

# Bidirectional Introgression between *Mus musculus domesticus* and *Mus spretus*

Sarah E. Banker<sup>1,\*</sup>, François Bonhomme<sup>2</sup>, and Michael W. Nachman <sup>1</sup>

<sup>1</sup>Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, California, USA

<sup>2</sup>ISEM, Institut des Sciences de l'Evolution, CNRS, EPHE, IRD, Univ Montpellier, Montpellier, France

\*Corresponding author: E-mail: sarah\_banker@berkeley.edu.

Accepted: 22 December 2021

## Abstract

Introgressed variants from other species can be an important source of genetic variation because they may arise rapidly, can include multiple mutations on a single haplotype, and have often been pretested by selection in the species of origin. Although introgressed alleles are generally deleterious, several studies have reported introgression as the source of adaptive alleles—including the rodenticide-resistant variant of *Vkorc1* that introgressed from *Mus spretus* into European populations of *Mus musculus domesticus*. Here, we conducted bidirectional genome scans to characterize introgressed regions into one wild population of *M. spretus* from Spain and three wild populations of *M. m. domesticus* from France, Germany, and Iran. Despite the fact that these species show considerable intrinsic postzygotic reproductive isolation, introgression was observed in all individuals, including in the *M. musculus* reference genome (GRCm38). *Mus spretus* individuals had a greater proportion of introgression compared with *M. m. domesticus*, and within *M. m. domesticus*, the proportion of introgression decreased with geographic distance from the area of sympatry. Introgression was observed on all autosomes for both species, but not on the X-chromosome in *M. m. domesticus*, consistent with known X-linked hybrid sterility and inviability genes that have been mapped to the *M. spretus* X-chromosome. Tract lengths were generally short with a few outliers of up to 2.7 Mb. Interestingly, the longest introgressed tracts were in olfactory receptor regions, and introgressed tracts were significantly enriched for olfactory receptor genes in both species, suggesting that introgression may be a source of functional novelty even between species with high barriers to gene flow.

**Key words:** hybridization, adaptive introgression, genomics, olfactome, house mouse, population genomics.

## Significance

Although several recent studies have identified gene flow into a focal taxon, nearly all define one species as the “donor” of genetic material and the focal taxon as the “recipient.” Here, we describe bidirectional patterns of introgression between two species of *Mus* that exhibit substantial reproductive isolation. We find that introgression has occurred in both directions, but there is an asymmetry in the level of introgression, with *Mus spretus* containing more introgression overall. Notably, in both species, we found that introgressed tracts showed an enrichment of olfactory receptor genes. Our results highlight the importance of bidirectional introgression even between species that exhibit substantial barriers to gene flow.

## Introduction

The advent of whole-genome sequences for many organisms has led to a renewed interest in hybridization and its role in evolution (Taylor and Larson 2019). Although hybrid zones have been studied for decades (e.g., Endler 1977; Harrison

1990), the focus of these earlier studies, particularly in animals, has been on understanding barriers to gene flow. In contrast, the recent focus has been on the creative potential of hybridization, both for adaptive evolution and for

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

speciation. Indeed, genomic tracts that introgress between divergent taxa can be an important source of variation on which selection may act. Unlike new mutations, introgressed variants may be introduced rapidly and repeatedly, can include multiple mutations on a single haplotype, and have often been tested by selection in the taxon of origin. Although most introgressed genomic material is expected to be deleterious or neutral in the recipient species, some introgressed regions may be beneficial.

Whole-genome studies have now revealed many cases of hybridization and introgression between species of animals, including humans (Green et al. 2010), house mice (Liu et al. 2015), bears (Cahill et al. 2018), cichlid fish (Meier et al. 2017), sea bass (Duranton et al. 2020), *Heliconius* butterflies (Edelman et al. 2019), and many others (reviewed in Taylor and Larson 2019). In some cases, specific introgressed genes conferring an advantage have been identified. One of the first examples of adaptive introgression was in populations of house mice (*Mus musculus domesticus*) in Germany and Northern Spain which evolved resistance to the rodenticide warfarin. Mice that are resistant to warfarin carry an allele of *Vkorc1* that is highly diverged from other alleles seen in *M. m. domesticus*, but identical to *Vkorc1* haplotypes in the closely related species, *Mus spretus* (Song et al. 2011). Other notable examples of adaptive introgression include high elevation adaptation in humans conferred by an *Epas1* allele that came from Denisovans (Huerta-Sánchez et al. 2014), introgression of genes responsible for wing pattern mimicry in *Heliconius* butterflies (Dasmahapatra et al. 2012; Wallbank et al. 2016), and introgression from black-tailed jackrabbits to snowshoe hares of an *agouti* allele conferring seasonal color changes (Jones et al. 2018).

Despite the recent focus on hybridization, relatively few studies have addressed the directionality of gene flow between lineages or the geographic extent to which introgression extends across the range of a recipient species. If most introgressed loci are deleterious, they are expected to be eliminated by selection and thus are not expected to spread easily (e.g., Juric et al. 2016, Schumer et al. 2018). This prediction can be tested by sampling in areas of sympatry and areas of allopatry for species whose ranges partially overlap. House mice provide a good opportunity for studying introgression. *Mus musculus* consists of three subspecies that together have a nearly worldwide distribution: *M. m. musculus* is found in eastern Europe and northern Asia; *M. m. domesticus* is found in western Europe, the Middle East, North Africa and has recently been introduced to Australia, the Americas, and southern Africa in association with humans; and *Mus musculus castaneus* is found in southeast Asia (fig. 1) (Bonhomme and Searle 2012). These subspecies hybridize in several regions of secondary contact and are isolated by varying degrees of hybrid male sterility (reviewed in Phifer-Rixey and Nachman 2015). Introgression between subspecies has been studied extensively (e.g., Tucker et al. 1992; Teeter et al. 2008; Gerald et al. 2011; Staubach et al. 2012).

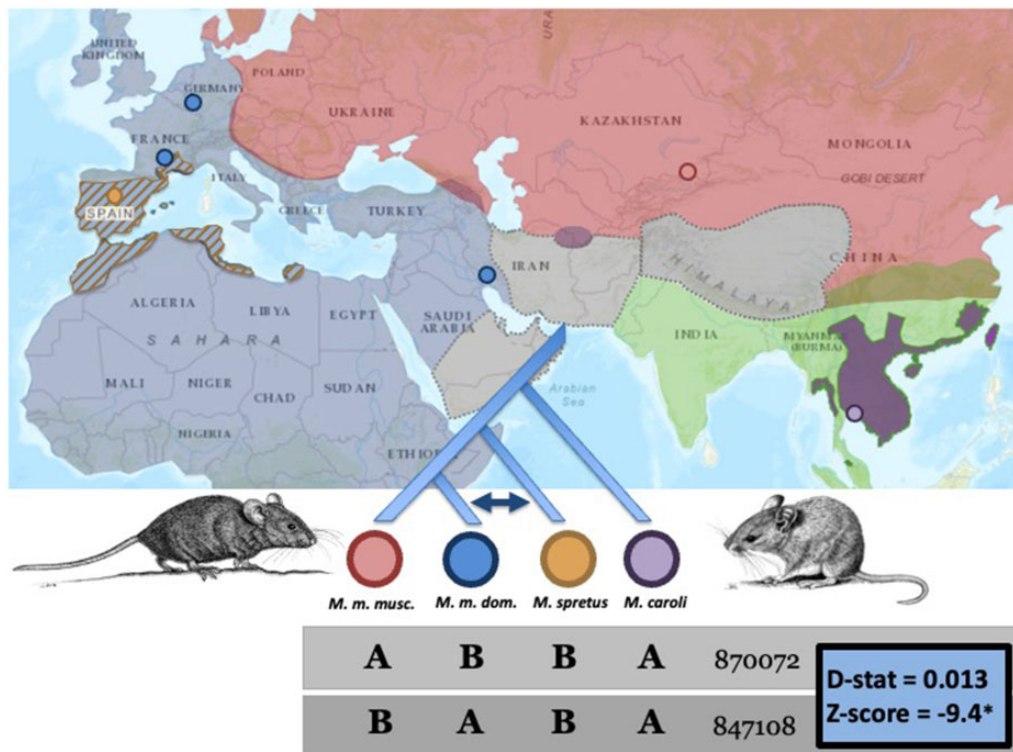
*Mus spretus* is a closely related species that is found in parts of North Africa and southern Europe; its range is wholly contained within part of the range of *M. m. domesticus* (fig. 1). *Mus spretus* is not commensal with humans and is believed to have diverged from *M. musculus* 1–3 Ma (She et al. 1990; Chevret et al. 2005). The first laboratory crosses between *M. musculus* and *M. spretus* were reported by Bonhomme et al. (1978). Reduced fitness of hybrids is a consequence of abnormal placental development due to disrupted imprinting (Zechner et al. 1996, 1997) as well as complete sterility of hybrid males (Bonhomme et al. 1978; Matsuda and Chapman 1992). Despite the sterility of hybrid males, viable and fertile F1 females can be recovered and backcrossed to either parent, and these crosses have been widely used to map specific genes in mice (reviewed in DeJager et al. 2009). No F1 hybrids between *M. musculus* and *M. spretus* have been observed in nature; however, Orth et al. (2002) reported reciprocal patterns of introgression between these species based on allozymes, mtDNA, and microsatellites. More recently, Song et al. (2011) showed that a *Vkorc1* allele underlying warfarin resistance in *M. m. domesticus* came from *M. spretus*. Liu et al. (2015) then used a genotyping array and found that introgression between *M. spretus* and *M. m. domesticus* was not limited to *Vkorc1*; other genomic regions were introgressed as well. Liu et al. (2015) included only a single representative *M. spretus* individual (reference genome SPRET-EIJ) and thus could not assess the directionality of introgression.

Here we extend the work of Orth et al. (2002) and Liu et al. (2015) by using whole-genome sequences of *M. spretus*, as well as two subspecies of *M. musculus* to assess the extent and direction of introgression between *M. spretus* and *M. musculus*. Specifically, we address the following questions: 1) How extensive is introgression between *M. spretus* and *M. m. domesticus* in natural populations? 2) Does gene flow occur in both directions, and if so, is the extent of introgression similar? 3) Does the amount of introgression decrease with increasing geographic distance from the area of sympatry? 4) Do regions that contain introgression have higher than average recombination rates, as expected if linked selection purges deleterious alleles more often in regions of low recombination? 5) How are introgressed tracts distributed across the genome, and what genes do they encode?

## Results

### Significant Introgression between *M. spretus* and *M. m. domesticus*

To identify historical gene flow between *M. spretus* and *M. m. domesticus*, we calculated the *D*-statistic using a population level ABBA–BABA test and assessed significance of the *D*-statistic with a *Z*-score. Populations included in the ABBA–BABA test consisted of *M. m. musculus* (8 Kazakhstan individuals), *M. m. domesticus* (24 Eurasian individuals), *M. spretus* (8 Spain



**Fig. 1.**—Geographic sampling and ABBA–BABA design. Solid-shaded regions indicate geographic ranges of *M. musculus* subspecies (modified from Phifer-Rixey and Nachman [2015] and Rajabi-Maham et al. [2008]) (blue = *M. m. domesticus*; red = *M. m. musculus*; green = *M. m. castaneus*; gray = further lineages and possible additional subspecies [Hardouin et al. 2015] and the Himalayan region). Areas of overlap represent regions of subspecific hybridization; range edges and areas of hybridization are approximate. *Mus spretus* is completely sympatric with *M. m. domesticus* in southern Europe and north Africa, range depicted in gold hatching. *Mus caroli* is completely sympatric with *M. m. castaneus* in southeast Asia, range depicted in purple. *Mus spretus* and *M. caroli* ranges were downloaded from IUCN Redlist. Colored dots on the map represent collection locality of reference genomes and wild populations used in this study (gold = Spanish *M. spretus* population; blue = French, German, and Iranian *M. m. domesticus* populations; red = Kazakhstani *M. m. musculus* population; purple = *M. caroli* reference genome). The phylogenetic tree shows the species tree, with the arrow indicating introgression between *M. m. domesticus* and *M. spretus*. Values indicate the genome total number of SNPs categorized as “ABBA” and “BABA,” with significance assessed with *D*-statistic and *Z*-score. Mouse drawings (Universitat de Valencia, ZooBot) show *M. m. domesticus* (left) and *M. spretus* (right).

individuals), and the outgroup *Mus caroli* (single Thailand individual) (fig. 1). *Mus musculus domesticus* had an excess number of shared derived single nucleotide polymorphisms (SNPs) with *M. spretus* compared with *M. m. musculus* (*D*-statistic = 0.013, *Z*-score =  $-9.4$ ,  $P < 0.001$ , fig. 1). The *D*-statistic was also significantly positive when each of the three *M. m. domesticus* population was tested individually (table 1). The ABBA–BABA test thus indicates that gene flow has occurred between *M. m. domesticus* and *M. spretus*. However, this test does not specify the directionality of gene flow between the two species, nor does it identify the genomic regions that are shared between the two species.

#### Introgression Is Bidirectional, with *M. spretus* Containing More Introgressed DNA Than *M. m. domesticus*

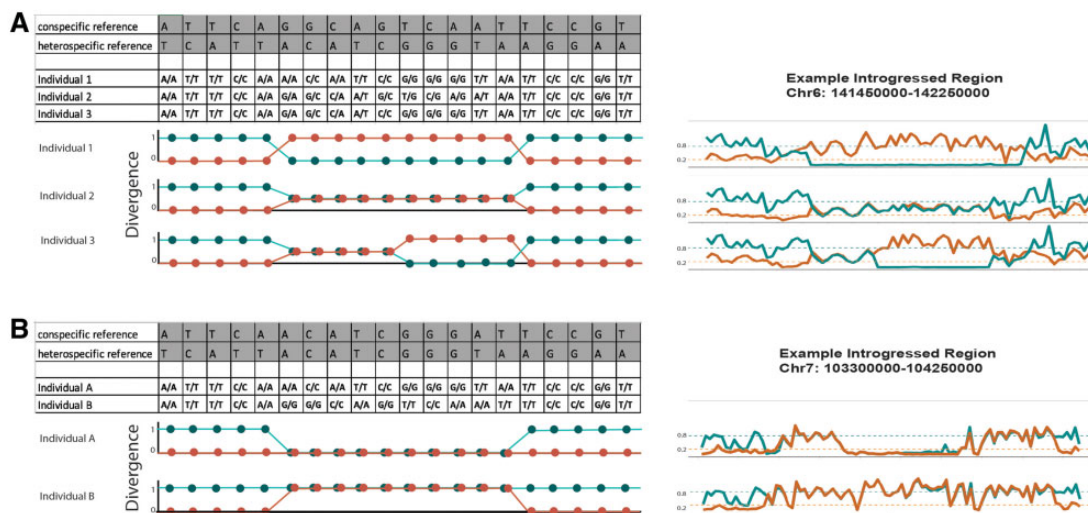
We identified introgressed genomic regions (IGRs) across the genome in 24 individuals of *M. m. domesticus* and 8 individuals of *M. spretus* using the comparison of absolute

nucleotide divergence ( $D_{xy}$ ) between each individual and the heterospecific and conspecific reference sequences (see fig. 2, Materials and Methods). We accounted for directionality by considering only sites that were invariant in the donor species. The genome-wide average heterospecific  $D_{xy}$  was approximately 1%, and the genome-wide average conspecific  $D_{xy}$  was approximately 0.2% (supplementary table 1, Supplementary Material online). A histogram of the difference in  $D_{xy}$  between the heterospecific and conspecific comparisons for different genomic regions revealed a mode close to the expected value of 0.8% (fig. 3). There was a second, smaller mode at 0%, and the distribution had a large asymmetrical tail with an excess of sites at or below  $-0.8\%$ . The presence of this smaller mode and the asymmetry provides clear evidence of introgression and cannot be explained by incomplete lineage sorting. The large asymmetry near  $-0.8\%$  represents homozygous introgressed regions, whereas the mode at 0 represents a combination of heterozygous introgressions and introgressions in the reference genome.

**Table 1**  
Summary Statistics of Genomic Introgression for Each Population

Population	<i>D</i> -Statistic (Z-Score)	Introgressed Windows (Collapsed by Population)					Percent Genome Introgressed		
		Total no.	Average IGR length (kb)	No. of IGR >100 kb	No. of IGR >500 kb	Average pop- ulation frequency	Average per individual (%)	Total (col- lapsed by population) (%)	
<i>M. spretus</i> Spain	N/A	139	114.13	58	4	0.15	0.078	0.58	
<i>M. m. domesticus</i> France	0.018 (12.07)	95	56.42	18	1	0.33	0.069	0.22	
<i>M. m. domesticus</i> Germany	0.010 (11.50)	89	61.57	25	1	0.29	0.068	0.20	
<i>M. m. domesticus</i> Iran	0.009 (11.46)	97	22.27	4	0	0.45	0.036	0.08	

N/A, not applicable.

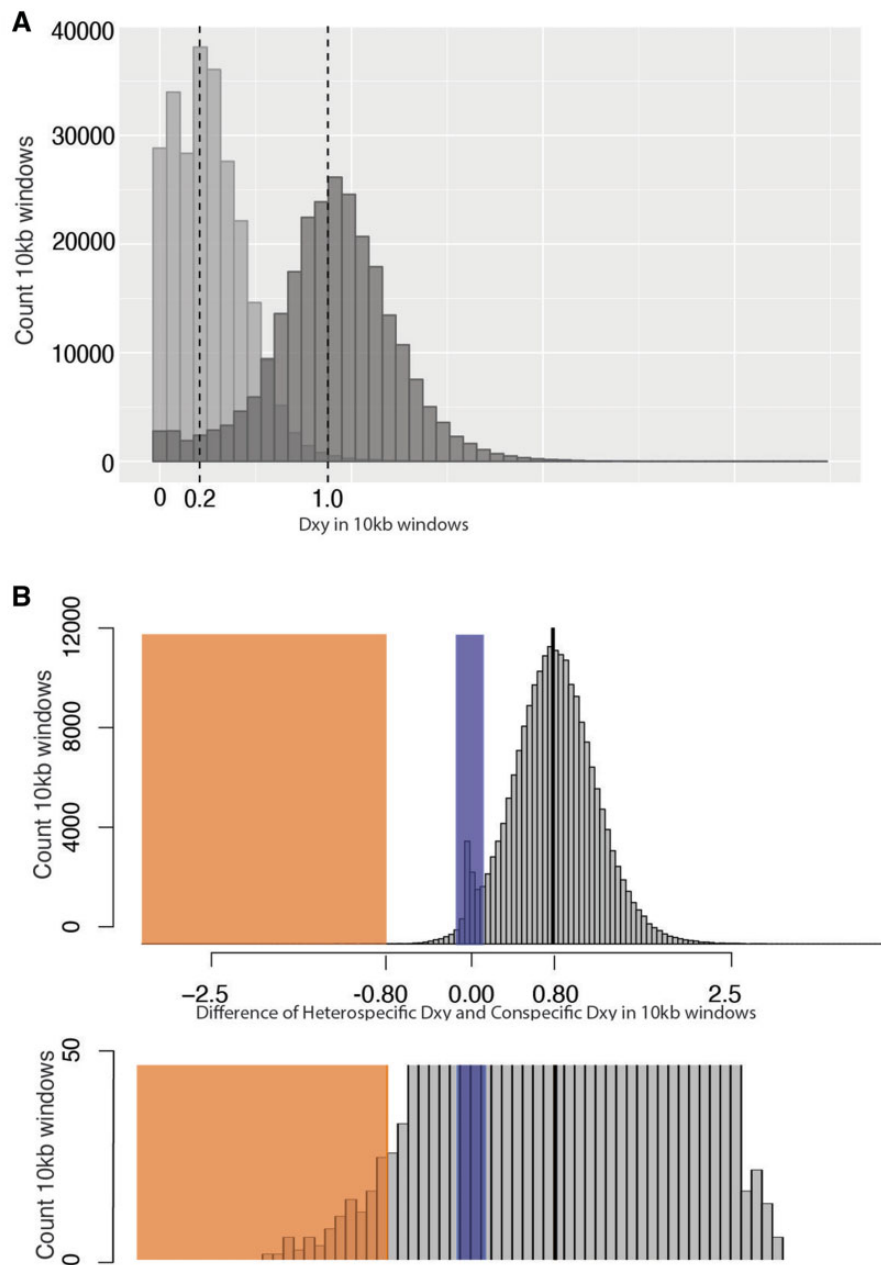


**FIG. 2.**—Identification of introgressed genomic regions. (A) Cartoon example (left) and actual data (right) showing comparisons used to identify introgressed genomic regions in focal individuals and not in the reference genome. For each individual,  $D_{xy}$  was calculated between the conspecific reference and the heterospecific reference genome sequences. In both the cartoon and real examples, conspecific divergence is shown in orange, and heterospecific divergence is shown in blue. Individual 1 contains a homozygous introgression at sites 6–14, individual 2 contains a heterozygous introgression at sites 6–14, individual 3 contains a heterozygous introgression at sites 6–9 and a homozygous introgression at sites 10–14. (B) Cartoon example (left) and actual data (right) showing comparisons used to identify introgressed genomic regions in focal individuals and in the conspecific reference genome. This example shows introgression at sites 6–14 in the conspecific reference genome and in individual A, but not in individual B.

We conducted computer simulations to assess the false-positive rate for identifying IGRs. We simulated population divergence without gene flow using estimates of population size, divergence time, mutation rate, and recombination rate from the literature (see Materials and Methods for details). Output from these simulations was analyzed using the same methods as with the actual data; we calculated heterospecific  $D_{xy}$ , conspecific  $D_{xy}$ , and the difference between them for each of 100,000 simulated 10 kb windows. The difference between heterospecific  $D_{xy}$  and conspecific  $D_{xy}$  was normally distributed with a mean of approximately 0.8. In 100,000 simulated 10 kb windows, none had a difference

less than or equal to 0, indicating that the false-positive rate for detecting IGRs is less than  $10^{-5}$  (supplementary fig. 2, Supplementary Material online). The low false-positive rate suggests that this approach is very conservative for detecting introgression. Next, we conducted simulations with gene flow to corroborate the intuition that introgression will lead to a shift in the distribution of the difference between heterospecific  $D_{xy}$  and conspecific  $D_{xy}$  values, leading to values below  $-0.8$  and an additional mode at 0. These simulations were identical to the first ones, but included a recent period of modest bidirectional migration. In 100,000 simulated 10 kb windows with gene flow, we observed a



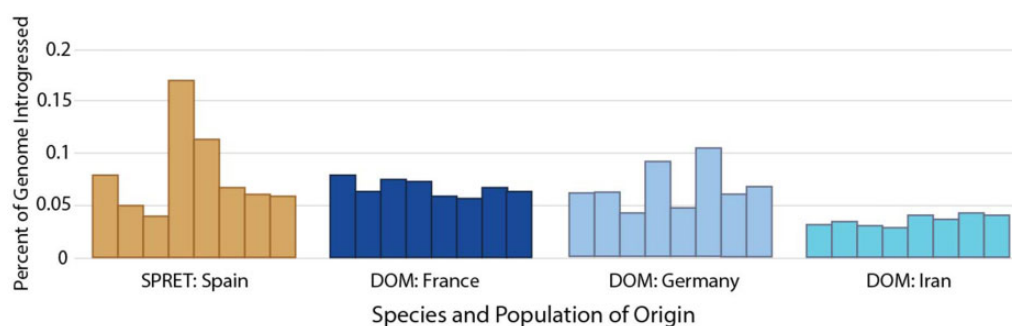


**Fig. 3.**— Histograms of  $D_{xy}$  values and their difference in comparisons between focal individuals and the heterospecific and conspecific reference genomes. (A) Histogram of conspecific and heterospecific  $D_{xy}$  in 10 kb windows. Light gray bars depict conspecific  $D_{xy}$  for all 10 kb windows (mean =  $\sim 0.2\%$ ). Dark gray bars depict heterospecific  $D_{xy}$  for all 10 kb windows (mean =  $\sim 1.0\%$ ). (B) Histogram of the difference in  $D_{xy}$  (heterospecific  $D_{xy}$  - conspecific  $D_{xy}$ ) per 10 kb window across the genome for a representative focal individual. The upper panel shows the entire distribution, whereas the lower panel is zoomed in to show the difference in the upper and lower tails of the distribution. The mean value shown as a solid black line ( $\sim 0.8\%$ ). The orange highlighted region represents windows with a value less than  $-0.8\%$  which are candidates for homozygous introgressions. The blue highlighted region centered around zero represents windows with nearly identical conspecific and heterospecific  $D_{xy}$  and includes candidates for heterozygous introgression.

secondary mode in the difference between heterospecific  $D_{xy}$  and conspecific  $D_{xy}$  at 0, as in the actual data (supplementary fig. 2, Supplementary Material online). A total of 6,450 10 kb windows had a difference between heterospecific  $D_{xy}$  and

conspecific  $D_{xy}$  less than or equal to 0. These simulations show that gene flow can produce the observed patterns.

IGRs were detected in all 32 individuals. The eight *M. spretus* individuals had an average of 0.08% of the



**Fig. 4.**—Proportion of genome introgressed per individual. The amount of introgressed DNA as a percentage of genome length for each of the 32 individuals, grouped by species and by population. (SPRET= *M. spretus*; DOM= *M. m. domesticus*).

genome putatively introgressed, ranging from 0.04 to 0.17% (fig. 4 and table 1). IGRs with overlapping breakpoints among individuals were collapsed to count the number of unique IGRs. *Mus spretus* had a total of 139 unique IGRs with a mean tract length of 114 kb. The three populations of *M. m. domesticus* had fewer unique IGRs (FRA, 95; GER, 89; and IRA, 97) and shorter tracts (average tract lengths: FRA, 56 kb; GER, 62 kb; and IRA, 22 kb) compared with *M. spretus*. *Mus musculus domesticus* individuals had an average of 0.06% of the genome identified as introgressed, ranging from 0.02 to 0.1% (fig. 4 and table 1). Thus, a small percentage of the genome has been introgressed in both directions between *M. m. domesticus* and *M. spretus*, with *M. spretus* genomes containing a greater amount of introgressed DNA.

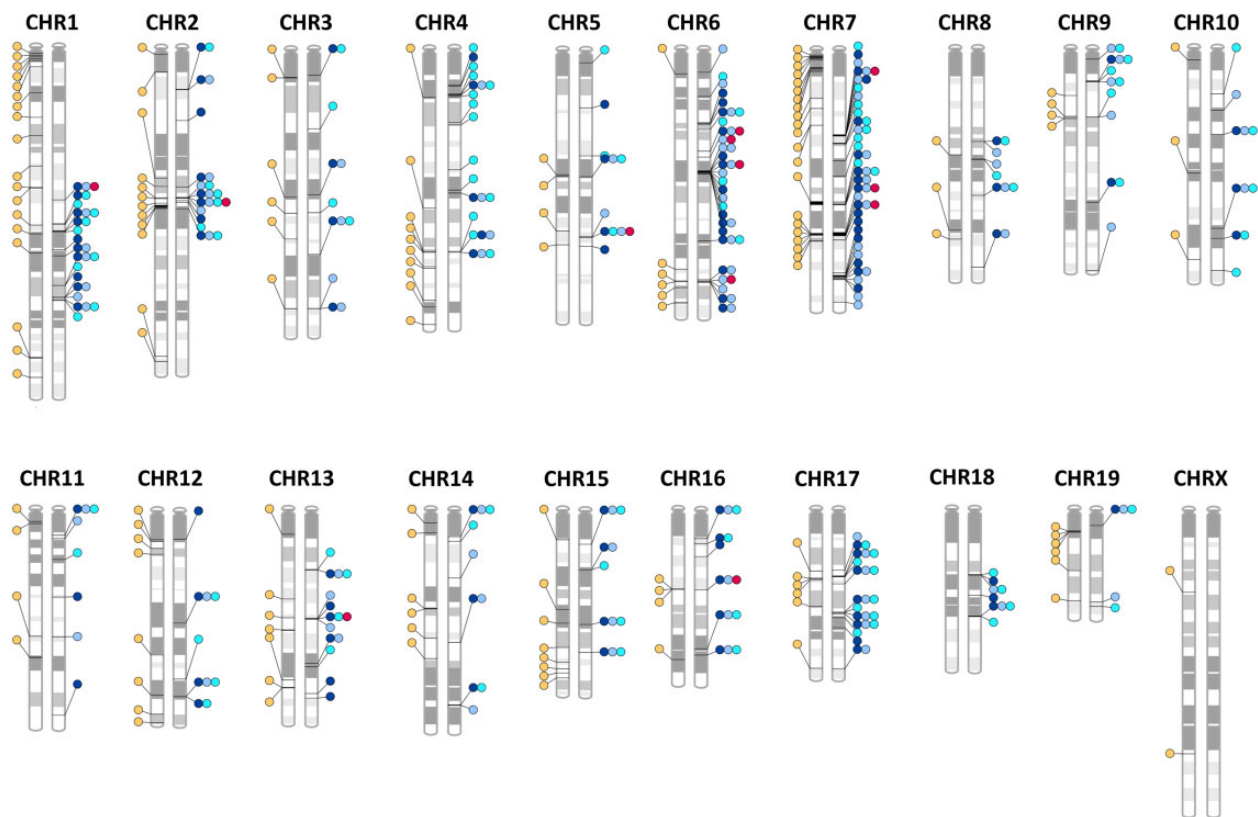
In all populations, the majority of IGRs were small (10 kb); however, there were several IGRs that were much longer (supplementary fig. 3, Supplementary Material online). *Mus spretus* had 58 unique introgressed tracts longer than 100 kb; the longest was a 1.87-Mb tract on chromosome 7 that contains a large cluster of olfactory receptor genes. The French and German *M. m. domesticus* populations shared their longest IGR, a 2.02-Mb tract on chromosome 7 that contains a large cluster of olfactory receptor genes. The Iranian *M. m. domesticus* population had only four IGRs that exceeded 100 kb, the longest being a 0.35-Mb region on chromosome 7 that contains no protein coding genes.

To identify the genomic position of introgressed regions, we mapped all IGRs identified in both species to the mouse genome (fig. 5). IGRs were located on nearly all 19 autosomes in both *M. spretus* and *M. m. domesticus* (absent only from chr18 in *M. spretus*). There were also two distinct IGRs on the X-chromosome in *M. spretus* (positions: 46330000–46350000 and 137000000–137340000), but there were no IGRs on the X-chromosome in any of the 24 *M. m. domesticus* individuals. Interestingly, chromosome 7 contained the largest fraction of total introgressed DNA in both species, corresponding to approximately 50% of IGRs from *M. spretus* into *M. m. domesticus* and approximately 33% of IGRs from *M. spretus* into *M. m. domesticus*. In addition to the large introgressed

tract on chromosome 7 in *M. m. domesticus* that contained a cluster of olfactory receptor genes, we also observed a 0.6-Mb tract on chromosome 7 containing *Vkorc1*, as previously described (Song et al. 2011). In both species, IGRs on chromosome 17 were not within or near the t-haplotype region, providing no evidence of introgression caused by t-haplotype meiotic drive. IGRs did not cluster near or far from centromeres or gene-rich regions, instead they appeared evenly distributed across chromosomes. Approximately half (45%) of all 10 kb windows across the genome contained genes. A similar proportion of IGRs contained genes (49–53%), suggesting that introgressed tracts are not enriched for gene-rich or gene-poor genomic regions.

#### Introgression from *M. spretus* in the *M. musculus* Reference Genome

We detected ten distinct genomic regions ranging in size from 30 kb to 0.44 Mb of *M. spretus* origin in the *M. musculus* reference genome, totaling 1.5 Mb (supplementary table 7, Supplementary Material online). In contrast, we did not detect any introgression from *M. musculus* in the reference genome of *M. spretus*. The method we used to detect introgressed regions in the reference genome required introgressions that were not fixed in the recipient species. In this study, the three geographically discrete wild populations of *M. m. domesticus* gave us greater power to detect nonfixed introgressions in the *M. m. domesticus* reference genome than in the *M. spretus* reference genome, because only a single population of wild *M. spretus* was sampled. Therefore, we cannot rule out the possibility that there is also introgression in the *M. spretus* reference genome. Moreover, the *M. spretus* reference genome was sequenced in a strain established from wild mice trapped in southern Spain. Population sampling of *M. spretus* from geographic regions farther from where the reference sample was collected would increase the power to detect introgressed regions in the *M. spretus* reference genome.



**FIG. 5.**—Genomic location of introgressed genomic regions (IGRs) between *M. m. domesticus* and *M. spretus*. The left chromosome in each pair represents *M. spretus*, with yellow dots indicating the location of IGRs ( $\geq 10$  kb) from *M. m. domesticus* into *M. spretus*. The right chromosome in each pair represents *M. m. domesticus*, with blue dots indicating the location of IGRs ( $\geq 10$  kb) from *M. spretus* into three populations of *M. m. domesticus* (dark blue = France; light blue = Germany; teal = Iran). Red dots on the right chromosome indicate the position of IGRs in the *M. musculus* reference genome GRCm38.

### Introgression Decreases with Increasing Geographic Distance from the Area of Sympatry

To assess the relationship between the geographic distance from the area of sympatry and the amount of introgression within each population, we estimated the amount and identity of genomic introgression from *M. spretus* into three *M. m. domesticus* populations. The three *M. m. domesticus* populations were all allopatric to *M. spretus* and were collected at varying distances from the area of sympatry (fig. 1). The distance from the nearest known *M. spretus* locality was approximately 550 km for the French *M. m. domesticus* population, approximately 1,050 km for the German population, and approximately 2,650 km for the Iranian population. ABBA-BABA tests were conducted for each *M. m. domesticus* population together with *M. m. musculus*, *M. spretus*, and *M. caroli*. The *D*-statistic was greatest for the French population, intermediate for the German population, and smallest for the Iranian population (table 1). Similarly, the average percent of the genome putatively introgressed was greatest in the French population, slightly less in the German population, and lowest in the Iranian population. There was a strong positive correlation ( $r^2 = 0.96$ ) between the average percent of

the genome introgressed and distance from the area of sympatry. The three *M. m. domesticus* populations have a similar number of introgressed tracts, and the differences in the introgressed genomic proportions were driven by the length, rather than by the number of introgressed tracts (table 1).

### Windows with IGRs Had Recombination Rates Equal to or Slightly Lower Than Windows without IGRs

We tested the prediction that introgressed regions preferentially fall in regions of the genome with higher rates of recombination (e.g., Nachman and Payseur 2012; Schumer et al. 2018). Contrary to expectations, 1 Mb windows that contained IGRs did not show higher recombination rates than regions without IGRs. The average recombination rate across nonoverlapping 1 Mb windows was 0.59 cM/Mb. The *M. spretus* population showed no significant difference in recombination rate between 1 Mb windows that contained IGRs (recombination rate = 0.53 cM/Mb) and 1 Mb windows that did not contain IGRs (recombination rate = 0.59 cM/Mb) (Mann-Whitney *U* test;  $P = 0.055$ ). All three *M. m. domesticus* populations had significantly lower recombination rates in 1 Mb windows containing IGRs (France, 0.46 cM/Mb;

Germany,  $0.43 \text{ cM/Mb}$ ; and Iran,  $0.37 \text{ cM/Mb}$ ) compared with those without (Mann–Whitney  $U$  test; France,  $P = 1.4 \times 10^{-4}$ ; Germany,  $P = 1.1 \times 10^{-4}$ ; and Iran,  $P = 1.8 \times 10^{-9}$ ).

### Introgressed Tracts Were Significantly Enriched for Olfactory Receptor Genes in Both Directions

To assess the functional significance of introgressed regions, genes encoded with IGRs were tested for overrepresentation of Gene Ontology (GO) categories. PANTHER overrepresentation tests of GO annotations of genes located within IGRs showed significant enrichment (Fisher's exact test,  $P < 0.001$ ) for olfactory-related GO categories in both *M. spretus* and *M. m. domesticus* (supplementary table 8, Supplementary Material online).

The olfactome in *Mus* consists of genomic regions ranging in size from solitary olfactory genes ( $< 1 \text{ kb}$ ) to large clusters of several hundred olfactory genes ( $> 1 \text{ Mb}$ ). Compared with the small proportion ( $< 1\%$ ) of the genome putatively introgressed in each *Mus* species, a significantly greater proportion of the olfactome is putatively introgressed (supplementary table 9, Supplementary Material online). This pattern is strongest in *M. spretus* with  $> 10\%$  of the olfactome introgressed. There is a striking overlap between the longest olfactory receptor clusters and the longest IGRs. The three longest olfactory receptor clusters (4.71 Mb on chr2, 2.48 Mb on chr9, and 1.66 Mb on chr7) all contain IGRs, and the three longest IGRs (2.02 Mb on chr7, 1.87 Mb on chr7, and 1.26 Mb on chr2) all contain olfactory receptor clusters.

## Discussion

### How Extensive Is Introgression between *M. spretus* and *M. m. domesticus*?

A small, but significant proportion of the genome is introgressed between natural populations of the house mouse (*M. m. domesticus*) and the Algerian mouse (*M. spretus*). These observations confirm and extend the results of previous studies (Orth et al. 2002; Liu et al. 2015). Using mtDNA, seven allozyme loci and four microsatellites, Orth et al. (2002) provided evidence of introgression in both directions between these species. Liu et al. (2015) used a genome-wide approach and documented introgression at a low level throughout the genome but did not assess the directionality of introgression. Our results show that introgression is both bidirectional and widespread throughout the genome. Despite this, the amount of introgression in this study (0.02–0.1% of the genome) is an order of magnitude less than has been reported between *Neanderthals* and modern non-African human populations (1–2%) (Green et al. 2010; Prüfer et al. 2014). The proportion of introgression is even higher for some young species groups that comprise adaptive radiations, such as *Heliconius* butterflies ( $\sim 20$ – $40\%$ ; Martin et al. 2013) and

Lake Tanganyika cichlid fishes (9–25%; Gante et al. 2016). The low level of introgression observed between *M. m. domesticus* and *M. spretus* may reflect the comparatively large degree of reproductive isolation between these taxa. First, they are sympatric over only a small portion of the range of *M. m. domesticus* (fig. 1). Even where they are sympatric, they are largely restricted to different habitats: *M. m. domesticus* is nearly always commensal with humans, whereas *M. spretus* is found in more natural settings. Consistent with these observations, no F1s have ever been observed in nature (Orth et al. 2002). Finally, these species display a high level of intrinsic postzygotic isolation due to both hybrid sterility and reduced hybrid viability. Crosses between male *M. spretus* and female *M. musculus* produce fertile F1 females and sterile F1 males (Bonhomme et al. 1978). F1 male sterility is due to lack of spermatozoa (Pelz and Niethammer 1978). Placentas of female F1s are hyperplastic when backcrossed with *M. musculus* and hypoplastic when backcrossed with *M. spretus* (Zechner et al. 1997), resulting in a moderate reduction in the viability of hybrid conceptuses (Arévalo et al. 2021). In light of these barriers to gene flow, the observation of low levels of introgression across the genome serves as a reminder that even good “biological species” may have semi-permeable genomes.

In addition to the detection of IGRs in wild mice, we also found evidence of ten IGRs in the *M. musculus* reference genome GRCm38, containing part or all of 27 protein coding genes (supplementary table 7, Supplementary Material online). The GRCm38 reference genome was generated from the classic laboratory strain C57BL/6J, which is mainly of *M. m. domesticus* subspecific origin with a few small regions of *M. m. musculus* and *M. m. castaneus* origin due to the establishment of lab lines from “fancy mice” (Yang et al. 2011). The regions of GRCm38 we identify here as introgressed from *M. spretus* are likely to have been present in the wild *M. m. domesticus* populations that were collected and bred as “fancy mice” preceding the 20th century.

### Bidirectional Introgression

Many studies measuring introgression assume gene flow from a donor population into a focal recipient population and do not consider the possibility of bidirectional gene flow. There are some situations where unidirectional gene flow may be expected, such as in cases where the donor species is rare, invasive, or a migrant to the recipient population (e.g., introgression between invasive and Gulf killifish, Oziolor and Matson 2018). A few recent studies have reported bidirectional gene flow. For example, Walsh et al. (2018) found bidirectional and largely symmetrical introgression between Nelson's sparrow and the Saltmarsh sparrow. Valencia-Montoya et al. (2020) found that initially after the introduction of an invasive moth (*Helicoverpa*) in Brazil, the invasive moth species harbored more introgressed tracts



compared with the native species. Over time, that pattern flipped, driven by the adaptive benefit of an allele conferring resistance to pesticide that introgressed from the invasive species into the native species.

Here, we document bidirectional introgression and found that *M. spretus* harbors a greater amount of introgressed DNA than *M. m. domesticus*. In a general sense, these observations confirm similar suggestions made by Orth et al. (2002) based on mtDNA and 12 nuclear loci. In that study, introgression was documented at nearly all loci, although more introgression was seen into *M. spretus* than into *M. m. domesticus*. The fact that introgression was seen at most of the loci surveyed by Orth et al. (2002) while less than 1% of the genome was identified as introgressed in the present study might seem surprising, although several factors could account for this difference. First, although most loci in Orth et al. (2002) showed some introgression when considering all individuals from both species, individual mice showed little introgression, as found here. For example, only 4 of the 64 *M. m. domesticus* showed any introgression from *M. spretus*, and each of these 4 mice harbored a *M. spretus* allele at only 1 of 13 loci. Second, Orth et al. (2002) reported higher proportions of introgressed alleles in Northern Africa than European samples of both species of *Mus*, while we sampled populations only from Europe. It is possible that some North African populations of *M. spretus* and *M. m. domesticus* have a greater proportion of IGRs, although Liu et al. (2015) found the opposite pattern in *M. m. domesticus*. Third, Orth et al. (2002) did not explicitly account for unsorted ancestral polymorphism as an explanation for shared variation. In contrast, our approach only included SNPs that were invariant within the donor population, and thus, the IGRs reported here are unlikely to be the result of incomplete lineage sorting. Fourth, our approach was very conservative in identifying windows as IGRs, and we do not suggest to have identified all genomic regions that have introgressed between the two species. Finally, the higher mutation rates of allozymes and microsatellites raise the possibility that some of the shared variation reported by Orth et al. (2002) is a consequence of recurrent mutation rather than gene flow. Nonetheless, our genome-wide results are broadly consistent with their finding of bidirectional and asymmetric introgression based on surveys of a few loci.

#### Does the Amount of Introgression Decrease with Increasing Geographic Distance from the Area of Sympatry?

The amount of introgression is predicted to decrease with increasing distance from the area of sympatry, either due to selection against deleterious introgressed variants or simply as a consequence of limited gene flow over long geographic distances. We tested this prediction and found that introgression in *M. m. domesticus* decreased with increasing distance from the area of overlap with *M. spretus*. This pattern is

consistent with other taxa in which introgression has been quantified in multiple allopatric populations. For example, desert fishes endemic to the Colorado river (*Gila robusta* and *G. cypha*) show spatially heterogeneous asymmetric introgression throughout their range, with the highest levels in areas of sympatry, but detectable introgression even in allopatric sites (Chafin et al. 2019). In European *Lissotriton* newts, introgression is highly localized, detected only at sites where they share the same habitat (i.e., syntopic), and absent at allotopic sites (Johanet et al. 2011). Similarly, in European seabass (*Dicentrarchus*), introgressed alleles were fixed in populations parapatric with the donor species, and absent in populations more geographically isolated, suggesting that these introgressed alleles may contribute to the reproductive isolation within the recipient species (Duranton et al. 2020). The *M. m. domesticus* populations included in this study are separated by hundreds of kilometers. It is striking that both the *D*-statistic and the analysis of IGRs provide evidence for significant introgression in the Iranian population of *M. m. domesticus*, 650 km from the range of *M. spretus*. As migration between the sampled populations is probably low, two other explanations for this finding should be considered. One is that these regions originally introgressed into populations of *M. m. domesticus* that are geographically closer to *M. spretus*, and gene flow within *M. m. domesticus* carried the alleles to Iran. Another possible explanation is that the signal of introgression comes from a closely related *Mus* species that is sympatric with *M. m. domesticus* in Iran, *M. macedonius*. The presence of multiple IGRs that are shared between the three populations of *M. m. domesticus* supports the gene-flow hypothesis. Future studies with increased sampling across the diversity of *Mus* species would help distinguish between introgression from *M. spretus* and introgression from *M. macedonius* in Iran. If IGRs have been carried long distances by migration, this would suggest that the observed *M. spretus* alleles are not strongly deleterious in *M. m. domesticus*.

#### Is the Recombination Rate Higher in IGRs?

Hybrids between *M. spretus* and *M. m. domesticus* have greatly reduced fitness due to intrinsic postzygotic isolation. Alleles causing hybrid sterility or inviability are deleterious and are thus expected to be eliminated by selection. Similarly, evidence from studies of humans suggests that introgressed loci from Neanderthals may often be deleterious (e.g., Sankararaman et al. 2014; Harris and Nielsen 2016). The deleterious effects of introgressed DNA can also be seen in time-series collections of invasive moths in Brazil which showed that after an initial pulse of hybridization, the proportion of introgressed DNA decreased (Valencia-Montoya et al. 2020). In genomic regions with low rates of recombination, gene flow between species is expected to be reduced because of linkage to deleterious alleles that are purged from the

population (Nachman and Payseur 2012; Schumer et al. 2018). Reduced introgression in genomic regions with low recombination has been seen in many species, including sunflowers (Rieseberg et al. 1999), *Drosophila* (Machado et al. 2002), rabbits (Carneiro et al. 2009), house mice (Geraldès et al. 2011), humans (Schumer et al. 2018), monkey flowers (Nelson et al. 2021), swordtail fishes (Schumer et al. 2018), and *Heliconius* butterflies (Edelman et al. 2019). In contrast, we were surprised to find that genomic windows with IGRs in *M. m. domesticus* showed a significantly reduced recombination rate compared with genomic windows without IGRs. This was particularly surprising in light of the fact that introgression between *M. m. musculus* and *M. m. domesticus* is greater in genomic regions with higher rates of recombination (Geraldès et al. 2011), as seen in many other species. In principle, the pattern that we document here could be caused by hitchhiking with beneficial variants, because hitchhiking effects are expected to be greater in regions of low recombination. However, at this point such a conclusion is speculative as we have no evidence that any of the introgressed variants are adaptive (apart from *Vkorc1*, discussed below).

#### How Are Introgressed Tracts Distributed across the Genome, and What Genes Do They Encode?

Despite the barriers to reproductive isolation between *M. spretus* and *M. m. domesticus*, IGRs were detected on nearly all autosomes of both species (missing only from chromosome 18 in *M. spretus*). We found less introgression on the X-chromosome than on the autosomes; no IGRs were detected on the *M. m. musculus* X-chromosome, and only two IGRs were detected on the *M. spretus* X-chromosome. The reduced introgression on the X-chromosome is consistent with other studies and may be explained by the “large X-effect,” the disproportionately large number of incompatibilities on the X-chromosome, as well as the exposure to selection of recessive X-linked incompatibilities in males (Coyne and Orr 1998; Presgraves 2018). One of the first documented cases of this pattern was seen in the hybrid zone between *M. m. domesticus* and *M. m. musculus*, where clines for X- and Y-linked loci were steeper than those for autosomal loci (Tucker et al. 1992). More recently, Presgraves (2018) reviewed over 100 studies and showed that the sex chromosomes were more differentiated than the autosomes in most cases, irrespective of whether males or females were the heterogametic sex. However, Presgraves (2018) also points out that such a pattern may arise in the absence of barriers to gene flow, as a simple consequence of differences in effective population size for the X-chromosome and autosomes. Indeed, Fraisse and Sachdeva (2021) summarized the results of 30 empirical studies that utilized genomic data and found that the sex chromosomes present a stronger barrier to gene flow than the autosomes in only about half of the cases. Thus, in principle, the patterns documented here could be caused by factors

other than barriers to gene flow. However, it is of considerable interest that two genes known to be involved in intrinsic postzygotic isolation between *M. musculus* and *M. spretus* map to the *M. spretus* X-chromosome (Guénet et al. 1990; Zechner et al. 1996). A major locus underlying abnormal placental development in *M. musculus* × *M. spretus* hybrids maps to the proximal portion of the *M. spretus* X-chromosome (Zechner et al. 1996), and a gene underlying hybrid male sterility maps to the distal portion of the *M. spretus* X-chromosome (Guénet et al. 1990). In both cases, it is the presence of an X-linked allele of *M. spretus* origin in a hybrid background that causes inviability or sterility. This is entirely consistent with our observation of asymmetric introgression of the X-chromosome between these species and strongly suggests that the patterns we document are caused by barriers to gene flow.

In *M. m. domesticus* and *M. spretus*, olfactory receptor genes were overrepresented in IGRs, and the longest introgressed regions contained olfactory receptor gene clusters. Odorant receptors (ORs) constitute the largest gene superfamily in mammals and are responsible for odor perception and detection of chemical cues. Mammals with greater reliance on olfaction have a greater number of functional OR genes than mammals with a greater emphasis on vision. For example, mice and rats, which rely heavily on olfaction, have over 1,100 intact OR genes, whereas humans have only 404 intact OR genes (Nei et al. 2008; Degl’Innocenti et al. 2019). Across mammals, particular OR gene families are associated with different ecotypes (aquatic, semiaquatic, terrestrial, and flying) suggesting that natural selection is driving patterns of OR diversity within clades (Hayden et al. 2010). Our finding that OR genes are overrepresented in introgressed regions is consistent with the findings of Liu et al. (2015) who reported an overrepresentation of OR genes introgressed into *M. m. domesticus* individuals. Here, we find evidence that this pattern is bidirectional between the species; *M. spretus* also harbors introgressed regions that are enriched for olfactory receptor genes. This pattern suggests that a beneficial effect of introgression may be the increased diversity of olfactory receptor genes.

The length of introgressed tracts is influenced by the number of generations that have elapsed since hybridization and the strength of positive selection on the region. Repeated backcrossing and recombination break down the size of introgressed tracts, whereas positive selection can lead to introgressed tracts remaining longer than would be expected given the time since hybridization. The timing of the first contact of *M. spretus* and *M. m. domesticus* is not precisely known. *Mus spretus* likely colonized Europe first from Northern Africa via the Gibraltar strait sometime during the last 10,000 years (Lalis et al. 2019). Cucchi et al. (2020) estimate that *M. m. domesticus* expanded their range westward to Europe from an ancestral population in the Iranian Plateau after the Bronze Age (~4,000 years ago). These

archeologically based estimates suggest that the first encounters and gene transfers between *M. spretus* and *M. m. domesticus* may date to approximately 3,000 years ago.

In general, genomic regions that are adaptively introgressed are expected to have long tract lengths and high population frequency. There are a few putatively introgressed regions that fit these conditions. One of the longest IGRs in *M. m. domesticus* contained the gene *Vkorc1*. The *M. spretus* haplotype of *Vkorc1* (*Vkorc1<sup>SPR</sup>*) confers resistance to warfarin rodenticide and therefore has a strong adaptive benefit in populations that are exposed to rodenticide. The identification of this strongly beneficial haplotype having a heterospecific origin in *M. m. domesticus* was one of the first compelling examples of an adaptive introgression (Song et al. 2011). Despite the well-documented benefits this introgressed haplotype confers (Endepols et al. 2013; Goulois et al. 2017), we found that the *Vkorc1<sup>SPR</sup>* allele was present at low frequency in *M. m. domesticus* populations. Only two heterozygous individuals were observed (FRA7 and GER4) out of 24, representing an overall frequency of 4.2%. Although a number of *M. m. domesticus* populations across Europe have high frequencies of *Vkorc1<sup>SPR</sup>*, this allele is absent or nearly absent in nearby populations (Rost et al. 2009; Song et al. 2011; Goulois et al. 2017). The highly localized pattern of frequency variation at this allele suggests that *Vkorc1<sup>SPR</sup>* may be deleterious in the absence of the strong selective pressure imposed by rodenticide, consistent with the known costs of resistance seen in other systems (e.g., Bergelson et al. 1996; Berticat et al. 2002; Janmaat and Myers 2003).

Interestingly, several IGRs contain overlapping introgressed tracts with distinct breakpoints, in which the central, shared region is at high frequency, raising the possibility that this central region may contain adaptive alleles. One such example in *M. spretus* involves the introgressed region on chromosome 6 containing *Slco1a1*, a gene associated with blood and urine homeostasis (supplementary fig. 1 and table 11, Supplementary Material online). *Slco1a1* has a frequency of 0.8125 in *M. spretus* with decreasing frequency of introgression on either side. Along with *Slco1a1*, a gene with a similar pattern in *M. spretus* is *Esx1* (X-chromosome). Genes with a similar pattern in *M. m. domesticus* include *skint5* (chromosome 4), a cluster of hemoglobin beta genes (chromosome 7), and *Apobec3* (chromosome 15) (supplementary table 11, Supplementary Material online). Although *Vkorc1* is a clear example of adaptive introgression, no agents of selection have been reported for these other genes or gene regions.

## Conclusion

We conducted a genome-wide scan of introgression between *M. m. domesticus* and *M. spretus* and observed widespread bidirectional and asymmetric introgression. Notably, the *M. musculus* reference genome harbors previously unidentified segments from *M. spretus*. No introgression was seen

from *M. spretus* to *M. m. domesticus* on the X-chromosome, consistent with the location of known hybrid sterility and inviability genes. Introgression between species decreased with increasing geographic distance from the areas where the two species co-occur. Finally, we observed an overrepresentation of introgressed olfactory receptor genes in both species, suggesting on-going selection for novel olfactory repertoires.

## Materials and Methods

### Samples

All population samples and whole-genome reference sequences of *M. musculus*, *M. spretus*, and *M. caroli* were from previously published work. The wild population samples were downloaded as the variant call format (VCF) file <AllMouse.vcf\_90\_recalibrated\_snps\_raw\_indels\_reheader\_PopSorted.PASS.vcf> (Harr et al. 2016) which is aligned to the mm10 reference genome (Waterston et al. 2002). The *M. caroli* genome (Thybert et al. 2018), which was used as an outgroup for ABBA–BABA tests (see below), was downloaded from European Nucleotide Archive GCA\_900094665.2, and converted to mm10 genome coordinates via liftOver with the UCSC chain file: <GCF\_900094665.1ToMm10.over.chain.gz>. The *M. spretus* genome (Keane et al. 2011), which was used to identify IGRs (see below), was downloaded from UCSC GCA\_001624865.1\_SPRET\_Eij\_v1, and converted to mm10 genome coordinates via liftOver with the UCSC chain file: <GCA\_001624865.1\_SPRET\_Eij\_v1ToMm10.over.chain.gz>.

### Population ABBA–BABA Analyses

To test for introgression between *M. spretus* and *M. musculus*, we first employed population-level ABBA–BABA tests (Green et al. 2010, Durand et al. 2011). The four populations that were used for this test were as follows: Pop1 (8 *M. m. musculus* individuals from Kazakhstan), Pop2 (24 *M. m. domesticus* individuals, including 8 each from France, Germany, and Iran), Pop3 (8 *M. spretus* individuals from Spain), and Outgroup (the *M. caroli* reference genome) (fig. 1). Each of the three *M. m. domesticus* populations was also tested individually, resulting in four separate ABBA–BABA analyses (table 1).

Variant calls for 43 *Mus* individuals (27 *M. m. domesticus*, 8 *M. m. musculus*, and 8 *M. spretus*) were extracted from the “PASS” vcf file published in Harr et al. (2016). This file was pruned with vcftools v0.1.15 (Danecek et al. 2011) to include only sites that were bi-allelic, and had alleles called for at least 23/43 individuals. The *M. caroli* allele (Thybert et al. 2018) was defined as the ancestral state; sites not called in the *M. caroli* reference genome were not included. The *D*-statistic is the difference of ABBA and BABA allelic patterns across the genome, divided by their sum, or  $D = (ABBA - BABA) / (ABBA + BABA)$  (Durand et al. 2011). A significantly positive

$D$ -statistic indicates an excess of SNPs in common between Pop2 and Pop3, whereas a significantly negative  $D$ -statistic indicates an excess of SNPs in common between Pop1 and Pop3. Significance was assessed by block jackknife procedure (window size 200 kb) implemented with the script jackKnife.R from the ANGSD package (Korneliussen et al. 2014).

### Identifying Putatively Introgressed Tracts

The ABBA–BABA tests described above can provide statistical support for introgression genome-wide but are underpowered to identify specific IGRs (Martin et al. 2015) or to identify the frequency of introgressed variants. To identify putatively IGRs in each wild *M. m. domesticus* and each *M. spretus* individual, we calculated the difference in genetic divergence between the focal individual and the conspecific and heterospecific reference genome sequences in sliding windows across the genome (fig. 2), as described in detail below. This method is conceptually similar to the metric  $G_{\min}$  (Geneva et al. 2015).

### Filtering Sites

To exclude ancient shared polymorphisms, we only included sites that were invariant among eight individuals of the “donor” species, and had allele calls for at least four of those. Thus, to identify introgression from *M. m. domesticus* to *M. spretus*, each of the eight wild *M. spretus* individuals was compared with the conspecific SpretEIJ reference genome (Keane et al. 2011) and to the heterospecific *M. m. domesticus* reference genome (GRCm38, Waterston et al. 2002) at sites that were invariant among eight *M. m. domesticus* samples representing all *M. m. domesticus* populations in the Harr et al. (2016) data set (FRA1, FRA2, GER1, GER2, HEL2, HEL3, IRA1, and IRA2). To identify introgression from *M. spretus* to *M. m. domesticus*, each of the 24 wild *M. m. domesticus* individuals was compared with the conspecific GRCm38 reference genome and to the heterospecific SpretEIJ reference genome at sites that were invariant among the 8 *M. spretus* individuals. Despite inbreeding, a small fraction (~4.8%) of the *M. spretus* reference genome SNPs are heterozygous; these were filtered out using vcfTools. (version 0.1.15, Danecek et al. 2011). Regions of the *M. musculus* genome of *M. m. castaneus* or *M. m. musculus* subspecific origin (Yang et al. 2011) were not included in analyses. Absolute nucleotide distance ( $D_{xy}$ ) for unphased diploid genotypes was calculated in nonoverlapping 10 kb sliding windows using custom scripts.

### Homozygous Introgression Tracts

To identify introgressed regions, we first consider the situation where the reference genomes do not contain introgressed tracts. We then consider the situation where the reference genomes also contain introgressed tracts. Genome-wide, the

average  $D_{xy}$  between *M. m. domesticus* and *M. spretus* is approximately 1%, but the average  $D_{xy}$  within each species is approximately 0.2%, and the difference between these values is 0.8% (fig. 3). In contrast, windows that are homozygous for an introgressed tract are expected to have a low level of divergence when compared with the reference genome of the donor species and a high level of divergence when compared with the reference genome of the same species (fig. 2A). Specifically, homozygous introgressed regions should exhibit a heterospecific  $D_{xy}$  of approximately 0.2% and a conspecific  $D_{xy}$  of approximately 1%, with an expected difference of approximately  $-0.8\%$  (supplementary table 1, Supplementary Material online). This expected value ( $-0.8\%$ ) is more than three standard deviations from the mean genome-wide value (0.8%), corresponding to a Z-score of  $-3.23$  for *M. musculus* and  $-3.65$  for *M. spretus* ( $P < 0.001$  in each case). Putatively introgressed homozygous windows were thus conservatively defined as those for which the difference in  $D_{xy}$  values between the focal individual and the two reference genomes matched or exceeded this threshold (fig. 3).

### Heterozygous Introgression Tracts

Windows that are heterozygous for an introgressed tract are expected to have levels of divergence that are the average of the two alleles. Thus, heterozygous regions are expected to exhibit nearly identical  $D_{xy}$  in comparison to the conspecific and heterospecific reference genomes at a value intermediate between approximately 0.2% and 1% (fig. 2A), and the difference between these values is expected to be 0 (fig. 3). This expected value (0) is nearly two standard deviations from the mean genome-wide value (0.8%), corresponding to a Z-score of  $-1.78$  ( $P < 0.05$ ). We identified windows within the 5% quantile centered on this expected value (fig. 3). To reduce the number of false positives, heterozygous windows were only defined as an IGR if they were greater than 30 kb in length (i.e., three contiguous 10 kb windows), or if the window was homozygous for the introgressed haplotype in another individual.

### Introgression in the Reference Genome

We also searched for regions that were introgressed in each reference genome. In cases where a focal individual is homozygous for an introgressed tract that is also shared with the conspecific reference genome,  $D_{xy}$  for the focal individual is expected to be approximately 0.2% compared with both the conspecific reference genome and the heterospecific reference genome, and the difference in  $D_{xy}$  values is expected to be 0. Moreover,  $D_{xy}$  for focal individuals not sharing this introgressed tract is expected to be approximately 1% compared with both the conspecific reference genome and the heterospecific reference genome, and the difference in  $D_{xy}$



values is also expected to be 0. Thus, as above, we identified windows within the 5% quantile centered on the expected value of 0. Among these, windows with  $D_{xy} \geq 1\%$  in focal individual A and  $D_{xy} \leq 0.2\%$  in focal individual B were defined as putatively introgressed in both the reference genome and individual B (fig. 2B). Finally, we note that we cannot identify introgressed regions that are fixed in the recipient species (i.e., found in all sampled individuals and the reference genome) because such regions are expected to show little variation within and between species and thus cannot be distinguished from regions of very high constraint.

### Collapsing Windows into Tracts

For each of the 32 individuals (24 *M. m. domesticus* and 8 *M. spretus*), adjacent putatively introgressed windows (10 kb each) were collapsed into IGRs and the population frequency of each IGR was calculated (supplementary tables 2–5, Supplementary Material online). Within populations, IGRs that had overlapping breakpoints between individuals were further collapsed to population-level IGRs (supplementary table 6, Supplementary Material online).

### Computer Simulations

To estimate the false-positive rate for detecting IGRs we generated 100,000 simulated 10 kb genomic windows with realistic population parameters using the program SLiM3 (Haller and Messer 2019) and custom python scripts. We simulated two populations, each with an effective population size ( $N_e$ ) of 100,000 diploid individuals (Phifer-Rixey et al. 2020), mutation rate per base per generation ( $\mu_0$ ) of  $6e^{-9}$  (Milholland et al. 2017), and recombination rate per bp ( $r_0$ ) of  $5.9842e^{-9}$  (Cox et al. 2009). The two simulated populations were isolated for 900,000 generations. A scaling parameter of  $Q = 50$  was used to make the simulations more computationally feasible. SLiM3 was used to generate tree files for each simulated 10 kb window. A custom python script used pslim and msprime to overlay mutations onto the SLiM-generated tree files. We then subsampled ten (diploid) individuals per population and generated one vcf file per simulated 10 kb window with the 20 genotype calls. Individual 1 from Population A was assigned to represent the “focal individual”; the first allele of Individual 2 from Population A was assigned to represent the heterospecific reference, and individuals 1–8 of Population B were assigned to represent the conspecific population.

A second simulation that included migration was run to corroborate the intuition that the patterns seen in the actual data are consistent with introgression. The second simulation was identical to the first, but included a recent period of modest bidirectional migration between the two populations (simulated migration rate:  $1e^{-5}$  from generation 750,000–900,000).

Vcf files from each of the two simulations were analyzed using the same methods as for the actual data to calculate heterospecific  $D_{xy}$ , conspecific  $D_{xy}$ , and the difference between them for each of the 100,000 simulated 10 kb windows. The results were summarized in histograms (supplementary fig. 3, Supplementary Material online).

### Overrepresentation Tests

#### PANTHER Analyses

To test for functional enrichment of Gene Ontology categories, genes located within putatively introgressed windows were compared with background genes from all genomic windows included in these analyses using PANTHER v11.1 (Mi et al. 2016). Significance was calculated using Fisher’s exact tests with a Bonferroni correction. PANTHER enrichment tests were done for each of the four populations.

#### Olfactome Proportion

Gene Ontology analyses have the weakness of reporting overrepresentations of gene categories that exist in clusters, such as olfactory genes. To support that the overrepresentation of olfactory genes is not due to a bias in the GO analyses, we calculated the percentage of the total olfactome that was introgressed. The olfactome was defined as all nonpseudogenized mouse olfactory genes or gene-clusters (Degl’Innocenti et al. 2019), totaling 1.46% of the mouse genome. A gene or gene-cluster was considered introgressed if it contained or overlapped with an IGR.

#### Recombination Rate

Several studies have reported a greater proportion of introgression in genomic regions experiencing a higher recombination rate, likely due to the reduced effects of selection at linked sites in such regions (Carneiro et al. 2009; Gerales et al. 2011; Schumer et al. 2018). To test for a difference in the recombination rate of genomic regions that contain IGRs and those without IGRs, recombination rate was estimated in nonoverlapping 1 Mb windows (Cox et al. 2009). We used the Mann–Whitney  $U$  test to compare the recombination rate of 1 Mb windows with and without putative introgression in all four populations.

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

### Acknowledgments

We thank members of the Nachman lab for valuable discussions. Katya Mack assisted with developing the computational pipeline. Elizabeth Beckman assisted with the SLiM

simulations. Sylvia Durkin, Elizabeth Beckman, Emilie Richards, Jimmy A. McGuire, and Mallory Ballinger gave helpful feedback on the manuscript. This work was supported by a National Institutes of Health (NIH) grant to MWN (NIGMS R01 GM127468).

## Data Availability

The data sets were derived from source: <https://doi.org/10.1038/sdata.2016.75>. The data underlying this article are available in the article and in its online [supplementary material](#). Code used in this project is available on GitHub: <https://github.com/sebanker27/SpretusIntrogression>

## Literature Cited

- Arévalo L, Gardner S, Campbell P. 2021. Haldane's rule in the placenta: sex-biased misregulation of the *Kcnq1* imprinting cluster in hybrid mice. *Evolution* 75(1):86–100.
- Bergelson J, Purrington CB, Palm CJ, Lopez-Gutierrez JC. 1996. Costs of resistance: a test using transgenic *Arabidopsis thaliana*. *Proc R Soc London Ser B Biol Sci.* 263(1377):1659–1663.
- Berticat C, Boquien G, Raymond M, Chevillon C. 2002. Insecticide resistance genes induce a mating competition cost in *Culex pipiens* mosquitoes. *Genet Res.* 79(1):41–47.
- Bonhomme F, Martin S, Thaler L. 1978. Hybridization between *Mus musculus* L. and *Mus spretus* Lataste under laboratory conditions (author's transl). *Experientia* 34(9):1140–1141.
- Bonhomme F, Searle JB. 2012. House mouse phylogeography. In: Macholán M, Baird S, Munclinger P, Piálék J, editors. *Evolution of the house mouse*. Cambridge: Cambridge University Press. p. 278–296.
- Cahill JA, et al. 2018. Genomic evidence of widespread admixture from polar bears into brown bears during the last ice age. *Mol Biol Evol.* 35(5):1120–1129.
- Carneiro M, Ferrand N, Nachman MW. 2009. Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* 181(2):593–606.
- Chafin TK, Douglas MR, Martin BT, Douglas ME. 2019. Hybridization drives genetic erosion in sympatric desert fishes of western North America. *Heredity (Edinb).* 123(6):759–773.
- Chevret P, Veyrunes F, Britton-Davidian J. 2005. Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biol J Linn Soc.* 84(3):417–427.
- Cox A, et al. 2009. A new standard genetic map for the laboratory mouse. *Genetics* 182(4):1335–1344.
- Coyne JA, Orr HA. 1998. The evolutionary genetics of speciation. *Philos Trans R Soc Lond B Biol Sci.* 353(1366):287–305.
- Cucchi T, et al. 2020. Tracking the Near Eastern origins and European dispersal of the western house mouse. *Sci Rep.* 10(1):8276.
- Danecek P, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
- Dasmahapatra KK, et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487:94–98.
- Degl'Innocenti A, Meloni G, Mazzolai B, Ciofani G. 2019. A purely bioinformatic pipeline for the prediction of mammalian odorant receptor gene enhancers. *BMC Bioinformatics* 20(1):474.
- Dejager L, Libert C, Montagutelli X. 2009. Thirty years of *Mus spretus*: a promising future. *Trends Genet.* 25(5):234–241.
- Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol.* 28(8):2239–2252.
- Duranton M, et al. 2020. The contribution of ancient admixture to reproductive isolation between European sea bass lineages. *Evol Lett.* 4(3):226–242.
- Edelman NB, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366(6465):594–599.
- Endepols S, Klemann N, Song Y, Kohn MH. 2013. *Vkorc1* variation in house mice during warfarin and difenacoum field trials. *Pest Manag Sci.* 69(3):409–413.
- Endler JA. 1977. *Geographic variation, speciation, and clines*. Princeton (NJ): Princeton University Press.
- Fraïsse C, Sachdeva H. 2021. The rates of introgression and barriers to genetic exchange between hybridizing species: sex chromosomes vs autosomes. *Genetics* 217(2):iyaa025. doi:10.1093/genetics/iyaa025.
- Gante HF, et al. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Mol Ecol.* 25(24):6143–6161.
- Geneva AJ, Muirhead CA, Kingan SB, Garrigan D. 2015. A new method to scan genomes for introgression in a secondary contact model. *PLoS One* 10(4):e0118621.
- Geraldes A, Basset P, Smith KL, Nachman MW. 2011. Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Mol Ecol.* 20(22):4722–4736.
- Goulois J, et al. 2017. Study of the efficiency of anticoagulant rodenticides to control *Mus musculus domesticus* introgressed with *Mus spretus Vkorc1*. *Pest Manag Sci.* 73(2):325–331.
- Guénet JL, Nagamine C, Simon-Chazottes D, Montagutelli X, Bonhomme F. 1990. *Hst-3*: an X-linked hybrid sterility gene. *Genet Res.* 56(2–3):163–165.
- Green RE, et al. 2010. A draft sequence of the Neanderthal genome. *Science* 328(5979):710–722.
- Haller BC, Messer PW. 2019. SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol Biol Evol.* 36(3):632–637.
- Hardouin EA, et al. 2015. Eurasian house mouse (*Mus musculus* L.) differentiation at microsatellite loci identifies the Iranian plateau as a phylogeographic hotspot. *BMC Evol Biol.* 15:26.
- Harr B, et al. 2016. Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*. *Sci Data.* 3:160075.
- Harris K, Nielsen R. 2016. The genetic cost of Neanderthal introgression. *Genetics* 203(2):881–891.
- Harrison RG. 1990. Hybrid zones: windows on evolutionary process. In: Futuyma DJ, Antonovics J, editors. *Oxford surveys in evolutionary biology*, vol. 7. Oxford: Oxford University Press. p. 69–128.
- Hayden S, et al. 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res.* 20(1):1–9.
- Huerta-Sánchez E, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512:194–197.
- Janmaat AF, Myers J. 2003. Rapid evolution and the cost of resistance to *Bacillus thuringiensis* in greenhouse populations of cabbage loopers, *Trichoplusia ni*. *Proc R Soc Lond B.* 270(1530):2263–2270.
- Johanet A, Secondi J, Lemaire C. 2011. Widespread introgression does not leak into allotopy in a broad sympatric zone. *Heredity (Edinb).* 106(6):962–972.
- Jones MR, et al. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* 360(6395):1355–1358.
- Juric I, Aeschbacher S, Coop G. 2016. The strength of selection against Neanderthal introgression. *PLoS Genet.* 12(11):e1006340.
- Keane TM, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477(7364):289–294.
- Korneliusson TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15(1):356.

- Lalis A, et al. 2019. Out of Africa: demographic and colonization history of the Algerian mouse (*Mus spretus* Lataste). *Heredity* (Edinb). 122(2):150–171.
- Liu KJ, et al. 2015. Interspecific introgressive origin of genomic diversity in the house mouse. *Proc Natl Acad Sci U S A*. 112(1):196–201.
- Machado CA, Kliman RM, Markert JA, Hey J. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol Biol Evol*. 19(4):472–488.
- Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol Biol Evol*. 32(1):244–257. 2014
- Martin SH, et al. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res*. 23(11):1817–1828.
- Matsuda Y, Chapman VM. 1992. Analysis of sex-chromosome aneuploidy in interspecific backcross progeny between the laboratory mouse strain C57BL/6 and *Mus spretus*. *Cytogenet Cell Genet*. 60(1):74–78.
- Meier JI, et al. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat Commun*. 8:14363.
- Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. 2016. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res*. 44(D1):D336–D342.
- Milholland B, et al. 2017. Differences between germline and somatic mutation rates in humans and mice. *Nat Commun*. 8:15183.
- Nachman MW, Payseur BA. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc Lond B Biol Sci*. 367(1587):409–421.
- Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet*. 9(12):951–963.
- Nelson TC, et al. 2021. Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLoS Genet*. 17(2):e1009095.
- Orth A, et al. 2002. Natural hybridization between 2 sympatric species of mice, *Mus musculus domesticus* L. and *Mus spretus* Lataste. *C R Biol*. 325(2):89–97.
- Oziol EM, Matson CW. 2018. Adaptation in polluted waters: lessons from Killifish. In: Burggren W, Dubansky B, editors. *Development and environment*. Cham: Springer. p. 355–375.
- Pelz HJ, Niethammer J. 1978. Cross-breeding of laboratory mice and *Mus spretus* from Portugal. *J Mammal Biol*. 43:302–304.
- Phifer-Rixey M, Harr B, Hey J. 2020. Further resolution of the house mouse (*Mus musculus*) phylogeny by integration over isolation-with-migration histories. *BMC Evol Biol*. 20(120):1–9.
- Phifer-Rixey M, Nachman MW. 2015. The Natural History of Model Organisms: insights into mammalian biology from the wild house mouse *Mus musculus*. *Elife* 4:e05959.
- Presgraves DC. 2018. Evaluating genomic signatures of “the large X-effect” during complex speciation. *Mol Ecol*. 27(19):3822–3830.
- Prüfer K, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49.
- Rajabi-Maham H, Orth A, Bonhomme F. 2008. Phylogeography and post-glacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe. *Mol Ecol*. 17(2):627–641.
- Rieseberg LH, Whitton J, Gardner K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152(2):713–727.
- Rost S, et al. 2009. Novel mutations in the *VKORC1* gene of wild rats and mice—a response to 50 years of selection pressure by warfarin? *BMC Genet*. 10:4.
- Sankararaman S, et al. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507(7492):354–357.
- Schumer M, et al. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360(6389):656–660.
- She JX, Bonhomme F, Boursot P, Thaler L, Catzeffis F. 1990. Molecular phylogenies in the genus *Mus*: comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. *Biol J Linn Soc*. 41(1–3):83–103.
- Song Y, et al. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Curr Biol*. 21(15):1296–1301.
- Staubach F, et al. 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genet*. 8(8):e1002891.
- Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat Ecol Evol*. 3(2):170–177.
- Teeter KC, et al. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res*. 18(1):67–76.
- Thybert D, et al. 2018. Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. *Genome Res*. 28(4):448–459.
- Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM. 1992. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution* 46(4):1146–1163.
- Valencia-Montoya WA, et al. 2020. Adaptive introgression across semipermeable species boundaries between local *Helicoverpa zea* and invasive *Helicoverpa armigera* moths. *Mol Biol Evol*. 37(9):2568–2583.
- Wallbank RWR, et al. 2016. Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLoS Biol*. 14(1):e1002353.
- Walsh J, Kovach AI, Olsen BJ, Shriver WG, Lovette IJ. 2018. Bidirectional adaptive introgression between two ecologically divergent sparrow species. *Evolution* 72(10):2076–2089.
- Waterston RH, et al.; Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520–562.
- Yang H, et al. 2011. Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet*. 43(7):648–655.
- Zechner U, et al. 1996. An X-chromosome linked locus contributes to abnormal placental development in mouse interspecific hybrids. *Nat Genet*. 12:398–403.
- Zechner U, et al. 1997. Paternal transmission of X-linked placental dysplasia in mouse interspecific hybrids. *Genetics* 146(4):1399–1405.

Associate editor: Adam Eyre-Walker