare Mitochondria: Multiple Introgressive Hybridization in Hares, Genus Lepus'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1505): 2831–39. https://doi.org/10.1098/rstb.2008.0053.

Alves, P.C., J. Melo-Ferreira, M. Branco, F. Suchentrunk, N. Ferrand, and D.J. Harris. 2008. '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): 1191–97. https://doi.org/10.1016/j.ympev.2007.11.010.

Amoutzias, Grigoris D., Themistoklis Giannoulis, Katerina A. Moutou, Anna Maria G Psarra, Costas Stamatis, Andreas Tsipourlianos, and Zissis Mamuris. 2016. 'SNP Identification through Transcriptome Analysis of the European Brown Hare (Lepus Europaeus): Cellular Energetics and Mother's Curse'. *PLoS ONE* 11 (7). https://doi.org/10.1371/journal.pone.0159939.

Andrews, Simon. 2010a. 'FastQC'. *Babraham Bioinformatics*. https://doi.org/citeulike-article-id:11583827.

———. 2010b. 'FastQC: A Quality Control Tool for High Throughput Sequence Data.' *Babraham Bioinformatics*, http://www.bioinformatics.babraham.ac.uk/projects/. https://doi.org/citeulike-article-id:11583827.

Arnegard, Matthew E., Matthew D. McGee, Blake Matthews, Kerry B. Marchinko, Gina L. Conte, Sahriar Kabir, Nicole Bedford, et al. 2014. 'Genetics of Ecological Divergence during Speciation'. *Nature* 511 (7509): 307–11. https://doi.org/10.1038/nature13301.

Arnold, Michael L. 1992. 'Natural Hybridization as an Evolutionary Process'. *Annual Review of Ecology and Systematics* 23 (1): 237–61. https://doi.org/10.1146/annurev.es.23.110192.001321.

Aubry, Sylvain, Steven Kelly, Britta M C Kümpers, Richard D. Smith-Unna, and Julian M. Hibberd. 2014. 'Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C4 Photosynthesis'. *PLoS Genetics* 10 (6). https://doi.org/10.1371/journal.pgen.1004365.

Baker, Christopher L., Shimpei Kajita, Michael Walker, Ruth L. Saxl, Narayanan Raghupathy, Kwangbom Choi, Petko M. Petkov, and Kenneth Paigen. 2015. 'PRDM9 Drives Evolutionary Erosion of Hotspots in Mus Musculus through Haplotype-Specific Initiation of Meiotic Recombination'. *PLOS Genetics* 11 (1): e1004916. https://doi.org/10.1371/journal.pgen.1004916.

Barton, N. H., and G. M. Hewitt. 1989. 'Adaptation, Speciation and Hybrid Zones'. *Nature* 341 (6242): 497–503. https://doi.org/10.1038/341497a0.

Barton, Nick, and Bengt Olle Bengtsson. 1986. 'The Barrier to Genetic Exchange between Hybridising Populations'. *Heredity* 57 (3): 357–76. https://doi.org/10.1038/hdy.1986.135.

Bateson. 1909. *Heredity and Variation in Modern Lights*. http://archive.org/details/cbarchive_36304_hereditydandvariationinmodernli1909.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop, and B. de Massy. 2010. 'PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice'. *Science (New York, N.Y.)* 327 (5967): 836–40. https://doi.org/10.1126/science.1183439.

Bayes, Joshua J., and Harmit S. Malik. 2009. 'Altered Heterochromatin Binding by a Hybrid Sterility Protein in Drosophila Sibling Species'. *Science (New York, N.Y.)* 326 (5959): 1538–41. https://doi.org/10.1126/science.1181756.

Beck, Jon A., Sarah Lloyd, Majid Hafezparast, Moyha Lennon-Pierce, Janan T. Eppig, Michael F. W. Festing, and Elizabeth M. C. Fisher. 2000. 'Genealogies of Mouse Inbred Strains'. *Nature Genetics* 24 (1): 23–25. https://doi.org/10.1038/71641.

Böhne, Astrid, Frédéric Brunet, Delphine Galiana-Arnoux, Christina Schultheis, and Jean-Nicolas Volff. 2008. 'Transposable Elements as Drivers of Genomic and Biological Diversity in Vertebrates'. *Chromosome Research* 16 (1): 203–15. https://doi.org/10.1007/s10577-007-1202-6.

Bolger, a. M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data'. *Bioinformatics* 30 (15): 2114–20.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data'. *Bioinformatics (Oxford, England)* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Boursot, P., J.-C. Auffray, J. Britton-Davidian, and F. Bonhomme. 1993. 'The Evolution of House Mice'. *Annual Review of Ecology and Systematics* 24: 119–52.

Boutet, Emmanuel, Damien Lieberherr, Michael Tognolli, Michel Schneider, Parit Bansal, Alan J Bridge, Sylvain Poux, Lydie Bougueleret, and Ioannis Xenarios. 2016. 'UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View.' *Methods in Molecular Biology (Clifton, N.J.)* 1374: 23–54. https://doi.org/10.1007/978-1-4939-3167-5_2.

Brandvain, Yaniv, Amanda M. Kenney, Lex Flagel, Graham Coop, and Andrea L. Sweigart. 2014. 'Speciation and Introgression between Mimulus Nasutus and Mimulus Guttatus'. Edited by Chris D. Jiggins. *PLoS Genetics* 10 (6): e1004410. https://doi.org/10.1371/journal.pgen.1004410.

Brennecke, Julius, Colin D. Malone, Alexei A. Aravin, Ravi Sachidanandam, Alexander Stark, and Gregory J. Hannon. 2008. 'An Epigenetic Role for Maternally Inherited PiRNAs in Transposon Silencing'. *Science (New York, N.Y.)* 322 (5906): 1387–92. https://doi.org/10.1126/science.1165171.

Brideau, Nicholas J., Heather A. Flores, Jun Wang, Shamoni Maheshwari, Xu Wang, and Daniel A. Barbash. 2006. 'Two Dobzhansky-Muller Genes Interact to Cause Hybrid Lethality in Drosophila'. *Science (New York, N.Y.)* 314 (5803): 1292–95. https://doi.org/10.1126/science.1133953.

Brown, Judith D., and Rachel J. O'Neill. 2010. 'Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited'. *Annual Review of Genomics and Human Genetics* 11 (1): 291–316. https://doi.org/10.1146/annurev-genom-082509-141554.

Brown, Judith D., Vanessa Piccuillo, and Rachel J. O'Neill. 2012. 'Retroelement Demethylation Associated with Abnormal Placentation in Mus Musculus × Mus Caroli Hybrids1'. *Biology of Reproduction* 86 (3): 88, 1–10. https://doi.org/10.1095/biolreprod.111.095273.

Burt, Austin, and Robert Trivers. 2006. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Cambridge, Mass: Belknap Press of Harvard University Press.

Butlin, Roger K, and Sean Stankowski. 2020. 'Is It Time to Abandon the Biological Species Concept? No'. *National Science Review* 7 (8): 1400–1401. https://doi.org/10.1093/nsr/nwaa109.

Calfee, Erin, Daniel Gates, Anne Lorant, M. Taylor Perkins, Graham Coop, and Jeffrey Ross-Ibarra. 2021. 'Selective Sorting of Ancestral Introgression in Maize and Teosinte along an Elevational Cline'. https://doi.org/10.1101/2021.03.05.434040.

Camus, M. Florencia, Jochen B. W. Wolf, Edward H. Morrow, and Damian K. Dowling. 2015. 'Single Nucleotides in the MtDNA Sequence Modify Mitochondrial Molecular Function and Are Associated with Sex-Specific Effects on Fertility and Aging'. *Current Biology* 25 (20): 2717–22. https://doi.org/10.1016/j.cub.2015.09.012.

Caravaggi, Anthony, Katie Leach, Francesco Santilli, Jukka Rintala, Pekka Helle, Juha Tiainen, Francesco Bisi, Adriano Martinoli, W Ian Montgomery, and Neil Reid. 2016. 'Niche Overlap of Mountain Hare Subspecies and the Vulnerability of Their Ranges to Invasion by the European Hare; the (Bad) Luck of the Irish'. *Biological Invasions*, 1–20. https://doi.org/10.1007/s10530-016-1330-z.

Caravaggi, Anthony, Katie Leach, Francesco Santilli, Jukka Rintala, Pekka Helle, Juha Tiainen, Francesco Bisi, Adriano Martinoli, W. Ian Montgomery, and Neil Reid. 2017. 'Niche Overlap of Mountain Hare Subspecies and the Vulnerability of Their Ranges to Invasion by the European Hare; the (Bad) Luck of the Irish'. *Biological Invasions* 19 (2): 655–74. https://doi.org/10.1007/s10530-016-1330-z.

Caravaggi, Anthony, W Ian Montgomery, and Neil Reid. 2015. 'Range Expansion and Comparative Habitat Use of Insular, Congeneric Lagomorphs: Invasive European Hares Lepus Europaeus and Endemic Irish Hares Lepus Timidus Hibernicus'. *Biological Invasions* 17 (2): 687–98. https://doi.org/10.1007/s10530-014-0759-1.

Case, Andrea L., Findley R. Finseth, Camille M. Barr, and Lila Fishman. 2016. 'Selfish Evolution of Cytonuclear Hybrid Incompatibility in Mimulus'. *Proceedings of the Royal Society B: Biological Sciences* 283 (1838): 20161493. https://doi.org/10.1098/rspb.2016.1493.

Chafee, Meghan E., Courtney N. Zecher, Michelle L. Gourley, Victor T. Schmidt, John H. Chen, Sarah R. Bordenstein, Michael E. Clark, and Seth R. Bordenstein. 2011. 'Decoupling of Host–Symbiont–Phage Coadaptations Following Transfer Between Insect Species'. *Genetics* 187 (1): 203–15. https://doi.org/10.1534/genetics.110.120675.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. 'The Effect of Deleterious Mutations on Neutral Molecular Variation'. *Genetics* 134 (4): 1289–1303.

Chase, Kevin, David R. Carrier, Frederick R. Adler, Elaine A. Ostrander, and Karl G. Lark. 2005. 'Interaction between the X Chromosome and an Autosome Regulates Size Sexual Dimorphism in Portuguese Water Dogs'. *Genome Research* 15 (12): 1820–24. https://doi.org/10.1101/gr.3712705.

Christie, Kyle, and Sharon Y. Strauss. 2018. 'Along the Speciation Continuum: Quantifying Intrinsic and Extrinsic Isolating Barriers across Five Million Years of Evolutionary Divergence in California Jewelflowers'. *Evolution; International Journal of Organic Evolution* 72 (5): 1063–79. https://doi.org/10.1111/evo.13477.

Christie, Paul, and Mark R. Macnair. 1987. 'The Distribution of Postmating Reproductive Isolating Genes in Populations of the Yellow Monkey Flower, Mimulus Guttatus'. *Evolution* 41 (3): 571–78. https://doi.org/10.1111/j.1558-5646.1987.tb05827.x.

Consortium, Heliconius Genome. 2012. 'Butterfly Genome Reveals Promiscuous Exchange of Mimicry Adaptations among Species.' *Nature* 487 (7405): 94–98. https://doi.org/10.1038/nature11041.

Corbett-Detig, Russell, and Rasmus Nielsen. 2017. 'A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy'. *PLOS Genetics* 13 (1): e1006529. https://doi.org/10.1371/journal.pgen.1006529.

Courret, Cécile, Ching-Ho Chang, Kevin H.-C. Wei, Catherine Montchamp-Moreau, and Amanda M. Larracuente. 2019. 'Meiotic Drive Mechanisms: Lessons from Drosophila'. *Proceedings of the Royal Society B: Biological Sciences* 286 (1913): 20191430. https://doi.org/10.1098/rspb.2019.1430.

Coyne, Jerry A., and H. Allen Orr. 1997. '"Patterns of Speciation in Drosophila" Revisited'. *Evolution* 51 (1): 295–303. https://doi.org/10.2307/2410984.

———. 2004. *Speciation*. Oxford, New York: Oxford University Press.

Crespi, Bernard. 2008. 'Turner Syndrome and the Evolution of Human Sexual Dimorphism'. *Evolutionary Applications* 1 (3): 449–61. https://doi.org/10.1111/j.1752-4571.2008.00017.x.

Crespi, Bernard, and Patrik Nosil. 2013. 'Conflictual Speciation: Species Formation via Genomic Conflict'. *Trends in Ecology & Evolution* 28 (1): 48–57. https://doi.org/10.1016/j.tree.2012.08.015.

Cruickshank, Tami E., and Matthew W. Hahn. 2014. 'Reanalysis Suggests That Genomic Islands of Speciation Are Due to Reduced Diversity, Not Reduced Gene Flow.' *Molecular Ecology* 23 (13): 3133–57. https://doi.org/10.1111/mec.12796.

Cucchi, Thomas, Jean-Christophe Auffray, and Jean-Denis Vigne. 2012. 'On the Origin of the House Mouse Synanthropy and Dispersal in the Near East and Europe: Zooarchaeological Review and Perspectives'. In , 65–93.

Cucchi, Thomas, Jean-Denis Vigne, and Jean-Christophe Auffray. 2005. 'First Occurrence of the House Mouse (Mus Musculus Domesticus Schwarz & Schwarz, 1943) in the Western Mediterranean: A Zooarchaeological Revision of Subfossil Occurrences'. *Biological Journal of the Linnean Society* 84 (3): 429–45. https://doi.org/10.1111/j.1095-8312.2005.00445.x.

Currat, Mathias, Manuel Ruedi, Rémy J. Petit, and Laurent Excoffier. 2008. 'The Hidden Side of Invasions: Massive Introgression by Local Genes'. *Evolution* 62 (8): 1908–20. https://doi.org/10.1111/j.1558-5646.2008.00413.x.

Darwin, Charles, and Leonard Kebler. 1859. *On the Origin of Species by Means of Natural Selection*. https://www.loc.gov/item/06017473/.

Delmore, Kira E., and Darren E. Irwin. 2014. 'Hybrid Songbirds Employ Intermediate Routes in a Migratory Divide'. *Ecology Letters* 17 (10): 1211–18. https://doi.org/10.1111/ele.12326.

Din, W., R. Anand, P. Boursot, D. Darviche, B. Dod, E. Jouvin-Marche, A. Orth, G. P. Talwar, P.-A. Cazenave, and F. Bonhomme. 1996. 'Origin and Radiation of the House Mouse: Clues from Nuclear Genes'. *Journal of Evolutionary Biology* 9 (5): 519–39. https://doi.org/10.1046/j.1420-9101.1996.9050519.x.

Dobzhansky, Theodosius, and Stephen Jay Gould. 1982. *Genetics and the Origin of Species*. New York: Columbia University Press.

Dod, Barbara, Lars S. Jermiin, Pierre Boursot, Verne H. Chapman, J. Tönnes Nielsen, and François Bonhomme. 1993. 'Counterselection on Sex Chromosomes in the Mus Musculus European Hybrid Zone'. *Journal of Evolutionary Biology* 6 (4): 529–46. https://doi.org/10.1046/j.1420-9101.1993.6040529.x.

Dumont, Beth L., and Bret A. Payseur. 2011. 'Genetic Analysis of Genome-Scale Recombination Rate Evolution in House Mice'. *PLoS Genet* 7 (6): e1002116.

Duplantier, J.-M., A. Orth, J. Catalan, and F. Bonhomme. 2002. 'Evidence for a Mitochondrial Lineage Originating from the Arabian Peninsula in the Madagascar House Mouse (Mus Musculus)'. *Heredity* 89 (2): 154–58. https://doi.org/10.1038/sj.hdy.6800122.

Duranton, Maud, and John Pool. 2021. 'Interactions between Natural Selection and Recombination Shape the Genomic Landscape of Introgression'. Preprint. Evolutionary Biology. https://doi.org/10.1101/2021.10.12.462572.

Duvaux, Ludovic, Khalid Belkhir, Matthieu Boulesteix, and Pierre Boursot. 2011. 'Isolation and Gene Flow: Inferring the Speciation History of European House Mice: HOUSE MOUSE SPECIATION HISTORY'. *Molecular Ecology* 20 (24): 5248–64. https://doi.org/10.1111/j.1365-294X.2011.05343.x.

Duvernell, David D., Eric Westhafer, and Jacob F. Schaefer. 2019. 'Late Pleistocene Range Expansion of North American Topminnows Accompanied by Admixture and Introgression'. *Journal of Biogeography* 46 (9): 2126–40. https://doi.org/10.1111/jbi.13597.

Edmonds, Christopher A., Anita S. Lillie, and L. Luca Cavalli-Sforza. 2004. 'Mutations Arising in the Wave Front of an Expanding Population'. *Proceedings of the National Academy of Sciences* 101 (4): 975–79. https://doi.org/10.1073/pnas.0308064100.

El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R Eddy, Aurélien Luciani, Simon C Potter, Matloob Qureshi, et al. 2019. 'The Pfam Protein Families Database in 2019'. *Nucleic Acids Research* 47 (D1): D427–32. https://doi.org/10.1093/nar/gky995.

Elgvin, Tore O., Cassandra N. Trier, Ole K. Tørresen, Ingerid J. Hagen, Sigbjørn Lien, Alexander J. Nederbragt, Mark Ravinet, Henrik Jensen, and Glenn-Peter Sætre. 2017. 'The Genomic Mosaicism of Hybrid Speciation'. *Science Advances* 3 (6): e1602996. https://doi.org/10.1126/sciadv.1602996.

Enk, Jacob, Alison Devault, Regis Debruyne, Christine E. King, Todd Treangen, Dennis O'Rourke, Steven L. Salzberg, Daniel Fisher, Ross MacPhee, and Hendrik Poinar. 2011. 'Complete Columbian Mammoth Mitogenome Suggests Interbreeding with Woolly Mammoths'. *Genome Biology* 12 (5): R51. https://doi.org/10.1186/gb-2011-12-5-r51.

Enk, Jacob, Alison Devault, Christopher Widga, Jeffrey Saunders, Paul Szpak, John Southon, Jean-Marie Rouillard, et al. 2016. 'Mammuthus Population Dynamics in Late Pleistocene North America: Divergence, Phylogeography, and Introgression'. *Frontiers in Ecology and Evolution* 4: 42. https://doi.org/10.3389/fevo.2016.00042.

Estonba, Andone, Ainhoa Solís, M. Iriondo, María J. Sanz-Martín, G. Pérez-Suárez, G. Markov, and F. Palacios. 2006. 'The Genetic Distinctiveness of the Three Iberian Hare Species: Lepus Europaeus, L. Granatensis, and L. Castroviejoi'. *Mammalian Biology* 71 (1): 52–59. https://doi.org/10.1016/j.mambio.2005.08.010.

Excoffier, Laurent, Matthieu Foll, and Rémy J. Petit. 2009. 'Genetic Consequences of Range Expansions'. *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414.

Excoffier, Laurent, and Nicolas Ray. 2008. 'Surfing during Population Expansions Promotes Genetic Revolutions and Structuration'. *Trends in Ecology & Evolution* 23 (7): 347–51. https://doi.org/10.1016/j.tree.2008.04.004.

Ferreira, M. S., P. C. Alves, C. M. Callahan, J. P. Marques, L. S. Mills, J. M. Good, J. Melo-Ferreira. 2017. The transcriptional landscape of seasonal coat colour moult in the snowshoe hare. Molecular Ecology. 26:4173-4185.

Ferreira, M. S., Alves, P. C., Callahan, C. M., Giska, I., Farelo, L., Jenny, H., Mills, L. S., Hackländer, K., Good, J. M., & Melo-Ferreira, J. (2020). Transcriptomic regulation of seasonal coat color change in hares. Ecology and evolution, 10(3), 1180–1192. https://doi.org/10.1002/ece3.5956

Ferreira, Mafalda S, Matthew R Jones, Colin M Callahan, Liliana Farelo, Zelalem Tolesa, Franz Suchentrunk, Pierre Boursot, et al. 2021. 'The Legacy of Recurrent

Introgression during the Radiation of Hares'. *Systematic Biology* 70 (3): 593–607. https://doi.org/10.1093/sysbio/syaa088.

Finn, Robert D., Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Andreas Heger, et al. 2014. 'Pfam: The Protein Families Database'. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt1223.

Finn, Robert D., Jody Clements, and Sean R. Eddy. 2011. 'HMMER Web Server: Interactive Sequence Similarity Searching'. *Nucleic Acids Research* 39 (SUPPL. 2). https://doi.org/10.1093/nar/gkr367.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. The Genetical Theory of Natural Selection. Oxford, England: Clarendon Press. https://doi.org/10.5962/bhl.title.27468.

Fishman, Lila, and John H. Willis. 2006. 'A CYTONUCLEAR INCOMPATIBILITY CAUSES ANTHER STERILITY IN MIMULUS HYBRIDS'. *Evolution* 60 (7): 1372–81. https://doi.org/10.1554/05-708.1.

Fontaine, Michael C., James B. Pease, Aaron Steele, Robert M. Waterhouse, Daniel E. Neafsey, Igor V. Sharakhov, Xiaofang Jiang, et al. 2015. 'Extensive Introgression in a Malaria Vector Species Complex Revealed by Phylogenomics'. *Science (New York, N.Y.)* 347 (6217): 1258524. https://doi.org/10.1126/science.1258524.

Forejt, J., and P. Iványi. 1974. 'Genetic Studies on Male Sterility of Hybrids between Laboratory and Wild Mice (Mus Musculus L.)'. *Genetics Research* 24 (2): 189–206. https://doi.org/10.1017/S0016672300015214.

Fraïsse, Christelle, P. Alexander Gunnarsson, Denis Roze, Nicolas Bierne, and John J. Welch. 2016. 'The Genetics of Speciation: Insights from Fisher's Geometric Model'. *Evolution* 70 (7): 1450–64. https://doi.org/10.1111/evo.12968.

Frank, Steven A., and Bernard J. Crespi. 2011. 'Pathology from Evolutionary Conflict, with a Theory of X Chromosome versus Autosome Conflict over Sexually Antagonistic Traits'. *Proceedings of the National Academy of Sciences* 108 (Supplement 2): 10886–93. https://doi.org/10.1073/pnas.1100921108.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012a. 'CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data'. *Bioinformatics* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

———. 2012b. 'CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data'. *Bioinformatics* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

Fujiwara, Kazumichi, Yosuke Kawai, Kazuo Moriwaki, Toyoyuki Takada, Toshihiko Shiroishi, Naruya Saitou, Hitoshi Suzuki, and Naoki Osada. 2021. 'Insights into Mus Musculus Subspecies Population Structure across Eurasia Revealed by Whole-Genome Sequence Analysis'. https://doi.org/10.1101/2021.02.05.429881.

Gavrilets, Sergey. 2003. 'Perspective: Models of Speciation: What Have We Learned in 40 Years?' *Evolution* 57 (10): 2197–2215. https://doi.org/10.1111/j.0014-3820.2003.tb00233.x.

Gayral, Philippe, José Melo-Ferreira, Sylvain Glémin, Nicolas Bierne, Miguel Carneiro, Benoit Nabholz, Joao M. Lourenco, et al. 2013. 'Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate-Invertebrate Gap'. *PLoS Genetics* 9 (4). https://doi.org/10.1371/journal.pgen.1003457.

Gene Ontology Consortium. 2000. 'The Gene Ontology Consortium. Gene Ontology: Tool for the Unification of Biology'. *Nature Genetics* 25 (1): 25–29. https://doi.org/10.1038/75556.

Geraldes, A., P. Basset, K. L. Smith, and M. W. Nachman. 2011. 'Higher Differentiation among Subspecies of the House Mouse (Mus Musculus) in Genomic Regions with Low Recombination: RECOMBINATION AND SPECIATION IN MICE'.

Molecular Ecology 20 (22): 4722–36. https://doi.org/10.1111/j.1365-294X.2011.05285.x.

Geraldes, Armando, Patrick Basset, Barbara Gibson, Kimberly L. Smith, Bettina Harr, Hon-Tsen Yu, Nina Bulatova, Yaron Ziv, and Michael W. Nachman. 2008. 'Inferring the History of Speciation in House Mice from Autosomal, X-Linked, Y-Linked and Mitochondrial Genes'. *Molecular Ecology* 17 (24): 5349–63. https://doi.org/10.1111/j.1365-294X.2008.04005.x.

Giska, Iwona, Liliana Farelo, João Pimenta, Fernando A. Seixas, Mafalda S. Ferreira, João P. Marques, Inês Miranda, et al. 2019. 'Introgression Drives Repeated Evolution of Winter Coat Color Polymorphism in Hares'. *Proceedings of the National Academy of Sciences*, November. https://doi.org/10.1073/pnas.1910471116.

Giska, Iwona, João Pimenta, Liliana Farelo, Pierre Boursot, Klaus Hackländer, Hannes Jenny, Neil Reid, et al. 2022. 'The Evolutionary Pathways for Local Adaptation in Mountain Hares'. *Molecular Ecology* n/a (n/a). https://doi.org/10.1111/mec.16338.

Good, Jeffrey M., Matthew D. Dean, and Michael W. Nachman. 2008. 'A Complex Genetic Basis to X-Linked Hybrid Male Sterility Between Two Species of House Mice'. *Genetics* 179 (4): 2213–28. https://doi.org/10.1534/genetics.107.085340.

Good, Jeffrey M., Thomas Giger, Matthew D. Dean, and Michael W. Nachman. 2010. 'Widespread Over-Expression of the X Chromosome in Sterile F1 Hybrid Mice'. *PLOS Genetics* 6 (9): e1001148. https://doi.org/10.1371/journal.pgen.1001148.

Gow, J. L., C. L. Peichel, and E. B. Taylor. 2007. 'Ecological Selection against Hybrids in Natural Populations of Sympatric Threespine Sticklebacks'. *Journal of Evolutionary Biology* 20 (6): 2173–80. https://doi.org/10.1111/j.1420-9101.2007.01427.x.

Grabherr, Manfred G, Brian J Haas, Moran Yassour, Joshua Z Levin, Dawn A Thompson, Ido Amit, Xian Adiconis, et al. 2011a. 'Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome.' *Nature Biotechnology* 29 (7): 644–52. https://doi.org/10.1038/nbt.1883.

———. 2011b. 'Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome'. *Nature Biotechnology* 29 (May): 644.

Guénet, Jean-Louis, and François Bonhomme. 2003. 'Wild Mice: An Ever-Increasing Contribution to a Popular Mammalian Model'. *Trends in Genetics* 19 (1): 24–31. https://doi.org/10.1016/S0168-9525(02)00007-0.

Haas & Papanicolaou et al. (2015) 2019. *TransDecoder Source. Contribute to TransDecoder/TransDecoder Development by Creating an Account on GitHub*. Perl. TransDecoder. https://github.com/TransDecoder/TransDecoder.

Haas, Brian J, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. 'De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis.' *Nature Protocols* 8 (8): 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hamid, Haddadian Shad, Jamshid Darvish, Eskandar Rastegar-Pouyani, and Ahmad Mahmoudi. 2017. 'Subspecies Differentiation of the House Mouse Mus Musculus Linnaeus, 1758 in the Center and East of the Iranian Plateau and Afghanistan'. *Mammalia* 81 (2). https://doi.org/10.1515/mammalia-2015-0041.

Hamill, R M, D Doyle, and E J Duke. 2006. 'Spatial Patterns of Genetic Diversity across European Subspecies of the Mountain Hare, Lepus Timidus L.' *Heredity* 97 (5): 355–65. https://doi.org/10.1038/sj.hdy.6800880.

Hardouin, Emilie A, Annie Orth, Meike Teschke, Jamshid Darvish, Diethard Tautz, and François Bonhomme. 2015. 'Eurasian House Mouse (Mus Musculus L.) Differentiation at Microsatellite Loci Identifies the Iranian Plateau as a

Phylogeographic Hotspot'. *BMC Evolutionary Biology* 15 (February). https://doi.org/10.1186/s12862-015-0306-4.

Harr, Bettina, Emre Karakoc, Rafik Neme, Meike Teschke, Christine Pfeifle, Željka Pezer, Hiba Babiker, et al. 2016. 'Genomic Resources for Wild Populations of the House Mouse, Mus Musculus and Its Close Relative Mus Spretus'. *Scientific Data* 3 (1): 1–14. https://doi.org/10.1038/sdata.2016.75.

Harris, Kelley, and Rasmus Nielsen. 2016. 'The Genetic Cost of Neanderthal Introgression'. *Genetics* 203 (2): 881–91. https://doi.org/10.1534/genetics.116.186890.

Harrison, D. L. 1970. 'SOME OBSERVATIONS ON THE STATUS OF THE SOUTH ARABIAN HOUSE MOUSE MUS MUSCULUS GENTILULUS THOMAS, 1919. (RODENTIA : MURIDAE)' 34 (2): 244–47. https://doi.org/10.1515/mamm.1970.34.2.244.

Hermansen, Jo S., Stein A. Sæther, Tore O. Elgvin, Thomas Borge, Elin Hjelle, and Glenn-Peter Sætre. 2011. 'Hybrid Speciation in Sparrows I: Phenotypic Intermediacy, Genetic Admixture and Barriers to Gene Flow'. *Molecular Ecology* 20 (18): 3812–22. https://doi.org/10.1111/j.1365-294X.2011.05183.x.

Hessenauer, Pauline, Anna Fijarczyk, Hélène Martin, Julien Prunier, Guillaume Charron, Jérôme Chapuis, Louis Bernier, Philippe Tanguay, Richard C. Hamelin, and Christian R. Landry. 2020. 'Hybridization and Introgression Drive Genome Evolution of Dutch Elm Disease Pathogens'. *Nature Ecology & Evolution* 4 (4): 626–38. https://doi.org/10.1038/s41559-020-1133-6.

Huerta-Sánchez, Emilia, Xin Jin, Asan, Zhuoma Bianba, Benjamin M. Peter, Nicolas Vinckenbosch, Yu Liang, et al. 2014. 'Altitude Adaptation in Tibetans Caused by Introgression of Denisovan-like DNA'. *Nature* 512 (7513): 194–97. https://doi.org/10.1038/nature13408.

Hughes, Maria, Neil Reid, Ian Montgomery, and Paulo Prodoehl. 2011. 'Verification of Hybridisation between Introduced European and Native Irish Hares'. *Northern Ireland Environment Agency Research and Development SeriesNorthern Ireland Environment Agency Research and Development Series* 11–11.

Hui, Ling, Jing Lu, Yibing Han, and Stephen H. Pilder. 2006. 'The Mouse T Complex Gene Tsga2, Encoding Polypeptides Located in the Sperm Tail and Anterior Acrosome, Maps to a Locus Associated with Sperm Motility and Sperm-Egg Interaction Abnormalities'. *Biology of Reproduction* 74 (4): 633–43. https://doi.org/10.1095/biolreprod.105.045963.

IUCN. 2016. 'IUCN Red List of Threatened Species'. *Version 2016-1*, www.iucnredlist.org.

Janoušek, Václav, Pavel Munclinger, Liuyang Wang, Katherine C. Teeter, and Priscilla K. Tucker. 2015a. 'Functional Organization of the Genome May Shape the Species Boundary in the House Mouse'. *Molecular Biology and Evolution* 32 (5): 1208–20. https://doi.org/10.1093/molbev/msv011.

———. 2015b. 'Functional Organization of the Genome May Shape the Species Boundary in the House Mouse'. *Molecular Biology and Evolution* 32 (5): 1208–20. https://doi.org/10.1093/molbev/msv011.

Janoušek, Václav, LIUYANG WANG, KEN LUZYNSKI, PETRA DUFKOVÁ, MARTINA M. VYSKOČILOVÁ, MICHAEL W. NACHMAN, PAVEL MUNCLINGER, MILOŠ MACHOLÁN, JAROSLAV PIÁLEK, and PRISCILLA K. TUCKER. 2012. 'Genome-Wide Architecture of Reproductive Isolation in a Naturally Occurring Hybrid Zone between Mus Musculus Musculus and M. m. Domesticus'. *Molecular Ecology* 21 (12): 10.1111/j.1365-294X.2012.05583.x. https://doi.org/10.1111/j.1365-294X.2012.05583.x.

Jones, Eleanor P., JEROEN Van Der Kooij, Roar Solheim, and Jeremy B. Searle. 2010. 'Norwegian House Mice (Mus Musculus Musculus/Domesticus): Distributions,

Routes of Colonization and Patterns of Hybridization'. *Molecular Ecology* 19 (23): 5252–64. https://doi.org/10.1111/j.1365-294X.2010.04874.x.

Jones, Matthew R., L. Scott Mills, Paulo Célio Alves, Colin M. Callahan, Joel M. Alves, Diana J. R. Lafferty, Francis M. Jiggins, Jeffrey D. Jensen, José Melo-Ferreira, and Jeffrey M. Good. 2018. 'Adaptive Introgression Underlies Polymorphic Seasonal Camouflage in Snowshoe Hares'. *Science* 360 (6395): 1355–58. https://doi.org/10.1126/science.aar5273.

Jones, M. R., L. S. Mills, J. D. Jensen, and J. M. Good. 2020. The origin and spread of locally adaptive seasonal camouflage in snowshoe hares. American Naturalist 196:316–332.

Juric, Ivan, Simon Aeschbacher, and Graham Coop. 2016. 'The Strength of Selection against Neanderthal Introgression'. *PLOS Genetics* 12 (11): e1006340. https://doi.org/10.1371/journal.pgen.1006340.

Jurka, Jerzy, Weidong Bao, and Kenji K Kojima. 2011. 'Families of Transposable Elements, Population Structure and the Origin of Species'. *Biology Direct* 6 (1): 44. https://doi.org/10.1186/1745-6150-6-44.

Kanehisa, Minoru, Susumu Goto, Yoko Sato, Miho Furumichi, and Mao Tanabe. 2012. 'KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets'. *Nucleic Acids Research* 40 (D1). https://doi.org/10.1093/nar/gkr988.

Khurana, Jaspreet S., Jie Wang, Jia Xu, Birgit S. Koppetsch, Travis C. Thomson, Anetta Nowosielska, Chengjian Li, Phillip D. Zamore, Zhiping Weng, and William E. Theurkauf. 2011. 'Adaptation to P Element Transposon Invasion in Drosophila Melanogaster'. *Cell* 147 (7): 1551–63. https://doi.org/10.1016/j.cell.2011.11.042.

Kirkpatrick, Mark, and Virginie Ravigné. 2002. 'Speciation by Natural and Sexual Selection: Models and Experiments.' *The American Naturalist* 159 (S3): S22–35. https://doi.org/10.1086/338370.

Klopfstein, Seraina, Mathias Currat, and Laurent Excoffier. 2006. 'The Fate of Mutations Surfing on the Wave of a Range Expansion'. *Molecular Biology and Evolution* 23 (3): 482–90. https://doi.org/10.1093/molbev/msj057.

Krogh, a, B Larsson, G von Heijne, and E L Sonnhammer. 2001. 'Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes.' *Journal of Molecular Biology* 305 (3): 567–80. https://doi.org/10.1006/jmbi.2000.4315.

Kunte, Krushnamegh, Cristina Shea, Matthew L. Aardema, J. Mark Scriber, Thomas E. Juenger, Lawrence E. Gilbert, and Marcus R. Kronforst. 2011. 'Sex Chromosome Mosaicism and Hybrid Speciation among Tiger Swallowtail Butterflies'. *PLOS Genetics* 7 (9): e1002274. https://doi.org/10.1371/journal.pgen.1002274.

Langdon, Quinn K., David Peris, Juan I. Eizaguirre, Dana A. Opulente, Kelly V. Buh, Kayla Sylvester, Martin Jarzyna, et al. 2020. 'Postglacial Migration Shaped the Genomic Diversity and Global Distribution of the Wild Ancestor of Lager-Brewing Hybrids'. *PLOS Genetics* 16 (4): e1008680. https://doi.org/10.1371/journal.pgen.1008680.

Larson, Erica L., Dan Vanderpool, Brice A. J. Sarver, Colin Callahan, Sara Keeble, Lorraine P. Provencio, Michael D. Kessler, et al. 2018. 'The Evolution of Polymorphic Hybrid Incompatibilities in House Mice'. *Genetics*, April, genetics.300840.2018. https://doi.org/10.1534/genetics.118.300840.

Lawson, Daniel J., Lucy van Dorp, and Daniel Falush. 2018. 'A Tutorial on How Not to Over-Interpret STRUCTURE and ADMIXTURE Bar Plots'. *Nature Communications* 9 (1): 3258. https://doi.org/10.1038/s41467-018-05257-7.

Leach, Katie, Ruth Kelly, Alison Cameron, W. Ian Montgomery, and Neil Reid. 2015. 'Expertly Validated Models and Phylogenetically-Controlled Analysis Suggests Responses to Climate Change Are Related to Species Traits in the Order Lagomorpha'. *PLoS ONE* 10 (4). https://doi.org/10.1371/journal.pone.0122267.

Leach, Katie, W. Ian Montgomery, and Neil Reid. 2015. 'Biogeography, Macroecology and Species' Traits Mediate Competitive Interactions in the Order Lagomorpha'. *Mammal Review*. https://doi.org/10.1111/mam.12035.

Levänen, Riikka, Carl-Gustaf Thulin, Göran Spong, and Jaakko L. O. Pohjoismäki. 2018. 'Widespread Introgression of Mountain Hare Genes into Fennoscandian Brown Hare Populations'. *PLOS ONE* 13 (1): e0191790. https://doi.org/10.1371/journal.pone.0191790.

Li, Heng, and Richard Durbin. 2010. 'Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform'. *Bioinformatics* 26 (5): 589–95. https://doi.org/10.1093/bioinformatics/btp698.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Lindholm, Anna K, Kerstin Musolf, Andrea Weidt, and Barbara König. 2013. 'Mate Choice for Genetic Compatibility in the House Mouse'. *Ecology and Evolution* 3 (5): 1231–47. https://doi.org/10.1002/ece3.534.

Liu, Kevin J., Ethan Steinberg, Alexander Yozzo, Ying Song, Michael H. Kohn, and Luay Nakhleh. 2015. 'Interspecific Introgressive Origin of Genomic Diversity in the House Mouse'. *Proceedings of the National Academy of Sciences* 112 (1): 196–201. https://doi.org/10.1073/pnas.1406298111.

Macholán, Miloš, Stuart J. E. Baird, Alena Fornůsková, Iva Martincová, Pavel Rubík, Ľudovít Ďureje, Emanuel Heitlinger, and Jaroslav Piálek. 2019. 'Widespread Introgression of the *Mus Musculus Musculus* Y Chromosome in Central Europe'. Preprint. Evolutionary Biology. https://doi.org/10.1101/2019.12.23.887471.

Macholán, Miloš, Stuart J. E. Baird, Pavel Munclinger, and Jaroslav Piálek. 2012. *Evolution of the House Mouse*. Cambridge University Press.

Macholán, Miloš, Stuart JE Baird, Pavel Munclinger, Petra Dufková, Barbora Bímová, and Jaroslav Piálek. 2008a. 'Genetic Conflict Outweighs Heterogametic Incompatibility in the Mouse Hybrid Zone?' *BMC Evolutionary Biology* 8 (1): 271. https://doi.org/10.1186/1471-2148-8-271.

———. 2008b. 'Genetic Conflict Outweighs Heterogametic Incompatibility in the Mouse Hybrid Zone?' *BMC Evolutionary Biology* 8 (1): 271. https://doi.org/10.1186/1471-2148-8-271.

Macholán, Miloš, Boris Kryštufek, and Vladimír Vohralík. 2003. 'The Location of TheMus Musculus/M. Domesticus Hybrid Zone in the Balkans: Clues from Morphology'. *Acta Theriologica* 48 (2): 177–88. https://doi.org/10.1007/BF03194157.

Maheshwari, Shamoni, and Daniel A. Barbash. 2011. 'The Genetics of Hybrid Incompatibilities'. *Annual Review of Genetics* 45 (1): 331–55. https://doi.org/10.1146/annurev-genet-110410-132514.

Mallet, James, Margarita Beltrán, Walter Neukirchen, and Mauricio Linares. 2007. 'Natural Hybridization in Heliconiine Butterflies: The Species Boundary as a Continuum'. *BMC Evolutionary Biology* 7 (1): 28. https://doi.org/10.1186/1471-2148-7-28.

Mallet, James, Nora Besansky, and Matthew W. Hahn. 2016. 'How Reticulated Are Species?' *BioEssays* 38 (2): 140–49. https://doi.org/10.1002/bies.201500149.

Mallet, James, and Mathieu Joron. 1999. 'Evolution of Diversity in Warning Color and Mimicry: Polymorphisms, Shifting Balance, and Speciation'. *Annual Review of Ecology and Systematics* 30: 201–33.

Marques, Joao P, Liliana Farelo, Joana Vilela, Dan Vanderpool, Paulo C Alves, Jeffrey M Good, Pierre Boursot, and Jose Melo-Ferreira. 2017. 'Range Expansion Underlies Historical Introgressive Hybridization in the Iberian Hare.' *Scientific Reports* 7 (January): 40788. https://doi.org/10.1038/srep40788.

Marques, João P., Liliana Farelo, Joana Vilela, Dan Vanderpool, Paulo C. Alves, Jeffrey M. Good, Pierre Boursot, and José Melo-Ferreira. 2017. 'Range Expansion Underlies Historical Introgressive Hybridization in the Iberian Hare'. *Scientific Reports* 7 (1): 40788. https://doi.org/10.1038/srep40788.

Marques, João P., Mafalda S. Ferreira, Liliana Farelo, Colin M. Callahan, Klaus Hackländer, Hannes Jenny, W. Ian Montgomery, et al. 2017. 'Mountain Hare Transcriptome and Diagnostic Markers as Resources to Monitor Hybridization with European Hares'. *Scientific Data* 4 (1): 170178. https://doi.org/10.1038/sdata.2017.178.

Marques, Joao P, Mafalda S Ferreira, Liliana Farelo, Colin M Callahan, Klaus Hacklander, Hannes Jenny, W Ian Montgomery, et al. 2017. 'Mountain Hare Transcriptome and Diagnostic Markers as Resources to Monitor Hybridization with European Hares.' *Scientific Data* 4 (December): 170178–170178. https://doi.org/10.1038/sdata.2017.178.

Marques, João P., Fernando A. Seixas, Liliana Farelo, Colin M. Callahan, Jeffrey M. Good, W. Ian Montgomery, Neil Reid, Paulo C. Alves, Pierre Boursot, and José Melo-Ferreira. 2020. 'An Annotated Draft Genome of the Mountain Hare (Lepus Timidus)'. *Genome Biology and Evolution* 12 (1): 3656–62. https://doi.org/10.1093/gbe/evz273.

Martienssen, Robert A. 2010. 'Heterochromatin, Small RNA and Post-Fertilization Dysgenesis in Allopolyploid and Interploid Hybrids of Arabidopsis'. *The New Phytologist* 186 (1): 46–53. https://doi.org/10.1111/j.1469-8137.2010.03193.x.

Martin, Simon H., Kanchon K. Dasmahapatra, Nicola J. Nadeau, Camilo Salazar, James R. Walters, Fraser Simpson, Mark Blaxter, Andrea Manica, James Mallet, and Chris D. Jiggins. 2013. 'Genome-Wide Evidence for Speciation with Gene Flow in Heliconius Butterflies'. *Genome Research* 23 (11): 1817–28. https://doi.org/10.1101/gr.159426.113.

Martin, Simon H., John W. Davey, Camilo Salazar, and Chris D. Jiggins. 2019. 'Recombination Rate Variation Shapes Barriers to Introgression across Butterfly Genomes'. *PLOS Biology* 17 (2): e2006288. https://doi.org/10.1371/journal.pbio.2006288.

Martincová, Iva, Ľudovít Ďureje, Jakub Kreisinger, Miloš Macholán, and Jaroslav Piálek. 2019. 'Phenotypic Effects of the Y Chromosome Are Variable and Structured in Hybrids among House Mouse Recombinant Lines'. *Ecology and Evolution*, May, ece3.5196. https://doi.org/10.1002/ece3.5196.

Masly, John P., and Daven C. Presgraves. 2007. 'High-Resolution Genome-Wide Dissection of the Two Rules of Speciation in Drosophila'. *PLOS Biology* 5 (9): e243. https://doi.org/10.1371/journal.pbio.0050243.

Matthee, Conrad A, Bettine Jansen Van Vuuren, Diana Bell, and Terence J Robinson. 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): 433–47. https://doi.org/10.1080/10635150490445715.

Maxwell, Colin S., Victoria E. Sepulveda, David A. Turissini, William E. Goldman, and Daniel R. Matute. 2018. 'Recent Admixture between Species of the Fungal Pathogen Histoplasma'. *Evolution Letters* 2 (3): 210–20. https://doi.org/10.1002/evl3.59.

Mayr, Ernst. 1942. 'Systematics And The Origin Of Species From The Viewpoint Of A Zoologist'. 1942.

———. 1963. *Animal Species and Evolution*. *Animal Species and Evolution*. Harvard University Press. https://www.degruyter.com/document/doi/10.4159/harvard.9780674865327/html.

McDermott, Shannon R., and Mohamed A. F. Noor. 2010. 'The Role of Meiotic Drive in Hybrid Male Sterility'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1544): 1265–72. https://doi.org/10.1098/rstb.2009.0264.

McKenna, A. H., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, et al. 2010. 'The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data'. *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

McLaughlin, Richard N., and Harmit S. Malik. 2017. 'Genetic Conflicts: The Usual Suspects and Beyond'. Edited by Joel D. Levine, Daniel J. C. Kronauer, and Michael H. Dickinson. *Journal of Experimental Biology* 220 (1): 6–17. https://doi.org/10.1242/jeb.148148.

Meier, Joana I., David A. Marques, Salome Mwaiko, Catherine E. Wagner, Laurent Excoffier, and Ole Seehausen. 2017. 'Ancient Hybridization Fuels Rapid Cichlid Fish Adaptive Radiations'. *Nature Communications* 8 (1): 14363. https://doi.org/10.1038/ncomms14363.

Melo, Maria C., Alicia Grealy, Beth Brittain, Greg M. Walter, and Daniel Ortiz-Barrientos. 2014. 'Strong Extrinsic Reproductive Isolation between Parapatric Populations of an Australian Groundsel'. *New Phytologist* 203 (1): 323–34. https://doi.org/10.1111/nph.12779.

Melo-Ferreira, J., P. C. Alves, H. Freitas, N. Ferrand, and P. Boursot. 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): 2643–58. https://doi.org/10.1111/j.1365-294X.2009.04221.x.

Melo-Ferreira, J., P. Boursot, M. Carneiro, P. J. Esteves, L. Farelo, and P. C. Alves. 2012a. 'Recurrent Introgression of Mitochondrial DNA among Hares (Lepus Spp.) Revealed by Species-Tree Inference and Coalescent Simulations'. *Systematic Biology* 61 (3): 367–81. https://doi.org/10.1093/sysbio/syr114.

———. 2012b. 'Recurrent Introgression of Mitochondrial DNA Among Hares (Lepus Spp.) Revealed by Species-Tree Inference and Coalescent Simulations'. *Systematic Biology* 61 (3): 367–367. https://doi.org/10.1093/sysbio/syr114.

Melo-Ferreira, J., P. Boursot, E. Randi, A. Kryukov, F. Suchentrunk, N. Ferrand, and P. C. Alves. 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): 605–18. https://doi.org/10.1111/j.1365-294X.2006.03166.x.

Melo-Ferreira, J., P. Boursot, F. Suchentrunk, N. Ferrand, and P. C. Alves. 2005. 'Invasion from the Cold Past: Extensive Introgression of Mountain Hare (Lepus Timidus) Mitochondrial DNA into Three Other Hare Species in Northern Iberia'. *Molecular Ecology* 14 (8): 2459–64. https://doi.org/10.1111/j.1365-294X.2005.02599.x.

Melo-Ferreira, J, L Farelo, H Freitas, F Suchentrunk, P Boursot, and P C Alves. 2014a. 'Home-Loving Boreal Hare Mitochondria Survived Several Invasions in Iberia: The Relative Roles of Recurrent Hybridisation and Allele Surfing.' *Heredity* 112 (3): 265–73. https://doi.org/10.1038/hdy.2013.102.

———. 2014b. 'Home-Loving Boreal Hare Mitochondria Survived Several Invasions in Iberia: The Relative Roles of Recurrent Hybridisation and Allele Surfing'. *Heredity* 112 (3): 265–73. https://doi.org/10.1038/hdy.2013.102.

Metcalfe, Cushla J., Kira V. Bulazel, Gianni C. Ferreri, Elizabeth Schroeder-Reiter, Gerhard Wanner, Willem Rens, Craig Obergfell, Mark D. B. Eldridge, and Rachel J. O'Neill. 2007. 'Genomic Instability Within Centromeres of Interspecific Marsupial Hybrids'. *Genetics* 177 (4): 2507–17. https://doi.org/10.1534/genetics.107.082313.

Mihola, Ondrej, Vladimir Landa, Florencia Pratto, Kevin Brick, Tatyana Kobets, Fitore Kusari, Srdjan Gasic, et al. 2021. 'Rat PRDM9 Shapes Recombination Landscapes, Duration of Meiosis, Gametogenesis, and Age of Fertility'. *BMC Biology* 19 (1): 86. https://doi.org/10.1186/s12915-021-01017-0.

Mihola, Ondrej, Zdenek Trachtulec, Cestmir Vlcek, John C. Schimenti, and Jiri Forejt. 2009. 'A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase'. *Science* 323 (5912): 373–75. https://doi.org/10.1126/science.1163601.

Moran, Benjamin M, Cheyenne Payne, Quinn Langdon, Daniel L Powell, Yaniv Brandvain, and Molly Schumer. 2021. 'The Genomic Consequences of Hybridization'. Edited by Patricia J Wittkopp. *ELife* 10 (August): e69016. https://doi.org/10.7554/eLife.69016.

Muller, H. 1942. 'Isolating Mechanisms, Evolution, and Temperature'. *Biol. Symp.* 6: 71–125.

Nachman, Michael W., and Bret A. Payseur. 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): 409–21. https://doi.org/10.1098/rstb.2011.0249.

Nielsen, Rasmus, Joshua M. Akey, Mattias Jakobsson, Jonathan K. Pritchard, Sarah Tishkoff, and Eske Willerslev. 2017. 'Tracing the Peopling of the World through Genomics'. *Nature* 541 (7637): 302–10. https://doi.org/10.1038/nature21347.

Norris, Laura C., Bradley J. Main, Yoosook Lee, Travis C. Collier, Abdrahamane Fofana, Anthony J. Cornel, and Gregory C. Lanzaro. 2015. 'Adaptive Introgression in an African Malaria Mosquito Coincident with the Increased Usage of Insecticide-Treated Bed Nets'. *Proceedings of the National Academy of Sciences* 112 (3): 815–20. https://doi.org/10.1073/pnas.1418892112.

Nosil, Patrik, and Jeffrey L. Feder. 2012. 'Genomic Divergence during Speciation: Causes and Consequences'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1587): 332–42. https://doi.org/10.1098/rstb.2011.0263.

Nosil, Patrik, Daniel J. Funk, and Daniel Ortiz-Barrientos. 2009. 'Divergent Selection and Heterogeneous Genomic Divergence'. *Molecular Ecology* 18 (3): 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x.

Nosil, Patrik, Luke J. Harmon, and Ole Seehausen. 2009. 'Ecological Explanations for (Incomplete) Speciation'. *Trends in Ecology & Evolution* 24 (3): 145–56. https://doi.org/10.1016/j.tree.2008.10.011.

Nosil, Patrik, and Dolph Schluter. 2011. 'The Genes Underlying the Process of Speciation'. *Trends in Ecology & Evolution* 26 (4): 160–67. https://doi.org/10.1016/j.tree.2011.01.001.

Orr, H. A. 1995. 'The Population Genetics of Speciation: The Evolution of Hybrid Incompatibilities'. *Genetics* 139 (4): 1805–13.

Patel, Maulik R, Ganesh K Miriyala, Aimee J Littleton, Heiko Yang, Kien Trinh, Janet M Young, Scott R Kennedy, Yukiko M Yamashita, Leo J Pallanck, and Harmit S Malik. 2016. 'A Mitochondrial DNA Hypomorph of Cytochrome Oxidase Specifically Impairs Male Fertility in Drosophila Melanogaster'. *ELife* 5: e16923. https://doi.org/10.7554/eLife.16923.

Payseur, Bret A., and Loren H. Rieseberg. 2016. 'A Genomic Perspective on Hybridization and Speciation'. *Molecular Ecology* 25 (11): 2337–60. https://doi.org/10.1111/mec.13557.

Petersen, Thomas Nordahl, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2011. 'SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions'. *Nature Methods* 8 (10): 785–86. https://doi.org/10.1038/nmeth.1701.

Phifer-Rixey, Megan, Matthew Bomhoff, and Michael W. Nachman. 2014. 'Genome-Wide Patterns of Differentiation Among House Mouse Subspecies'. *Genetics* 198 (1): 283–97. https://doi.org/10.1534/genetics.114.166827.

Phifer-Rixey, Megan, Bettina Harr, and Jody Hey. 2020. 'Further Resolution of the House Mouse (Mus Musculus) Phylogeny by Integration over Isolation-with-Migration Histories'. *BMC Evolutionary Biology* 20 (1): 120. https://doi.org/10.1186/s12862-020-01666-9.

Phifer-Rixey, Megan, and Michael W Nachman. 2015. 'Insights into Mammalian Biology from the Wild House Mouse Mus Musculus'. *ELife* 4 (April): e05959. https://doi.org/10.7554/eLife.05959.

Pool, John E. 2015. 'The Mosaic Ancestry of the Drosophila Genetic Reference Panel and the D. Melanogaster Reference Genome Reveals a Network of Epistatic Fitness Interactions'. *Molecular Biology and Evolution* 32 (12): 3236–51. https://doi.org/10.1093/molbev/msv194.

Powell, Sean, Damian Szklarczyk, Kalliopi Trachana, Alexander Roth, Michael Kuhn, Jean Muller, Roland Arnold, et al. 2012. 'EggNOG v3.0: Orthologous Groups Covering 1133 Organisms at 41 Different Taxonomic Ranges'. *Nucleic Acids Research* 40 (D1). https://doi.org/10.1093/nar/gkr1060.

Prager, E M, C Orrego, and R D Sage. 1998. 'Genetic Variation and Phylogeography of Central Asian and Other House Mice, Including a Major New Mitochondrial Lineage in Yemen.' *Genetics* 150 (2): 835–61.

Presgraves, Daven. 2009. '12 - Drive and Sperm: The Evolution and Genetics of Male Meiotic Drive'. In *Sperm Biology*, edited by Tim R. Birkhead, David J. Hosken, and Scott Pitnick, 471–506. London: Academic Press. https://doi.org/10.1016/B978-0-12-372568-4.00012-4.

Presgraves, Daven C. 2010. 'The Molecular Evolutionary Basis of Species Formation'. *Nature Reviews Genetics* 11 (3): 175–80. https://doi.org/10.1038/nrg2718.

Price, Trevor D., and Michelle M. Bouvier. 2002. 'The Evolution of F1 Postzygotic Incompatibilities in Birds'. *Evolution* 56 (10): 2083–89. https://doi.org/10.1111/j.0014-3820.2002.tb00133.x.

Queiroz, Kevin de. 1998. 'The General Lineage Concept of Species, Species Chteria, and the Process of Speciation', 19.

Quilodrán, Claudio S., Alexandros Tsoupas, and Mathias Currat. 2020. 'The Spatial Signature of Introgression After a Biological Invasion With Hybridization'. *Frontiers in Ecology and Evolution* 8 (September): 569620. https://doi.org/10.3389/fevo.2020.569620.

Raufaste, Nathalie, Annie Orth, Khalid Belkhir, David Senet, Carole Smadja, Stuart Baird, François Bonhomme, Barbara Dod, and Pierre Boursot. 2005. 'Inferences of Selection and Migration in the Danish House Mouse Hybrid Zone'. *Biological Journal of the Linnean Society* 84 (3): 593–616. https://doi.org/10.1111/j.1095-8312.2005.00457.x.

Rebollo, Rita, Béatrice Horard, Benjamin Hubert, and Cristina Vieira. 2010. 'Jumping Genes and Epigenetics: Towards New Species'. *Gene* 454 (1): 1–7. https://doi.org/10.1016/j.gene.2010.01.003.

Reid, Neil. 2011. 'European Hare (Lepus Europaeus) Invasion Ecology: Implication for the Conservation of the Endemic Irish Hare (Lepus Timidus Hibernicus)'. *Biological Invasions* 13 (3): 559–69. https://doi.org/10.1007/s10530-010-9849-x.

Reid, Neil, and W. Ian Montgomery. 2007. 'Is Naturalisation of the Brown Hare in Ireland a Threat to the Endemic Irish Hare?' *Biology and Environment* 107 (3): 129–38. https://doi.org/10.3318/BIOE.2007.107.3.129.

Rice, William R., Sergey Gavrilets, and Urban Friberg. 2008. 'Sexually Antagonistic "Zygotic Drive" of the Sex Chromosomes'. *PLOS Genetics* 4 (12): e1000313. https://doi.org/10.1371/journal.pgen.1000313.

Rieseberg, L. H. 2003. 'Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization'. *Science* 301 (5637): 1211–16. https://doi.org/10.1126/science.1086949.

Rieseberg, Loren H., Margaret A. Archer, and Robert K. Wayne. 1999. 'Transgressive Segregation, Adaptation and Speciation'. *Heredity* 83 (4): 363–72. https://doi.org/10.1038/sj.hdy.6886170.

Rieseberg, Loren H., and Benjamin K. Blackman. 2010. 'Speciation Genes in Plants'. *Annals of Botany* 106 (3): 439–55. https://doi.org/10.1093/aob/mcq126.

Rieseberg, Loren H., and John H. Willis. 2007. 'Plant Speciation'. *Science (New York, N.Y.)* 317 (5840): 910–14. https://doi.org/10.1126/science.1137729.

Rosenberg, N A, T Burke, K Elo, M W Feldman, P J Freidlin, M A Groenen, J Hillel, et al. 2001. 'Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes from 20 Chicken Breeds.' *Genetics* 159 (2): 699–713.

Rosenberg, Noah A, Jonathan K Pritchard, James L Weber, Howard M Cann, Kenneth K Kidd, Lev A Zhivotovsky, and Marcus W Feldman. 2002. 'Genetic Structure of Human Populations' 298: 6.

Roux, Camille, Christelle Fraïsse, Jonathan Romiguier, Yoann Anciaux, Nicolas Galtier, and Nicolas Bierne. 2016. 'Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence'. Edited by Craig Moritz. *PLOS Biology* 14 (12): e2000234. https://doi.org/10.1371/journal.pbio.2000234.

Rundle, Howard D., and Patrik Nosil. 2005. 'Ecological Speciation'. *Ecology Letters* 8 (3): 336–52. https://doi.org/10.1111/j.1461-0248.2004.00715.x.

Sackton, Timothy B., Horacio Montenegro, Daniel L. Hartl, and Bernardo Lemos. 2011. 'Interspecific Y Chromosome Introgressions Disrupt Testis-Specific Gene Expression and Male Reproductive Phenotypes in Drosophila'. *Proceedings of the National Academy of Sciences* 108 (41): 17046–51. https://doi.org/10.1073/pnas.1114690108.

Sankararaman, Sriram, Swapan Mallick, Michael Dannemann, Kay Prüfer, Janet Kelso, Svante Pääbo, Nick Patterson, and David Reich. 2014. 'The Genomic Landscape of Neanderthal Ancestry in Present-Day Humans'. *Nature* 507 (7492): 354–57. https://doi.org/10.1038/nature12961.

Sanz-Martín, María J., Andone Estonba, Carmen Manzano, Mikel Iriondo, Gonzalo Pérez-Suárez, Alfonso Fernández, and Fernando Palacios. 2014. 'Genetic Structure of Brown and Iberian Hare Populations in Northern Iberia: Implications for Conservation of Genetic Diversity'. *The Journal of Wildlife Management* 78 (4): 632–44. https://doi.org/10.1002/jwmg.713.

Schluter, Dolph, and Associate Editor: Mark A. McPeek. 2000. 'Ecological Character Displacement in Adaptive Radiation'. *The American Naturalist* 156 (S4): S4–16. https://doi.org/10.1086/303412.

Schumer, Molly, Rongfeng Cui, Gil G. Rosenthal, and Peter Andolfatto. 2015. 'Reproductive Isolation of Hybrid Populations Driven by Genetic Incompatibilities'. Edited by Bret A. Payseur. *PLOS Genetics* 11 (3): e1005041. https://doi.org/10.1371/journal.pgen.1005041.

Schumer, Molly, Chenling Xu, Daniel L. Powell, Arun Durvasula, Laurits Skov, Chris Holland, John C. Blazier, et al. 2018. 'Natural Selection Interacts with Recombination to Shape the Evolution of Hybrid Genomes'. *Science* 360 (6389): 656–60. https://doi.org/10.1126/science.aar3684.

Schwarz, Ernst, and Henriette K. Schwarz. 1943. 'The Wild and Commensal Stocks of the House Mouse, Mus Musculus Linnaeus'. *Journal of Mammalogy* 24 (1): 59–72. https://doi.org/10.2307/1374781.

Scordato, Elizabeth S. C., Chris C. R. Smith, Georgy A. Semenov, Yu Liu, Matthew R. Wilkins, Wei Liang, Alexander Rubtsov, et al. 2020. 'Migratory Divides Coincide with Reproductive Barriers across Replicated Avian Hybrid Zones above the

Tibetan Plateau'. *Ecology Letters* 23 (2): 231–41. https://doi.org/10.1111/ele.13420.

Ségurel, Laure, Ellen Miranda Leffler, and Molly Przeworski. 2011. 'The Case of the Fickle Fingers: How the PRDM9 Zinc Finger Protein Specifies Meiotic Recombination Hotspots in Humans'. *PLOS Biology* 9 (12): e1001211. https://doi.org/10.1371/journal.pbio.1001211.

Seixas, Fernando A., Pierre Boursot, and José Melo-Ferreira. 2018. 'The Genomic Impact of Historical Hybridization with Massive Mitochondrial DNA Introgression'. *Genome Biology* 19 (1): 91. https://doi.org/10.1186/s13059-018-1471-8.

Selz, O. M., and O. Seehausen. 2019. 'Interspecific Hybridization Can Generate Functional Novelty in Cichlid Fish'. *Proceedings of the Royal Society B: Biological Sciences* 286 (1913): 20191621. https://doi.org/10.1098/rspb.2019.1621.

Shan, Xiaohui, Zhenlan Liu, Zhenying Dong, Yongming Wang, Yu Chen, Xiuyun Lin, Likun Long, Fangpu Han, Yingshan Dong, and Bao Liu. 2005. 'Mobilization of the Active MITE Transposons MPing and Pong in Rice by Introgression from Wild Rice (Zizania Latifolia Griseb.)'. *Molecular Biology and Evolution* 22 (4): 976–90. https://doi.org/10.1093/molbev/msi082.

Shi, Wei, Antje Krella, Annie Orth, Yang Yu, and Reinald Fundele. 2005. 'Widespread Disruption of Genomic Imprinting in Adult Interspecies Mouse (Mus) Hybrids'. *Genesis (New York, N.Y.: 2000)* 43 (3): 100–108. https://doi.org/10.1002/gene.20161.

Simon, Alexis, Nicolas Bierne, and John J. Welch. 2018. 'Coadapted Genomes and Selection on Hybrids: Fisher's Geometric Model Explains a Variety of Empirical Patterns'. *Evolution Letters* 2 (5): 472–98. https://doi.org/10.1002/evl3.66.

Sjodin, Bryson M F, Kurt E Galbreath, Hayley C Lanier, and Michael A Russello. 2021. 'Chromosome-Level Reference Genome Assembly for the American Pika (Ochotona Princeps)'. *Journal of Heredity* 112 (6): 549–57. https://doi.org/10.1093/jhered/esab031.

Slatkin, M., and R. Lande. 1994. 'Segregation Variance after Hybridization of Isolated Populations'. *Genetical Research* 64 (1): 51–56. https://doi.org/10.1017/s0016672300032547.

Smith, Andrew T., ed. 2018. *Lagomorphs: Pikas, Rabbits, and Hares of the World*. Baltimore: Johns Hopkins University Press.

Smith-Unna, Richard, Chris Boursnell, Rob Patro, Julian Hibberd, and Steven Kelly. 2016a. 'TransRate: Reference Free Quality Assessment of de Novo Transcriptome Assemblies'. *Genome Research*, June, gr.196469.115. https://doi.org/10.1101/gr.196469.115.

Smith-Unna, Richard, Chris Boursnell, Rob Patro, Julian M. Hibberd, and Steven Kelly. 2016b. 'TransRate: Reference-Free Quality Assessment of de Novo Transcriptome Assemblies'. *Genome Research* 26 (8): 1134–44. https://doi.org/10.1101/gr.196469.115.

Song, Ying, Stefan Endepols, Nicole Klemann, Dania Richter, Franz-Rainer Matuschka, Ching-Hua Shih, Michael W. Nachman, and Michael H. Kohn. 2011. 'Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice'. *Current Biology: CB* 21 (15): 1296–1301. https://doi.org/10.1016/j.cub.2011.06.043.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, et al. 2014. 'Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation'. *Science* 344 (6185): 738–42. https://doi.org/10.1126/science.1252136.

Srinivasa, Ananya Nidamangala, and Sarah E. Zanders. 2020. 'Meiotic Drive'. *Current Biology* 30 (11): R627–29. https://doi.org/10.1016/j.cub.2020.04.023.

Steensma, David P., Robert A. Kyle, and Marc A. Shampo. 2010. 'Abbie Lathrop, the "Mouse Woman of Granby": Rodent Fancier and Accidental Genetics Pioneer'. *Mayo Clinic Proceedings* 85 (11): e83. https://doi.org/10.4065/mcp.2010.0647.

Stelkens, Rike, and Ole Seehausen. 2009. 'Genetic Distance Between Species Predicts Novel Trait Expression in Their Hybrids'. *Evolution* 63 (4): 884–97. https://doi.org/10.1111/j.1558-5646.2008.00599.x.

Suchentrunk, F, Z Mamuris, C Stamatis, H Ben Slimen, K Hackländer, G Haerer, and M Giacometti. 2005. 'Introgressive Hybridization in Wild Living Mountain Hares (L. Timidus Varronis) and Brown Hares (L. Europaeus) and Morphological Consequences'. *Mamm. Biol* 70 (Suppl.): 39–40.

Suzuki, H, M Nunome, G Kinoshita, K P Aplin, P Vogel, A P Kryukov, M-L Jin, et al. 2013. 'Evolutionary and Dispersal History of Eurasian House Mice Mus Musculus Clarified by More Extensive Geographic Sampling of Mitochondrial DNA'. *Heredity* 111 (5): 375–90. https://doi.org/10.1038/hdy.2013.60.

Suzuki, Hitoshi, Lyudmila V. Yakimenko, Daiki Usuda, and Liubov V. Frisman. 2015. 'Tracing the Eastward Dispersal of the House Mouse, Mus Musculus'. *Genes and Environment* 37 (August). https://doi.org/10.1186/s41021-015-0013-9.

Svardal, Hannes, Fu Xiang Quah, Milan Malinsky, Benjamin P Ngatunga, Eric A Miska, Walter Salzburger, Martin J Genner, George F Turner, and Richard Durbin. 2020. 'Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation'. *Molecular Biology and Evolution* 37 (4): 1100–1113. https://doi.org/10.1093/molbev/msz294.

Svedin, Nina, Chris Wiley, Thor Veen, Lars Gustafsson, and Anna Qvarnström. 2008. 'Natural and Sexual Selection against Hybrid Flycatchers'. *Proceedings of the Royal Society B: Biological Sciences* 275 (1635): 735–44. https://doi.org/10.1098/rspb.2007.0967.

Sweigart, Andrea L. 2010. 'The Genetics of Postmating, Prezygotic Reproductive Isolation Between Drosophila Virilis and D. Americana'. *Genetics* 184 (2): 401–10. https://doi.org/10.1534/genetics.109.111245.

Sweigart, Andrea L., and Lex E. Flagel. 2015. 'Evidence of Natural Selection Acting on a Polymorphic Hybrid Incompatibility Locus in Mimulus'. *Genetics* 199 (2): 543–54. https://doi.org/10.1534/genetics.114.171819.

Takada, Toyoyuki, Toshinobu Ebata, Hideki Noguchi, Thomas M. Keane, David J. Adams, Takanori Narita, Tadasu Shin-I, et al. 2013. 'The Ancestor of Extant Japanese Fancy Mice Contributed to the Mosaic Genomes of Classical Inbred Strains'. *Genome Research* 23 (8): 1329–38. https://doi.org/10.1101/gr.156497.113.

Tang, Shanwu, and Daven C. Presgraves. 2009. 'Evolution of the Drosophila Nuclear Pore Complex Results in Multiple Hybrid Incompatibilities'. *Science (New York, N.Y.)* 323 (5915): 779–82. https://doi.org/10.1126/science.1169123.

Taylor, Scott A., and Erica L. Larson. 2019. 'Insights from Genomes into the Evolutionary Importance and Prevalence of Hybridization in Nature'. *Nature Ecology & Evolution* 3 (2): 170. https://doi.org/10.1038/s41559-018-0777-y.

Teeter, Katherine C., Bret A. Payseur, Leslie W. Harris, Margaret A. Bakewell, Lisa M. Thibodeau, Janelle E. O'Brien, James G. Krenz, Maria A. Sans-Fuentes, Michael W. Nachman, and Priscilla K. Tucker. 2008. 'Genome-Wide Patterns of Gene Flow across a House Mouse Hybrid Zone'. *Genome Research* 18 (1): 67–76. https://doi.org/10.1101/gr.6757907.

Thulin, C. G., M. Jaarola, and H. Tegelström. 1997. 'The Occurrence of Mountain Hare Mitochondrial DNA in Wild Brown Hares'. *Molecular Ecology* 6 (5): 463–67.

Thulin, C G, M Jaarola, and H Tegelstrom. 1997. 'The Occurrence of Mountain Hare Mitochondrial DNA in Wild Brown Hares.' *Molecular Ecology* 6 (5): 463–67.

Thulin, Carl Gustaf. 2003. 'The Distribution of Mountain Hares Lepus Timidus in Europe: A Challenge from Brown Hares L. Europaeus?' *Mammal Review* 33 (1): 29–42. https://doi.org/10.1046/j.1365-2907.2003.00008.x.

Thulin, Carl Gustaf, Meiying Fang, and Alexander O. Averianov. 2006. 'Introgression from Lepus Europaeus to L. Timidus in Russia Revealed by Mitochondrial Single Nucleotide Polymorphisms and Nuclear Microsatellites'. *Hereditas* 143 (2006): 68–76. https://doi.org/10.1111/j.2006.0018-0661.01952.x.

Thybert, David, Maša Roller, Fábio C. P. Navarro, Ian Fiddes, Ian Streeter, Christine Feig, David Martin-Galvez, et al. 2018. 'Repeat Associated Mechanisms of Genome Evolution and Function Revealed by the Mus Caroli and Mus Pahari Genomes'. *Genome Research*, March. https://doi.org/10.1101/gr.234096.117.

Tishkoff, Sarah A., Floyd A. Reed, Françoise R. Friedlaender, Christopher Ehret, Alessia Ranciaro, Alain Froment, Jibril B. Hirbo, et al. 2009. 'The Genetic Structure and History of Africans and African Americans'. *Science (New York, N.Y.)* 324 (5930): 1035–44. https://doi.org/10.1126/science.1172257.

Turner, Leslie M, and Bettina Harr. 2014. 'Genome-Wide Mapping in a House Mouse Hybrid Zone Reveals Hybrid Sterility Loci and Dobzhansky-Muller Interactions'. Edited by Gil McVean. *ELife* 3 (December): e02504. https://doi.org/10.7554/eLife.02504.

Turner, Thomas L., and Matthew W. Hahn. 2010. 'Genomic Islands of Speciation or Genomic Islands and Speciation?' *Molecular Ecology* 19 (5): 848–50. https://doi.org/10.1111/j.1365-294X.2010.04532.x.

Unckless, R. L., and H. A. Orr. 2009. 'Dobzhansky–Muller Incompatibilities and Adaptation to a Shared Environment'. *Heredity* 102 (3): 214–17. https://doi.org/10.1038/hdy.2008.129.

Ungerer, Mark C., Suzanne C. Strakosh, and Ying Zhen. 2006. 'Genome Expansion in Three Hybrid Sunflower Species Is Associated with Retrotransposon Proliferation'. *Current Biology* 16 (20): R872–73. https://doi.org/10.1016/j.cub.2006.09.020.

Valen, Leigh M. van. 1973. 'A New Evolutionary Law'. In .

Valk, Tom van der, Patrícia Pečnerová, David Díez-del-Molino, Anders Bergström, Jonas Oppenheimer, Stefanie Hartmann, Georgios Xenikoudakis, et al. 2021. 'Million-Year-Old DNA Sheds Light on the Genomic History of Mammoths'. *Nature* 591 (7849): 265–69. https://doi.org/10.1038/s41586-021-03224-9.

Vanlerberghe, F., B. Dod, P. Boursot, M. Bellis, and F. Bonhomme. 1986. 'Absence of Y-Chromosome Introgression across the Hybrid Zone between Mus Musculus Domesticus and Mus Musculus Musculus'. *Genetics Research* 48 (3): 191–97. https://doi.org/10.1017/S0016672300025003.

Veller, Carl, Nathaniel B. Edelman, Pavitra Muralidhar, and Martin A. Nowak. 2019. 'Recombination, Variance in Genetic Relatedness, and Selection against Introgressed DNA'. https://doi.org/10.1101/846147.

Vrana, Paul B. 2007. 'Genomic Imprinting as a Mechanism of Reproductive Isolation in Mammals'. *Journal of Mammalogy* 88 (1): 5–23. https://doi.org/10.1644/06-MAMM-S-013R1.1.

Vyskočilová, MARTINA, ZDENĚK TRACHTULEC, JIŘÍ FOREJT, and JAROSLAV PIÁLEK. 2005. 'Does Geography Matter in Hybrid Sterility in House Mice?' *Biological Journal of the Linnean Society* 84 (3): 663–74. https://doi.org/10.1111/j.1095-8312.2005.00463.x.

Wang, Wenjuan, Yafang Wang, Fumin Lei, Yang Liu, Haitao Wang, and Jiakuan Chen. 2019. 'Incomplete Lineage Sorting and Introgression in the Diversification of Chinese Spot-Billed Ducks and Mallards'. *Current Zoology* 65 (5): 589–97. https://doi.org/10.1093/cz/zoy074.

Weissbrod, Lior, Fiona B. Marshall, François R. Valla, Hamoudi Khalaily, Guy Bar-Oz, Jean-Christophe Auffray, Jean-Denis Vigne, and Thomas Cucchi. 2017. 'Origins of House Mice in Ecological Niches Created by Settled Hunter-Gatherers in the Levant 15,000 y Ago'. *Proceedings of the National Academy of Sciences* 114 (16): 4099–4104. https://doi.org/10.1073/pnas.1619137114.

White, Michael A., Brian Steffy, Tim Wiltshire, and Bret A. Payseur. 2011. 'Genetic Dissection of a Key Reproductive Barrier Between Nascent Species of House Mice'. *Genetics* 189 (1): 289–304. https://doi.org/10.1534/genetics.111.129171.

White, Michael A., Maria Stubbings, Beth L. Dumont, and Bret A. Payseur. 2012. 'Genetics and Evolution of Hybrid Male Sterility in House Mice'. *Genetics* 191 (3): 917–34. https://doi.org/10.1534/genetics.112.140251.

Whitney, Kenneth D., Rebecca A. Randell, and Loren H. Rieseberg. 2010. 'Adaptive Introgression of Abiotic Tolerance Traits in the Sunflower Helianthus Annuus'. *The New Phytologist* 187 (1): 230–39. https://doi.org/10.1111/j.1469-8137.2010.03234.x.

Wright, Kevin M., Deborah Lloyd, David B. Lowry, Mark R. Macnair, and John H. Willis. 2013. 'Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in Mimulus Guttatus'. *PLOS Biology* 11 (2): e1001497. https://doi.org/10.1371/journal.pbio.1001497.

Wu, Chung-I. 2001. 'The Genic View of the Process of Speciation'. *Journal of Evolutionary Biology* 14 (6): 851–65. https://doi.org/10.1046/j.1420-9101.2001.00335.x.

Yamada, Fumio, Mika Takaki, and Hitoshi Suzuki. 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): 107–16. https://doi.org/10.1266/ggs.77.107.

Yonekawa, H, K Moriwaki, O Gotoh, N Miyashita, Y Matsushima, L M Shi, W S Cho, X L Zhen, and Y Tagashira. 1988. 'Hybrid Origin of Japanese Mice "Mus Musculus Molossinus": Evidence from Restriction Analysis of Mitochondrial DNA.' *Molecular Biology and Evolution* 5 (1): 63–78. https://doi.org/10.1093/oxfordjournals.molbev.a040476.

Yonekawa, Hiromichi, Kazuo Moriwaki, Osamu Gotoh, Jun-Ichi Hayashi, Junko Watanabe, Nobumoto Miyashita, Michael L. Petras, and Yusaku Tagashira. 1981. 'Evolutionary Relationships among Five Subspecies of MUS MUSCULUS Based on Restriction Enzyme Cleavage Patterns of Mitochondrial DNA'. *Genetics* 98 (4): 801–16.

Zachos, F. E., H. Ben Slimen, K. Hackländer, M. Giacometti, and F. Suchentrunk. 2010. 'Regional Genetic in Situ Differentiation despite Phylogenetic Heterogeneity in Alpine Mountain Hares'. *Journal of Zoology* 282 (1): 47–53. https://doi.org/10.1111/j.1469-7998.2010.00710.x.

Zeh, David W., Jeanne A. Zeh, and Yoichi Ishida. 2009. 'Transposable Elements and an Epigenetic Basis for Punctuated Equilibria'. *BioEssays* 31 (7): 715–26. https://doi.org/10.1002/bies.200900026.

Chapter II

Generating genomic resources to tackle fundamental evolutionary questions

This chapter is composed by two main articles, a first one where we assemble and characterize the first mountain hare transcriptome and produced a panel of ~5k putative diagnostic SNPs between the mountain and the European hare. This species-diagnostic panel can be used to design population assessing tools to monitor hybridization between the two species in several contact zones across its distribution.

In the second paper, we present the first annotated genome of a hare species, the mountain hare, and we briefly evaluate the use of pseudo-references to analyse genomic data of species without a species-specific reference genome.

In this chapter, genomic resources were produced that can in the future be used for precise population and evolutionary inferences in hares. In addition, these works were important to shape my critical thinking, train to writing and submiting articles, build my bioinformatic skills and strengthen the teamwork between me and my supervisors but also, between me and the team in which I am incorporated.

Both works were published in the SCI indexed journals, the first on *Scientific Data* (João P. Marques, Ferreira, et al. 2017) and the second on *Genome Biology and Evolution* (João P. Marques et al. 2020).

38 FCUP and U. Montpellier
Using genomic tools to understand species differentiation and admixture in hares and mice

Mountain hare transcriptome and diagnostic markers as resources to monitor hybridization with European hares

Scientific Data 4:170178
doi: 10.1038/sdata.2017.178

João P. Marques[1,2], Mafalda S. Ferreira[1,2], Colin M. Callahan[3], Klaus Hackländer[4], Hannes Jenny[5], W. Ian Montgomery[6,7], Neil Reid[6,7], Jeffrey M. Good[3], Paulo C. Alves[1,2], José Melo-Ferreira[1,2]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, 4485-661 Vairão, Portugal
[2]Departamento de Biologia, Faculdade de Ciências do Porto, 4169-007 Porto, Portugal
[3]Division of Biological Sciences, University of Montana, 32 Campus Drive, Missoula, MT 59812, USA
[4]Institute of Wildlife Biology and Game Management, BOKU-University of Natural Resources and Life Sciences, 1180 Vienna, Austria
[5]Amt für Jagd und Fischerei Graubünden, 7001 Chur, Switzerland
[6]Institute of Global Food Security, School of Biological Sciences, Queen's University Belfast, United Kingdom
[7]School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7BL, United Kingdom

Corresponding author: José Melo-Ferreira (jmeloferreira@cibio.up.pt)

Published: 5 December 2017

Abstract

We report the first mountain hare (*Lepus timidus*) transcriptome, produced by *de novo* assembly of RNA-sequencing reads. Data were obtained from eight specimens sampled in two localities, Alps and Ireland. The mountain hare tends to be replaced by the invading European hare (*Lepus europaeus*) in their numerous contact zones where the species hybridize, which affects their gene pool to a yet unquantified degree. We characterize and annotate the mountain hare transcriptome, detect polymorphism in the two analysed populations and use previously published data on the European hare (three specimens, representing the European lineage of the species) to identify 4 672 putative diagnostic sites between the species. These valuable genomic resources can be used to design tools to assess population status and monitor hybridization between species.

Design Type(s)	individual genetic characteristics comparison design • strain comparison design • population dynamics analysis objective
Measurement Type(s)	transcription profiling assay
Technology Type(s)	RNA sequencing
Factor Type(s)	geographic location
Sample Characteristic(s)	Lepus timidus • liver • Republic of Ireland • Switzerland

Background & Summary

The mountain hare (*Lepus timidus*) is an Arcto-alpine species that was the most common and widely distributed hare species across Europe during the last glacial periods(Carl Gustaf Thulin 2003). Nowadays, the mountain hare is distributed from Fennoscandia to Eastern Siberia, but also occurs in isolated/refuge populations (e.g. Ireland, Scotland, the Alps, Poland, the Baltics and Japan), and in places where it has been introduced (Iceland, England, Faroe Islands and New Zealand) (see Fig. 1). Even though they are a popular game species and abundant within its range, mountain hares have sharply declined in some regions, particularly in areas of contact with the European hare (*Lepus europaeus*), where the latter tends to invade and replace the range of the former (Carl Gustaf Thulin 2003; Reid and Montgomery 2007; Reid 2011; Caravaggi, Montgomery, and Reid 2015). Mountain and European hares share extensive natural and human-induced contact zones in Western Europe, from the British Isles to

Scandinavia and Central Europe (Fig.1). Climate change is predicted to affect lagomorphs extensively(Leach et al. 2015; Leach, Montgomery, and Reid 2015) and, in particular, to accelerate the replacement of mountain hares by European hares in the contact zones, such as the Alps, Sweden or Ireland (Acevedo et al. 2012a; Caravaggi et al. 2016). The two species may hybridize when in contact, resulting in some genetic introgression(C G Thulin, Jaarola, and Tegelstrom 1997; Carl Gustaf Thulin, Fang, and Averianov 2006; Suchentrunk et al. 2005; Melo-Ferreira et al. 2009; Zachos et al. 2010), with potential effects on local adaptation(Hughes et al. 2011).

Even though the mountain hare and other hare species have been the subject of several population genetics studies, these have been mostly based on a few markers(Hamill, Doyle, and Duke 2006; Carl Gustaf Thulin, Fang, and Averianov 2006; Melo-Ferreira et al. 2012a; 2014a). Therefore, permanent genomic resources provide fundamental information to develop monitoring tools to evaluate population status and implement protective policies. In this work, we use high-throughput RNA sequencing to: i) generate genomic resources for the mountain hare; and, ii) use published data on the European hare(Amoutzias et al. 2016) to pinpoint candidate fixed differences between the species that can be used to build genotyping tools to monitor gene exchange in the contact zones. We here present the first mountain hare transcriptome, and the most complete among the currently available *Lepus* transcriptomes.

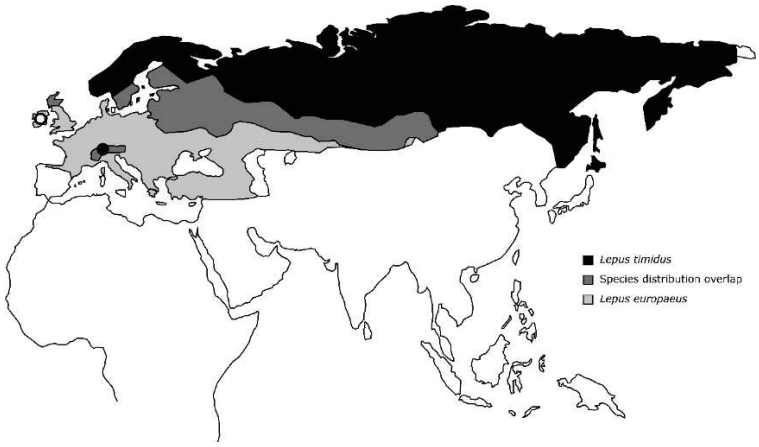


Figure 1: **Mountain and European hare distribution** - Distribution of the mountain hare, *Lepus timidus*, and the European hare, *L. europaeus*, in Eurasia with indication of the areas of contact and of broad geographic overlap between the species (approximate distribution ranges were adapted from IUCN Spatial Data Resources; IUCN 2016(IUCN 2016)). Circles indicate the sampling locations for this work (open circle – Ireland; closed circle – Alps).

Methods

A summary of the methodological workflow is shown in the flowchart of Fig. 2.

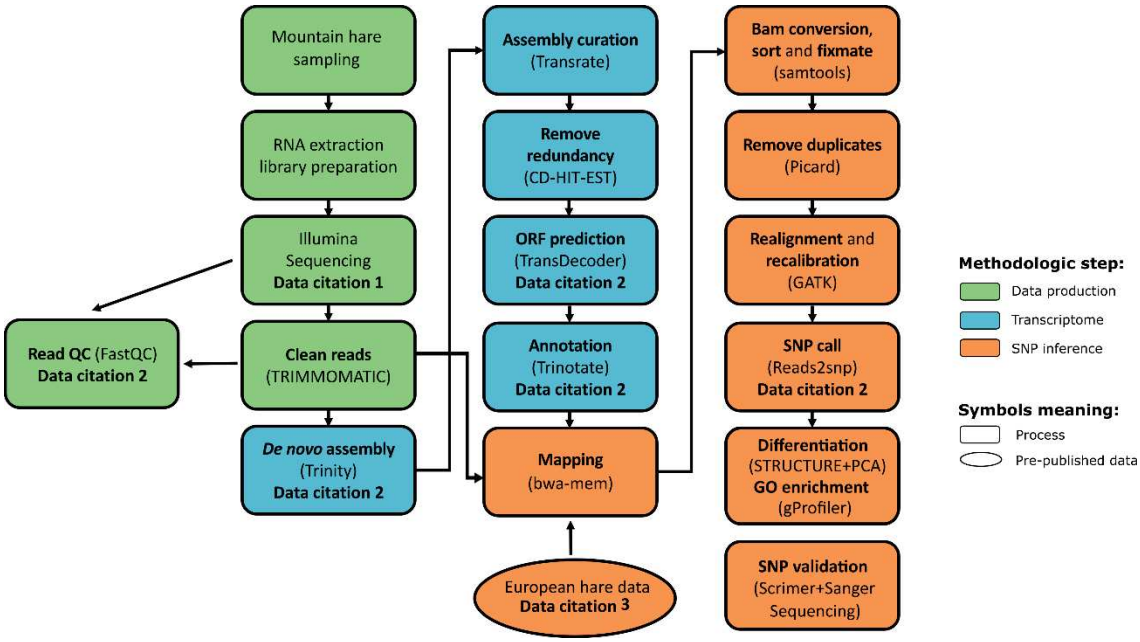


Figure 2: Methodological workflow – Flowchart of the RNA-seq setup and data analysis steps.

Sampling procedure and locations

Specimens from the Alps were sampled during regular permit hunting in Grisons, Switzerland. Specimens from Ireland were captured from the wild in Borris-in-Ossory, by the Irish Coursing Club (ICC) for scientific research purposes under National Parks & Wildlife (NPWS) licence No. C 337/2012 issued by the Department of Arts, Heritage and the Gaeltacht (dated 31/10/2012). Irish hares were dispatched humanely and in accordance with the licence conditions by means of lethal injection administered by Mr William Fitzgerald, Veterinary Laboratory Service Follow (MVB MVM CertCSM), from the Department of Agriculture, Food and the Marine, Regional Veterinary Laboratory, Hebron Road, Kilkenny, R95 TX39. Total RNA was isolated from 8 individuals.

RNA extraction

Liver tissue was freshly collected, immediately preserved in RNAlater and then stored at –80ºC until RNA extraction. Prior to extraction, frozen samples were ground in liquid nitrogen with a ceramic mortar and pestle. Mortar and pestle were washed prior to extraction using a 6-step wash that includes the following washing reagents in order: 70% ethanol, tap water, 10% bleach, milli-Q water, RNase away (Thermo Fisher

Scientific) and finishing with molecular grade H_2O. RNA extraction was performed using RNeasy® Mini Kit according to manufacturer instructions.

RNA sequencing library preparation

The SureSelect Strand-Specific RNA Library Prep for Illumina Multiplexed Sequencing (Agilent Technologies) kit was used to prepare cDNA libraries for all samples. Library sizes were estimated using a Bioanalyzer 2100 and quantified using KAPA Library quantification kit (KAPA BIOSYSTEMS). Equal molar concentrations of each library were pooled together for sequencing.

Sequence data processing and *de novo* transcriptome assembly

A first quality evaluation of obtained sequence reads (Data citation 1) was performed with FastQC v0.11.5(Andrews 2010b). After read quality inspection, adapters were removed and quality trimming performed using TRIMMOMATIC v0.36(a. M. Bolger, Lohse, and Usadel 2014), with instructions to remove the first ten bases, Illumina adapters, reads below 25 bp long and bases in the ends of reads with quality below 10, and to perform a 4-base sliding window trimming and cutting fragments with an average quality below 10. Trimmed-read quality was rechecked with FastQC (Data citation 2). A *de novo* transcriptome assembly was then performed using all properly paired reads from the eight individuals in the dataset using TRINITY v2.2.0(Grabherr et al. 2011a), establishing RF as read orientation for a strand-specific assembly. In addition, as a complementary resource, *de novo* transcriptome assemblies for each of the two sampling localities were also performed. Transrate v1.0.3(Smith-Unna et al. 2016b) was used to evaluate assembly quality and completeness and to remove possible chimeras and poorly supported contigs. Cleaned reads were mapped back to the produced assembly and only the well-supported contigs were retained (Transrate optimal cut-off > 0.024). In order to remove redundancy produced by using multi-sample data to perform the assembly, all contigs were clustered using CD-HIT-EST v4.6.4(Fu et al. 2012a) with a 95% similarity threshold. Open reading frames were predicted with TransDecoder v3.0.0(Haas et al. 2013) to remove possible contaminants such as non-coding RNA and DNA contamination. The final filtered transcriptome comprised contigs with predicted open reading frame and/or rabbit (*Oryctolagus cuniculus*) or pfam annotation. Filtered transcriptome as well as raw assemblies are available in *Figshare* (Data citation 2).

Transcriptome annotation

Transcriptome annotation was performed adapting the protocol of Trinotate v3.0.1(Haas et al. 2013), using i) Conditional Reciprocal Best BLAST (crb-blast) v0.6.6(Aubry et al. 2014) against the rabbit transcriptome reference (release 86) and Swiss-Prot database(Boutet et al. 2016); ii) protein domain identification by HMMER v3.1b2(Finn, Clements, and Eddy 2011) onto the PFAM database(Finn et al. 2014); iii) protein signal peptide through signal v 4.1(Petersen et al. 2011); iv) transmembrane domain prediction using tmHMM v2.0(Krogh et al. 2001); and v) eggNOG(Powell et al. 2012), GO(Gene Ontology Consortium 2000)and Kegg(Kanehisa et al. 2012) databases annotation. Annotation information was incorporated into a xls database (Data citation 2).

SNP inference

SNP calling was performed separately for mountain hares (Data citation 1) and European hares (Data citation 3, from Amoutzias *et al.*[18]). The three European hare specimens represent the European lineage of the species[18]. First, reads from all the individuals were mapped to the filtered mountain hare *de novo* transcriptome with bwa-mem v0.7.15(Li and Durbin 2010) with default parameters and read group information added to each sequencing lane-sample pair. The resulting alignments were converted to a binary file (bam format), sorted and submitted to fixmate step using SAMtools v1.3.1(Li et al. 2009). Duplicate reads were removed using Picard v1.140 (http://broadinstitute.github.io/picard) with the option MarkDuplicates. Realignment and recalibration was performed with Genome Analysis Toolkit v3.6-0(McKenna et al. 2010). Finally, SNP call was carried out using Reads2snp v2.0.64(Gayral et al. 2013) using a threshold of 20 for site and mapping qualities, the paralog filter, a minimum coverage of 10X and a genotype probability >0.95. The resulting VCF file was deposited in *Figshare* (Data citation 2).

Differentiation, admixture and Gene Ontology enrichment analysis

A set of random 5 502 SNPs, selected from independent contigs in order to reduce the linkage probability, was identified with VCFtools v0.1.14[38]. PGDSpyder v2.1.1.0[39] was used to convert this file to the required file formats. Partitions of genetic diversity in the dataset were investigated with a Principal Components Analysis, using PLINK v1.90b3.45[40] and ggplot2 R package[41] to plot the results. Additionally, the data were analysed using the admixture model implemented in STRUCTURE 2.3.4[42], with

three replicate runs with 1 million steps after a burn-in period of 200 000, and K=2. Results were plotted using CLUMPACK43. A Gene Ontology enrichment analysis was performed for the collection of contigs/genes with fixed differences between mountain and European hare samples and between mountain hare sampling localities. The analysis was based on the rabbit proteome annotations and performed with gProfiler[34], applying the g:SCS multiple test correction and the “best per parent group” hierarchical filter. The background set of genes was reduced to contigs with SNP information.

Independent SNP genotyping

A random set of 110 SNPs, inferred as potentially diagnostic between L. timidus and L. europaeus (Data citation 2, Supplementary Table 1, deposited in Figshare), was selected for independent validation using Sanger sequencing. DNA was extracted from two of the previously analysed mountain hare samples (one Alpine, Sample_3112, and one Irish, Sample_3103) and two other European hare specimens (sampled in Clermont-Ferrand – Sample 1569 – Font-Romeu, Pyrenees – Sample 1550 – in France during the regular hunting season). DNA extraction was performed using JETQUICK Tissue DNA Purification kit (Genomed). PCR primers were designed to be anchored in a single exon (taking into account intron-exon boundaries from the European rabbit reference genome) and to amplify a portion of 110 independent contigs containing at least one putative diagnostic SNP. The Primer sets were designed using the Scrimer pipeline44 which depends on Primer345 to design and set the primer conditions. A third internal sequencing primer was designed. PCRs were performed using QIAGEN Multiplex PCR Master Mix (Qiagen) and the following thermal cycling profile: initial denaturation at 95°C for 15', 35 cycles of denaturation at 95°C for 30'', annealing at 60-67°C for 20'' and elongation at 72°C for 30'', and a final extension step at 72°C for 5'. PCR products were visually inspected under UV-light after electrophoresis in agarose gels stained with GelRed (Biotium), purified with Exonuclease I (New England Biolabs) and FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), and sequenced using internal or, in a few cases, PCR primers in a ABI 3130xl genetic analyzer.

Code availability

Analyses in this work were performed with freely available, open access tools mainly using command line versions (detailed list in Table 1). Parameters are described in the methods section and software versions and commands used are detailed in Table 1.

Table 1: Open access tools and commands used to perform data analyses (analytical steps and colours correspond to those in Fig. 2).

Analytical Step	Description	Software/Version	Command
Read QC	Read quality control	FastQC v0.11.5	fastqc /path_to/raw.fq.gz (Data citation 1,2)
Clean Reads	Adaptor and low quality trimming	TRIMMOMATIC v0.36	java -jar /path_to/trimmomatic-0.36.jar PE -phred33 -threads 8 raw_R1.fq.gz raw_R2.fq.gz clean_FP.fq.gz clean_FU.fq.gz clean_RP.fq.gz clean_RU.fq.gz HEADCROP:10 ILLUMINACLIP:/path_to/adapters_list.fa:2:30:10 TRAILING:10 SLIDINGWINDOW:4:10 MINLEN:25
De novo assembly	Transcriptome assembly	Trinity v2.2.0	Trinity --seqType fq --left clean_FP.fq.gz --right clean_RP.fq.gz --CPU 20 --max_memory 150G --SS_lib_type RF --output trinity_assembly
Assembly curation	Filtering out contigs with low read support	Transrate v1.0.3	transrate --assembly Ltimidus_Trinity.fasta --left clean_FP.fq.gz --right clean_RP.fq.gz --threads 10 --reference Oryctolagus_cuniculus.OryCun2.0.81.pep.all.fa --output transrate_Ltimidus_Trinity
Remove redundancy	Clustering of highly homologous sequences	CD-HIT-EST v4.6.4	cd-hit-est -i good.Ltimidus_Trinity.fasta -c 0.95 -o AlpsIrel.fasta
ORF prediction	Filtering based on candidate coding regions and pfam annotation	TransDecoder v3.0.0	TransDecoder.LongOrfs -t AlpsIrel.fasta
		HMMER v3.1b2	hmmscan --cpu 8 --domtblout pfam.domtblout /path_to/Pfam-A.hmm transdecoder_dir/longest_orfs.pep
		TransDecoder v3.0.0	TransDecoder.Predict -t AlpsIrel.fasta --cpu 2 --retain_pfam_hits pfam.domtblout
Annotation	Annotation assessment	Trinotate v3.0.1	wget "https://data.broadinstitute.org/Trinity/Trinotate_v3_RESOURCES/Trinotate_v3.sqlite.gz" -O Trinotate.sqlite.gz
		Gunzip	gunzip Trinotate.sqlite.gz
	Conditional reciprocal best blast annotation	crb-blast v0.6.6	crb-blast --query AlpsIrel.cds --target database(SP and Ocun) --threads 4 --split 4 --output blastx.outfmt6
		crb-blast v0.6.6	crb-blast --query AlpsIrel.pep --target database(SP and Ocun) --threads 4 --split 4 --output blastp.outfmt6
	Signalp annotation	signalp v4.1	signalp -f short -n signalp.out AlpsIrel.pep
	Pfam annotation	HMMER v3.1b2	hmmscan --cpu 2 --domtblout TrinotatePFAM.out Pfam-A.hmm AlpsIrel.pep
	tmhmm annotation	tmHMM v2.0	tmhmm --short < AlpsIrel.pep > tmhmm.out
	Combine annotations	Trinity utilities v2.2.0	/path_to/trinityrnaseq-2.2.0/util/support_scripts/get_Trinity_gene_to_trans_map.pl AlpsIrel.fasta >AlpsIrel.gene_trans_map
		Trinotate v3.0.1	Trinotate Trinotate.sqlite init --gene_trans_map AlpsIrel.gene_trans_map --transcript_fasta AlpsIrel.fasta --transdecoder_pep AlpsIrel.pep

	SwissProt annotation load	Trinotate v3.0.1	Trinotate Trinotate.sqlite LOAD_swissprot_blastp SP.blastp.outfmt6 #and# Trinotate Trinotate.sqlite LOAD_swissprot_blastx SP.blastx.outfmt6
	O.cuniculus annotation load	Trinotate v3.0.1	1. Trinotate Trinotate.sqlite LOAD_custom_blast --outfmt6 Ocun.blastp.outfmt6 --prog blastp --dbtype Ocun; 2. Trinotate Trinotate.sqlite LOAD_custom_blast --outfmt6 Ocun.blastx.outfmt6 --prog blastx --dbtype Ocun
	Pfam annotation load	Trinotate v3.0.1	Trinotate Trinotate.sqlite LOAD_pfam TrinotatePFAM.out
	tmhmm annotation load	Trinotate v3.0.1	Trinotate Trinotate.sqlite LOAD_tmhmm tmhmm.out
	Signalp annotation load	Trinotate v3.0.1	Trinotate Trinotate.sqlite LOAD_signalp signalp.out
	Joint annotation file	Trinotate v3.0.1	Trinotate Trinotate.sqlite report > LtimidusTranscriptome.xls
Mapping	Read mapping into the curated reference	bwa-mem v0.7.15	bwa index AlpsIrel.cds
		bwa-mem v0.7.15	bwa mem -t 10 -R '@RG\tID:pop_sample_lane\tSM:popsample\tLB:LIBsample' AlpsIrel.cds Sample_L*_FP.fq.gz Sample_L*_RP.fq.gz > Sample_lane.sam
Bam conversion,sort and fixmate	Fixmate and BAM conversion	SAMtools v1.3.1	samtools fixmate --output-fmt BAM sample_lane.sam sample_lane_fixmate.bam
	BAM sort	SAMtools v1.3.1	samtools sort -O bam -o sample_lane_sorted.bam -T /path_to/temp/ sample_lane_fixmate.bam
Remove duplicates	Mark and remove duplicates	Picard v1.140	java -jar /path_to/picard.jar MarkDuplicates REMOVE_DUPLICATES=True MAX_FILE_HANDLES_FOR_READ_ENDS_MAP=950 ASSUME_SORTED=true VALIDATION_STRINGENCY=SILENT I=sample_lane_sorted.bam I=sample_lane_sorted.bam I=sample_lane_sorted.bam O=sample_rmdup.bam M=duplic_stats_sample TMP_DIR=/path_to/temp
Realignment and recalibration	Realignment	GATK v3.6-0	java -jar /path_to/GenomeAnalysisTK.jar -T RealignerTargetCreator -R AlpsIrel.cds -I sample_rmdup.bam -o sample_int.list
	Recalibration	GATK v3.6-0	java -jar /path_to/GenomeAnalysisTK.jar -T IndelRealigner -R AlpsIrel.cds -I sample_rmdup.bam -targetIntervals sample_int.list -o sample_realign.bam
SNP call	SNP call	Reads2snp v2.0.64	reads2snp_2.0.64.bin -bamlist LtimLeur_list.txt -bamref AlpsIrel.cds -out LtimVsLeur -min 10 -nbth 12 -th1 0.95 -par 1 -th2 0.01 -opt bfgs -fis 0.0 -pre 0.001 -rqt 20
Differentiation analysis	Remove indels and missing data	VCFtools v0.1.14	vcftools --vcf LtimVsLeur.vcf --recode --recode-INFO-all --remove-indels --max-missing-count 0 --out LtimVsLeur_noindels
	Extract 1SNP per contig	VCFtools v0.1.14	vcftools --vcf LtimVsLeur_noindels.recode.vcf --recode --recode-INFO-all --thin 10000 --min-alleles 2 --out LtimVsLeur_1SNPperContig
	VCF to STRUCTURE convertion	PGDSpyder v2.1.1.0	java -Xmx1024m -Xms512m -jar /path_to/PGDSpider2-cli.jar -inputfile LtimVsLeur_1SNPperContig.recode.vcf -inputformat VCF -outputfile LtimVsLeur_SNPs -outputformat STRUCTURE -spid VCF_to_STRUCTURE.spid
	Structure analysis	STRUCTURE v2.3.4	structure -m mainparams (standard parameters except 1 million steps after a burn-in period of 200 000, K=2 and admixture model)
		CLUMPACK v42089	The Web version was used - http://clumpak.tau.ac.il/
	PCA analysis	PLINK v1.90b3.45	plink --file LtimVsLeur_1SNPperContig --pca 3
		ggplot2 R package v2.2.1	1. R; 2. library(ggfortify); 3. pca <- read.table('plink.eigenvec', header=TRUE); 4. df <- pca[c(3, 4)]; 5. autoplot(prcomp(df), data = pca, colour = 'Species.Pop', size = 5)

Data Records

Forty-eight raw FASTQ files were submitted to *NCBI Sequence Read Archive*, with BioProject accession PRJNA358867 (Data Citation 1 and Tables 2-3). FASTQ files were divided in two sets, corresponding to the sampling localities (Ltim_Ireland and Ltim_Alps), and by biosample-specimen (SAMN06186748-3101, SAMN06186761-3102, SAMN06186762-3103 and SAMN06186763-3105; SAMN06186727-3112, SAMN06186728-3113, SAMN06186729-3114 and SAMN06186738-3116). In each biosample, six files were submitted, corresponding to three different Illumina HiSeq sequencing lanes and two read directions. Pre/post-cleaning FASTQC base quality pdf report (FASTQC.pdf) can be accessed in *Figshare* (Data Citation 2). This dataset is the core of this work and has not been released or analysed previously.

Trinity raw assemblies (Ltimidus_Trinity.fasta, LtimidusIreland_Trinity.fasta and LtimidusAlps_Trinity.fasta) were deposited on *Figshare* (Data citation 2). The curated transcriptome assembly fasta files (LtimidusTranscriptome.cds.fasta and LtimidusTranscriptome.pep.fasta) and the annotated xls database file (LtimidusTranscriptome.xls) can also be found in *Figshare* (Data citation 2).

The European hare data used here (Data Citation 3) was previously published by Amoutzias *et al.*(Amoutzias et al. 2016) (*GenBank* Project SRP055741, samples SRR1823098, SRR1863103 and SRR1863605).

Mapping statistics (Table 5), SNP call VCF file (LtimVsLeur.vcf) and population/species diagnostic SNPs tables (Supplementary Tables 1-4) were deposited in *Figshare* (Data citation 2).

Table 2: Summary of sample data information deposited in the NCBI database.

Sample ID	Species (population)	Tissue	Method	NCBI BioSample ID
Sample_3101	*Lepus timidus* (Ireland)	liver	RNA-seq	SAMN06186748
Sample_3102	*Lepus timidus* (Ireland)	liver	RNA-seq	SAMN06186761
Sample_3103	*Lepus timidus* (Ireland)	liver	RNA-seq	SAMN06186762
Sample_3105	*Lepus timidus* (Ireland)	liver	RNA-seq	SAMN06186763
Sample_3112	*Lepus timidus* (Alps)	liver	RNA-seq	SAMN06186727
Sample_3113	*Lepus timidus* (Alps)	liver	RNA-seq	SAMN06186728
Sample_3114	*Lepus timidus* (Alps)	liver	RNA-seq	SAMN06186729
Sample_3116	*Lepus timidus* (Alps)	liver	RNA-seq	SAMN06186738

Table 3: Illumina RNA-seq data deposited in the NCBI database.

Sample ID	NCBI SRA runs accession	Raw reads	Mbytes
Sample_3101	SRR5133282	26,598,712	2,525
Sample_3102	SRR5133280	26,128,525	2,532
Sample_3103	SRR5133285	24,469,456	2,414
Sample_3105	SRR5133283	26,662,182	2,582
Sample_3112	SRR5133287	22,444,667	2,263
Sample_3113	SRR5133281	20,825,930	2,100
Sample_3114	SRR5133286	32,749,011	3,294
Sample_3116	SRR5133284	21,690,965	2,189

Technical Validation

RNA integrity

The quality and quantity of each RNA sample was assessed using the 260/280 and 260/230 absorbance ratios estimated by an IMPLEN P330 NanoPhotometer and RNA Integrity Number (RIN) and concentration (μg/μL) with a Bioanalyzer 2100 (Agilent Technologies). All samples had RIN values above 8.

RNA-Seq data quality

The Illumina HiSeq run produced a total raw output of 103 941 215 100 bp paired-end reads (207 882 430 total reads). Adapter removal and quality trimming decreased this number to 201 569 448 reads (97%) (Table 4). Final analysed reads passed the minimum quality parameters as established by FastQC.

Transcriptome assembly curation, annotation and quality

Cleaned reads were assembled into 272 183 contigs with a mean length of 594 bp and a N50 length of 839 bp (Table 4). After assembly curation with Transrate optimal cut-off > 0.024, clustering with a 95% similarity threshold and open reading frame prediction, were retained 25 868 transcripts with a mean length of 842 bp and a N50 length of 1 182 (Table 4).

Annotation using a conditional reciprocal best blast hit approach results in 16 772 (65 %) annotated transcripts, of which 13 641 were annotated to the rabbit transcriptome and 15 955 to the Swiss-Prot database (Fig. 3). In order to reduce the number of non-

annotated transcripts, the less stringent unidirectional blast hit was added to the database. Hits were recovered for 25 549 transcripts (99%) (Fig. 3).

The mountain hare transcriptome produced in this study represents an important improvement compared to the currently available transcriptomic resources for *Lepus* – *L. granatensis*(Joao P Marques, Farelo, et al. 2017) and *L. europaeus*(Amoutzias et al. 2016) transcriptomes – as it performs better on several assembly statistics, such as reference coverage (42% vs. 32% in *L. granatensis* and 40% in *L. europaeus*; using the rabbit transcriptome as reference).

Table 4: Mountain hare transcriptome assembly and curation characterisation.

***Lepus timidus* transcriptome**	**Value**
Raw Reads	207,882,430
Clean Reads	201,569,448
Mapped Reads	136,511,846
Raw *de novo* assembly (Trinity)	
Number of contigs	272,183
Largest (bp)	14,048
Smallest (bp)	201
N50 (bp)	839
Mean (bp)	594
Post assembly curation (TransRate)	
Number of contigs	113,694
Largest (bp)	14,048
Smallest (bp)	201
N50 (bp)	801
Mean (bp)	567
Post redundancy removal (CD-HIT-EST)	
Number of contigs	109,239
Largest (bp)	14,048
Smallest (bp)	201
N50 (bp)	765
Mean (bp)	554
Post open reading frame prediction (TransDecoder)	
Number of contigs	25,868
Largest (bp)	13,728
Smallest (bp)	297
N50 (bp)	1,182
Mean (bp)	842
Reference Coverage (%)	42

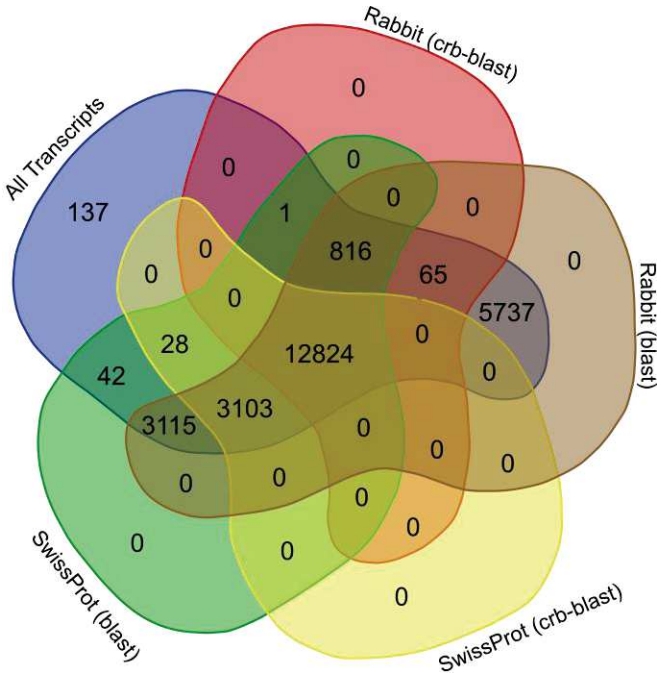


Figure 3: Annotation summary - Number of transcripts annotated with different combinations of methods and databases: all transcripts; transcripts annotated with crb-blast against rabbit transcriptome; transcripts annotated with a unidirectional BLASTx against rabbit transcriptome; transcripts annotated with crb-blast against the Swiss-Prot database; and transcripts annotated with a unidirectional BLASTx against the Swiss-Prot database.

Genetic variation, differentiation and gene ontology enrichment

In total, 218 057 526 reads (63%) were mapped to the filtered transcriptome – 136 511 846 mountain hare reads (68%) and 81 545 680 European hare reads (57%) (see statistics in Table 5). After filtering, 159 629 high-quality SNPs were inferred, of which 41 182 (26%) were sequenced in all eleven specimens. A summary of polymorphic, shared and fixed SNPs is shown in Fig. 3. 4 672 putative species-diagnostic SNPs (considered when species presented alternative fixed alleles) were inferred (Data citation 2, Supplementary Tables 1-3, also deposited in Figshare). The diagnostic power of our SNP set could be strongly reduced if any of the sequenced specimens was admixed (namely from the Alps, where the species overlap). We therefore conducted a Principal Component Analysis and a Bayesian Assignment analysis to assess our ability to separate the species. The results suggest that the analysed mountain and European hares are well differentiated with our SNP set, and only possible limited levels of admixture were found for individual 3116 (Fig. 4). An extra table of species-diagnostic SNPs excluding that individual was therefore produced (Data citation 2, Supplementary

Table 4, also deposited in Figshare). 25 269 SNPs were inferred in the mountain hare, of which 12 548 and 18 591 were polymorphic in the Irish and Alpine samples respectively, and 126 were fixed between sampling localities (Data citation 2, Supplementary Tables 5-7, deposited in Figshare). The “membrane part” gene ontology term was found enriched in the collection of genes with fixed differences between the Irish and Alpine mountain hare samples, while terms “lipid metabolic process”, “small molecule catabolic process”, “extracellular space and acyl-CoA dehydrogenase activity” were found enriched in genes with fixed differences between samples of the two species. Note however that even though the background gene set was controlled for, RNA-sequencing data does not provide an unbiased sample of information across different genes and these results may represent tissue-related functions.

Table 5: Mapping statistics.

Sample ID	Species (population)	Raw reads #	Mapped reads #	Mapped reads %
Sample_3101	*Lepus timidus* (Ireland)	26,598,712	19,648,435	74
Sample_3102	*Lepus timidus* (Ireland)	26,128,525	18,781,893	72
Sample_3103	*Lepus timidus* (Ireland)	24,469,456	16,102,091	66
Sample_3105	*Lepus timidus* (Ireland)	26,662,182	18,429,333	69
Sample_3112	*Lepus timidus* (Alps)	22,444,667	13,913,982	62
Sample_3113	*Lepus timidus* (Alps)	20,825,930	13,935,177	67
Sample_3114	*Lepus timidus* (Alps)	32,749,011	21,360,771	65
Sample_3116	Lepus timidus (Alps)	21,690,965	14,340,164	66
Sample_H1	Lepus europaeus	20,825,930	14,100,961	62
Sample_H2	*Lepus europaeus*	32,749,011	28,922,352	57
Sample_H3	*Lepus europaeus*	21,690,965	38,522,367	55

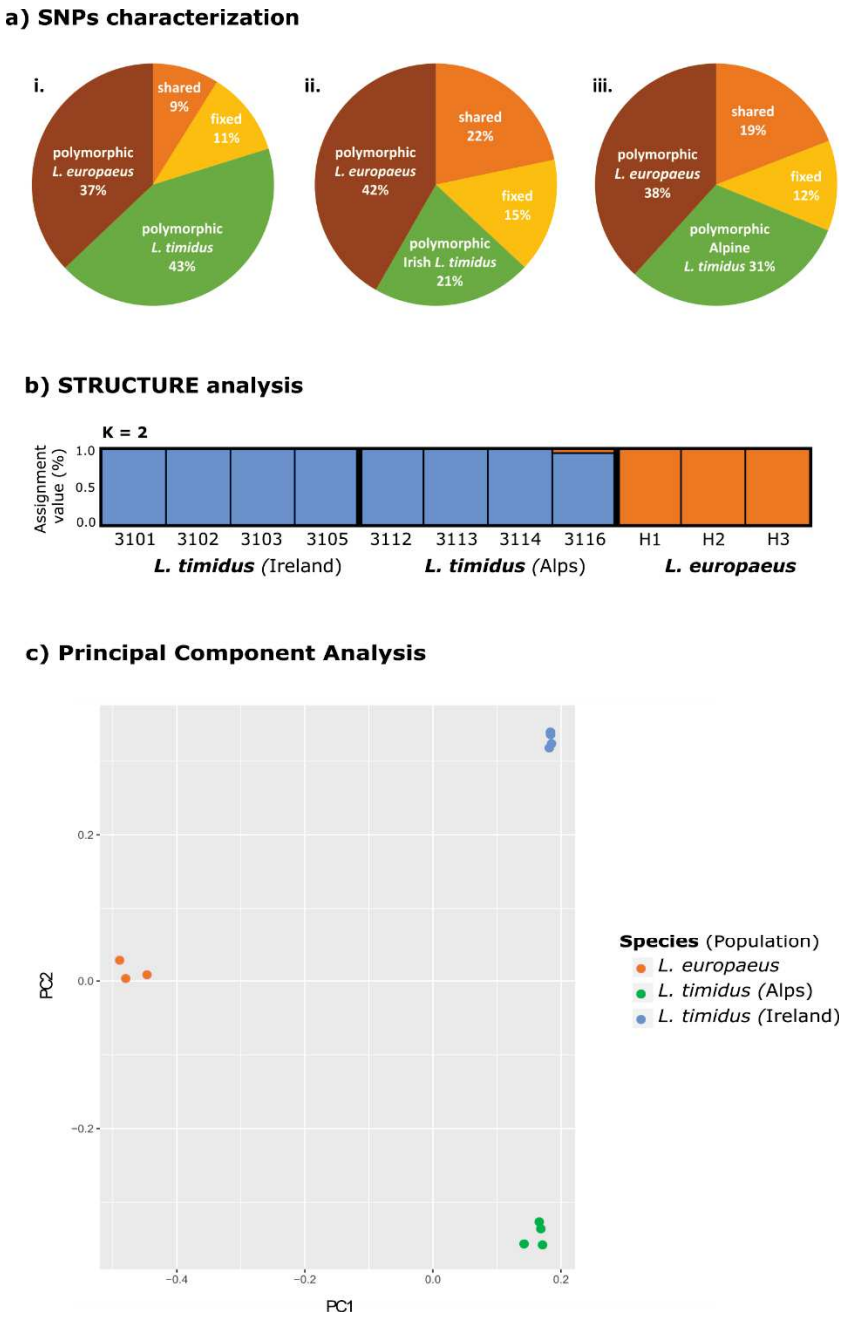


Figure 4: Characterization of inferred SNPs in the sampled populations and species – **a)** Relative proportion of the 41 182 SNPs mapped to the mountain hare transcriptome, summarized as polymorphic within each species and fixed or shared between *L. timidus* (mountain hare) and *L. europaeus* (European hare). The proportion is shown considering the complete *L. timidus* dataset (**i**) and only the Irish (**ii**) and Alpine (**iii**) populations. **b)** STRUCTURE analysis to evaluate cluster membership and admixture proportions. Individuals are sorted by population and species. Mountain hare populations are shown in blue and European hare individuals in orange. **c)** Principal Component Analysis (PCA) plot using one SNP per contig. The first principal component (PC1) splits species and the second (PC2) the sampled populations.

SNP Validation

Independent SNP genotyping was performed for a random subset of 110 putative species-diagnostic SNPs, which laid on different contigs. Technical validation was considered successful for SNPs showing the expected alternative alleles, being one fixed in L. timidus (note that the sequenced L. europaeus specimens differed from the RNA-sequencing). PCR amplification was successful for 96 of the 110 target contigs (87%), 88 amplicons were successfully sequenced in both species (92%), and concordance between sequences and expected SNPs was obtained for 85 of the sequenced fragments (97%). This represents an overall validation success of 77%, which compares to studies using similar approaches47–49 (Data citation 2; see Supplementary Table 8 for full genotyping results and Supplementary Table 9 with the list of all primers, both deposited in Figshare). This technical validation proportion is conservative, as it is reduced by technical issues in PCR amplification and sequencing, and potential intraspecific polymorphism in the European hare (given the use of two different samples for validation), in addition to real false positives. From the validated SNPs, 71 confirmed alternate alleles in the species, but their diagnostic utility should be tested with larger population sampling.

Usage Notes

These genomic resources will be useful for a variety of studies, particularly in the characterization of genetic diversity in mountain hare populations and on the development of hybridization monitoring tools. Note that SNPs were here inferred from an uneven and small species sample, and therefore any diagnostic genotyping assay built from this data should be first tested with adequate sample sizes from pure parental populations of the species, before being applied to hybrid zones.

Acknowledgements

This work was partially funded by project HybridAdapt with reference FCT-ANR/BIA-EVF/0250/2012, supported by Portuguese National Funds through the Fundação para a Ciência e a Tecnologia, FCT. Additional funding was obtained from POPH-QREN funds, from the European Social Fund and FCT (IF/00033/2014 FCT Investigator grant to JMF and PD/BD/108131/2015 PhD grant to MSF); from project "Genomics and Evolutionary Biology", co-financed by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte) under the National Strategic Reference Framework (NSRF) and the European Regional Development Fund (ERDF); and from an EU FP7 REGPOT grant (CIBIO-New-Gen) (286431). Laboratory work at the University of Montana was performed in the UM Genomics Core, supported with a grant from the M.J. Murdock Charitable Trust. Transcriptome sequencing was performed through the Vincent J. Coates Genomics Sequencing Laboratory at University of California Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. We are grateful to René Gadient and local hunters in Grisons for their tremendous effort in collecting samples under Alpine conditions. Laura Claffey issued the licence to take animals for scientific purposes in Ireland (dated 31/10/2012) from the Wildlife Licensing Unit, National Parks & Wildlife Service (NPWS), Department of Arts, Heritage & the Gaeltacht, and Mr. Jimi Conroy, the local NPWS Conservation Officer consented to sampling within his jurisdiction. We thank the Irish Coursing Club (ICC) for providing hares, and in particular Mr D.J. Histon, Chief Executive and Secretary, for supporting the research. Thanks also to Mr William Fitzgerald, Department of Agriculture, Food and the Marine for dispatching animals in a humane manner.

Author contributions

JPM and JMF conceived the study. JMF, PCA, NR, WIM, KH and HJ organized and performed the sampling. CC and MSF performed the laboratory work at the University of Montana under the supervision of JMG. JPM analysed the data. JPM wrote the first draft with contributions from JMF and MSF. All authors read, revised and approved the manuscript.

Competing interests

The authors declare no competing financial interests.

Supplementary Tables (Annexes)

Supplementary Table 1 - Species-diagnostic SNPs table considering the two *L. timidus* populations.

Supplementary Table 2 - Species-diagnostic SNPs table considering only Irish *L. timidus* individuals.

Supplementary Table 3 - Species-diagnostic SNPs table considering only Alpine *L. timidus* individuals.

Supplementary Table 4 - Species-diagnostic SNPs table considering the two *L. timidus* populations, but excluding individual 3116.

References

Acevedo, Pelayo, Alberto Jiménez-Valverde, José Melo-Ferreira, Raimundo Real, and Paulo Célio Alves. 2012a. 'Parapatric Species and the Implications for Climate Change Studies: A Case Study on Hares in Europe'. *Global Change Biology* 18 (5): 1509–19. https://doi.org/10.1111/j.1365-2486.2012.02655.x.

———. 2012b. 'Parapatric Species and the Implications for Climate Change Studies: A Case Study on Hares in Europe'. *Global Change Biology* 18 (5): 1509–19. https://doi.org/10.1111/j.1365-2486.2012.02655.x.

Acevedo, Pelayo, José Melo-Ferreira, Raimundo Real, and Paulo C. Alves. 2014. 'Evidence for Niche Similarities in the Allopatric Sister Species Lepus Castroviejoi and Lepus Corsicanus'. *Journal of Biogeography* 41 (5): 977–86. https://doi.org/10.1111/jbi.12270.

Altuna, Echave. 1970. 'Hallazgo de una liebre ártica (Lepus timidus L.) en el yacimiento prehistórico de Urtiaga (Guipúzcoa)', 4.

Alves, Paulo C., José Melo-Ferreira, Hélder Freitas, and Pierre Boursot. 2008. 'The Ubiquitous Mountain Hare Mitochondria: Multiple Introgressive Hybridization in Hares, Genus Lepus'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1505): 2831–39. https://doi.org/10.1098/rstb.2008.0053.

Alves, P.C., J. Melo-Ferreira, M. Branco, F. Suchentrunk, N. Ferrand, and D.J. Harris. 2008. '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): 1191–97. https://doi.org/10.1016/j.ympev.2007.11.010.

Amoutzias, Grigoris D., Themistoklis Giannoulis, Katerina A. Moutou, Anna Maria G Psarra, Costas Stamatis, Andreas Tsipourlianos, and Zissis Mamuris. 2016. 'SNP Identification through Transcriptome Analysis of the European Brown Hare (Lepus Europaeus): Cellular Energetics and Mother's Curse'. *PLoS ONE* 11 (7). https://doi.org/10.1371/journal.pone.0159939.

Andrews, Simon. 2010a. 'FastQC'. *Babraham Bioinformatics*. https://doi.org/citeulike-article-id:11583827.

———. 2010b. 'FastQC: A Quality Control Tool for High Throughput Sequence Data.' *Babraham Bioinformatics*, http://www.bioinformatics.babraham.ac.uk/projects/. https://doi.org/citeulike-article-id:11583827.

Arnegard, Matthew E., Matthew D. McGee, Blake Matthews, Kerry B. Marchinko, Gina L. Conte, Sahriar Kabir, Nicole Bedford, et al. 2014. 'Genetics of Ecological

Divergence during Speciation'. *Nature* 511 (7509): 307–11. https://doi.org/10.1038/nature13301.

Arnold, Michael L. 1992. 'Natural Hybridization as an Evolutionary Process'. *Annual Review of Ecology and Systematics* 23 (1): 237–61. https://doi.org/10.1146/annurev.es.23.110192.001321.

Aubry, Sylvain, Steven Kelly, Britta M C Kümpers, Richard D. Smith-Unna, and Julian M. Hibberd. 2014. 'Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C4 Photosynthesis'. *PLoS Genetics* 10 (6). https://doi.org/10.1371/journal.pgen.1004365.

Baker, Christopher L., Shimpei Kajita, Michael Walker, Ruth L. Saxl, Narayanan Raghupathy, Kwangbom Choi, Petko M. Petkov, and Kenneth Paigen. 2015. 'PRDM9 Drives Evolutionary Erosion of Hotspots in Mus Musculus through Haplotype-Specific Initiation of Meiotic Recombination'. *PLOS Genetics* 11 (1): e1004916. https://doi.org/10.1371/journal.pgen.1004916.

Barton, N. H., and G. M. Hewitt. 1989. 'Adaptation, Speciation and Hybrid Zones'. *Nature* 341 (6242): 497–503. https://doi.org/10.1038/341497a0.

Barton, Nick, and Bengt Olle Bengtsson. 1986. 'The Barrier to Genetic Exchange between Hybridising Populations'. *Heredity* 57 (3): 357–76. https://doi.org/10.1038/hdy.1986.135.

Bateson. 1909. *Heredity and Variation in Modern Lights*. http://archive.org/details/cbarchive_36304_hereditydandvariationinmodernli1909.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop, and B. de Massy. 2010. 'PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice'. *Science (New York, N.Y.)* 327 (5967): 836–40. https://doi.org/10.1126/science.1183439.

Bayes, Joshua J., and Harmit S. Malik. 2009. 'Altered Heterochromatin Binding by a Hybrid Sterility Protein in Drosophila Sibling Species'. *Science (New York, N.Y.)* 326 (5959): 1538–41. https://doi.org/10.1126/science.1181756.

Beck, Jon A., Sarah Lloyd, Majid Hafezparast, Moyha Lennon-Pierce, Janan T. Eppig, Michael F. W. Festing, and Elizabeth M. C. Fisher. 2000. 'Genealogies of Mouse Inbred Strains'. *Nature Genetics* 24 (1): 23–25. https://doi.org/10.1038/71641.

Böhne, Astrid, Frédéric Brunet, Delphine Galiana-Arnoux, Christina Schultheis, and Jean-Nicolas Volff. 2008. 'Transposable Elements as Drivers of Genomic and Biological Diversity in Vertebrates'. *Chromosome Research* 16 (1): 203–15. https://doi.org/10.1007/s10577-007-1202-6.

Bolger, a. M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data'. *Bioinformatics* 30 (15): 2114–20.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data'. *Bioinformatics (Oxford, England)* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Boursot, P., J.-C. Auffray, J. Britton-Davidian, and F. Bonhomme. 1993. 'The Evolution of House Mice'. *Annual Review of Ecology and Systematics* 24: 119–52.

Boutet, Emmanuel, Damien Lieberherr, Michael Tognolli, Michel Schneider, Parit Bansal, Alan J Bridge, Sylvain Poux, Lydie Bougueleret, and Ioannis Xenarios. 2016. 'UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View.' *Methods in Molecular Biology (Clifton, N.J.)* 1374: 23–54. https://doi.org/10.1007/978-1-4939-3167-5_2.

Brandvain, Yaniv, Amanda M. Kenney, Lex Flagel, Graham Coop, and Andrea L. Sweigart. 2014. 'Speciation and Introgression between Mimulus Nasutus and Mimulus Guttatus'. Edited by Chris D. Jiggins. *PLoS Genetics* 10 (6): e1004410. https://doi.org/10.1371/journal.pgen.1004410.

Brennecke, Julius, Colin D. Malone, Alexei A. Aravin, Ravi Sachidanandam, Alexander Stark, and Gregory J. Hannon. 2008. 'An Epigenetic Role for Maternally Inherited PiRNAs in Transposon Silencing'. *Science (New York, N.Y.)* 322 (5906): 1387–92. https://doi.org/10.1126/science.1165171.

Brideau, Nicholas J., Heather A. Flores, Jun Wang, Shamoni Maheshwari, Xu Wang, and Daniel A. Barbash. 2006. 'Two Dobzhansky-Muller Genes Interact to Cause Hybrid Lethality in Drosophila'. *Science (New York, N.Y.)* 314 (5803): 1292–95. https://doi.org/10.1126/science.1133953.

Brown, Judith D., and Rachel J. O'Neill. 2010. 'Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited'. *Annual Review of Genomics and Human Genetics* 11 (1): 291–316. https://doi.org/10.1146/annurev-genom-082509-141554.

Brown, Judith D., Vanessa Piccuillo, and Rachel J. O'Neill. 2012. 'Retroelement Demethylation Associated with Abnormal Placentation in Mus Musculus × Mus Caroli Hybrids1'. *Biology of Reproduction* 86 (3): 88, 1–10. https://doi.org/10.1095/biolreprod.111.095273.

Burt, Austin, and Robert Trivers. 2006. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Cambridge, Mass: Belknap Press of Harvard University Press.

Butlin, Roger K, and Sean Stankowski. 2020. 'Is It Time to Abandon the Biological Species Concept? No'. *National Science Review* 7 (8): 1400–1401. https://doi.org/10.1093/nsr/nwaa109.

Calfee, Erin, Daniel Gates, Anne Lorant, M. Taylor Perkins, Graham Coop, and Jeffrey Ross-Ibarra. 2021. 'Selective Sorting of Ancestral Introgression in Maize and Teosinte along an Elevational Cline'. https://doi.org/10.1101/2021.03.05.434040.

Camus, M. Florencia, Jochen B. W. Wolf, Edward H. Morrow, and Damian K. Dowling. 2015. 'Single Nucleotides in the MtDNA Sequence Modify Mitochondrial Molecular Function and Are Associated with Sex-Specific Effects on Fertility and Aging'. *Current Biology* 25 (20): 2717–22. https://doi.org/10.1016/j.cub.2015.09.012.

Caravaggi, Anthony, Katie Leach, Francesco Santilli, Jukka Rintala, Pekka Helle, Juha Tiainen, Francesco Bisi, Adriano Martinoli, W Ian Montgomery, and Neil Reid. 2016. 'Niche Overlap of Mountain Hare Subspecies and the Vulnerability of Their Ranges to Invasion by the European Hare; the (Bad) Luck of the Irish'. *Biological Invasions*, 1–20. https://doi.org/10.1007/s10530-016-1330-z.

Caravaggi, Anthony, Katie Leach, Francesco Santilli, Jukka Rintala, Pekka Helle, Juha Tiainen, Francesco Bisi, Adriano Martinoli, W. Ian Montgomery, and Neil Reid. 2017. 'Niche Overlap of Mountain Hare Subspecies and the Vulnerability of Their Ranges to Invasion by the European Hare; the (Bad) Luck of the Irish'. *Biological Invasions* 19 (2): 655–74. https://doi.org/10.1007/s10530-016-1330-z.

Caravaggi, Anthony, W Ian Montgomery, and Neil Reid. 2015. 'Range Expansion and Comparative Habitat Use of Insular, Congeneric Lagomorphs: Invasive European Hares Lepus Europaeus and Endemic Irish Hares Lepus Timidus Hibernicus'. *Biological Invasions* 17 (2): 687–98. https://doi.org/10.1007/s10530-014-0759-1.

Case, Andrea L., Findley R. Finseth, Camille M. Barr, and Lila Fishman. 2016. 'Selfish Evolution of Cytonuclear Hybrid Incompatibility in Mimulus'. *Proceedings of the Royal Society B: Biological Sciences* 283 (1838): 20161493. https://doi.org/10.1098/rspb.2016.1493.

Chafee, Meghan E., Courtney N. Zecher, Michelle L. Gourley, Victor T. Schmidt, John H. Chen, Sarah R. Bordenstein, Michael E. Clark, and Seth R. Bordenstein. 2011. 'Decoupling of Host–Symbiont–Phage Coadaptations Following Transfer Between Insect Species'. *Genetics* 187 (1): 203–15. https://doi.org/10.1534/genetics.110.120675.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. 'The Effect of Deleterious Mutations on Neutral Molecular Variation'. *Genetics* 134 (4): 1289–1303.

Chase, Kevin, David R. Carrier, Frederick R. Adler, Elaine A. Ostrander, and Karl G. Lark. 2005. 'Interaction between the X Chromosome and an Autosome Regulates Size Sexual Dimorphism in Portuguese Water Dogs'. *Genome Research* 15 (12): 1820–24. https://doi.org/10.1101/gr.3712705.

Christie, Kyle, and Sharon Y. Strauss. 2018. 'Along the Speciation Continuum: Quantifying Intrinsic and Extrinsic Isolating Barriers across Five Million Years of Evolutionary Divergence in California Jewelflowers'. *Evolution; International Journal of Organic Evolution* 72 (5): 1063–79. https://doi.org/10.1111/evo.13477.

Christie, Paul, and Mark R. Macnair. 1987. 'The Distribution of Postmating Reproductive Isolating Genes in Populations of the Yellow Monkey Flower, Mimulus Guttatus'. *Evolution* 41 (3): 571–78. https://doi.org/10.1111/j.1558-5646.1987.tb05827.x.

Consortium, Heliconius Genome. 2012. 'Butterfly Genome Reveals Promiscuous Exchange of Mimicry Adaptations among Species.' *Nature* 487 (7405): 94–98. https://doi.org/10.1038/nature11041.

Corbett-Detig, Russell, and Rasmus Nielsen. 2017. 'A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy'. *PLOS Genetics* 13 (1): e1006529. https://doi.org/10.1371/journal.pgen.1006529.

Courret, Cécile, Ching-Ho Chang, Kevin H.-C. Wei, Catherine Montchamp-Moreau, and Amanda M. Larracuente. 2019. 'Meiotic Drive Mechanisms: Lessons from Drosophila'. *Proceedings of the Royal Society B: Biological Sciences* 286 (1913): 20191430. https://doi.org/10.1098/rspb.2019.1430.

Coyne, Jerry A., and H. Allen Orr. 1997. '"Patterns of Speciation in Drosophila" Revisited'. *Evolution* 51 (1): 295–303. https://doi.org/10.2307/2410984.

———. 2004. *Speciation*. Oxford, New York: Oxford University Press.

Crespi, Bernard. 2008. 'Turner Syndrome and the Evolution of Human Sexual Dimorphism'. *Evolutionary Applications* 1 (3): 449–61. https://doi.org/10.1111/j.1752-4571.2008.00017.x.

Crespi, Bernard, and Patrik Nosil. 2013. 'Conflictual Speciation: Species Formation via Genomic Conflict'. *Trends in Ecology & Evolution* 28 (1): 48–57. https://doi.org/10.1016/j.tree.2012.08.015.

Cruickshank, Tami E., and Matthew W. Hahn. 2014. 'Reanalysis Suggests That Genomic Islands of Speciation Are Due to Reduced Diversity, Not Reduced Gene Flow.' *Molecular Ecology* 23 (13): 3133–57. https://doi.org/10.1111/mec.12796.

Cucchi, Thomas, Jean-Christophe Auffray, and Jean-Denis Vigne. 2012. 'On the Origin of the House Mouse Synanthropy and Dispersal in the Near East and Europe: Zooarchaeological Review and Perspectives'. In , 65–93.

Cucchi, Thomas, Jean-Denis Vigne, and Jean-Christophe Auffray. 2005. 'First Occurrence of the House Mouse (Mus Musculus Domesticus Schwarz & Schwarz, 1943) in the Western Mediterranean: A Zooarchaeological Revision of Subfossil Occurrences'. *Biological Journal of the Linnean Society* 84 (3): 429–45. https://doi.org/10.1111/j.1095-8312.2005.00445.x.

Currat, Mathias, Manuel Ruedi, Rémy J. Petit, and Laurent Excoffier. 2008. 'The Hidden Side of Invasions: Massive Introgression by Local Genes'. *Evolution* 62 (8): 1908–20. https://doi.org/10.1111/j.1558-5646.2008.00413.x.

Darwin, Charles, and Leonard Kebler. 1859. *On the Origin of Species by Means of Natural Selection*. https://www.loc.gov/item/06017473/.

Delmore, Kira E., and Darren E. Irwin. 2014. 'Hybrid Songbirds Employ Intermediate Routes in a Migratory Divide'. *Ecology Letters* 17 (10): 1211–18. https://doi.org/10.1111/ele.12326.

Din, W., R. Anand, P. Boursot, D. Darviche, B. Dod, E. Jouvin-Marche, A. Orth, G. P. Talwar, P.-A. Cazenave, and F. Bonhomme. 1996. 'Origin and Radiation of the House Mouse: Clues from Nuclear Genes'. *Journal of Evolutionary Biology* 9 (5): 519–39. https://doi.org/10.1046/j.1420-9101.1996.9050519.x.

Dobzhansky, Theodosius, and Stephen Jay Gould. 1982. *Genetics and the Origin of Species*. New York: Columbia University Press.

Dod, Barbara, Lars S. Jermiin, Pierre Boursot, Verne H. Chapman, J. Tönnes Nielsen, and François Bonhomme. 1993. 'Counterselection on Sex Chromosomes in the Mus Musculus European Hybrid Zone'. *Journal of Evolutionary Biology* 6 (4): 529–46. https://doi.org/10.1046/j.1420-9101.1993.6040529.x.

Dumont, Beth L., and Bret A. Payseur. 2011. 'Genetic Analysis of Genome-Scale Recombination Rate Evolution in House Mice'. *PLoS Genet* 7 (6): e1002116.

Duplantier, J.-M., A. Orth, J. Catalan, and F. Bonhomme. 2002. 'Evidence for a Mitochondrial Lineage Originating from the Arabian Peninsula in the Madagascar House Mouse (Mus Musculus)'. *Heredity* 89 (2): 154–58. https://doi.org/10.1038/sj.hdy.6800122.

Duranton, Maud, and John Pool. 2021. 'Interactions between Natural Selection and Recombination Shape the Genomic Landscape of Introgression'. Preprint. Evolutionary Biology. https://doi.org/10.1101/2021.10.12.462572.

Duvaux, Ludovic, Khalid Belkhir, Matthieu Boulesteix, and Pierre Boursot. 2011. 'Isolation and Gene Flow: Inferring the Speciation History of European House Mice: HOUSE MOUSE SPECIATION HISTORY'. *Molecular Ecology* 20 (24): 5248–64. https://doi.org/10.1111/j.1365-294X.2011.05343.x.

Duvernell, David D., Eric Westhafer, and Jacob F. Schaefer. 2019. 'Late Pleistocene Range Expansion of North American Topminnows Accompanied by Admixture and Introgression'. *Journal of Biogeography* 46 (9): 2126–40. https://doi.org/10.1111/jbi.13597.

Edmonds, Christopher A., Anita S. Lillie, and L. Luca Cavalli-Sforza. 2004. 'Mutations Arising in the Wave Front of an Expanding Population'. *Proceedings of the National Academy of Sciences* 101 (4): 975–79. https://doi.org/10.1073/pnas.0308064100.

El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R Eddy, Aurélien Luciani, Simon C Potter, Matloob Qureshi, et al. 2019. 'The Pfam Protein Families Database in 2019'. *Nucleic Acids Research* 47 (D1): D427–32. https://doi.org/10.1093/nar/gky995.

Elgvin, Tore O., Cassandra N. Trier, Ole K. Tørresen, Ingerid J. Hagen, Sigbjørn Lien, Alexander J. Nederbragt, Mark Ravinet, Henrik Jensen, and Glenn-Peter Sætre. 2017. 'The Genomic Mosaicism of Hybrid Speciation'. *Science Advances* 3 (6): e1602996. https://doi.org/10.1126/sciadv.1602996.

Enk, Jacob, Alison Devault, Regis Debruyne, Christine E. King, Todd Treangen, Dennis O'Rourke, Steven L. Salzberg, Daniel Fisher, Ross MacPhee, and Hendrik Poinar. 2011. 'Complete Columbian Mammoth Mitogenome Suggests Interbreeding with Woolly Mammoths'. *Genome Biology* 12 (5): R51. https://doi.org/10.1186/gb-2011-12-5-r51.

Enk, Jacob, Alison Devault, Christopher Widga, Jeffrey Saunders, Paul Szpak, John Southon, Jean-Marie Rouillard, et al. 2016. 'Mammuthus Population Dynamics in Late Pleistocene North America: Divergence, Phylogeography, and Introgression'. *Frontiers in Ecology and Evolution* 4: 42. https://doi.org/10.3389/fevo.2016.00042.

Estonba, Andone, Ainhoa Solís, M. Iriondo, María J. Sanz-Martín, G. Pérez-Suárez, G. Markov, and F. Palacios. 2006. 'The Genetic Distinctiveness of the Three Iberian Hare Species: Lepus Europaeus, L. Granatensis, and L. Castroviejoi'.

Mammalian Biology 71 (1): 52–59. https://doi.org/10.1016/j.mambio.2005.08.010.

Excoffier, Laurent, Matthieu Foll, and Rémy J. Petit. 2009. 'Genetic Consequences of Range Expansions'. *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414.

Excoffier, Laurent, and Nicolas Ray. 2008. 'Surfing during Population Expansions Promotes Genetic Revolutions and Structuration'. *Trends in Ecology & Evolution* 23 (7): 347–51. https://doi.org/10.1016/j.tree.2008.04.004.

Ferreira, Mafalda S, Matthew R Jones, Colin M Callahan, Liliana Farelo, Zelalem Tolesa, Franz Suchentrunk, Pierre Boursot, et al. 2021. 'The Legacy of Recurrent Introgression during the Radiation of Hares'. *Systematic Biology* 70 (3): 593–607. https://doi.org/10.1093/sysbio/syaa088.

Finn, Robert D., Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Andreas Heger, et al. 2014. 'Pfam: The Protein Families Database'. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt1223.

Finn, Robert D., Jody Clements, and Sean R. Eddy. 2011. 'HMMER Web Server: Interactive Sequence Similarity Searching'. *Nucleic Acids Research* 39 (SUPPL. 2). https://doi.org/10.1093/nar/gkr367.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. The Genetical Theory of Natural Selection. Oxford, England: Clarendon Press. https://doi.org/10.5962/bhl.title.27468.

Fishman, Lila, and John H. Willis. 2006. 'A CYTONUCLEAR INCOMPATIBILITY CAUSES ANTHER STERILITY IN MIMULUS HYBRIDS'. *Evolution* 60 (7): 1372–81. https://doi.org/10.1554/05-708.1.

Fontaine, Michael C., James B. Pease, Aaron Steele, Robert M. Waterhouse, Daniel E. Neafsey, Igor V. Sharakhov, Xiaofang Jiang, et al. 2015. 'Extensive Introgression in a Malaria Vector Species Complex Revealed by Phylogenomics'. *Science (New York, N.Y.)* 347 (6217): 1258524. https://doi.org/10.1126/science.1258524.

Forejt, J., and P. Iványi. 1974. 'Genetic Studies on Male Sterility of Hybrids between Laboratory and Wild Mice (Mus Musculus L.)'. *Genetics Research* 24 (2): 189–206. https://doi.org/10.1017/S0016672300015214.

Fraïsse, Christelle, P. Alexander Gunnarsson, Denis Roze, Nicolas Bierne, and John J. Welch. 2016. 'The Genetics of Speciation: Insights from Fisher's Geometric Model'. *Evolution* 70 (7): 1450–64. https://doi.org/10.1111/evo.12968.

Frank, Steven A., and Bernard J. Crespi. 2011. 'Pathology from Evolutionary Conflict, with a Theory of X Chromosome versus Autosome Conflict over Sexually Antagonistic Traits'. *Proceedings of the National Academy of Sciences* 108 (Supplement 2): 10886–93. https://doi.org/10.1073/pnas.1100921108.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012a. 'CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data'. *Bioinformatics* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

———. 2012b. 'CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data'. *Bioinformatics* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

Fujiwara, Kazumichi, Yosuke Kawai, Kazuo Moriwaki, Toyoyuki Takada, Toshihiko Shiroishi, Naruya Saitou, Hitoshi Suzuki, and Naoki Osada. 2021. 'Insights into Mus Musculus Subspecies Population Structure across Eurasia Revealed by Whole-Genome Sequence Analysis'. https://doi.org/10.1101/2021.02.05.429881.

Gavrilets, Sergey. 2003. 'Perspective: Models of Speciation: What Have We Learned in 40 Years?' *Evolution* 57 (10): 2197–2215. https://doi.org/10.1111/j.0014-3820.2003.tb00233.x.

Gayral, Philippe, José Melo-Ferreira, Sylvain Glémin, Nicolas Bierne, Miguel Carneiro, Benoit Nabholz, Joao M. Lourenco, et al. 2013. 'Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate-Invertebrate Gap'. *PLoS Genetics* 9 (4). https://doi.org/10.1371/journal.pgen.1003457.

Gene Ontology Consortium. 2000. 'The Gene Ontology Consortium. Gene Ontology: Tool for the Unification of Biology'. *Nature Genetics* 25 (1): 25–29. https://doi.org/10.1038/75556.

Geraldes, A., P. Basset, K. L. Smith, and M. W. Nachman. 2011. 'Higher Differentiation among Subspecies of the House Mouse (Mus Musculus) in Genomic Regions with Low Recombination: RECOMBINATION AND SPECIATION IN MICE'. *Molecular Ecology* 20 (22): 4722–36. https://doi.org/10.1111/j.1365-294X.2011.05285.x.

Geraldes, Armando, Patrick Basset, Barbara Gibson, Kimberly L. Smith, Bettina Harr, Hon-Tsen Yu, Nina Bulatova, Yaron Ziv, and Michael W. Nachman. 2008. 'Inferring the History of Speciation in House Mice from Autosomal, X-Linked, Y-Linked and Mitochondrial Genes'. *Molecular Ecology* 17 (24): 5349–63. https://doi.org/10.1111/j.1365-294X.2008.04005.x.

Giska, Iwona, Liliana Farelo, João Pimenta, Fernando A. Seixas, Mafalda S. Ferreira, João P. Marques, Inês Miranda, et al. 2019. 'Introgression Drives Repeated Evolution of Winter Coat Color Polymorphism in Hares'. *Proceedings of the National Academy of Sciences*, November. https://doi.org/10.1073/pnas.1910471116.

Giska, Iwona, João Pimenta, Liliana Farelo, Pierre Boursot, Klaus Hackländer, Hannes Jenny, Neil Reid, et al. 2022. 'The Evolutionary Pathways for Local Adaptation in Mountain Hares'. *Molecular Ecology* n/a (n/a). https://doi.org/10.1111/mec.16338.

Good, Jeffrey M., Matthew D. Dean, and Michael W. Nachman. 2008. 'A Complex Genetic Basis to X-Linked Hybrid Male Sterility Between Two Species of House Mice'. *Genetics* 179 (4): 2213–28. https://doi.org/10.1534/genetics.107.085340.

Good, Jeffrey M., Thomas Giger, Matthew D. Dean, and Michael W. Nachman. 2010. 'Widespread Over-Expression of the X Chromosome in Sterile F1 Hybrid Mice'. *PLOS Genetics* 6 (9): e1001148. https://doi.org/10.1371/journal.pgen.1001148.

Gow, J. L., C. L. Peichel, and E. B. Taylor. 2007. 'Ecological Selection against Hybrids in Natural Populations of Sympatric Threespine Sticklebacks'. *Journal of Evolutionary Biology* 20 (6): 2173–80. https://doi.org/10.1111/j.1420-9101.2007.01427.x.

Grabherr, Manfred G, Brian J Haas, Moran Yassour, Joshua Z Levin, Dawn A Thompson, Ido Amit, Xian Adiconis, et al. 2011a. 'Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome.' *Nature Biotechnology* 29 (7): 644–52. https://doi.org/10.1038/nbt.1883.

———. 2011b. 'Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome'. *Nature Biotechnology* 29 (May): 644.

Guénet, Jean-Louis, and François Bonhomme. 2003. 'Wild Mice: An Ever-Increasing Contribution to a Popular Mammalian Model'. *Trends in Genetics* 19 (1): 24–31. https://doi.org/10.1016/S0168-9525(02)00007-0.

Haas & Papanicolaou et al. (2015) 2019. *TransDecoder Source. Contribute to TransDecoder/TransDecoder Development by Creating an Account on GitHub*. Perl. TransDecoder. https://github.com/TransDecoder/TransDecoder.

Haas, Brian J, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. 'De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for

Reference Generation and Analysis.' *Nature Protocols* 8 (8): 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hamid, Haddadian Shad, Jamshid Darvish, Eskandar Rastegar-Pouyani, and Ahmad Mahmoudi. 2017. 'Subspecies Differentiation of the House Mouse Mus Musculus Linnaeus, 1758 in the Center and East of the Iranian Plateau and Afghanistan'. *Mammalia* 81 (2). https://doi.org/10.1515/mammalia-2015-0041.

Hamill, R M, D Doyle, and E J Duke. 2006. 'Spatial Patterns of Genetic Diversity across European Subspecies of the Mountain Hare, Lepus Timidus L.' *Heredity* 97 (5): 355–65. https://doi.org/10.1038/sj.hdy.6800880.

Hardouin, Emilie A, Annie Orth, Meike Teschke, Jamshid Darvish, Diethard Tautz, and François Bonhomme. 2015. 'Eurasian House Mouse (Mus Musculus L.) Differentiation at Microsatellite Loci Identifies the Iranian Plateau as a Phylogeographic Hotspot'. *BMC Evolutionary Biology* 15 (February). https://doi.org/10.1186/s12862-015-0306-4.

Harr, Bettina, Emre Karakoc, Rafik Neme, Meike Teschke, Christine Pfeifle, Željka Pezer, Hiba Babiker, et al. 2016. 'Genomic Resources for Wild Populations of the House Mouse, Mus Musculus and Its Close Relative Mus Spretus'. *Scientific Data* 3 (1): 1–14. https://doi.org/10.1038/sdata.2016.75.

Harris, Kelley, and Rasmus Nielsen. 2016. 'The Genetic Cost of Neanderthal Introgression'. *Genetics* 203 (2): 881–91. https://doi.org/10.1534/genetics.116.186890.

Harrison, D. L. 1970. 'SOME OBSERVATIONS ON THE STATUS OF THE SOUTH ARABIAN HOUSE MOUSE MUS MUSCULUS GENTILULUS THOMAS, 1919. (RODENTIA : MURIDAE)' 34 (2): 244–47. https://doi.org/10.1515/mamm.1970.34.2.244.

Hermansen, Jo S., Stein A. Sæther, Tore O. Elgvin, Thomas Borge, Elin Hjelle, and Glenn-Peter Sætre. 2011. 'Hybrid Speciation in Sparrows I: Phenotypic Intermediacy, Genetic Admixture and Barriers to Gene Flow'. *Molecular Ecology* 20 (18): 3812–22. https://doi.org/10.1111/j.1365-294X.2011.05183.x.

Hessenauer, Pauline, Anna Fijarczyk, Hélène Martin, Julien Prunier, Guillaume Charron, Jérôme Chapuis, Louis Bernier, Philippe Tanguay, Richard C. Hamelin, and Christian R. Landry. 2020. 'Hybridization and Introgression Drive Genome Evolution of Dutch Elm Disease Pathogens'. *Nature Ecology & Evolution* 4 (4): 626–38. https://doi.org/10.1038/s41559-020-1133-6.

Huerta-Sánchez, Emilia, Xin Jin, Asan, Zhuoma Bianba, Benjamin M. Peter, Nicolas Vinckenbosch, Yu Liang, et al. 2014. 'Altitude Adaptation in Tibetans Caused by Introgression of Denisovan-like DNA'. *Nature* 512 (7513): 194–97. https://doi.org/10.1038/nature13408.

Hughes, Maria, Neil Reid, Ian Montgomery, and Paulo Prodoehl. 2011. 'Verification of Hybridisation between Introduced European and Native Irish Hares'. *Northern Ireland Environment Agency Research and Development SeriesNorthern Ireland Environment Agency Research and Development Series* 11–11.

Hui, Ling, Jing Lu, Yibing Han, and Stephen H. Pilder. 2006. 'The Mouse T Complex Gene Tsga2, Encoding Polypeptides Located in the Sperm Tail and Anterior Acrosome, Maps to a Locus Associated with Sperm Motility and Sperm-Egg Interaction Abnormalities'. *Biology of Reproduction* 74 (4): 633–43. https://doi.org/10.1095/biolreprod.105.045963.

IUCN. 2016. 'IUCN Red List of Threatened Species'. *Version 2016-1*, www.iucnredlist.org.

Janoušek, Václav, Pavel Munclinger, Liuyang Wang, Katherine C. Teeter, and Priscilla K. Tucker. 2015a. 'Functional Organization of the Genome May Shape the Species Boundary in the House Mouse'. *Molecular Biology and Evolution* 32 (5): 1208–20. https://doi.org/10.1093/molbev/msv011.

———. 2015b. ‘Functional Organization of the Genome May Shape the Species Boundary in the House Mouse’. *Molecular Biology and Evolution* 32 (5): 1208–20. https://doi.org/10.1093/molbev/msv011.

Janoušek, Václav, LIUYANG WANG, KEN LUZYNSKI, PETRA DUFKOVÁ, MARTINA M. VYSKOČILOVÁ, MICHAEL W. NACHMAN, PAVEL MUNCLINGER, MILOŠ MACHOLÁN, JAROSLAV PIÁLEK, and PRISCILLA K. TUCKER. 2012. ‘Genome-Wide Architecture of Reproductive Isolation in a Naturally Occurring Hybrid Zone between Mus Musculus Musculus and M. m. Domesticus’. *Molecular Ecology* 21 (12): 10.1111/j.1365-294X.2012.05583.x. https://doi.org/10.1111/j.1365-294X.2012.05583.x.

Jones, Eleanor P., JEROEN Van Der Kooij, Roar Solheim, and Jeremy B. Searle. 2010. ‘Norwegian House Mice (Mus Musculus Musculus/Domesticus): Distributions, Routes of Colonization and Patterns of Hybridization’. *Molecular Ecology* 19 (23): 5252–64. https://doi.org/10.1111/j.1365-294X.2010.04874.x.

Jones, Matthew R., L. Scott Mills, Paulo Célio Alves, Colin M. Callahan, Joel M. Alves, Diana J. R. Lafferty, Francis M. Jiggins, Jeffrey D. Jensen, José Melo-Ferreira, and Jeffrey M. Good. 2018. ‘Adaptive Introgression Underlies Polymorphic Seasonal Camouflage in Snowshoe Hares’. *Science* 360 (6395): 1355–58. https://doi.org/10.1126/science.aar5273.

Juric, Ivan, Simon Aeschbacher, and Graham Coop. 2016. ‘The Strength of Selection against Neanderthal Introgression’. *PLOS Genetics* 12 (11): e1006340. https://doi.org/10.1371/journal.pgen.1006340.

Jurka, Jerzy, Weidong Bao, and Kenji K Kojima. 2011. ‘Families of Transposable Elements, Population Structure and the Origin of Species’. *Biology Direct* 6 (1): 44. https://doi.org/10.1186/1745-6150-6-44.

Kanehisa, Minoru, Susumu Goto, Yoko Sato, Miho Furumichi, and Mao Tanabe. 2012. ‘KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets’. *Nucleic Acids Research* 40 (D1). https://doi.org/10.1093/nar/gkr988.

Khurana, Jaspreet S., Jie Wang, Jia Xu, Birgit S. Koppetsch, Travis C. Thomson, Anetta Nowosielska, Chengjian Li, Phillip D. Zamore, Zhiping Weng, and William E. Theurkauf. 2011. ‘Adaptation to P Element Transposon Invasion in Drosophila Melanogaster’. *Cell* 147 (7): 1551–63. https://doi.org/10.1016/j.cell.2011.11.042.

Kirkpatrick, Mark, and Virginie Ravigné. 2002. ‘Speciation by Natural and Sexual Selection: Models and Experiments.’ *The American Naturalist* 159 (S3): S22–35. https://doi.org/10.1086/338370.

Klopfstein, Seraina, Mathias Currat, and Laurent Excoffier. 2006. ‘The Fate of Mutations Surfing on the Wave of a Range Expansion’. *Molecular Biology and Evolution* 23 (3): 482–90. https://doi.org/10.1093/molbev/msj057.

Krogh, a, B Larsson, G von Heijne, and E L Sonnhammer. 2001. ‘Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes.’ *Journal of Molecular Biology* 305 (3): 567–80. https://doi.org/10.1006/jmbi.2000.4315.

Kunte, Krushnamegh, Cristina Shea, Matthew L. Aardema, J. Mark Scriber, Thomas E. Juenger, Lawrence E. Gilbert, and Marcus R. Kronforst. 2011. ‘Sex Chromosome Mosaicism and Hybrid Speciation among Tiger Swallowtail Butterflies’. *PLOS Genetics* 7 (9): e1002274. https://doi.org/10.1371/journal.pgen.1002274.

Langdon, Quinn K., David Peris, Juan I. Eizaguirre, Dana A. Opulente, Kelly V. Buh, Kayla Sylvester, Martin Jarzyna, et al. 2020. ‘Postglacial Migration Shaped the Genomic Diversity and Global Distribution of the Wild Ancestor of Lager-Brewing Hybrids’. *PLOS Genetics* 16 (4): e1008680. https://doi.org/10.1371/journal.pgen.1008680.

Larson, Erica L., Dan Vanderpool, Brice A. J. Sarver, Colin Callahan, Sara Keeble, Lorraine P. Provencio, Michael D. Kessler, et al. 2018. ‘The Evolution of

Polymorphic Hybrid Incompatibilities in House Mice'. *Genetics*, April, genetics.300840.2018. https://doi.org/10.1534/genetics.118.300840.

Lawson, Daniel J., Lucy van Dorp, and Daniel Falush. 2018. 'A Tutorial on How Not to Over-Interpret STRUCTURE and ADMIXTURE Bar Plots'. *Nature Communications* 9 (1): 3258. https://doi.org/10.1038/s41467-018-05257-7.

Leach, Katie, Ruth Kelly, Alison Cameron, W. Ian Montgomery, and Neil Reid. 2015. 'Expertly Validated Models and Phylogenetically-Controlled Analysis Suggests Responses to Climate Change Are Related to Species Traits in the Order Lagomorpha'. *PLoS ONE* 10 (4). https://doi.org/10.1371/journal.pone.0122267.

Leach, Katie, W. Ian Montgomery, and Neil Reid. 2015. 'Biogeography, Macroecology and Species' Traits Mediate Competitive Interactions in the Order Lagomorpha'. *Mammal Review*. https://doi.org/10.1111/mam.12035.

Levänen, Riikka, Carl-Gustaf Thulin, Göran Spong, and Jaakko L. O. Pohjoismäki. 2018. 'Widespread Introgression of Mountain Hare Genes into Fennoscandian Brown Hare Populations'. *PLOS ONE* 13 (1): e0191790. https://doi.org/10.1371/journal.pone.0191790.

Li, Heng, and Richard Durbin. 2010. 'Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform'. *Bioinformatics* 26 (5): 589–95. https://doi.org/10.1093/bioinformatics/btp698.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Lindholm, Anna K, Kerstin Musolf, Andrea Weidt, and Barbara König. 2013. 'Mate Choice for Genetic Compatibility in the House Mouse'. *Ecology and Evolution* 3 (5): 1231–47. https://doi.org/10.1002/ece3.534.

Liu, Kevin J., Ethan Steinberg, Alexander Yozzo, Ying Song, Michael H. Kohn, and Luay Nakhleh. 2015. 'Interspecific Introgressive Origin of Genomic Diversity in the House Mouse'. *Proceedings of the National Academy of Sciences* 112 (1): 196–201. https://doi.org/10.1073/pnas.1406298111.

Macholán, Miloš, Stuart J. E. Baird, Alena Fornůsková, Iva Martincová, Pavel Rubík, Ľudovít Ďureje, Emanuel Heitlinger, and Jaroslav Piálek. 2019. 'Widespread Introgression of the *Mus Musculus Musculus* Y Chromosome in Central Europe'. Preprint. Evolutionary Biology. https://doi.org/10.1101/2019.12.23.887471.

Macholán, Miloš, Stuart J. E. Baird, Pavel Munclinger, and Jaroslav Piálek. 2012. *Evolution of the House Mouse*. Cambridge University Press.

Macholán, Miloš, Stuart JE Baird, Pavel Munclinger, Petra Dufková, Barbora Bímová, and Jaroslav Piálek. 2008a. 'Genetic Conflict Outweighs Heterogametic Incompatibility in the Mouse Hybrid Zone?' *BMC Evolutionary Biology* 8 (1): 271. https://doi.org/10.1186/1471-2148-8-271.

———. 2008b. 'Genetic Conflict Outweighs Heterogametic Incompatibility in the Mouse Hybrid Zone?' *BMC Evolutionary Biology* 8 (1): 271. https://doi.org/10.1186/1471-2148-8-271.

Macholán, Miloš, Boris Kryštufek, and Vladimír Vohralík. 2003. 'The Location of TheMus Musculus/M. Domesticus Hybrid Zone in the Balkans: Clues from Morphology'. *Acta Theriologica* 48 (2): 177–88. https://doi.org/10.1007/BF03194157.

Maheshwari, Shamoni, and Daniel A. Barbash. 2011. 'The Genetics of Hybrid Incompatibilities'. *Annual Review of Genetics* 45 (1): 331–55. https://doi.org/10.1146/annurev-genet-110410-132514.

Mallet, James, Margarita Beltrán, Walter Neukirchen, and Mauricio Linares. 2007. 'Natural Hybridization in Heliconiine Butterflies: The Species Boundary as a Continuum'. *BMC Evolutionary Biology* 7 (1): 28. https://doi.org/10.1186/1471-2148-7-28.

Mallet, James, Nora Besansky, and Matthew W. Hahn. 2016. 'How Reticulated Are Species?' *BioEssays* 38 (2): 140–49. https://doi.org/10.1002/bies.201500149.

Mallet, James, and Mathieu Joron. 1999. 'Evolution of Diversity in Warning Color and Mimicry: Polymorphisms, Shifting Balance, and Speciation'. *Annual Review of Ecology and Systematics* 30: 201–33.

Marques, Joao P, Liliana Farelo, Joana Vilela, Dan Vanderpool, Paulo C Alves, Jeffrey M Good, Pierre Boursot, and Jose Melo-Ferreira. 2017. 'Range Expansion Underlies Historical Introgressive Hybridization in the Iberian Hare.' *Scientific Reports* 7 (January): 40788. https://doi.org/10.1038/srep40788.

Marques, João P., Liliana Farelo, Joana Vilela, Dan Vanderpool, Paulo C. Alves, Jeffrey M. Good, Pierre Boursot, and José Melo-Ferreira. 2017. 'Range Expansion Underlies Historical Introgressive Hybridization in the Iberian Hare'. *Scientific Reports* 7 (1): 40788. https://doi.org/10.1038/srep40788.

Marques, João P., Mafalda S. Ferreira, Liliana Farelo, Colin M. Callahan, Klaus Hackländer, Hannes Jenny, W. Ian Montgomery, et al. 2017. 'Mountain Hare Transcriptome and Diagnostic Markers as Resources to Monitor Hybridization with European Hares'. *Scientific Data* 4 (1): 170178. https://doi.org/10.1038/sdata.2017.178.

Marques, Joao P, Mafalda S Ferreira, Liliana Farelo, Colin M Callahan, Klaus Hacklander, Hannes Jenny, W Ian Montgomery, et al. 2017. 'Mountain Hare Transcriptome and Diagnostic Markers as Resources to Monitor Hybridization with European Hares.' *Scientific Data* 4 (December): 170178–170178. https://doi.org/10.1038/sdata.2017.178.

Marques, João P., Fernando A. Seixas, Liliana Farelo, Colin M. Callahan, Jeffrey M. Good, W. Ian Montgomery, Neil Reid, Paulo C. Alves, Pierre Boursot, and José Melo-Ferreira. 2020. 'An Annotated Draft Genome of the Mountain Hare (Lepus Timidus)'. *Genome Biology and Evolution* 12 (1): 3656–62. https://doi.org/10.1093/gbe/evz273.

Martienssen, Robert A. 2010. 'Heterochromatin, Small RNA and Post-Fertilization Dysgenesis in Allopolyploid and Interploid Hybrids of Arabidopsis'. *The New Phytologist* 186 (1): 46–53. https://doi.org/10.1111/j.1469-8137.2010.03193.x.

Martin, Simon H., Kanchon K. Dasmahapatra, Nicola J. Nadeau, Camilo Salazar, James R. Walters, Fraser Simpson, Mark Blaxter, Andrea Manica, James Mallet, and Chris D. Jiggins. 2013. 'Genome-Wide Evidence for Speciation with Gene Flow in Heliconius Butterflies'. *Genome Research* 23 (11): 1817–28. https://doi.org/10.1101/gr.159426.113.

Martin, Simon H., John W. Davey, Camilo Salazar, and Chris D. Jiggins. 2019. 'Recombination Rate Variation Shapes Barriers to Introgression across Butterfly Genomes'. *PLOS Biology* 17 (2): e2006288. https://doi.org/10.1371/journal.pbio.2006288.

Martincová, Iva, Ľudovít Ďureje, Jakub Kreisinger, Miloš Macholán, and Jaroslav Piálek. 2019. 'Phenotypic Effects of the Y Chromosome Are Variable and Structured in Hybrids among House Mouse Recombinant Lines'. *Ecology and Evolution*, May, ece3.5196. https://doi.org/10.1002/ece3.5196.

Masly, John P., and Daven C. Presgraves. 2007. 'High-Resolution Genome-Wide Dissection of the Two Rules of Speciation in Drosophila'. *PLOS Biology* 5 (9): e243. https://doi.org/10.1371/journal.pbio.0050243.

Matthee, Conrad A, Bettine Jansen Van Vuuren, Diana Bell, and Terence J Robinson. 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): 433–47. https://doi.org/10.1080/10635150490445715.

Maxwell, Colin S., Victoria E. Sepulveda, David A. Turissini, William E. Goldman, and Daniel R. Matute. 2018. 'Recent Admixture between Species of the Fungal Pathogen Histoplasma'. *Evolution Letters* 2 (3): 210–20. https://doi.org/10.1002/evl3.59.

Mayr, Ernst. 1942. 'Systematics And The Origin Of Species From The Viewpoint Of A Zoologist'. 1942.

———. 1963. *Animal Species and Evolution*. *Animal Species and Evolution*. Harvard University Press. https://www.degruyter.com/document/doi/10.4159/harvard.9780674865327/html.

McDermott, Shannon R., and Mohamed A. F. Noor. 2010. 'The Role of Meiotic Drive in Hybrid Male Sterility'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1544): 1265–72. https://doi.org/10.1098/rstb.2009.0264.

McKenna, A. H., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, et al. 2010. 'The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data'. *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

McLaughlin, Richard N., and Harmit S. Malik. 2017. 'Genetic Conflicts: The Usual Suspects and Beyond'. Edited by Joel D. Levine, Daniel J. C. Kronauer, and Michael H. Dickinson. *Journal of Experimental Biology* 220 (1): 6–17. https://doi.org/10.1242/jeb.148148.

Meier, Joana I., David A. Marques, Salome Mwaiko, Catherine E. Wagner, Laurent Excoffier, and Ole Seehausen. 2017. 'Ancient Hybridization Fuels Rapid Cichlid Fish Adaptive Radiations'. *Nature Communications* 8 (1): 14363. https://doi.org/10.1038/ncomms14363.

Melo, Maria C., Alicia Grealy, Beth Brittain, Greg M. Walter, and Daniel Ortiz-Barrientos. 2014. 'Strong Extrinsic Reproductive Isolation between Parapatric Populations of an Australian Groundsel'. *New Phytologist* 203 (1): 323–34. https://doi.org/10.1111/nph.12779.

Melo-Ferreira, J., P. C. Alves, H. Freitas, N. Ferrand, and P. Boursot. 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): 2643–58. https://doi.org/10.1111/j.1365-294X.2009.04221.x.

Melo-Ferreira, J., P. Boursot, M. Carneiro, P. J. Esteves, L. Farelo, and P. C. Alves. 2012a. 'Recurrent Introgression of Mitochondrial DNA among Hares (Lepus Spp.) Revealed by Species-Tree Inference and Coalescent Simulations'. *Systematic Biology* 61 (3): 367–81. https://doi.org/10.1093/sysbio/syr114.

———. 2012b. 'Recurrent Introgression of Mitochondrial DNA Among Hares (Lepus Spp.) Revealed by Species-Tree Inference and Coalescent Simulations'. *Systematic Biology* 61 (3): 367–367. https://doi.org/10.1093/sysbio/syr114.

Melo-Ferreira, J., P. Boursot, E. Randi, A. Kryukov, F. Suchentrunk, N. Ferrand, and P. C. Alves. 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): 605–18. https://doi.org/10.1111/j.1365-294X.2006.03166.x.

Melo-Ferreira, J., P. Boursot, F. Suchentrunk, N. Ferrand, and P. C. Alves. 2005. 'Invasion from the Cold Past: Extensive Introgression of Mountain Hare (Lepus Timidus) Mitochondrial DNA into Three Other Hare Species in Northern Iberia'. *Molecular Ecology* 14 (8): 2459–64. https://doi.org/10.1111/j.1365-294X.2005.02599.x.

Melo-Ferreira, J, L Farelo, H Freitas, F Suchentrunk, P Boursot, and P C Alves. 2014a. 'Home-Loving Boreal Hare Mitochondria Survived Several Invasions in Iberia:

The Relative Roles of Recurrent Hybridisation and Allele Surfing.' *Heredity* 112 (3): 265–73. https://doi.org/10.1038/hdy.2013.102.

———. 2014b. 'Home-Loving Boreal Hare Mitochondria Survived Several Invasions in Iberia: The Relative Roles of Recurrent Hybridisation and Allele Surfing'. *Heredity* 112 (3): 265–73. https://doi.org/10.1038/hdy.2013.102.

Metcalfe, Cushla J., Kira V. Bulazel, Gianni C. Ferreri, Elizabeth Schroeder-Reiter, Gerhard Wanner, Willem Rens, Craig Obergfell, Mark D. B. Eldridge, and Rachel J. O'Neill. 2007. 'Genomic Instability Within Centromeres of Interspecific Marsupial Hybrids'. *Genetics* 177 (4): 2507–17. https://doi.org/10.1534/genetics.107.082313.

Mihola, Ondrej, Vladimir Landa, Florencia Pratto, Kevin Brick, Tatyana Kobets, Fitore Kusari, Srdjan Gasic, et al. 2021. 'Rat PRDM9 Shapes Recombination Landscapes, Duration of Meiosis, Gametogenesis, and Age of Fertility'. *BMC Biology* 19 (1): 86. https://doi.org/10.1186/s12915-021-01017-0.

Mihola, Ondrej, Zdenek Trachtulec, Cestmir Vlcek, John C. Schimenti, and Jiri Forejt. 2009. 'A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase'. *Science* 323 (5912): 373–75. https://doi.org/10.1126/science.1163601.

Moran, Benjamin M, Cheyenne Payne, Quinn Langdon, Daniel L Powell, Yaniv Brandvain, and Molly Schumer. 2021. 'The Genomic Consequences of Hybridization'. Edited by Patricia J Wittkopp. *ELife* 10 (August): e69016. https://doi.org/10.7554/eLife.69016.

Muller, H. 1942. 'Isolating Mechanisms, Evolution, and Temperature'. *Biol. Symp.* 6: 71–125.

Nachman, Michael W., and Bret A. Payseur. 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): 409–21. https://doi.org/10.1098/rstb.2011.0249.

Nielsen, Rasmus, Joshua M. Akey, Mattias Jakobsson, Jonathan K. Pritchard, Sarah Tishkoff, and Eske Willerslev. 2017. 'Tracing the Peopling of the World through Genomics'. *Nature* 541 (7637): 302–10. https://doi.org/10.1038/nature21347.

Norris, Laura C., Bradley J. Main, Yoosook Lee, Travis C. Collier, Abdrahamane Fofana, Anthony J. Cornel, and Gregory C. Lanzaro. 2015. 'Adaptive Introgression in an African Malaria Mosquito Coincident with the Increased Usage of Insecticide-Treated Bed Nets'. *Proceedings of the National Academy of Sciences* 112 (3): 815–20. https://doi.org/10.1073/pnas.1418892112.

Nosil, Patrik, and Jeffrey L. Feder. 2012. 'Genomic Divergence during Speciation: Causes and Consequences'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1587): 332–42. https://doi.org/10.1098/rstb.2011.0263.

Nosil, Patrik, Daniel J. Funk, and Daniel Ortiz-Barrientos. 2009. 'Divergent Selection and Heterogeneous Genomic Divergence'. *Molecular Ecology* 18 (3): 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x.

Nosil, Patrik, Luke J. Harmon, and Ole Seehausen. 2009. 'Ecological Explanations for (Incomplete) Speciation'. *Trends in Ecology & Evolution* 24 (3): 145–56. https://doi.org/10.1016/j.tree.2008.10.011.

Nosil, Patrik, and Dolph Schluter. 2011. 'The Genes Underlying the Process of Speciation'. *Trends in Ecology & Evolution* 26 (4): 160–67. https://doi.org/10.1016/j.tree.2011.01.001.

Orr, H. A. 1995. 'The Population Genetics of Speciation: The Evolution of Hybrid Incompatibilities'. *Genetics* 139 (4): 1805–13.

Patel, Maulik R, Ganesh K Miriyala, Aimee J Littleton, Heiko Yang, Kien Trinh, Janet M Young, Scott R Kennedy, Yukiko M Yamashita, Leo J Pallanck, and Harmit S Malik. 2016. 'A Mitochondrial DNA Hypomorph of Cytochrome Oxidase

Specifically Impairs Male Fertility in Drosophila Melanogaster'. *ELife* 5: e16923. https://doi.org/10.7554/eLife.16923.

Payseur, Bret A., and Loren H. Rieseberg. 2016. 'A Genomic Perspective on Hybridization and Speciation'. *Molecular Ecology* 25 (11): 2337–60. https://doi.org/10.1111/mec.13557.

Petersen, Thomas Nordahl, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2011. 'SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions'. *Nature Methods* 8 (10): 785–86. https://doi.org/10.1038/nmeth.1701.

Phifer-Rixey, Megan, Matthew Bomhoff, and Michael W. Nachman. 2014. 'Genome-Wide Patterns of Differentiation Among House Mouse Subspecies'. *Genetics* 198 (1): 283–97. https://doi.org/10.1534/genetics.114.166827.

Phifer-Rixey, Megan, Bettina Harr, and Jody Hey. 2020. 'Further Resolution of the House Mouse (Mus Musculus) Phylogeny by Integration over Isolation-with-Migration Histories'. *BMC Evolutionary Biology* 20 (1): 120. https://doi.org/10.1186/s12862-020-01666-9.

Phifer-Rixey, Megan, and Michael W Nachman. 2015. 'Insights into Mammalian Biology from the Wild House Mouse Mus Musculus'. *ELife* 4 (April): e05959. https://doi.org/10.7554/eLife.05959.

Pool, John E. 2015. 'The Mosaic Ancestry of the Drosophila Genetic Reference Panel and the D. Melanogaster Reference Genome Reveals a Network of Epistatic Fitness Interactions'. *Molecular Biology and Evolution* 32 (12): 3236–51. https://doi.org/10.1093/molbev/msv194.

Powell, Sean, Damian Szklarczyk, Kalliopi Trachana, Alexander Roth, Michael Kuhn, Jean Muller, Roland Arnold, et al. 2012. 'EggNOG v3.0: Orthologous Groups Covering 1133 Organisms at 41 Different Taxonomic Ranges'. *Nucleic Acids Research* 40 (D1). https://doi.org/10.1093/nar/gkr1060.

Prager, E M, C Orrego, and R D Sage. 1998. 'Genetic Variation and Phylogeography of Central Asian and Other House Mice, Including a Major New Mitochondrial Lineage in Yemen.' *Genetics* 150 (2): 835–61.

Presgraves, Daven. 2009. '12 - Drive and Sperm: The Evolution and Genetics of Male Meiotic Drive'. In *Sperm Biology*, edited by Tim R. Birkhead, David J. Hosken, and Scott Pitnick, 471–506. London: Academic Press. https://doi.org/10.1016/B978-0-12-372568-4.00012-4.

Presgraves, Daven C. 2010. 'The Molecular Evolutionary Basis of Species Formation'. *Nature Reviews Genetics* 11 (3): 175–80. https://doi.org/10.1038/nrg2718.

Price, Trevor D., and Michelle M. Bouvier. 2002. 'The Evolution of F1 Postzygotic Incompatibilities in Birds'. *Evolution* 56 (10): 2083–89. https://doi.org/10.1111/j.0014-3820.2002.tb00133.x.

Queiroz, Kevin de. 1998. 'The General Lineage Concept of Species, Species Chteria, and the Process of Speciation', 19.

Quilodrán, Claudio S., Alexandros Tsoupas, and Mathias Currat. 2020. 'The Spatial Signature of Introgression After a Biological Invasion With Hybridization'. *Frontiers in Ecology and Evolution* 8 (September): 569620. https://doi.org/10.3389/fevo.2020.569620.

Raufaste, Nathalie, Annie Orth, Khalid Belkhir, David Senet, Carole Smadja, Stuart Baird, François Bonhomme, Barbara Dod, and Pierre Boursot. 2005. 'Inferences of Selection and Migration in the Danish House Mouse Hybrid Zone'. *Biological Journal of the Linnean Society* 84 (3): 593–616. https://doi.org/10.1111/j.1095-8312.2005.00457.x.

Rebollo, Rita, Béatrice Horard, Benjamin Hubert, and Cristina Vieira. 2010. 'Jumping Genes and Epigenetics: Towards New Species'. *Gene* 454 (1): 1–7. https://doi.org/10.1016/j.gene.2010.01.003.

Reid, Neil. 2011. 'European Hare (Lepus Europaeus) Invasion Ecology: Implication for the Conservation of the Endemic Irish Hare (Lepus Timidus Hibernicus)'. *Biological Invasions* 13 (3): 559–69. https://doi.org/10.1007/s10530-010-9849-x.

Reid, Neil, and W. Ian Montgomery. 2007. 'Is Naturalisation of the Brown Hare in Ireland a Threat to the Endemic Irish Hare?' *Biology and Environment* 107 (3): 129–38. https://doi.org/10.3318/BIOE.2007.107.3.129.

Rice, William R., Sergey Gavrilets, and Urban Friberg. 2008. 'Sexually Antagonistic "Zygotic Drive" of the Sex Chromosomes'. *PLOS Genetics* 4 (12): e1000313. https://doi.org/10.1371/journal.pgen.1000313.

Rieseberg, L. H. 2003. 'Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization'. *Science* 301 (5637): 1211–16. https://doi.org/10.1126/science.1086949.

Rieseberg, Loren H., Margaret A. Archer, and Robert K. Wayne. 1999. 'Transgressive Segregation, Adaptation and Speciation'. *Heredity* 83 (4): 363–72. https://doi.org/10.1038/sj.hdy.6886170.

Rieseberg, Loren H., and Benjamin K. Blackman. 2010. 'Speciation Genes in Plants'. *Annals of Botany* 106 (3): 439–55. https://doi.org/10.1093/aob/mcq126.

Rieseberg, Loren H., and John H. Willis. 2007. 'Plant Speciation'. *Science (New York, N.Y.)* 317 (5840): 910–14. https://doi.org/10.1126/science.1137729.

Rosenberg, N A, T Burke, K Elo, M W Feldman, P J Freidlin, M A Groenen, J Hillel, et al. 2001. 'Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes from 20 Chicken Breeds.' *Genetics* 159 (2): 699–713.

Rosenberg, Noah A, Jonathan K Pritchard, James L Weber, Howard M Cann, Kenneth K Kidd, Lev A Zhivotovsky, and Marcus W Feldman. 2002. 'Genetic Structure of Human Populations' 298: 6.

Roux, Camille, Christelle Fraïsse, Jonathan Romiguier, Yoann Anciaux, Nicolas Galtier, and Nicolas Bierne. 2016. 'Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence'. Edited by Craig Moritz. *PLOS Biology* 14 (12): e2000234. https://doi.org/10.1371/journal.pbio.2000234.

Rundle, Howard D., and Patrik Nosil. 2005. 'Ecological Speciation'. *Ecology Letters* 8 (3): 336–52. https://doi.org/10.1111/j.1461-0248.2004.00715.x.

Sackton, Timothy B., Horacio Montenegro, Daniel L. Hartl, and Bernardo Lemos. 2011. 'Interspecific Y Chromosome Introgressions Disrupt Testis-Specific Gene Expression and Male Reproductive Phenotypes in Drosophila'. *Proceedings of the National Academy of Sciences* 108 (41): 17046–51. https://doi.org/10.1073/pnas.1114690108.

Sankararaman, Sriram, Swapan Mallick, Michael Dannemann, Kay Prüfer, Janet Kelso, Svante Pääbo, Nick Patterson, and David Reich. 2014. 'The Genomic Landscape of Neanderthal Ancestry in Present-Day Humans'. *Nature* 507 (7492): 354–57. https://doi.org/10.1038/nature12961.

Sanz-Martín, María J., Andone Estonba, Carmen Manzano, Mikel Iriondo, Gonzalo Pérez-Suárez, Alfonso Fernández, and Fernando Palacios. 2014. 'Genetic Structure of Brown and Iberian Hare Populations in Northern Iberia: Implications for Conservation of Genetic Diversity'. *The Journal of Wildlife Management* 78 (4): 632–44. https://doi.org/10.1002/jwmg.713.

Schluter, Dolph, and Associate Editor: Mark A. McPeek. 2000. 'Ecological Character Displacement in Adaptive Radiation'. *The American Naturalist* 156 (S4): S4–16. https://doi.org/10.1086/303412.

Schumer, Molly, Rongfeng Cui, Gil G. Rosenthal, and Peter Andolfatto. 2015. 'Reproductive Isolation of Hybrid Populations Driven by Genetic Incompatibilities'. Edited by Bret A. Payseur. *PLOS Genetics* 11 (3): e1005041. https://doi.org/10.1371/journal.pgen.1005041.

Schumer, Molly, Chenling Xu, Daniel L. Powell, Arun Durvasula, Laurits Skov, Chris Holland, John C. Blazier, et al. 2018. 'Natural Selection Interacts with Recombination to Shape the Evolution of Hybrid Genomes'. *Science* 360 (6389): 656–60. https://doi.org/10.1126/science.aar3684.

Schwarz, Ernst, and Henriette K. Schwarz. 1943. 'The Wild and Commensal Stocks of the House Mouse, Mus Musculus Linnaeus'. *Journal of Mammalogy* 24 (1): 59–72. https://doi.org/10.2307/1374781.

Scordato, Elizabeth S. C., Chris C. R. Smith, Georgy A. Semenov, Yu Liu, Matthew R. Wilkins, Wei Liang, Alexander Rubtsov, et al. 2020. 'Migratory Divides Coincide with Reproductive Barriers across Replicated Avian Hybrid Zones above the Tibetan Plateau'. *Ecology Letters* 23 (2): 231–41. https://doi.org/10.1111/ele.13420.

Ségurel, Laure, Ellen Miranda Leffler, and Molly Przeworski. 2011. 'The Case of the Fickle Fingers: How the PRDM9 Zinc Finger Protein Specifies Meiotic Recombination Hotspots in Humans'. *PLOS Biology* 9 (12): e1001211. https://doi.org/10.1371/journal.pbio.1001211.

Seixas, Fernando A., Pierre Boursot, and José Melo-Ferreira. 2018. 'The Genomic Impact of Historical Hybridization with Massive Mitochondrial DNA Introgression'. *Genome Biology* 19 (1): 91. https://doi.org/10.1186/s13059-018-1471-8.

Selz, O. M., and O. Seehausen. 2019. 'Interspecific Hybridization Can Generate Functional Novelty in Cichlid Fish'. *Proceedings of the Royal Society B: Biological Sciences* 286 (1913): 20191621. https://doi.org/10.1098/rspb.2019.1621.

Shan, Xiaohui, Zhenlan Liu, Zhenying Dong, Yongming Wang, Yu Chen, Xiuyun Lin, Likun Long, Fangpu Han, Yingshan Dong, and Bao Liu. 2005. 'Mobilization of the Active MITE Transposons MPing and Pong in Rice by Introgression from Wild Rice (Zizania Latifolia Griseb.)'. *Molecular Biology and Evolution* 22 (4): 976–90. https://doi.org/10.1093/molbev/msi082.

Shi, Wei, Antje Krella, Annie Orth, Yang Yu, and Reinald Fundele. 2005. 'Widespread Disruption of Genomic Imprinting in Adult Interspecies Mouse (Mus) Hybrids'. *Genesis (New York, N.Y.: 2000)* 43 (3): 100–108. https://doi.org/10.1002/gene.20161.

Simon, Alexis, Nicolas Bierne, and John J. Welch. 2018. 'Coadapted Genomes and Selection on Hybrids: Fisher's Geometric Model Explains a Variety of Empirical Patterns'. *Evolution Letters* 2 (5): 472–98. https://doi.org/10.1002/evl3.66.

Sjodin, Bryson M F, Kurt E Galbreath, Hayley C Lanier, and Michael A Russello. 2021. 'Chromosome-Level Reference Genome Assembly for the American Pika (Ochotona Princeps)'. *Journal of Heredity* 112 (6): 549–57. https://doi.org/10.1093/jhered/esab031.

Slatkin, M., and R. Lande. 1994. 'Segregation Variance after Hybridization of Isolated Populations'. *Genetical Research* 64 (1): 51–56. https://doi.org/10.1017/s0016672300032547.

Smith, Andrew T., ed. 2018. *Lagomorphs: Pikas, Rabbits, and Hares of the World*. Baltimore: Johns Hopkins University Press.

Smith-Unna, Richard, Chris Boursnell, Rob Patro, Julian Hibberd, and Steven Kelly. 2016a. 'TransRate: Reference Free Quality Assessment of de Novo Transcriptome Assemblies'. *Genome Research*, June, gr.196469.115. https://doi.org/10.1101/gr.196469.115.

Smith-Unna, Richard, Chris Boursnell, Rob Patro, Julian M. Hibberd, and Steven Kelly. 2016b. 'TransRate: Reference-Free Quality Assessment of de Novo Transcriptome Assemblies'. *Genome Research* 26 (8): 1134–44. https://doi.org/10.1101/gr.196469.115.

Song, Ying, Stefan Endepols, Nicole Klemann, Dania Richter, Franz-Rainer Matuschka, Ching-Hua Shih, Michael W. Nachman, and Michael H. Kohn. 2011. 'Adaptive

Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice'. *Current Biology: CB* 21 (15): 1296–1301. https://doi.org/10.1016/j.cub.2011.06.043.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, et al. 2014. 'Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation'. *Science* 344 (6185): 738–42. https://doi.org/10.1126/science.1252136.

Srinivasa, Ananya Nidamangala, and Sarah E. Zanders. 2020. 'Meiotic Drive'. *Current Biology* 30 (11): R627–29. https://doi.org/10.1016/j.cub.2020.04.023.

Steensma, David P., Robert A. Kyle, and Marc A. Shampo. 2010. 'Abbie Lathrop, the "Mouse Woman of Granby": Rodent Fancier and Accidental Genetics Pioneer'. *Mayo Clinic Proceedings* 85 (11): e83. https://doi.org/10.4065/mcp.2010.0647.

Stelkens, Rike, and Ole Seehausen. 2009. 'Genetic Distance Between Species Predicts Novel Trait Expression in Their Hybrids'. *Evolution* 63 (4): 884–97. https://doi.org/10.1111/j.1558-5646.2008.00599.x.

Suchentrunk, F, Z Mamuris, C Stamatis, H Ben Slimen, K Hackländer, G Haerer, and M Giacometti. 2005. 'Introgressive Hybridization in Wild Living Mountain Hares (L. Timidus Varronis) and Brown Hares (L. Europaeus) and Morphological Consequences'. *Mamm. Biol* 70 (Suppl.): 39–40.

Suzuki, H, M Nunome, G Kinoshita, K P Aplin, P Vogel, A P Kryukov, M-L Jin, et al. 2013. 'Evolutionary and Dispersal History of Eurasian House Mice Mus Musculus Clarified by More Extensive Geographic Sampling of Mitochondrial DNA'. *Heredity* 111 (5): 375–90. https://doi.org/10.1038/hdy.2013.60.

Suzuki, Hitoshi, Lyudmila V. Yakimenko, Daiki Usuda, and Liubov V. Frisman. 2015. 'Tracing the Eastward Dispersal of the House Mouse, Mus Musculus'. *Genes and Environment* 37 (August). https://doi.org/10.1186/s41021-015-0013-9.

Svardal, Hannes, Fu Xiang Quah, Milan Malinsky, Benjamin P Ngatunga, Eric A Miska, Walter Salzburger, Martin J Genner, George F Turner, and Richard Durbin. 2020. 'Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation'. *Molecular Biology and Evolution* 37 (4): 1100–1113. https://doi.org/10.1093/molbev/msz294.

Svedin, Nina, Chris Wiley, Thor Veen, Lars Gustafsson, and Anna Qvarnström. 2008. 'Natural and Sexual Selection against Hybrid Flycatchers'. *Proceedings of the Royal Society B: Biological Sciences* 275 (1635): 735–44. https://doi.org/10.1098/rspb.2007.0967.

Sweigart, Andrea L. 2010. 'The Genetics of Postmating, Prezygotic Reproductive Isolation Between Drosophila Virilis and D. Americana'. *Genetics* 184 (2): 401–10. https://doi.org/10.1534/genetics.109.111245.

Sweigart, Andrea L., and Lex E. Flagel. 2015. 'Evidence of Natural Selection Acting on a Polymorphic Hybrid Incompatibility Locus in Mimulus'. *Genetics* 199 (2): 543–54. https://doi.org/10.1534/genetics.114.171819.

Takada, Toyoyuki, Toshinobu Ebata, Hideki Noguchi, Thomas M. Keane, David J. Adams, Takanori Narita, Tadasu Shin-I, et al. 2013. 'The Ancestor of Extant Japanese Fancy Mice Contributed to the Mosaic Genomes of Classical Inbred Strains'. *Genome Research* 23 (8): 1329–38. https://doi.org/10.1101/gr.156497.113.

Tang, Shanwu, and Daven C. Presgraves. 2009. 'Evolution of the Drosophila Nuclear Pore Complex Results in Multiple Hybrid Incompatibilities'. *Science (New York, N.Y.)* 323 (5915): 779–82. https://doi.org/10.1126/science.1169123.

Taylor, Scott A., and Erica L. Larson. 2019. 'Insights from Genomes into the Evolutionary Importance and Prevalence of Hybridization in Nature'. *Nature Ecology & Evolution* 3 (2): 170. https://doi.org/10.1038/s41559-018-0777-y.

Teeter, Katherine C., Bret A. Payseur, Leslie W. Harris, Margaret A. Bakewell, Lisa M. Thibodeau, Janelle E. O'Brien, James G. Krenz, Maria A. Sans-Fuentes, Michael W. Nachman, and Priscilla K. Tucker. 2008. 'Genome-Wide Patterns of Gene Flow across a House Mouse Hybrid Zone'. *Genome Research* 18 (1): 67–76. https://doi.org/10.1101/gr.6757907.

Thulin, C. G., M. Jaarola, and H. Tegelström. 1997. 'The Occurrence of Mountain Hare Mitochondrial DNA in Wild Brown Hares'. *Molecular Ecology* 6 (5): 463–67.

Thulin, C G, M Jaarola, and H Tegelstrom. 1997. 'The Occurrence of Mountain Hare Mitochondrial DNA in Wild Brown Hares.' *Molecular Ecology* 6 (5): 463–67.

Thulin, Carl Gustaf. 2003. 'The Distribution of Mountain Hares Lepus Timidus in Europe: A Challenge from Brown Hares L. Europaeus?' *Mammal Review* 33 (1): 29–42. https://doi.org/10.1046/j.1365-2907.2003.00008.x.

Thulin, Carl Gustaf, Meiying Fang, and Alexander O. Averianov. 2006. 'Introgression from Lepus Europaeus to L. Timidus in Russia Revealed by Mitochondrial Single Nucleotide Polymorphisms and Nuclear Microsatellites'. *Hereditas* 143 (2006): 68–76. https://doi.org/10.1111/j.2006.0018-0661.01952.x.

Thybert, David, Maša Roller, Fábio C. P. Navarro, Ian Fiddes, Ian Streeter, Christine Feig, David Martin-Galvez, et al. 2018. 'Repeat Associated Mechanisms of Genome Evolution and Function Revealed by the Mus Caroli and Mus Pahari Genomes'. *Genome Research*, March. https://doi.org/10.1101/gr.234096.117.

Tishkoff, Sarah A., Floyd A. Reed, Françoise R. Friedlaender, Christopher Ehret, Alessia Ranciaro, Alain Froment, Jibril B. Hirbo, et al. 2009. 'The Genetic Structure and History of Africans and African Americans'. *Science (New York, N.Y.)* 324 (5930): 1035–44. https://doi.org/10.1126/science.1172257.

Turner, Leslie M, and Bettina Harr. 2014. 'Genome-Wide Mapping in a House Mouse Hybrid Zone Reveals Hybrid Sterility Loci and Dobzhansky-Muller Interactions'. Edited by Gil McVean. *ELife* 3 (December): e02504. https://doi.org/10.7554/eLife.02504.

Turner, Thomas L., and Matthew W. Hahn. 2010. 'Genomic Islands of Speciation or Genomic Islands and Speciation?' *Molecular Ecology* 19 (5): 848–50. https://doi.org/10.1111/j.1365-294X.2010.04532.x.

Unckless, R. L., and H. A. Orr. 2009. 'Dobzhansky–Muller Incompatibilities and Adaptation to a Shared Environment'. *Heredity* 102 (3): 214–17. https://doi.org/10.1038/hdy.2008.129.

Ungerer, Mark C., Suzanne C. Strakosh, and Ying Zhen. 2006. 'Genome Expansion in Three Hybrid Sunflower Species Is Associated with Retrotransposon Proliferation'. *Current Biology* 16 (20): R872–73. https://doi.org/10.1016/j.cub.2006.09.020.

Valen, Leigh M. van. 1973. 'A New Evolutionary Law'. In .

Valk, Tom van der, Patrícia Pečnerová, David Díez-del-Molino, Anders Bergström, Jonas Oppenheimer, Stefanie Hartmann, Georgios Xenikoudakis, et al. 2021. 'Million-Year-Old DNA Sheds Light on the Genomic History of Mammoths'. *Nature* 591 (7849): 265–69. https://doi.org/10.1038/s41586-021-03224-9.

Vanlerberghe, F., B. Dod, P. Boursot, M. Bellis, and F. Bonhomme. 1986. 'Absence of Y-Chromosome Introgression across the Hybrid Zone between Mus Musculus Domesticus and Mus Musculus Musculus'. *Genetics Research* 48 (3): 191–97. https://doi.org/10.1017/S0016672300025003.

Veller, Carl, Nathaniel B. Edelman, Pavitra Muralidhar, and Martin A. Nowak. 2019. 'Recombination, Variance in Genetic Relatedness, and Selection against Introgressed DNA'. https://doi.org/10.1101/846147.

Vrana, Paul B. 2007. 'Genomic Imprinting as a Mechanism of Reproductive Isolation in Mammals'. *Journal of Mammalogy* 88 (1): 5–23. https://doi.org/10.1644/06-MAMM-S-013R1.1.

Vyskočilová, MARTINA, ZDENĚK TRACHTULEC, JIŘÍ FOREJT, and JAROSLAV PIÁLEK. 2005. 'Does Geography Matter in Hybrid Sterility in House Mice?' *Biological Journal of the Linnean Society* 84 (3): 663–74. https://doi.org/10.1111/j.1095-8312.2005.00463.x.

Wang, Wenjuan, Yafang Wang, Fumin Lei, Yang Liu, Haitao Wang, and Jiakuan Chen. 2019. 'Incomplete Lineage Sorting and Introgression in the Diversification of Chinese Spot-Billed Ducks and Mallards'. *Current Zoology* 65 (5): 589–97. https://doi.org/10.1093/cz/zoy074.

Weissbrod, Lior, Fiona B. Marshall, François R. Valla, Hamoudi Khalaily, Guy Bar-Oz, Jean-Christophe Auffray, Jean-Denis Vigne, and Thomas Cucchi. 2017. 'Origins of House Mice in Ecological Niches Created by Settled Hunter-Gatherers in the Levant 15,000 y Ago'. *Proceedings of the National Academy of Sciences* 114 (16): 4099–4104. https://doi.org/10.1073/pnas.1619137114.

White, Michael A., Brian Steffy, Tim Wiltshire, and Bret A. Payseur. 2011. 'Genetic Dissection of a Key Reproductive Barrier Between Nascent Species of House Mice'. *Genetics* 189 (1): 289–304. https://doi.org/10.1534/genetics.111.129171.

White, Michael A., Maria Stubbings, Beth L. Dumont, and Bret A. Payseur. 2012. 'Genetics and Evolution of Hybrid Male Sterility in House Mice'. *Genetics* 191 (3): 917–34. https://doi.org/10.1534/genetics.112.140251.

Whitney, Kenneth D., Rebecca A. Randell, and Loren H. Rieseberg. 2010. 'Adaptive Introgression of Abiotic Tolerance Traits in the Sunflower Helianthus Annuus'. *The New Phytologist* 187 (1): 230–39. https://doi.org/10.1111/j.1469-8137.2010.03234.x.

Wright, Kevin M., Deborah Lloyd, David B. Lowry, Mark R. Macnair, and John H. Willis. 2013. 'Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in Mimulus Guttatus'. *PLOS Biology* 11 (2): e1001497. https://doi.org/10.1371/journal.pbio.1001497.

Wu, Chung-I. 2001. 'The Genic View of the Process of Speciation'. *Journal of Evolutionary Biology* 14 (6): 851–65. https://doi.org/10.1046/j.1420-9101.2001.00335.x.

Yamada, Fumio, Mika Takaki, and Hitoshi Suzuki. 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): 107–16. https://doi.org/10.1266/ggs.77.107.

Yonekawa, H, K Moriwaki, O Gotoh, N Miyashita, Y Matsushima, L M Shi, W S Cho, X L Zhen, and Y Tagashira. 1988. 'Hybrid Origin of Japanese Mice "Mus Musculus Molossinus": Evidence from Restriction Analysis of Mitochondrial DNA.' *Molecular Biology and Evolution* 5 (1): 63–78. https://doi.org/10.1093/oxfordjournals.molbev.a040476.

Yonekawa, Hiromichi, Kazuo Moriwaki, Osamu Gotoh, Jun-Ichi Hayashi, Junko Watanabe, Nobumoto Miyashita, Michael L. Petras, and Yusaku Tagashira. 1981. 'Evolutionary Relationships among Five Subspecies of MUS MUSCULUS Based on Restriction Enzyme Cleavage Patterns of Mitochondrial DNA'. *Genetics* 98 (4): 801–16.

Zachos, F. E., H. Ben Slimen, K. Hackländer, M. Giacometti, and F. Suchentrunk. 2010. 'Regional Genetic in Situ Differentiation despite Phylogenetic Heterogeneity in Alpine Mountain Hares'. *Journal of Zoology* 282 (1): 47–53. https://doi.org/10.1111/j.1469-7998.2010.00710.x.

Zeh, David W., Jeanne A. Zeh, and Yoichi Ishida. 2009. 'Transposable Elements and an Epigenetic Basis for Punctuated Equilibria'. *BioEssays* 31 (7): 715–26. https://doi.org/10.1002/bies.200900026.

Data Citations

1. Marques, J. P., Ferreira, M. S., Callahan, C. M., Hackländer, K., Jenny, H., Montgomery, W. I., Reid, N., Good, J. M., Alves, P. C. & Melo-Ferreira, J. *GenBank* Project PRJNA358867 (2017)

2. Marques, J. P., Ferreira, M. S., Callahan, C. M., Hackländer, K., Jenny, H., Montgomery, W. I., Reid, N., Good, J. M., Alves, P. C. & Melo-Ferreira, J. *Figshare* https://*Figshare*.com/s/97c2240fe0b5b320eb0f (2017)

3. *GenBank* Project SRP055741 (2016).

FCUP and U. Montpellier

Using genomic tools to understand species differentiation and admixture in hares and mice

An Annotated Draft Genome of the Mountain Hare (*Lepus timidus*)

Genome Biology and Evolution 12(1):3656–3662
doi:10.1093/gbe/evz273

João P. Marques[1,2,3,†], Fernando A. Seixas[1,2,3,†], Liliana Farelo[1], Colin M. Callahan[4], Jeffrey M. Good[4,5], W. Ian Montgomery[6], Neil Reid[6], Paulo C. Alves[1,2,5], Pierre Boursot[3,*] and José Melo-Ferreira[1,2,*]

[1] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal
[2] Departamento de Biologia, Faculdade de Ciências do Porto, Portugal
[3] Institut des Sciences de l'Evolution Montpellier (ISEM), Univ Montpellier, CNRS, IRD, EPHE, France
[4] Division of Biological Sciences, University of Montana
[5] Wildlife Biology Program, University of Montana
[6] Institute of Global Food Security, School of Biological Sciences, Queen's University Belfast, United Kingdom

[†] These authors contributed equally to this work.
[*] Corresponding authors: E-mails: pierre.boursot@umontpellier.fr and jmeloferreira@cibio.up.pt.

Accepted: December 7, 2019

Data deposition: This project has been deposited in the NCBI Sequence Read Achive Database under BioProject PRJNA561582. The hare pseudoreference genome is available in Dryad: https://doi.org/10.5061/dryad.x95x69pd8.

Abstract

Hares (genus *Lepus*) provide clear examples of repeated and often massive introgressive hybridization and striking local adaptations. Genomic studies on this group have so far relied on comparisons to the European rabbit (*Oryctolagus cuniculus*) reference genome. Here, we report the first de novo draft reference genome for a hare species, the mountain hare (*Lepus timidus*), and evaluate the efficacy of whole-genome re-sequencing analyses using the new reference versus using the rabbit reference genome. The genome was assembled using the ALLPATHS-LG protocol with a combination of overlapping pair and mate-pair Illumina sequencing (77x coverage). The assembly contained 32,294 scaffolds with a total length of 2.7 Gb and a scaffold N50 of 3.4Mb. Re-scaffolding based on the rabbit reference reduced the total number of scaffolds to 4,205 with a scaffold N50 of 194Mb. A correspondence was found between 22 of these hare scaffolds and the rabbit chromosomes, based on gene content and direct alignment. We annotated 24,578 protein coding genes by combining *ab-initio* predictions, homology search and transcriptome data, of which 683 were solely derived from hare-specific transcriptome data. The hare reference genome is therefore a new resource to discover and investigate hare-specific variation. Similar estimates of heterozygosity and inferred demographic history profiles were obtained when mapping hare whole-genome re-sequencing data to the new hare draft genome or to alternative references based on the rabbit genome. Our results validate previous reference-based strategies and suggest that the chromosome-scale hare draft genome should enable chromosome-wide analyses and genome scans on hares.

Key words

Whole-genome sequencing, de novo assembly, annotation, hares, Leporids, Lagomorpha.

Introduction

The ability to sequence whole genomes has revolutionized our power to study the evolution of non-model organisms. Hares (genus *Lepus*) have recently emerged as useful evolutionary models to understand introgressive hybridization and local adaptation (Alves et al. 2008; Jones et al. 2018; Seixas et al. 2018; Giska et al. 2019). Genomic analyses on this group have primarily relied on comparisons to the high-quality reference genome of another leporid, the more extensively studied European rabbit (*Oryctolagus cuniculus*) (Carneiro et al. 2014), estimated to share a most recent common ancestor with hares 12 million years ago (Matthee et al. 2004). Although these studies have generally used iterative mapping approaches to reduce divergence and increase mapping efficiency (e.g., Jones et al. 2018; Seixas et al. 2018), it remains unclear to what extent reliance on an outgroup reference may have limited genomic inferences.

We extend the genomic resources of Leporids by assembling the first draft genome of a hare species, the mountain hare (*Lepus timidus*). The mountain hare is an arcto-alpine species widely distributed in the northern Palearctic, from western Europe to eastern Asia, with some isolated populations, as in the Alps, Poland, Great Britain, and Ireland. The current distribution of the species reflects the colonization of previously glaciated areas in the north, and the retreat from southernmost regions in the south (Waltari and Cook 2005; Hamill et al. 2006; Melo-Ferreira et al. 2007; Smith et al. 2018).

The species has been implicated in recurrent events of introgressive hybridization with other hare species from Europe (Thulin et al. 1997; Alves et al. 2003; Melo-Ferreira et al. 2009; Seixas et al. 2018; Giska et al. 2019), and displays important locally adapted traits, such as varying ecologies (Caravaggi et al. 2017), size differences among regions, or distinctive coat color (Smith et al. 2018; Giska et al. 2019). Furthermore, genus Lepus is distributed worldwide with more than 30 classified species, which show adaptions to contrasting environments, from arctic to arid regions. Detailed investigation of relevant evolutionary processes in the genus can benefit from the availability of hare-specific genomic resources (Fontanesi et al. 2016).

Materials and Methods

DNA Sampling, Extraction, and Sequencing

One female mountain hare (Lepus timidus hibernicus) specimen (NCBI BioSample ID SAMN12621015) was captured from the wild for scientific research purposes by the Irish Coursing Club (ICC) at Borris-in-Ossory, County Laois under National Parks & Wildlife (NPWS) license no. C 337/2012 issued by the Department of Arts, Heritage and the Gaeltacht (dated October 31, 2012). Genomic DNA was extracted from kidney, muscle, and ear tissue using the JETquick Tissue DNA Spin Kit (GENOMED), with RNAse and proteinase K to remove RNA and protein contamination. Genomic libraries of different insert lengths were generated following the standard ALLPATHS-LG protocol (Gnerre et al. 2011): one Illumina TruSeq DNA library of 180 bp fragments was sequenced with overlapping paired-end (OPE) reads, and three Illumina TruSeq DNA mate-pair (MP) libraries of 2.5, 4.5, and 8.0 kb insert sizes. Whole-genome sequencing was performed at The Genome Analysis Center (TGAC, currently Earlham Institute, Norwich, UK)—seven HiSeq2000 lanes (five OPE and two 4.5 kb MP)—and CIBIO's New-Gen sequencing platform—three HiSeq1500 lanes (2.5 and 8.0 kb MP). Raw sequencing reads were deposited in the Sequence Read Archive (details in supplementary table S1, Supplementary Material online).

Read Quality Assessment and Filtering

Exact duplicates were removed both from OPE and MP libraries using PRINSEQ v0.20.4 (Schmieder and Edwards 2011b). PhiX sequences were identified using Bowtie2-v2.2.3 (Langmead and Salzberg 2012) and removed. Adapter sequences were removed using Cutadapt v1.4.1 (Martin 2011) for OPE reads and Skewer v1.3.1 (Jiang et al. 2014) for mate-pairs. For the latter, only pairs in the correct orientation determined by the presence of the junction adapter were retained.

Genome Size Estimation

Genome size was estimated using a k-mer-based approach (Marc¸ais and Kingsford 2011). First, the frequency distribution of 17 bp k-mers was obtained using jellyfish v2.2.6 (Marçais and Kingsford 2011) based on the OPE raw reads—

supplementary figure S1, Supplementary Material online. The sequencing depth was then calculated following M = N* (L – k + 1)/L, where M is the peak of the k-mer depth frequency distribution, L is the read length, and k is the chosen k-mer length in bp. Finally, the genome size was estimated by dividing the total number of bases sequenced by the sequencing depth.

Genome Assembly and Annotation

De novo assembly was performed using ALLPATHS-LG (Gnerre et al. 2011) with default parameters using OPE and mate-pair reads. The resulting assembly was evaluated with REAPR v1.0.18 (Hunt et al. 2013) to break incorrect scaffolds, by mapping the paired-end and the 4.5 kb mate-pair reads on the assembled genome. Another round of scaffolding was then performed using SSPACE v3.0 (Boetzer et al. 2011), with a minimum overlap of 32 bp and supported by a minimum of 20 reads. Finally, we leveraged the existence of the high-quality assembly of the genome of the European rabbit (Oryctolagus cuniculus—Ensembl OryCun2.0), to improve the contiguity of the assembly using the reference-based scaffolder MeDuSa v.1.6 with five iterations (Bosi et al. 2015). This re-scaffolding orders and re-orientates scaffolds without affecting intra-scaffold sequence. Quality control of the assembly at different stages was assessed based on metrics obtained with QUAST v.3.2 (Mikheenko et al. 2016). The completeness of the L. timidus re-scaffolded genome was evaluated using BUSCO v.3.0.2 (Sim~ ao et al. 2015), based on the presence and absence of core single-copy genes (from mammalia_odb9 database). We then checked consistency of gene content in the larger chromosome-like scaffolds and rabbit chromosomes using blastp from NCBI BLAST v2.7.1þ (Camacho et al. 2009), considering the best hit per gene with similarity above 90% over 500 bp. The 22 rabbit chromosomes were aligned against inferred corresponding *L. timidus* re-scaffolded scaffolds using D-Genies v. 1.2.0 Mashmap (Cabanettes and Klopp 2018).

Repetitive regions were identified using RepeatModeler v.1.0.11 (Smit and Hubley 2008) and masked using RepeatMasker v.4.0.7 (Smit et al. 2013). The masked genome was used as input for gene prediction in MAKER v.3.01.02 (Cantarel et al. 2008), using ab-initio predictions, *L. timidus* transcriptome data, and rabbit protein annotations (*O. cuniculus*) (supplementary text, Supplementary Material online). Functional inference for genes and transcripts was performed using the translated CDS features of each coding transcript. Each predicted protein sequence was based on blastp searches against the Uniprot-Swissprot database to retrieve gene name and function,

and InterProscan v5.30-69 (Jones et al. 2014) to retrieve Interpro, Pfam v31.0 (Finn et al. 2016), GO (Mi et al. 2017), KEGG (Kanehisa et al. 2016), and Reactome (Fabregat et al. 2018) information.

Analyses of Whole-Genome Re-Sequencing Data

To compare the performance of using the *L. timidus* genome or other strategies based on the rabbit genome for wholegenome analyses, we analysed re-sequencing data from the mountain hare and another hare species, the Iberian hare, *L. granatensis*, mapping the reads to 1) the new *L. timidus* re-scaffolded genome, 2) the rabbit reference genome (available from Ensembl—OryCun2.0, release 80), and 3) a hare pseudo-reference genome built through iterative mapping of hare sequence reads on the rabbit genome, followed by reference updating (from Seixas et al. 2018). For the resequencing data (NCBI Sequence Read Archive Biosamples SAMN07526960 and SAMN07526963; Seixas et al. 2018), adapters were removed using cutadapt version 1.8 (Martin 2011) and low quality bases (quality < 20 at the end of reads, and 4 consecutive bp with average quality < 30) were trimmed using Trimmomatic v0.33 (Bolger et al. 2014). Mapping was done using BWA-MEM v0.7.15 (Li 2013). Mapped reads were sorted with samtools v1.3.1 (Li et al. 2009) and read duplicates removed using Picard Markduplicates (Picard toolkit 2019). Realignment around INDELs was performed using GATK v3.2-2 (Van der Auwera et al. 2013). Genotype calling was performed for each species independently using bcftools v.1.5 (Li 2011), with minimum site (QUAL) and RMS mapping (MQ) qualities of 20, coverage (FMT/DP) between 6X and twice the average genomic coverage, and minimum genotype quality (FMT/GQ) of 20. Indels and flanking 10 bp were coded as missing data. Only sites covered in the two analysed individuals were retained. Heterozygosity was calculated in sliding windows of 50 kb, using the popgenWindows.py script available at https://github.com/simonhmartin/genomics_general, and 500 windows were randomly sampled per references and species.

The Pairwise Sequentially Markovian Coalescent (PSMC) model (Li and Durbin 2011) was then used to compare single-genome demographic inferences of the mountain hare using alternative genome references (*L. timidus* assembled genome, prior to re-scaffolding, was also included, to control for potential biases arising from the reference-based re-scaffolding process), and to infer the demographic profiles of *L. timidus* and *L. granatensis* using the *L. timidus* rescaffolded reference (as in Seixas et

al. 2018, who used the hare pseudo-reference). Changes in the density of called variants among references should cause important differences in the inferred profiles. Diploid consensus sequences were built using samtools v1.3.1 mpileup and call modules, and only sites with minimum base and mapping qualities of 20, and coverage between 8X and twice the average depth were considered (atomic time intervals were set to 4 + 50*2 + 2 + 4 as in Seixas et al. 2018). Results were scaled using a generation time of 2 years (Marboutin and Peroux 1995) and a mutation rate (l) of 2.8×10^{-9} substitutions/ site/generation (Seixas et al. 2018). The variance of effective population size (Ne) estimates was assessed by 50 bootstraps in randomly sampled segments with replacement.

Results and Discussion

De Novo Reference Genome Assembly and Annotation

Genome assembly and sequencing metrics are in table 1 and supplementary table S1, Supplementary Material online. The assembly length, 2.70 Gb, was consistent with the k-mer estimate (2.75 Gb) and the assembled length of the rabbit genome (2.74 Gb; Carneiro et al. 2014). The *L. timidus* rescaffolded genome contained 4,205 scaffolds, being 99.9% of the total assembly size comprised in 605 scaffolds with a minimum length of 35 kb. Of the 4,104 mammalian core genes, 3,793 (92.4%) were present in our assembly, 3,445 of which (90.8%) were found as complete single copies. The number of predicted and annotated genes (29,238 and 24,578, respectively—table 1 and supplementary fig. S2, Supplementary Material online) are in line with several published mammalian genomes that used similar sequencing approaches (Keane et al. 2015; Li et al. 2017; Koepfli et al. 2019; Ming et al. 2019), and with the extrapolation from the BUSCO completeness assessment, suggesting that the majority of genes present in our draft genome was covered by the annotation process. A total of 683 predicted genes were uniquely annotated based on the hare transcriptome and possibly represent hare-specific genes (supplementary fig. S2, Supplementary Material online).

Through the characterization of gene content (supplementary fig. S3, Supplementary Material online) and chromosome-scaffold alignment (supplementary Figs. S4 and S5, Supplementary Material online), we were able to establish correspondence between the rabbit chromosomes (2N = 44) and 22 scaffolds of the re-scaffolded version of the *L. timidus* assembly. The 22 scaffolds correspond to 83% of the total length of the assembly (2.24 Gb). It should be noted that hares have 24 chromosomes, since rabbit's chromosomes 1 and 2 are each split into two in hares (presumably resulting from two fusions in the rabbit lineage; Robinson et al. 2002). This karyotype difference was naturally not recovered in our re-scaffolded assembly, which highlights the inherent shortcomings of reference-guided scaffolding. While the new genome should be accurate in resolving small-scale structural variation (small insertions–-deletions, repeats, and/or inversions as recovered by the original assembled contigs/scaffolds; Salzberg et al. 2012), larger genomic rearrangements, should be missed due to the assumption of synteny with the reference.

The Impact of Alternative Reference Mapping Strategies on Genomic Analyses

The proportion of mapped reads from whole-genome re-sequencing was higher using the hare pseudo-reference, but the number of uniquely mapped reads was larger using the *L. timidus* reference genome (supplementary table S2, Supplementary Material online). These statistics suggest that the hare pseudo-reference increases both mapping proportion and efficiency, but the new hare reference allows increased confidence in mapping, measured as the proportion of uniquely mapped reads. The distributions of heterozygosity estimated in 50 kb windows along the genome did not differ significantly across analyses with different references (P> 0.05, Wilcoxon Ranked-sum test; supplementary fig.S7, Supplementary Material online). In agreement, the PSMC demographic profiles also displayed similar shapes across references, with only slight differences of inferred effective population sizes (fig. 1a). Also, the demographic profiles of both *L. timidus* and *L. granatensis* inferred using the new *L. timidus* re-scaffolded genome are similar to those inferred by Seixas et al. (2018) using the hare pseudo-reference (fig. 1b). These results suggest that the use of the alternative tested references does not impact heterozygosity tract patterns, and thus that approaches based on hare pseudoreferences has not limited evolutionary inference and genome scans on hares. It also shows that re-scaffolding our *de novo* assembly using the rabbit genome enables the use of the new hare genome reference for genomic scale scans, where the ordering along the chromosome is important. Finally, the new hare draft genome can be useful to reveal hare-specific variation, reflected for example in the putative hare-specific genes annotated here, which needs to be evaluated and investigated.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments

In Ireland, a license (no. C 337/2012 dated October 31, 2012) to take animals for scientific purposes under the Wildlife Acts 1976–-2012 Sections 9, 23, and 34 was issued by Laura Claffey, Wildlife Licensing Unit, National Parks & Wildlife Service (NPWS), Department of Arts, Heritage & the Gaeltacht (DAHG). Thanks to Jimi Conroy, the local NPWS Conservation Officer, who consented to sampling within his jurisdiction. The Irish Coursing Club (ICC) captured animals from the wild and we thank D.J. Histon, Chief Executive and Secretary, who supported the research. William Fitzgerald, Department of Agriculture, Food and the Marine (DAFM) dispatched animals in a humane manner and provided dissection skills. This work was funded by the project HybridAdapt, supported by Portuguese National Funds through the Fundação para a Ciência e a Tecnologia (FCT) (reference FCT-ANR/BIA-EVF/0250/2012), and by French Agence Nationale de la Recherche (ANR) (reference ANR-12-ISV7–0002-01). Additional support was obtained from Programa Operacional Potencial Humano – Quadro de Referência Estratégica Nacional (POPH-QREN) funds from the European Social Fund (ESF) and FCT (IF/00033/2014/CP1256/CT0005 FCT Investigator grant to J.M.-F., and SFRH/BD/115089/2016 and SFRH/BD/87126/2012 PhD grants to J.P.M. and F.A.S., respectively), and from a National Science Foundation (NSF) EPSCoR grant (OIA-1736249 to J.M.G.). Instrumentation, laboratory, and/or computational support were provided by CIBIO NEW-GEN sequencing platform, supported by the European Union under grant agreement no. 286431, by the Montpellier Bioinformatics Biodiversity platform supported by LabEx CeMEB, an ANR “Investissements d’avenir” program (ANR-10-LABX-04-01), and by the University of Montana Genomics Core.

Additional support was obtained from the projects NORTE-01-0145-FEDER-000007 (NORTE2020 Programme, PORTUGAL 2020, through the European Regional Development Fund – ERDF), and POCI-01-0145-FEDER-022184 (COMPETE2020, PORTUGAL2020, and ERDF). It was also supported by Portugal–France bilateral cooperation programs PESSOA (FCT in Portugal; Ministère de l’Europe et des Affaires etrange`res (MEAE) and Ministère de l’Education nationale, de l’Enseignement sup erieur et de la Recherche (MESRI) in France) and Cotutelles/Cotutelas PESSOA (FCT through Conselho de Reitores das Universidades Portuguesas (CRUP) in Portugal; MEAE and MESRI through Conference des Présidents d’Universit e (CPU) in France).

This research was conducted in the frame of the Laboratoire International Associé (LIA) “Biodiversity and Evolution” supported by Institut ecologie et environnement – InEE (Centre national de la recherche scientifique – CNRS, France) and FCT (Portugal).

Author Contributions

J.M.-F., P.B., F.S., and P.C.A. conceived the study. N.R., W.I.M., P.C.A., and J.M.-F. organized and performed sampling. L.F. and C.M.C. performed laboratory work under supervision of J.M.-F. and J.M.G., respectively. J.P.M. and F.A.S. analysed the data. J.P.M., F.A.S., and J.M.-F. wrote the manuscript, with contributions from P.B. and J.M.G. and comments from the other authors. All authors read, revised, and approved the manuscript.

Literature Cited

Alves PC, Ferrand N, Suchentrunk F, Harris DJ. 2003. Ancient introgression of Lepus timidus mtDNA into L. granatensis and L. europaeus in the Iberian Peninsula. Mol Phylogenet Evol. 27(1):70–80.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P. 2008. The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus Lepus. Philos Trans R Soc B Biol Sci. 363(1505):2831–2839.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27(4):578–579.

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

Bosi E, et al. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31(15):2443–2451.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 6:e4958.

Camacho C, et al. 2009. BLASTþ: architecture and applications. BMC Bioinformatics 10(1):421.

Cantarel BL, et al. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18(1):188–196.

Caravaggi A, et al. 2017. Niche overlap of mountain hare subspecies and the vulnerability of their ranges to invasion by the European hare; the (bad) luck of the Irish. Biol Invasions 19(2):655–674.

Carneiro M, et al. 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science 345(6200):1074–1079.

Fabregat A, et al. 2018. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 46(D1):D649–D655.

Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44(D1):D279–D285.

Fontanesi L, et al. 2016. LaGomiCs—lagomorph genomics consortium: an international collaborative effort for sequencing the genomes of an entire mammalian order. J Hered. 107(4):295–308.

Giska I, et al. 2019. Introgression drives repeated evolution of winter coat color polymorphism in hares. Proc Natl Acad Sci U S A. 116(14):24150–24156.

Gnerre S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 108(4):1513–1518.

Hamill RM, Doyle D, Duke EJ. 2006. Spatial patterns of genetic diversity across European subspecies of the mountain hare, Lepus timidus L. Heredity (Edinb). 97(5):355–365.

Hunt M, et al. 2013. REAPR: a universal tool for genome assembly evaluation. Genome Biol. 14(5):R47.

Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15(1):182.

Jones MR, et al. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science 360(6395): 1355–1358.

Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44(D1):D457–D462.

Keane M, et al. 2015. Insights into the evolution of longevity from the bowhead whale genome. Cell Rep. 10(1):112–122.

Koepfli K-P, et al. 2019. Whole genome sequencing and re-sequencing of the sable antelope (Hippotragus niger): a resource for monitoring diversity in ex situ and in situ populations. G3 9(6):1785–1793.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359.

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. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv 00: 3.

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

Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.

Li Z, et al. 2017. Draft genome of the reindeer (*Rangifer tarandus*). Gigascience 6(12):1–5.

Marboutin E, Peroux R. 1995. Survival pattern of European hare in a decreasing population. J Appl Ecol. 32(4):809–816.

Marc¸ais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6):764–770.

Martin M. 2011. Cutadapt removes adapter sequences from highthroughput sequencing reads. Embnet J. 17(1):10.

Matthee CA, Van Vuuren BJ, Bell D, Robinson TJ. 2004. A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the miocene. System Biol. 53(3):433–447.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P. 2009. The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. Mol Ecol. 18(12):2643–2658.

Melo-Ferreira J, et al. 2007. The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. Mol Ecol. 16(3):605–618.

Mi H, et al. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45(D1):D183–D189.

Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. Bioinformatics 32(7):1088–1090.

Ming Y, Jian J, Yu X, Wang J, Liu W. 2019. The genome resources for conservation of Indo-Pacific humpback dolphin, Sousa chinensis. Sci Data 6(1):68.

Picard toolkit. 2019. Broad Institute, available at http://broadinstitute.github.io/picard/.

Robinson TJ, Yang F, Harrison WR. 2002. Chromosome painting refines the history of genome evolution in hares and rabbits (order Lagomorpha). Cytogenet. Genome. Res. 96(1–4):223–227.

Salzberg SL, et al. 2012. GAGE: a critical evaluation of genome assemblies and assembly algorithms. Genome Res. 22(3):557–567.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6):863–864.

Seixas FA, Boursot P, Melo-Ferreira J. 2018. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. Genome Biol. 19(1):91.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.

Smit A, Hubley R. 2008. RepeatModeler Open-1.0, available at http://www.repeatmasker.org.

Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0, available at http://www.repeatmasker.org.

Smith, A.T., Johnston, C.H., Alves, P.C., Hackl€ ander, K., editors. 2018. Lagomorphs: pikas, rabbits, and hares of the world. Baltimore (MD): Johns Hopkins University Press.

Thulin CG, Jaarola M, Tegelstro¨ m H. 1997. The occurrence of mountain hare mitochondrial DNA in wild brown hares. Mol Ecol. 6(5): 463–467.

Van der Auwera GA, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1–33.

Waltari E, Cook JA. 2005. Hares on ice: phylogeography and historical demographics of Lepus arcticus, L. othus, and L. timidus (Mammalia: Lagomorpha). Mol Ecol. 14(10):3005–3016.

Associate Editor: Judith Mank

Supplementary Material for:

An annotated draft genome of the mountain hare (*Lepus timidus*)

João P. Marques[1,2,3#], Fernando A. Seixas[1,2,3#], Liliana Farelo[1], Colin M. Callahan[4], Jeffrey M. Good[4,5], W. Ian Montgomery[6], Neil Reid[6], Paulo C. Alves[1,2,5], Pierre Boursot[3*], José Melo-Ferreira[1,2*]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, 4485-661 Vairão, Portugal.
[2]Departamento de Biologia, Faculdade de Ciências do Porto, 4169-007 Porto, Portugal.
[3]Institut des Sciences de l'Évolution, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France.
[4]Division of Biological Sciences, University of Montana, Missoula, MT 59812, United States of America
[5]Wildlife Biology Program, University of Montana, Missoula, MT 59812, United States of America.
[6]Institute of Global Food Security, School of Biological Sciences, Queen's University Belfast, 19 Chlorine Gardens, Belfast, BT9 5DL, UK.

* **Corresponding authors**: José Melo-Ferreira, CIBIO-InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal, tel: +351 252660411, e-mail: jmeloferreira@cibio.up.pt; Pierre Boursot, Institut des Sciences de l'Évolution, Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier, France, pierre.boursot@umontpellier.fr.

This PDF file includes:

Supplementary Text

Materials and methods – transcriptomic data for gene annotation

RNA extraction and sequencing

One female mountain hare specimen (SAMN12621015) was captured from the wild in Borris-in-Ossory, by the Irish Coursing Club (ICC) for scientific research purposes under National Parks & Wildlife (NPWS) licence No. C 337/2012 issued by the Department of Arts, Heritage and the Gaeltacht (dated 31/10/2012). RNA was extracted from freshly collected liver and skin tissue from the same specimen used for genomic DNA analyses (SAMN12621015), immediately preserved in RNAlater and stored at −80 °C until RNA extraction. Prior to extraction, frozen samples were ground in liquid nitrogen with a ceramic mortar and pestle. Mortar and pestle were washed prior to extraction using a 6-step wash that included the following washing reagents: 70% ethanol, tap water, 10% bleach, milli-Q water, RNase away (Thermo Fisher Scientific) and molecular grade water. RNA extraction was performed using the RNeasy Mini Kit (Qiagen) according to manufacturer instructions.

cDNA libraries were built using the SureSelect Strand-Specific RNA Library Prep kit for Illumina Multiplexed Sequencing (Agilent Technologies). Library sizes were estimated using a Bioanalyzer 2,100 and quantified using KAPA Library quantification kit (KAPA BIOSYSTEMS). Equimolar concentrations of each library were pooled and sequenced (100 bp paired-end) using three lanes of an Illumina HiSeq 2000 at the QB3 facility at the University of California, Berkeley. Raw reads were deposited in the Sequence Read Archive (Accession Number details in Supplementary Table 1).

The quality of sequencing reads was evaluated with FastQC v0.11.519 (Andrews 2010a). TRIMMOMATIC v0.3620 (A. M. Bolger, Lohse, and Usadel 2014) was then used to trim reads, by removing the first ten nucleotides, Illumina adapters, reads below 25 bp long, and trailing nucleotides with quality under 10, and using 4-base sliding windows to ensure average quality above 10.

Transcriptome assembly and quality control

A *de novo* transcriptome assembly was performed using paired reads from the skin (ventral and dorsal) and liver libraries using TRINITY v2.2.021 (Grabherr et al. 2011b), establishing RF as read orientation for a strand-specific assembly. Assembly accuracy and completeness was evaluated with Transrate v1.0.1 (Smith-Unna et al. 2016a). This tool defines an optimal cut-off for the assembly by mapping back the cleaned reads against the generated reference, which consequently enables the removal of possible chimeras and poorly supported contigs, retaining only the best-supported contigs. Redundancy was further reduced using CD-HIT-EST v4.6.4 (Fu et al. 2012b) with a 95% similarity threshold. Finally, TransDecoder v3.0.0 (Haas & Papanicolaou et al. [2015] 2019) was employed to predict candidate coding regions and homology with PFAM common protein domains (El-Gebali et al. 2019) and *O. cuniculus* transcriptome reference (accessed on May 11, 2018), discarding possible non-coding RNA and DNA contamination. The final transcriptome comprised transcripts with predicted open reading frames or homology with the *O. cuniculus* transcriptome or PFAM database. Final transcriptome assembly available under the accession code GHXQ00000000.

Supplementary figures

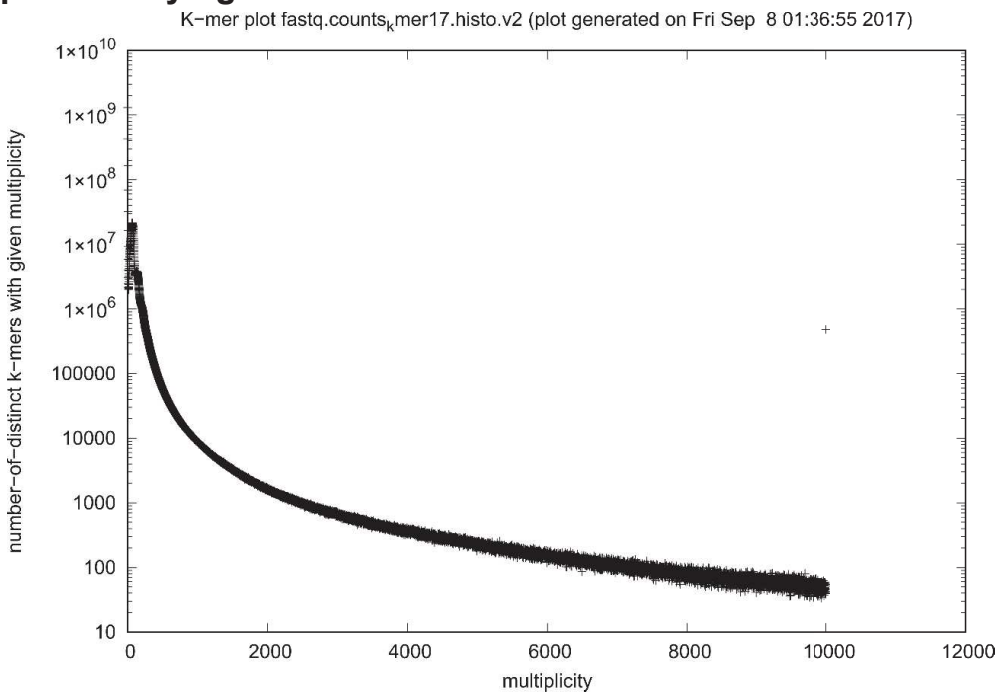


Fig. S1 - Frequency distribution of 17 bp k-mers obtained using jellyfish and based only the overlapping paired-end raw reads. A peak was found at 65.

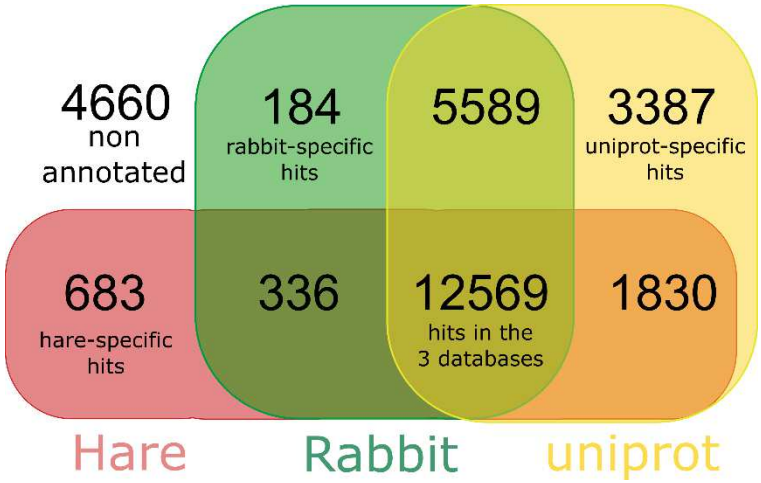


Fig. S2 – Annotation summary of predicted protein-coding genes. Annotation was performed using three independent sources (*L. timidus* transcriptome – Hare; the *O. cuniculus* transcriptome – Rabbit; and the UniProt-SwissProt database (uniprot). The diagram displays the overlapping and database-specific annotations, and the predicted genes that remained non-annotated. Of the 683 hare-specific hits, 454 are located in the 22 scaffolds with correspondence to the rabbit chromosomes.

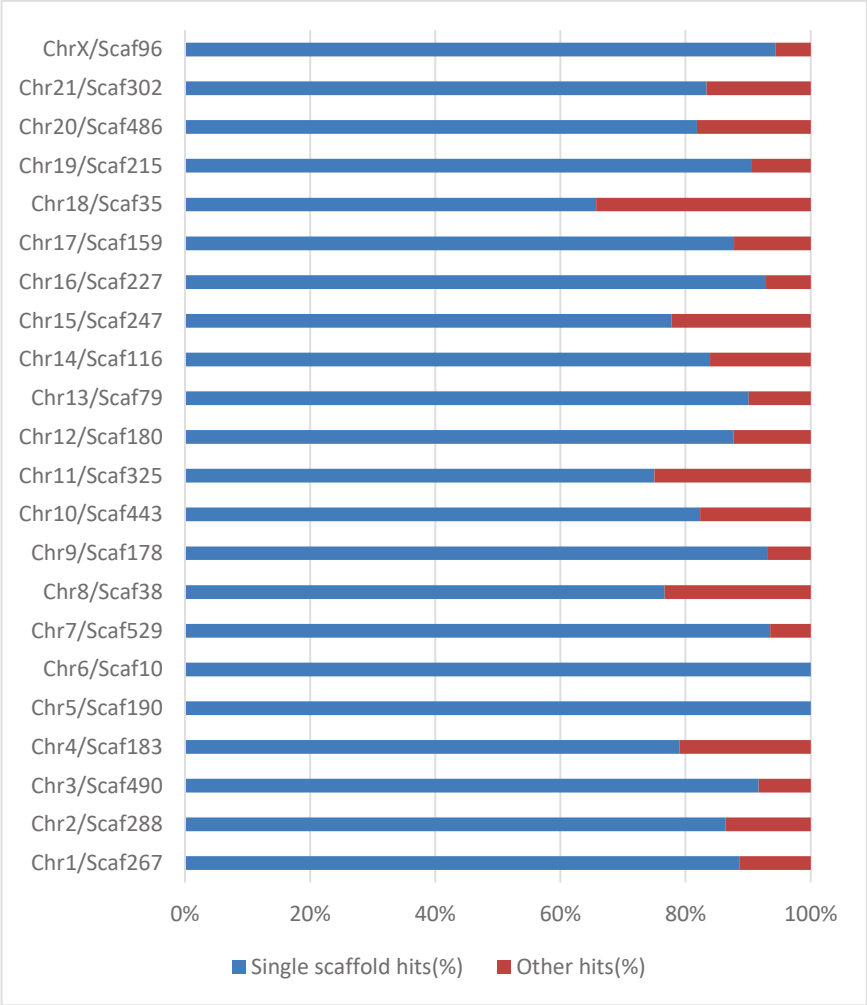


Fig. S3 – Gene content evaluation of the 22 longest scaffolds of the *timidus* re-scaffolded reference genome, using a standard BLASTN homology-based search of the rabbit genes in the hare reference. This figure shows the proportion of rabbit genes present in a given chromosome (Chr) having the longest hit predominantly in a single hare reference scaffold (Scaf). Blue represents the proportion of genes with one to one Chr/Scaf correspondence, while orange depicts the proportion of genes with the longest hit on another scaffold.

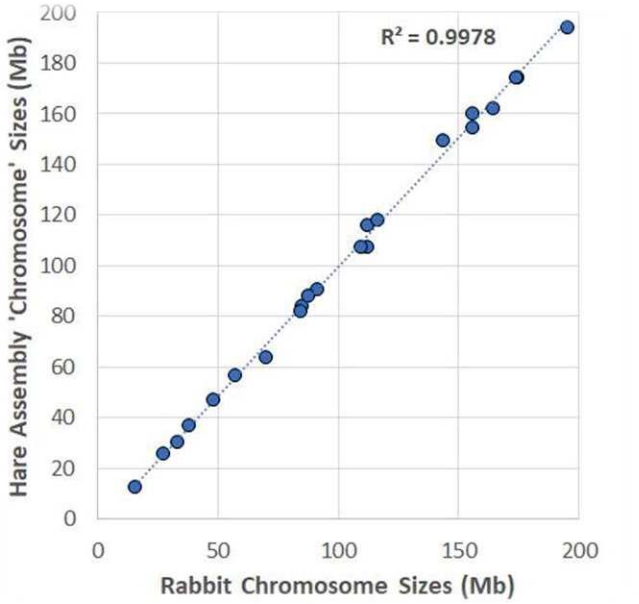


Fig. S4 - Size correlation between the 22 largest scaffolds of the *timidus* re-scaffolded genome and the corresponding rabbit chromosomes (Spearman's rank correlation p=0.9978) (correspondence established based on gene content evaluation; Fig. S3). The dashed line indicates a linear regression trendline.

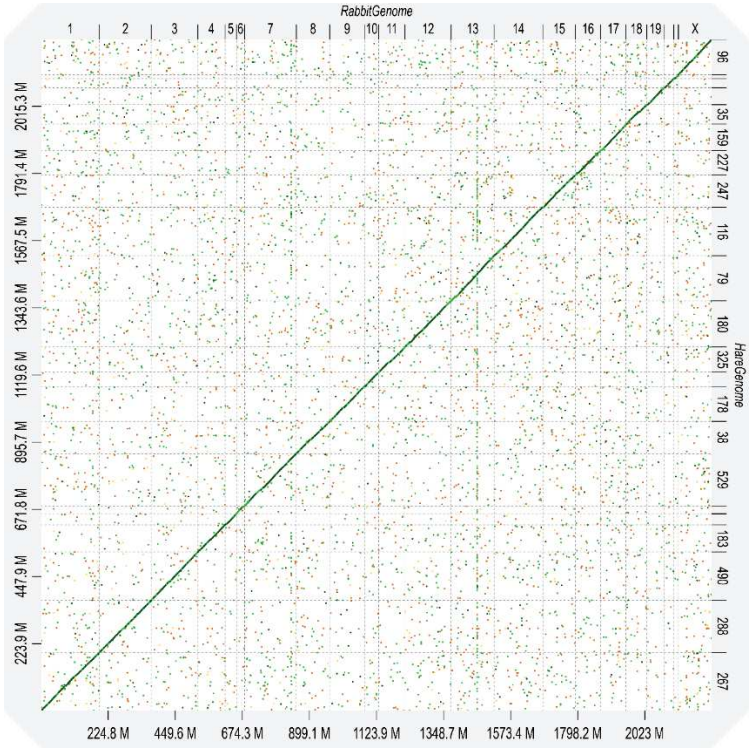


Fig. S5 – Syntenic similarity overview comparing the rabbit genome with the re-scaffolded version of the hare genome.

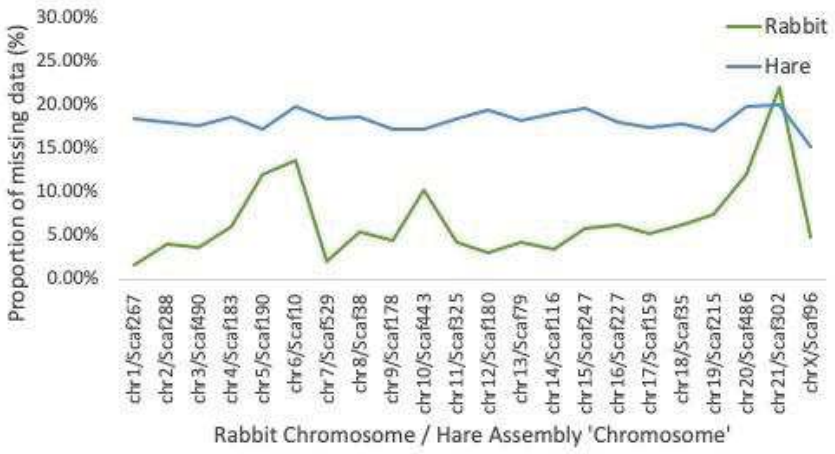


Fig. S6 – Proportion of missing data in the rabbit reference chromosomes and the corresponding scaffolds of the *timidus* re-scaffolded genome (correspondence established based on gene content; Fig. S3).

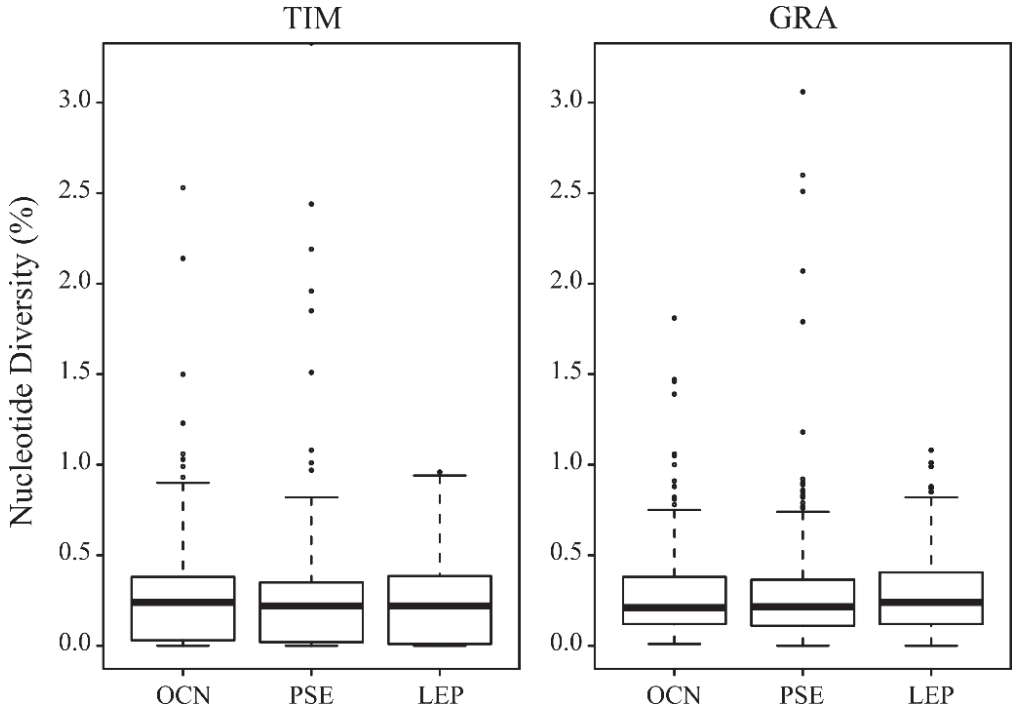


Fig. S7 – Nucleotide diversity estimates based on 500 randomly sampled 50-kb windows for each hare species (TIM – *L. timidus*; GRA – *L. granatensis*), calculated after mapping the reads on three alternative references – OCN – Rabbit; PSE – pseudo-hare; LEP – re-scaffolded hare. Only the 21 autosomes in the rabbit and pseudo-hare reference genomes and the putative autosomes in the hare reference genome were considered in the analysis. No significant differences were found between distributions (Wilcoxon rank sum test P-value > 0.05).

Supplementary tables

Table S1 - Illumina sequencing data deposited in the NCBI database.

Genomic Data (NCBI BioSample ID SAMN12621015**)**					
NCBI SRA runs accession	**Library type**	**Raw reads**	**Estimate insert size**	**Read length**	**Raw Coverage**
SRR10020059	Overlapping PE 180	212,616,812	180 ± 53	100	12.76
SRR10020059	Overlapping PE 180	207,650,218	180 ± 53	100	12.46
SRR10020059	Overlapping PE 180	208,737,561	180 ± 53	100	12.52
SRR10020059	Overlapping PE 180	212,778,842	180 ± 53	100	12.77
SRR10020059	Overlapping PE 180	210,899,916	180 ± 53	100	12.65
SRR10020057	Mate Pair 4.5kb	169,117,267	4,666 ± 1533	150	16.91
SRR10020057	Mate Pair 4.5kb	168,379,472	4,666 ± 1533	150	16.84
SRR10020056	Mate Pair 2.5kb	26,865,187	2,710 ± 471	100	1.79
SRR10020056	Mate Pair 2.5kb	31,997,329	2,710 ± 471	100	2.13
SRR10020056	Mate Pair 2.5kb	127,480,636	2,696 ± 341	125	10.62
SRR10020058	Mate Pair 8.5kb	14,284,714	8,720 ± 1519	100	0.95
SRR10020058	Mate Pair 8.5kb	17,548,328	8,122 ± 2408	100	1.17
SRR10020058	Mate Pair 8.5kb	89,754,713	8,122 ± 2408	125	7.48
Transcriptomic Data (RNA-seq**) -** Read length 100 bp					
Sample ID	**NCBI BioSample ID**	**Population**	**NCBI SRA accession**	**Raw reads**	**Tissue**
Sample_3101	SAMN06186748*	Ireland	SRR5133282	26,598,712	Liver
Sample_3102	SAMN06186761*	Ireland	SRR5133280	26,128,525	Liver
Sample_3103	SAMN06186762*	Ireland	SRR5133285	24,469,456	Liver
Sample_3104	SAMN12621015	Ireland	SRR10020060	24,526,380	Liver
Sample_3104	SAMN12621015	Ireland	SRR10020054/5	54,198,960	Skin
Sample_3105	SAMN06186763*	Ireland	SRR5133283	26,662,182	Liver
Sample_3112	SAMN06186727*	Alps	SRR5133287	22,444,667	Liver
Sample_3113	SAMN06186728*	Alps	SRR5133281	20,825,930	Liver
Sample_3114	SAMN06186729*	Alps	SRR5133286	32,749,011	Liver
Sample_3116	SAMN06186738*	Alps	SRR5133284	21,690,965	Liver

*Data from (Joao P Marques, Ferreira, et al. 2017)

Table S2 - Mapping statistics from resequencing data used in genome comparisons

		Species (Sample ID)	***L. timidus*** (SAMN07526960) *	***L. granatensis*** (SAMN07526963) *
		Cleaned reads	585,526,956	674,193,394
Reference	Rabbit genome (OCN)	Mapped **(%)**	550,980,866 (**94.1**)	637,786,951 (**94.6**)
		Uniquely Mapped **(%)**	534,451,440 (**91.3**)	609,724,325 (**90.4**)
	Hare Pseudo-Reference (PSE)	Mapped **(%)**	556,250,608 (**95.0**)	644,528,885 (**95.6**)
		Uniquely Mapped **(%)**	537,338,088 (**91.8**)	612,946,969 (**90.9**)
	Hare genome (LEP)	Mapped **(%)**	548,638,758 (**93.7**)	622,954,696 (**92.4**)
		Uniquely Mapped **(%)**	543,152,370 (**92.8**)	614,856,285 (**91.2**)

*Data from (Seixas, Boursot, and Melo-Ferreira 2018)

References

Acevedo, Pelayo, Alberto Jiménez-Valverde, José Melo-Ferreira, Raimundo Real, and Paulo Célio Alves. 2012a. 'Parapatric Species and the Implications for Climate Change Studies: A Case Study on Hares in Europe'. *Global Change Biology* 18 (5): 1509–19. https://doi.org/10.1111/j.1365-2486.2012.02655.x.

———. 2012b. 'Parapatric Species and the Implications for Climate Change Studies: A Case Study on Hares in Europe'. *Global Change Biology* 18 (5): 1509–19. https://doi.org/10.1111/j.1365-2486.2012.02655.x.

Acevedo, Pelayo, José Melo-Ferreira, Raimundo Real, and Paulo C. Alves. 2014. 'Evidence for Niche Similarities in the Allopatric Sister Species Lepus Castroviejoi and Lepus Corsicanus'. *Journal of Biogeography* 41 (5): 977–86. https://doi.org/10.1111/jbi.12270.

Altuna, Echave. 1970. 'Hallazgo de una liebre ártica (Lepus timidus L.) en el yacimiento prehistórico de Urtiaga (Guipúzcoa)', 4.

Alves, Paulo C., José Melo-Ferreira, Hélder Freitas, and Pierre Boursot. 2008. 'The Ubiquitous Mountain Hare Mitochondria: Multiple Introgressive Hybridization in Hares, Genus Lepus'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1505): 2831–39. https://doi.org/10.1098/rstb.2008.0053.

Alves, P.C., J. Melo-Ferreira, M. Branco, F. Suchentrunk, N. Ferrand, and D.J. Harris. 2008. '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): 1191–97. https://doi.org/10.1016/j.ympev.2007.11.010.

Amoutzias, Grigoris D., Themistoklis Giannoulis, Katerina A. Moutou, Anna Maria G Psarra, Costas Stamatis, Andreas Tsipourlianos, and Zissis Mamuris. 2016. 'SNP Identification through Transcriptome Analysis of the European Brown Hare (Lepus Europaeus): Cellular Energetics and Mother's Curse'. *PLoS ONE* 11 (7). https://doi.org/10.1371/journal.pone.0159939.

Andrews, Simon. 2010a. 'FastQC'. *Babraham Bioinformatics*. https://doi.org/citeulike-article-id:11583827.

———. 2010b. 'FastQC: A Quality Control Tool for High Throughput Sequence Data.' *Babraham Bioinformatics*, http://www.bioinformatics.babraham.ac.uk/projects/. https://doi.org/citeulike-article-id:11583827.

Arnegard, Matthew E., Matthew D. McGee, Blake Matthews, Kerry B. Marchinko, Gina L. Conte, Sahriar Kabir, Nicole Bedford, et al. 2014. 'Genetics of Ecological Divergence during Speciation'. *Nature* 511 (7509): 307–11. https://doi.org/10.1038/nature13301.

Arnold, Michael L. 1992. 'Natural Hybridization as an Evolutionary Process'. *Annual Review of Ecology and Systematics* 23 (1): 237–61. https://doi.org/10.1146/annurev.es.23.110192.001321.

Aubry, Sylvain, Steven Kelly, Britta M C Kümpers, Richard D. Smith-Unna, and Julian M. Hibberd. 2014. 'Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C4 Photosynthesis'. *PLoS Genetics* 10 (6). https://doi.org/10.1371/journal.pgen.1004365.

Baker, Christopher L., Shimpei Kajita, Michael Walker, Ruth L. Saxl, Narayanan Raghupathy, Kwangbom Choi, Petko M. Petkov, and Kenneth Paigen. 2015. 'PRDM9 Drives Evolutionary Erosion of Hotspots in Mus Musculus through Haplotype-Specific Initiation of Meiotic Recombination'. *PLOS Genetics* 11 (1): e1004916. https://doi.org/10.1371/journal.pgen.1004916.

Barton, N. H., and G. M. Hewitt. 1989. 'Adaptation, Speciation and Hybrid Zones'. *Nature* 341 (6242): 497–503. https://doi.org/10.1038/341497a0.

Barton, Nick, and Bengt Olle Bengtsson. 1986. 'The Barrier to Genetic Exchange between Hybridising Populations'. *Heredity* 57 (3): 357–76. https://doi.org/10.1038/hdy.1986.135.

Bateson. 1909. *Heredity and Variation in Modern Lights*. http://archive.org/details/cbarchive_36304_hereditydandvariationinmodernli1909.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop, and B. de Massy. 2010. 'PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice'. *Science (New York, N.Y.)* 327 (5967): 836–40. https://doi.org/10.1126/science.1183439.

Bayes, Joshua J., and Harmit S. Malik. 2009. 'Altered Heterochromatin Binding by a Hybrid Sterility Protein in Drosophila Sibling Species'. *Science (New York, N.Y.)* 326 (5959): 1538–41. https://doi.org/10.1126/science.1181756.

Beck, Jon A., Sarah Lloyd, Majid Hafezparast, Moyha Lennon-Pierce, Janan T. Eppig, Michael F. W. Festing, and Elizabeth M. C. Fisher. 2000. 'Genealogies of Mouse Inbred Strains'. *Nature Genetics* 24 (1): 23–25. https://doi.org/10.1038/71641.

Böhne, Astrid, Frédéric Brunet, Delphine Galiana-Arnoux, Christina Schultheis, and Jean-Nicolas Volff. 2008. 'Transposable Elements as Drivers of Genomic and Biological Diversity in Vertebrates'. *Chromosome Research* 16 (1): 203–15. https://doi.org/10.1007/s10577-007-1202-6.

Bolger, a. M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data'. *Bioinformatics* 30 (15): 2114–20.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data'. *Bioinformatics (Oxford, England)* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Boursot, P., J.-C. Auffray, J. Britton-Davidian, and F. Bonhomme. 1993. 'The Evolution of House Mice'. *Annual Review of Ecology and Systematics* 24: 119–52.

Boutet, Emmanuel, Damien Lieberherr, Michael Tognolli, Michel Schneider, Parit Bansal, Alan J Bridge, Sylvain Poux, Lydie Bougueleret, and Ioannis Xenarios. 2016. 'UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View.' *Methods in Molecular Biology (Clifton, N.J.)* 1374: 23–54. https://doi.org/10.1007/978-1-4939-3167-5_2.

Brandvain, Yaniv, Amanda M. Kenney, Lex Flagel, Graham Coop, and Andrea L. Sweigart. 2014. 'Speciation and Introgression between Mimulus Nasutus and Mimulus Guttatus'. Edited by Chris D. Jiggins. *PLoS Genetics* 10 (6): e1004410. https://doi.org/10.1371/journal.pgen.1004410.

Brennecke, Julius, Colin D. Malone, Alexei A. Aravin, Ravi Sachidanandam, Alexander Stark, and Gregory J. Hannon. 2008. 'An Epigenetic Role for Maternally Inherited PiRNAs in Transposon Silencing'. *Science (New York, N.Y.)* 322 (5906): 1387–92. https://doi.org/10.1126/science.1165171.

Brideau, Nicholas J., Heather A. Flores, Jun Wang, Shamoni Maheshwari, Xu Wang, and Daniel A. Barbash. 2006. 'Two Dobzhansky-Muller Genes Interact to Cause Hybrid Lethality in Drosophila'. *Science (New York, N.Y.)* 314 (5803): 1292–95. https://doi.org/10.1126/science.1133953.

Brown, Judith D., and Rachel J. O'Neill. 2010. 'Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited'. *Annual Review of Genomics and Human Genetics* 11 (1): 291–316. https://doi.org/10.1146/annurev-genom-082509-141554.

Brown, Judith D., Vanessa Piccuillo, and Rachel J. O'Neill. 2012. 'Retroelement Demethylation Associated with Abnormal Placentation in Mus Musculus × Mus Caroli Hybrids1'. *Biology of Reproduction* 86 (3): 88, 1–10. https://doi.org/10.1095/biolreprod.111.095273.

Burt, Austin, and Robert Trivers. 2006. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Cambridge, Mass: Belknap Press of Harvard University Press.

Butlin, Roger K, and Sean Stankowski. 2020. 'Is It Time to Abandon the Biological Species Concept? No'. *National Science Review* 7 (8): 1400–1401. https://doi.org/10.1093/nsr/nwaa109.

Calfee, Erin, Daniel Gates, Anne Lorant, M. Taylor Perkins, Graham Coop, and Jeffrey Ross-Ibarra. 2021. 'Selective Sorting of Ancestral Introgression in Maize and Teosinte along an Elevational Cline'. https://doi.org/10.1101/2021.03.05.434040.

Camus, M. Florencia, Jochen B. W. Wolf, Edward H. Morrow, and Damian K. Dowling. 2015. 'Single Nucleotides in the MtDNA Sequence Modify Mitochondrial Molecular Function and Are Associated with Sex-Specific Effects on Fertility and Aging'. *Current Biology* 25 (20): 2717–22. https://doi.org/10.1016/j.cub.2015.09.012.

Caravaggi, Anthony, Katie Leach, Francesco Santilli, Jukka Rintala, Pekka Helle, Juha Tiainen, Francesco Bisi, Adriano Martinoli, W Ian Montgomery, and Neil Reid. 2016. 'Niche Overlap of Mountain Hare Subspecies and the Vulnerability of Their Ranges to Invasion by the European Hare; the (Bad) Luck of the Irish'. *Biological Invasions*, 1–20. https://doi.org/10.1007/s10530-016-1330-z.

Caravaggi, Anthony, Katie Leach, Francesco Santilli, Jukka Rintala, Pekka Helle, Juha Tiainen, Francesco Bisi, Adriano Martinoli, W. Ian Montgomery, and Neil Reid. 2017. 'Niche Overlap of Mountain Hare Subspecies and the Vulnerability of Their Ranges to Invasion by the European Hare; the (Bad) Luck of the Irish'. *Biological Invasions* 19 (2): 655–74. https://doi.org/10.1007/s10530-016-1330-z.

Caravaggi, Anthony, W Ian Montgomery, and Neil Reid. 2015. 'Range Expansion and Comparative Habitat Use of Insular, Congeneric Lagomorphs: Invasive European Hares Lepus Europaeus and Endemic Irish Hares Lepus Timidus Hibernicus'. *Biological Invasions* 17 (2): 687–98. https://doi.org/10.1007/s10530-014-0759-1.

Case, Andrea L., Findley R. Finseth, Camille M. Barr, and Lila Fishman. 2016. 'Selfish Evolution of Cytonuclear Hybrid Incompatibility in Mimulus'. *Proceedings of the Royal Society B: Biological Sciences* 283 (1838): 20161493. https://doi.org/10.1098/rspb.2016.1493.

Chafee, Meghan E., Courtney N. Zecher, Michelle L. Gourley, Victor T. Schmidt, John H. Chen, Sarah R. Bordenstein, Michael E. Clark, and Seth R. Bordenstein. 2011. 'Decoupling of Host–Symbiont–Phage Coadaptations Following Transfer Between Insect Species'. *Genetics* 187 (1): 203–15. https://doi.org/10.1534/genetics.110.120675.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. 'The Effect of Deleterious Mutations on Neutral Molecular Variation'. *Genetics* 134 (4): 1289–1303.

Chase, Kevin, David R. Carrier, Frederick R. Adler, Elaine A. Ostrander, and Karl G. Lark. 2005. 'Interaction between the X Chromosome and an Autosome Regulates Size Sexual Dimorphism in Portuguese Water Dogs'. *Genome Research* 15 (12): 1820–24. https://doi.org/10.1101/gr.3712705.

Christie, Kyle, and Sharon Y. Strauss. 2018. 'Along the Speciation Continuum: Quantifying Intrinsic and Extrinsic Isolating Barriers across Five Million Years of Evolutionary Divergence in California Jewelflowers'. *Evolution; International Journal of Organic Evolution* 72 (5): 1063–79. https://doi.org/10.1111/evo.13477.

Christie, Paul, and Mark R. Macnair. 1987. 'The Distribution of Postmating Reproductive Isolating Genes in Populations of the Yellow Monkey Flower, Mimulus Guttatus'. *Evolution* 41 (3): 571–78. https://doi.org/10.1111/j.1558-5646.1987.tb05827.x.

Consortium, Heliconius Genome. 2012. 'Butterfly Genome Reveals Promiscuous Exchange of Mimicry Adaptations among Species.' *Nature* 487 (7405): 94–98. https://doi.org/10.1038/nature11041.

Corbett-Detig, Russell, and Rasmus Nielsen. 2017. 'A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next

Generation Sequence Data in Samples of Arbitrary Ploidy'. *PLOS Genetics* 13 (1): e1006529. https://doi.org/10.1371/journal.pgen.1006529.

Courret, Cécile, Ching-Ho Chang, Kevin H.-C. Wei, Catherine Montchamp-Moreau, and Amanda M. Larracuente. 2019. 'Meiotic Drive Mechanisms: Lessons from Drosophila'. *Proceedings of the Royal Society B: Biological Sciences* 286 (1913): 20191430. https://doi.org/10.1098/rspb.2019.1430.

Coyne, Jerry A., and H. Allen Orr. 1997. '"Patterns of Speciation in Drosophila" Revisited'. *Evolution* 51 (1): 295–303. https://doi.org/10.2307/2410984.

———. 2004. *Speciation*. Oxford, New York: Oxford University Press.

Crespi, Bernard. 2008. 'Turner Syndrome and the Evolution of Human Sexual Dimorphism'. *Evolutionary Applications* 1 (3): 449–61. https://doi.org/10.1111/j.1752-4571.2008.00017.x.

Crespi, Bernard, and Patrik Nosil. 2013. 'Conflictual Speciation: Species Formation via Genomic Conflict'. *Trends in Ecology & Evolution* 28 (1): 48–57. https://doi.org/10.1016/j.tree.2012.08.015.

Cruickshank, Tami E., and Matthew W. Hahn. 2014. 'Reanalysis Suggests That Genomic Islands of Speciation Are Due to Reduced Diversity, Not Reduced Gene Flow.' *Molecular Ecology* 23 (13): 3133–57. https://doi.org/10.1111/mec.12796.

Cucchi, Thomas, Jean-Christophe Auffray, and Jean-Denis Vigne. 2012. 'On the Origin of the House Mouse Synanthropy and Dispersal in the Near East and Europe: Zooarchaeological Review and Perspectives'. In , 65–93.

Cucchi, Thomas, Jean-Denis Vigne, and Jean-Christophe Auffray. 2005. 'First Occurrence of the House Mouse (Mus Musculus Domesticus Schwarz & Schwarz, 1943) in the Western Mediterranean: A Zooarchaeological Revision of Subfossil Occurrences'. *Biological Journal of the Linnean Society* 84 (3): 429–45. https://doi.org/10.1111/j.1095-8312.2005.00445.x.

Currat, Mathias, Manuel Ruedi, Rémy J. Petit, and Laurent Excoffier. 2008. 'The Hidden Side of Invasions: Massive Introgression by Local Genes'. *Evolution* 62 (8): 1908–20. https://doi.org/10.1111/j.1558-5646.2008.00413.x.

Darwin, Charles, and Leonard Kebler. 1859. *On the Origin of Species by Means of Natural Selection*. https://www.loc.gov/item/06017473/.

Delmore, Kira E., and Darren E. Irwin. 2014. 'Hybrid Songbirds Employ Intermediate Routes in a Migratory Divide'. *Ecology Letters* 17 (10): 1211–18. https://doi.org/10.1111/ele.12326.

Din, W., R. Anand, P. Boursot, D. Darviche, B. Dod, E. Jouvin-Marche, A. Orth, G. P. Talwar, P.-A. Cazenave, and F. Bonhomme. 1996. 'Origin and Radiation of the House Mouse: Clues from Nuclear Genes'. *Journal of Evolutionary Biology* 9 (5): 519–39. https://doi.org/10.1046/j.1420-9101.1996.9050519.x.

Dobzhansky, Theodosius, and Stephen Jay Gould. 1982. *Genetics and the Origin of Species*. New York: Columbia University Press.

Dod, Barbara, Lars S. Jermiin, Pierre Boursot, Verne H. Chapman, J. Tönnes Nielsen, and François Bonhomme. 1993. 'Counterselection on Sex Chromosomes in the Mus Musculus European Hybrid Zone'. *Journal of Evolutionary Biology* 6 (4): 529–46. https://doi.org/10.1046/j.1420-9101.1993.6040529.x.

Dumont, Beth L., and Bret A. Payseur. 2011. 'Genetic Analysis of Genome-Scale Recombination Rate Evolution in House Mice'. *PLoS Genet* 7 (6): e1002116.

Duplantier, J.-M., A. Orth, J. Catalan, and F. Bonhomme. 2002. 'Evidence for a Mitochondrial Lineage Originating from the Arabian Peninsula in the Madagascar House Mouse (Mus Musculus)'. *Heredity* 89 (2): 154–58. https://doi.org/10.1038/sj.hdy.6800122.

Duranton, Maud, and John Pool. 2021. 'Interactions between Natural Selection and Recombination Shape the Genomic Landscape of Introgression'. Preprint. Evolutionary Biology. https://doi.org/10.1101/2021.10.12.462572.

Duvaux, Ludovic, Khalid Belkhir, Matthieu Boulesteix, and Pierre Boursot. 2011. 'Isolation and Gene Flow: Inferring the Speciation History of European House Mice: HOUSE MOUSE SPECIATION HISTORY'. *Molecular Ecology* 20 (24): 5248–64. https://doi.org/10.1111/j.1365-294X.2011.05343.x.

Duvernell, David D., Eric Westhafer, and Jacob F. Schaefer. 2019. 'Late Pleistocene Range Expansion of North American Topminnows Accompanied by Admixture and Introgression'. *Journal of Biogeography* 46 (9): 2126–40. https://doi.org/10.1111/jbi.13597.

Edmonds, Christopher A., Anita S. Lillie, and L. Luca Cavalli-Sforza. 2004. 'Mutations Arising in the Wave Front of an Expanding Population'. *Proceedings of the National Academy of Sciences* 101 (4): 975–79. https://doi.org/10.1073/pnas.0308064100.

El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R Eddy, Aurélien Luciani, Simon C Potter, Matloob Qureshi, et al. 2019. 'The Pfam Protein Families Database in 2019'. *Nucleic Acids Research* 47 (D1): D427–32. https://doi.org/10.1093/nar/gky995.

Elgvin, Tore O., Cassandra N. Trier, Ole K. Tørresen, Ingerid J. Hagen, Sigbjørn Lien, Alexander J. Nederbragt, Mark Ravinet, Henrik Jensen, and Glenn-Peter Sætre. 2017. 'The Genomic Mosaicism of Hybrid Speciation'. *Science Advances* 3 (6): e1602996. https://doi.org/10.1126/sciadv.1602996.

Enk, Jacob, Alison Devault, Regis Debruyne, Christine E. King, Todd Treangen, Dennis O'Rourke, Steven L. Salzberg, Daniel Fisher, Ross MacPhee, and Hendrik Poinar. 2011. 'Complete Columbian Mammoth Mitogenome Suggests Interbreeding with Woolly Mammoths'. *Genome Biology* 12 (5): R51. https://doi.org/10.1186/gb-2011-12-5-r51.

Enk, Jacob, Alison Devault, Christopher Widga, Jeffrey Saunders, Paul Szpak, John Southon, Jean-Marie Rouillard, et al. 2016. 'Mammuthus Population Dynamics in Late Pleistocene North America: Divergence, Phylogeography, and Introgression'. *Frontiers in Ecology and Evolution* 4: 42. https://doi.org/10.3389/fevo.2016.00042.

Estonba, Andone, Ainhoa Solís, M. Iriondo, María J. Sanz-Martín, G. Pérez-Suárez, G. Markov, and F. Palacios. 2006. 'The Genetic Distinctiveness of the Three Iberian Hare Species: Lepus Europaeus, L. Granatensis, and L. Castroviejoi'. *Mammalian Biology* 71 (1): 52–59. https://doi.org/10.1016/j.mambio.2005.08.010.

Excoffier, Laurent, Matthieu Foll, and Rémy J. Petit. 2009. 'Genetic Consequences of Range Expansions'. *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414.

Excoffier, Laurent, and Nicolas Ray. 2008. 'Surfing during Population Expansions Promotes Genetic Revolutions and Structuration'. *Trends in Ecology & Evolution* 23 (7): 347–51. https://doi.org/10.1016/j.tree.2008.04.004.

Ferreira, Mafalda S, Matthew R Jones, Colin M Callahan, Liliana Farelo, Zelalem Tolesa, Franz Suchentrunk, Pierre Boursot, et al. 2021. 'The Legacy of Recurrent Introgression during the Radiation of Hares'. *Systematic Biology* 70 (3): 593–607. https://doi.org/10.1093/sysbio/syaa088.

Finn, Robert D., Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Andreas Heger, et al. 2014. 'Pfam: The Protein Families Database'. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt1223.

Finn, Robert D., Jody Clements, and Sean R. Eddy. 2011. 'HMMER Web Server: Interactive Sequence Similarity Searching'. *Nucleic Acids Research* 39 (SUPPL. 2). https://doi.org/10.1093/nar/gkr367.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. The Genetical Theory of Natural Selection. Oxford, England: Clarendon Press. https://doi.org/10.5962/bhl.title.27468.

Fishman, Lila, and John H. Willis. 2006. 'A CYTONUCLEAR INCOMPATIBILITY CAUSES ANTHER STERILITY IN MIMULUS HYBRIDS'. *Evolution* 60 (7): 1372–81. https://doi.org/10.1554/05-708.1.

Fontaine, Michael C., James B. Pease, Aaron Steele, Robert M. Waterhouse, Daniel E. Neafsey, Igor V. Sharakhov, Xiaofang Jiang, et al. 2015. 'Extensive Introgression in a Malaria Vector Species Complex Revealed by Phylogenomics'. *Science (New York, N.Y.)* 347 (6217): 1258524. https://doi.org/10.1126/science.1258524.

Forejt, J., and P. Iványi. 1974. 'Genetic Studies on Male Sterility of Hybrids between Laboratory and Wild Mice (Mus Musculus L.)'. *Genetics Research* 24 (2): 189–206. https://doi.org/10.1017/S0016672300015214.

Fraïsse, Christelle, P. Alexander Gunnarsson, Denis Roze, Nicolas Bierne, and John J. Welch. 2016. 'The Genetics of Speciation: Insights from Fisher's Geometric Model'. *Evolution* 70 (7): 1450–64. https://doi.org/10.1111/evo.12968.

Frank, Steven A., and Bernard J. Crespi. 2011. 'Pathology from Evolutionary Conflict, with a Theory of X Chromosome versus Autosome Conflict over Sexually Antagonistic Traits'. *Proceedings of the National Academy of Sciences* 108 (Supplement 2): 10886–93. https://doi.org/10.1073/pnas.1100921108.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012a. 'CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data'. *Bioinformatics* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

———. 2012b. 'CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data'. *Bioinformatics* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

Fujiwara, Kazumichi, Yosuke Kawai, Kazuo Moriwaki, Toyoyuki Takada, Toshihiko Shiroishi, Naruya Saitou, Hitoshi Suzuki, and Naoki Osada. 2021. 'Insights into Mus Musculus Subspecies Population Structure across Eurasia Revealed by Whole-Genome Sequence Analysis'. https://doi.org/10.1101/2021.02.05.429881.

Gavrilets, Sergey. 2003. 'Perspective: Models of Speciation: What Have We Learned in 40 Years?' *Evolution* 57 (10): 2197–2215. https://doi.org/10.1111/j.0014-3820.2003.tb00233.x.

Gayral, Philippe, José Melo-Ferreira, Sylvain Glémin, Nicolas Bierne, Miguel Carneiro, Benoit Nabholz, Joao M. Lourenco, et al. 2013. 'Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate-Invertebrate Gap'. *PLoS Genetics* 9 (4). https://doi.org/10.1371/journal.pgen.1003457.

Gene Ontology Consortium. 2000. 'The Gene Ontology Consortium. Gene Ontology: Tool for the Unification of Biology'. *Nature Genetics* 25 (1): 25–29. https://doi.org/10.1038/75556.

Geraldes, A., P. Basset, K. L. Smith, and M. W. Nachman. 2011. 'Higher Differentiation among Subspecies of the House Mouse (Mus Musculus) in Genomic Regions with Low Recombination: RECOMBINATION AND SPECIATION IN MICE'. *Molecular Ecology* 20 (22): 4722–36. https://doi.org/10.1111/j.1365-294X.2011.05285.x.

Geraldes, Armando, Patrick Basset, Barbara Gibson, Kimberly L. Smith, Bettina Harr, Hon-Tsen Yu, Nina Bulatova, Yaron Ziv, and Michael W. Nachman. 2008. 'Inferring the History of Speciation in House Mice from Autosomal, X-Linked, Y-Linked and Mitochondrial Genes'. *Molecular Ecology* 17 (24): 5349–63. https://doi.org/10.1111/j.1365-294X.2008.04005.x.

Giska, Iwona, Liliana Farelo, João Pimenta, Fernando A. Seixas, Mafalda S. Ferreira, João P. Marques, Inês Miranda, et al. 2019. 'Introgression Drives Repeated Evolution of Winter Coat Color Polymorphism in Hares'. *Proceedings of the National Academy of Sciences*, November. https://doi.org/10.1073/pnas.1910471116.

Giska, Iwona, João Pimenta, Liliana Farelo, Pierre Boursot, Klaus Hackländer, Hannes Jenny, Neil Reid, et al. 2022. 'The Evolutionary Pathways for Local Adaptation in Mountain Hares'. *Molecular Ecology* n/a (n/a). https://doi.org/10.1111/mec.16338.

Good, Jeffrey M., Matthew D. Dean, and Michael W. Nachman. 2008. 'A Complex Genetic Basis to X-Linked Hybrid Male Sterility Between Two Species of House Mice'. *Genetics* 179 (4): 2213–28. https://doi.org/10.1534/genetics.107.085340.

Good, Jeffrey M., Thomas Giger, Matthew D. Dean, and Michael W. Nachman. 2010. 'Widespread Over-Expression of the X Chromosome in Sterile F1 Hybrid Mice'. *PLOS Genetics* 6 (9): e1001148. https://doi.org/10.1371/journal.pgen.1001148.

Gow, J. L., C. L. Peichel, and E. B. Taylor. 2007. 'Ecological Selection against Hybrids in Natural Populations of Sympatric Threespine Sticklebacks'. *Journal of Evolutionary Biology* 20 (6): 2173–80. https://doi.org/10.1111/j.1420-9101.2007.01427.x.

Grabherr, Manfred G, Brian J Haas, Moran Yassour, Joshua Z Levin, Dawn A Thompson, Ido Amit, Xian Adiconis, et al. 2011a. 'Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome.' *Nature Biotechnology* 29 (7): 644–52. https://doi.org/10.1038/nbt.1883.

———. 2011b. 'Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome'. *Nature Biotechnology* 29 (May): 644.

Guénet, Jean-Louis, and François Bonhomme. 2003. 'Wild Mice: An Ever-Increasing Contribution to a Popular Mammalian Model'. *Trends in Genetics* 19 (1): 24–31. https://doi.org/10.1016/S0168-9525(02)00007-0.

Haas & Papanicolaou et al. (2015) 2019. *TransDecoder Source. Contribute to TransDecoder/TransDecoder Development by Creating an Account on GitHub*. Perl. TransDecoder. https://github.com/TransDecoder/TransDecoder.

Haas, Brian J, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. 'De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis.' *Nature Protocols* 8 (8): 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hamid, Haddadian Shad, Jamshid Darvish, Eskandar Rastegar-Pouyani, and Ahmad Mahmoudi. 2017. 'Subspecies Differentiation of the House Mouse Mus Musculus Linnaeus, 1758 in the Center and East of the Iranian Plateau and Afghanistan'. *Mammalia* 81 (2). https://doi.org/10.1515/mammalia-2015-0041.

Hamill, R M, D Doyle, and E J Duke. 2006. 'Spatial Patterns of Genetic Diversity across European Subspecies of the Mountain Hare, Lepus Timidus L.' *Heredity* 97 (5): 355–65. https://doi.org/10.1038/sj.hdy.6800880.

Hardouin, Emilie A, Annie Orth, Meike Teschke, Jamshid Darvish, Diethard Tautz, and François Bonhomme. 2015. 'Eurasian House Mouse (Mus Musculus L.) Differentiation at Microsatellite Loci Identifies the Iranian Plateau as a Phylogeographic Hotspot'. *BMC Evolutionary Biology* 15 (February). https://doi.org/10.1186/s12862-015-0306-4.

Harr, Bettina, Emre Karakoc, Rafik Neme, Meike Teschke, Christine Pfeifle, Željka Pezer, Hiba Babiker, et al. 2016. 'Genomic Resources for Wild Populations of the House Mouse, Mus Musculus and Its Close Relative Mus Spretus'. *Scientific Data* 3 (1): 1–14. https://doi.org/10.1038/sdata.2016.75.

Harris, Kelley, and Rasmus Nielsen. 2016. 'The Genetic Cost of Neanderthal Introgression'. *Genetics* 203 (2): 881–91. https://doi.org/10.1534/genetics.116.186890.

Harrison, D. L. 1970. 'SOME OBSERVATIONS ON THE STATUS OF THE SOUTH ARABIAN HOUSE MOUSE MUS MUSCULUS GENTILULUS THOMAS, 1919. (RODENTIA : MURIDAE)' 34 (2): 244–47. https://doi.org/10.1515/mamm.1970.34.2.244.

Hermansen, Jo S., Stein A. Sæther, Tore O. Elgvin, Thomas Borge, Elin Hjelle, and Glenn-Peter Sætre. 2011. 'Hybrid Speciation in Sparrows I: Phenotypic Intermediacy, Genetic Admixture and Barriers to Gene Flow'. *Molecular Ecology* 20 (18): 3812–22. https://doi.org/10.1111/j.1365-294X.2011.05183.x.

Hessenauer, Pauline, Anna Fijarczyk, Hélène Martin, Julien Prunier, Guillaume Charron, Jérôme Chapuis, Louis Bernier, Philippe Tanguay, Richard C. Hamelin, and Christian R. Landry. 2020. 'Hybridization and Introgression Drive Genome Evolution of Dutch Elm Disease Pathogens'. *Nature Ecology & Evolution* 4 (4): 626–38. https://doi.org/10.1038/s41559-020-1133-6.

Huerta-Sánchez, Emilia, Xin Jin, Asan, Zhuoma Bianba, Benjamin M. Peter, Nicolas Vinckenbosch, Yu Liang, et al. 2014. 'Altitude Adaptation in Tibetans Caused by Introgression of Denisovan-like DNA'. *Nature* 512 (7513): 194–97. https://doi.org/10.1038/nature13408.

Hughes, Maria, Neil Reid, Ian Montgomery, and Paulo Prodoehl. 2011. 'Verification of Hybridisation between Introduced European and Native Irish Hares'. *Northern Ireland Environment Agency Research and Development SeriesNorthern Ireland Environment Agency Research and Development Series* 11–11.

Hui, Ling, Jing Lu, Yibing Han, and Stephen H. Pilder. 2006. 'The Mouse T Complex Gene Tsga2, Encoding Polypeptides Located in the Sperm Tail and Anterior Acrosome, Maps to a Locus Associated with Sperm Motility and Sperm-Egg Interaction Abnormalities'. *Biology of Reproduction* 74 (4): 633–43. https://doi.org/10.1095/biolreprod.105.045963.

IUCN. 2016. 'IUCN Red List of Threatened Species'. *Version 2016-1*, www.iucnredlist.org.

Janoušek, Václav, Pavel Munclinger, Liuyang Wang, Katherine C. Teeter, and Priscilla K. Tucker. 2015a. 'Functional Organization of the Genome May Shape the Species Boundary in the House Mouse'. *Molecular Biology and Evolution* 32 (5): 1208–20. https://doi.org/10.1093/molbev/msv011.

———. 2015b. 'Functional Organization of the Genome May Shape the Species Boundary in the House Mouse'. *Molecular Biology and Evolution* 32 (5): 1208–20. https://doi.org/10.1093/molbev/msv011.

Janoušek, Václav, LIUYANG WANG, KEN LUZYNSKI, PETRA DUFKOVÁ, MARTINA M. VYSKOČILOVÁ, MICHAEL W. NACHMAN, PAVEL MUNCLINGER, MILOŠ MACHOLÁN, JAROSLAV PIÁLEK, and PRISCILLA K. TUCKER. 2012. 'Genome-Wide Architecture of Reproductive Isolation in a Naturally Occurring Hybrid Zone between Mus Musculus Musculus and M. m. Domesticus'. *Molecular Ecology* 21 (12): 10.1111/j.1365-294X.2012.05583.x. https://doi.org/10.1111/j.1365-294X.2012.05583.x.

Jones, Eleanor P., JEROEN Van Der Kooij, Roar Solheim, and Jeremy B. Searle. 2010. 'Norwegian House Mice (Mus Musculus Musculus/Domesticus): Distributions, Routes of Colonization and Patterns of Hybridization'. *Molecular Ecology* 19 (23): 5252–64. https://doi.org/10.1111/j.1365-294X.2010.04874.x.

Jones, Matthew R., L. Scott Mills, Paulo Célio Alves, Colin M. Callahan, Joel M. Alves, Diana J. R. Lafferty, Francis M. Jiggins, Jeffrey D. Jensen, José Melo-Ferreira, and Jeffrey M. Good. 2018. 'Adaptive Introgression Underlies Polymorphic

Seasonal Camouflage in Snowshoe Hares'. *Science* 360 (6395): 1355–58. https://doi.org/10.1126/science.aar5273.

Juric, Ivan, Simon Aeschbacher, and Graham Coop. 2016. 'The Strength of Selection against Neanderthal Introgression'. *PLOS Genetics* 12 (11): e1006340. https://doi.org/10.1371/journal.pgen.1006340.

Jurka, Jerzy, Weidong Bao, and Kenji K Kojima. 2011. 'Families of Transposable Elements, Population Structure and the Origin of Species'. *Biology Direct* 6 (1): 44. https://doi.org/10.1186/1745-6150-6-44.

Kanehisa, Minoru, Susumu Goto, Yoko Sato, Miho Furumichi, and Mao Tanabe. 2012. 'KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets'. *Nucleic Acids Research* 40 (D1). https://doi.org/10.1093/nar/gkr988.

Khurana, Jaspreet S., Jie Wang, Jia Xu, Birgit S. Koppetsch, Travis C. Thomson, Anetta Nowosielska, Chengjian Li, Phillip D. Zamore, Zhiping Weng, and William E. Theurkauf. 2011. 'Adaptation to P Element Transposon Invasion in Drosophila Melanogaster'. *Cell* 147 (7): 1551–63. https://doi.org/10.1016/j.cell.2011.11.042.

Kirkpatrick, Mark, and Virginie Ravigné. 2002. 'Speciation by Natural and Sexual Selection: Models and Experiments.' *The American Naturalist* 159 (S3): S22–35. https://doi.org/10.1086/338370.

Klopfstein, Seraina, Mathias Currat, and Laurent Excoffier. 2006. 'The Fate of Mutations Surfing on the Wave of a Range Expansion'. *Molecular Biology and Evolution* 23 (3): 482–90. https://doi.org/10.1093/molbev/msj057.

Krogh, a, B Larsson, G von Heijne, and E L Sonnhammer. 2001. 'Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes.' *Journal of Molecular Biology* 305 (3): 567–80. https://doi.org/10.1006/jmbi.2000.4315.

Kunte, Krushnamegh, Cristina Shea, Matthew L. Aardema, J. Mark Scriber, Thomas E. Juenger, Lawrence E. Gilbert, and Marcus R. Kronforst. 2011. 'Sex Chromosome Mosaicism and Hybrid Speciation among Tiger Swallowtail Butterflies'. *PLOS Genetics* 7 (9): e1002274. https://doi.org/10.1371/journal.pgen.1002274.

Langdon, Quinn K., David Peris, Juan I. Eizaguirre, Dana A. Opulente, Kelly V. Buh, Kayla Sylvester, Martin Jarzyna, et al. 2020. 'Postglacial Migration Shaped the Genomic Diversity and Global Distribution of the Wild Ancestor of Lager-Brewing Hybrids'. *PLOS Genetics* 16 (4): e1008680. https://doi.org/10.1371/journal.pgen.1008680.

Larson, Erica L., Dan Vanderpool, Brice A. J. Sarver, Colin Callahan, Sara Keeble, Lorraine P. Provencio, Michael D. Kessler, et al. 2018. 'The Evolution of Polymorphic Hybrid Incompatibilities in House Mice'. *Genetics*, April, genetics.300840.2018. https://doi.org/10.1534/genetics.118.300840.

Lawson, Daniel J., Lucy van Dorp, and Daniel Falush. 2018. 'A Tutorial on How Not to Over-Interpret STRUCTURE and ADMIXTURE Bar Plots'. *Nature Communications* 9 (1): 3258. https://doi.org/10.1038/s41467-018-05257-7.

Leach, Katie, Ruth Kelly, Alison Cameron, W. Ian Montgomery, and Neil Reid. 2015. 'Expertly Validated Models and Phylogenetically-Controlled Analysis Suggests Responses to Climate Change Are Related to Species Traits in the Order Lagomorpha'. *PLoS ONE* 10 (4). https://doi.org/10.1371/journal.pone.0122267.

Leach, Katie, W. Ian Montgomery, and Neil Reid. 2015. 'Biogeography, Macroecology and Species' Traits Mediate Competitive Interactions in the Order Lagomorpha'. *Mammal Review*. https://doi.org/10.1111/mam.12035.

Levänen, Riikka, Carl-Gustaf Thulin, Göran Spong, and Jaakko L. O. Pohjoismäki. 2018. 'Widespread Introgression of Mountain Hare Genes into Fennoscandian Brown Hare Populations'. *PLOS ONE* 13 (1): e0191790. https://doi.org/10.1371/journal.pone.0191790.

Li, Heng, and Richard Durbin. 2010. 'Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform'. *Bioinformatics* 26 (5): 589–95. https://doi.org/10.1093/bioinformatics/btp698.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Lindholm, Anna K, Kerstin Musolf, Andrea Weidt, and Barbara König. 2013. 'Mate Choice for Genetic Compatibility in the House Mouse'. *Ecology and Evolution* 3 (5): 1231–47. https://doi.org/10.1002/ece3.534.

Liu, Kevin J., Ethan Steinberg, Alexander Yozzo, Ying Song, Michael H. Kohn, and Luay Nakhleh. 2015. 'Interspecific Introgressive Origin of Genomic Diversity in the House Mouse'. *Proceedings of the National Academy of Sciences* 112 (1): 196–201. https://doi.org/10.1073/pnas.1406298111.

Macholán, Miloš, Stuart J. E. Baird, Alena Fornůsková, Iva Martincová, Pavel Rubík, Ľudovít Ďureje, Emanuel Heitlinger, and Jaroslav Piálek. 2019. 'Widespread Introgression of the *Mus Musculus Musculus* Y Chromosome in Central Europe'. Preprint. Evolutionary Biology. https://doi.org/10.1101/2019.12.23.887471.

Macholán, Miloš, Stuart J. E. Baird, Pavel Munclinger, and Jaroslav Piálek. 2012. *Evolution of the House Mouse*. Cambridge University Press.

Macholán, Miloš, Stuart JE Baird, Pavel Munclinger, Petra Dufková, Barbora Bímová, and Jaroslav Piálek. 2008a. 'Genetic Conflict Outweighs Heterogametic Incompatibility in the Mouse Hybrid Zone?' *BMC Evolutionary Biology* 8 (1): 271. https://doi.org/10.1186/1471-2148-8-271.

———. 2008b. 'Genetic Conflict Outweighs Heterogametic Incompatibility in the Mouse Hybrid Zone?' *BMC Evolutionary Biology* 8 (1): 271. https://doi.org/10.1186/1471-2148-8-271.

Macholán, Miloš, Boris Kryštufek, and Vladimír Vohralík. 2003. 'The Location of TheMus Musculus/M. Domesticus Hybrid Zone in the Balkans: Clues from Morphology'. *Acta Theriologica* 48 (2): 177–88. https://doi.org/10.1007/BF03194157.

Maheshwari, Shamoni, and Daniel A. Barbash. 2011. 'The Genetics of Hybrid Incompatibilities'. *Annual Review of Genetics* 45 (1): 331–55. https://doi.org/10.1146/annurev-genet-110410-132514.

Mallet, James, Margarita Beltrán, Walter Neukirchen, and Mauricio Linares. 2007. 'Natural Hybridization in Heliconiine Butterflies: The Species Boundary as a Continuum'. *BMC Evolutionary Biology* 7 (1): 28. https://doi.org/10.1186/1471-2148-7-28.

Mallet, James, Nora Besansky, and Matthew W. Hahn. 2016. 'How Reticulated Are Species?' *BioEssays* 38 (2): 140–49. https://doi.org/10.1002/bies.201500149.

Mallet, James, and Mathieu Joron. 1999. 'Evolution of Diversity in Warning Color and Mimicry: Polymorphisms, Shifting Balance, and Speciation'. *Annual Review of Ecology and Systematics* 30: 201–33.

Marques, Joao P, Liliana Farelo, Joana Vilela, Dan Vanderpool, Paulo C Alves, Jeffrey M Good, Pierre Boursot, and Jose Melo-Ferreira. 2017. 'Range Expansion Underlies Historical Introgressive Hybridization in the Iberian Hare.' *Scientific Reports* 7 (January): 40788. https://doi.org/10.1038/srep40788.

Marques, João P., Liliana Farelo, Joana Vilela, Dan Vanderpool, Paulo C. Alves, Jeffrey M. Good, Pierre Boursot, and José Melo-Ferreira. 2017. 'Range Expansion Underlies Historical Introgressive Hybridization in the Iberian Hare'. *Scientific Reports* 7 (1): 40788. https://doi.org/10.1038/srep40788.

Marques, João P., Mafalda S. Ferreira, Liliana Farelo, Colin M. Callahan, Klaus Hackländer, Hannes Jenny, W. Ian Montgomery, et al. 2017. 'Mountain Hare Transcriptome and Diagnostic Markers as Resources to Monitor Hybridization

with European Hares'. *Scientific Data* 4 (1): 170178. https://doi.org/10.1038/sdata.2017.178.

Marques, Joao P, Mafalda S Ferreira, Liliana Farelo, Colin M Callahan, Klaus Hacklander, Hannes Jenny, W Ian Montgomery, et al. 2017. 'Mountain Hare Transcriptome and Diagnostic Markers as Resources to Monitor Hybridization with European Hares.' *Scientific Data* 4 (December): 170178–170178. https://doi.org/10.1038/sdata.2017.178.

Marques, João P., Fernando A. Seixas, Liliana Farelo, Colin M. Callahan, Jeffrey M. Good, W. Ian Montgomery, Neil Reid, Paulo C. Alves, Pierre Boursot, and José Melo-Ferreira. 2020. 'An Annotated Draft Genome of the Mountain Hare (Lepus Timidus)'. *Genome Biology and Evolution* 12 (1): 3656–62. https://doi.org/10.1093/gbe/evz273.

Martienssen, Robert A. 2010. 'Heterochromatin, Small RNA and Post-Fertilization Dysgenesis in Allopolyploid and Interploid Hybrids of Arabidopsis'. *The New Phytologist* 186 (1): 46–53. https://doi.org/10.1111/j.1469-8137.2010.03193.x.

Martin, Simon H., Kanchon K. Dasmahapatra, Nicola J. Nadeau, Camilo Salazar, James R. Walters, Fraser Simpson, Mark Blaxter, Andrea Manica, James Mallet, and Chris D. Jiggins. 2013. 'Genome-Wide Evidence for Speciation with Gene Flow in Heliconius Butterflies'. *Genome Research* 23 (11): 1817–28. https://doi.org/10.1101/gr.159426.113.

Martin, Simon H., John W. Davey, Camilo Salazar, and Chris D. Jiggins. 2019. 'Recombination Rate Variation Shapes Barriers to Introgression across Butterfly Genomes'. *PLOS Biology* 17 (2): e2006288. https://doi.org/10.1371/journal.pbio.2006288.

Martincová, Iva, Ľudovít Ďureje, Jakub Kreisinger, Miloš Macholán, and Jaroslav Piálek. 2019. 'Phenotypic Effects of the Y Chromosome Are Variable and Structured in Hybrids among House Mouse Recombinant Lines'. *Ecology and Evolution*, May, ece3.5196. https://doi.org/10.1002/ece3.5196.

Masly, John P., and Daven C. Presgraves. 2007. 'High-Resolution Genome-Wide Dissection of the Two Rules of Speciation in Drosophila'. *PLOS Biology* 5 (9): e243. https://doi.org/10.1371/journal.pbio.0050243.

Matthee, Conrad A, Bettine Jansen Van Vuuren, Diana Bell, and Terence J Robinson. 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): 433–47. https://doi.org/10.1080/10635150490445715.

Maxwell, Colin S., Victoria E. Sepulveda, David A. Turissini, William E. Goldman, and Daniel R. Matute. 2018. 'Recent Admixture between Species of the Fungal Pathogen Histoplasma'. *Evolution Letters* 2 (3): 210–20. https://doi.org/10.1002/evl3.59.

Mayr, Ernst. 1942. 'Systematics And The Origin Of Species From The Viewpoint Of A Zoologist'. 1942.

———. 1963. *Animal Species and Evolution*. *Animal Species and Evolution*. Harvard University Press. https://www.degruyter.com/document/doi/10.4159/harvard.9780674865327/html.

McDermott, Shannon R., and Mohamed A. F. Noor. 2010. 'The Role of Meiotic Drive in Hybrid Male Sterility'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1544): 1265–72. https://doi.org/10.1098/rstb.2009.0264.

McKenna, A. H., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, et al. 2010. 'The Genome Analysis Toolkit: A MapReduce Framework

for Analyzing next-Generation DNA Sequencing Data'. *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

McLaughlin, Richard N., and Harmit S. Malik. 2017. 'Genetic Conflicts: The Usual Suspects and Beyond'. Edited by Joel D. Levine, Daniel J. C. Kronauer, and Michael H. Dickinson. *Journal of Experimental Biology* 220 (1): 6–17. https://doi.org/10.1242/jeb.148148.

Meier, Joana I., David A. Marques, Salome Mwaiko, Catherine E. Wagner, Laurent Excoffier, and Ole Seehausen. 2017. 'Ancient Hybridization Fuels Rapid Cichlid Fish Adaptive Radiations'. *Nature Communications* 8 (1): 14363. https://doi.org/10.1038/ncomms14363.

Melo, Maria C., Alicia Grealy, Beth Brittain, Greg M. Walter, and Daniel Ortiz-Barrientos. 2014. 'Strong Extrinsic Reproductive Isolation between Parapatric Populations of an Australian Groundsel'. *New Phytologist* 203 (1): 323–34. https://doi.org/10.1111/nph.12779.

Melo-Ferreira, J., P. C. Alves, H. Freitas, N. Ferrand, and P. Boursot. 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): 2643–58. https://doi.org/10.1111/j.1365-294X.2009.04221.x.

Melo-Ferreira, J., P. Boursot, M. Carneiro, P. J. Esteves, L. Farelo, and P. C. Alves. 2012a. 'Recurrent Introgression of Mitochondrial DNA among Hares (Lepus Spp.) Revealed by Species-Tree Inference and Coalescent Simulations'. *Systematic Biology* 61 (3): 367–81. https://doi.org/10.1093/sysbio/syr114.

———. 2012b. 'Recurrent Introgression of Mitochondrial DNA Among Hares (Lepus Spp.) Revealed by Species-Tree Inference and Coalescent Simulations'. *Systematic Biology* 61 (3): 367–367. https://doi.org/10.1093/sysbio/syr114.

Melo-Ferreira, J., P. Boursot, E. Randi, A. Kryukov, F. Suchentrunk, N. Ferrand, and P. C. Alves. 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): 605–18. https://doi.org/10.1111/j.1365-294X.2006.03166.x.

Melo-Ferreira, J., P. Boursot, F. Suchentrunk, N. Ferrand, and P. C. Alves. 2005. 'Invasion from the Cold Past: Extensive Introgression of Mountain Hare (Lepus Timidus) Mitochondrial DNA into Three Other Hare Species in Northern Iberia'. *Molecular Ecology* 14 (8): 2459–64. https://doi.org/10.1111/j.1365-294X.2005.02599.x.

Melo-Ferreira, J, L Farelo, H Freitas, F Suchentrunk, P Boursot, and P C Alves. 2014a. 'Home-Loving Boreal Hare Mitochondria Survived Several Invasions in Iberia: The Relative Roles of Recurrent Hybridisation and Allele Surfing.' *Heredity* 112 (3): 265–73. https://doi.org/10.1038/hdy.2013.102.

———. 2014b. 'Home-Loving Boreal Hare Mitochondria Survived Several Invasions in Iberia: The Relative Roles of Recurrent Hybridisation and Allele Surfing'. *Heredity* 112 (3): 265–73. https://doi.org/10.1038/hdy.2013.102.

Metcalfe, Cushla J., Kira V. Bulazel, Gianni C. Ferreri, Elizabeth Schroeder-Reiter, Gerhard Wanner, Willem Rens, Craig Obergfell, Mark D. B. Eldridge, and Rachel J. O'Neill. 2007. 'Genomic Instability Within Centromeres of Interspecific Marsupial Hybrids'. *Genetics* 177 (4): 2507–17. https://doi.org/10.1534/genetics.107.082313.

Mihola, Ondrej, Vladimir Landa, Florencia Pratto, Kevin Brick, Tatyana Kobets, Fitore Kusari, Srdjan Gasic, et al. 2021. 'Rat PRDM9 Shapes Recombination Landscapes, Duration of Meiosis, Gametogenesis, and Age of Fertility'. *BMC Biology* 19 (1): 86. https://doi.org/10.1186/s12915-021-01017-0.

Mihola, Ondrej, Zdenek Trachtulec, Cestmir Vlcek, John C. Schimenti, and Jiri Forejt. 2009. 'A Mouse Speciation Gene Encodes a Meiotic Histone H3

Methyltransferase'. *Science* 323 (5912): 373–75. https://doi.org/10.1126/science.1163601.

Moran, Benjamin M, Cheyenne Payne, Quinn Langdon, Daniel L Powell, Yaniv Brandvain, and Molly Schumer. 2021. 'The Genomic Consequences of Hybridization'. Edited by Patricia J Wittkopp. *ELife* 10 (August): e69016. https://doi.org/10.7554/eLife.69016.

Muller, H. 1942. 'Isolating Mechanisms, Evolution, and Temperature'. *Biol. Symp.* 6: 71–125.

Nachman, Michael W., and Bret A. Payseur. 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): 409–21. https://doi.org/10.1098/rstb.2011.0249.

Nielsen, Rasmus, Joshua M. Akey, Mattias Jakobsson, Jonathan K. Pritchard, Sarah Tishkoff, and Eske Willerslev. 2017. 'Tracing the Peopling of the World through Genomics'. *Nature* 541 (7637): 302–10. https://doi.org/10.1038/nature21347.

Norris, Laura C., Bradley J. Main, Yoosook Lee, Travis C. Collier, Abdrahamane Fofana, Anthony J. Cornel, and Gregory C. Lanzaro. 2015. 'Adaptive Introgression in an African Malaria Mosquito Coincident with the Increased Usage of Insecticide-Treated Bed Nets'. *Proceedings of the National Academy of Sciences* 112 (3): 815–20. https://doi.org/10.1073/pnas.1418892112.

Nosil, Patrik, and Jeffrey L. Feder. 2012. 'Genomic Divergence during Speciation: Causes and Consequences'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1587): 332–42. https://doi.org/10.1098/rstb.2011.0263.

Nosil, Patrik, Daniel J. Funk, and Daniel Ortiz-Barrientos. 2009. 'Divergent Selection and Heterogeneous Genomic Divergence'. *Molecular Ecology* 18 (3): 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x.

Nosil, Patrik, Luke J. Harmon, and Ole Seehausen. 2009. 'Ecological Explanations for (Incomplete) Speciation'. *Trends in Ecology & Evolution* 24 (3): 145–56. https://doi.org/10.1016/j.tree.2008.10.011.

Nosil, Patrik, and Dolph Schluter. 2011. 'The Genes Underlying the Process of Speciation'. *Trends in Ecology & Evolution* 26 (4): 160–67. https://doi.org/10.1016/j.tree.2011.01.001.

Orr, H. A. 1995. 'The Population Genetics of Speciation: The Evolution of Hybrid Incompatibilities'. *Genetics* 139 (4): 1805–13.

Patel, Maulik R, Ganesh K Miriyala, Aimee J Littleton, Heiko Yang, Kien Trinh, Janet M Young, Scott R Kennedy, Yukiko M Yamashita, Leo J Pallanck, and Harmit S Malik. 2016. 'A Mitochondrial DNA Hypomorph of Cytochrome Oxidase Specifically Impairs Male Fertility in Drosophila Melanogaster'. *ELife* 5: e16923. https://doi.org/10.7554/eLife.16923.

Payseur, Bret A., and Loren H. Rieseberg. 2016. 'A Genomic Perspective on Hybridization and Speciation'. *Molecular Ecology* 25 (11): 2337–60. https://doi.org/10.1111/mec.13557.

Petersen, Thomas Nordahl, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2011. 'SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions'. *Nature Methods* 8 (10): 785–86. https://doi.org/10.1038/nmeth.1701.

Phifer-Rixey, Megan, Matthew Bomhoff, and Michael W. Nachman. 2014. 'Genome-Wide Patterns of Differentiation Among House Mouse Subspecies'. *Genetics* 198 (1): 283–97. https://doi.org/10.1534/genetics.114.166827.

Phifer-Rixey, Megan, Bettina Harr, and Jody Hey. 2020. 'Further Resolution of the House Mouse (Mus Musculus) Phylogeny by Integration over Isolation-with-Migration Histories'. *BMC Evolutionary Biology* 20 (1): 120. https://doi.org/10.1186/s12862-020-01666-9.

Phifer-Rixey, Megan, and Michael W Nachman. 2015. 'Insights into Mammalian Biology from the Wild House Mouse Mus Musculus'. *ELife* 4 (April): e05959. https://doi.org/10.7554/eLife.05959.

Pool, John E. 2015. 'The Mosaic Ancestry of the Drosophila Genetic Reference Panel and the D. Melanogaster Reference Genome Reveals a Network of Epistatic Fitness Interactions'. *Molecular Biology and Evolution* 32 (12): 3236–51. https://doi.org/10.1093/molbev/msv194.

Powell, Sean, Damian Szklarczyk, Kalliopi Trachana, Alexander Roth, Michael Kuhn, Jean Muller, Roland Arnold, et al. 2012. 'EggNOG v3.0: Orthologous Groups Covering 1133 Organisms at 41 Different Taxonomic Ranges'. *Nucleic Acids Research* 40 (D1). https://doi.org/10.1093/nar/gkr1060.

Prager, E M, C Orrego, and R D Sage. 1998. 'Genetic Variation and Phylogeography of Central Asian and Other House Mice, Including a Major New Mitochondrial Lineage in Yemen.' *Genetics* 150 (2): 835–61.

Presgraves, Daven. 2009. '12 - Drive and Sperm: The Evolution and Genetics of Male Meiotic Drive'. In *Sperm Biology*, edited by Tim R. Birkhead, David J. Hosken, and Scott Pitnick, 471–506. London: Academic Press. https://doi.org/10.1016/B978-0-12-372568-4.00012-4.

Presgraves, Daven C. 2010. 'The Molecular Evolutionary Basis of Species Formation'. *Nature Reviews Genetics* 11 (3): 175–80. https://doi.org/10.1038/nrg2718.

Price, Trevor D., and Michelle M. Bouvier. 2002. 'The Evolution of F1 Postzygotic Incompatibilities in Birds'. *Evolution* 56 (10): 2083–89. https://doi.org/10.1111/j.0014-3820.2002.tb00133.x.

Queiroz, Kevin de. 1998. 'The General Lineage Concept of Species, Species Chteria, and the Process of Speciation', 19.

Quilodrán, Claudio S., Alexandros Tsoupas, and Mathias Currat. 2020. 'The Spatial Signature of Introgression After a Biological Invasion With Hybridization'. *Frontiers in Ecology and Evolution* 8 (September): 569620. https://doi.org/10.3389/fevo.2020.569620.

Raufaste, Nathalie, Annie Orth, Khalid Belkhir, David Senet, Carole Smadja, Stuart Baird, François Bonhomme, Barbara Dod, and Pierre Boursot. 2005. 'Inferences of Selection and Migration in the Danish House Mouse Hybrid Zone'. *Biological Journal of the Linnean Society* 84 (3): 593–616. https://doi.org/10.1111/j.1095-8312.2005.00457.x.

Rebollo, Rita, Béatrice Horard, Benjamin Hubert, and Cristina Vieira. 2010. 'Jumping Genes and Epigenetics: Towards New Species'. *Gene* 454 (1): 1–7. https://doi.org/10.1016/j.gene.2010.01.003.

Reid, Neil. 2011. 'European Hare (Lepus Europaeus) Invasion Ecology: Implication for the Conservation of the Endemic Irish Hare (Lepus Timidus Hibernicus)'. *Biological Invasions* 13 (3): 559–69. https://doi.org/10.1007/s10530-010-9849-x.

Reid, Neil, and W. Ian Montgomery. 2007. 'Is Naturalisation of the Brown Hare in Ireland a Threat to the Endemic Irish Hare?' *Biology and Environment* 107 (3): 129–38. https://doi.org/10.3318/BIOE.2007.107.3.129.

Rice, William R., Sergey Gavrilets, and Urban Friberg. 2008. 'Sexually Antagonistic "Zygotic Drive" of the Sex Chromosomes'. *PLOS Genetics* 4 (12): e1000313. https://doi.org/10.1371/journal.pgen.1000313.

Rieseberg, L. H. 2003. 'Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization'. *Science* 301 (5637): 1211–16. https://doi.org/10.1126/science.1086949.

Rieseberg, Loren H., Margaret A. Archer, and Robert K. Wayne. 1999. 'Transgressive Segregation, Adaptation and Speciation'. *Heredity* 83 (4): 363–72. https://doi.org/10.1038/sj.hdy.6886170.

Rieseberg, Loren H., and Benjamin K. Blackman. 2010. 'Speciation Genes in Plants'. *Annals of Botany* 106 (3): 439–55. https://doi.org/10.1093/aob/mcq126.

Rieseberg, Loren H., and John H. Willis. 2007. 'Plant Speciation'. *Science (New York, N.Y.)* 317 (5840): 910–14. https://doi.org/10.1126/science.1137729.

Rosenberg, N A, T Burke, K Elo, M W Feldman, P J Freidlin, M A Groenen, J Hillel, et al. 2001. 'Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes from 20 Chicken Breeds.' *Genetics* 159 (2): 699–713.

Rosenberg, Noah A, Jonathan K Pritchard, James L Weber, Howard M Cann, Kenneth K Kidd, Lev A Zhivotovsky, and Marcus W Feldman. 2002. 'Genetic Structure of Human Populations' 298: 6.

Roux, Camille, Christelle Fraïsse, Jonathan Romiguier, Yoann Anciaux, Nicolas Galtier, and Nicolas Bierne. 2016. 'Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence'. Edited by Craig Moritz. *PLOS Biology* 14 (12): e2000234. https://doi.org/10.1371/journal.pbio.2000234.

Rundle, Howard D., and Patrik Nosil. 2005. 'Ecological Speciation'. *Ecology Letters* 8 (3): 336–52. https://doi.org/10.1111/j.1461-0248.2004.00715.x.

Sackton, Timothy B., Horacio Montenegro, Daniel L. Hartl, and Bernardo Lemos. 2011. 'Interspecific Y Chromosome Introgressions Disrupt Testis-Specific Gene Expression and Male Reproductive Phenotypes in Drosophila'. *Proceedings of the National Academy of Sciences* 108 (41): 17046–51. https://doi.org/10.1073/pnas.1114690108.

Sankararaman, Sriram, Swapan Mallick, Michael Dannemann, Kay Prüfer, Janet Kelso, Svante Pääbo, Nick Patterson, and David Reich. 2014. 'The Genomic Landscape of Neanderthal Ancestry in Present-Day Humans'. *Nature* 507 (7492): 354–57. https://doi.org/10.1038/nature12961.

Sanz-Martín, María J., Andone Estonba, Carmen Manzano, Mikel Iriondo, Gonzalo Pérez-Suárez, Alfonso Fernández, and Fernando Palacios. 2014. 'Genetic Structure of Brown and Iberian Hare Populations in Northern Iberia: Implications for Conservation of Genetic Diversity'. *The Journal of Wildlife Management* 78 (4): 632–44. https://doi.org/10.1002/jwmg.713.

Schluter, Dolph, and Associate Editor: Mark A. McPeek. 2000. 'Ecological Character Displacement in Adaptive Radiation'. *The American Naturalist* 156 (S4): S4–16. https://doi.org/10.1086/303412.

Schumer, Molly, Rongfeng Cui, Gil G. Rosenthal, and Peter Andolfatto. 2015. 'Reproductive Isolation of Hybrid Populations Driven by Genetic Incompatibilities'. Edited by Bret A. Payseur. *PLOS Genetics* 11 (3): e1005041. https://doi.org/10.1371/journal.pgen.1005041.

Schumer, Molly, Chenling Xu, Daniel L. Powell, Arun Durvasula, Laurits Skov, Chris Holland, John C. Blazier, et al. 2018. 'Natural Selection Interacts with Recombination to Shape the Evolution of Hybrid Genomes'. *Science* 360 (6389): 656–60. https://doi.org/10.1126/science.aar3684.

Schwarz, Ernst, and Henriette K. Schwarz. 1943. 'The Wild and Commensal Stocks of the House Mouse, Mus Musculus Linnaeus'. *Journal of Mammalogy* 24 (1): 59–72. https://doi.org/10.2307/1374781.

Scordato, Elizabeth S. C., Chris C. R. Smith, Georgy A. Semenov, Yu Liu, Matthew R. Wilkins, Wei Liang, Alexander Rubtsov, et al. 2020. 'Migratory Divides Coincide with Reproductive Barriers across Replicated Avian Hybrid Zones above the Tibetan Plateau'. *Ecology Letters* 23 (2): 231–41. https://doi.org/10.1111/ele.13420.

Ségurel, Laure, Ellen Miranda Leffler, and Molly Przeworski. 2011. 'The Case of the Fickle Fingers: How the PRDM9 Zinc Finger Protein Specifies Meiotic Recombination Hotspots in Humans'. *PLOS Biology* 9 (12): e1001211. https://doi.org/10.1371/journal.pbio.1001211.

Seixas, Fernando A., Pierre Boursot, and José Melo-Ferreira. 2018. 'The Genomic Impact of Historical Hybridization with Massive Mitochondrial DNA Introgression'. *Genome Biology* 19 (1): 91. https://doi.org/10.1186/s13059-018-1471-8.

Selz, O. M., and O. Seehausen. 2019. 'Interspecific Hybridization Can Generate Functional Novelty in Cichlid Fish'. *Proceedings of the Royal Society B: Biological Sciences* 286 (1913): 20191621. https://doi.org/10.1098/rspb.2019.1621.

Shan, Xiaohui, Zhenlan Liu, Zhenying Dong, Yongming Wang, Yu Chen, Xiuyun Lin, Likun Long, Fangpu Han, Yingshan Dong, and Bao Liu. 2005. 'Mobilization of the Active MITE Transposons MPing and Pong in Rice by Introgression from Wild Rice (Zizania Latifolia Griseb.)'. *Molecular Biology and Evolution* 22 (4): 976–90. https://doi.org/10.1093/molbev/msi082.

Shi, Wei, Antje Krella, Annie Orth, Yang Yu, and Reinald Fundele. 2005. 'Widespread Disruption of Genomic Imprinting in Adult Interspecies Mouse (Mus) Hybrids'. *Genesis (New York, N.Y.: 2000)* 43 (3): 100–108. https://doi.org/10.1002/gene.20161.

Simon, Alexis, Nicolas Bierne, and John J. Welch. 2018. 'Coadapted Genomes and Selection on Hybrids: Fisher's Geometric Model Explains a Variety of Empirical Patterns'. *Evolution Letters* 2 (5): 472–98. https://doi.org/10.1002/evl3.66.

Sjodin, Bryson M F, Kurt E Galbreath, Hayley C Lanier, and Michael A Russello. 2021. 'Chromosome-Level Reference Genome Assembly for the American Pika (Ochotona Princeps)'. *Journal of Heredity* 112 (6): 549–57. https://doi.org/10.1093/jhered/esab031.

Slatkin, M., and R. Lande. 1994. 'Segregation Variance after Hybridization of Isolated Populations'. *Genetical Research* 64 (1): 51–56. https://doi.org/10.1017/s0016672300032547.

Smith, Andrew T., ed. 2018. *Lagomorphs: Pikas, Rabbits, and Hares of the World*. Baltimore: Johns Hopkins University Press.

Smith-Unna, Richard, Chris Boursnell, Rob Patro, Julian Hibberd, and Steven Kelly. 2016a. 'TransRate: Reference Free Quality Assessment of de Novo Transcriptome Assemblies'. *Genome Research*, June, gr.196469.115. https://doi.org/10.1101/gr.196469.115.

Smith-Unna, Richard, Chris Boursnell, Rob Patro, Julian M. Hibberd, and Steven Kelly. 2016b. 'TransRate: Reference-Free Quality Assessment of de Novo Transcriptome Assemblies'. *Genome Research* 26 (8): 1134–44. https://doi.org/10.1101/gr.196469.115.

Song, Ying, Stefan Endepols, Nicole Klemann, Dania Richter, Franz-Rainer Matuschka, Ching-Hua Shih, Michael W. Nachman, and Michael H. Kohn. 2011. 'Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice'. *Current Biology: CB* 21 (15): 1296–1301. https://doi.org/10.1016/j.cub.2011.06.043.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, et al. 2014. 'Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation'. *Science* 344 (6185): 738–42. https://doi.org/10.1126/science.1252136.

Srinivasa, Ananya Nidamangala, and Sarah E. Zanders. 2020. 'Meiotic Drive'. *Current Biology* 30 (11): R627–29. https://doi.org/10.1016/j.cub.2020.04.023.

Steensma, David P., Robert A. Kyle, and Marc A. Shampo. 2010. 'Abbie Lathrop, the "Mouse Woman of Granby": Rodent Fancier and Accidental Genetics Pioneer'. *Mayo Clinic Proceedings* 85 (11): e83. https://doi.org/10.4065/mcp.2010.0647.

Stelkens, Rike, and Ole Seehausen. 2009. 'Genetic Distance Between Species Predicts Novel Trait Expression in Their Hybrids'. *Evolution* 63 (4): 884–97. https://doi.org/10.1111/j.1558-5646.2008.00599.x.

Suchentrunk, F, Z Mamuris, C Stamatis, H Ben Slimen, K Hackländer, G Haerer, and M Giacometti. 2005. 'Introgressive Hybridization in Wild Living Mountain Hares (L. Timidus Varronis) and Brown Hares (L. Europaeus) and Morphological Consequences'. *Mamm. Biol* 70 (Suppl.): 39–40.

Suzuki, H, M Nunome, G Kinoshita, K P Aplin, P Vogel, A P Kryukov, M-L Jin, et al. 2013. 'Evolutionary and Dispersal History of Eurasian House Mice Mus Musculus Clarified by More Extensive Geographic Sampling of Mitochondrial DNA'. *Heredity* 111 (5): 375–90. https://doi.org/10.1038/hdy.2013.60.

Suzuki, Hitoshi, Lyudmila V. Yakimenko, Daiki Usuda, and Liubov V. Frisman. 2015. 'Tracing the Eastward Dispersal of the House Mouse, Mus Musculus'. *Genes and Environment* 37 (August). https://doi.org/10.1186/s41021-015-0013-9.

Svardal, Hannes, Fu Xiang Quah, Milan Malinsky, Benjamin P Ngatunga, Eric A Miska, Walter Salzburger, Martin J Genner, George F Turner, and Richard Durbin. 2020. 'Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation'. *Molecular Biology and Evolution* 37 (4): 1100–1113. https://doi.org/10.1093/molbev/msz294.

Svedin, Nina, Chris Wiley, Thor Veen, Lars Gustafsson, and Anna Qvarnström. 2008. 'Natural and Sexual Selection against Hybrid Flycatchers'. *Proceedings of the Royal Society B: Biological Sciences* 275 (1635): 735–44. https://doi.org/10.1098/rspb.2007.0967.

Sweigart, Andrea L. 2010. 'The Genetics of Postmating, Prezygotic Reproductive Isolation Between Drosophila Virilis and D. Americana'. *Genetics* 184 (2): 401–10. https://doi.org/10.1534/genetics.109.111245.

Sweigart, Andrea L., and Lex E. Flagel. 2015. 'Evidence of Natural Selection Acting on a Polymorphic Hybrid Incompatibility Locus in Mimulus'. *Genetics* 199 (2): 543–54. https://doi.org/10.1534/genetics.114.171819.

Takada, Toyoyuki, Toshinobu Ebata, Hideki Noguchi, Thomas M. Keane, David J. Adams, Takanori Narita, Tadasu Shin-I, et al. 2013. 'The Ancestor of Extant Japanese Fancy Mice Contributed to the Mosaic Genomes of Classical Inbred Strains'. *Genome Research* 23 (8): 1329–38. https://doi.org/10.1101/gr.156497.113.

Tang, Shanwu, and Daven C. Presgraves. 2009. 'Evolution of the Drosophila Nuclear Pore Complex Results in Multiple Hybrid Incompatibilities'. *Science (New York, N.Y.)* 323 (5915): 779–82. https://doi.org/10.1126/science.1169123.

Taylor, Scott A., and Erica L. Larson. 2019. 'Insights from Genomes into the Evolutionary Importance and Prevalence of Hybridization in Nature'. *Nature Ecology & Evolution* 3 (2): 170. https://doi.org/10.1038/s41559-018-0777-y.

Teeter, Katherine C., Bret A. Payseur, Leslie W. Harris, Margaret A. Bakewell, Lisa M. Thibodeau, Janelle E. O'Brien, James G. Krenz, Maria A. Sans-Fuentes, Michael W. Nachman, and Priscilla K. Tucker. 2008. 'Genome-Wide Patterns of Gene Flow across a House Mouse Hybrid Zone'. *Genome Research* 18 (1): 67–76. https://doi.org/10.1101/gr.6757907.

Thulin, C. G., M. Jaarola, and H. Tegelström. 1997. 'The Occurrence of Mountain Hare Mitochondrial DNA in Wild Brown Hares'. *Molecular Ecology* 6 (5): 463–67.

Thulin, C G, M Jaarola, and H Tegelstrom. 1997. 'The Occurrence of Mountain Hare Mitochondrial DNA in Wild Brown Hares.' *Molecular Ecology* 6 (5): 463–67.

Thulin, Carl Gustaf. 2003. 'The Distribution of Mountain Hares Lepus Timidus in Europe: A Challenge from Brown Hares L. Europaeus?' *Mammal Review* 33 (1): 29–42. https://doi.org/10.1046/j.1365-2907.2003.00008.x.

Thulin, Carl Gustaf, Meiying Fang, and Alexander O. Averianov. 2006. 'Introgression from Lepus Europaeus to L. Timidus in Russia Revealed by Mitochondrial Single Nucleotide Polymorphisms and Nuclear Microsatellites'. *Hereditas* 143 (2006): 68–76. https://doi.org/10.1111/j.2006.0018-0661.01952.x.

Thybert, David, Maša Roller, Fábio C. P. Navarro, Ian Fiddes, Ian Streeter, Christine Feig, David Martin-Galvez, et al. 2018. 'Repeat Associated Mechanisms of Genome Evolution and Function Revealed by the Mus Caroli and Mus Pahari Genomes'. *Genome Research*, March. https://doi.org/10.1101/gr.234096.117.

Tishkoff, Sarah A., Floyd A. Reed, Françoise R. Friedlaender, Christopher Ehret, Alessia Ranciaro, Alain Froment, Jibril B. Hirbo, et al. 2009. 'The Genetic Structure and History of Africans and African Americans'. *Science (New York, N.Y.)* 324 (5930): 1035–44. https://doi.org/10.1126/science.1172257.

Turner, Leslie M, and Bettina Harr. 2014. 'Genome-Wide Mapping in a House Mouse Hybrid Zone Reveals Hybrid Sterility Loci and Dobzhansky-Muller Interactions'. Edited by Gil McVean. *ELife* 3 (December): e02504. https://doi.org/10.7554/eLife.02504.

Turner, Thomas L., and Matthew W. Hahn. 2010. 'Genomic Islands of Speciation or Genomic Islands and Speciation?' *Molecular Ecology* 19 (5): 848–50. https://doi.org/10.1111/j.1365-294X.2010.04532.x.

Unckless, R. L., and H. A. Orr. 2009. 'Dobzhansky–Muller Incompatibilities and Adaptation to a Shared Environment'. *Heredity* 102 (3): 214–17. https://doi.org/10.1038/hdy.2008.129.

Ungerer, Mark C., Suzanne C. Strakosh, and Ying Zhen. 2006. 'Genome Expansion in Three Hybrid Sunflower Species Is Associated with Retrotransposon Proliferation'. *Current Biology* 16 (20): R872–73. https://doi.org/10.1016/j.cub.2006.09.020.

Valen, Leigh M. van. 1973. 'A New Evolutionary Law'. In .

Valk, Tom van der, Patrícia Pečnerová, David Díez-del-Molino, Anders Bergström, Jonas Oppenheimer, Stefanie Hartmann, Georgios Xenikoudakis, et al. 2021. 'Million-Year-Old DNA Sheds Light on the Genomic History of Mammoths'. *Nature* 591 (7849): 265–69. https://doi.org/10.1038/s41586-021-03224-9.

Vanlerberghe, F., B. Dod, P. Boursot, M. Bellis, and F. Bonhomme. 1986. 'Absence of Y-Chromosome Introgression across the Hybrid Zone between Mus Musculus Domesticus and Mus Musculus Musculus'. *Genetics Research* 48 (3): 191–97. https://doi.org/10.1017/S0016672300025003.

Veller, Carl, Nathaniel B. Edelman, Pavitra Muralidhar, and Martin A. Nowak. 2019. 'Recombination, Variance in Genetic Relatedness, and Selection against Introgressed DNA'. https://doi.org/10.1101/846147.

Vrana, Paul B. 2007. 'Genomic Imprinting as a Mechanism of Reproductive Isolation in Mammals'. *Journal of Mammalogy* 88 (1): 5–23. https://doi.org/10.1644/06-MAMM-S-013R1.1.

Vyskočilová, MARTINA, ZDENĚK TRACHTULEC, JIŘÍ FOREJT, and JAROSLAV PIÁLEK. 2005. 'Does Geography Matter in Hybrid Sterility in House Mice?' *Biological Journal of the Linnean Society* 84 (3): 663–74. https://doi.org/10.1111/j.1095-8312.2005.00463.x.

Wang, Wenjuan, Yafang Wang, Fumin Lei, Yang Liu, Haitao Wang, and Jiakuan Chen. 2019. 'Incomplete Lineage Sorting and Introgression in the Diversification of Chinese Spot-Billed Ducks and Mallards'. *Current Zoology* 65 (5): 589–97. https://doi.org/10.1093/cz/zoy074.

Weissbrod, Lior, Fiona B. Marshall, François R. Valla, Hamoudi Khalaily, Guy Bar-Oz, Jean-Christophe Auffray, Jean-Denis Vigne, and Thomas Cucchi. 2017. 'Origins of House Mice in Ecological Niches Created by Settled Hunter-Gatherers in the Levant 15,000 y Ago'. *Proceedings of the National Academy of Sciences* 114 (16): 4099–4104. https://doi.org/10.1073/pnas.1619137114.

White, Michael A., Brian Steffy, Tim Wiltshire, and Bret A. Payseur. 2011. 'Genetic Dissection of a Key Reproductive Barrier Between Nascent Species of House Mice'. *Genetics* 189 (1): 289–304. https://doi.org/10.1534/genetics.111.129171.

White, Michael A., Maria Stubbings, Beth L. Dumont, and Bret A. Payseur. 2012. 'Genetics and Evolution of Hybrid Male Sterility in House Mice'. *Genetics* 191 (3): 917–34. https://doi.org/10.1534/genetics.112.140251.

Whitney, Kenneth D., Rebecca A. Randell, and Loren H. Rieseberg. 2010. 'Adaptive Introgression of Abiotic Tolerance Traits in the Sunflower Helianthus Annuus'. *The New Phytologist* 187 (1): 230–39. https://doi.org/10.1111/j.1469-8137.2010.03234.x.

Wright, Kevin M., Deborah Lloyd, David B. Lowry, Mark R. Macnair, and John H. Willis. 2013. 'Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in Mimulus Guttatus'. *PLOS Biology* 11 (2): e1001497. https://doi.org/10.1371/journal.pbio.1001497.

Wu, Chung-I. 2001. 'The Genic View of the Process of Speciation'. *Journal of Evolutionary Biology* 14 (6): 851–65. https://doi.org/10.1046/j.1420-9101.2001.00335.x.

Yamada, Fumio, Mika Takaki, and Hitoshi Suzuki. 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): 107–16. https://doi.org/10.1266/ggs.77.107.

Yonekawa, H, K Moriwaki, O Gotoh, N Miyashita, Y Matsushima, L M Shi, W S Cho, X L Zhen, and Y Tagashira. 1988. 'Hybrid Origin of Japanese Mice "Mus Musculus Molossinus": Evidence from Restriction Analysis of Mitochondrial DNA.' *Molecular Biology and Evolution* 5 (1): 63–78. https://doi.org/10.1093/oxfordjournals.molbev.a040476.

Yonekawa, Hiromichi, Kazuo Moriwaki, Osamu Gotoh, Jun-Ichi Hayashi, Junko Watanabe, Nobumoto Miyashita, Michael L. Petras, and Yusaku Tagashira. 1981. 'Evolutionary Relationships among Five Subspecies of MUS MUSCULUS Based on Restriction Enzyme Cleavage Patterns of Mitochondrial DNA'. *Genetics* 98 (4): 801–16.

Zachos, F. E., H. Ben Slimen, K. Hackländer, M. Giacometti, and F. Suchentrunk. 2010. 'Regional Genetic in Situ Differentiation despite Phylogenetic Heterogeneity in Alpine Mountain Hares'. *Journal of Zoology* 282 (1): 47–53. https://doi.org/10.1111/j.1469-7998.2010.00710.x.

Zeh, David W., Jeanne A. Zeh, and Yoichi Ishida. 2009. 'Transposable Elements and an Epigenetic Basis for Punctuated Equilibria'. *BioEssays* 31 (7): 715–26. https://doi.org/10.1002/bies.200900026.

Chapter III

How species originate and admix

This chapter is composed by three main manuscripts. In the first one we contributed to the understanding of the documented massive introgression of the mitochondrial genome from one hare species into another. We demonstrate that this introgression occurred at the favour of the invasive replacement of the donor species by the receiver during the last deglaciation. This study contextualizes the importance of demographic and biogeographic history in promoting introgression and was published in the SCI indexed journal *Scientific Reports* (João P. Marques, Farelo, et al. 2017).

In the second manuscript we present the discovery of a population resulting from ancient admixture of two house mouse subspecies that are presently reproductively isolated.

Finally, in the third manuscript we demonstrate an extensive Y chromosome introgression from one house mouse subspecies to another. In this study we tested the potential link with an arms-race between ampliconic regions on the X and Y chromosomes that would manipulate sex-ratio. We therefore address the question of the potential role of genetic conflict in promoting introgression.

The last two manuscripts are in currently under preparation for submission to a peer reviewed journal.

Range expansion underlies historical introgressive hybridization in the Iberian hare

Scientific Reports 2017 7:40788
doi: 10.1038/srep40788

João P. Marques[1], Liliana Farelo[1], Joana Vilela[1], Paulo C. Alves[1,2], Jeffrey M. Good[3], Pierre Boursot[4], José Melo-Ferreira[1,2,*]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, 4485-661 Vairão, Portugal.
[2]Departamento de Biologia, Faculdade de Ciências do Porto, 4169-007 Porto, Portugal.
[3]Division of Biological Sciences, University of Montana, 32 Campus Drive, Missoula, MT 59812, USA.
[4]Institut des Sciences de l'Évolution, Université Montpellier, CNRS IRD, Case Courrier 063, Place Eugene Bataillon, 34095 Montpellier, France.

*Author for correspondence: José Melo-Ferreira, CIBIO-InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal, tel: +351 252 660 411, fax: +351 252 661 780, e-mail: jmeloferreira@cibio.up.pt.

Published: 25 January 2017

Abstract

Introgressive hybridization is an important and widespread evolutionary process, but the relative roles of neutral demography and natural selection in promoting massive introgression are difficult to assess and an important matter of debate. Hares from the Iberian Peninsula provide an appropriate system to study this question. In its northern range, the Iberian hare, *Lepus granatensis*, shows a northwards gradient of increasing mitochondrial DNA (mtDNA) introgression from the arctic/boreal *L. timidus,* which it presumably replaced after the last glacial maximum. Here, we asked whether a south-north expansion wave of *L. granatensis* into *L. timidus* territory could underlie mtDNA introgression, and whether nuclear genes interacting with mitochondria ("mitonuc" genes) were affected. We extended previous RNA-sequencing and produced a comprehensive annotated transcriptome assembly for *L. granatensis*. We then genotyped 100 discovered nuclear SNPs in 317 specimens spanning the species range. The distribution of allele frequencies across populations suggests a northwards range expansion, particularly in the region of mtDNA introgression. We found no correlation between variants at 39 mitonuc genes and mtDNA introgression frequency. Whether the nuclear and mitochondrial genomes coevolved will need a thorough investigation of the hundreds of mitonuc genes, but range expansion and species replacement likely promoted massive mtDNA introgression.

Introduction

Hybridization and genetic introgression between closely related species is a relevant evolutionary process that is widespread in nature[1]. Particularly frequent are cases of mitochondrial DNA introgression, often with apparent little or no nuclear DNA introgression, such as in elephants[2], hares[3] or chipmunks[4] (see Toews et al.[5] for a review). This raises important questions related to the demographic or adaptive processes underlying such a common phenomenon[5]. Many of the described cases of introgression may have resulted from range shifts and population replacements, which have presumably accompanied interactions between species during the drastic climate changes accompanying Pleistocene glacial cycles[6]. Understanding the causes and consequences of such introgression is of great interest to evolutionary biology[7], but is an exceptionally difficult endeavor. On the one hand, invasion and range replacement with hybridization may promote massive gene flow from the resident species into the invading species in its newly colonized territories. This demographic model predicts that high frequencies of introgressed variants are caused by their fixation on the invasion front due to genetic drift, as suggested both by simulated[8] and empirical data[9]. On the other hand, incorporating variants of a locally adapted resident species could facilitate colonization of new niches by the invading one[10]. The Pleistocene glacial oscillations strongly induced these interactions, by forcing species to change their ranges and promoting novel secondary contacts during the process[11,12].

Hares from Western Europe illustrate this range shift phenomenon, and appear as ideal models to study the causes and consequences of historical gene flow during the range shifts of the Pleistocene. The fossil record shows that the distribution of species greatly changed during glacial oscillations. The mountain hare, *Lepus timidus*, is currently distributed in the northern Palearctic and in some isolated populations such as Ireland, Scotland, Poland and the Alps, but fossils from southern France or northern Iberian Peninsula show that it inhabited southern Europe during the Pleistocene[13]. The three extant species of the Iberian Peninsula, *L. granatensis*, *L. europaeus* and *L. castroviejoi*, now show high frequencies of mitochondrial DNA (mtDNA) haplotypes introgressed from *L. timidus*, acquired through hybridization at the end of the last glacial period before the latter went locally extinct[14].

In the Iberian hare, *Lepus granatensis* (a species that is distributed across the Iberian Peninsula, except in the northernmost part) mtDNA introgression is strongly structured, being absent from the south and increasing in frequency towards the north. This pattern may be compatible with a northwards expansion of the species after the last glacial maximum, replacing and hybridizing with *L. timidus,* and spreading the traces of

introgression to the north[15,16]. However, the demographic history of *L. granatensis*, which could account for patterns of introgression, has not yet been inferred. Ecological niche modeling partially supports the northwards range expansion scenario, showing that areas where mtDNA introgression is found are those with highest habitat favorability for *L. timidus* at the last glacial maximum[17]. Still, the highest climatic favorability for *L. granatensis* in the same period is more dispersed across the Iberian Peninsula[18]. Moreover, evidence from nuclear markers (10 autosomal, two X-linked and one Y-linked) suggest that introgression is rare (but see pattern of another X-linked locus[19]) and occurs all over the range of *L. granatensis*[3]. This contrast questions the plausibility of a purely demographic scenario as accounting for the structure of mtDNA introgression; a striking pattern that is also repeated in the two other hare species of the Iberian Peninsula, *L. europaeus* and *L. castroviejoi*[20]. Considering the strong functional role of mtDNA-encoded peptides in energy metabolism, their close interactions with nuclear encoded peptides, and evidence for positive selection during the evolution of the *L. timidus* mtDNA lineage[21], cytonuclear coevolution and co-introgression in the northern range of *L. granatensis* remains a strong alternative hypothesis.

In this work we put together published and newly obtained RNA-sequencing data from *L. granatensis*, reconstruct a high-quality transcriptome for the species, and then genotype a subset of 100 ascertained single nucleotide polymorphisms (SNPs). Using these data, we i) test for correlation between variation at nuclear genes with mitochondrion-associated functions and mitochondrial DNA introgression, and ii) infer the demographic history of the species.

Results

RNA sequencing, transcriptome assembly and functional annotation

The cDNA library prepared from liver and kidney tissues of *Lepus granatensis* specimen "o" (Fig. 1), harvested in Navarra during the hunting season, was sequenced in 1/12th of an Illumina HiSeq 2000 lane, produced a raw output of 14,645,969 100bp paired-end reads. Adapter removal and quality trimming resulted in 13,052,770 paired-end reads for a total of 2,496,629,898 bp. Previously published data from 10 other *L. granatensis* specimens[22,23] (Fig. 1 and Supplementary Table S1; a total of 42,529,555 single-end 100bp reads) were added to the dataset, which was used for *de novo* transcriptome assembly. A total of 54,838 contigs, ranging from 224 to 12,481 bp were obtained, which was reduced to 50,580 contigs after removing redundancy considering sequence similarity (Table 1). The transcriptome was further filtered by retaining only contigs with

reciprocal best blast hit annotation against the rabbit genome (21,833 contigs) and/or containing a predicted open reading frame (ORF), resulting in a final assembly with 24,608 contigs, an N50 of 1,724 bp and a total length of 26,161,714 bp (Table 1). When compared with the available transcriptome produced by Cahais et al.[22], the addition of relevant paired-end data caused important improvements in several statistics of the reconstructed assembly, such as unfiltered N50 (1334 vs. 909) and reference coverage (32% vs. 23%) (Supplementary Table S2).

To predict the potential functions of the assembled transcripts, the retained unigenes were aligned with various protein databases in addition to the *O. cuniculus* protein collection: NCBI, SwissProt, Gene Ontology, InterProScan and KEGG. The resulting functional annotation proportions are depicted in Supplementary Table S3, and the most represented GO terms shown in Supplementary Fig. S1.

SNP inference and genotyping

After filtering for missing data and minimum allele frequency, a total of 3,532 SNPs were inferred in the eleven sequenced specimens. 3,402 of these SNPs lay in 1,196 genes annotated through reciprocal best blast hit with rabbit cDNA and peptide collection, while 130 were present in unannotated contigs. After alignment with rabbit genomic sequences, 1,119 genes remained, on which selection of SNPs for further genotyping was done in three distinct classes: A) randomly selected, avoiding only the selection of SNPs laying on the same gene (44 SNPs); B) laying in nuclear genes involved in mitochondrial functions and ordered according to differentiation between the RNA-seq samples from regions with and without mtDNA haplotypes of *L. timidus* origin (see Fig. 1c) (63 SNPs with F_{ST} ranging between 0.6641 and 0); and C) with the same criteria as B but no functions on mitochondria (44 SNPs with F_{ST} ranging between 0.7869 and 0.3803).

Four Sequenom multiplexes were constructed with the 151 selected SNPs, which laid on 133 genes (redundant SNPs for genes of category B were included). These loci were genotyped in 317 *L. granatensis* (which included the 11 sequenced specimens) and 30 *L. timidus* (see Fig. 1 for sampling localities). Population genetic analyses were performed using 314 *L. granatensis* specimens that were organized in sample localities with more than 12 individuals each. After filtering out loci that were invariant, had more than 20% missing data, had inconsistent genotype composition for the 11 RNA-sequenced specimens (used as controls in the genotyping), or laid on the same gene/contig, 100 SNPs remained for the population genetics analyses (31, 39 and 30

from categories A, B and C respectively). Minimum allele frequency in the genotyped loci ranged between 13 and 50%. All SNPs were unlinked and in Hardy-Weinberg proportions.

Correlation with mtDNA introgression

First we tested whether variation among *L. granatensis* populations was correlated with the prevalence of mitochondrial DNA introgression from *L. timidus*. Such correlation would suggest coevolution of the nuclear genome in response to the prevalence of the alien mitochondrial genome. This would be of particular interest if it concerned gene(s) of category B, involved in mitochondrial functions, as compared to those of category C, not involved but chosen with the same ascertainment method. We considered mtDNA introgression frequency as an environmental variable and tested its correlation with variations of allele frequencies among populations, using the Bayenv2 method[24]. We used the set of random SNPs to determine the covariance of allele frequencies due to population history, and then tested loci in categories B and C for association with mtDNA introgression prevalence. No significant associations were found.

Population structure and range expansion

We then sought to infer the population differentiation history over the species range, for which a set of presumably neutrally evolving loci is needed. Loci in category A (randomly selected) can be considered appropriate in this respect (note also that these SNPs either represent synonymous substitutions or lay in a non-coding portion). Loci in the two other categories were chosen with an ascertainment method that could introduce a bias towards loci with a certain differentiation pattern. However, although this ascertainment method was meant to increase the chance to find genes whose evolution was correlated to mitochondrial DNA introgression, we found no sign of such parallelism in any of the analyzed loci (see above). Furthermore, we estimated the differentiation between mtDNA introgressed and non-introgressed populations on our much larger sample in an analysis of molecular variance (using an AMOVA[25]). We found no significant difference in the distribution of the Φ_{CT} statistic between category A and either category B or C SNPs (Mann-Whitney test on pairs and Kruskal-Wallis on all three; mean Φ_{CT} of 0.01623, 0.0149 and 0.0155 for classes A, B and C respectively). Therefore, the results based on the 11 RNA-sequenced samples were very poor predictors of the genetic structure of the species, and no ascertainment bias towards SNPs with the targeted differentiation pattern seemed to prevail in our SNP dataset.

Using BayeScan[26], we inspected whether some of our loci showed any outlier pattern of differentiation among sampling localities. We found four loci with outlier patterns, three showing decreased differentiation (negative alpha; SNP033, SNP121 and SNP145) and one showing increased F_{ST} (positive alpha; SNP119). However, this method is expected to produce many false positives in situations of range expansion[27], and this result should thus be interpreted carefully.

Below we present results based on the whole set of SNPs to infer population structure and history. However, all tests were also performed separately on the randomly selected subset (31 SNPs) and the set of non-outlier loci from the BayeScan analysis (96 SNPs), and produced similar results (see complete results in Supplementary Tables S4 and S5 and Supplementary Figs. S2-4).

We tested for population subdivisions using the Bayesian methods implemented in STRUCTURE[28]. Partition into three clusters was the most favored hypothesis (Supplementary Table S4), with clusters being predominant in different geographic regions: K1 in the Southwestern part, K2 in the central and Northeastern part and K3 in the Northwestern part, around population GAL (Fig. 1a). Most localities appeared admixed between at least two clusters and, apart from a possible distinct genetic cluster in the Northwest (population GAL), a geographical gradient of individual assignment was apparent (Fig. 2a). This coincided with a pattern of isolation by distance (Fig. 2d, Spearman ranked correlation one-tailed P-value = 0.00 in the Mantel test of correlation between genetic and geographic distance). A principal components analysis confirmed these results and showed strong correlation of the two first axes of variation with longitude and latitude (Spearman ranked correlation one-tailed p-value = 0.00; Figs. 2b and 2c). Again, specimens from GAL were suggested to have some level of genetic differentiation from the rest (stars in Fig. 2b).

These patterns of weak clustering and correlation of genetic and geographic distances could reflect isolation by distance in an equilibrium population, vast admixture between historically differentiated populations, or past expansion over the species range. To explore the two latter possibilities we used the method of Peter and Slatkin[29] to infer range expansion and its direction. The secondary contact hypothesis predicts independent and geographically convergent expansions in different areas (corresponding to the areas of prevalence of the different STRUCTURE clusters), while the range expansion hypothesis predicts a single expansion over the species range. We thus applied the test of expansion species-wide, as well as separately on subsets of populations (regions) grouped according to the cluster they were assigned to in majority (region R1 in the Southwest where K1 was predominant, R2 in the Centre and Northeast

where K2 was the most frequent, and R3 in the Northwest where K3 was the major cluster, see Fig. 1a). Taking all three regions together, a signal of range expansion was inferred, with an origin near locality ALT (Fig. 1a, $P<0.05$) (see detailed results for all analyzed datasets in Supplementary Table S5). Allowing for the possibility of multiple origins, significant support for range expansion was only obtained when considering R1 and R2 together with origin again near ALT ($P<0.001$), and when taking R2 alone, with origin near CRE ($P<0.001$), thus also suggesting expansion from South to North. In summary, these results are neither compatible with a model of isolation by distance, nor with a model of secondary contact between the K1 and K2 clusters. They are compatible with a global range expansion from Southwest Iberia into the rest of the Peninsula. The south-north signal of expansion is most pronounced in the northern half of the Peninsula, the region where mitochondrial introgression from *L. timidus* prevails. The direction of expansion parallels the gradient of mtDNA introgression, with higher introgression in the direction of expansion. The Northwestern populations (around GAL) may not result from this global expansion, and are not affected by mitochondrial introgression[14]. There appears to have been an isolated pocket in this region, which secondarily admixed with the other populations. Similar conclusions are drawn from the analysis of the randomly selected loci and the dataset removing potential F_{ST} outliers (Supplementary Table S5).

Discussion

Extensive mitochondrial DNA introgression between species is very common in animals and cytonuclear co-evolution is often proposed as a likely explanation for these striking patterns[5]. In addition to providing further insights into the evolutionary history of hares, our study provides a relatively rare test of the alternative hypotheses that may account for mtDNA introgression.

We first produced a *de novo* assembly and annotation of the transcriptome of *L. granatensis,* and improved its quality when compared to a previous study[22]. The availability of RNA-sequencing from 11 individuals distributed across the range of the species allowed discovering numerous SNPs, which is a valuable resource for future work on this and related species. We could successfully validate a subset of the discovered SNPs on a large sample of individuals and then used them to address the question of general interest related with the causes and consequences of massive mitochondrial DNA introgression. One of the first objectives of our study was to discover nuclear genes involved in mitochondrial function, i.e. interacting with the mitochondrial genome or its products, which evolved under the influence of the massive prevalence of

an alien mitochondrial genome in some populations. This is a difficult statistical question because shared population history can cause correlations between allele frequencies at different loci that are not related with adaptation or coevolution. We therefore used a method that accounts for such correlations and removes their confounding effects, and treated mitochondrial DNA introgression frequency as an environmental variable, but found no correlation. We attempted to increase the chance of finding such genes by selecting SNPs with apparent high differentiation between mitochondrial DNA introgressed and non-introgressed regions based on the initial set of 11 sequenced specimens. We also applied the same ascertainment bias to genes not involved in mitochondrial metabolism as a control. However, enrichment on this basis was unsuccessful because the allele frequency estimates derived from the small ascertainment sample were shown to be very poor predictors of the species-wide patterns. This weakened our ability to find evidence for coevolution of the nuclear genome accompanying the massive mtDNA introgression. We also did not find any case where the SNP variant present in *L. timidus* is predominant only in the northern range of *L. granatensis* where the mtDNA haplotypes of the former are frequent, which suggests no cytonuclear co-introgression among our genotyped loci. A more thorough investigation of cytonuclear coevolution in this system is therefore needed, including all genes potentially interacting with the mitochondrial genome and its products. Interest in this question is revived by recently reported suspicion of interspecific cytonuclear incompatibilities in a mammal species[30].

Given the lack of association between mtDNA and our nuclear SNPs, we were able to consider the chosen loci as neutral to make inferences on population history and demography. We obtained similar results when using both the randomly selected loci and the dataset removing potential F_{ST} outlier loci (that could actually result from the range expansion we also inferred[27]), which suggests that the complete SNP set is appropriate for this purpose. The relative demography of interacting species is an important determinant of the rate and direction of introgression between them[8]. Importantly, it can explain unequal levels of introgression along the genome and geographic gradients of introgression, similar to patterns resulting from introgression promoted by natural selection along an environmental gradient. Hares from the Iberian Peninsula have been widely affected by introgression from the arctic/boreal *L. timidus*, but genetic reconstruction of demographic histories can only be performed in the receiver species, because the donor is locally extinct. In this work, we provide the first insights into the population history of *L. granatensis*, which can be used to interpret patterns of historical introgression of *L. timidus* origin.

Interpreting geographical variations of allele frequencies of particular loci as resulting from either selection or demographic and historical contingencies is notoriously difficult. This question can be addressed in different ways, depending on priors about species history and ecology and the function of the genes studied. Here we had a strong hypothesis of a south-north range expansion that could account for the geographically restricted and massive mitochondrial DNA introgression[3,15,16,19]. The signal of range expansion is clear in the Northern part where introgression is present, and weak considering only the South of the Peninsula, at the origin of range expansion and where mitochondrial introgression is absent (Fig. 1; Supplementary Table S5). It follows the direction expected to have created a South-North gradient of increasing introgression, due to allele surfing and potentially repeated introgression along the invasion and hybridization front[8], as evidenced in *L. europaeus*, in its Iberian range[16]. However, the sample of loci we used to infer range expansion is relatively small[29] and should be treated as preliminary. Interestingly, we also found that the Northwestern populations appear to have evolved separately and not to have been involved in the South-North expansion that underlies the hybridization events. Concordantly, they are not affected by mitochondrial DNA introgression.

Our results provide support for the important role of purely demographic processes in promoting massive mitochondrial DNA introgression, and suggest that this phenomenon may explain the common nature of mtDNA reticulation. Even though the phenomenon is expected to particularly affect regions of the genome linked to the least dispersing sex[8] (which in mammals are often females), similar gradients of introgression may occur along the genome. Therefore, understanding whether the patterns of mtDNA introgression in the affected species is generally accompanied by concomitant frequencies and distribution of introgressed nuclear DNA variants, as a result of the very same demographic dynamics, depends on a thorough inspection of the genome, which in most of the model systems for mtDNA reticulation has been poorly explored (Good et al.[4] showed that 95% of the models of cytonuclear discordance reviewed by Toews and Brelsford[5] analysed 20 or less nuclear markers). In addition, evolutionary histories are complex and we cannot dismiss the possibility that massive introgression had selective consequences on the nuclear genome, either by promoting coevolution of some mitochondrion nuclear genes or favoring co-introgression of such genes (hypotheses we had little power to test due to the small number of genes screened here). Introgression of locally adapted genes may facilitate range replacements and co-contribute to patterns of widespread reticulation. Again, only genome-wide surveys will allow addressing this question appropriately.

Methods

RNA-sequencing, de novo transcriptome assembly, validation and annotation

Liver and kidney samples from one *Lepus granatensis* (specimen "o" in Fig. 1) harvested in Navarra, Spain, during the hunting season were collected shortly after the death of the animal and immediately placed in RNAlater and then stored at -20ºC until RNA extraction. Grinding of the tissues was done separately, with liquid nitrogen and a ceramic mortar and pestle. RNA extraction was then performed using the QIAGEN RNeasy® Mini Kit following the manufacturer's protocol. The concentration of the RNA extracts and RNA Integrity Number (RIN) were estimated using a Bioanalyzer 2100 (Agilent Technologies). The RNA extracts of the two tissues were then pooled in equal proportions, cDNA libraries were produced using the TruSeq RNA Sample Preparation Kit (Illumina) and the fragment size distribution checked using a Bioanalyzer 2100 (Agilent Technologies). The library was then pooled in equimolar proportions with cDNA libraries produced for other purposes and sequenced in 1/12th of one lane of an HiSeq2000 at the QB3 Computational Genomics Resource Laboratory (CGRL), University of California, Berkeley, producing 100bp paired-end reads. Low quality reads were removed using CASAVA-1.8 FASTQ Filter, adapter sequences were identified and removed using Cutadapt v1.3[31] and reads were trimmed for quality using Trimmomatic v0.32[32] with options LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:20. Previously published transcriptome single-end sequence reads from 10 *L. granatensis* specimens[22,23] (data available from ftp://ngsisem.mbb.univ-montp2.fr/phylogenie_et_evolution_moleculaire/pub/popphyl/reads/Lepus_granatensis/) were added to the dataset (see geographic distribution of the 11 sampled specimens in Fig. 1a).

De novo transcriptome assembly was constructed using Trinity v.2.1.0[33] with default parameters and data from the 11 specimens (single-end data from 10 specimens and paired-end data from 1 specimen). Large amounts of chimeras and poorly supported contigs can arise during the assembly process. Therefore, in order to evaluate the quality of the produced assembly, sequenced reads were mapped back to the reference assembly and quantitative and qualitative measures were applied using TransRate[34]. A redundancy filter was then applied using CD-HIT-EST v 4.6.4[35] to remove contigs with more than 95% similarity. To evaluate the transcriptome completeness, cDNA and peptide collections of *Oryctolagus cuniculus* (European rabbit) were downloaded from ENSEMBL (http://www.ensembl.org/info/data/ftp; release 2.0.81), and used as reference (the average genomic divergence between rabbits and hares is expected to be around

5%[36]). TransDecoder[37] was then used to predict the open reading frames and remove transcriptome noise (e.g. non-coding RNA, DNA contamination or erroneously assembled contigs). Filtered assembly was aligned against the rabbit transcriptome using Conditional Reciprocal Best BLAST[38], which selects the reciprocal best hits from a bi-directional BLAST+ alignment. The final transcriptomes consisted of unigenes with rabbit annotation and/or predicted open reading frame. Similar statistics were then obtained for the existing *L. granatensis* transcriptome produced by Cahais et al.[22] to compare the quality and completeness of the assemblies.

Functional annotation was further performed to identify putative mRNA functions. Unigenes were additionally annotated using the blastx algorithm against Swiss-Prot[39] and NCBI non-redundant protein databases[40] applying an e-value cut-off of 1e-5. Additionally, NCBI non-redundant annotation output was incorporated on Blast2GO[41] in order to perform a functional classification of the transcripts, through the assigning of Gene Ontology (GO) terms and prediction of metabolic pathways using the Kyoto Encyclopedia of Genes and Genome (KEGG)[42,43]. Complementarily, InterProScan[44] was used to identify protein domains and the output was quantified.

Inference of single nucleotide polymorphisms and genotyping

Mapping of the paired-end and single-end reads of the 11 *L. granatensis* specimens onto the *de novo* assembled transcriptome was performed using the bwa-mem v0.7.12 algorithm with default parameters[45], and the resulting alignments were sorted using SAMtools v0.1.18[46]. SNP call was performed using Reads2snp v2.0.64[23,47] using a threshold of 20 for site and mapping qualities, a minimum coverage of 8X and a genotype probability >0.95.

In order to guarantee an overall representation of the species range, only sites represented in at least 8 out of the 11 specimens and with a minimum of 4 specimens from each of the southern or northern halves of Iberia were retained for further analyses. In addition, low frequency variants were filtered out, retaining only sites with a minimum allele frequency of 0.2.

Spliced alignments of the assembled contigs with the rabbit genome were produced using sim4db v1.896[48] in order to identify intron-exon boundaries. Only sequences with more than 90% identity and SNPs laying within the inferred alignment coordinates were retained for further analyses.

One first class of SNPs, selected randomly among independent contigs was identified for further genotyping – class A. Two additional classes were built, favoring

SNPs with higher F_{ST} between regions with and without mtDNA introgression considering the 11 sequenced specimens. Among these SNPs, those laying on genes with functions related with the mitochondrion were ascribed to class B. These are candidates for potential coevolution or cointrogression with mtDNA. A third class was then considered among SNPs on genes with no mitochondria functions – class C – which served as control for SNPs from class B. Genes with functions in the mitochondria (mitonuc genes) were defined according to the InterMitoBase database list, based on human information[49] (708 genes). Given that human gene annotation is extensively more complete than the rabbit one, gene codes were obtained using the human-rabbit 1:1 orthologs obtained from Biomart, or using blastx alignment against the human collection of peptides, when orthology information was missing.

Genotyping was performed using Sequenom MassARRAY at the Centre de Génomique Fonctionelle de Bordeaux, Plateforme Génome Transcriptome, Université Bordeaux 2. Preference was given for SNPs with at least one 100bp flank with no polymorphism among the 11 sequenced specimens. In some cases, polymorphism in the flanking regions was taken into account with degenerate nucleotide symbols. Whenever possible two SNPs were genotyped at certain focal genes from SNP class B. After multiplex construction, the final selection included 151 SNPs laying in 133 genes (class A: 49 genes; class B: 42 genes; class C: 42 genes), which were genotyped in 317 *L. granatensis* (including the 11 sequenced specimens) and 30 *L. timidus* (see geographic distribution of sampling in Fig. 1).

Genotype data analyses

Population genetics analyses were performed using 314 genotyped *L. granatensis* specimens that were organized in 20 sampling localities with at least 12 individuals each. Each genotyped locus was checked for conformation to Hardy-Weinberg proportions using genepop v4.2[50]. The same software was used to test for linkage disequilibrium between pairs of loci.

Given that the selections of SNPs from classes B and C were ascertained taking into account the geographic distribution of mtDNA introgression from *L. timidus*, we tested whether we could find any correlation using Bayenv2[24] and treating introgression frequencies as an environmental variable. The method initially estimates a null covariance matrix from a set of randomly selected loci, and then assigns Bayes Factors to each SNP that measures whether allele frequencies co-vary with an environmental variable, over the null model of population structure. The set of randomly selected loci

(class A) was used to estimate the covariance matrix based on 10 million iterations. Three replicate runs were performed to ensure the consistency of the estimates, and Bayes factors were averaged over the independent runs. The variable was normalized following the recommendation. A Bayes Factor > 3 was considered as indicative of correlation between the allele frequencies at a given locus and the variable.

We then checked whether the ascertainment based on differentiation between regions with and without mtDNA introgression based on the 11 sequenced specimens, used to select SNP classes B and C, indeed produced an enrichment of differentiation. This was tested using a locus-by-locus AMOVA[25], as implemented in Arlequin v3.5[51].

Also, we tested whether F_{ST} outliers could be found in our dataset, using the Bayesian method implemented in BayeScan v2.116. This method estimates whether subpopulation-specific allele frequencies differ from the common gene pool, as measured by an F_{ST} coefficient. The F_{ST} coefficient is decomposed in a locus-specific parameter, alpha, and a population-specific parameter, beta. A false discovery rate of 0.05 was used. Note however that this method is expected to produce many false positives when the true underlying demographic model is range expansion[27].

Analyses of population structure were further performed using the randomly selected loci, the dataset without F_{ST} outliers and the full dataset. To investigate the partition of genetic diversity in *Lepus granatensis* the admixture model implemented in STRUCTURE 2.3.4[28] was applied. A variable number of K populations, from 2 to 10, was considered and three replicate runs per partition with 1 million steps after a burn-in period of 500 000 were performed. The model considering sampling locations as prior information (LOC prior) was applied because it is expected to better detect shallow structure. Replicate runs were analyzed using CLUMPP v1.1.2[52] and DISTRUCT v1.1[53] was used to plot the results. The best number of populations, K, was inferred using Evanno's delta K method[54], as implemented in STRUCTURE HARVESTER[55].

Possible partitions of genetic diversity in the dataset were further investigated using principal components analyses, as implemented in Eigenstrat[56]. In addition, the existence of a correlation between population differentiation and geographic distance was verified using the ISOLDE method implemented in Genepop[50].

Finally, evidence for a range expansion and its putative origin was tested using the R library rangeExpansion[29]. This method estimates a directionality index that detects the clines of allele frequencies produced during range expansions. Range expansion was tested for the complete datasets and for subsets defined according to the STRUCTURE results, to detect possible multiple range expansions. Results were summarized and visualized using the summary and plot functions.

Acknowledgements

This work was supported by Portuguese National Funds through the Fundação para a Ciência e a Tecnologia, FCT (HybridAdapt, FCT-ANR/BIA-EVF/0250/2012), by POPH-QREN funds from the European Social Fund and FCT (IF/00033/2014 FCT Investigator grant to JMF), and by project ‘‘Genomics and Evolutionary Biology’’ co-financed by North Portugal Regional Operational Programme 2007/2013 (ON.2—O Novo Norte) under the National Strategic Reference Framework (NSRF) and the European Regional Development Fund (ERDF). This research was conducted in the scope of the International Associated Laboratory (LIA) "Biodiversity and Evolution" supported by InEE (Institut Ecologie et Environnement, CNRS, France).

Author Contribution

JMF and PB designed the study with insights from PCA. LF and JV performed the laboratory work at CIBIO-InBIO, University of Porto, supervised by JMF and at the University of Montana supervised by JMG, respectively. JPM and JMF analyzed the data. JMF wrote the paper with contributions from PB. All authors read, revised and approved the manuscript.

Additional Information

Accession codes: Raw RNA-sequencing data and the assembled transcriptomes are available at ###. The table of inferred SNPs from RNA-sequencing data and the genotyping results are available at ###.

Competing financial interests: The authors declare no competing financial interests.

References

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

2 Roca, A. L., Georgiadis, N. & O'Brien, S. J. Cytonuclear genomic dissociation in African elephant species. *Nat Genet* **37**, 96-100 (2005).

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

4 Good, J. M., Vanderpool, D., Keeble, S. & Bi, K. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution* **69**, 1961–1972, doi:10.1111/evo.12712 (2015).

5 Toews, D. P. L. & Brelsford, A. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21**, 3907-3930, doi:10.1111/j.1365-294X.2012.05664.x (2012).

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

7 Canestrelli, D. *et al.* The Tangled Evolutionary Legacies of Range Expansion and Hybridization. *Trends Ecol Evol* **31**, 677-688, doi:10.1016/j.tree.2016.06.010 (2016).

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

9 Mastrantonio, V., Porretta, D., Urbanelli, S., Crasta, G. & Nascetti, G. Dynamics of mtDNA introgression during species range expansion: insights from an experimental longitudinal study. *Scientific reports* **6**, 30355, doi:10.1038/srep30355 (2016).

10 Lucek, K., Lemoine, M. & Seehausen, O. Contemporary ecotypic divergence during a recent range expansion was facilitated by adaptive introgression. *Journal of Evolutionary Biology* **27**, 2233-2248, doi:10.1111/jeb.12475 (2014).

11 Hewitt, G. M. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112 (1999).

12 Rowe, K. C., Heske, E. J., Brown, P. W. & Paige, K. N. Surviving the ice: Northern refugia and postglacial colonization. *Proceeding of the National Academy of Sciences U S A* **101**, 10355-10359, doi:10.1073/pnas.0401338101 (2004).

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

14 Melo-Ferreira, J., Boursot, P., Suchentrunk, F., Ferrand, N. & Alves, P. C. Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology* **14**, 2459-2464 (2005).

15 Melo-Ferreira, J. *et al.* The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Molecular Ecology* **16**, 605-618 (2007).

16 Melo-Ferreira, J. *et al.* Home-loving boreal hare mitochondria survived several invasions in Iberia: the relative roles of recurrent hybridisation and allele surfing. *Heredity* **112**, 265-273, doi:10.1038/hdy.2013.102 (2014).

17 Acevedo, P. *et al.* Range dynamics driven by Quaternary climate oscillations explain the distribution of introgressed mtDNA ofLepus timidus origin in hares from the Iberian Peninsula. *J Biogeogr* **42**, 1727-1735, doi:10.1111/jbi.12556 (2015).

18 Acevedo, P., Melo-Ferreira, J., Real, R. & Alves, P. C. Past, Present and Future Distributions of an Iberian Endemic, Lepus granatensis: Ecological and Evolutionary Clues from Species Distribution Models. *PLoS One* **7**, e51529, doi:10.1371/journal.pone.0051529 (2012).

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

20 Melo-Ferreira, J. *et al.* Recurrent Introgression of Mitochondrial DNA Among Hares (Lepus spp.) revealed by Species-tree Inference and Coalescent Simulations. *Syst Biol* **61**, 367-381, doi:10.1093/sysbio/syr114 (2012).

21 Melo-Ferreira, J. *et al.* The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biology and Evolution* **6**, 886-896, doi:10.1093/gbe/evu059 (2014).

22 Cahais, V. *et al.* Reference-free transcriptome assembly in non-model organisms from next generation sequencing data. *Molecular Ecology Resources* **12**, 834-845 (2012).

23 Gayral, P. *et al.* Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. *PLoS Genet* **9**, e1003457, doi:10.1371/journal.pgen.1003457 (2013).

24 Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J. K. Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics* **185**, 1411-1423, doi:10.1534/genetics.110.114819 (2010).

25 Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes - Application to Human Mitochondrial-DNA Restriction Data. *Genetics* **131**, 479-491 (1992).

26 Foll, M. & Gaggiotti, O. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977-993, doi:10.1534/genetics.108.092221 (2008).

27 Lotterhos, K. E. & Whitlock, M. C. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol* **23**, 2178-2192, doi:10.1111/mec.12725 (2014).

28 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).

29 Peter, B. M. & Slatkin, M. Detecting Range Expansions from Genetic Data. *Evolution* **67**, 3274-3289, doi:Doi 10.1111/Evo.12202 (2013).

30 Ma, H. *et al.* Incompatibility between Nuclear and Mitochondrial Genomes Contributes to an Interspecies Reproductive Barrier. *Cell metabolism* **24**, 283-294 (2016).

31 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2010).

32 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

33 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644-652, doi:10.1038/nbt.1883 (2011).

34 Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. & Kelly, S. TransRate: reference free quality assessment of de novo transcriptome assemblies. *Genome Res*, doi:10.1101/gr.196469.115 (2016).

35 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, doi:10.1093/bioinformatics/btl158 (2006).

36 Carneiro, M., Blanco-Aguiar, J. A., Villafuerte, R., Ferrand, N. & Nachman, M. W. Speciation in the European rabbit (Oryctolagus cuniculus): islands of differentiation on the X chromosome and autosomes. *Evolution* **64**, 3443-3460 (2010).

37 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512, doi:10.1038/nprot.2013.084 (2013).

38 Aubry, S., Kelly, S., Kumpers, B. M., Smith-Unna, R. D. & Hibberd, J. M. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C4 photosynthesis. *PLoS Genet* **10**, e1004365, doi:10.1371/journal.pgen.1004365 (2014).

39 Gasteiger, E., Jung, E. & Bairoch, A. SWISS-PROT: connecting biomolecular knowledge via a protein database. *Curr Issues Mol Biol* **3**, 47-55 (2001).

40 Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61-65, doi:10.1093/nar/gkl842 (2007).

41 Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676, doi:10.1093/bioinformatics/bti610 (2005).

42 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

43 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457-462, doi:10.1093/nar/gkv1070 (2016).

44 Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res* **33**, W116-120, doi:10.1093/nar/gki442 (2005).

45 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:DOI 10.1093/bioinformatics/btp324 (2009).

46 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

47 Tsagkogeorga, G., Cahais, V. & Galtier, N. The Population Genomics of a Fast Evolver: High Levels of Diversity, Functional Constraint, and Molecular Adaptation in the Tunicate Ciona intestinalis. *Genome Biology and Evolution* **4**, 740-749, doi:10.1093/gbe/evs054 (2012).

48 Walenz, B. & Florea, L. Sim4db and Leaff: utilities for fast batch spliced alignment and sequence indexing. *Bioinformatics* **27**, 1869-1870, doi:10.1093/bioinformatics/btr285 (2011).

49 Gu, Z. G. *et al.* InterMitoBase: An annotated database and analysis platform of protein-protein interactions for human mitochondria. *BMC genomics* **12** (2011).

50 Raymond, M. & Rousset, F. GENEPOP (version 1.2): a population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249 (1995).

51 Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564-567, doi:10.1111/j.1755-0998.2010.02847.x (2010).

52 Jakobsson, M. & Rosenberg, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806, doi:10.1093/bioinformatics/btm233 (2007).

53 Rosenberg, N. A. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138, doi:10.1046/j.1471-8286.2003.00566.x (2004).

54 Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611-2620, doi:10.1111/j.1365-294X.2005.02553.x (2005).

55 Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4**, 359-361, doi:10.1007/s12686-011-9548-7 (2012).

56 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909, doi:http://www.nature.com/ng/journal/v38/n8/suppinfo/ng1847_S1.html (2006).

57 Mitchell-Jones, A. *et al. Atlas of European Mammals.*, (Academic Press, 1999).

Table 1: Summary statistics of *L. granatensis de novo* transcriptome assembly, considering three filtering levels: raw assembly, removing redundancy and retaining contigs with ORF and/or reciprocal best blast hit onto rabbit transcripts and peptides.

	Raw	Non-redundant	ORF and/or blast hit
Number of contigs	54,838	50,580	24,608
Average contig length (bp)	800	761	1,063
Total length (bp)	43,877,813	38,480,389	26,161,714
Maximum contig length (bp)	12,481	12,481	12,481
Minimum contig length (bp)	224	224	224
N50 (bp)	1,334	1,247	1,724
Number of contigs > 1 kb	13,340	11,299	9,361
Proportion of contigs > 1 kb (%)	24.3	22.3	38.0
Reference Proteins with blast hit (%)	51	51	51
Reference coverage (%)	32	32	32

Figure 1:

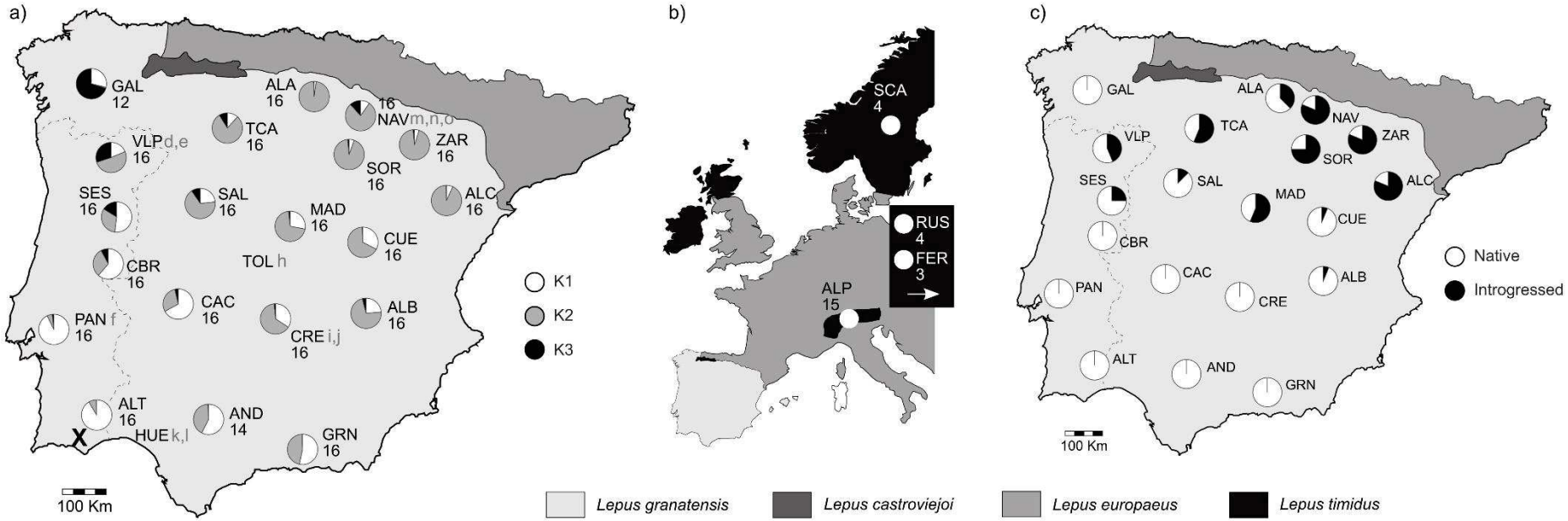


Figure 1: Geographic distribution of four hare species in the Iberian Peninsula (a, c) and Western Europe (b) [57]. a) Localities sampled for *L. granatensis* (see Supplementary Table S1 for a detailed description); numbers indicate sample size, pie charts the proportion of STRUCTURE assignment to each of 3 clusters using the LOC prior and the complete SNP dataset (100 loci); the "X" marks the inferred origin of the range expansion; grey letters indicate specimens for which RNA was sequenced and used to build the transcriptome ("f-n" from Cahais *et al.*[22] and Gayral et al.[23], and "o" from this work). b) Localities sampled for *L. timidus*, indicating the sample sizes; two sampling localities are not shown on this map (RUS – Russia; and FER – Far East Russia). c) Proportion of the mitochondrial DNA lineages, native *L. granatensis* or introgressed from *L. timidus*, in the genotyped samples. Maps were generated in vectorial format using Inkscape v0.91 (https://inkscape.org).

Figure 2:

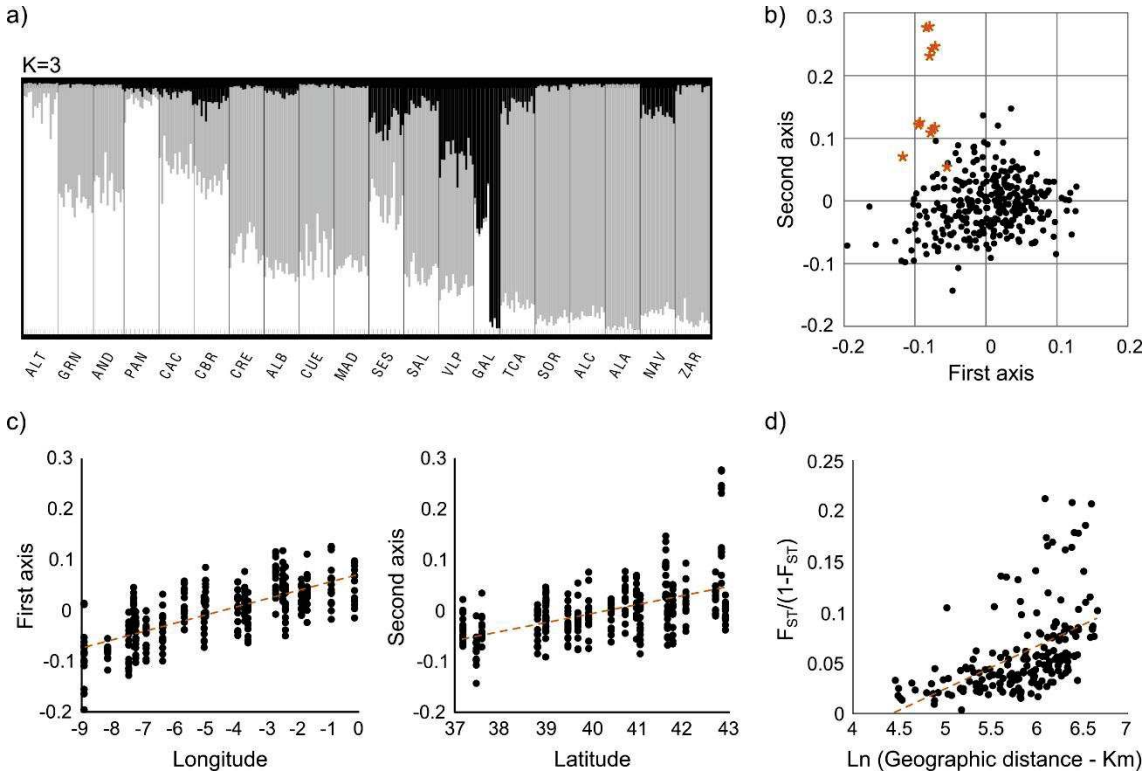


Figure 2: Organization of genetic diversity in *Lepus granatensis* from the analysis of 100 SNP loci (see Supplementary Figs. S2-S5 for a complete description of the results obtained with several subsets of the dataset): a) Structure plots with individual assignment to 3 clusters, as determined using Evanno's delta K method, and using the sample locations as priors of the admixture model; population codes as in Fig. 1; b) coordinates of samples on the first two axes of variation determined with a Principal Component Analysis (PCA) (stars correspond to specimens from Northwestern Iberian population GAL; see Fig. 1); c) Correlation between the first two PCA axes of variation and geographical coordinates of sample localities (Spearman rank correlation, p=0.00 for both analyses; dashed line indicates a linear regression trendline); d) correlation between genetic differentiation and geographic distance among pairs of populations (Spearman rank correlation, p=0.00; dashed line indicates a linear regression trendline).

Supplementary Information

Range expansion underlies historical introgressive hybridization in the Iberian hare

João P. Marques[1], Liliana Farelo[1], Joana Vilela[1], Paulo C. Alves[1,2], Jeffrey M. Good[3], Pierre Boursot[4], José Melo-Ferreira[1,2]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, 4485-661 Vairão, Portugal.
[2]Departamento de Biologia, Faculdade de Ciências do Porto, 4169-007 Porto, Portugal.
[3]Division of Biological Sciences, University of Montana, 32 Campus Drive, Missoula, MT 59812, USA.
[4]Institut des Sciences de l'Évolution, Université Montpellier, CNRS IRD, 34095 Montpellier, France.

Table S1: Sampling localities, sample sizes for transcriptome assembly and SNP genotyping and proportion of mtDNA introgression in the genotyped samples.

Code	Locality	Latitude	Longitude	n Transcriptome	n Genotyping	Proportion Introgression
Lepus granatensis						
ALT	Alcoutim, Portugal	37.469978	-7.473078	-	16	0
ALA	Álava, Spain	42.910000	-2.698387	-	16	0.375
ALB	Albacete, Spain	38.994350	-1.858542	-	16	0.0625
ALC	Alcañiz, Spain	41.051037	-0.133537	-	16	0.8125
AND	Andaluzia, Spain	37.590711	-5.019765	-	14	0
CAC	Cáceres, Spain	39.471329	-6.370961	-	16	0
CBR	Castelo Branco, Portugal	39.924751	-7.241590	-	16	0
CRE	Ciudad Real, Spain	38.984829	-3.927378	2	16	0
CUE	Cuenca, Spain	39.690079	-2.381535	-	16	0.0625
GAL	Galicia, Spain	42.826070	-8.157443	-	12	0
GRN	Granada, Spain	37.177338	-3.598557	-	16	0
HUE	Huesca, Spain	-	-	2	2	-
MAD	Madrid, Spain	40.416775	-3.703790	-	16	0.5625
NAV	Navarra, Spain	42.695393	-1.676069	3	16	0.8125
PAN	Pancas, Portugal	38.809101	-8.918929	1	16	0
SAL	Salamanca, Spain	40.970104	-5.663540	-	16	0.125
SES	Serra da Estrela, Portugal	40.725407	-6.905594	-	16	0.25
SOR	Soria, Spain	41.764431	-2.463772	-	16	0.75
TCA	Tierra de Campos, Spain	42.049622	-4.976654	-	16	0.5625
TOL	Toledo, Spain	-	-	1	1	-
VLP	Valpaços, Portugal	41.608715	-7.310906	2	16	0.4375
ZAR	Zaragoza, Spain	41.648792	-0.889581	-	16	0.8125
Lepus timidus						
ALP	Alps	-	-	-	15	-

FAR	Far East Russia	-	-	-	3	-
RUS	Western Russia	-	-	-	4	-
SCA	Scandinavia	-	-	-	4	-

Table S2: Comparison of summary statistics of *L. granatensis de novo* transcriptome assemblies produced by Cahais *et al.* (2012)* and by this work.

Statistics	Cahais *et al.* 2011	This work	Variation (%)
Number of contigs	45151	54838	+18
Average contig length (bp)	657	800	+18
Total length (bp)	29676032	43877813	+32
Maximum contig length (bp)	13780	12481	-10
Minimum contig length (bp)	201	224	+10
N50 (bp)	909	1334	+32
Number of contigs > 1 kb	7526	13340	+44
Proportion of contigs > 1 kb (%)	16.7	24.3	+31
Reference Proteins with blast hit (%)	46	51	+10
Reference coverage (%)	24	32	+25

*Cahais, V. *et al.* Reference-free transcriptome assembly in non-model organisms from next generation sequencing data. *Molecular Ecology Resources* 12, 834-845 (2012).

Table S3: Number of unigenes resulting from functional annotation of the *L. granatensis* transcriptome with different protein databases.

	Number of unigenes	Percentage	200 ≤ Length < 1000	Length ≥ 1000
O. cuniculus genome (Ensembl, 2.0.81)	21833	88.7	13097	8736
SwissProt	21933	89.1	13028	8905
NCBI NR	22362	90.9	13383	8979
InterProScan	3580	14.5	1456	2124
KEGG	17072	69.4	8960	8112
Gene Ontology	16867	68.5	9599	7268
Annotated	22740	92.4	13679	9061
Non annotated	1868	7.6	1563	305

Table S4: Results of the Evanno method, indicating the mean likelihoods of the STRUCTURE runs for each tested K value.

K	Replicates	Mean LnP(K)	StD LnP(K)	Ln'(K)[1]	|Ln''(K)|[2]	DeltaK[3]
All						
1	3	-36867.4000	0.1000	-	-	-
2	3	-36530.2667	6.4003	337.1333	16.7667	2.619685
3	3	-36176.3667	3.6950	353.9000	216.000	58.456720
4	3	-36038.4667	10.6651	137.9000	71.4000	6.694763
5	3	-35829.1667	129.1684	209.3000	0.3000	0.002323
6	3	-35619.5667	13.7143	209.6000	149.6667	10.913150
7	3	-35559.6333	41.2723	59.9333	38.4333	0.931214
8	3	-35461.2667	26.0108	98.3667	332.1667	12.770320
9	3	-35695.0667	273.7944	-233.8000	533.7333	1.949395
10	3	-35395.1333	32.1808	299.9333	-	-
No F_{ST} outliers						
1	3	-35384.7333	0.0577	-	-	-
2	3	-35070.5333	9.0224	314.2000	17.000	1.884204
3	3	-34739.3333	6.1712	331.2000	187.5333	30.388620
4	3	-34595.6667	9.4691	143.6667	80.3000	8.480239
5	3	-34371.7000	77.1213	223.9667	29.6333	0.384243
6	3	-34177.3667	19.4526	194.3333	139.9000	7.191844
7	3	-34122.9333	4.5369	54.4333	88.2000	19.440650
8	3	-34156.7000	150.4748	-33.7667	104.6333	0.695354
9	3	-34085.8333	60.7023	70.8667	56.0333	0.923085
10	3	-34071.0000	116.5166	14.8333	-	-
Random						
1	3	-11491.3000	0.1000	-	-	-
2	3	-11340.2667	2.8042	151.0333	50.0000	17.830630
3	3	-11239.2333	2.8361	101.0333	85.5000	30.147280
4	3	-11223.7000	80.6865	15.5333	5.5333	0.068578
5	3	-11213.7000	70.4360	10.0000	83.0000	1.178375
6	3	-11120.7000	12.0926	93.0000	112.0667	9.267407
7	3	-11139.7667	18.5971	-19.0667	72.4333	3.894866
8	3	-11231.2667	94.6392	-91.5000	68.2000	0.720632
9	3	-11254.5667	50.2674	-23.3000	18.0667	0.359411
10	3	-11259.8000	115.1112	-5.2333	-	-

All – 100 loci; No F_{ST} outliers – removing the 4 FST outliers (retaining 96 loci); Random – using only the randomly selected loci (31 loci).

[1]Rate of change of the likelihood distribution (mean).

[2]Absolute value of the 2nd order rate of change of the likelihood distribution (mean).

[3]DeltaK = mean(|Ln''(K)|)/StD(Ln(K))

Table S5: Inferred origin of range expansion considering the complete or random datasets and population partitions according to STRUCTURE results.

			Origin of Range Expansion[4]			
Dataset[1]	K[2]	Partition[3]	Longitude	Latitude	Closest Sampling Locality	Significance[5]
All	3	R1+R2+R3	-7.942774	37.17734	ALT	*
All	3	R1+R2	-7.942774	37.17734	ALT	**
All	3	R1+R3	-8.918929	42.82607	GAL	-
All	3	R2+R3	-8.157443	42.91	GAL	-
All	3	R1	-8.918929	38.48452	PAN	-
All	3	R2	-4.120964	38.98483	CRE	**
All	3	R3	-	-	-	-
No F_{ST} outliers	3	R1+R2+R3	-7.942774	37.17734	ALT	-
No F_{ST} outliers	3	R1+R2	-7.942774	37.17734	ALT	*
No F_{ST} outliers	3	R1+R3	-8.918929	42.82607	GAL	-
No F_{ST} outliers	3	R2+R3	-8.157443	42.91	GAL	-
No F_{ST} outliers	3	R1	-8.918929	38.48452	PAN	-
No F_{ST} outliers	3	R2	-4.120964	38.98483	CRE	**
No F_{ST} outliers	3	R3	-	-	-	-
Random	3	R1+R2+R3	-8.918929	38.68593	PAN	-
Random	3	R1+R2	-8.918929	38.68593	PAN	**
Random	3	R1+R3	-8.918929	38.66385	PAN	-
Random	3	R2+R3	-8.157443	42.91	GAL	-
Random	3	R1	-8.918929	38.54131	PAN	-
Random	3	R2	-3.820206	39.81118	CRE	-
Random	3	R3	-8.157443	42.82607	GAL	-

[1]All 100 loci, removing the 4 FST outliers (retaining 96 loci) and using only the randomly selected loci (31 loci).

[2]Number of K clusters inferred with STRUCTURE; best K, as inferred using the Evanno deltaK method.

[3]Clusters of populations defined according to STRUCTURE assignment, by grouping sampling localities with predominant assignment to each K cluster.

[4]Origin of range expansion estimated with the rangeExpansion method.

[5]Significance of range expansion inference; *P<0.05, **P<0.001.

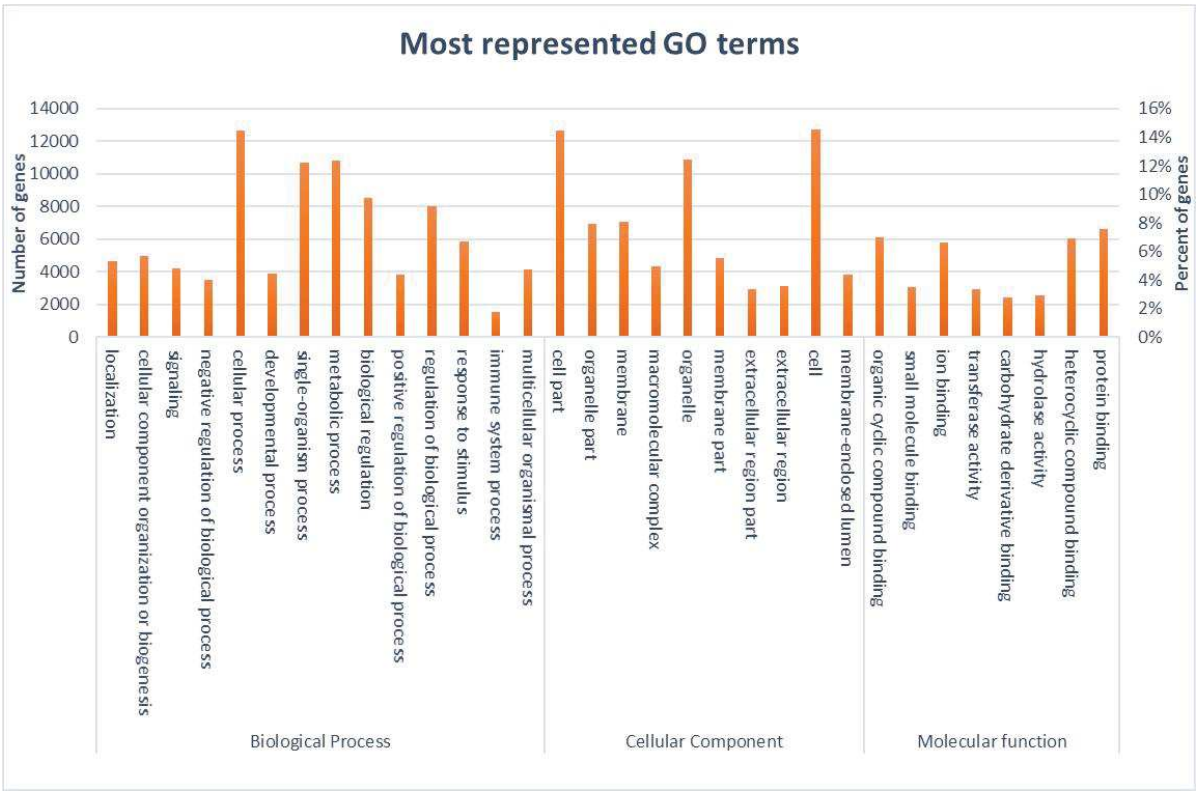


Figure S1: Most represented Gene Ontology (GO) terms in the *L. granatensis* transcriptome.

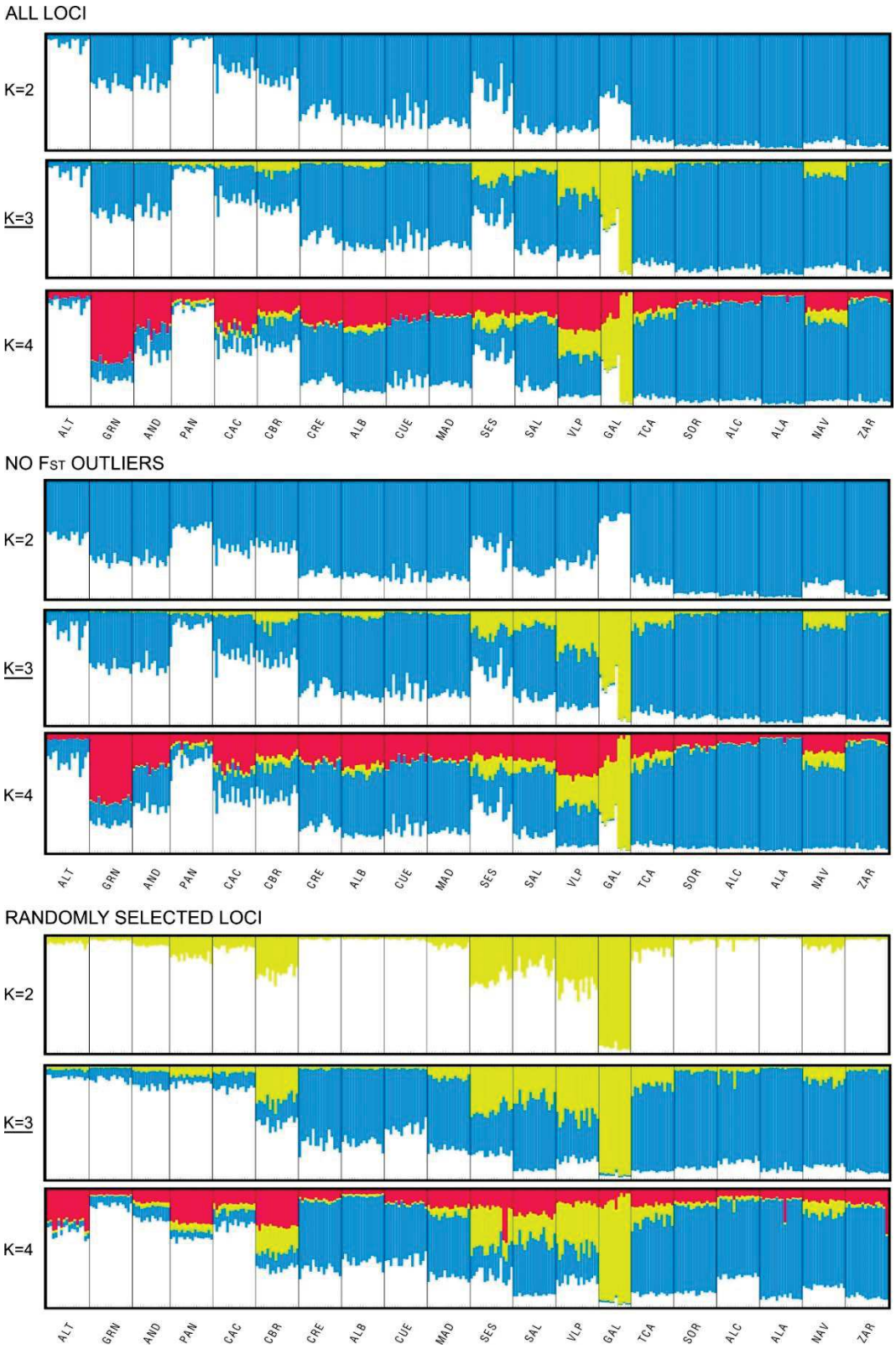


Figure S2: STRUCUTURE plots for K=2, 3 and 4 for analyses including all genotyped loci (100 SNPs), removing putative F_{ST} outliers (96 SNPs) and the randomly selected subset (31 SNPs). The best K, assessed using Evanno's delta K method, is underlined.

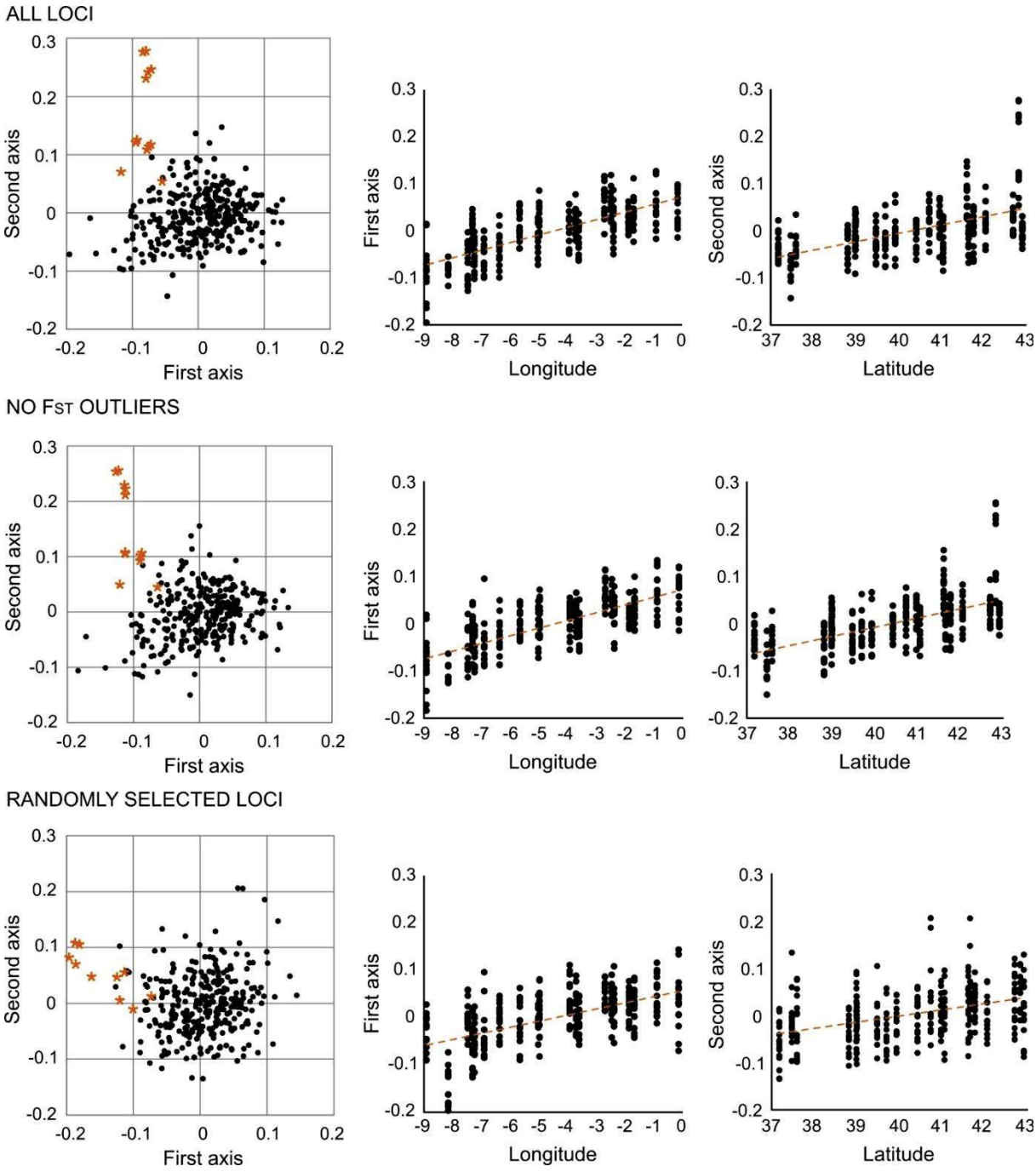


Figure S3: Principal component analysis (PCA) plots and correlation of the first two axes with longitude and latitude for the complete genotyped loci (100 SNPs), removing putative F_{ST} outliers (retaining 96 SNPs) and for the randomly selected subset (31 SNPs). Stars indicate samples from population GAL. All correlations are significant (Spearman rank correlation, p=0.00; dashed line indicates a linear regression trendline).

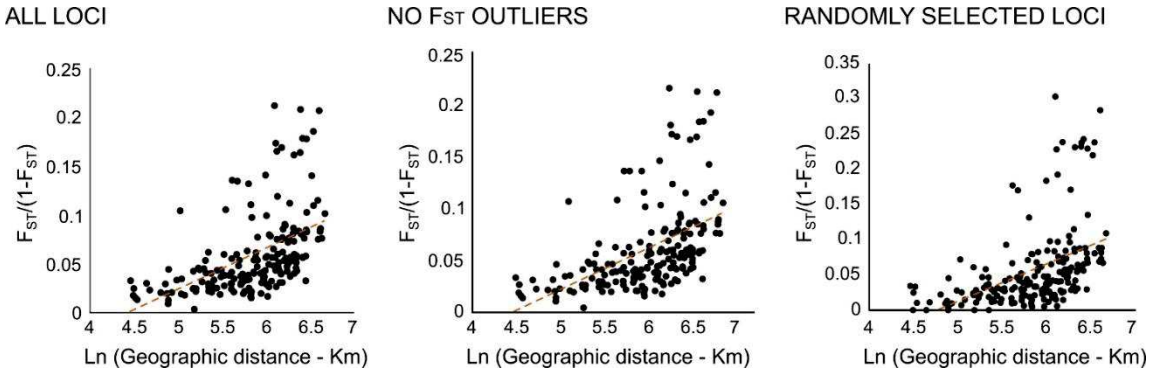


Figure S4: Correlation between genetic differentiation and geographic distance among pairs of populations for analyses conducted using all genotyped loci (100 SNPs), removing putative F_{ST} outliers (retaining 96 SNPs) and using the randomly selected dataset (31) (Spearman rank correlation, p=0.00 in both cases; dashed line indicates linear regression trendlines).

Extensive admixture in the cradle of differentiation and speciation of house mice.

Author affiliations:

João P. Marques[1,2,3]José Melo-Ferreira[1,2*], Pierre Boursot[3*]

[1]CIBIO, Research Center in Biodiversity and Genetic Resources, InBIO Associate Laboratory, University of Porto, 4485-661 Vairão, Portugal;
[2]Department of Biology, FCUP, Faculty of Sciences of the University of Porto, 4169-007 Porto, Portugal;
[3]Institut des Sciences de l'Évolution, Université de Montpellier, CNRS, IRD, , Montpellier, France;

Corresponding authors:

José Melo-Ferreira, CIBIO-InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal, tel: +351 252660411, e-mail: jmeloferreira@cibio.up.pt; Pierre Boursot, Institut des Sciences de l'Évolution, Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier, France, e-mail: pierre.boursot@umontpellier.fr.

Keywords: *Mus musculus,* speciation, genomic incompatibilities, central Iranian lineage

Running title: Admixture in the house mouse cradle of differentiation

Abstract

European house mice (*Mus musculus domesticus* and *M. m. musculus*) are distributed parapatrically and separated by a narrow tension zone, which witnesses the existence of numerous hybrid incompatibilities that prevent free admixture of their genomes. The formation of this hybrid zone results from the very recent expansion of the two subspecies, in association with human settlements and trade. Here, we analyse whole genome sequencing data from populations representing the three known subspecies (including the Asian subspecies *M. m. castaneus*), together with populations close to the presumed cradle of their differentiation, in and around Iran, a region where the three subspecies are in contact. We describe in Central Iran a population (CEI) that is distinct from the three subspecies. Based on the correlations of allele frequencies, and distributions of X chromosome cross-coalescence times between populations, we infer that CEI results from secondary admixture between Iranian *domesticus* and a population related to *musculus*. Methods exploiting allele frequency correlations or inferring ancestry from linkage disequilibrium estimate a *domesticus* contribution of 30-40% for the autosomes and 20% for the X. The mitochondrial DNA and Y chromosome are of the non-*domesticus* types in CEI. This situation may help later studies pinpoint ancient incompatible gene interactions that were sorted in CEI, but could still participate in the isolation between *domesticus* and *musculus* in Europe, as well as incompatibilities that have evolved during the expansion of these two subspecies from their cradle. We propose a biogeographic scenario for the differentiation of *domesticus* and *musculus*, and argue that the resulting picture resembles that of a ring species. The scenario also suggests that a selective advantage may have maintained the non-*domesticus* Y chromosome in the admixed population CEI.

Introduction

That populations evolving independently from each other can accumulate genetic divergence ultimately leading to genetic incompatibilities preventing free secondary admixture, thus leading to the creation of new species, is a well-established pattern (Coyne and Orr 2004). Understanding the underlying evolutionary processes is the focus of intensive research, aiming at addressing fundamental questions such as the genetic origin and architecture of hybrid incompatibilities and the drivers of their onset (drift vs. natural selection, be it adaptive or not). Several approaches of this question have been attempted, each suffering its own limitations. One is based on crosses in the laboratory, aiming at revealing associations between genetic markers and phenotypes witnessing lower hybrid fitness (most often infertility or unviability) (Threadgill et al. 2011). The major limitations are that the results are dependent on the phenotype examined and on the genetic background of the subjects of the crosses, and the success dependent on the complexity of its genetic determination. Another approach is based on the study of genetic exchanges across natural hybrid zones, the expectation being that genomic regions implicated in hybrid incompatibilities should show the most limited exchanges (Macholán et al. 2008). Although this approach is agnostic of the phenotypes concerned, its limitations are the stochasticity of variations in allele frequencies and natural population structure, and possibly the non-fixation of incompatibly factors. A third approach consists in examining the patterns of divergence and differentiation along the genome, the underlying idea being that regions of highest differentiation should contain the factors responsible for the inability to re-admix (Ravinet et al. 2017). This approach is potentially valid in cases where the sister populations have continued to exchange significantly after their separation. The underlying principle is similar to the case of hybrid zones, but several confounding factors prevent a simple interpretation of landscapes of differentiation in the light of sole variations of gene flow. Finally, a situation of interest in this context is when populations that result from extensive admixture of previously differentiated parental populations (or species) exist (Elgvin et al. 2017). Observed patterns of genomic admixture can help pinpoint incompatible combinations of ancestries. The limitation here is to disentangle random from deterministic associations. Given the limitations of each of these approaches, a combination of several, when possible, seems a possible way to progress.

The house mouse (*Mus musculus*) is an interesting model in this context. It has long been known to be structured in several entities (that we will here call subspecies, following the major consensus trend in the literature) that can be recognized genetically and are parapatrically distributed (Boursot et al. 1993). Note we will here ignore the New World, where this species was introduced by humans in historical times. *M. musculus domesticus* thrives

from the Near East to the Mediterranean basin and Western Europe, *Mus musculus musculus* from central Europe to North-East Asia, and *M. m. castaneus* from Indo-Pakistan to South-East Asia. The latter two subspecies have colonized the Japanese archipelago, their admixture making up what is referred to as *M. m. molossinus* (Yonekawa et al. 2012). In East Asia, they are largely parapatric, but seem to admix over a relatively wide area in China (Li et al. 2020). We will here focus on the case that has been studied most extensively, the pair *domesticus-musculus*. These subspecies have recently colonized Western Europe and form a secondary hybrid zone that is narrow (about 20 km), and has all the characteristics of a tension zone, maintained by a balance between dispersal, recombination and selection at many loci against hybrid genotypes (Raufaste et al. 2005, Macholan et al. 2007). There is also evidence for behavioral reinforcement of reproductive isolation in this hybrid zone (Smadja and Ganem 2005, Bimova et al. 2011). There is evidence of lower fertility of hybrid males, both in the natural hybrid zone and in artificial hybrids (Britton Davidian et al. 2005). The X chromosome is heavily involved in hybrid incompatibilities, as witnessed both by very narrow clines in the hybrid zone (Macholán et al. 2011, Dod et al. 1993), and association with hybrid male sterility in laboratory crosses (Good, Dean, and Nachman 2008, White et al. 2011). Several autosomal factors have also been associated with this hybrid infertility syndrome, but they can vary depending on the genetic background. In one specific cross, an autosomal gene involved in male hybrid sterility was identified: *Prdm9*, which controls the position of DNA double strand breaks during meiosis, and the mechanism responsible for meiotic arrest in hybrid males has been deciphered (Baudat et al. 2010). In this cross, an interacting X chromosome factor was also identified, but the causative gene still needs to be confirmed (Forejt, Jansa, and Parvanov 2021). The factors responsible for male hybrid sterility however vary depending on the genetic backgrounds of the parental genomes, indicating that the incompatibility alleles are not fixed in natural populations (Vyskočilová, Pražanová, and Piálek 2009). Geographic and genomic clines have been intensely studied across the European hybrid zone with a number of diagnostic markers along the genome (Janousek et al. 2012). Clines for the X chromosome tend to be consistently narrow. Some autosomal clines also show narrow clines, but only rarely consistently between different transects in different geographical parts of the hybrid zone. Altogether, these results show that *domesticus* and *musculus* in Western Europe are partially genetically isolated (two identified factors being male hybrid infertility and behavioral reinforcement), and that the genetic basis of hybrid incompatibility is complex and variable, involving a large number of autosomal loci and a large part of the X chromosome.

Several studies have attempted to infer the history of the differentiation of the house mouse subspecies, based on DNA sequence variations at multiple loci. All studies have

rejected differentiation in isolation, and proposed a model of divergence with gene flow (Phifer-Rixey, Harr, and Hey 2020). Most studies applied a model of continuous gene flow since the initial divergence, but Duvaux et al. (2011) tested more complex models of gene flow for the *domesticus-musculus* pair and found the best fit for a model with an initial period of isolation followed by secondary admixture (which fits a model of allopatric speciation). Their results however suggested that the admixture had started well before the formation of the European hybrid zone, which results from the very recent expansion of the species in Western Europe (a few thousand years at most). They therefore inferred that admixture must have occurred in the region of origin of this European expansion, namely Iran and its surroundings.

Motivated by this hypothesis, we studied mice from central Iran, a region that is adjacent to known occurrences of both *domesticus* (in the Fertile Crescent, SW Iran) and *musculus* (Northern Iran). The genetic makeup of the mice from Central Iran is still poorly known, because studies were based on a limited number of markers. The mitochondrial DNA lineages found in this region do not belong to the monophyletic lineages characteristic of either *domesticus* or *musculus*, but to a third very diversified lineage, that is usually referred to as *castaneus* (Rajabi-Maham et al. 2012). The haplogroup composition of Central Iran also distinguishes it from *castaneus* from eastern Iran or India. Based on microsatellite variations, the central Iranian population is also distinct from *domesticus* and *musculus*, as well as from *castaneus* populations from Eastern Iran, Indo-Pakistan and SE Asia (Hardouin et al. 2015). There are indications of morphological distinctiveness of the central Iranian population as compared to other regions of Iran, which led some authors to propose a new subspecies name (*M. m. isaticus*, or *M. m. isatissus*, depending on the publications by the same authors (Hamid et al. 2017, Molavi et al. 2015). However, mitochondrial DNA represents a single marker, microsatellites convey little information about ancient ancestries, and morphological data are difficult to interpret in a phylogenetic context. Therefore, we used whole genome sequencing of representatives of the three subspecies and of central Iran. We found evidence that the central Iranian population is indeed distinct from *domesticus* and *musculus* (as well as *castaneus*), but appears to result from an ancient and thorough admixture between two populations related to *domesticus* and *musculus*. We discuss the possible causes of this contrast with the situation in Europe, where the two subspecies do not admix extensively, and the opportunity this offers to better understand the buildup of reproductive isolation between these two subspecies.

Results

Sequence data

We sequenced the whole genome of three mice from SW Iran (*M. m. domesticus*), six mice from Northern Pakistan (*M. m. castaneus*), and seven from Central Iran, as well as two samples of *Mus cypriacus* (a species endemic to the Island of Cyprus, and related to *Mus macedonicus*). We also used whole genome sequences from Harr et al. (2016): *M. m. domesticus* from Ahvaz, SW Iran, and France, *M. m. musculus* from Afghanistan and Kazakhstan, and *M. m. castaneus* from Northern India. Sup. Fig. 1 shows the geographical origin of the sampled populations, and the codes used hereafter to designate them.

After mapping to the house mouse reference genome, calling and filtering variants, we retained ~2.3M high-quality sites. Sup. Table 1 gives the list of samples sequenced, coverage information, sex, whether they carry a *t*-haplotype and whether they are used in the present study. After eliminating related individuals (Supl. Table 2), and applying our coverage criterium (see material and methods), we retained six individuals per population, except for Pakistan, where only four met the criteria. We also identified the individuals bearing a *t*-haplotype (see material and methods - Sup. Fig. 2).

Status of the central Iranian population

We used a subset of linkage disequilibrium (LD) pruned 287k biallelic single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) > 0.02 for two unsupervised methods of population structure inference. The PCA (Figure 1) based on autosomal markers clearly separates the three subspecies along axes 1 and 2, with little dispersion inside each (X and Y chromosomes results - Sup. Fig. 3-4). The CEI sample lies in between the *domesticus* and *musculus* samples. This could reflect that CEI is a descendant of an ancestor of both *domesticus* and *musculus*, or that CEI results from a secondary admixture *domesticus* and *musculus*. Bayesian clustering using ADMIXTURE was not able to clearly distinguish between the two possibilities (Figure 2 and Sup. Fig. 5). The method clearly distinguished the three subspecies, without evidence of admixture, but suggested two possibilities for CEI. For k=4, the most supported value, CEI is modelled as an entity distinct from the three subspecies. However, k=3 is only slightly disfavored, and models CEI as an admixture between *domesticus* and *musculus*, with rather balanced contributions (40 and 60 %, respectively).

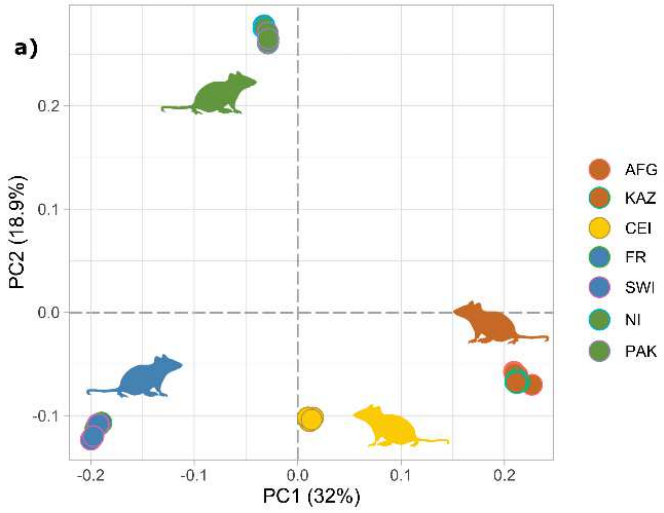


Figure 1- PCA of the LD-pruned high-quality SNP set. Intermediate central Iranian house mouse position relative to the well-known house mouse subspecies.

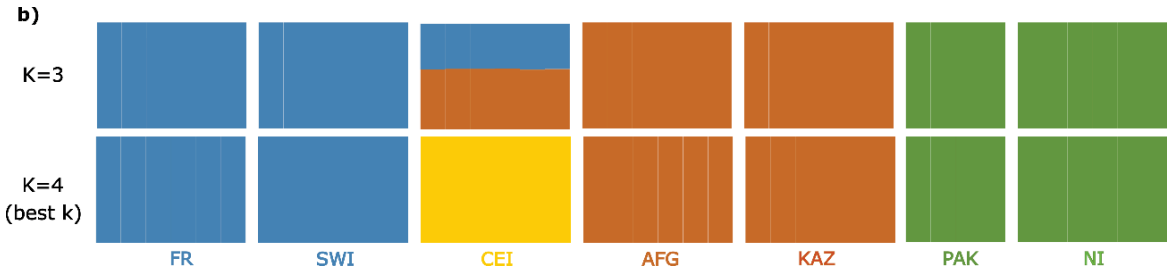


Figure 2 - Population structure based on the admixture analysis for the three well-accepted subspecies and the new lineage. The two best supported values of k (3 and 4) suggest two different scenarios (simple divergence without gene flow vs admix lineage between musculus and domesticus).

We note that neither the PCA nor the Admixture analyses show any sign of a contribution of *castaneus* to CEI, so we will hereafter concentrate on the relationship of CEI with *domesticus* and *musculus* only. Divergence (*Dxy*) across autosomal genomic regions is on average higher between *domesticus* and *musculus* than between any of them and CEI (Figure 3a). This could result from *domesticus* and *musculus* having independently and quasi-simultaneously derived from the CEI branch. If such was the case, a similar pattern should also be seen on the X, but this is not the case since for this chromosome, *Dxy* between *musculus* and CEI is much lower than for the two other pairs, which involve *domesticus* and have similar distributions (Sup. Fig. 6). The X pattern appears more compatible with *musculus* and CEI being sister populations, and the autosomal pattern should then be interpreted as resulting from introgression from *domesticus* into CEI. The distributions of pairwise differentiation (*Fst*) are not informative to distinguish these two major hypotheses (Figure 3b and Sup. Fig. 6).

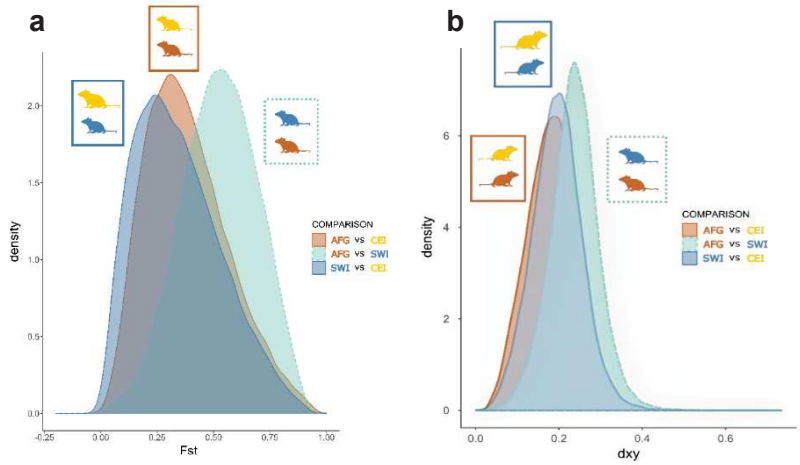


Figure 3 - a) Density plot of autosomal Fst pairwise-estimates; b) Density plot of autosomal Dxy pairwise-estimates

We further investigated these alternative hypotheses using the information carried by allele frequency covariances. We estimated a population tree using the graph-based model implemented in TreeMix, by running simulations of 0 to 5 migrations events (with three replicates per event). Without modelling migration (admixture) events, and according to the likelihood estimates, none of the models fit the data (residuals > 52 s.e.). We however obtain a substantial fit improvement when modelling one migration event (admixture weight = 31 %, residuals < 1.8 s.e.) which indicates significant migration between *domesticus* from SWI and CEI (Figure 4). Note the direction of migration is not identifiable by the model. Adding further migration events does not contribute to better explain the data (Figure 4 and Sup. Fig. 7).

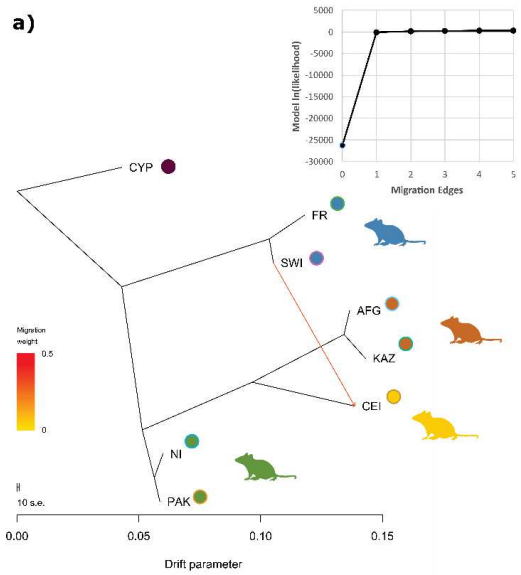


Figure 4 - TreeMix phylogram with one migration event. Migration edge was coloured according to migration weight, the per cent ancestry received from the source population. Branch lengths are proportional to the evolutionary change (the drift parameter) and terminal nodes were labelled with population codes and colours (see Sup. Table 1).

In order to orientate the migration thus detected, and quantify its contribution, we also applied statistical tests based on *Patterson's D* (ABBA-BABA) statistics, which quantify the distribution of derived alleles among populations and its imbalance caused by genetic exchanges. Assuming a scenario with four populations related by the rooted tree (((P1,P2),P3),O) and defining the ancestral allele as A and derived as B, it is expected that under a scenario without gene flow the ABBA and BABA patterns should occur with equal frequencies (due to incomplete lineage sorting) (Green et al. 2010; Durand et al. 2011). On the other hand, a significant deviation from that expectation is consistent with introgression between P3 and either P1 or P2. Here we have used the *Dsuite* toolkit (Malinsky, Matschiner, and Svardal 2021) to calculate *D* and *f4*-ratio statistics for all possible P1, P2 and P3 combinations (the Pakistan population was excluded to prevent bias due to its low number of individuals), with *cypriacus* fixed as the outgroup (O). We found a significant excess of shared derived alleles between *domesticus* and CEI – Table 1 (see Sup. Table 3 for complete results). The *f4*-ratio gives an estimate of roughly 30% of *domesticus* contribution to CEI. This is comparable to the estimate of migration weight derived from TreeMix and helps clarifying its directionality.

Table 1 - Results from Patterson's D and f4-ratio tests between populations with significance values

	P1	P2	P3	Patterson's *D*	*Z* score	*P-value*	*F4-ratio*
Autosomes	AFG	CEI	SWI	0.387	114.6	<0.000001	0.291
Autosomes	KAZ	CEI	SWI	0.386	121.5	<0.000001	0.289

We applied another related method of inference of admixture graphs, that uses correlations of allele frequencies (*f*-statistics), and used a heuristic (qpBrute) that can explore all possible topologies while allowing a maximum of two admixture events among sampled populations. We thus explored all possible relationships of the CEI population to the three subspecies (using *M. cypriacus* as an outgroup). Only one of the models provided a good fit, explaining all 68 f4-statistics without significant outliers. This model is shown in Figure 5, and differs from the TreeMix model in several respects. The position of *domesticus* as an outgroup is here little supported, except through an inferred (and modest, 8%) contribution of an ancestral lineage to the composition of SWI. Another difference is the inference of a contribution of *castaneus* (25%) to the ancestor of the two *musculus* populations. However, the model agrees with our major point of interest, which is the hybrid origin of CEI. In this admixture graph, CEI is modelled as the descendant of a population deriving for 60% from a

common ancestor with the two *musculus* populations, and for 40% from a common ancestor with the SWI *domesticus* population. We note that, as in the Treemix model, the *domesticus*-like ancestor of CEI is quite close (in terms of drift distance) to the SWI *domesticus* population, while its *musculus*-like ancestor is relatively far from the sampled *musculus* populations. This latter point might explain why the ADMIXTURE analysis tended to favour k=4, the solution where CEI is considered a different population rather than an admixture. However, under the ADMIXTURE model with k=3, the inferred contributions of *domesticus* and *musculus* to CEI are similar to those inferred in this admixture graph.

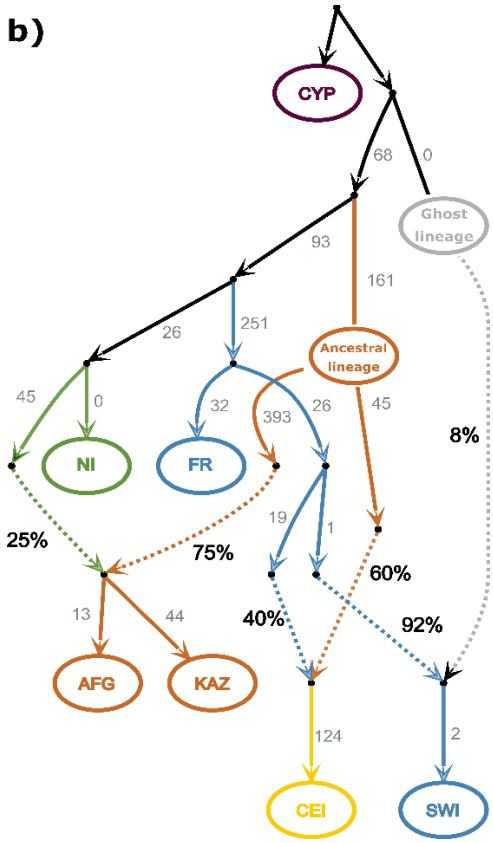


Figure 5 - Best-fitting admixture graph model suggests a hybrid origin for the Iranian lineage (CEI).

All methods presented above converge towards the hypothesis of an admixed origin of CEI. However, they are all somewhat related and based on the same type of data (allele frequencies). We thus challenged this conclusion using a very different method. The PSMC method infers the piecewise time distribution of past coalescence events based on variations of heterozygosity along a diploid chromosome. When the two haplotypes making up this diploid genotype are sampled from two different populations, the PSMC results can be used to date the divergence between the populations (corresponding to the time of rapid decrease of coalescence rates). We applied this approach on the X chromosome, because using males we could have good quality phased haplotypes from each population. We could then apply

the PSMC method to artificial diploids made of one haplotype from each of any two populations. The results are shown in Figure 6 for all pairs of populations (displaying the results from the haplotype with the best sequence coverage in each population). Note that this graph does not represent the rate of coalescence through time, but its inverse which, if the haplotypes were sampled from a panmictic population, would represent the effective size of the population. Therefore, in this graph, population divergence is inferred to have occurred at the time when population size grows to infinity (i.e. when instantaneous coalescence rate decreases towards zero). Focusing on the pairs including CEI in Figure 6, we see that CEI appears to have separated from *castaneus* (NI) relatively anciently, roughly at the same time as all pairs of peripheral subspecies from one another (65-200 kY according to the mutation rate and generation times chosen here). The graph may suggest that *domesticus* and *castaneus* diverged before the other pairs (100-200 kY) but in any case, all these divergences are ancient. The interesting result for our purpose concerns the divergences of CEI from *domesticus* and *musculus*. The profiles are very different between these two pairs, and inform us about the relationships between the three. The separation of CEI from *domesticus* appears to have started at the same time as that between *domesticus* and *musculus*. This is coherent with the population tree inferred with Treemix, where CEI and *musculus* are sister populations. However, this ancient decrease in coalescence rate then plateaus, which can be interpreted as the result of readmixture of *domesticus* and CEI. The PSMC profile between CEI and *musculus* is compatible with a simple divergence, that seems to have occurred 30-70 kY ago and to have started roughly at the same time as the admixture with *domesticus*. This analysis is compatible with our hypothesis that CEI descends from a population that admixed with the *domesticus* branch as it was diverging from the *musculus* branch, in agreement with all other analyses presented above.

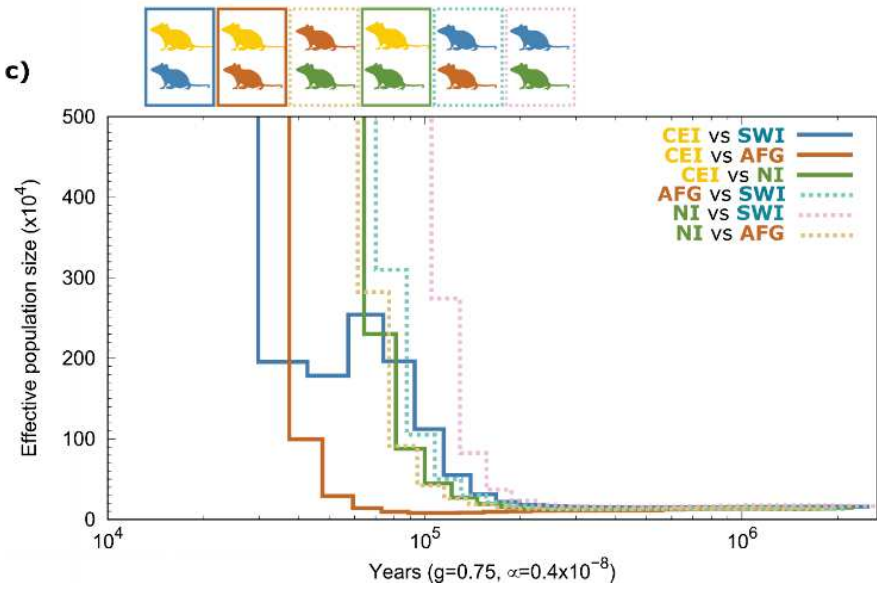


Figure 6 - PSMC plot based on the artificial pseudo-diploid X- hybrid chromosomes based on males from different lineages.

Ancestry deconvolution in the Central Iranian population

The discovery of extensive admixture in CEI between populations related to the *domesticus* and *musculus* branches contrasts with the situation found in Europe, where these two subspecies form a narrow tension zone preventing extensive admixture, and suggesting the existence of numerous hybrid incompatibility loci. It is therefore interesting to determine which genomic regions were able to admix in CEI (and are thus unlikely to participate in hybrid incompatibilities), and which were not (and are thus more likely to contain loci involved in incompatibilities).

For that purpose, we used ELAI, a local ancestry inference method based on an HMM and the partition of linkage disequilibrium between intra and interpopulation layers. We assigned the AFG and SWI as parental *musculus* and *domesticus* populations, respectively, and CEI as the target population. This method assigned on average across autosomes about 60% ancestry to *musculus* and 40% to *domesticus*, in good agreement with the other methods (ADMIXTURE, admixture graph and f4-ratio statistics, see above) – Sup. Fig. 8. There were variations of these ancestry proportions among autosomes, and chromosome 17 appeared as an outlier with 70-30% proportions, however the results for this chromosome are less reliable due to the exclusion of t-haplotype bearers and thus lower sample size. The X-chromosome deviates from the autosomal pattern, with 80% assigned to *musculus*-like and 20% to

domesticus-like ancestry. This limited admixture of the X is in line with the demonstrated strong involvement of the X in hybrid incompatibilities, and may indicate that the ancestral populations that admixed in CEI carried such incompatibilities, thus reinforcing the interest for this population.

To analyse these results in more details, we partitioned the autosomes and X-chromosome into segments according to their inferred local ancestry in the CEI sample (see material and methods), as estimated by the number of haplotypes of *domesticus* ancestry among the haplotypes sampled. We then calculated three population statistics (*Fst*, *Dxy* and Pi) in each of these categories of segments. As can be seen on Figure 7, average *Fst* between the parental populations (*domesticus* and *musculus*) did not vary among these categories, indicating no overall relationship between the level of differentiation between the parentals and the level of admixture in CEI (indicating the absence of an overall bias of the ancestry inference method). As expected, for both parental populations, *Fst* between a parental and CEI was strongly negatively correlated with the proportion of ancestry of CEI from this parental: *Fst* between CEI segments of pure *domesticus* (resp. *musculus*) origin and *musculus* (resp. *domesticus*) was similar to that between *domesticus* and *musculus*, and decreased with increasing *musculus* (resp. *domesticus*) ancestry. Note however that this overall decrease was not monotonous, with a relative increase at the end of this overall gradient. Similarly, there was no global effect of the admixture category on the average *Dxy* between the parental populations. As expected, we observe a clear negative correlation between *Dxy* from CEI to either parental and the contribution of this parental to its ancestry: *Dxy* between CEI segments of pure *domesticus* (resp. *musculus*) ancestry and *musculus* is similar to that between *domesticus* and *musculus*, and decreases with increasing *musculus* (resp. *domesticus*) ancestry.

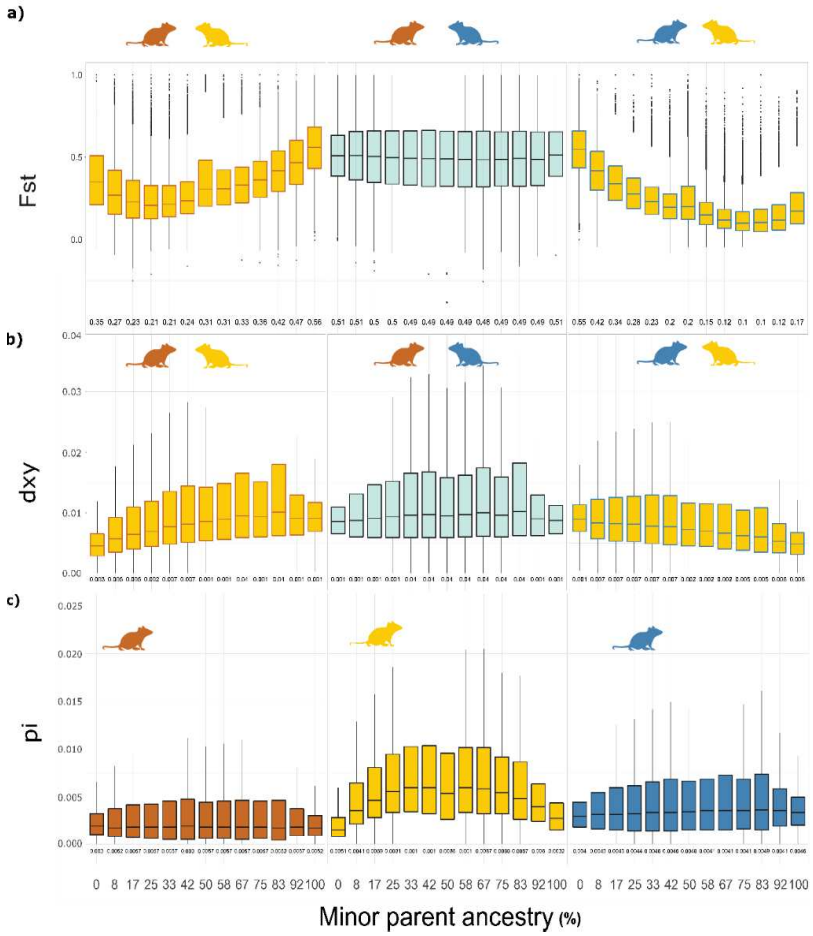


Figure 7 - Parental contributions to the Iranian lineage. a) Comparison between minor parental ancestry and genetic differentiation (*Fst*): as expected *Fst* increases for the pairwise musculus - central Iranian and negative between domesticus - central Iranian and negative between *domesticus* - central Iranian. No correlation on the pair *musculus* - *domesticus*. b) Comparison between minor parental ancestry and absolute sequence divergence (*dxy*), as expected it follows the differentiation pattern. c) Lineage heterozygosity (pi). High heterozygosity on the hybrid lineage

Finally, nucleotide diversity (heterozygosity) did not vary across categories of segments in the parental populations, again showing that the ancestry assignation was not biased by this characteristic of the parental genomes. In CEI, heterozygosity of the segments of pure ancestry was similar to that of their corresponding inferred parent, but increased above these levels with increasing levels of inferred admixture of the segments, up to values higher than those of the parental populations. Note that this does not correspond to an overall higher heterozygosity in CEI than in the parentals (Sup. Fig. 9), thus reinforcing the hypothesis that the elevated heterozygosity of some CEI segments results from admixture of the differentiated *domesticus* and *musculus* populations.

We also verified the coherence of the results of the ancestry inference by inspecting phylogenetic signals in the genomic segments of pure or quasi-pure inferred ancestry in CEI.

We first selected SNPs that were diagnostic of the parental populations and could verify that their majority state in CEI corresponds to the inferred major ancestry of the genomic segments (Sup. Fig. 10). We also applied the TWISST method, which determines in genomic segments the majority topology among all possible combinations of one haplotype from each population. We found that overall, the topologies were concordant with the inferred ancestries*, i.e*. CEI was most often sister to the inferred major ancestry population of the pure or quasi-pure segments (Sup. Fig. 11 and Sup. Fig. 12).

All results presented above appear coherent with the results of the local ancestry inference, and thus reinforce the evidence for CEI being an admixed population.

Selection and ancestry in Central Iran

The local ancestry inference we performed allows us to address the question of the determinants of the patterns of admixture in CEI, and in particular the role of selection. Because the effects of selection at a locus on the diversity at neighbouring loci are influenced by the local intensity of recombination, we explored the relationship between recombination and admixture along the genome. The reduced admixture of the X chromosome may suggest that incompatibilities between the X and factors elsewhere in the genome have prevented free admixture. If numerous loci are involved in hybrid incompatibilities, one would predict a lower representation of the genome of the minor parent (*domesticus*) in regions of low recombination, because of tighter linkage to the incompatibility loci in these regions. We however found a slight but significant negative correlation between *domesticus* ancestry and recombination, the pattern being slightly stronger on the X chromosome than on the autosomes. (Table 2 and Sup. Fig. 13-15). We found similar slight negative correlations between gene density and recombination Sup. Fig. 16. Therefore, genomic regions with lower recombination and lower gene density tend to display more *domesticus*-like contribution in the CEI population. Note that gene density is weakly, though significantly, positively correlated with recombination.

Table 2 - Spearman's correlation between domesticus-like ancestry and recombination rate and gene density

domesticus-like ancestry **vs**	**Recombination rate**	**Gene density**
Autosomes	ρ = -0.065 p = 0. 00000002	ρ = -0.052 p = 0. 00000002
Chromosome X	ρ = -0.12 p = 0.000057	ρ = -0.17 p = 0. 0000002

Mitochondrial DNA variation

We assembled the full mitochondrial sequences of all individuals and inferred their phylogeny (Sup. Fig. 17). As expected, *domesticus* and *musculus* formed two monophyletic groups. A third group, sister to *musculus*, included all *castaneus* samples. Inside this group, one branch is exclusive to Pakistan, but contains one CEI sample. Its sister branch splits into two branches, one containing all Indian samples, and the other all remaining CEI samples. These results reinforce the idea of the distinctiveness between *musculus* and the *musculus*-like ancestor of CEI, suggested by the analysis of allele frequencies (Treemix and Admixture graph reported above).

Discussion

House mice have long been known to be structured in several subspecies, whose definition and genetic relationships had been established based on geographic sampling mostly focusing on allopatric populations (mostly Europe for *domesticus* and *musculus*, and India or East Asia for *castaneus*), far away from the supposed cradle of differentiation. This led to the recognition of three major and well differentiated, parapatric, subspecies (Boursot et al. 1993). Studies of their contacts and admixture have concerned the periphery of the range of the species, namely Europe and East Asia, where secondary contacts result from the recent expansion of the species from its cradle (Boursot et al. 1993). Here we extended sampling to a geographical region, Iran, where all subspecies are, or may have anciently been, in contact. The complexity and originality of the genetic makeup in this region had been suspected based on mtDNA and microsatellite variation, from which central Iran and South-East Iran appeared as two new entities with their own genetic characteristics (Rajabi-Maham et al. 2012, Hardouin et al. 2015). The analysis of morphological characters of Iranian mice had also revealed specificities of Central Iranian mice, leading to the proposal of a new subspecies (that was named either *M. m. isaticus* or *M. m. isatissus* in different papers by the same authors) (Hamid et al. 2017, Molavi et al. 2015). Based on whole genome sequences, we here confirm Central Iran as a genetic entity distinct from the three previously described subspecies (in the PCA or admixture analysis). Whether this deserves revising or expanding the taxonomy can be somewhat arbitrary, and premature in the absence of a good understanding of the relationship of this new entity with existing ones. We have tried here to gain such understanding, leaving taxonomic issues aside.

We clearly demonstrate that the history of differentiation of Central Iran from the three "canonical" subspecies cannot be explained by a simple tree-like history of successive differentiation without substantial admixture. All methods employed here suggest that a major event of admixture between the *domesticus* and *musculus* lineages is at the origin of the Central Iranian population. For this reason, most of the results discussed below imply only *domesticus*, *musculus* and CEI.

The simplest, although not the strongest, evidence for this major admixture event are the distributions of pairwise divergences (*Dxy*) for the autosomes and the X. In both cases, the *domesticus*-*musculus* divergence is the highest. For the X, the *domesticus*-CEI divergence is only slightly lower, and the *musculus*-CEI divergence much lower. This fits a history of introgression from *domesticus* into CEI, the latter being sister to *musculus.* The apparent bimodality of the *musculus*-CEI distribution might result from this admixture, the higher mode corresponding to genomic regions with a *domesticus* contribution in CEI*.* For the autosomes, the *domesticus*-CEI distribution is more shifted to the left, and the *musculus*-CEI less so, making them closer to one another. This fits the same interpretation, but with more contribution of *domesticus* to CEI for the autosomes than the X, making the autosomal distributions of the two pairs involving CEI closer to each other.

Strong confirmation for a contribution of the *domesticus* branch to CEI comes from the hybrid PSMC analysis of the X chromosome, namely the shape of the curve for the *domesticus*-CEI pair. CEI appears to have started separating from *domesticus* roughly at the same time as it did from *castaneus*, but to have later admixed secondarily with *domesticus*, before separating again. This analysis also confirms that *musculus* and CEI are sister since they separated more recently than all other pairs. These two conclusions are supported by the Treemix analysis, based on a very different type of data (allele frequencies), and the f4 analysis confirms the direction of the introgression. The admixture graph gives a more complex picture, since the best model implies several admixture events. It however agrees with all the other methods to model CEI as admixed between the *domesticus* and *musculus* lineages.

The methods based on allele frequencies (Treemix and Admixture graph) allow us to better understand the relationship between the parental populations of the admixture leading to CEI and our sampled *domesticus* and *musculus* populations. The domesticus-like donor appears closely related to our SWI population, but slightly more distantly related to the FR population. This could be due to drift during the expansion of *domesticus* to the West of its Iranian cradle, or to the ancestral *domesticus* population having been structured, only one of

its components contributing to CEI. The *musculus*-like donor appears quite distinct from our sampled *musculus* populations in terms of allele frequency divergence. This is also attested by the mitochondrial DNA data, showing that CEI and *musculus* are reciprocally monophyletic, and relatively distant. It is furthermore reflected in the distributions of autosomal *Fst*, since the *domesticus*-CEI distribution is shifted to lower values as compared to the *musculus*-CEI distribution, presumably because SWI is a better ancestor proxy than is AFG. Note this pattern is not seen in the X-chromosome distribution of *Fst*, probably due to the lower contribution of *domesticus* to CEI for the X. And the three pairwise X distributions are equally centred around high values.

We have attempted to partition the genomes of the CEI individuals among the two ancestries. The performance of the ancestry deconvolution method used (ELAI) is likely to have been negatively impacted by the *musculus* parental population used being an imperfect proxy of this ancestor. However, the *domesticus* population used as parental seems a much better proxy, and the results obtained point to genome-wide proportions of ancestries that are comparable to those obtained with allele-frequency methods, which is reassuring. Furthermore, the distributions of classical statistics among genomic segments according to their level of *domesticus* contribution in CEI fit well the expectations of the admixture model for all three statistics (*Fst*, *Dxy* and Pi). This fit is remarkable since the ancestry deconvolution method is based on linkage disequilibrium, which is not accounted for in the statistics whose distributions are examined. This provides additional strong support for the admixture hypothesis, and is reassuring about the performance of the ancestry inference, despite the inadequacy of the parental populations used for the ancestry inference method used.

Based on the ancestry deconvolution performed, we correlated variations of ancestry proportions along the genome with those of genomic characteristics such as recombination rate or gene density, because this could inform us about the evolutionary forces modulating admixture along the genome. If the dominant force was incompatibilities preventing admixture, one would expect (if incompatibility loci were numerous and spread across the genome) a positive correlation between recombination and minor ancestry proportion in the admixed population. The reason is because higher recombination would allow neutral loci to more easily become independent of the negative influence of neighbouring incompatibility loci. If on the other hand the major driving force was positive selection facilitating introgression, one would expect a negative correlation between recombination and minor ancestry proportion. The reason is that in regions of lower recombination, hitch-hiking by favoured introgressing alleles would cause a greater than average minor ancestry level at neighbouring neutral loci across longer genomic segments, thus inflating the proportion of minor ancestry in low

recombining regions. Here we observed moderate, but significant, negative correlations of both recombination and gene density with *domesticus* ancestry in CEI. Note that gene density and recombination rate are known to be positively correlated. Whether this indicates that the dominant force determining the proportion of *domesticus* in CEI is positive selection will however need further investigation. In fact, it is likely that both phenomena predicted above influence such correlations simultaneously and antagonistically. The resulting intensity and direction of the correlations are difficult to predict since this depends on several parameters such as the proportion of minor ancestry, the number of loci subject to positive or negative selection, and the intensity of selection (Duranton and Pool 2021). Therefore, it is difficult to interpret the weak negative correlation we find between recombination and admixture at this stage.

We have found that the *domesticus* contribution to CEI is lower for the X than the autosomes (roughly half). This could be taken as an indication that there is counterselection on *domesticus* X introgression, which would reveal the existence of X-linked hybrid incompatibilities between the progenitors of CEI. Reduced admixture of the X chromosome is a common phenomenon, and reflects the "large X effect", *i.e*. stronger involvement of the X in hybrid incompatibilities, which underlies Haldane's rule. However, sex-biased contributions to the admixture could also contribute to such differences, since male migration has a lower impact on X introgression than on the autosomes, due to male X hemizygosity, while female migration affects all chromosomes equally (leaving the Y aside of course). The fact that *domesticus* did not contribute mtDNA to CEI while it contributed 40% of the autosomal genome would be in line with this hypothesis. However, there is no *domesticus* contribution to CEI for the Y-chromosome either (Sup. Fig. 18). A way to reconcile these apparently contradictory observations by minimally invoking *ad hoc* selection would be to suppose that there was invasion of the territory of the *musculus*-like ancestor of the CEI population by *domesticus*, with strong male-biased migration (explaining the absence of mtDNA admixture and lower X invasion), and an advantage of the non-*domesticus* Y chromosome (or a disadvantage of the *domesticus* Y chromosome). Further investigation would be needed to evaluate the likelihood that such purely demographic processes could generate an absence of mtDNA introgression together with 40% autosomal introgression, and to estimate the expected level of X introgression. We could also hypothesize that the admixture occurred by an invasive process in the other direction, *i.e.* invasion of a *musculus*-like population into a resident *domesticus* population. The interpretation would thus have to be reversed, implying that female biased migration accounts for the fixation of the invading mitochondrial genome and the excess of X invasion as compared to autosomes. However, the fixation of the non-*domesticus* Y from the

invader population would not be expected without selection in this case either. However, the propensity of males to migrate more than females is the most common pattern in Mammals, and several inferences of the history of colonisation of new territories by the house mouse have suggested that present mitochondrial DNA variation witnesses primary colonisation events rather than consecutive secondary invasions (Jones et al. 2012), suggesting either higher male migration or the higher difficulty of newly immigrating females to establish and reproduce in resident populations than males. We therefore favour the hypothesis of a *domesticus* invasion into an area previously occupied by the *musculus*-like ancestor of CEI, and retaining the hypothesis of an advantage of the non-*domesticus* Y in such a situation of hybridization. In several parts of the European hybrid zone between *domesticus* and *musculus*, the *musculus* Y chromosome is predominant in the centre of the hybrid zone and slightly beyond (Macholan et al. 2008), and it has been shown that it contributes to rescue the fertility of hybrid males, which is reduced by incompatibilities (Albrechtova et al. 2012).

Although the potential power of such purely demographic models to explain the dependency of the admixture pattern to sex-linked transmission must be evaluated, it would be surprising that admixture in CEI be solely controlled by such stochastic processes. The divergence between the progenitors of the CEI population was already quite substantial (as attested by the long branch between SWI and CEI, despite the SWI contribution to CEI), leaving time for incompatibilities to have accumulated. An interesting question is then how the admixture pattern in CEI could help pinpoint genomic regions involved in such incompatibilities, which would be likely to also affect the interaction between *domesticus* and *musculus* in their European hybrid zone. Given the high level of admixture of the CEI population, it is likely that most incompatibilities have been sorted, the CEI population retaining mostly compatible combinations of alleles, be they of *domesticus* or *musculus* origin. A simple prediction of the Dobzhansky-Muller model is that incompatibility loci should lie in regions of pure ancestry in CEI (or inversely that they should not lie in regions of admixed ancestry). The potential of this prediction to be used to narrow down the list of candidates for incompatibilities is however hampered by the stochasticity of lineage sorting (in addition to that of sampling of course), potentially leading to many false positives. The combined ancestry at pairs of loci might not be very informative either, because the sorting of incompatibilities could have led to retain either the ancestral alleles at both interacting loci (in which case the two loci would have different ancestries in CEI), or to retain the ancestral allele at one locus and the derived at the other (in which case they would have the same ancestry in CEI). It could however be that a fine analysis of the genomic distribution and gene content of the regions of pure ancestry in CEI be suggestive of the genomic architecture and biology of reproductive isolation. Further

information could also be gathered from the study of the present contact zones of CEI with *domesticus* and *musculus*. However, their genetic analysis would be rendered difficult by the hybrid nature of CEI, making the population ancestry inference in the hybrid zones challenging, with variable power across genomic regions depending on their ancestries in CEI. Comparisons of introgression patterns among genomic regions of a single pure ancestry in CEI might point to potential candidates.

It is striking that *domesticus* and *musculus* form a tension zone in Europe and were able to admix to such an extent in Iran. There could be several reasons for this contrast. One could be linked to ecology and environment. The contact in Europe occurred when the mice were already commensal with humans. Therefore, their habitat and migration were at least partly, and perhaps predominantly, determined by human installations and human movements, and little dependent on climatic conditions, that are buffered by this lifestyle (and are not very contrasted across their zone of interaction anyway). The admixture in Central Iran is much older than commensalism with human, and could have lasted for much longer than the few thousand years of secondary contact in Europe. Climatic fluctuations during this long period could have participated in favouring invasions such as the ones we proposed above. Another possibility would be that the degree of reproductive isolation has increased after European *domesticus* and *bona fide musculus* derived from their Iranian ancestors, the progenitors of CEI. Indeed, we found some divergence of the French *domesticus* population from SWI *domesticus* (a close relative of the *domesticus* contributor to CEI), and even more between *musculus* and the non-*domesticus* ancestor of CEI (Admixture graph results). Some of the incompatibility loci of the European hybrid zone could thus be mined in parts of the genome that have diverged during the separation of *domesticus* and *musculus* from their Iranian ancestors. This model of differentiation would give the situation some of the characteristics of a ring species, with gradual and independent differentiation of two lineages from a common ancestor, and increasing reproductive isolation along this gradient. The first step of the gradient would be Iran, where the ancestors of *domesticus* and *musculus* were able to admix extensively, and the second Europe, where the ring closes with a clear tension zone. Behavioural reinforcement, that was demonstrated to occur in the European hybrid zone, might have evolved recently and participate in the greater isolation in Europe.

Our data allow us to examine the broader picture of the history of differentiation of the whole species into its recognised subspecies, and compare our results with those of previous analyses. Attempts to reconstruct the history of differentiation of these three entities have been based on simple models of divergence with gene flow, and all pointed to models of divergence with gene flow. The most recent such attempt, based on the largest number of loci, and on a

sampling largely overlapping with ours, suggested a model of successive divergence of *domesticus*, soon followed by the divergence between *musculus* and *castaneus*, with significant amounts of gene flow, particularly between *castaneus* and both *domesticus* and *musculus*, with more migration from the latter two into *castaneus* than the reverse (Phifer-Rixey, Harr, and Hey 2020). They also inferred migration from *domesticus* into *musculus*. Apart from the underlying population tree, our results have little in common with these results. The analyses of these authors are based on the likelihood of coalescence patterns at many loci under simple models of divergence with gene flow. The methods we used are based on using correlations of allele frequencies at very many loci to fit models allowing divergence and secondary admixtures. Furthermore, our sampling includes CEI, and the strong admixture signal in this population has a strong influence on the final model retained. It is therefore difficult to compare the results of the two studies. By construction our model is more complex in terms of population history, and we can try to evaluate the plausibility of this complex model. One of the complexities is the contribution of a ghost lineage to the ancestry of SWI. Previous studies have described that house mice from the Arabian Peninsula possess a mitochondrial lineage branching deeply into the mtDNA phylogeny of the species (the so called "*gentilulus*" branch) (Duplantier et al. 2002, Suzuki et al. 2015). This indicates past isolation of this region, the population of which could have more recently admixed with *domesticus* from Iran, which would account for this ghost population, that is inferred to have contributed to SWI but not FR *domesticus*. Another interesting suggestion of the admixture graph is the existence of a proto-*musculus* population, at the origin of CEI, and from which *musculus* would be derived (rather anciently given the mtDNA phylogeny and the length of the branch in the admixture graph). A final interesting suggestion is a contribution of *castaneus* to the makeup of the *musculus* populations. Based on all this we could propose the following model: the ancient presence of the ancestor of *domesticus* in the Arabian Peninsula, and the Near East, with some degree of differentiation between Southwest Iran and further West towards the Mediterranean (the population that will later colonise Europe). Central Iran was anciently colonised by a population related to *castaneus* (as witnessed by mtDNA variation). Rather anciently (according to the mtDNA phylogeny), *musculus* separated from this population (presumably to the north, where it is now found). The *musculus* mtDNA lineage was then lost from CEI and fixed in *musculus*. Our results also suggest that after diverging from the CEI branch, *musculus* admixed with *castaneus* (presumably through contacts in Eastern Iran). More recently (perhaps in association with humans), *musculus* invaded NE Europe and NE Asia. Relatively recently (as compared to the differentiation of *musculus* from the CEI branch), *domesticus* from SWI invaded the territory of the CEI population, leading to the present heavily admixed population in Central Iran.

Many aspects of the above scenario remain uncertain and will have to be specifically tested. However, the major result of our study, which is the admixed nature of CEI, appears well established. It is interesting that two lineages were able to extensively admix in the region close to their origin, but formed a tension zone preventing admixture after they geographically expanded. Some major questions now are to what extent these contrasted outcomes result from the different conditions under which the two admixtures occurred, the different times of admixture, the different times of divergence and degree of differentiation of the parental populations.

Material and methods

Additional detailed information on materials and methods with associated references is provided in *Supplementary Material*.

Data filtering, read mapping and genotype calling

Individual sequencing reads were processed following *Bettina et al*. (2016). In brief, filtered reads were mapped against the mouse genome reference sequence - *mm10* ('GRCm38 - Mm10 - Genome - Assembly - NCBI' n.d.) using bwa-mem (Li 2013). The sorting, marking and duplicates removal use performed with Picard tools software suite ('Picard Tools - By Broad Institute' n.d.) was used for. Raw SNP and indel calls were obtained following the GATK (Auwera and O'Connor 2020) 'Best Practice' instructions on joint genotyping. The raw .vcf files were subjected to the GATK VSQR SNP filtering step, which uses known variants as training data to predict whether a new variant is likely a true positive, or a false positive. As training data we used the file 'mgp.v5.merged.snps_all.dbSNP142.vcf' downloaded from ('Mouse Genomes Project - Wellcome Sanger Institute' n.d.) which was filtered for 'PASS' SNPs. In addition, we used very stringent hard filtering criteria on our own dataset, and included these SNPs as training sets as well (see details in *Bettina et al.* (2016). Due to an absence of a reliable indel reference dataset we decided to exclude all indels from our final dataset. Highly related individuals were removed, and t-carriers were identified due to its impact on the chromosome 17.

Population Structure and Genetic Relationships

An initial unsupervised population structure analysis was performed using the non-parametric principal component analysis (PCA), as implemented in PLINK 2 (Chang et al. 2015). The PCA was based on a subsample of bi-allelic SNPs at least 25 kb apart and without missing genotypes. Additionally, we apply ADMIXTURE v1.3.0 (Alexander, Novembre, and Lange 2009) and its implemented Bayesian Markov Chain Monte Carlo model (MCMC) on the pruned datasets and the cross-validation error was calculated for identifying the best K value. Five replicate runs were performed for each number of populations (k) set from 1 to 6. Replicate runs were analyzed using CLUMPAK (Kopelman et al. 2015) and DISTRUCT (Rosenberg 2004) was used to plot the results. The best number of populations, K, was

inferred using Evanno's delta K method (Evanno, Regnaut, and Goudet 2005), as implemented in CLUMPAK.

We estimated genetic relationships and admixture among the different populations using TreeMix v.1.13 (Pickrell et al. 2012; Pickrell and Pritchard 2012). We estimated the allele frequencies among the randomly sampled alleles and subsequently ran the TreeMix model accounting for linkage disequilibrium by grouping sites in blocks of 1,000 single-nucleotide polymorphisms (-k 1,000) setting the *Mus cypriacus* as root. Standard errors (-SE) and bootstrap replicates (-bootstrap) were used to evaluate the confidence in the inferred tree topology. After constructing a maximum-likelihood tree, migration events were added (−m) and iterated 10 times for each value of m (1–10) to check for convergence in the likelihood of the model as well as the explained variance following each addition of a migration event. The inferred maximum-likelihood trees were visualized with the in-built TreeMix R script plotting functions.

Dsuite toolkit (Malinsky, Matschiner, and Svardal 2021) was used to test gene flow among the different populations through the calculation of genome-wide *D*-statistics (ABBA-BABA test) (Green et al. 2010; Durand et al. 2011) and f4-ratio (Patterson et al. 2012) for all possible P1, P2 and P3 combinations. *Mus cypriacus* was fixed as the outgroup. A z-score with an absolute value of 3 or more was considered to be evidence of significant gene flow.

An admixturegraph analysis was performed using qpBrute (Liu et al. 2019; Leathlobhair et al. 2018), which enabled us to estimate shared genetic drift using f2, f3 and f4 statistics. At each step, insertion of a new node was tested at all branches of the graph, except the outgroup branch. In cases in which a node could not be inserted without producing f4 outliers (that is, $|Z| \geq 3$), all possible admixture combinations were also attempted. The resulting list of all fitted graphs was then passed to the MCMC algorithm implemented in the admixturegraph R package, to compute the marginal likelihood of the models and their Bayes factors.

X-hybrid Pairwise sequentially Markovian coalescent (hPSMC)

We used seqtk (Li [2012] 2022) to combine X haploid male X-chromosomes, to construct pseudo-diploid sequences. The PSMC model estimates the Time to Most Recent Common Ancestor (TMRCA) of segmental blocks of the genome and uses information from the rates of the coalescent events to infer Ne at a given time, thereby providing a direct estimate of the past demographic changes of a population (Li and Durbin 2011). The method has been validated by its successful reconstructions of demographic histories using simulated data and genome sequences from modern human populations (Li and Durbin 2011). A consensus sequence of each bam file was then generated in fastq format sequentially using the SAMtools

mpileup command with the –C50 option to reduce the effect of reads with excessive mismatches (Li et al. 2009); bcftools view –c to call variants; lastly, vcfutils.pl vcf2fq to convert the vcf file of called variants to fastq format. Pairs of fastq files were then merged using seqtk and PSMC inference carried out using the recommended input parameters for human autosomal data (Li and Durbin 2011), i.e. 25 iterations, with maximum TMRCA (Tmax) = 15, number of atomic time intervals (n) = 64 (following the pattern 1*4 + 25*2 + 1*4 + 1*6), and initial theta ratio (r) =5. Plots were scaled to real time as per, $20\mu X=\mu A[2(2+\alpha)]/[3(1+\alpha)]$, assuming a ratio of male-to-female mutation rate of $\alpha = 2$ (Miyata et al. 1987) and an autosomal mutation rate (μA) of 4.1x10-9 substitutions/nucleotide/generation. This gave us an estimated $\mu X = 3.3 \times 10\text{-}9$ substitutions/nucleotide/generation. Only males were used in these analyses

Ancestry inference in the central Iranian population

To perform the ancestry deconvolution analysis of central Iranian genomic segments we used the Efficient Local Ancestry Inference (ELAI) method (Guan 2014). This method implements a two-layer HMM (hidden Markov model) to infer local ancestry of admixed individuals without prior definition of window sizes, by looking at two layers of linkage-disequilibrium—within and among defined groups. It returns at each variable position in the genome the most likely proportions of ancestries (true values being expected to take values 0, 1, or 2 in two-way admixture). We ran ELAI on the unphased dataset and two population samples: CEI defined as the admixed population, and the Afghan *musculus* defined as one of the donors in the admixture and the southwest Iranian population of *domesticus* as the other one. We set the number of upper-layer groups to 2, representing *musculus* and *domesticus*, and that of lower-layer clusters to 10 (five times the number of upper-layer clusters, as recommended). We performed three different expectation maximization (EM) runs of 20 steps with mixture generation values of 25k, 50k and 100k and different random seeds. ELAI results were averaged over the five independent runs. Sites with a proportion of *musculus* or *domesticus* ancestry between 0.5 and 1.5 were considered heterozygous for and those with values over >1.5 homozygous for introgression. For each individual, single ancestry fragments were defined as consecutive sites defined according to the above criteria.

Relationship between introgression and recombination rate and gene density

Spearman's rank correlation was applied to test the statistical relation of *domesticus* introgression prevalence with recombination and gene density. The introgression frequency at a given SNP position in the genome was measured as the number of ELAI introgressed fragments across individuals overlapping that SNP. Average recombination rate was estimate for each introgression frequency segment change, using the local recombination rate inferred in *Booker et al.* (Booker, Ness, and Keightley 2017). Gene density was calculated as number of coding sites (information extracted from Ensembl) along each given introgression frequency segment change. To consider introgression frequency, recombination rate, and gene density jointly, we calculated the partial correlation between local ancestry *domesticus* introgression frequency and the recombination rate, controlling for the number of coding bp in a given position. To ensure some degree of independence, we subsampled SNPs that were at least 10 kb apart. Chromosome 17 was excluded given their known structural differences between *t*-carriers.

References

Albrechtova, Jana, Tomàs Albrecht, Stuart J. E. Baird, Milos Macholan, Geir Rudolfsen, Pavel Munclinger, Priscilla K. Tucker, and Jaroslav Pialek. 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. doi: 10.1098/rspb.2012.1802.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. 'Fast Model-Based Estimation of Ancestry in Unrelated Individuals'. *Genome Research* 19 (9): 1655–64. https://doi.org/10.1101/gr.094052.109.

Auwera, Geraldine A. Van der, and Brian D. O'Connor. 2020. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Incorporated.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science. 2010 Feb 12;327(5967):836-40. doi: 10.1126/science.1183439. Epub 2009 Dec 31. Erratum in: Science. 2010 May 7;328(5979):690. PMID: 20044539; PMCID: PMC4295902.

Bimova, B. V., M. Macholan, S. J. Baird, P. Munclinger, P. Dufkova, C. M. Laukaitis, R. C. Karn, K. Luzynski, P. K. Tucker, and J. Pialek. 2011. "Reinforcement selection acting on the European house mouse hybrid zone." *Mol Ecol* 20 (11):2403-24. doi: 10.1111/j.1365-294X.2011.05106.x.

Booker, Tom R., Rob W. Ness, and Peter D. Keightley. 2017. 'The Recombination Landscape in Wild House Mice Inferred Using Population Genomic Data'. *Genetics*. https://doi.org/10.1534/genetics.117.300063.

Boursot, P., J.-C. Auffray, J. Britton-Davidian, and F. Bonhomme. 1993. "The evolution of house mice." *Annual Review of Ecology and Systematics* 24:119-152. doi: 10.1146/annurev.es.24.110193.001003.

Britton Davidian, J., Fabienne Fel-Clair, J. Lopez, Paul Alibert, and P Boursot. 2005. "Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids." *Biol J Linn Soc* 84 (3):379-393.

Chang, Christopher C, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. 2015. 'Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets'. *GigaScience* 4 (1): s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8.

Coyne, Jerry A., and H. Allen Orr. Speciation. Vol. 37. Sunderland, MA: Sinauer Associates, 2004.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. 'The Variant Call Format and VCFtools'. *Bioinformatics* 27 (15): 2156–58. https://doi.org/10.1093/bioinformatics/btr330.

Dierckxsens, Nicolas, Patrick Mardulyn, and Guillaume Smits. 2016. 'NOVOPlasty: *De Novo* Assembly of Organelle Genomes from Whole Genome Data'. *Nucleic Acids Research*, October, gkw955. https://doi.org/10.1093/nar/gkw955.

Dod, B., L.S; Jermiin, P. Boursot, V.M. Chapman, J.T. Nielsen, and F. Bonhomme. 1993. "Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone." *J Evol Biol* 6:529-546.

Duplantier, J-M, A Orth, J Catalan, and F Bonhomme. 2002. "Evidence for a mitochondrial lineage originating from the Arabian peninsula in the Madagascar House Mouse (Mus musculus)." *Heredity* 89 (2):154-158.

Durand, Eric Y., Nick Patterson, David Reich, and Montgomery Slatkin. 2011. 'Testing for Ancient Admixture between Closely Related Populations'. *Molecular Biology and Evolution* 28 (8): 2239–52. https://doi.org/10.1093/molbev/msr048.

Duranton, Maud, and John E. Pool. 2021. "Interactions between natural selection and recombination shape the genomic landscape of introgression." *bioRxiv*:2021.10.12.462572. doi: 10.1101/2021.10.12.462572.

Duvaux, Ludovic, Khalid Belkhir, Matthieu Boulesteix, and Pierre Boursot. 2011. "Isolation and gene flow: inferring the speciation history of European house mice." *Molecular Ecology* 20 (24):5248-5264. doi: 10.1111/j.1365-294X.2011.05343.x.

Elgvin, Tore O., Cassandra N. Trier, Ole K. Tørresen, Ingerid J. Hagen, Sigbjørn Lien, Alexander J. Nederbragt, Mark Ravinet, Henrik Jensen, and Glenn-Peter Sætre. 2017. 'The Genomic Mosaicism of Hybrid Speciation'. Science Advances 3 (6): e1602996. https://doi.org/10.1126/sciadv.1602996.

Evanno, G., S. Regnaut, and J. Goudet. 2005. 'Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study'. *Molecular Ecology* 14 (8): 2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Forejt, Jiri, Petr Jansa, and Emil Parvanov. 2021. "Hybrid sterility genes in mice (Mus musculus): a peculiar case of PRDM9 incompatibility." *Trends in Genetics* 37 (12):1095-1108. doi: https://doi.org/10.1016/j.tig.2021.06.008.

Good, Jeffrey M., Matthew D. Dean, and Michael W. Nachman. 2008. "A Complex Genetic Basis to X-Linked Hybrid Male Sterility Between Two Species of House Mice." *Genetics* 179 (4):2213-2228. doi: 10.1534/genetics.107.085340.

'GRCm38 - Mm10 - Genome - Assembly - NCBI'. n.d. Accessed 19 January 2022. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20/.

Green, Richard E., Johannes Krause, Adrian W. Briggs, Tomislav Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, et al. 2010. 'A Draft Sequence of the Neandertal Genome'. *Science* 328 (5979): 710–22. https://doi.org/10.1126/science.1188021.

Guan, Yongtao. 2014. 'Detecting Structure of Haplotypes and Local Ancestry'. *Genetics* 196 (3): 625–42. https://doi.org/10.1534/genetics.113.160697.

Hamid, H. S., J. Darvish, E. Rastegar-Pouyani, and A. Mahmoudi. 2017. "Subspecies differentiation of the house mouse Mus musculus Linnaeus, 1758 in the center and east of the Iranian plateau and Afghanistan." *Mammalia* 81 (2):147-168. doi: 10.1515/mammalia-2015-0041.

Hardouin, E. A., A. Orth, M. Teschke, J. Darvish, D. Tautz, and F. Bonhomme. 2015. "Eurasian house mouse (*Mus musculus* L.) differentiation at microsatellite loci identifies the Iranian plateau as a phylogeographic hotspot." *BMC Evol Biol* 15 (1):306. doi: 10.1186/s12862-015-0306-4.

Harr, Bettina, Emre Karakoc, Rafik Neme, Meike Teschke, Christine Pfeifle, Željka Pezer, Hiba Babiker, et al. 2016. 'Genomic Resources for Wild Populations of the House Mouse, Mus Musculus and Its Close Relative Mus Spretus'. *Scientific Data* 3 (1): 1–14. https://doi.org/10.1038/sdata.2016.75.

Janousek, V., L. Wang, K. Luzynski, P. Dufkova, M. M. Vyskocilova, M. W. Nachman, P. Munclinger, M. Macholan, J. Pialek, and P. K. Tucker. 2012. "Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between Mus musculus musculus and M. m. domesticus." *Mol Ecol* 21 (12):3032-47. doi: 10.1111/j.1365-294X.2012.05583.x.

Jones, E. P., K. Skirnisson, T. H. McGovern, M. T. Gilbert, E. Willerslev, and J. B. Searle. 2012. "Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region." *BMC Evol Biol* 12:35. doi: 10.1186/1471-2148-12-35.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms'. *Molecular Biology and Evolution* 35 (6): 1547–49. https://doi.org/10.1093/molbev/msy096.

Kopelman, Naama M., Jonathan Mayzel, Mattias Jakobsson, Noah A. Rosenberg, and Itay Mayrose. 2015. 'Clumpak: A Program for Identifying Clustering Modes and Packaging Population Structure Inferences across K'. *Molecular Ecology Resources* 15 (5): 1179–91. https://doi.org/10.1111/1755-0998.12387.

Leathlobhair, Máire Ní, Angela R. Perri, Evan K. Irving-Pease, Kelsey E. Witt, Anna Linderholm, James Haile, Ophelie Lebrasseur, et al. 2018. 'The Evolutionary History of Dogs in the Americas'. *Science*, July. https://doi.org/10.1126/science.aao4776.

Li, Heng. 2013. 'Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM'. *ArXiv:1303.3997 [q-Bio]*, May. http://arxiv.org/abs/1303.3997.

———. (2012) 2022. *Lh3/Seqtk*. C. https://github.com/lh3/seqtk.

Li, Heng, and Richard Durbin. 2011. 'Inference of Human Population History from Individual Whole-Genome Sequences'. *Nature* 475 (7357): 493–96. https://doi.org/10.1038/nature10231.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352

Li, Yue, Kazumichi Fujiwara, Naoki Osada, Yosuke Kawai, Toyoyuki Takada, Alexey P. Kryukov, Kuniya Abe, Hiromichi Yonekawa, Toshihiko Shiroishi, Kazuo Moriwaki, Naruya Saitou, and Hitoshi Suzuki. 2020. "House mouse Mus musculus dispersal in East Eurasia inferred from 98 newly determined complete mitochondrial genome sequences." *Heredity*. doi: 10.1038/s41437-020-00364-y.

Liu, Langqing, Mirte Bosse, Hendrik-Jan Megens, Laurent A. F. Frantz, Young-Lim Lee, Evan K. Irving-Pease, Goutam Narayan, Martien A. M. Groenen, and Ole Madsen. 2019. 'Genomic Analysis on Pygmy Hog Reveals Extensive Interbreeding during Wild Boar Expansion'. *Nature Communications* 10 (1): 1992. https://doi.org/10.1038/s41467-019-10017-2.

Macholan, M., S.J. Baird, P. Munclinger, P. Dufkova, B. Bimova, and J. Pialek. 2008. "Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone?" *BMC Evolutionary Biology* 8 (1):271-.

Macholán, Miloš, Stuart J. E. Baird, Petra Dufková, Pavel Munclinger, Barbora Vošlajerová Bímová, and Jaroslav Piálek. 2011. "ASSESSING MULTILOCUS INTROGRESSION PATTERNS: A CASE STUDY ON THE MOUSE X CHROMOSOME IN CENTRAL EUROPE." *Evolution* 65 (5):1428-1446. doi: 10.1111/j.1558-5646.2011.01228.x.

Macholan, Milos, Pavel Munclinger, Monika Sugerkova, Petra Dufkova, Barbora Bimova, Eva Bozikova, Jan Zima, and Jaroslav Pialek. 2007. "GENETIC ANALYSIS OF AUTOSOMAL AND X-LINKED MARKERS ACROSS A MOUSE HYBRID ZONE." *Evolution* 61 (4):746-771. doi: doi:10.1111/j.1558-5646.2007.00065.x.

Malinsky, Milan, Michael Matschiner, and Hannes Svardal. 2021. 'Dsuite - Fast D-Statistics and Related Admixture Evidence from VCF Files'. *Molecular Ecology Resources* 21 (2): 584–95. https://doi.org/10.1111/1755-0998.13265.

Miyata, T., H. Hayashida, K. Kuma, K. Mitsuyasu, and T. Yasunaga. 1987. 'Male-Driven Molecular Evolution: A Model and Nucleotide Sequence Analysis'. *Cold Spring Harbor Symposia on Quantitative Biology* 52 (January): 863–67. https://doi.org/10.1101/SQB.1987.052.01.094.

Molavi, Farahnaz, Jamshid Darvish, Farhang Haddad, and Maryam Matin. 2015. "Chromosome C-banding in Mus musculus L.1766 Strains Shows a Fixed Position for the Centromere and Variable Amounts in Different Populations." *Journal of Genetic Resources* 1 (2):83-88. doi: 10.22080/jgr.2015.1167.

'Mouse Genomes Project - Wellcome Sanger Institute'. n.d. Accessed 19 January 2022. https://www.sanger.ac.uk/data/mouse-genomes-project/.

Patterson, Nick, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping Zhan, Teri Genschoreck, Teresa Webster, and David Reich. 2012. 'Ancient Admixture in Human History'. *Genetics* 192 (3): 1065–93. https://doi.org/10.1534/genetics.112.145037.

Phifer-Rixey, Megan, Bettina Harr, and Jody Hey. 2020. "Further resolution of the house mouse (Mus musculus) phylogeny by integration over isolation-with-migration histories." *BMC Evolutionary Biology* 20 (1):120. doi: 10.1186/s12862-020-01666-9.

'Picard Tools - By Broad Institute'. n.d. Accessed 19 January 2022. https://broadinstitute.github.io/picard/.

Pickrell, Joseph K., Nick Patterson, Chiara Barbieri, Falko Berthold, Linda Gerlach, Tom Güldemann, Blesswell Kure, et al. 2012. 'The Genetic Prehistory of Southern Africa'. *Nature Communications* 3 (October): 1143. https://doi.org/10.1038/ncomms2140.

Pickrell, Joseph K., and Jonathan K. Pritchard. 2012. 'Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data'. Edited by Hua Tang. *PLoS Genetics* 8 (11): e1002967. https://doi.org/10.1371/journal.pgen.1002967.

Rajabi-Maham, Hassan, Annie Orth, Roohollah Siahsarvie, Pierre Boursot, Jamshid Darvish, and François Bonhomme. 2012. "The south-eastern house mouse Mus musculus castaneus (Rodentia: Muridae) is a polytypic subspecies." *Biological Journal of the Linnean Society* 107 (2):295-306. doi: 10.1111/j.1095-8312.2012.01957.x.

Raufaste, N., A. Orth, K. Belkhir, D. Senet, C. Smadja, S. J. E. Baird, F. Bonhomme, B. Dod, and P. Boursot. 2005. "Inferences of selection and migration in the Danish house mouse hybrid zone." *Biological Journal of the Linnean Society* 84 (3):593-616. doi: DOI 10.1111/j.1095-8312.2005.00457.x.

Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F., Mehlig, B. and Westram, A.M. (2017), Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol., 30: 1450-1477. https://doi.org/10.1111/jeb.13047

Reimand, Juri, Raivo Kolde, and Tambet Arak. 2019. *GProfileR: Interface to the 'g:Profiler' Toolkit* (version 0.7.0). https://CRAN.R-project.org/package=gProfileR.

Rosenberg, Noah A. 2004. 'Distruct: A Program for the Graphical Display of Population Structure'. *Molecular Ecology Notes* 4 (1): 137–38. https://doi.org/10.1046/j.1471-8286.2003.00566.x.

Smadja, C., and G. Ganem. 2005. "Asymmetrical reproductive character displacement in the house mouse." *Journal of Evolutionary Biology* 18 (6):1485-1493.

Suzuki, H., L. V. Yakimenko, D. Usuda, and L. V. Frisman. 2015. "Tracing the eastward dispersal of the house mouse, Mus musculus." *Genes Environ* 37:20. doi: 10.1186/s41021-015-0013-9.

Threadgill DW, Miller DR, Churchill GA, de Villena FP. The collaborative cross: a recombinant inbred mouse population for the systems genetic era. ILAR J. 2011;52(1):24-31. doi: 10.1093/ilar.52.1.24. PMID: 21411855.

Vyskočilová, Martina, Gabriela Pražanová, and Jaroslav Piálek. 2009. "Polymorphism in hybrid male sterility in wild-derived *Mus musculus musculus* strains on proximal chromosome 17." *Mammalian Genome* 20 (2):83-91. doi: 10.1007/s00335-008-9164-3.

White, Michael A., Brian Steffy, Tim Wiltshire, and Bret A. Payseur. 2011. "Genetic Dissection of a Key Reproductive Barrier Between Nascent Species of House Mice." *Genetics* 189 (1):289-304. doi: 10.1534/genetics.111.129171.

Yonekawa, H., J.J. Sato, H. Suzuki, and K. Moriwaki. 2012. "Origin and genetic status of *Mus musculus molossinus*: a typical example of reticulate evolution in the genus *Mus*." In *Evolution of the House Mouse*, edited by M. Macholán, S.J.E. Baird, P. Munclinger and J. Pialek, 94-113. New York: Cambridge University Press.

Supplementary tables

Sup. Table 1 – Sampling details (in bold samples included in the final dataset)

Sample ID	POP ID	Lineage	Colour	Sex	Country	t- haplotype	Raw Coverage	Source
afg396	AFG	*musculus*	Orange	Male	Afghanistan	Yes	14	(Harr et al. 2016)
afg413	AFG	*musculus*	Orange	Male	Afghenistan	No	21	(Harr et al. 2016)
afg416	AFG	*musculus*	Orange	Male	Afghanistan	Yes	16	(Harr et al. 2016)
afg424	AFG	*musculus*	Orange	Male	Afghanistan	No	17	(Harr et al. 2016)
afg435	AFG	*musculus*	Orange	Female	Afghanistan	No	19	(Harr et al. 2016)
afg444	AFG	*musculus*	Orange	Male	Afghanistan	yes	18	(Harr et al. 2016)
cei18774	CEI	uncertain	Gold	Male	Iran	No	19	This study
cei18775	CEI	uncertain	Gold	Male	Iran	No	16	This study
cei18784	CEI	uncertain	Gold	Male	Iran	No	17	This study
cei18795	CEI	uncertain	Gold	Male	Iran	No	16	This study
cei18798	CEI	uncertain	Gold	Female	Iran	No	15	This study
cei18799	CEI	uncertain	Gold	Male	Iran	No	17	This study
cei18806	CEI	uncertain	Gold	Male	Iran	Yes	16	This study
fr14	FR	*domesticus*	Blue	Male	France	No	24	(Harr et al. 2016)
fr15B	FR	*domesticus*	Blue	Male	France	No	23	(Harr et al. 2016)
Fr16B	FR	*domesticus*	Blue	Male	France	No	24	(Harr et al. 2016)
Fr18B	FR	*domesticus*	Blue	Male	France	No	24	(Harr et al. 2016)
frB2C	FR	*domesticus*	Blue	Male	France	Yes	14	(Harr et al. 2016)
frC1	FR	*domesticus*	Blue	Male	France	Yes	20	(Harr et al. 2016)
frE1	FR	*domesticus*	Blue	Male	France	No	22	(Harr et al. 2016)
frF1B	FR	*domesticus*	Blue	Male	France	Yes	23	(Harr et al. 2016)
KazAL1	KAZ	*musculus*	Orange	Female	Kazakhstan	No	23	(Harr et al. 2016)

KazAL16	KAZ	*musculus*	Orange	Male	Kazakhstan	No	25	(Harr et al. 2016)
KazAL19	KAZ	*musculus*	Orange	Female	Kazakhstan	Yes	24	(Harr et al. 2016)
KazAL33	KAZ	*musculus*	Orange	Female	Kazakhstan	No	25	(Harr et al. 2016)
KazAL38	KAZ	*musculus*	Orange	Male	Kazakhstan	No	25	(Harr et al. 2016)
KazAL40	KAZ	*musculus*	Orange	Female	Kazakhstan	No	26	(Harr et al. 2016)
KazAL41	KAZ	*musculus*	Orange	Male	Kazakhstan	Yes	26	(Harr et al. 2016)
KazAL42	KAZ	*musculus*	Orange	Female	Kazakhstan	No	25	(Harr et al. 2016)
swi40	SWI	*domesticus*	Blue	Female	Iran	No	17	This study
swi84	SWI	*domesticus*	Blue	Female	Iran	Yes	16	This study
swi86	SWI	*domesticus*	Blue	Male	Iran	No	17	This study
swiAH15	SWI	*domesticus*	Blue	Male	Iran	No	22	(Harr et al. 2016)
swiAH23	SWI	*domesticus*	Blue	Male	Iran	No	24	(Harr et al. 2016)
swiJR11	SWI	*domesticus*	Blue	Male	Iran	No	23	(Harr et al. 2016)
swiJR15	SWI	*domesticus*	Blue	Male	Iran	No	22	(Harr et al. 2016)
swiJR2	SWI	*domesticus*	Blue	Male	Iran	No	23	(Harr et al. 2016)
swiJR5	SWI	*domesticus*	Blue	Male	Iran	No	17	(Harr et al. 2016)
swiJR7	SWI	*domesticus*	Blue	Male	Iran	No	18	(Harr et al. 2016)
swiJR8	SWI	*domesticus*	Blue	Male	Iran	No	17	(Harr et al. 2016)
niH12	NI	*castaneus*	Green	Male	India	Yes	20	(Harr et al. 2016)
niH14	NI	*castaneus*	Green	Female	India	Yes	17	(Harr et al. 2016)
niH15	NI	*castaneus*	Green	Female	India	No	13	(Harr et al. 2016)
niH24	NI	*castaneus*	Green	Female	India	No	13	(Harr et al. 2016)
niH26	NI	*castaneus*	Green	Female	India	No	17	(Harr et al. 2016)
niH27	NI	*castaneus*	Green	Female	India	Yes	13	(Harr et al. 2016)
niH28	NI	*castaneus*	Green	Male	India	No	15	(Harr et al. 2016)
niH30	NI	*castaneus*	Green	Female	India	No	21	(Harr et al. 2016)
niH34	NI	*castaneus*	Green	Male	India	No	21	(Harr et al. 2016)
niH36	NI	*castaneus*	Green	Female	India	No	19	(Harr et al. 2016)

pak10338	PAK	*castaneus*	Green	Female	Pakisthan	Yes	17	This study
pak10342	PAK	*castaneus*	Green	Male	Pakisthan	Yes	14	This study
pak10348	PAK	*castaneus*	Green	Male	Pakisthan	Yes	19	This study
pak10354	PAK	*castaneus*	Green	Female	Pakisthan	Yes	17	This study
pak10358	PAK	*castaneus*	Green	Male	Pakisthan	Yes	19	This study
pak10363	PAK	*castaneus*	Green	Male	Pakisthan	Yes	18	This study
cyp	CYP	cypriacus	Purple		Cyprus	Not tested		This study
cyp	CYP	cypriacus	Purple		Cyprus	Not tested		This study

Sup. Table 2 – Relatedness analysis

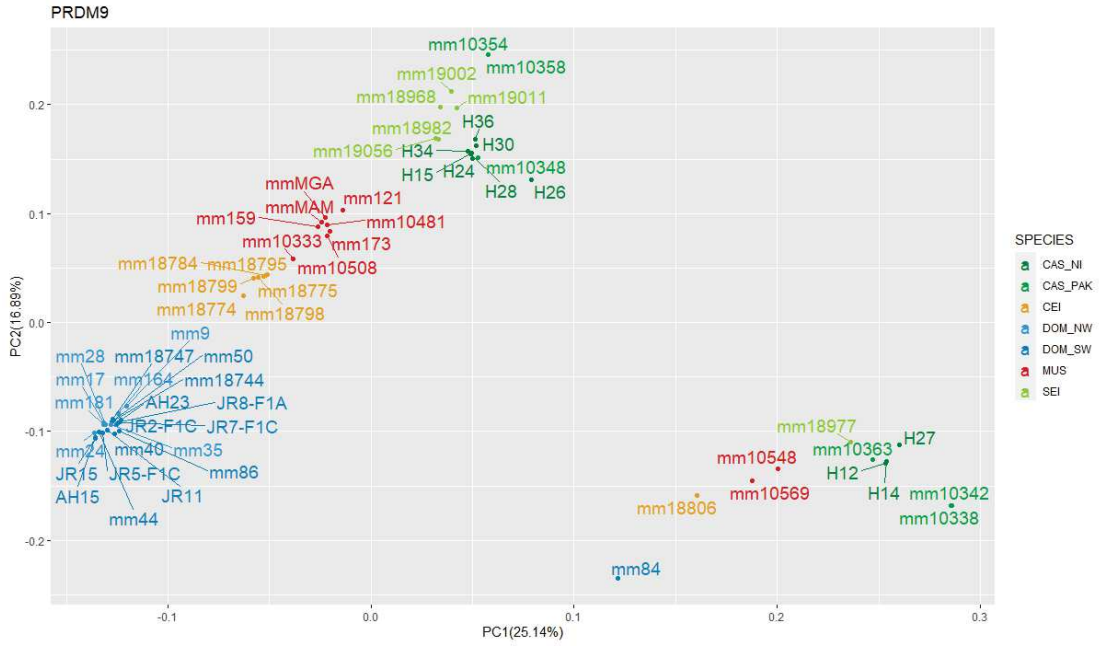
INDV1	INDV2	N_AaAa	N_AAaa	N1_Aa	N2_Aa	RELATEDNESS_PHI
kazAL19	kazAL16	1860	110	2891	2925	0.281981
kazAL16	kazAL19	1860	110	2925	2891	0.281981
pak10342	pak10338	7659	583	11994	11609	0.275092
pak10338	pak10342	7659	583	11609	11994	0.275092
pak10358	pak10354	7437	511	12026	12051	0.266437
pak10354	pak10358	7437	511	12051	12026	0.266437
swiJR5	swiAH15	2962	37	5528	5633	0.258758
swiAH15	swiJR5	2962	37	5633	5528	0.258758
kazAL42	kazAL33	1773	119	3339	2781	0.250817
kazAL33	kazAL42	1773	119	2781	3339	0.250817
cei18799	cei18798	2956	51	5876	5966	0.241007
cei18798	cei18799	2956	51	5966	5876	0.241007
spr68	spr69	280	11	590	564	0.22357
spr69	spr68	280	11	564	590	0.22357
swiJR7	swiJR11	2443	294	5245	5445	0.173527
swiJR11	swiJR7	2443	294	5445	5245	0.173527
kazAL42	kazAL40	1397	230	3339	2577	0.158384
kazAL40	kazAL42	1397	230	2577	3339	0.158384
swiJR15	swiAH15	2269	439	5569	5633	0.124174
swiAH15	swiJR15	2269	439	5633	5569	0.124174
swiJR15	swiJR11	2206	422	5569	5445	0.123661
swiJR11	swiJR15	2206	422	5445	5569	0.123661
kazAL42	kazAL41	1204	270	3339	2226	0.119317
kazAL41	kazAL42	1204	270	2226	3339	0.119317
kazAL42	kazAL38	1198	275	3339	2298	0.114955
kazAL38	kazAL42	1198	275	2298	3339	0.114955
swiJR5	swiJR15	2152	449	5528	5569	0.113004
swiJR15	swiJR5	2152	449	5569	5528	0.113004
fr15B	fr14	1474	387	3658	3495	0.097861
fr14	fr15B	1474	387	3495	3658	0.097861
kazAL38	kazAL33	1153	346	2298	2781	0.0907659
kazAL33	kazAL38	1153	346	2781	2298	0.0907659

Sup. Table 3 - D statistics for all combinations of populations with six individuals

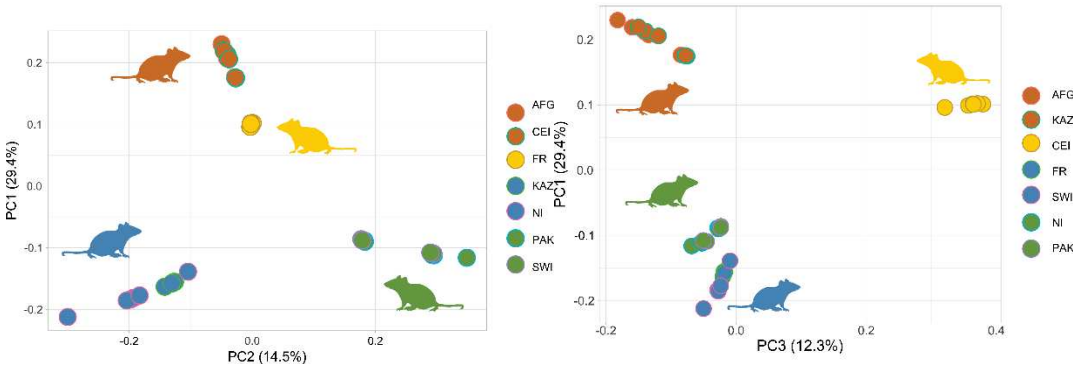
P1	P2	P3	Dstatistic	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
KAZ	CEI	SWI	0.38744	114.6	0.000000	0.291	1684600	1384870	611432
AFG	CEI	SWI	0.38601	121.5	0.000000	0.289	1680590	1377890	610398
KAZ	CEI	FR	0.38072	108.1	0.000000	0.216	1720580	1367180	613210
AFG	CEI	FR	0.37919	117.0	0.000000	0.215	1716540	1360160	612250
FR	CEI	NI	0.12884	63.2	0.000000	0.390	1696380	1117540	862447
SWI	CEI	NI	0.12520	70.7	0.000000	0.379	1726560	1094050	850589
FR	SWI	CEI	0.05737	22.4	0.000000	0.029	2378040	494605	440930
NI	AFG	SWI	0.04776	19.0	0.000000	0.039	1327340	1189870	1081390
NI	KAZ	SWI	0.04487	16.5	0.000000	0.037	1322020	1193750	1091220
FR	SWI	KAZ	0.04327	15.9	0.000000	0.009	3069240	412370	378166
FR	SWI	AFG	0.04323	16.0	0.000000	0.010	3062240	411314	377224
NI	AFG	FR	0.03800	14.8	0.000000	0.024	1346300	1174740	1088730
NI	KAZ	FR	0.03511	12.5	0.000000	0.022	1341000	1178520	1098560
KAZ	AFG	NI	0.03040	14.2	0.000000	0.037	3450400	256603	241464
FR	SWI	NI	0.01549	6.6	0.000000	0.018	3140610	381213	369580
KAZ	AFG	FR	0.01247	3.4	0.000374	0.002	3611170	245810	239754
KAZ	AFG	SWI	0.01217	3.5	0.000207	0.002	3578280	247003	241061
KAZ	CEI	NI	0.00714	4.2	0.000011	0.031	1824630	892391	879732
KAZ	AFG	CEI	0.00496	1.8	0.039988	0.002	2562350	301272	298299
CEI	AFG	NI	0.00140	0.8	0.197742	0.006	1818940	886214	883734

Supplementary figures

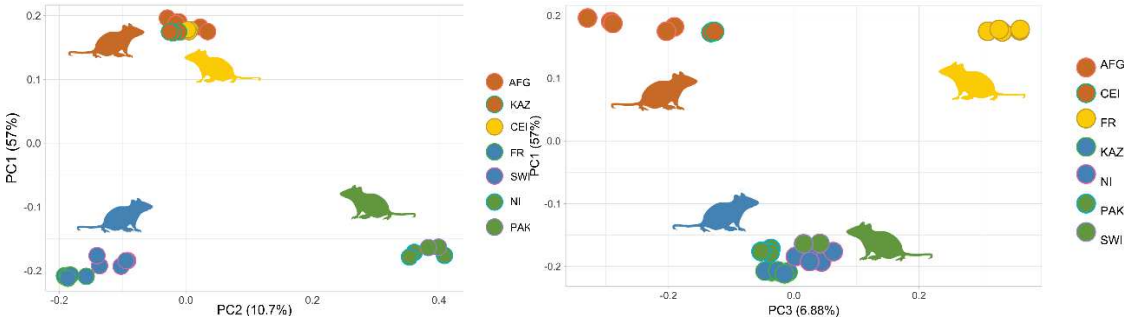
Sup. Fig. 1 – Map with the rough location of populations sampling. Colours represent the separation in populations. Being: Yellow - central Iranian lineage (CEI); Dark blue – southwest domesticus (SWI); red – Afghan musculus (AFG) and Kazak musculus (KAZ), Green – Pakistan castaneus (PAK); Light Green – North Indian castaneus (NI).



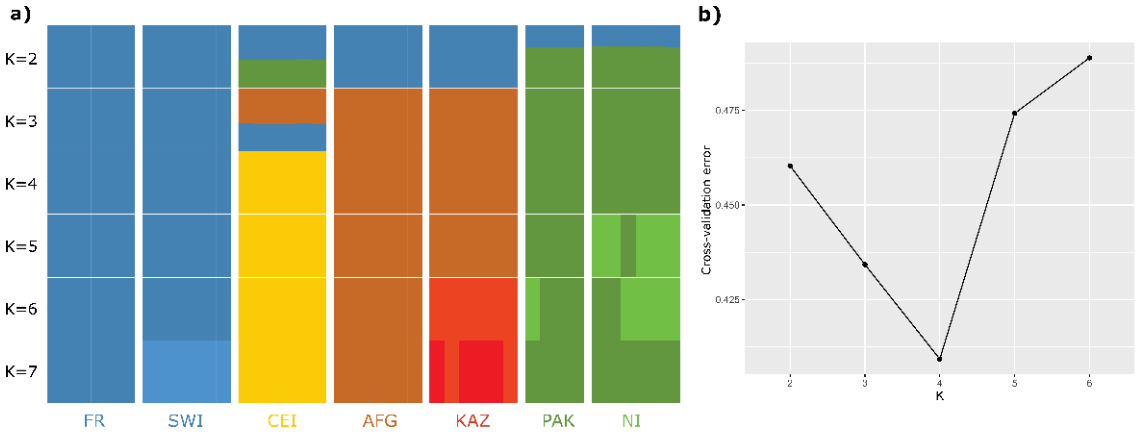
Sup. Fig. 2 - PRDM9 PCA analysis



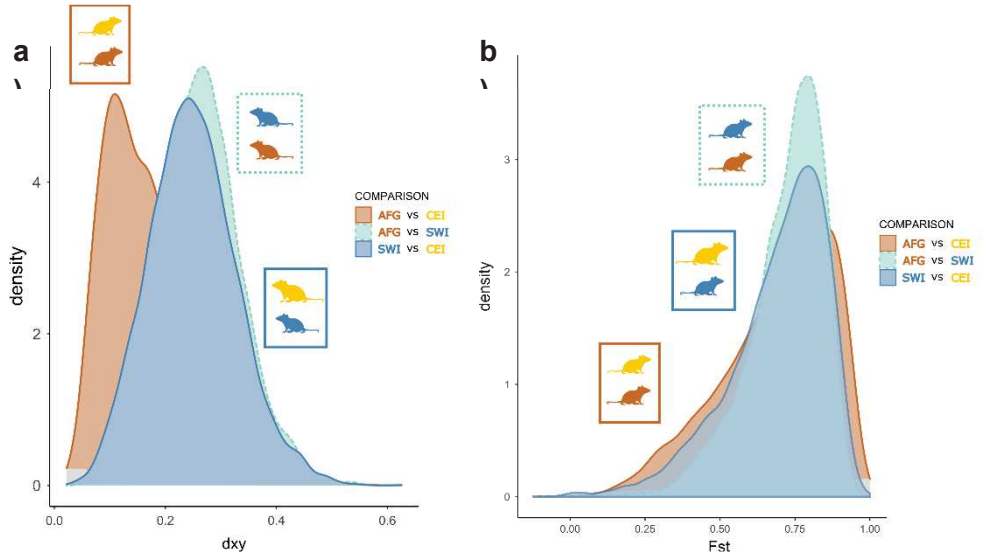
Sup. Fig. 3 – X chromosome PCA's. a) PC1 vs PC2; b) PC1 vs PC3



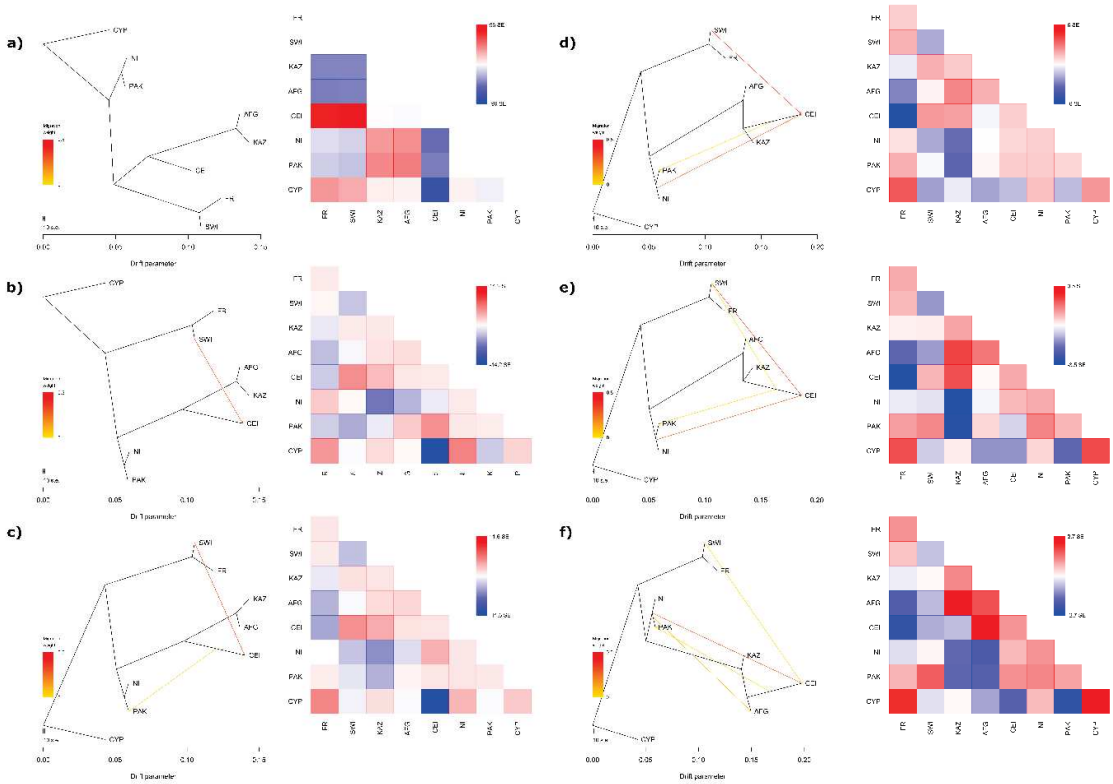
Sup. Fig. 4 - Y chromosome PCA's. a) PC1 vs PC2; b) PC1 vs PC3



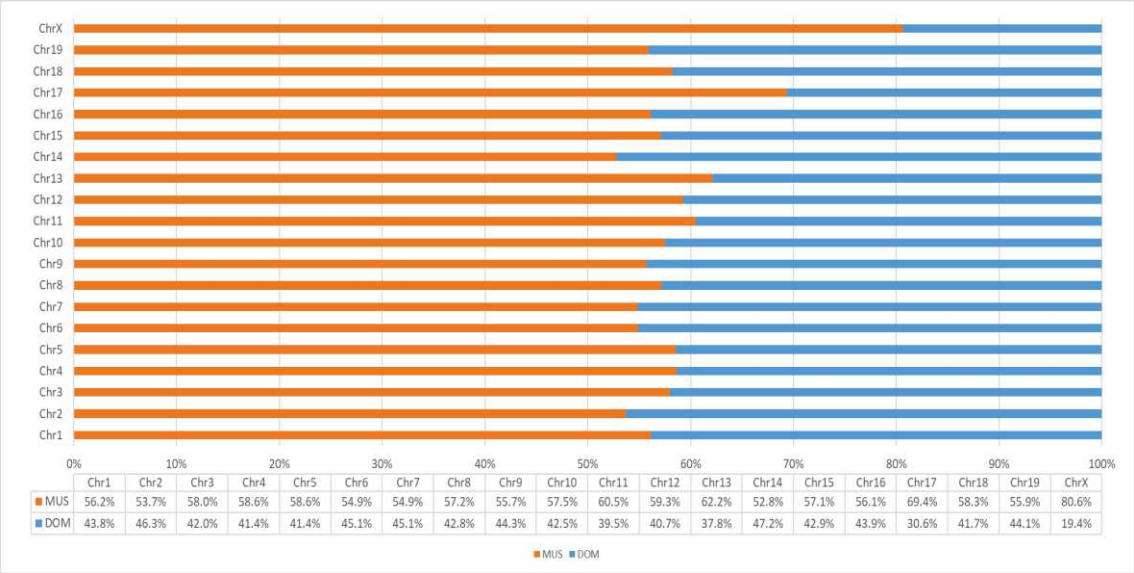
Sup. Fig. 5 – a) ADMIXTURE analysis for K=2 to K=7. b) Cross-validation error



Sup. Fig. 6 - a) Density plot of chromosome X Fst pairwise-estimates; b) Density plot of chromosome X Dxy pairwise-estimates

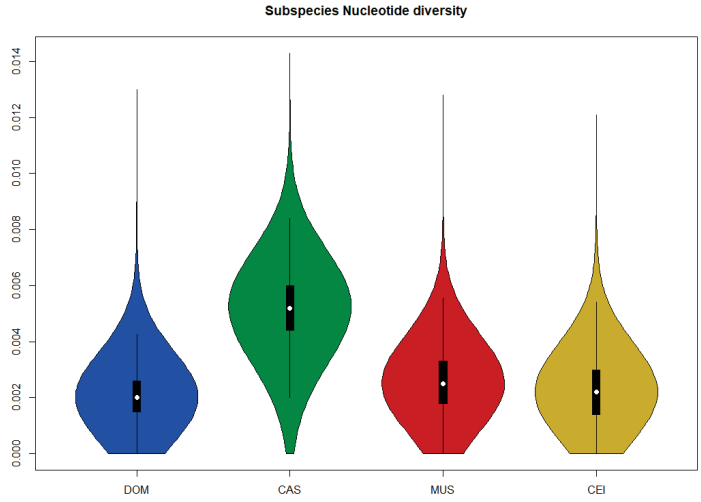


Sup. Fig. 7 - Admixture graph inferred using Treemix and 25kb apart SNPs from all available individuals. a) A simple tree-like model without admixture fits the data poorly, as can be seen from the matrix of residuals between empirical and modelled allele frequency covariance on the right. b) The placement of one admixture event from the common ancestor of M. m. musculus to central Iran population. c) The optimal placement of two admixture event are from the common ancestor of M. m. musculus to central Iran population, as well as from the common ancestor of central Iran population to the M. m. musculus population of Caucasus.

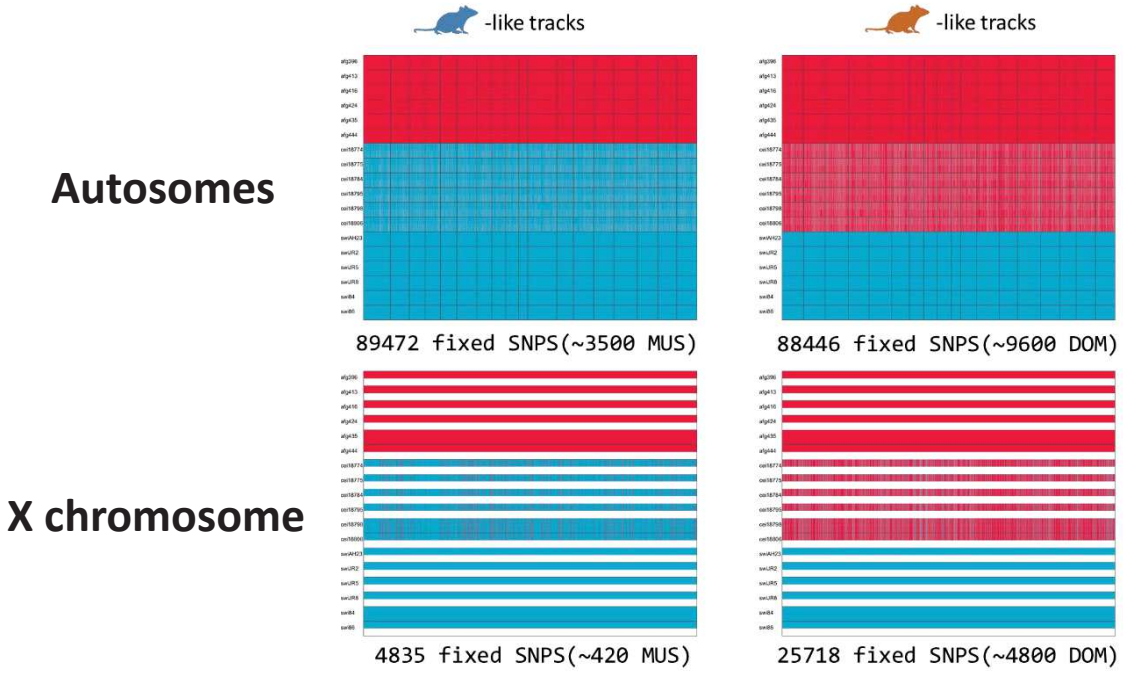


	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10	Chr11	Chr12	Chr13	Chr14	Chr15	Chr16	Chr17	Chr18	Chr19	ChrX
MUS	56.2%	53.7%	58.0%	58.6%	58.6%	54.9%	54.9%	57.2%	55.7%	57.5%	60.5%	59.3%	62.2%	52.8%	57.1%	56.1%	69.4%	58.3%	55.9%	80.6%
DOM	43.8%	46.3%	42.0%	41.4%	41.4%	45.1%	45.1%	42.8%	44.3%	42.5%	39.5%	40.7%	37.8%	47.2%	42.9%	43.9%	30.6%	41.7%	44.1%	19.4%

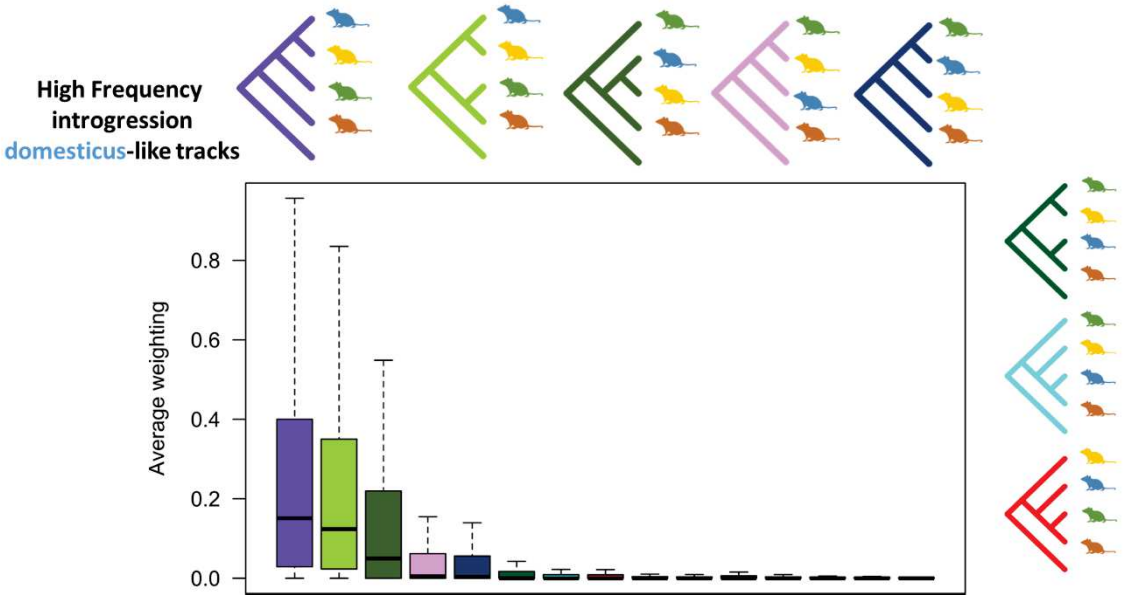
Sup. Fig. 8 - ELAI chromosome-wide ancestry assignment proportions



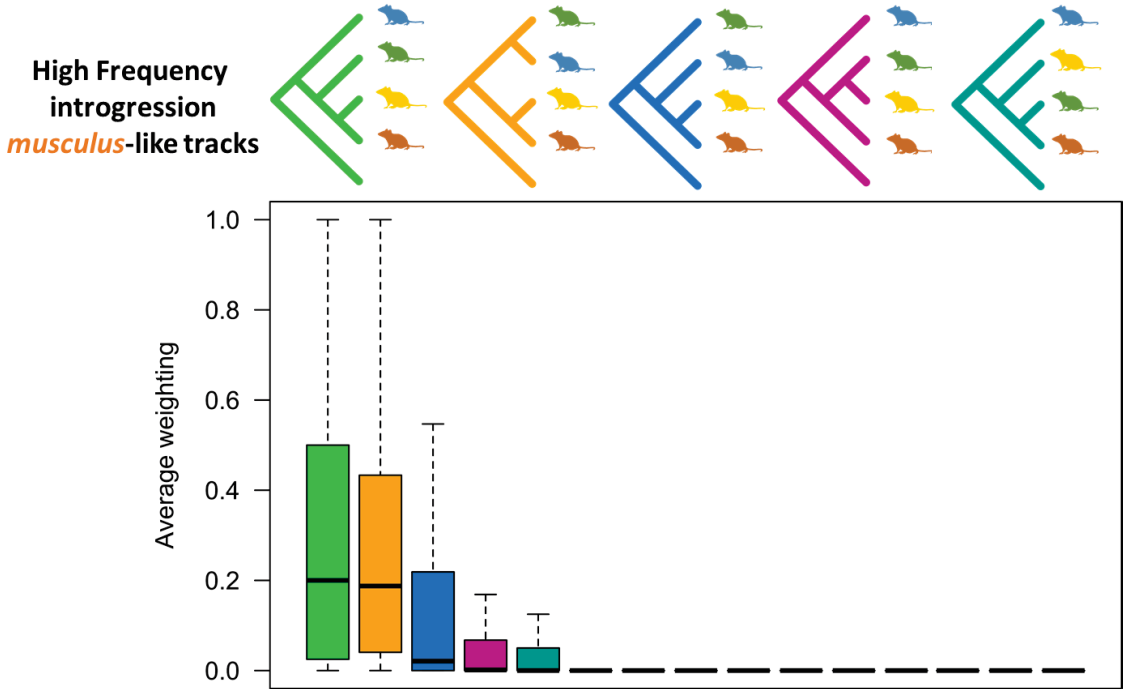
Sup. Fig. 9 - Nucleotide diversity per subspecies



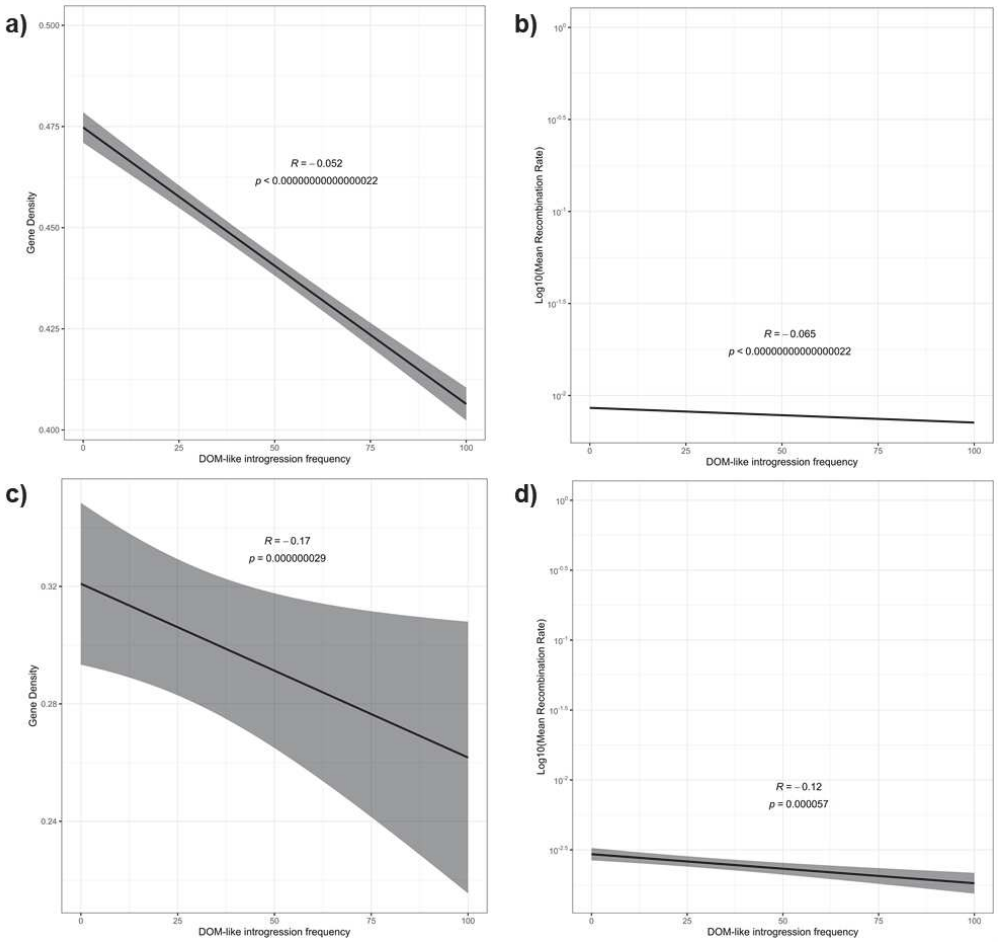
Sup. Fig. 10 – Fixed SNPs segregation analysis on high frequency introgression tracks for the autosomes and X-chromosome



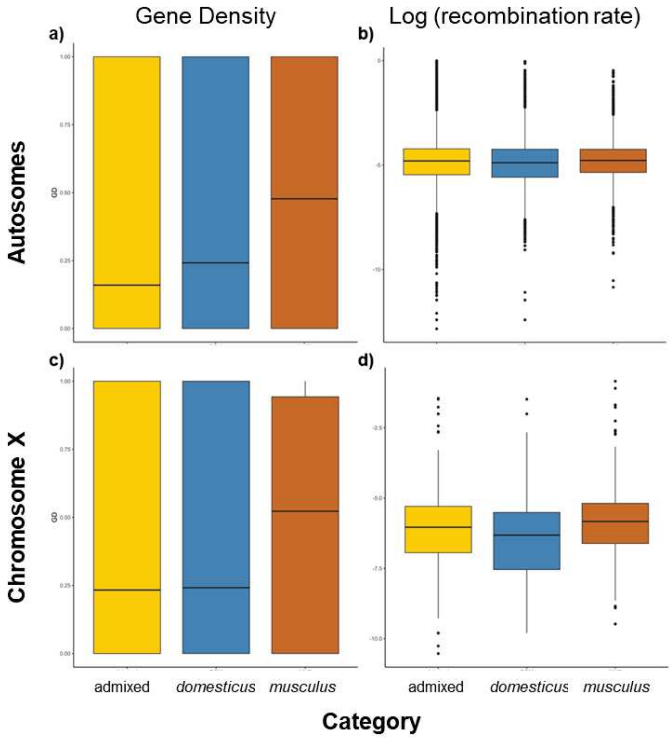
Sup. Fig. 11 - TWISST analysis on the high frequency domesticus-like tracks



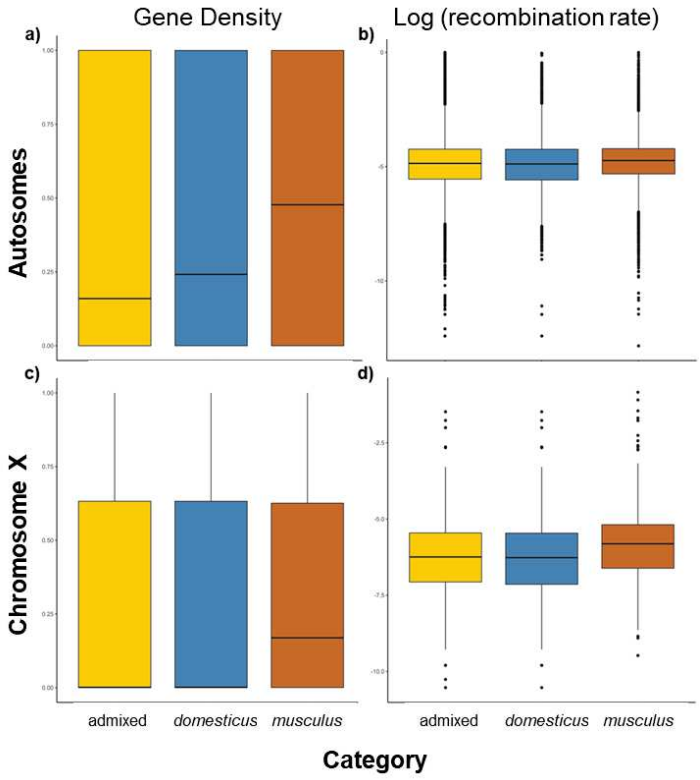
Sup. Fig. 12 – TWISST analysis on the high frequency musculus-like tracks



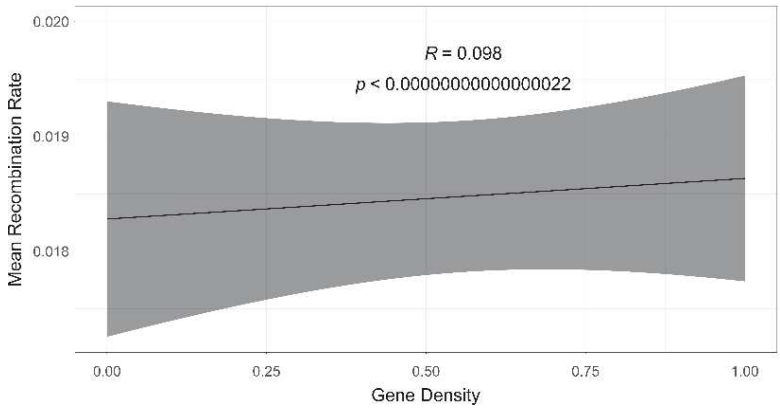
Sup. Fig. 13 -DOM-like introgression correlations with Gene density (a) and c)) and recombination rate (b) and d)). a)/b) Autosomes (except chromosome 17) and b)/c) Chromosome X



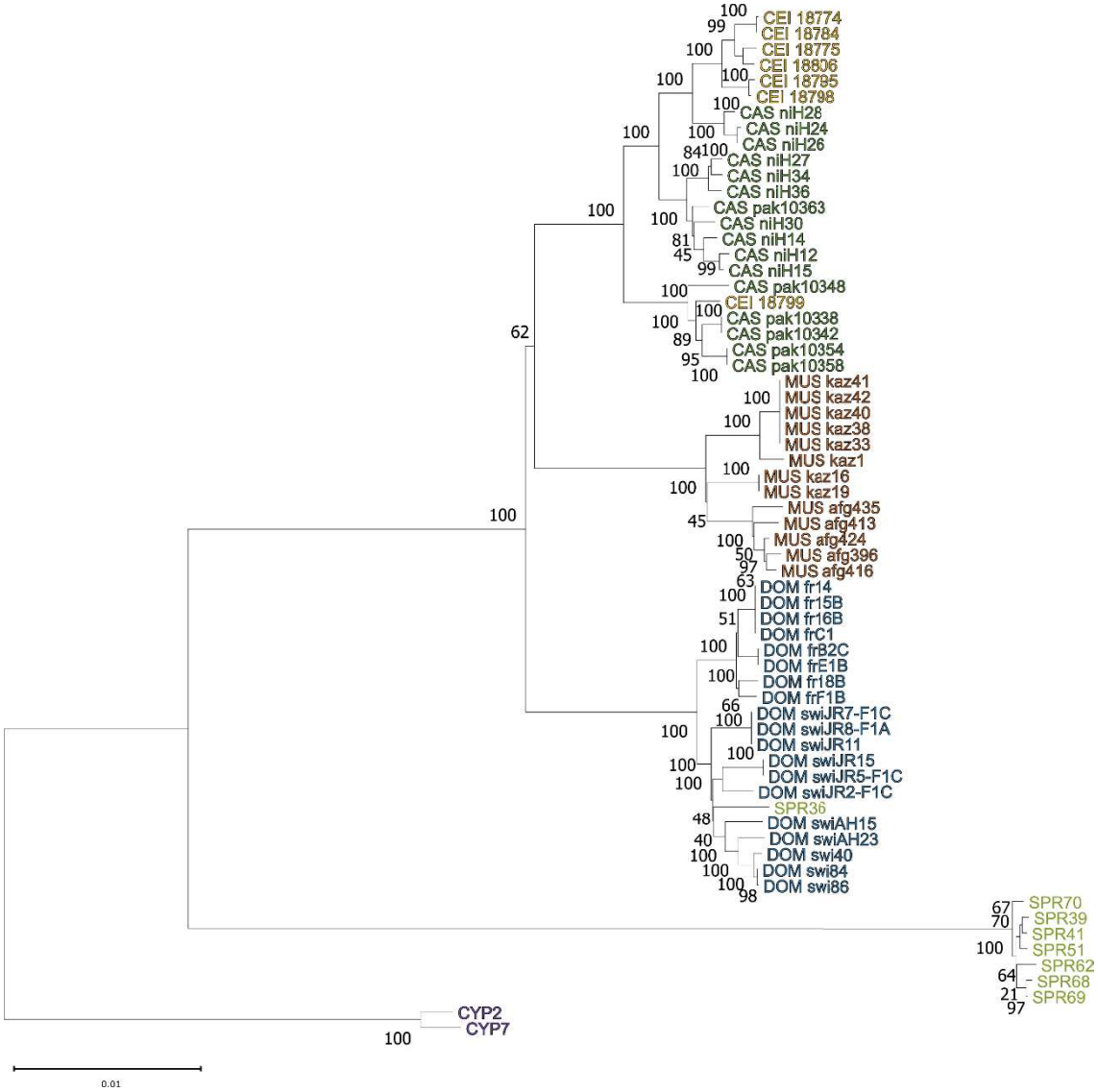
Sup. Fig. 14 – Gene density (left) and log recombination rate (right) boxplots divided by fixed ancestry (MUS or DOM) and admixed (not fixed). Autosomes - Upper panel. Chromosome X - Lower panel. Values based on random sites sampled along the genome.



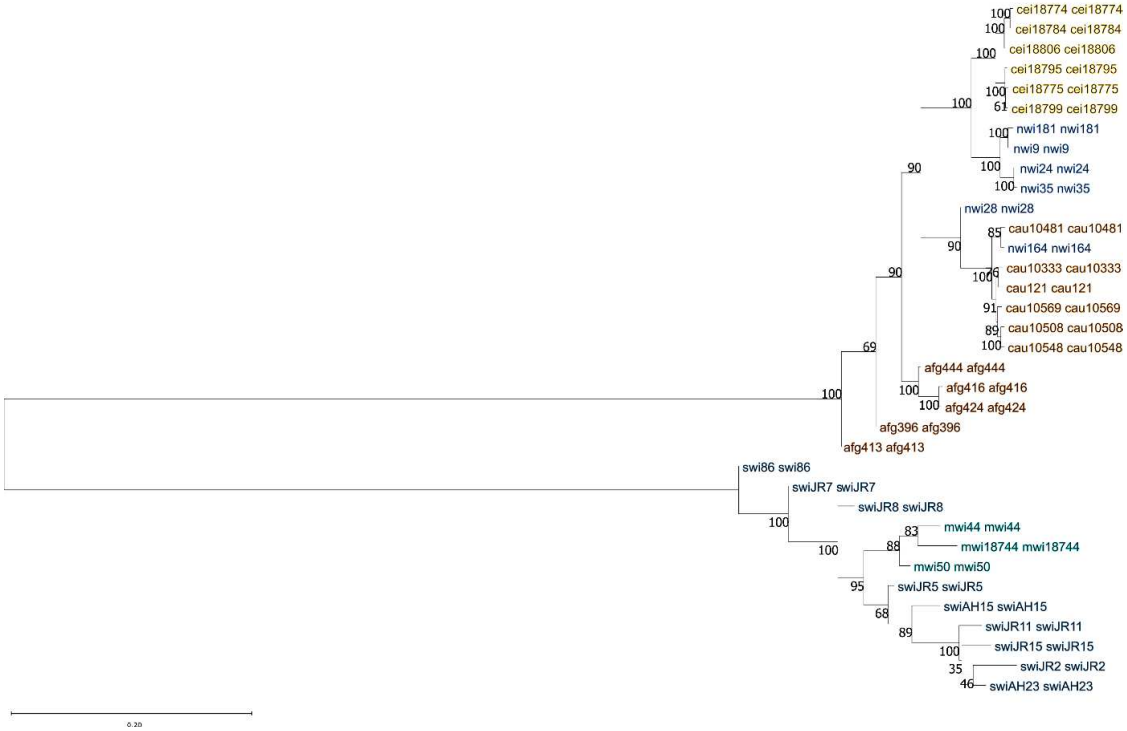
Sup. Fig. 15 - Gene density (left) and log recombination rate (right) boxplots divided by almost fixed (>80%) ancestry (MUS or DOM) and admixed (not fixed). Autosomes - Upper panel. Chromosome X - Lower panel. Values based on random sites sampled along the genome.



Sup. Fig. 16 - Genome-wide mean recombination rate correlation with gene density

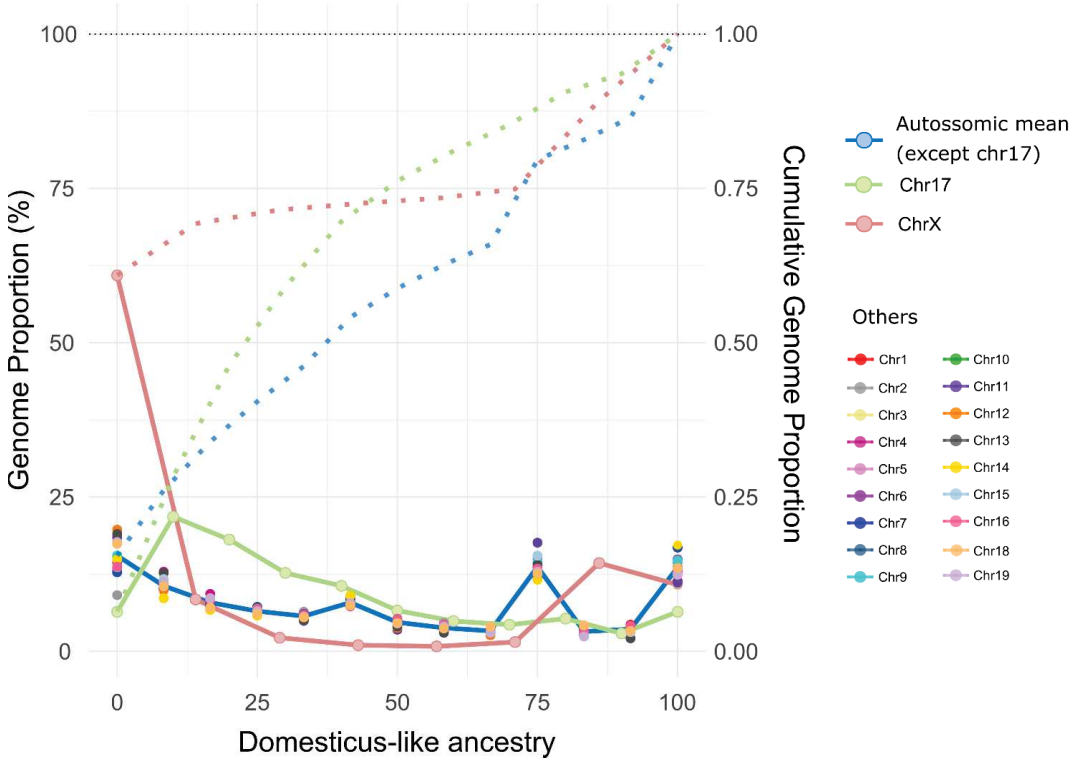


Sup. Fig. 17 – Mitochondrial neighbour joining tree

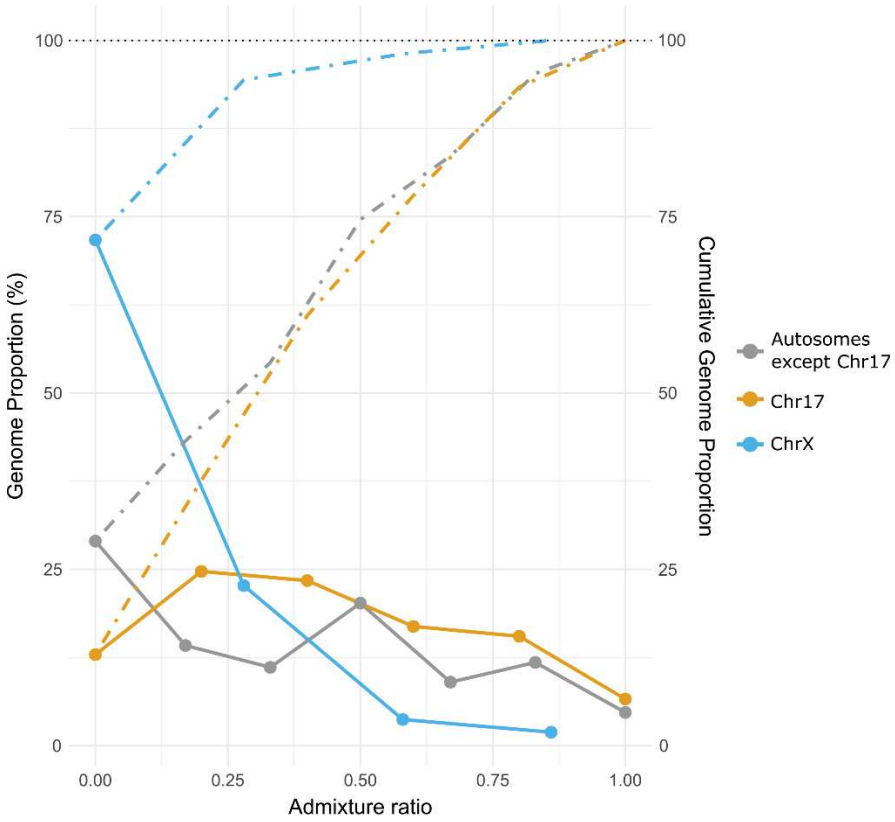


Sup. Fig. 18 - Chromosome Y negihbour joining phylogeny from Marques et al. (in preparation – Thesis Publication V)

Extra figures not included in this preliminary version



Sup. Fig. 19 – Chromosomal proportion assignment according to domesticus-like ancestry. The cumulative chromosomal proportion is indicated on the secondary scale.



Sup. Fig. 20 - Chromosomal proportion assignment according to the ratio of admixture (being 0 fixed for either musculus or domesticus and 1 an equative mixture of musculus and domesticus-like ancestry). The cumulative chromosomal proportion is indicated on the secondary scale.

Supplementary Text

Sampling relatedness

We used the relatedness2 option of vcftools to assess pairwise individual relatedness among all mice in the dataset, using the KING method (Danecek et al. 2011). This analysis is based on GATK called genotypes and the 90% tranche PASS-filtered SNPs. We restricted the dataset to only include autosomal SNPs, thinned to 1 SNP every 25kb. We also removed sites that had more than 20% missing data, expected ranges of kinship coefficients ('Phi') are >0.354 for duplicate samples/monozygotic twins, [0.177–0.354] for 1st degree relatives, [0.0884–0.177] for 2nd degree relative, [0.0442–0.0884] for 3rd degree relatives and <0.0442 for unrelated - Sup. Table 2). No duplicate samples were detected. Most related animals were found in the populations from Iran and Kazakhstan. In the case of the Iranian population the increased relatedness within the sample can be explained by the fact that some breeding adults were used in multiple cross. The relatedness observed in the population from Kazakhstan is best explained by the fact that mice were collected in close proximity, rather than over a larger regional scale (Harr et al. 2016). We only consider first and second degree relatedness relevant here and related individuals were removed from the final analysed dataset.

t- haplotype individuals' identification

The *t*-haplotype is a complex set of 4 inversions, comprising a 30–40 Mbp long region of chromosome 17. It causes transmission ratio distortion, and heterozygous t-haplotype carriers tend to predominantly transmit the *t*-haplotype carrying chromosome to their offspring. Despite their massive transmission advantage, *t*-haplotype carrying individuals are rare in natural populations of mice but have been found in all recognized subspecies. We have identified our *t*-carriers individuals by leverage on a PRDM9 gene region PCA analysis. The PCA segregates *t*-carriers from non-carriers rather than by subspecies. These results agreed with what was previously described for the northern Indian individuals (confirmed by PCR on *Bettina et al.* (2016). We have identified approximately a *t*-carrier individual per population, including one individual in central Iran.

Mitochondrial DNA phylogeny

Whole mitochondrial genomes (except some *D-loop* sequences) were recovered for all individuals with NOVOPlasty v.2.7.2 (Dierckxsens, Mardulyn, and Smits 2016). The mitochondrial genome phylogeny was reconstructed using a neighbour joining tree method implemented on MEGAX (Kumar et al. 2018) with 100 iterations.

Y chromosome evaluation

Due to Y chromosome lack of recombination, we estimated a phylogeny using a neighbour joining tree method implemented on MEGAX (Kumar et al. 2018). We used a concatenated alignment of single copy Y genes and run 100 iterations to construct a phylogenetic tree (Sup. Fig. 22).

Classification: Biological Sciences - Evolution

A successful Y chromosome lineage in regions of admixture between house mouse subspecies.

Author affiliations:
João P. Marques[1,2,3], José Melo-Ferreira[1,2*], Pierre Boursot[3*]

[1]CIBIO, Research Center in Biodiversity and Genetic Resources, InBIO Associate Laboratory, University of Porto, 4485-661 Vairão, Portugal;
[2]Department of Biology, FCUP, Faculty of Sciences of the University of Porto, 4169-007 Porto, Portugal;
[3]Institut des Sciences de l'Évolution, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France;

Corresponding authors:
José Melo-Ferreira, CIBIO-InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal, tel: +351 252660411, e-mail: jmeloferreira@cibio.up.pt; Pierre Boursot, Institut des Sciences de l'Évolution, Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier, France, e-mail: pierre.boursot@umontpellier.fr.

Keywords: *Mus musculus,* sex chromosomes, hybrid constraints, genomic incompatibilities, Y introgression, copy number variation

Abstract

The interaction between the two house mouse subspecies *M. m. domesticus* and *M. m. musculus* has been intensely studied in their European hybrid zone, which results from a very recent secondary contact. Here we study, using whole genome sequencing, their interactions in Iran, the presumed cradle of their expansion, where contacts are more ancient, including contacts with another newly described genetic entity (CEI), itself resulting from ancient thorough admixture of the *domesticus* and *musculus* branches. We find in Northwestern Iran three-way admixtures in populations of either *musculus* or *domesticus* major ancestry. In the latter case, we find complete and presumably selective introgression of a Y chromosome related to the *musculus* lineage. The genes following this pattern of massive introgression are enriched in male fertility ontology terms. We find a correlation between copy numbers of Y and X ampliconic families (*Sly/Slx*) whose interaction is known to control sex chromosome transmission in a dosage dependent manner. We argue that the invasive success of the *musculus* Y is not due to its transmission distortion power linked to high *Sly* copy number, but rather to its contribution to rescue infertility in hybrids, thus explaining its prevalence in all known admixed house mouse populations.

Introduction

The European house mouse subspecies (*Mus musculus domesticus* and *M. m. musculus*) are a famous model of incipient speciation, because after they recently expanded to Europe (in association with human settlement and trading), they met and formed a narrow tension zone across which gene flow is limited by the existence of numerous genetic incompatibilities (Raufaste et al. 2005, Macholan et al. 2007), as well as by the existence of prezygotic isolation. Prezygotic isolation is thought to have arisen after the secondary contact, by a reinforcement mechanism (Bimova et al. 2011, Smadja and Ganem 2005). However, the nature of the intrinsic incompatibilities and the circumstances under which they accumulated remain poorly understood. Inferences of the divergence history of these subspecies from their present genetic differentiation has suggested that their secondary contact and genetic exchanges started well before the formation of the European hybrid zone, and must have thus occurred in the cradle of the European expansion, presumably in Iran (Duvaux et al. 2011). It is therefore interesting to compare the patterns of admixture in these regions of ancient and presumably repeated contact with those following the more recent and extensive European geographic expansion. We previously reported the occurrence in Central Iran of a distinct genetic entity that appears to result from extensive past admixture between populations that were closely related to *domesticus* and *musculus*. Here we extend the exploration of genetic variation in Iran to regions of potential contact and admixture between *domesticus*, *musculus* and this newly described Central Iranian population. We document three-way admixture in two populations from North-Western Iran, and massive Y chromosome introgression in one of them.

Results

Population structure assessment and mitochondrial DNA phylogeny

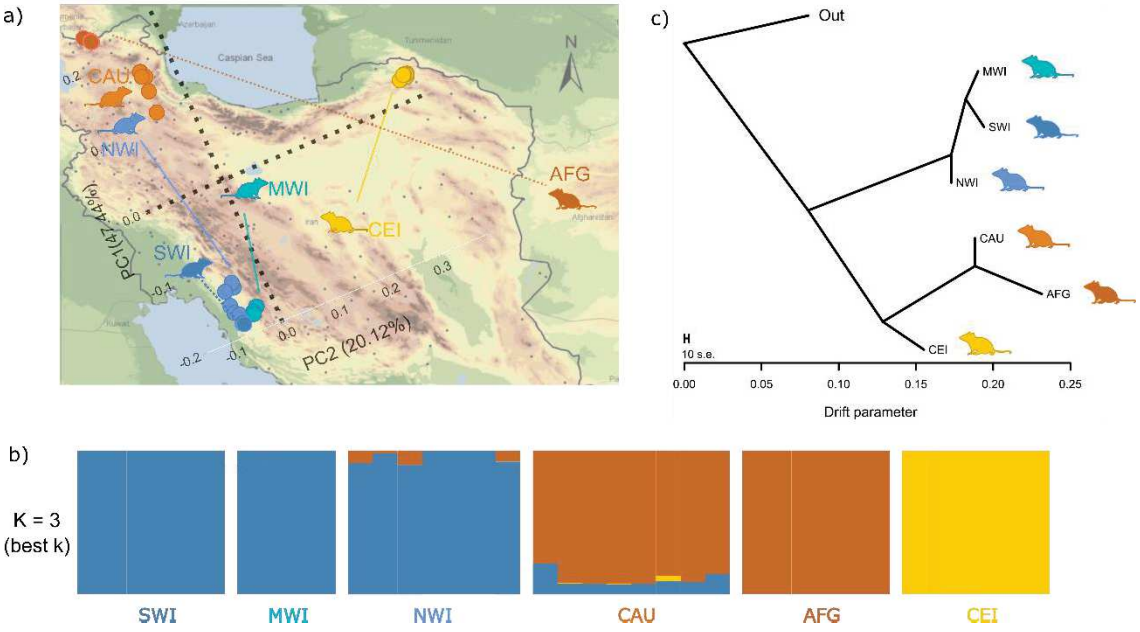


Figure 1 – Genetic strutuctue of populations on the Iranian plateau. (a) Principal compontent analysis. The percentage of variance explained by each component accompanies the titles of the axes. (b) ADMIXTURE plot for the best supported K = 3. (c) Treemix population tree.

We generated whole genome sequence data for 17 specimens from Northwestern Iran and the neighbouring countries in the Caucasus and 4 from the Zagros mountains, at an average 17x coverage per individual. These were put together with published sequence data for a Southwest Iranian population of *domesticus* (herafter referred to as SWI), a *musculus* population from Afghanistan (AFG), a recently described Central Iranian lineage (CEI) and two *Mus cypriacus* specimens used as outgroup (OUT). A rough location of these samples, marked by mouse colourful icons near population codenames is shown on Figure 1a). Note that on this Figure our new sample is divided into three, with three codenames. This is a post-hoc subdivision based on a combination of geography and overall genomic composition (see below). These names will be used throughout: NWI (for Northwestern Iran), MWI (for Medium Western Iran) and CAU (for Caucasus). A more precise location of each sample is represented on Sup. Fig. 1, and a full description of the samples is provided in in Sup. Table 1. After mapping the reads to the reference genome we retained ~2.62M high-quality sites. The final main dataset was composed by 39 high coverage (14-24x) individuals, 19 from our newly sampled areas (further details Sup. Table 1), six individuals for each of the previously characterised populations, representing the three major lineages present in Iran (SWI for

domesticus, AFG for musculus, and CEI for the newly described lineage from Central Iran), and two *Mus cypriacus* used as outgroup (Sup. Fig. 1). *M. cypriacus* is endemic to the island of Cyprus and related to *Mus macedonicus*. Although most of the analyses were performed using this dataset, we also included additional data to further support some results. This is mentioned along the analyses and is fully described in Sup. Tables 2-4.

We used a subset of ~254k linkage disequilibrium (*LD*) pruned biallelic single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) > 0.02 to infer population structure. On a Principal component analysis (PCA), the first component (PC1) explained 47.4% of the variation among house mouse specimens, and divided the individuals into three well separated clusters (Figure1a), representing the two subspecies *domesticus* and *musculus*, as well as the Central Iranian population (CEI), previously described as resulting from an admixture between *domesticus* and another branch related to musculus. The second axis (PC2) separated the central Iranian lineage from *musculus* and *domesticus* and represented ~20% of the variation found. As mentioned above, our new samples can be divided into three groups to which we give names: CAU samples cluster in the PCA with *musculus* (AFG sample), NWI and MWI samples with *domesticus*. According to their positions in the PCA, these samples however appear to have multiple ancestries: CAU appears to have a substantial contribution from *domesticus* and a smaller one from CEI, NWI some contribution from *musculus*, and MWI some contribution from CEI.

The subdivision suggested by the PCA is confirmed by the whole-genome ADMIXTURE analysis which strongly supports K = 3 as the optimal number of clusters (Sup. Fig. 2). This analysis also confirms the admixtures suggested by the PCA, except for MWI which is not inferred as admixed.

We further took advantage of the information carried by allele frequencies to investigate the population relationship and estimate a population tree using the graph-based model implemented in TreeMix (Pickrell et al. 2012; Pickrell and Pritchard 2012). The phylogram (Figure 1c) shows a clear division between *domesticus* and the other populations. Note the populations inferred as admixed between *domesticus* and *musculus* (NWI, CAU and CEI) all have short branches, and two of them (CEI and NWI) branch closer to the root than the others. Both patterns are expected (models with more migration events presented in Sup. Fig. 3).

In order to further test for admixture, we performed an *f3* statistics test (Reich et al. 2009; Patterson et al. 2012) applied to all combinations of analysed populations – Table 7 and extended results in Sup. Table 5. The results suggest that the NWI

population is likely the result of admixture between *domesticus*, *musculus* and the CEI lineage.

The mitochondrial genome phylogeny is globally in line with what is expected from the above results, but reveals some level of mito-nuclear discordance (Sup. Fig. 4). Although most CAU samples lie on the *musculus* branch, as expected, two (cau10508 and cau10333) have a *domesticus* mitochondrial type and one (cau121) the CEI type. Two of the 6 NWI samples (nwi164 and nwi181) lie on the CEI branch rather than the expected *domesticus* one. Mitochondrial admixture in this region was further confirmed by a compilation of the data in the literature, representing 422 mitochondrial *D-loop* sequences (Sup. Fig. 5 and Sup. Text).

Table 1 – Admixture f3-statistics of the form f3 (NWI; SWI, "Admixture Source"). These statistics represent the northwestern Iranian population as a mix of two populations with a more negative result signifying the more likely admixture event.

f3(A; B, C)	Target	Source1	Source2	f3	std. error	Z-Score
NWI; SWI,CAU	NWI	SWI	CAU	-0.00593	0.000271	-21.8531
NWI; SWI,AFG	NWI	SWI	AFG	-0.00583	0.000309	-18.8921
NWI; SWI,CEI	NWI	SWI	CEI	-0.00077	0.000212	-3.61722

Sex chromosome characterization

The X-chromosome PCA arrangement generally follows the autosomal broad pattern, with PC1-PC2 separating *domesticus* from *musculus*-Iranian lineage, and PC3 *musculus* and the Iranian lineage. The Y phylogeny showed a clear separation between *domesticus* SWI and MWI populations from all the other samples. In the non-*domesticus* clade, one subclade contains all CEI samples, and another all CAU samples, so that AFG appears paraphyletic as compared to these two (Figure 2). A striking result is that the NWI samples all lie in the non-*domesticus* clade. Among the 6 NWI samples, 4 (nwi181, nwi9, nwi24 and nwi35) are related to the CEI subclade, although distinct from it, and 2 (nwi164 and nwi28) are included in the CAU subclade. Given the genomic composition of NWI, this therefore witnesses introgression of alien non-domesticus Y chromosomes into NWI. This apparent complete Y replacement was confirmed on a larger sample from this region by a PCR assay of an indel in an intron of the Zfy2 gene, that is diagnostic of the *domesticus* Y lineage (Boissinot and Boursot 1997) (Sup. Fig. 6).

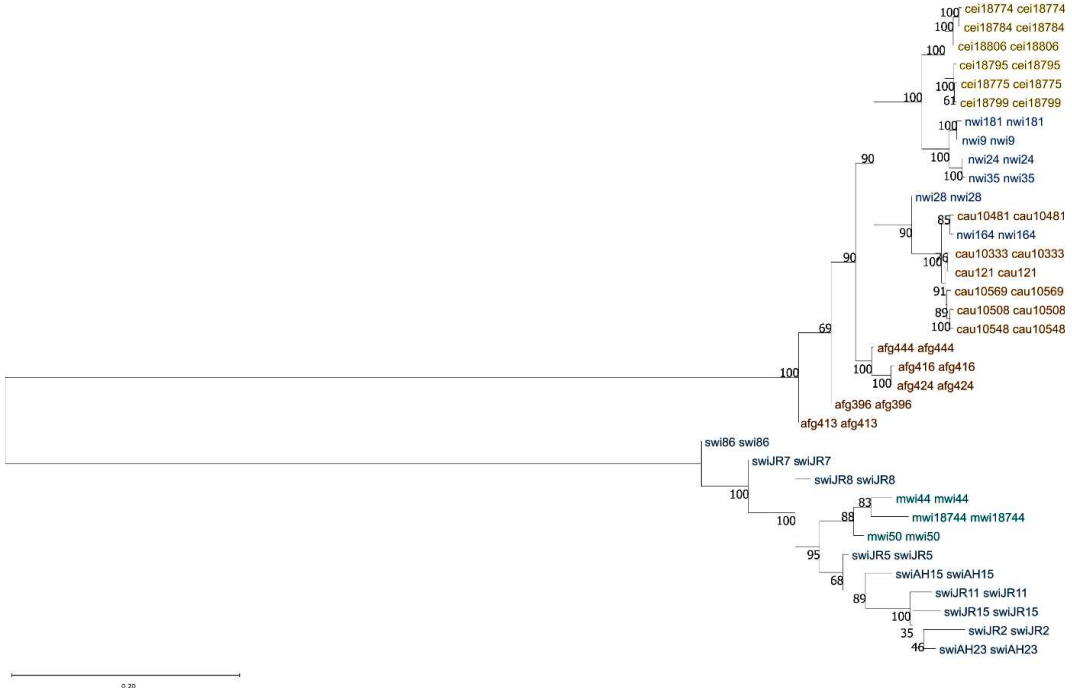


Figure 2 – Chromosome Y negihbour joining phylogeny

Ancestry and introgression inference

We then sought for other genomic regions displaying patterns of extensive introgression into the NWI population, paralleling the Y pattern. We applied a Hidden Markov Model (HMM) linkage disequilibrium-based method - ELAI (Guan 2014), to partition the NWI genomes according to their ancestries, using AFG, SWI and CEI as putative parental populations. ELAI was able to assign a possible ancestry to most of the autosomes (less than 5% unassigned autosomal tracks), with higher uncertainty for chromosomes 1 (13.7%), 2 (6.6%), 13 (12.7%) and 17 (7.8%). As suggested by the previous results the ELAI analysis revealed a predominantly *domesticus*-like overall autosomal ancestry (82% of the entire length). On average 13% was assigned to another ancestry and 5% remained unassigned. Among autosomes, domesticus ancestry varied from a minimum of 78.8% for chromosome 19 to a maximum of 84.1% for chromosome 12. The contributions of the two other parents was balanced between AFG (chromosomal average 7%, minimum 3.3% on chromosome 1 and maximum 9.4% on chromosome 19) and CEI (chromosomal average 7%, minimum 3.7% on chromosome 1 and maximum 8.3% on chromosome 16). A higher proportion of *domesticus* ancestry was inferred for the X chromosome (91.5%), with more contribution of CEI (5.5%) than AFG (1.6%; 1.4% of the tracts were not assigned) - Sup. Fig. 7. We found a correlation between ancestry and geography, with a north-south decrease of *musculus*-like

introgression and an East-West decrease of CEI contribution (Sup. Fig. 8). These decreases correspond to an increase in the distance from the potential source of the introgression and show that frequencies of introgression are not at equilibrium across the region studied. We performed a gene enrichment analysis for all genomic regions (autosomal or X-chromosome) with at least half of the haplotypes assigned to non-*domesticus* origin (*musculus* or CEI-like). Among terms associated with human pathologies, the analysis revealed an enrichment in terms associated with fertility (particularly male fertility) – Table 2. The X was the chromosome with the highest number of genes showing high introgression frequencies (>50%). Among these genes, 46% had exclusive contribution from CEI, only 18% exclusively *musculus* contribution, while 37% showed both parental contributions – Sup. Fig. 9-12.

Table 2 - Gprofiler Gene Ontology analysis of genes showing non-domesticus introgression frequencies of at least 50% (Full results in https://biit.cs.ut.ee/gplink/l/g-fjKvmWRy).

Term name	Term ID	Adjusted *p*-value	*p*-value -log10	Term size	Intersection size
Obstructive azoospermia	HP:0011962	4.21E-11	10.375	45	22
Recurrent spontaneous abortion	HP:0200067	1.96E-10	9.7088	39	20
Synovial sarcoma	HP:0012570	1.58E-09	8.8001	14	12
Spontaneous abortion	HP:0005268	1.58E-08	7.8018	57	22
Increased circulating gonadotropin level	HP:0000837	4.91E-08	7.3090	95	28
Non-obstructive azoospermia	HP:0011961	1.59E-07	6.7994	63	22
Abnormal circulating gonadotropin level	HP:0030338	2.23E-07	6.6507	114	30
Azoospermia	HP:0000027	1.05E-06	5.9760	128	31
Abnormal spermatogenesis	HP:0008669	1.40E-06	5.8528	144	33
Hyperpituitarism	HP:0010514	5.25E-05	4.2792	126	28
Abnormal delivery	HP:0001787	5.57E-05	4.2536	97	24
Autosomal dominant inheritance	HP:0000006	6.24E-05	4.2045	1818	187
Phenotypic abnormality	HP:0000118	0.00014	3.8537	4520	401
Decreased testicular size	HP:0008734	0.00024	3.6317	167	32
Mode of inheritance	HP:0000005	0.00024	3.6270	4257	379
Abnormality of the testis size	HP:0045058	0.00024	3.6237	201	36
Aplasia/Hypoplasia of the testes	HP:0010468	0.00041	3.3866	188	34
Functional abnormality of male internal genitalia	HP:0000025	0.00115	2.9402	232	38
Acetabular spurs	HP:0010454	0.00256	2.5853	5	5
Abnormality of the anterior pituitary	HP:0011747	0.00597	2.2240	296	43
Abnormal male reproductive system physiology	HP:0012874	0.00813	2.0898	280	41
Abnormality of the pituitary gland	HP:0012503	0.02448	1.6111	323	44
Abnormal circulating hormone level	HP:0003117	0.02725	1.5646	460	57
Abnormality of the hypothalamus-pituitary axis	HP:0000864	0.04667	1.3309	373	48

X chromosome Copy-Number Variations

Because ampliconic regions on the mouse X and Y chromosomes are poorly assembled in the reference, and are known to vary a lot in copy-number, we designed a strategy based on the analysis of k-mer coverage after mapping sequencing reads onto the reference genome. In total we recovered 233 different CNVs with at least a single significant blast hit against X and Y references, to which we mapped all sequencing reads of each specimen. After controlling for specimen sex (conversion to haploid copy number) and re-scaling copy number values according to single copy autosomal genes coverages across the genome, we identified 53 significantly variable CNVs on Y and X chromosome libraries. We were able to retrieve an annotation for 28 of them and we performed a multivariate analysis to identify CNVs that mimic the Y invasion pattern, i.e. that had copy numbers resembling those of the Y donors. Note that for this analysis we considered additional reference populations, from published data (*domesticus* populations from France and Germany, and *musculus* populations from the Czech Republic and Kazakhstan).

We identified a CNV, corresponding to the *Slx* gene, with an interesting pattern. The *domesticus* populations (barring NWI) had the lowest estimated number of copies (2-16 per haploid genome). Interestingly, NWI was estimated to have a higher number of copies (13-25), barely overlapping with the *domesticus* range, but largely overlapping with the higher ranges of CEI (17-28) and *musculus* (15-35). – Figure 3. We note that we found concordant results for males and females, which reinforces the confidence that the repeat libraries we used to estimate copy numbers were X-specific, because significant cross-mapping of Y sequences onto these libraries would have led to higher copy number estimates in males than females, which is not the case. These data do not however allow determining whether the resemblance between NWI and the non-*domesticus* populations results from introgression or convergence.

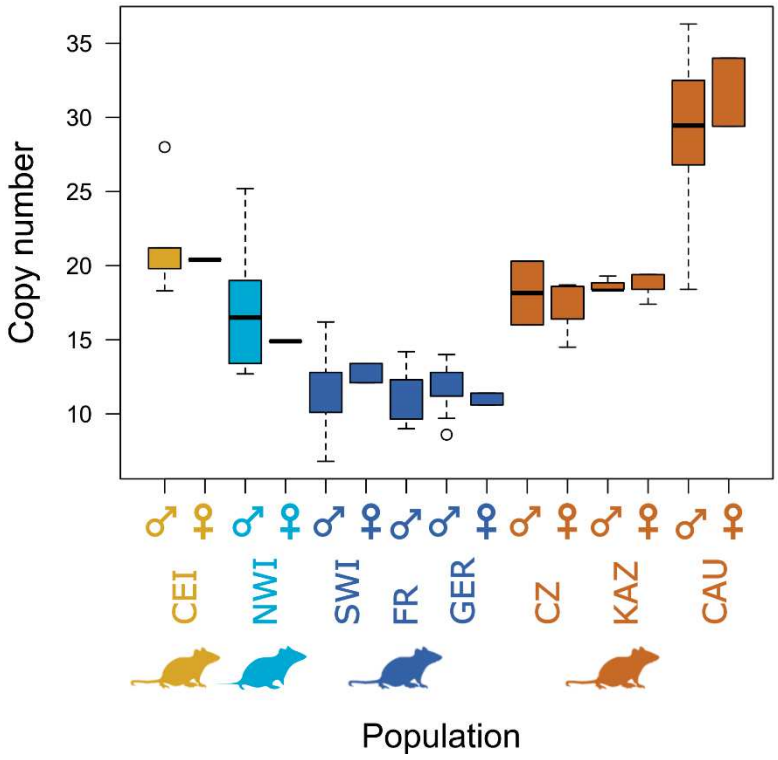


Figure 3 - Haploid copy number per population of a repeat contained on the Slx gene.

Discussion

In this study we explored the genetic composition of house mice from a poorly characterised region of its distribution area, in Northwestern Iran and neighbouring countries south of the Caucasus, as well as the Zagros Mountains. As we anticipated, we found contributions of the three previously defined genetic entities known to thrive in the surroundings of this region, namely *M. m. domesticus*, *M. m. musculus* and a lineage newly described in Central Iran. However, the concordance between the different methods used was not complete when inferring the relative contributions to the admixture. According to the descriptive analysis with the least assumptions (PCA based on individual genotypes), our northwesternmost samples could clearly be partitioned into *domesticus*-like (samples grouped under the name NWI), and *musculus*-like (grouped under the name CAU). The PCA suggested a possible minor contribution of *musculus* to NWI and of *domesticus* to CAU, a result in agreement with the ADMIXTURE analysis. The CAU results agree with a previous analysis of populations from south of the Caucasus using allozymes (Orth et al. 1996). The result in NWI is newly described here. In both cases this is in apparent contrast with the situation in Europe, where admixture between *domesticus* and *musculus* appears to be limited to a narrow hybrid zone, the transition occurring in 20-30 km (Raufaste et al. 2005, Macholan et al. 2007). We note however that the studies of the European hybrid zone have focused on markers with contrasting frequencies in the parental populations (most often diagnostic), which could introduce a bias as compared to our analyses. However, the study on allozymes in South Caucasus also used such diagnostic markers. Combining genome-wide and unbiased genetic typing (as we did here) with a finer geographic sampling of the contact zone (as was done in the European hybrid zone) would be necessary to fully compare the two situations, and to test the existence of tension zones between *domesticus* and *musculus* in Iran as well. The distribution of the assignment of our individual samples from this region to either *domesticus* or *musculus* is clearly geographically structured, but our sampling is too scarce and geographically loose to reliably fit models of high complexity such as clines (Sup. Fig. 8). In any case, this region is clearly a zone of admixture between at least these two subspecies.

The contribution of the Central Iranian lineage (represented by the CEI sample) was more difficult to detect concordantly across all methods used. In the PCA, our three new samples (CAU, NWI and MWI) are slightly closer to CEI than the other *domesticus* and *musculus* populations on the axis that differentiates CEI from all others (PC2). However, a CEI contribution was only suggested for CAU in the ADMIXTURE analysis.

None of these admixtures was detected in the Treemix analysis, but their occurrence would be compatible with the branches leading to these three new populations being shorter or closer to the root in the tree produced. We believe that the methods that make simplifying assumptions, such as ADMIXTURE with its simple underlying model, or that are constrained by an algorithm (such as the sequential addition of admixture events in TreeMix) fail to capture the full complexity of the relationships between these populations. One of the complexities is that, according to previous results, one of the parental populations considered here (CEI) is itself an admixture between *domesticus* and an ancestral population related to (although clearly differentiated from) *musculus*. If, as suggested by the PCA, our three new populations are three-way admixtures with different parental proportions, and given one of the parents (CEI) was previously shown to be admixed (one of its parents being unsampled and presumably extinct), none of the rather simplistic model-based methods applied here may be able to properly reflect such complexity.

The uniparentally transmitted non-recombining markers (mtDNA and the Y chromosome) however give us clear indications on the admixed nature of our newly sampled populations. The mitochondrial phylogeny clearly identifies three lineages characteristic of the three entities in contact in this geographic region (*domesticus*, *musculus* and CEI), which we interpret as witnessing past differentiation and phylogeographic pattern of these entities. We therefore interpret deviations from this pattern as resulting from secondary admixture. Thus, the mitochondrial complete sequences reported here suggest that the CAU population is a three-way admixture with a majority of *musculus* origin, and minor contributions of both the *domesticus* and CEI lineages. They also show some CEI contribution to the mostly *domesticus* NWI population. The compilation of published D-Loop sequencing results (Rajabi-Maham et al. 2012, Bonhomme et al. 2011), on a much better sampling of this region, enriches this view by showing that, although the three lineages are geographically structured, three-way admixture occurs in the regions corresponding to our CAU and NWI samples. Such is not the case however for the region corresponding to our MWI samples, which appears to be pure *domesticus* for mtDNA.

The Y chromosome phylogeny reveals the existence of two major, very divergent, lineages. One of them is characteristic of *domesticus* and is fixed in SWI and MWI. Surprisingly however, it is not found in the NWI sample, which we interpret as resulting from introgression into NWI, given the high divergence between the two Y lineages and the mostly *domesticus* autosomal makeup of NWI. The other major Y lineage is

phylogeographically well structured, which allows us to infer the origins of the Y introgression in NWI. Some of the NWI Ys belong to a subclade that contains all CAU samples, while the others belong to a sublineage containing all CEI samples, which suggests the contribution of these two populations to the genetic makeup of NWI. The denser sampling of the region using a simpler assay confirmed that the *domesticus* Y lineage appears fixed in the regions corresponding to SWI and MWI, while the non-*domesticus* lineage is fixed in the regions corresponding to all our other samples.

As we have discussed above, extending minor ancestry inference genome-wide appeared difficult using allele frequency-, model-based, methods. In waiting for more sophisticated analyses (for instance based on ancestral recombination graphs), and proper modelling of this complexity (for instance based on simulations), we attempted to simplify the question by focusing on the NWI population, which appeared particularly interesting due to the discordance between the Y and autosomal major ancestries. We also turned to an ancestry inference method (ELAI) based on linkage disequilibrium rather than allele frequencies. The choice of appropriate parental populations was however delicate. Our objective was to identify the contributions of three major evolutionary lineages to the NWI population, since the PCA pointed to three major poles of differentiation when combining axes 1 and 2, which together explain 67% of the variance. The method did infer a triple ancestry in NWI, that was not suggested by the other methods, apart from the f3 analysis. We note however that one of the parental populations used in this inference (CEI) has been shown to be an ancient admixture between populations related to *domesticus* and *musculus*. Therefore, one could question the ability of the method to properly distinguish the contributions of CEI from those of *domesticus* and *musculus*. The gradients of introgression along the geographic directions predicted by the geographic positions of the inferred donor populations seem to indicate that the method was able to at least partly disentangle this problem. We note that genomic regions of inferred CEI ancestry in NWI, based on linkage disequilibrium, could originally be of either *domesticus*-like or *musculus*-like origin in CEI. A finer analysis of these genomic regions would be interesting to understand the relationship between this ancient origin and the propensity to introgress into the NWI, mostly *domesticus,* population. With the present results, we can however conclude that NWI is a three-way admixture between the entities that surround it geographically, with a major *domesticus* contribution.

Our most remarkable result was the complete replacement, in NWI, of the *domesticus* Y by *musculus*-related Y lineages, apparently coming from the neighbouring

populations (CAU and mostly CEI). This contrasts with the relatively modest inferred autosomal and mitochondrial contributions of these populations to NWI, and suggests that this massive Y introgression is driven by selection. The propensity of the *musculus* Y chromosome to cross the hybrid zone between *domesticus* and *musculus* has been described in great detail in one transect in the Czech Republic (Macholan et al. 2008), and shown to also prevail over most of the European hybrid zone (Macholan et al. 2019), including in Scandinavia (Jones et al. 2010). A possible origin of the advantage of this Y chromosome comes from the observation that it contributes to partially rescue hybrid male infertility in the Czech transect (Albrechtova et al. 2012). Hybrid male infertility is frequent in the centre of the European hybrid zone (Albrechtova et al. 2012, Britton Davidian et al. 2005), and is the major known source of hybrid unfitness and postzygotic isolation between these subspecies. We have shown that Northwestern Iran is a zone of admixture between several differentiated genomes, and although hybrid male fertility has not been studied in this region, it is likely to be affected. We therefore emit the hypothesis that the *musculus* Y chromosome lineage has an advantage in situations of admixture between house mouse subspecies, such as in NWI. The analysis of the genetic composition of the admixed CEI population had also led us to the conclusion of an advantage maintaining the non-*domesticus* Y in this population, based on the conjunction of 30-40% *domesticus* autosomal contribution with a lower *domesticus* contribution for the X and the absence of contribution for both mtDNA and the Y. We note that the prevalence of the *musculus* Y lineage was also reported in other zones of admixture between house mouse subspecies, such as between *M. m. musculus* and *M. m. castaneus* (for which hybrid male sterility is also reported) in SE Asia (Boissinot and Boursot 1997) and Japan. We also note that old inbred mouse laboratory strains, although they are essentially of *domesticus* origin, all carry a Y chromosome variant of the *musculus* lineage (Bishop et al. 1985), also found in Japan (Nagamine et al. 1992). The introduction and selective fixation of this Y chromosome type could thus also be linked to episodes of admixture during the derivation of these strains, presumably with male Japanese fancy mice used to introduce interesting characters.

We found, among highly introgressed genomic regions in NWI, an enrichment in genes related to male fertility, particularly on the X. This could be an indication that these genes are interacting with the introgressed *musculus* Y to control male fertility, explaining their co-introgession with the Y. Other good candidates for co-introgression with the *musculus* Y are X ampliconic regions. Various experiments have demonstrated functional interactions between X and Y ampliconic regions, particularly those containing

the *Slx* and *Sly* genes, respectively. The Y amplicons were shown to repress the expression of the X amplicons, in a dosage sensitive manner. Dosage imbalance causes modifications of the sex-ratio of the progeny, a relative excess of Y copies leading to more males (and thus a transmission advantage for the Y), a relative deficit to more females (transmission advantage for the X) (Morgan and Pardo Manuel de Villena 2017). The *musculus* Y is known to harbour more copies of the *Sly* amplicon than the *domesticus* Y (Morgan and Pardo Manuel de Villena 2017). In accordance, in the European hybrid zone, a comparatively male-biased sex-ratio was reported in geographic regions of incursion of the *musculus* Y into *domesticus* territory (Macholan et al. 2008). This conflict for transmission should lead to a coevolution of copy numbers between the X and Y to maintain transmission rates in accordance with the control of optimal sex ratio by natural selection in the populations. One would thus predict a correlation between X and Y copy numbers across populations/subspecies, a prediction we attempted to test among our samples.

Our short-read sequencing method and the incompleteness of the mouse reference genome in ampliconic regions led us to use a method based on k-mer abundance to estimate copy numbers (CNs). For the *Sly* family, we relied on already reported results, showing a higher CN on the *musculus* Y than the *domesticus* one. For the *Slx* ampliconic region, we found relatively and consistently low CNs in unadmixed *domesticus* samples. Second, we found relatively higher CNs in all representatives of the *musculus* branch of the Y tree, including NWI. The estimated CNs in this population is clearly higher than in *domesticus*, and only slightly lower than in CEI (one of the inferred Y donors). However, none of the NWI samples reaches values as high as those inferred in CAU (the other inferred Y donor). Whether this resemblance of NWI *Slx* CN with that of the populations with a *musculus* Y is due to introgression or convergence cannot be determined. An examination of the ancestry of the sequences flanking the ampliconic region could potentially be informative in this respect. An estimation of *Sly* CN is also lacking to fully characterise the correlation between X and Y CNs across populations. If the presumed fertility-rescuing property of the *musculus* Y is not linked to *Sly* CN (which is likely), one would expect Y variants with the lowest CN to preferentially introgress into populations with low *Slx* CN, because laboratory experiments have shown that high imbalance between *Slx* and *Sly* CNs impairs male fertility (Morgan and Pardo Manuel de Villena 2017). Under this hypothesis, transmission distortion could be the major force driving gradual CN inflation in isolated populations, but not the force driving extensive introgression during admixture between populations with contrasted CNs. A

prediction of this model would be that introgressing Ys should tend to be in the low range of CN variation in the donor population. If X and Y CNs are correlated inside populations, we predict very high *Sly* CN in CAU (paralleling the measured very high *Slx* CN), which could explain the apparently low frequency of Y introgression from CAU into NWI. Overall, the model proposes that the introgression of the *musculus* Y in all regions of admixture is driven by its hybrid male fertility rescuing properties, but that its distorting properties may rather be a disadvantage. There could be two reasons for the invasion not to apparently occur beyond the regions of admixture: that the positive effect on male fertility is restricted to admixture conditions, and that the transmission distortion advantage is counterbalanced by its impairment of male fertility. The successfully introgressing Y haplotypes would be those for which the hybrid rescuing property is not overwhelmed by the collateral negative effects of a too strong transmission distortion (reduced male fertility). An open related question is to understand why CN inflation has occurred at different speeds in different taxa (slow in *domesticus* and fast in *musculus*). Although this could be by chance, there could be deterministic reasons linked to varying mutation processes, or to varying modalities of the control of the sex-ratio, according to population structure and dynamics inducing varying degrees of group selection.

Material and Methods

Additional detailed information on materials and methods with associated references is provided in supplementary material.

Data filtering, read mapping, genotype calling, and iterative mapping

Individual sequencing data were processed following *Bettina et al*. (2016). In brief, filtered reads were mapped against the mouse genome reference sequence - *mm10* ('GRCm38 - Mm10 - Genome - Assembly - NCBI' n.d.) using bwa-mem (Li 2013). The sorting, marking and duplicates removal use performed with Picard tools software suite ('Picard Tools - By Broad Institute' n.d.) was used for. Raw SNP and indel calls were obtained following the GATK (Auwera and O'Connor 2020) 'Best Practice' instructions on joint genotyping. The raw .vcf files were subjected to the GATK VSQR SNP filtering step, which uses known variants as training data to predict whether a new variant is likely a true positive, or a false positive. The *Mus musculus* SNP database (dbSNP) ('mgp.v5.merged.snps_all.dbSNP142.vcf' downloaded from ('Mouse Genomes Project - Wellcome Sanger Institute' n.d.) was used as training dataset and the vcf was filtered for 'PASS' SNPs. Additional very stringent hard filtering criteria were applied and included these SNPs as training sets as well (see details in *Bettina et al.* (2016)). Due to an absence of a reliable indel reference dataset we decided to exclude all indels from our final dataset. Highly related individuals were removed, and t-carriers were identified due to its impact on the chromosome 17 (Sup. Fig. 13).

Analysis of Population structure

An initial unsupervised population structure analysis was performed using the non-parametric principal component analysis (PCA), as implemented in PLINK 2 (Chang et al. 2015). The PCA was based on a subsample of bi-allelic SNPs at least 25 kb apart and without missing genotypes. Additionally, we apply Admixture v1.3.0 (Alexander, Novembre, and Lange 2009) and its implemented Bayesian Markov Chain Monte Carlo model (MCMC) on the pruned datasets and the cross-validation error was calculated for identifying the best K value. Five replicate runs were performed for each number of populations (k) set from 1 to 6. Replicate runs were analyzed using CLUMPAK (Kopelman et al. 2015) and DISTRUCT (Rosenberg 2004) was used to plot the results. The best number of populations, K, was inferred using Evanno's delta K method (Evanno, Regnaut, and Goudet 2005), as implemented in CLUMPAK.

Inferring genetic relationships and gene flow between populations

Genetic relationships and gene flow were inferred using the Treemix v.1.13 approach (Pickrell et al. 2012; Pickrell and Pritchard 2012) and f3-statistics (Reich et al. 2009; Patterson et al. 2012). First TreeMix was used to estimate the genetic relationships among the different populations. First allele frequencies were estimated among the randomly sampled alleles and subsequently TreeMix model was run accounting for linkage disequilibrium by grouping sites in blocks of 500 single-nucleotide polymorphisms (-k 500) setting the *Mus cypriacus* as root. Standard errors (-SE) and bootstrap replicates (-bootstrap) were used to evaluate the confidence in the inferred tree topology. After constructing a maximum-likelihood tree, migration events were added (−m) and iterated 5 times for each value of m (1–5) to check for convergence in the likelihood of the model as well as the explained variance following each addition of a migration event. The inferred maximum-likelihood trees were visualized with the in-built TreeMix R script plotting functions.

The f3-statistics were calculated in the form of f3 (target, source 1, source 2) using ADMIXTOOLS v.7.0.2 (Patterson et al. 2012), and provide evidence that the target population is derived from an admixture of populations related to sources 1 and 2. We tested all possible combinations of target and source populations in our dataset. Standard errors were obtained using blocks of 500 SNPs. Tests with a Z-score < −3 were considered significant.

Mitochondrial DNA phylogeny

Whole mitochondrial genomes (except some *D-loop* sequences) were recovered for all individuals with NOVOPlasty 3.7.2 (Dierckxsens, Mardulyn, and Smits 2016). The mitochondrial genome phylogeny was reconstructed using a neighbour joining tree method implemented on MEGAX (Kumar et al. 2018) with 100 iterations.

To the study dataset, we add up *castaneus* and *Mus spretus* from previously published data (Harr et al. 2016). Results were further confirmed by a more extensive survey of 422 individuals D-loop sequences (Sup. Fig. 4 and Sup. Text).

Y chromosome analysis

Due to Y chromosome lack of recombination, we used the neighbour joining tree method implemented on MEGAX to reconstruct the chromosome phylogeny (Kumar et al. 2018). We used a concatenated alignment of single copy Y genes and run 100 iterations to construct a phylogenetic tree (Figure2).

Inference of introgression—Efficient Local Ancestry Inference (ELAI)

To perform the ancestry deconvolution analysis of central Iranian genomic segments we used the Efficient Local Ancestry Inference (ELAI) method (Guan 2014). This method implements a two-layer HMM (hidden Markov model) to infer local ancestry of admixed individuals without prior definition of window sizes, by looking at two layers of linkage-disequilibrium—within and among defined groups. It returns at each variable position in the genome the most likely proportions of ancestries (true values being expected to take values 0, 1, or 2 in two-way admixture). We ran ELAI following Marques et al. (unpublished) on the unphased dataset with four population samples: NWI defined as the admixed population, while the Afghan *musculus*, the southwest Iranian population of *domesticus* and the recently described CEI were defined as the possible donors in the admixture.

. We define the number of upper-layer as 3, representing *musculus, domesticus* and CEI, and that of lower-layer clusters to 15 (five times the number of upper-layer clusters, as recommended). We performed three different expectation maximization (EM) runs of 20 steps with mixture generation values of 25k,50k and 100k and different random seeds. ELAI results were averaged over the three independent runs. Sites with a proportion of *musculus, domesticus* or CEI ancestry between 0.5 and 1.5 were considered heterozygous for and those with values over >1.5 homozygous for introgression. For each individual, single ancestry fragments were defined as consecutive sites defined according to the above criteria.

GO enrichment analyses

We tested for functional enrichment of genes with high introgression frequencies using the g:Profiler R package (Reimand, Kolde, and Arak 2019). Categories with less than five genes were excluded and the SCS algorithm was applied for computing multiple testing correction for *p*-values. Only genes within segments windows with more than 20 informative sites were considered for the background list of genes. We used both the mouse GO term annotation and the more complete human one.

Relationship between geography and introgression

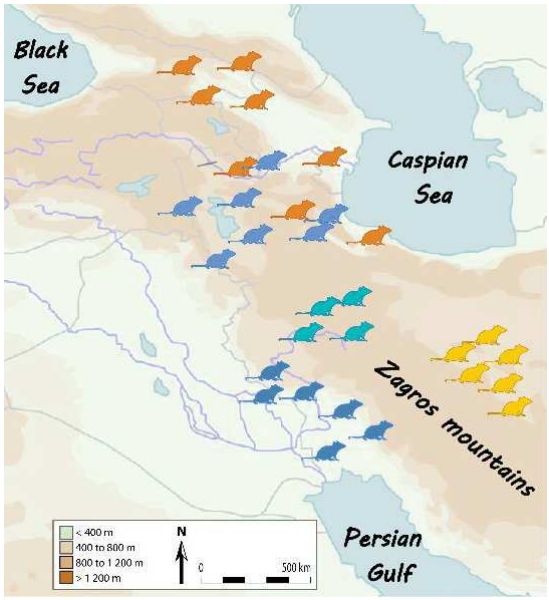
We tested the statistical relation of introgression prevalence with geographical distance to the introgression source. Introgression prevalence was defined by the proportion of introgressed fragments per individual divided the chromosomal length. Chromosome 17 was excluded given their known structural differences between *t*-carriers. Spearman's rank correlation was applied to test the correlation between introgression prevalence and geographical distance to the possible source populations.

De novo repeat database and copy number variation (CNV) analyses

REPdenovo v0.1.0 (Chu, Nielsen, and Wu 2016) was run on per individual pre-processed genomic short reads in default mode with a minimum repeat frequency of 100× (i.e., the squared mean genome coverage of 20×). In order to have a global representation of different worldwide populations we add up already published data of four additional populations (two *musculus* – Kazakhstan (KAZ) and Czech Republic (CZ) and two *domesticus* – France (FR) and Germany (GER) - sample details on Sup. Table 13. We create two subsets one by combining all male repeats and another one for females. We then used tblastn (Altschul et al. 1990) with the parameters -evalue 1e-6, -numalignments 1, and -numdescriptions 1 to blast the REPdenovo repeat libraries against the *Mus musculus* genome. We retained all repeats with blast hits on sex chromosomes. Redundancy on each dataset was removed using cd-hit (Fu et al. 2012) with standard parameters. We then used Bowtie v2 (Langmead and Salzberg 2012) in the sensitive local alignment mode to map all reads to the filtered repeated database. The mapping results were then used to calculate mean per-contig coverage per individual with bamtools (Barnett et al. 2011). Copy numbers were estimated dividing the mean mapped read coverage per contig by the sex-specific single copy genes coverage. Haploid copy numbers were calculated for each detected CNV according to the individual sex.

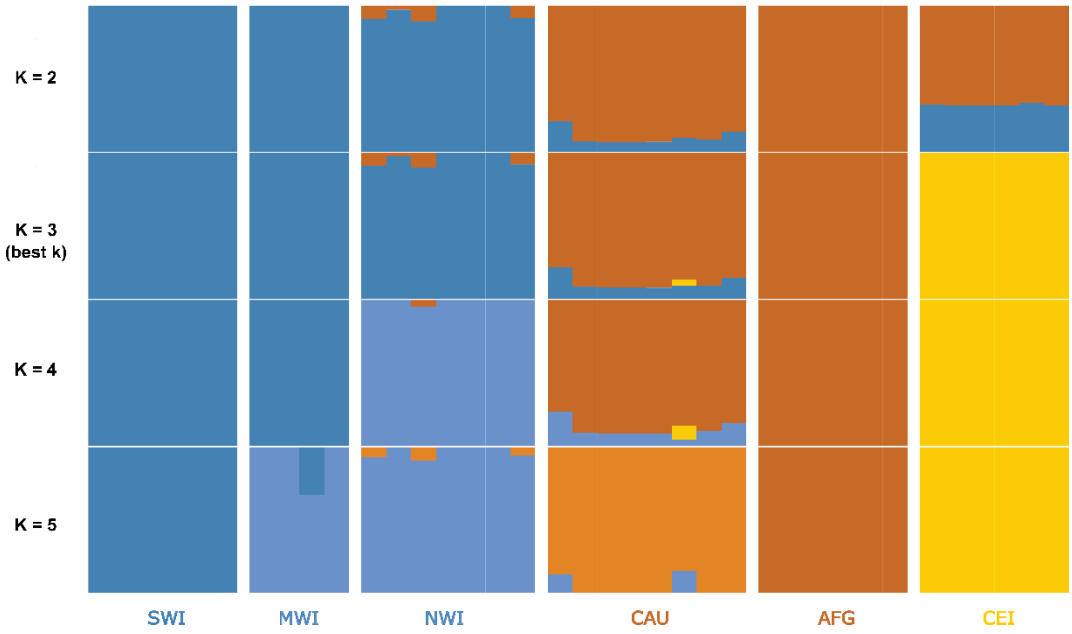
Supplementary Figures

Sampling details (full information Sup. Table 1)



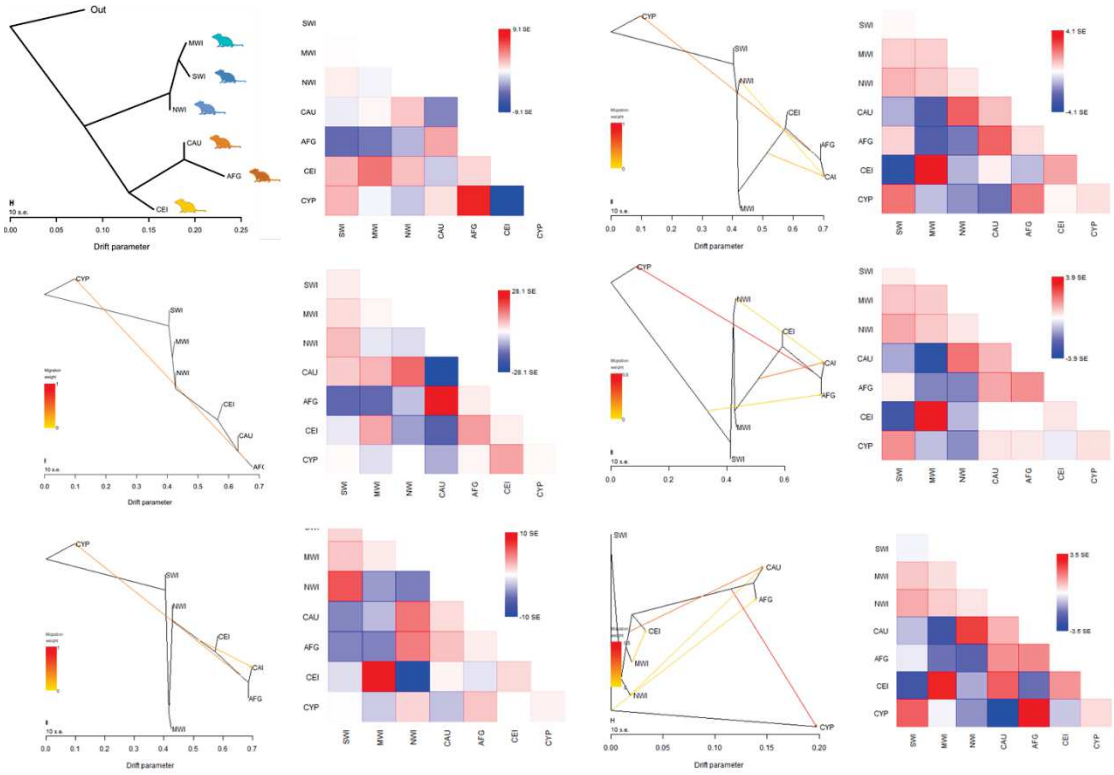
Sup. Fig. 1 – Map with the rough location of specimens sampling. Colours represent the separation in populations. Being: Yellow - central Iranian lineage; Dark blue – southwest domesticus; Blue – northwest domesticus; Lightblue – zagroos domesticus; dark orange – Caucasus musculus.

Admixture analysis



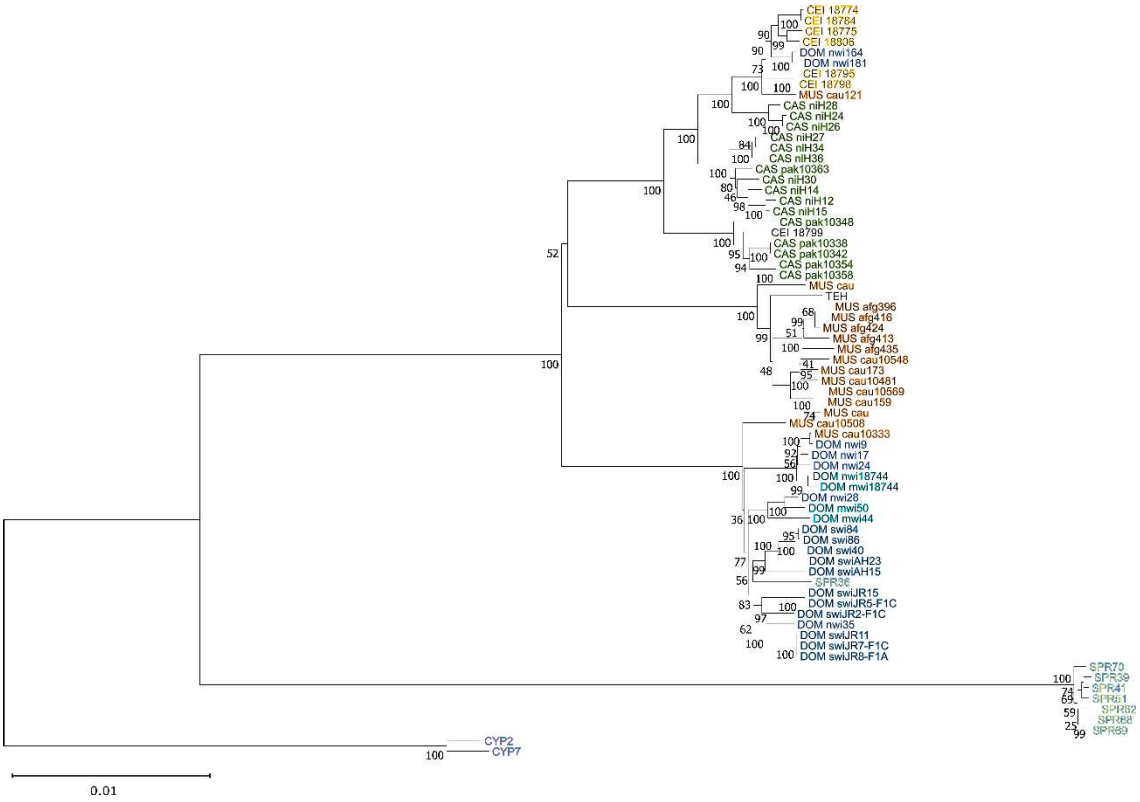
Sup. Fig. 2 - ADMIXTURE analysis for K=2 to K=5.

Treemix

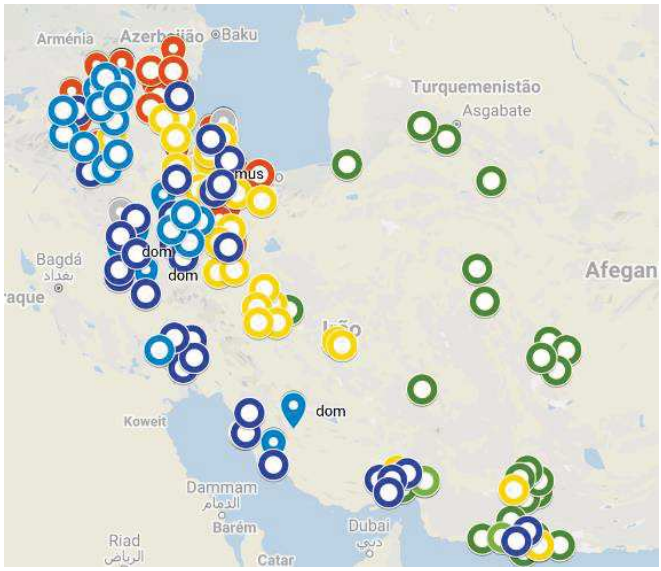


Sup. Fig. 3 – Treemix runs for m=0 to m=5 and respective residual plots.

Mitochondrial data – phylogeny and *D*-loop extensive analysis



Sup. Fig. 4 – Neighbour-joining mitochondrial genome phylogeny

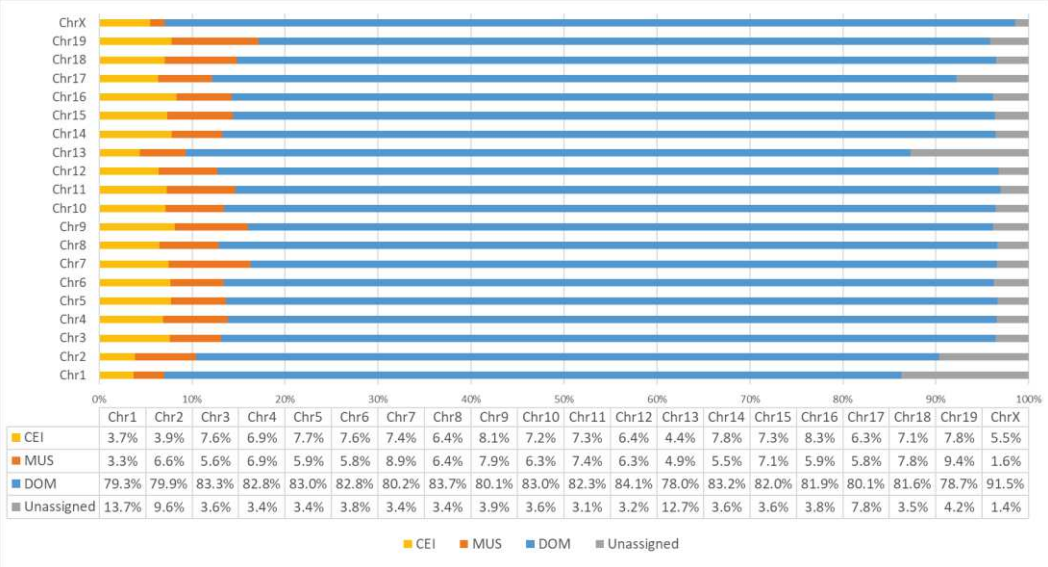


Sup. Fig. 5 – Mitochondrial D-loop types distribution map. In green the castaneus like, Blue – domesticus-like, Orange – musculus-like and Yellow-central Iranian lineage-like

ChrY indel data

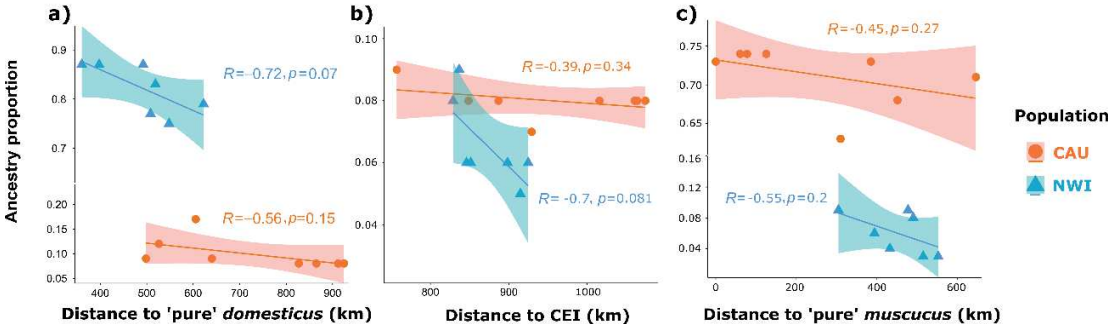
Sup. Fig. 6 – Distribution map summarizing the results of the Y diagnostic INDEL PCR. In blue the domesticus-like Ys and in dark orange the musculus and CEI-like Y's

Introgression frequency and its correlation with geography

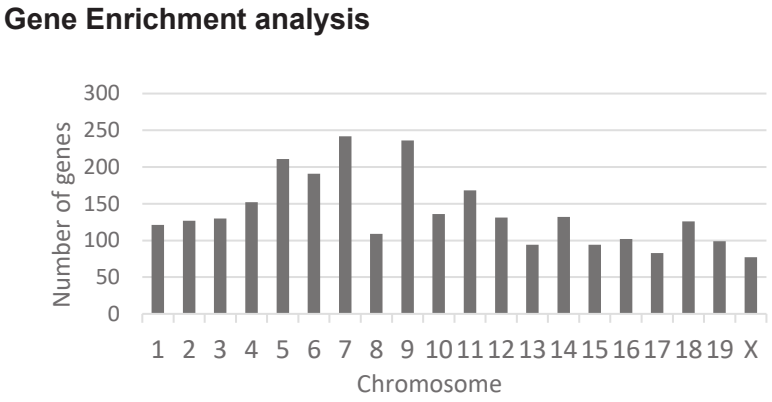


	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10	Chr11	Chr12	Chr13	Chr14	Chr15	Chr16	Chr17	Chr18	Chr19	ChrX
CEI	3.7%	3.9%	7.6%	6.9%	7.7%	7.6%	7.4%	6.4%	8.1%	7.2%	7.3%	6.4%	4.4%	7.8%	7.3%	8.3%	6.3%	7.1%	7.8%	5.5%
MUS	3.3%	6.6%	5.6%	6.9%	5.9%	5.8%	8.9%	6.4%	7.9%	6.3%	7.4%	6.3%	4.9%	5.5%	7.1%	5.9%	5.8%	7.8%	9.4%	1.6%
DOM	79.3%	79.9%	83.3%	82.8%	83.0%	82.8%	80.2%	83.7%	80.1%	83.0%	82.3%	84.1%	78.0%	83.2%	82.0%	81.9%	80.1%	81.6%	78.7%	91.5%
Unassigned	13.7%	9.6%	3.6%	3.4%	3.4%	3.8%	3.4%	3.4%	3.9%	3.6%	3.1%	3.2%	12.7%	3.6%	3.6%	3.8%	7.8%	3.5%	4.2%	1.4%

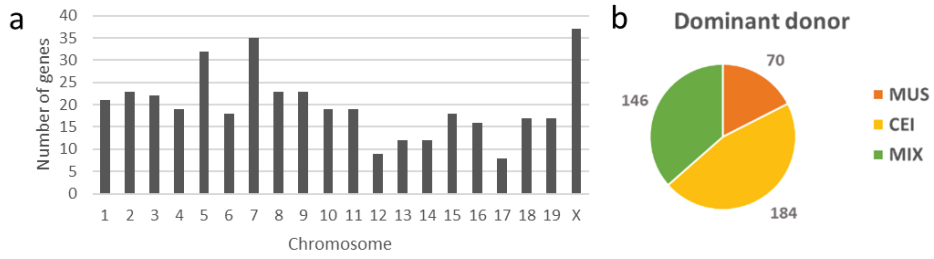
Sup. Fig. 7 - Northwest Iranian individuals' genome partition according to each parental contribution inferred by ELAI.



Sup. Fig. 8 - Correlation between introgression and distance of CAU and NWI to the most likely source (set as the less admixed individual of each population). (a) Correlation to the domesticus-like ancestry (NWI). (a) Correlation to the musculus-like ancestry (CAU). (a) Correlation to the central iranian lineage-like ancestry (CEI).



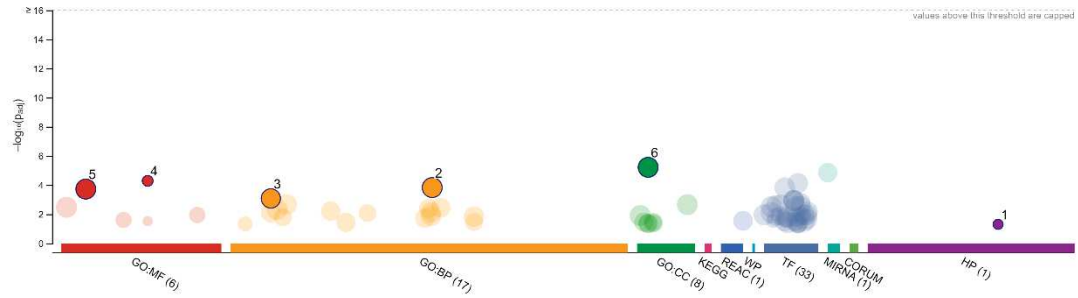
Sup. Fig. 9 – High-frequency (>50%) non-domesticus introgression genes distribution



Sup. Fig. 10 – High-frequency (>50%) non-domesticus introgression genes at the enriched categories. a) Number of genes per chromosome. b) Enrichment categories genes ancestry assignment

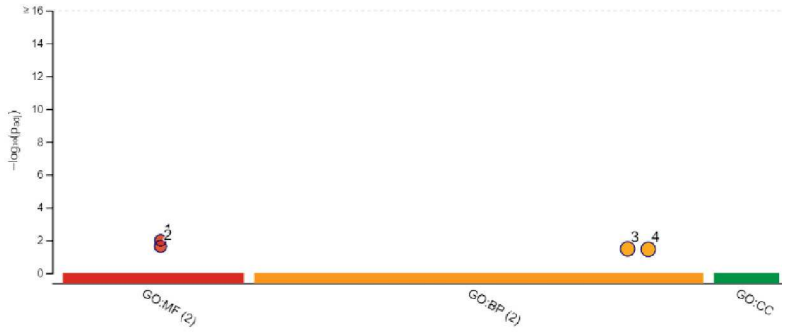
GO enrichment analysis and the presence of several male fertility terms

(https://biit.cs.ut.ee/gplink/l/g-fjKvmWRy)



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	HP	HP:0025546	Abnormal mean corpuscular hemoglobin concentr...	4.971×10^{-2}
2	GO:BP	GO:0050896	response to stimulus	1.485×10^{-4}
3	GO:BP	GO:0007165	signal transduction	8.289×10^{-4}
4	GO:MF	GO:0042287	MHC protein binding	5.156×10^{-5}
5	GO:MF	GO:0005515	protein binding	1.877×10^{-4}
6	GO:CC	GO:0016020	membrane	6.063×10^{-6}

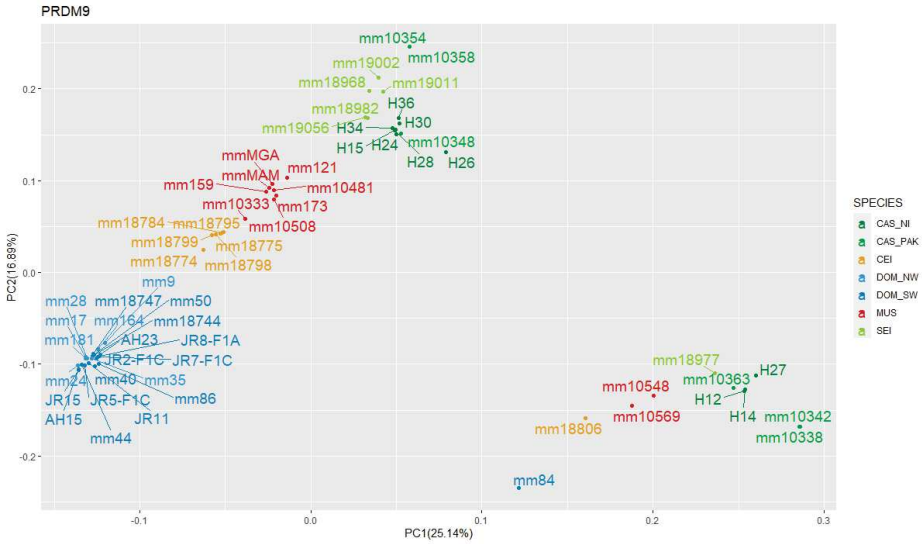
Sup. Fig. 11 – Gprofiler Gene Ontology analysis of genes showing non-domesticus introgression frequencies of at least 80%.



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0042288	MHC class I protein binding	1.000×10^{-2}
2	GO:MF	GO:0042287	MHC protein binding	2.367×10^{-2}
3	GO:BP	GO:1902106	negative regulation of leukocyte differentiation	3.380×10^{-2}
4	GO:BP	GO:1903707	negative regulation of hemopoiesis	3.584×10^{-2}

Sup. Fig. 12 - Gprofiler Gene Ontology analysis of genes showing fixation of non-domesticus introgression (100% introgression frequency).

t-haplotype detection



Sup. Fig. 13 - PRDM9 principal component analysis

Supplementary tables

Sup. Table 1 - Sampling details (in bold samples included in the final dataset)

Sample ID	POP ID	Lineage	Colour	Sex	Region	*t*-haplotype	Coverage	Source
afg396	AFG	*musculus*	Orange	Male	Afghanistan	Yes	14	(Harr et al. 2016)
afg413	AFG	*musculus*	Orange	Male	Afghenistan	No	21	(Harr et al. 2016)
afg416	AFG	*musculus*	Orange	Male	Afghanistan	Yes	16	(Harr et al. 2016)
afg424	AFG	*musculus*	Orange	Male	Afghanistan	No	17	(Harr et al. 2016)
afg435	AFG	*musculus*	Orange	Female	Afghanistan	No	19	(Harr et al. 2016)
afg444	AFG	*musculus*	Orange	Male	Afghanistan	Yes	18	(Harr et al. 2016)
cau121	CAU	undetermined	Light Orange	Male	Caucasus	No	17	This study
cau159	CAU	undetermined	Light Orange	Female	Caucasus	No	15	This study
cau173	CAU	undetermined	Light Orange	Female	Caucasus	No	17	This study
cau10333	CAU	undetermined	Light Orange	Male	Caucasus	No	21	This study
cau10481	CAU	undetermined	Light Orange	Male	Caucasus	No	18	This study
cau10508	CAU	undetermined	Light Orange	Male	Caucasus	No	19	This study
cau10548	CAU	undetermined	Light Orange	Male	Caucasus	Yes	20	This study
cau10569	CAU	undetermined	Light Orange	Male	Caucasus	Yes	19	This study
cei18774	CEI	Central Iran	Gold	Male	Iran - Central	No	19	Marques et al.(Publication IV)
cei18775	CEI	Central Iran	Gold	Male	Iran - Central	No	16	Marques et al.(Publication IV)
cei18784	CEI	Central Iran	Gold	Male	Iran - Central	No	17	Marques et al.(Publication IV)
cei18795	CEI	Central Iran	Gold	Male	Iran - Central	No	16	Marques et al.(Publication IV)
cei18798	CEI	Central Iran	Gold	Female	Iran - Central	No	15	Marques et al.(Publication IV)
cei18799	CEI	Central Iran	Gold	Male	Iran - Central	No	17	Marques et al.(Publication IV)
cei18806	CEI	Central Iran	Gold	Male	Iran - Central	Yes	16	Marques et al.(Publication IV)
mwi44	MWI	undetermined	Light Blue	Male	Iran - Zagroos	No	15	This study
mwi50	MWI	undetermined	Light Blue	Male	Iran - Zagroos	No	14	This study
mwi18744	MWI	undetermined	Light Blue	Male	Iran - Zagroos	No	17	This study

mwi18747	MWI	undetermined	Light Blue	Female	Iran - Zagroos	No	18	This study
nwi9	NWI	undetermined	Blue	Male	Iran - NW	No	15	This study
nwi17	NWI	undetermined	Blue	Female	Iran – NW	No	15	This study
nwi24	NWI	undetermined	Blue	Male	Iran – NW	No	18	This study
nwi28	NWI	undetermined	Blue	Male	Iran - NW	No	18	This study
nwi35	NWI	undetermined	Blue	Male	Iran – NW	No	17	This study
nwi164	NWI	undetermined	Blue	Male	Iran – NW	No	17	This study
nwi181	NWI	undetermined	Blue	Male	Iran – NW	No	16	This study
swi40	SWI	*domesticus*	Navy Blue	Female	Iran – SW	No	17	Marques et al.(Publication IV)
swi84	SWI	*domesticus*	Navy Blue	Female	Iran – SW	Yes	16	Marques et al.(Publication IV)
swi86	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	17	Marques et al.(Publication IV)
swiAH15	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	22	(Harr et al. 2016)
swiAH23	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	24	(Harr et al. 2016)
swiJR11	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	23	(Harr et al. 2016)
swiJR15	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	22	(Harr et al. 2016)
swiJR2	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	23	(Harr et al. 2016)
swiJR5	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	17	(Harr et al. 2016)
swiJR7	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	18	(Harr et al. 2016)
swiJR8	SWI	*domesticus*	Navy Blue	Male	Iran – SW	No	17	(Harr et al. 2016)
cyp17368	CYP	cypriacus	Purple	Male	Cyprus	Not tested	15	Marques et al.(Publication IV)
cyp17373	CYP	cypriacus	Purple	Male	Cyprus	Not tested	16	Marques et al.(Publication IV)

Sup. Table 2 – Sampling details of additional data included in the CNVs analysis

Sample ID	POP ID	Lineage	Data type	Colour	Sex	Region	*t*-haplotype	Coverage	Source
KazAL1	KAZ	*musculus*	WGS	Orange	Female	Kazakhstan	No	23	(Harr et al. 2016)
KazAL16	KAZ	*musculus*	WGS	Orange	Male	Kazakhstan	No	25	(Harr et al. 2016)
KazAL19	KAZ	*musculus*	WGS	Orange	Female	Kazakhstan	Yes	24	(Harr et al. 2016)
KazAL33	KAZ	*musculus*	WGS	Orange	Female	Kazakhstan	No	25	(Harr et al. 2016)
KazAL38	KAZ	*musculus*	WGS	Orange	Male	Kazakhstan	No	25	(Harr et al. 2016)
KazAL40	KAZ	*musculus*	WGS	Orange	Female	Kazakhstan	No	26	(Harr et al. 2016)
KazAL41	KAZ	*musculus*	WGS	Orange	Male	Kazakhstan	Yes	26	(Harr et al. 2016)
KazAL42	KAZ	*musculus*	WGS	Orange	Female	Kazakhstan	No	25	(Harr et al. 2016)
fr14	FR	*domesticus*	WGS	Blue	Male	France	No	24	(Harr et al. 2016)
fr15B	FR	*domesticus*	WGS	Blue	Male	France	No	23	(Harr et al. 2016)
Fr16B	FR	*domesticus*	WGS	Blue	Male	France	No	24	(Harr et al. 2016)
Fr18B	FR	*domesticus*	WGS	Blue	Male	France	No	24	(Harr et al. 2016)
frB2C	FR	*domesticus*	WGS	Blue	Male	France	Yes	14	(Harr et al. 2016)
frC1	FR	*domesticus*	WGS	Blue	Male	France	Yes	20	(Harr et al. 2016)
frE1	FR	*domesticus*	WGS	Blue	Male	France	No	22	(Harr et al. 2016)
frF1B	FR	*domesticus*	WGS	Blue	Male	France	Yes	23	(Harr et al. 2016)
fr14	FR	*domesticus*	WGS	Blue	Male	France	No	24	(Harr et al. 2016)
fr15B	FR	*domesticus*	WGS	Blue	Male	France	No	23	(Harr et al. 2016)
Fr16B	FR	*domesticus*	WGS	Blue	Male	France	No	24	(Harr et al. 2016)
gerHG06	GER	*domesticus*	WGS	Blue	Female	Germany	Not tested	11	(Harr et al. 2016)

gerHG08	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	14	(Harr et al. 2016)
gerHG13	GER	*domesticus*	WGS	Blue	Female	Germany	Not tested	12	(Harr et al. 2016)
gerTP1	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	23	(Harr et al. 2016)
gerTP12	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	22	(Harr et al. 2016)
gerTP17	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	24	(Harr et al. 2016)
gerTP3	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	23	(Harr et al. 2016)
gerTP4	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	24	(Harr et al. 2016)
gerTP5	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	20	(Harr et al. 2016)
gerTP7	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	20	(Harr et al. 2016)
gerTP8	GER	*domesticus*	WGS	Blue	Male	Germany	Not tested	22	(Harr et al. 2016)

Sup. Table 3 – Sampling details of additional data included in the chromosome Y analysis

Region	Locality	Latitude	Longitude	Number of individuals	Y-type
Ardabile	Ardabile	38.249	48.293	3	*musculus*
Ardabile	Bilehsavar	39.379	48.362	2	*musculus*
Ardabile	Khalkhal	37.595	48.554	3	*musculus*
Ardabile	Meshkinshahr	38.399	47.682	3	*musculus*
Ardabile	Namin	38.427	48.484	4	*musculus*
Azar. East	Jolfa	38.930	45.672	1	*musculus*
Azar. East	Komar-1	38.997	46.521	1	*musculus*
Azar. East	Komar-2	38.997	46.521	1	*musculus*
Azar. East	Komar-2	38.997	46.521	1	n.d.
Azar. East	Komar-3	38.997	46.521	2	*musculus*
Azar. East	Maragheh	37.382	46.254	2	*musculus*
Azar. East	Marand	38.428	45.775	2	*musculus*
Azar. West	Mahabad	36.765	45.722	3	*musculus*
Azar. West	Salmass	38.198	44.768	2	*musculus*
Azar. West	Uromieh	37.555	45.103	1	*musculus*
Boushehr	Rahdar	27.970	51.850	2	*domesticus*
Fars	Shiraz	29.060	52.540	1	*domesticus*
Gilan	Langerud	37.195	50.149	1	*musculus*
Gilan	Sangar	37.178	49.694	1	*musculus*

Gilan	Talesh	37.365	50.093	3	*musculus*
Gilan	Talesh	37.365	50.093	1	n.d.
Hamadan	Saamen	34.199	48.704	12	*domesticus*
Hamadan	Saamen	34.199	48.704	3	n.d.
Ilam	Dareh shahr	33.140	47.380	7	*domesticus*
Ilam	Shirvan cherdavel	33.850	46.460	1	*domesticus*
Kermanshah	Javanroud	34.810	46.490	1	n.d.
Kermanshah	Kermanshah	34.314	47.065	1	*domesticus*
Kordestan	Sanandaj	35.320	47.994	1	*domesticus*
Markazi	Ashtian	34.550	50.000	3	*musculus*
Markazi	Mahalat	33.880	50.500	1	*musculus*
Markazi	Saveh	35.020	50.330	2	*musculus*
Markazi	Tafresh	34.624	49.987	3	*musculus*
Qazvin	Abyek	36.050	50.530	1	*musculus*
Qazvin	Qazvin	36.270	50.000	1	*musculus*
Qazvin	Takestan	36.070	49.700	3	*musculus*
Tehran	Tehran	35.809	51.433	3	*musculus*
Zanjan	Abhar	36.141	49.215	1	*musculus*
Zanjan	Zanjan	36.670	48.480	6	*musculus*

Sup. Table 4 – Sampling details of additional data included in the D-loop mitochondrial analysis with the type counts per locality

Province	**Locality**	**Country**	**Long O**	**Lat N**	***domesticus* G2**	***domesticus* G3**	***castaneus* HG1B**	***castaneus* HG2**	***castaneus* HG3**	***musculus***
Ardabile	Ardabile	Iran	48.29	38.25			5			
Ardabile	Aslandooz	Iran	47.41	39.44						1
Ardabile	Bilehsavar	Iran	48.36	39.38						3
Ardabile	Khalkhal	Iran	48.55	37.60			5			
Ardabile	Meshkinshahr	Iran	47.68	38.40						5
Ardabile	Namin	Iran	48.48	38.43		3	2			1
Azarbaijan-e Gharbi	Mahabad	Iran	45.72	36.76	1	3	2			
Azarbaijan-e Gharbi	Salmass	Iran	44.77	38.20	1	4				
Azarbaijan-e Gharbi	Serow	Iran	44.68	37.73	1					
Azarbaijan-e Gharbi	Uromieh	Iran	45.10	37.56	1					
Azarbaijan-e Sharghi	Jolfa	Iran	45.67	38.93	2					
Azarbaijan-e Sharghi	Komar-1	Iran	46.50	39.00	1					
Azarbaijan-e Sharghi	Komar-2	Iran	46.54	39.00	6					

Azarbaijan-e Sharghi	Komar-3	Iran	46.51	39.00	3				
Azarbaijan-e Sharghi	Maragheh	Iran	46.25	37.38	3		2		
Azarbaijan-e Sharghi	Marand	Iran	45.78	38.43	3				
Azarbaijan-e Sharghi	Tabriz	Iran	46.29	38.08	1				
Boushehr	Boushehr	Iran	50.84	28.97		1	2		
Boushehr	Rahdar	Iran	51.85	27.97		3			
Fars	Kenar takhte	Iran	51.04	29.54		1			
Gilan	Langerud	Iran	50.15	37.20		1	1		
Gilan	Manjil	Iran	49.40	36.74			1		
Gilan	Sangar	Iran	49.69	37.18			2		
Gilan	Saravan	Iran	49.64	37.03			1		
Gilan	Talesh	Iran	50.09	37.37		2	4		
Golestan	Gorgan	Iran	54.44	36.84				1	
Hamadan	Famenin1	Iran	48.97	35.12	12				
Hamadan	Famenin2	Iran	48.98	35.11	2		7		
Hamadan	Hamadan	Iran	48.54	34.81	6	1			
Hamadan	JahanAbad	Iran	48.97	35.11	1	1			
Hamadan	Saamen	Iran	48.70	34.20		18			
Hormozgan	Bandarabbas1	Iran	56.25	27.20		3		1	

Hormozgan	Bandarabbas2	Iran	56.31	27.19	1				
Hormozgan	Banu	Iran	56.50	27.33	1	6			
Hormozgan	Deh barez	Iran	57.17	27.48			5	4	
Hormozgan	Kalat	Iran	56.34	27.32	2				
Ilam	Dareh shahr	Iran	47.38	33.14	9				
Ilam	Ilam	Iran	46.43	33.64	1				
Ilam	Shirvan cherdavel	Iran	46.46	33.85	1				
Isfahan	Dowlat Abad	Iran	51.61	32.74		6			
Isfahan	Khorzough	Iran	51.60	32.71		3		1	
Isfahan	Mahmoud Abad	Iran	51.57	32.77		9			
Isfahan	Shahin Shahr	Iran	51.54	32.82		2			
Isfahan	Shahrak Montazeri	Iran	51.58	32.79		8			
Kerman	Kerman	Iran	57.07	30.28				1	
Kermanshah	Javanroud	Iran	46.49	34.81	4				
Kermanshah	Kermanshah	Iran	47.07	34.31	4				
Khorasan-e jonoubi	Birdjand	Iran	59.21	32.87				2	
Khorasan-e jonoubi	Noghabe	Iran	59.06	33.87				5	
Khorasan-e razavi	Mashhad	Iran	59.5	36.39				2	

Khorasan-e shomali	Gouy-e Nik	Iran	57.09	37.94					3	
Khorasan-e shomali	Shirvan	Iran	57.93	37.53					1	
Khuzestan	Ahvaz	Iran	48.64	31.31		9		2		
Khuzestan	GamishAbad	Iran	48.66	31.26		2				
Khuzestan	Hamidieh	Iran	48.49	31.46	1	9				
Khuzestan	Shavour	Iran	48.46	31.84		11				
Kordestan	Sanandaj	Iran	47.00	35.32		2				
Markazi	Ashtian	Iran	50.00	34.55		3				
Markazi	Khomein	Iran	50.05	33.63			1			
Markazi	Mahalat	Iran	50.50	33.88			2			
Markazi	Saveh	Iran	50.33	35.02			5			
Markazi	Tafresh	Iran	49.99	34.62			6			
Mazandaran	Chalus	Iran	51.42	36.66						1
Qazvin	Abyek	Iran	50.53	36.05			1			
Qazvin	Qazvin	Iran	50.00	36.27		3				
Qazvin	Takestan	Iran	49.70	36.07		2	2			
Sistan va Balouchestan	Asadabad	Iran	60.72	27.22					15	
Sistan va Balouchestan	Bampur	Iran	60.45	27.20			3		13	

Sistan va Balouchestan	Chahnime	Iran	61.67	31.25			1
Sistan va Balouchestan	Iranshahr1	Iran	60.68	27.17			11
Sistan va Balouchestan	Iranshahr2	Iran	60.58	27.20			2
Sistan va Balouchestan	Katamak	Iran	61.65	31.25			1
Sistan va Balouchestan	Khane-Koute	Iran	61.64	31.25			3
Sistan va Balouchestan	Kombaki	Iran	59.20	25.70			2
Sistan va Balouchestan	Negur	Iran	61.90	25.50			20
Sistan va Balouchestan	Nikshahr	Iran	60.22	26.22			7
Sistan va Balouchestan	Now Bandian	Iran	61.18	25.50		1	3
Sistan va Balouchestan	Rikapout	Iran	60.53	27.20			1
Sistan va Balouchestan	Takht-e-edalate	Iran	61.72	31.32			1
Sistan va Balouchestan	Tchabahar1	Iran	60.63	25.37	7		3

Sistan va Balouchestan	Tchabahar2	Iran	60.63	25.30	4		1	3
Sistan va Balouchestan	Zabol	Iran	61.62	31.25				3
Tehran	Tehran	Iran	51.43	35.81		9		
Yazd	Eslamie	Iran	54.10	31.73		6		
Yazd	FakhrAbad	Iran	54.25	31.61		7		
Zanjan	Abhar	Iran	49.22	36.14		2		
Zanjan	Zanjan	Iran	48.48	36.67	1	18		

Sup. Table 5 – f3 statistics for all combinations of populations considering the Caucasus (CAU) and the Northwest Iranian (NWI) populations as targets

f3(A; B, C)	Target	Source1	Source2	f3	std. error	Z-Score
CAU; AFG,MWI	CAU	AFG	MWI	-0.02351	0.000266	-88.3447
CAU; AFG,NWI	CAU	AFG	NWI	-0.02301	0.000254	-90.6009
CAU; SWI,AFG	CAU	SWI	AFG	-0.02291	0.000262	-87.2982
CAU; AFG,CEI	CAU	AFG	CEI	-0.00744	0.000182	-40.9542
NWI; SWI,CAU	NWI	SWI	CAU	-0.00593	0.000271	-21.8531
NWI; SWI,AFG	NWI	SWI	AFG	-0.00583	0.000309	-18.8921
NWI; AFG,MWI	NWI	AFG	MWI	-0.00308	0.000328	-9.3977
NWI; CAU,MWI	NWI	CAU	MWI	-0.00258	0.00029	-8.89157
NWI; SWI,CEI	NWI	SWI	CEI	-0.00077	0.000212	-3.61722

Supplementary Text

Sampling relatedness

We used the relatedness2 option of vcftools to assess pairwise individual relatedness among all mice in the dataset, using the KING method (Danecek et al. 2011). This analysis is based on GATK called genotypes and the 90% tranche PASS-filtered SNPs. We restricted the dataset to only include autosomal SNPs, thinned to 1 SNP every 25kb. We also removed sites that had more than 20% missing data, expected ranges of kinship coefficients ('Phi') are >0.354 for duplicate samples/monozygotic twins, [0.177–0.354] for 1st degree relatives, [0.0884–0.177] for 2nd degree relative, [0.0442–0.0884] for 3rd degree relatives and <0.0442 for unrelated). No duplicate samples were detected. Most related animals were found in the populations from Iran and Kazakhstan. In the case of the Iranian population the increased relatedness within the sample can be explained by the fact that some breeding adults were used in multiple cross. The relatedness observed in the population from Kazakhstan is best explained by the fact that mice were collected in close proximity, rather than over a larger regional scale(Harr et al. 2016). We only consider first and second degree relatedness relevant here and related individuals were removed from the final analysed dataset.

t- haplotype individuals' identification

The *t*-haplotype is a complex set of 4 inversions, comprising a 30–40 Mbp long region of chromosome 17. It causes transmission ratio distortion, and heterozygous t-haplotype carriers tend to predominantly transmit the *t*-haplotype carrying chromosome to their offspring. Despite their massive transmission advantage, *t*-haplotype carrying individuals are rare in natural populations of mice but have been found in all recognized subspecies. We have identified our *t*-carriers individuals by leverage on a PRDM9 gene region PCA analysis (Sup. Fig. 13). The PCA segregates *t*-carriers from non-carriers rather than by subspecies. These results agreed with what was previously described for the northern Indian individuals (confirmed by PCR on *Bettina et al.* (2016). We have identified approximately a *t*-carrier individual per population, except on the MWI and NWI *domesticus* populations.

References

Albrechtova, Jana, Tomàs Albrecht, Stuart J. E. Baird, Milos Macholan, Geir Rudolfsen, Pavel Munclinger, Priscilla K. Tucker, and Jaroslav Pialek. 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. doi: 10.1098/rspb.2012.1802.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. 'Fast Model-Based Estimation of Ancestry in Unrelated Individuals'. Genome Research 19 (9): 1655–64. https://doi.org/10.1101/gr.094052.109.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. 'Basic Local Alignment Search Tool'. *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Auwera, Geraldine A. Van der, and Brian D. O'Connor. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media, Incorporated.

Barnett, Derek W., Erik K. Garrison, Aaron R. Quinlan, Michael P. Strömberg, and Gabor T. Marth. 2011. 'BamTools: A C++ API and Toolkit for Analyzing and Managing BAM Files'. Bioinformatics 27 (12): 1691–92. https://doi.org/10.1093/bioinformatics/btr174.

Bimova, B. V., M. Macholan, S. J. Baird, P. Munclinger, P. Dufkova, C. M. Laukaitis, R. C. Karn, K. Luzynski, P. K. Tucker, and J. Pialek. 2011. "Reinforcement selection acting on the European house mouse hybrid zone." *Mol Ecol* 20 (11):2403-24. doi: 10.1111/j.1365-294X.2011.05106.x.

Bishop, C. E. , P. Boursot, F. Bonhomme, and D. Hatat. 1985. "Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome." *Nature* 315:70-72.

Boissinot, S, and P Boursot. 1997. "Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: selection on the Y chromosome?" *Genetics* 146:1019-1033.

Bonhomme, François, Annie Orth, Thomas Cucchi, Hassan Rajabi-Maham, Josette Catalan, Pierre Boursot, Jean-Christophe Auffray, and Janice Britton-Davidian. 2011. "Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization." *Proceedings of the Royal Society B: Biological Sciences* 278 (1708):1034-1043. doi: 10.1098/rspb.2010.1228.

Britton Davidian, J., Fabienne Fel-Clair, J. Lopez, Paul Alibert, and P Boursot. 2005. "Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids." Biol J Linn Soc 84 (3):379-393.

Chang, Christopher C, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. 2015. 'Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets'. GigaScience 4 (1): s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8.

Chu, Chong, Rasmus Nielsen, and Yufeng Wu. 2016. 'REPdenovo: Inferring De Novo Repeat Motifs from Short Sequence Reads'. Edited by Christophe Antoniewski. PLOS ONE 11 (3): e0150719. https://doi.org/10.1371/journal.pone.0150719.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. 'The Variant Call Format and VCFtools'. Bioinformatics 27 (15): 2156–58. https://doi.org/10.1093/bioinformatics/btr330.

Dierckxsens, Nicolas, Patrick Mardulyn, and Guillaume Smits. 2016. 'NOVOPlasty: De Novo Assembly of Organelle Genomes from Whole Genome Data'. Nucleic Acids Research, October, gkw955. https://doi.org/10.1093/nar/gkw955.

Duvaux, Ludovic, Khalid Belkhir, Matthieu Boulesteix, and Pierre Boursot. 2011. "Isolation and gene flow: inferring the speciation history of European house mice." Molecular Ecology 20 (24):5248-5264. doi: 10.1111/j.1365-294X.2011.05343.x.

Evanno, G., S. Regnaut, and J. Goudet. 2005. 'Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study'. Molecular Ecology 14 (8): 2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. 'CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data'. Bioinformatics 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

'GRCm38 - Mm10 - Genome - Assembly - NCBI'. n.d. Accessed 19 January 2022. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20/.

Guan, Yongtao. 2014. 'Detecting Structure of Haplotypes and Local Ancestry'. Genetics 196 (3): 625–42. https://doi.org/10.1534/genetics.113.160697.

Harr, Bettina, Emre Karakoc, Rafik Neme, Meike Teschke, Christine Pfeifle, Željka Pezer, Hiba Babiker, et al. 2016. 'Genomic Resources for Wild Populations of the House Mouse, Mus Musculus and Its Close Relative Mus Spretus'. Scientific Data 3 (1): 1–14. https://doi.org/10.1038/sdata.2016.75.

Jones, Eleanor P., Jeroen Van Der Kooij, Roar Solheim, and Jeremy B. Searle. 2010. "Norwegian house mice (Mus musculus musculus/domesticus): distributions, routes of colonization and patterns of hybridization." Molecular Ecology 19 (23):5252-5264. doi: 10.1111/j.1365-294X.2010.04874.x.

Kopelman, Naama M., Jonathan Mayzel, Mattias Jakobsson, Noah A. Rosenberg, and Itay Mayrose. 2015. 'Clumpak: A Program for Identifying Clustering Modes and Packaging Population Structure Inferences across K'. Molecular Ecology Resources 15 (5): 1179–91. https://doi.org/10.1111/1755-0998.12387.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms'. Molecular Biology and Evolution 35 (6): 1547–49. https://doi.org/10.1093/molbev/msy096.

Langmead, Ben, and Steven L Salzberg. 2012. 'Fast Gapped-Read Alignment with Bowtie 2'. Nature Methods 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Li, Heng. 2013. 'Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM'. ArXiv:1303.3997 [q-Bio], May. http://arxiv.org/abs/1303.3997.

Macholan, M., S.J. Baird, P. Munclinger, P. Dufkova, B. Bimova, and J. Pialek. 2008. "Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone?" BMC Evolutionary Biology 8 (1):271-.

Macholan, Milos, Stuart J. E. Baird, Alena Fornuskova, Iva Martincov, Pavel Rubik, Ludovit Dureje, Emanuel Heitlinger, and Jaroslav Pialek. 2019. "Widespread introgression of the Mus musculus musculus Y chromosome in Central Europe." bioRxiv:2019.12.23.887471. doi: 10.1101/2019.12.23.887471.

Macholan, Milos, Pavel Munclinger, Monika Sugerkova, Petra Dufkova, Barbora Bimova, Eva Bozikova, Jan Zima, and Jaroslav Pialek. 2007. "GENETIC ANALYSIS OF AUTOSOMAL AND X-LINKED MARKERS ACROSS A MOUSE HYBRID ZONE." Evolution 61 (4):746-771. doi: doi:10.1111/j.1558-5646.2007.00065.x.

Morgan, A. P., and F. Pardo Manuel de Villena. 2017. "Sequence and structural diversity of mouse Y chromosomes." Mol Biol Evol 34 (12):3186-3204. doi: 10.1093/molbev/msx250.

'Mouse Genomes Project - Wellcome Sanger Institute'. n.d. Accessed 19 January 2022. https://www.sanger.ac.uk/data/mouse-genomes-project/.

Nagamine, C.M., Y. Nishioka, K. Moriwaki, P. Boursot, F. Bonhomme, and Y.F.C. Lau. 1992. "The musculus-type Y chromosome of the laboratory mouse is of Asian origin." Mammalian Genome 3:84-91.

Orth, A, E. Lyapunova, A. Kandaurov, S. Boissinot, P. Boursot, N. Vorontsov, and F. Bonhomme. 1996. "L'espèce polytypique Mus musculus en Transcaucasie." Comptes Rendus de l'Académie des Sciences, Paris 319:435-441.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich. 2012. 'Ancient Admixture in Human History'. Genetics 192 (3): 1065–93. https://doi.org/10.1534/genetics.112.145037.

'Picard Tools - By Broad Institute'. n.d. Accessed 19 January 2022. https://broadinstitute.github.io/picard/.

Pickrell, Joseph K., Nick Patterson, Chiara Barbieri, Falko Berthold, Linda Gerlach, Tom Güldemann, Blesswell Kure, et al. 2012. 'The Genetic Prehistory of Southern Africa'. Nature Communications 3 (October): 1143. https://doi.org/10.1038/ncomms2140.

Pickrell, Joseph K., and Jonathan K. Pritchard. 2012. 'Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data'. Edited by Hua Tang. PLoS Genetics 8 (11): e1002967. https://doi.org/10.1371/journal.pgen.1002967.

Rajabi-Maham, Hassan, Annie Orth, Roohollah Siahsarvie, Pierre Boursot, Jamshid Darvish, and François Bonhomme. 2012. "The south-eastern house mouse Mus musculus castaneus (Rodentia: Muridae) is a polytypic subspecies." Biological Journal of the Linnean Society 107 (2):295-306. doi: 10.1111/j.1095-8312.2012.01957.x.

Raufaste, N., A. Orth, K. Belkhir, D. Senet, C. Smadja, S. J. E. Baird, F. Bonhomme, B. Dod, and P. Boursot. 2005. "Inferences of selection and migration in the Danish house mouse hybrid zone." Biological Journal of the Linnean Society 84 (3):593-616. doi: DOI 10.1111/j.1095-8312.2005.00457.x.

Reich, David, Kumarasamy Thangaraj, Nick Patterson, Alkes L. Price, and Lalji Singh. 2009. 'Reconstructing Indian Population History'. Nature 461 (7263): 489–94. https://doi.org/10.1038/nature08365.

Reimand, Juri, Raivo Kolde, and Tambet Arak. 2019. GProfileR: Interface to the 'g:Profiler' Toolkit (version 0.7.0). https://CRAN.R-project.org/package=gProfileR.

Rosenberg, Noah A. 2004. 'Distruct: A Program for the Graphical Display of Population Structure'. Molecular Ecology Notes 4 (1): 137–38. https://doi.org/10.1046/j.1471-8286.2003.00566.x.

Smadja, C., and G. Ganem. 2005. "Asymmetrical reproductive character displacement in the house mouse." Journal of Evolutionary Biology 18 (6):1485-1493.

Chapter IV

Final considerations and future prospects

The use of laboratory-based model organisms, as beans, mice or drosophila, has allowed a rapid progress on the understanding of major evolutionary principles (e.g. Haldane's rule or Bateson-Dobzhansky-Muller Incompatibilities model), but provide limited insights into the underpinnings of speciation in nature, and the mechanisms under which species can continue to exchange genetic variation via introgression despite the establishment of partial reproductive isolation. The advent of next-generation sequencing has revolutionized our ability to sample genetic variation from natural populations, and the establishment of models to understand organismal diversification, such as several iconic adaptive radiation systems, like cichlids, anole lizards, hares, or sticklebacks. With the increased use of genomics tools, the sampling of biological systems that can provide a more complete view of the diversification mechanisms is expected in the near future. The work developed in this thesis sought to contribute for the general debate on the importance and complexity of admixture during speciation. The conclusions drawn were based on the study of natural populations of hares (*Lepus spp.*) and house mice (*Mus musculus*). Both biological systems have documented cases of hybridization between closely related taxa, which allowed electing specific models supported by previously published results and/or preliminary inferences made in an exploratory phase of the work.

4.1- Standard analysis of big genomic datasets

The advances in high throughput sequencing technologies have revolutionized our ability to collect and analyse massive amounts of genetic information. The affordable production costs together with the development of user-friendly software/pipelines for data treatment and analysis has enabled a rapid and uncontrolled expansion of population genomic studies. Such an impressive data production rate and publication-pressure causes huge challenges for a careful data analysis and following interpretation. Blindly relying on the outputs of these standard tools can render an accumulation of errors from the initial inferences to the final study conclusions. Unfortunately, this situation is particularly critical on the initial exploratory analysis, where genetic clustering algorithms such as STRUCTURE or ADMIXTURE have been extensively used to characterize individuals and populations. Indeed, STRUCTURE-like software's alongside with Principal Component Analysis (PCA) have become a quasi-universal method to analyse genetic data. The numerous successful examples of population structure inference (e.g. Rosenberg et al. 2001; Tishkoff et al. 2009; Rosenberg et al. 2002) created a flow of result over-interpretation (Lawson, van Dorp, and Falush 2018). The problem is even more pronounced in the case of large and diverse datasets, where the risk of over or under interpreting results is more likely. During the development of this work, I have faced many challenges, but the definition of a demographic history of divergence on the mouse system was/is the biggest. The uncertainty about lineage status on the Iranian plateau created a challenge difficult to reconcile across analyses, as the initial unsupervised ADMIXTURE and PCA analysis suggest the presence of a new subspecies (apart from the three well accepted ones). As most of the subsequent methods required user-defined parameters, everything was initially based on this idea of four completely independent subspecies. Due to some inconsistent outcomes based on this assumption, we felt the need of adding further exploratory methods such as Admixture Graphs and D_{xy}, F_{st} genome wide distributions that allow us to realize that the central Iran population, is indeed a new lineage with its own gene pool but which originally had resulted from secondary admixture between Iranian *domesticus* and a population related to *musculus* (Publication IV). This example stresses the need of a careful exploration of all range of possible hypothesis.

To summarize, the application of previous cited standard exploratory tools is very useful, especially because they are easy to implement, fast and usually supported by a vast bibliography and numerous examples. However, we stress the fact that this kind of

methods should be wisely used and complemented with less standard approaches. It is fundamental to define a biologically relevant question adjusted in the light of each new analysis. The interpretation of results should be made with scepticism and scrutiny, and based on carefully defined hypotheses (Johri et al. 2021).

4.2- Models to study speciation with gene flow

Hares and mice comprise a vast number of recently diverged species with current or ancestral evidence of gene flow among them. These models have been extensively studied in the last decades and present striking similarities that make them appropriate to study the genomic underpinnings of speciation and gene flow (Ferreira et al. 2021 and Duvaux et al. 2011). Each genus is composed by species (or subspecies) with distinct degrees of divergence, gene flow and ecological differentiation, and thus constitute an exceptional framework for a comparative approach. The parallel study of species or subspecies pairs of hares and mice provided us with the ability to compare independent realizations of the evolutionary process and to complement the analysis of speciation at two timescales of divergence and gene flow (both are more recent in mice) (Wolf and Ellegren 2016). The limited extent of this work prevented extracting the maximum potential from the combined use of these two systems, but greatly contributed to our understanding of their biology and disclosed new natural laboratories to study the contribution and complexity of gene flow to speciation. The joint analysis of both systems promises to be a rich contribution to the ongoing general debate on the genomics of speciation and hybridization.

4.2-1. Advancing genomic resources to reinforce hares as model systems to evolutionary and conservation studies

The genomic resources generated for the hare system (Chapter II) provide important assets to study relevant questions related with the biology and evolution of these organisms. We have characterized and annotated three hare transcriptomes (Publications I, II and III), detected polymorphism in two populations of mountain hare and identified putative diagnostic sites between mountain and European hares (Publication I). These resources will be useful for a variety of studies, particularly in the characterization of genetic diversity in mountain hare and the diagnostic sites database can be used to design tools to assess population status and monitor hybridization between species, particularly in places where the mountain and European hare have

overlap distributions, such as the Alps, Sweden or Ireland (Caravaggi et al. 2017; Acevedo et al. 2012). The two species are known to hybridize when in contact, resulting in some described genetic introgression (C G Thulin, Jaarola, and Tegelstrom 1997; Carl Gustaf Thulin, Fang, and Averianov 2006; Suchentrunk et al. 2005; Melo-Ferreira et al. 2009; Zachos et al. 2010), with potential effects on local adaptation (Hughes et al. 2011) and on their conservation status (Levänen et al. 2018). Yet, the degree of genetic exchanges in these contacts is still unclear, and the newly identified markers can make an important contribution to these quantifications in the future.

The *de novo* mountain hare reference genome generated on this thesis (Publication II), will allow future studies on hares to use a closer reference genome, which is specifically relevant in cases where hare-specific variation is needed. This resource was already incorporated in the analysis of local adaptation in mountain hare populations (Giska et al. 2022). Additionally, the developed strategy to achieve a chromosome level assembly was followed by Sjodin et al (2021) on the production of the American Pika reference genome assembly and can be valuable for similar frameworks. To understand to what extent reliance on an outgroup reference may have limited genomic inferences we briefly evaluate its use. The results suggest that the use of the alternative references does not impact heterozygosity tract patterns, and thus that approaches based on hare pseudoreferences has not limited evolutionary inference and genome scans on hares. To what extent this validation is true for methods based on different genomic features is unknown and needs to be accessed in the future. Additionally, and despite our chromosome level assembly has been produced by relying on the anchored of *de novo* short read assembly on the rabbit genome, it still allows to inspect some minor structural variants that can be useful for a better understanding of the evolutionary history of the species. The technical limitations found at the beginning of this thesis, when the use of long read sequencing was still limited, are now overcome by the cost-effective use of technologies such as Pacbio HiFi, that in combination with chromosomal conformation captures technologies (e.g. Hi-C) makes the production of *de novo* chromosome level assemblies a regular practice that should be extended to the system (see e.g. the Vertebrate Genome Project assembly pipeline (https://galaxyproject.github.io/training-material//topics/assembly/tutorials/vgp_genome_assembly/tutorial.html).

Despite the great progress made in the last decade, and the contribution made in this work, the hare system still lacks essential genomic resources usually available for

model species. Thus, it is essential to provide the system with more and better tools in order to explore all its potential. Particularly, it would be important to generate:

- species-specific chromosome level reference-genomes – this would be essential to infer species-structural differences, as well as species-specific variation.
- Species hybridization tools – to understand the impact of hybridization in species contact zones is fundamental to have resources to characterize its extent.
- Genetic maps – as discussed throughout this document, the recombination landscape fundamental to comprehend the hybridization dynamics.
- Population-based data – to reconstruct the history of gene flow as well as its magnitude and timing of hybridization.

Moreover, this work contributed to establish an underlying demographic model to explain the dynamics of ancient genetic exchanges affecting the current gene pool of hares from the Iberian Peninsula. Specifically, we show that the distribution of genetic variation across the range of the Iberian hare is compatible with a post-glacial northwards expansion, clarifying the demographic dynamics during the ancient hybridization events with mountain hares, before the latter went locally extinct from the region. This model set out the hypothesis that purely demographic processes may have promoted massive introgression through a process of allele surfing in the front of the range expansion of the Iberian hare during the species replacement process (Publication III). Indeed, this model served as the basis for simulations performed by Seixas et al. (2018), who showed that surfing of introgressed variants during the range replacement of the mountain hare by the Iberian hare explain general genomic patterns of introgression and, importantly, that this model can also explain the strong northwards gradients of mtDNA introgression from the mountain hare into the Iberian hare, including the quasi-fixation of introgressed haplotypes in northern populations. Altogether, these are important contributions to understand the importance of the underlying demographic models to understand patterns of introgression, which could be interpreted as resulting from selective processes, such as strong structure across populations, or high frequencies of introgressed variants. Furthermore, the demonstration that hares could play an important role in deciphering the effects of species contacts driven by environmental change and range shifts, make them a valuable model to evaluate the genomic impacts of recent climate change driven by anthropogenic actions. With the support of the unprecedented power of DNA sequencing that currently allows a cost-effective generation of modern but

also historical data, the use of a temporal sampling analysis could allow understand to what extent introgression can be involved in rapid adaptation and range replacements of species in face of present fast environmental changes.

4.2-2. The house mouse cradle of differentiation as a natural laboratory to study speciation with gene flow

The Iranian plateau has been proposed to be the house mouse cradle of differentiation, and it is currently the place in the world with the highest known diversity of mouse lineages within the same geographical area. Studies based on microsatellite loci confirmed the presence of the three well described subspecies and proposed the existence of two new genetic entities in central and southeast Iran (Hardouin et al. 2015). In this thesis, we confirmed the existence of a unique genomic entity, distinct from the three subspecies, in Central Iran (Publication IV). Remarkably this new entity was inferred to be the result of secondary admixture between populations related to current *domesticus* and *musculus*, which are known to form a tension zone in Europe. We hypothesize that this admixture could have been possible if the degree of reproductive isolation has increased after the “out-of-Iran”. In this case the resulting picture would be analogous to a ring species, with a gradual differentiation of both lineages from a common ancestor and increasing reproductive isolation along a gradient that starts in Iran (cradle of differentiation) and closes in the European hybrid zone (tension zone). Despite the closer affinity to the current *musculus* the central Iranian lineage presents a balanced parental contribution. The way in which species with balanced genomic mosaicism are able to escape genomic incompatibilities is poorly characterized, especially due to the few well-documented cases, mostly restricted to plants (e.g. wild sunflowers (Rieseberg 2003)), insects (e.g. tiger swallowtail butterflies (Kunte et al. 2011)) and birds (e.g. Italian sparrows (Elgvin et al. 2017), or golden-crowned manakins (Barrera-Guzmán et al. 2018)). To my knowledge, the central Iranian lineage represents the first genome-wide study of a mammal population with balanced genomic contributions from both parentals and is a living example of the hybridization potential to produce diversity. This outstanding case study may help later works to identify ancient incompatible loci interactions that were sorted in the “hybrid” lineage but could still contribute to the isolation between current *domesticus* and *musculus*, as well as incompatibilities that have arisen throughout the expansion of these two subspecies from the Iranian plateau. In addition, the central Iranian population inhabits a region marked

by a special topography and different climatic environments. The region is surrounded by natural geographical barriers, as the Central Playa (Esfahan Plain) to the north, the Lut desert to the east known to be one of the most arid areas of the world, and the Zagros mountains at southwest. Although nowadays a harsh environment prevails, it is believed that during the glacial periods the region was humid and suitable for acting as a refugium for several species (Shad and Darvish 2018). The region's difficult access may have limited the continued contacts of central Iranian population with surrounding populations across climate oscillations periods and may gradually contribute to its confinement and genetic differentiation from its close relatives. On the other hand, the current harsh environment may have involved some degree of specialization and adaptation. Whether this has left evidence on the genome variation of the population is a question to which we could not answer, as properly characterizing selection signals would require a larger sample size. Unfortunately, the geo-political situation in the region makes extra sampling a very challenging task, but still vital for a better understanding of this and related questions.

We further extended the exploration of the Iranian plateau to areas of potential contact and admixture between *musculus, domesticus* and this newly described Central Iranian lineage. We documented three-way admixture in two populations from North-Western Iran, and massive *musculus*-related Y chromosome introgression in the *domesticus* population (Publication V). We find co-introgression of male fertility genes along the genome and a correlation between copy numbers of Y and X ampliconic families (*Sly*/*Slx*). Whether this has implications on hybrid fertility still needs to be tested and, if true, functionally validated. The complete introgression of a Y chromosome related to the *musculus* lineage, presumably due to a selective advantage, resembles the situation on the European hybrid zone where a unidirectional Y introgression from *musculus* into *domesticus* (Macholán et al. 2019; 2008) is also detected, suggesting that the *musculus* Y can be advantageous in a hybrid context. This invasive success of the *musculus*-related Y in hybrid context has been linked to its transmission distortion power associated high *Sly* copy number (Morgan and Pardo Manuel de Villena 2017). On the contrary we argue that this success may be related with its contribution to rescue infertility in hybrids (see Albrechtová et al. 2012), therefore explaining its occurrence in all known admixed house mouse populations. This remarkable similarity between the European and Iranian (here described) contact zones provides a powerful replicate of the Y introgression phenomenon and promises to be a natural laboratory of excellence to help understand the process behind this asymmetrical invasion. Regardless of the

great number of resources available to the house mouse system, the sex-chromosomes analysis is still very constrained by the available reference genome. The house mouse reference genome is an “hybrid” between a major *domesticus* contribution for the assembly of the autosomes and X-chromosome and a *musculus*-derived Y chromosome assembly (Soh et al. 2014). This difference in the source of each genomic component adds some noise to the analysis of multicopy regions, as the uncertainty in the sex chromosome calls is amplified by the homology between the two sex-chromosomes. In order to reduce this bias effect, we produced a sex-specific *de novo* repeat database. The strategy allows us to control for sex-bias but not the lineage-specific effect. The production and comparison of sex-lineage-specific assemblies could mitigate the problem, or at least be used to identify relevant regions involved in the proposed evolutionary dynamics. Moreover, due to time constraints we did not deeply investigate the significant *domesticus* contribution to the *musculus* population around Caucasus. This population genetic diversity has been a matter of debate due to its distinctive mitochondrial genetic composition (Tembotova et al. 2021) and mating behaviour (Ambaryan and Kotenkova 2020). Whether both observations are related or independent of the described *domesticus* contribution still needs to be explored.

Despite the progresses made in this study, the Iranian plateau and surrounding areas promises to contain additional lineages, particularly in southeast Iran (as suggested by Hardouin et al. 2015 and Hamid et al. 2016), and tension zones (e.g. on the northeast between *castaneus* and *musculus*) that may help to understand the species history of divergence and degree of isolation between populations. Preliminary unpublished results suggest that the southeast Iran (the Zabol-Baluch region) harbours a different Y lineage and some evidence of admixture with central Iranian and *castaneus* lineages. Whether this is a new lineage (as suggested by Hardouin et al. 2015 and Hamid et al. 2016) or represents the result of the admixture of lineages still needs to be investigated. A vast genome-wide population survey across the plateau, supported by our initial findings, may help to clarify many aspects of the models described above that remain uncertain, such as the timeframe of speciation events, the existence of other intermediate lineages as suggested by the ring species concept or the extent of sex chromosomes invasion. Although we need further multidisciplinary approaches (from genomics to behaviour and physiology) to clarify some of the results achieved in this thesis, we can already attest the house mouse living in the Iranian plateau as an outstanding model to study divergence with gene-flow.

4.3- Final note - science outreach

Science education provides students with valuable skills on critical thinking, which is fundamental for their education and life. Yet, science outreach is often absent from academic activities and doctoral programmes. Bringing science to students is an efficient way of promoting science literacy, while providing a clear understanding on current avenues of research. During the last years I co-coordinated a scientific programme on ecology and evolution that has been developed in the framework of a Portuguese high school community. This programme was composed of several year-long projects where the students were challenged to go through a conventional framework of a research project, from its critical conception to the final public release of the results. The project contributed with several scientific and applied outputs that went from management and control of species reintroduction in a public urban park, water sources quality control, to the description of the hare immune related gene polymorphism repertoire. This last work culminated with a scientific poster accepted to be presented at the "6th World Lagomorph Conference" at Montpellier in 2022.

It is fundamental that each scientist can contribute to the scientific literacy of the society we are inserted in. It is also particularly important that this kind of actions are recognized as fundamental by doctoral schools and not seen as a waste of "publication-time". The world needs to change towards a better scientific knowledge, and we need to make the first move!

4.4- References

Acevedo, Pelayo, Alberto Jiménez-Valverde, José Melo-Ferreira, Raimundo Real, and Paulo Célio Alves. 2012. 'Parapatric Species and the Implications for Climate Change Studies: A Case Study on Hares in Europe'. *Global Change Biology* 18 (5): 1509–19. https://doi.org/10.1111/j.1365-2486.2012.02655.x.

Albrechtová, Jana, Tomáš Albrecht, Stuart J. E. Baird, Miloš Macholán, Geir Rudolfsen, Pavel Munclinger, Priscilla K. Tucker, and Jaroslav Piálek. 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): 4803–10. https://doi.org/10.1098/rspb.2012.1802.

Ambaryan, A.V., and E.V. Kotenkova. 2020. 'Mating Behavior of House Mice of Trans-Caucasian Hybrid Zone: A Comparative Study with Parent Species Mus Musculus'. *Russian Journal of Theriology* 19 (2): 149–60. https://doi.org/10.15298/rusjtheriol.19.2.05.

Barrera-Guzmán, Alfredo O., Alexandre Aleixo, Matthew D. Shawkey, and Jason T. Weir. 2018. 'Hybrid Speciation Leads to Novel Male Secondary Sexual Ornamentation of an Amazonian Bird'. *Proceedings of the National Academy of Sciences* 115 (2): E218–25. https://doi.org/10.1073/pnas.1717319115.

Caravaggi, Anthony, Katie Leach, Francesco Santilli, Jukka Rintala, Pekka Helle, Juha Tiainen, Francesco Bisi, Adriano Martinoli, W. Ian Montgomery, and Neil Reid. 2017. 'Niche Overlap of Mountain Hare Subspecies and the Vulnerability of Their Ranges to Invasion by the European Hare; the (Bad) Luck of the Irish'. *Biological Invasions* 19 (2): 655–74. https://doi.org/10.1007/s10530-016-1330-z.

Duvaux, Ludovic, Khalid Belkhir, Matthieu Boulesteix, and Pierre Boursot. 2011. 'Isolation and Gene Flow: Inferring the Speciation History of European House Mice: HOUSE MOUSE SPECIATION HISTORY'. Molecular Ecology 20 (24): 5248–64. https://doi.org/10.1111/j.1365-294X.2011.05343.x.

Elgvin, Tore O., Cassandra N. Trier, Ole K. Tørresen, Ingerid J. Hagen, Sigbjørn Lien, Alexander J. Nederbragt, Mark Ravinet, Henrik Jensen, and Glenn-Peter Sætre. 2017. 'The Genomic Mosaicism of Hybrid Speciation'. *Science Advances* 3 (6): e1602996. https://doi.org/10.1126/sciadv.1602996.

Ferreira, Mafalda S, Matthew R Jones, Colin M Callahan, Liliana Farelo, Zelalem Tolesa, Franz Suchentrunk, Pierre Boursot, et al. 2021. 'The Legacy of Recurrent Introgression during the Radiation of Hares'. Systematic Biology 70 (3): 593–607. https://doi.org/10.1093/sysbio/syaa088.

Giska, Iwona, João Pimenta, Liliana Farelo, Pierre Boursot, Klaus Hackländer, Hannes Jenny, Neil Reid, et al. 2022. 'The Evolutionary Pathways for Local Adaptation in Mountain Hares'. *Molecular Ecology* n/a (n/a). https://doi.org/10.1111/mec.16338.

Hamid, Haddadian Shad, Jamshid Darvish, Eskandar Rastegar-Pouyani, and Ahmad Mahmoudi. 2016. 'Subspecies Differentiation of the House Mouse Mus Musculus Linnaeus, 1758 in the Center and East of the Iranian Plateau and Afghanistan'. *Mammalia* 81 (2): 147–68. https://doi.org/10.1515/mammalia-2015-0041.

Hardouin, Emilie A, Annie Orth, Meike Teschke, Jamshid Darvish, Diethard Tautz, and François Bonhomme. 2015. 'Eurasian House Mouse (Mus Musculus L.) Differentiation at Microsatellite Loci Identifies the Iranian Plateau as a Phylogeographic Hotspot'. *BMC Evolutionary Biology* 15 (February). https://doi.org/10.1186/s12862-015-0306-4.

Hughes, Maria, Neil Reid, Ian Montgomery, and Paulo Prodoehl. 2011. 'Verification of Hybridisation between Introduced European and Native Irish Hares'. *Northern Ireland Environment Agency Research and Development SeriesNorthern Ireland Environment Agency Research and Development Series* 11–11.

Kunte, Krushnamegh, Cristina Shea, Matthew L. Aardema, J. Mark Scriber, Thomas E. Juenger, Lawrence E. Gilbert, and Marcus R. Kronforst. 2011. 'Sex Chromosome Mosaicism and Hybrid Speciation among Tiger Swallowtail Butterflies'. *PLOS Genetics* 7 (9): e1002274. https://doi.org/10.1371/journal.pgen.1002274.

Lawson, Daniel J., Lucy van Dorp, and Daniel Falush. 2018. 'A Tutorial on How Not to Over-Interpret STRUCTURE and ADMIXTURE Bar Plots'. *Nature Communications* 9 (1): 3258. https://doi.org/10.1038/s41467-018-05257-7.

Levänen, Riikka, Carl-Gustaf Thulin, Göran Spong, and Jaakko L. O. Pohjoismäki. 2018. 'Widespread Introgression of Mountain Hare Genes into Fennoscandian Brown Hare Populations'. *PLOS ONE* 13 (1): e0191790. https://doi.org/10.1371/journal.pone.0191790.

Macholán, Miloš, Stuart J. E. Baird, Alena Fornůsková, Iva Martincová, Pavel Rubík, Ľudovít Ďureje, Emanuel Heitlinger, and Jaroslav Piálek. 2019. 'Widespread Introgression of the *Mus Musculus Musculus* Y Chromosome in Central Europe'. Preprint. Evolutionary Biology. https://doi.org/10.1101/2019.12.23.887471.

Macholán, Miloš, Stuart JE Baird, Pavel Munclinger, Petra Dufková, Barbora Bímová, and Jaroslav Piálek. 2008. 'Genetic Conflict Outweighs Heterogametic Incompatibility in the Mouse Hybrid Zone?' *BMC Evolutionary Biology* 8 (1): 271. https://doi.org/10.1186/1471-2148-8-271.

Melo-Ferreira, J., P. C. Alves, H. Freitas, N. Ferrand, and P. Boursot. 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): 2643–58. https://doi.org/10.1111/j.1365-294X.2009.04221.x.

Rieseberg, L. H. 2003. 'Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization'. *Science* 301 (5637): 1211–16. https://doi.org/10.1126/science.1086949.

Rosenberg, N A, T Burke, K Elo, M W Feldman, P J Freidlin, M A Groenen, J Hillel, et al. 2001. 'Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes from 20 Chicken Breeds.' *Genetics* 159 (2): 699–713.

Rosenberg, Noah A, Jonathan K Pritchard, James L Weber, Howard M Cann, Kenneth K Kidd, Lev A Zhivotovsky, and Marcus W Feldman. 2002. 'Genetic Structure of Human Populations' 298: 6.

Seixas, Fernando A., Pierre Boursot, and José Melo-Ferreira. 2018. 'The Genomic Impact of Historical Hybridization with Massive Mitochondrial DNA Introgression'. *Genome Biology* 19 (1): 91. https://doi.org/10.1186/s13059-018-1471-8.

Shad, Hamid Haddadian, and Jamshid Darvish. 2018. 'Rodents Diversity in Protected Areas of Yazd Province, Central of Iran' 1 (1): 13.

Sjodin, Bryson M F, Kurt E Galbreath, Hayley C Lanier, and Michael A Russello. 2021. 'Chromosome-Level Reference Genome Assembly for the American Pika (Ochotona Princeps)'. *Journal of Heredity* 112 (6): 549–57. https://doi.org/10.1093/jhered/esab031.

Soh, Y. Q. Shirleen, Jessica Alföldi, Tatyana Pyntikova, Laura G. Brown, Tina Graves, Patrick J. Minx, Robert S. Fulton, et al. 2014. 'Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes'. *Cell* 159 (4): 800–813. https://doi.org/10.1016/j.cell.2014.09.052.

Suchentrunk, F, Z Mamuris, C Stamatis, H Ben Slimen, K Hackländer, G Haerer, and M Giacometti. 2005. 'Introgressive Hybridization in Wild Living Mountain Hares (L. Timidus Varronis) and Brown Hares (L. Europaeus) and Morphological Consequences'. *Mamm. Biol* 70 (Suppl.): 39–40.

Tembotova, Fatimat, Ekaterina Kuchinova, Albina Amshokova, and Ekaretina Kononenko. 2021. 'The Genetic Diversity of the Genus Mus (Linnaeus, 1758) in the Eastern Part of the North Caucasus'. *BMC Zoology* 6 (1): 29. https://doi.org/10.1186/s40850-021-00093-7.

Thulin, C G, M Jaarola, and H Tegelstrom. 1997. 'The Occurrence of Mountain Hare Mitochondrial DNA in Wild Brown Hares.' *Molecular Ecology* 6 (5): 463–67.

Thulin, Carl Gustaf, Meiying Fang, and Alexander O. Averianov. 2006. 'Introgression from Lepus Europaeus to L. Timidus in Russia Revealed by Mitochondrial Single Nucleotide Polymorphisms and Nuclear Microsatellites'. *Hereditas* 143 (2006): 68–76. https://doi.org/10.1111/j.2006.0018-0661.01952.x.

Tishkoff, Sarah A., Floyd A. Reed, Françoise R. Friedlaender, Christopher Ehret, Alessia Ranciaro, Alain Froment, Jibril B. Hirbo, et al. 2009. 'The Genetic Structure and History of Africans and African Americans'. *Science (New York, N.Y.)* 324 (5930): 1035–44. https://doi.org/10.1126/science.1172257.

Wolf JB, Ellegren H. Making sense of genomic islands of differentiation in light of speciation. Nat Rev Genet. 2017 Feb;18(2):87-100. doi: 10.1038/nrg.2016.133. Epub 2016 Nov 14. PMID: 27840429.

Zachos, F. E., H. Ben Slimen, K. Hackländer, M. Giacometti, and F. Suchentrunk. 2010. 'Regional Genetic in Situ Differentiation despite Phylogenetic Heterogeneity in Alpine Mountain Hares'. *Journal of Zoology* 282 (1): 47–53. https://doi.org/10.1111/j.1469-7998.2010.00710.x.