On the demographic history of the western house mouse, *Mus musculus domesticus*

by

Kennedy Agwamba

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computational Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael W Nachman, Chair
Professor Rasmus Nielsen
Professor Jeffrey Wall

Spring 2024

On the demographic history of the western house mouse, *Mus musculus domesticus*

Kennedy Agwamba

Abstract

On the demographic history of the western house mouse, *Mus musculus domesticus*

by

Kennedy Agwamba

Doctor of Philosophy in Computational Biology

University of California, Berkeley

Professor Michael W Nachman, Chair

The human commensal, *Mus musculus domesticus*, is native to the Eurasian continent, having spread from north of the Indo-Pakistani subcontinent and the Middle East to western Europe by the end of the Iron Age. Wild populations of *M. m. domesticus* are now distributed across Africa, the Americas, and Oceania, a range consistent with the global migration of western Europeans that began in the late 15th century. Despite its standing as the premier mammalian model organism for biomedical, ecological, and evolutionary research, important details surrounding the population history of wild house mice remain a mystery. To investigate patterns of genetic structure and infer the demographic history of the western house mouse, I analyze collections of genomic and exomic sequences generated from mice sampled in western Europe and the Americas. First, I present evidence that connects the colonization history of house mice in eastern North America to the colonization history of the Americas by European settlers which began in the late 15th century. Then, I analyzed a genomic dataset from an expanded sampling of Western Europe, including 59 new whole genome sequences from historically relevant regions of Western Europe, and present evidence of population structure, population splits, and gene flow between distinct populations of Western European house mice. Lastly, I combine datasets comprised of Western European genomes and North and South American exomes to detail the recent colonization history of house mice in both American continents. Altogether, these results elucidate details around the recent introduction of Western European house mice to North and South America, highlighting the effects of human migration and global colonization on the concurrent spread of an invasive human commensal.

To Crenshaw, To Van Ness, To El Segundo, and Rose

Contents

List of Figures

List of Tables

July 2017.

I remember the first time I visited my parent's house after spending nearly a year in New York City. I had moved to the Bronx from Seattle in the fall of 2016 without telling a soul. I went the entire year without seeing anyone I knew from my first 18 years of life.

At the age of 22, most people would refer to visiting their parent's house as "coming home". But that house had not been my home since I left for college. My father had harped all of his children to be out of that house when we turned 18. Given the way he ran his household, I was enthusiastic to make this a reality.

I was a visitor in the house I grew up in. My younger sibling had fully colonized the room we once shared. To allow him to continue to revel in the experience of having his own room, I opted to sleep on the couch. Sleeping on the couch allowed me to settle into my new relationship with a once familiar environment. My mother, who I hadn't seen in a year, slept on the couch perpendicular to the one I settled into. At the time, I wasn't sure if this was an act of solidarity, or just her new regular sleeping arrangement.

In the morning, I would wake up around 4:30am to the sound of my mother's voice singing Igbo hymns. It was a practice she picked up when I was in high school. Back then, I would wake up every day around the same time to catch the Metro 210 North on Crenshaw and El Segundo for the first leg of my two-hour commute to North Hollywood High School. I had no idea that she would continue the practice for years to come.

My sister once told me our mother prayed every day after I would leave the house for high school. She would pray that God would bring me home safely. We were like many other American families in this way: what we lacked in resources and personal agency, we made up for with prayer and hope.

During this time at my parent's house, I would learn some incredible details about my mother and her character. I realized it was from her that I inherited my sense of baseless optimism. And I believe this moment I have chosen to share speaks to her brand of humanity, and the humanity she instilled in her children.

******

I would like to thank my advisor, Michael Nachman, for accepting me into your lab, and sharing your enthusiasm of scientific theory and the natural world with me. I would like to thank the members of the Nachman lab and the greater MVZ for creating a welcoming community for scientific discourse. I am thankful to my

collaborators, and previous members of the Nachman and Searle lab who collected the mice we sequenced and analyzed for these studies, and in particular, Lydia Smith, whose wet lab and sequencing experience proved invaluable for the completion of my second and third chapters. I am thankful to Kate Chase for being responsive, understanding, and patient with all of my administrative needs. I am also grateful to my committee members, Jeff Wall and Rasmus Nielsen, for all of your support and guidance.

Lastly, to my friends, who constantly acknowledge my humanity, and to my mother, who bestowed me with it, please enjoy 15,000+ words of scientific prose about house mouse demography.

Chapter 1

The demographic history of house mice (*Mus musculus domesticus*) in eastern North America

This chapter is adapted from content published in G3 (Agwamba and Nachman 2023).

Introduction

Untangling the invasion histories of human commensals is helpful for understanding the impact of human migration on global ecosystems. Rats, pigeons, sparrows, zebra mussels and fruit flies are a few of the many invasive species that humans have accidentally or intentionally helped disperse around the world (Robbins 1973; Astanei et al. 2005; Stringham et al. 2012; Jones et al. 2013; Puckett et al. 2016; Mallez and McCartney 2018; Arguello et al. 2019; Cucchi et al. 2020). The presence of these species often dramatically alters ecosystems, affecting the evolutionary trajectories of the organisms that interact with them and their shared ecological resources (Early et al. 2016; Doherty et al. 2016; Dueñas et al. 2018).

One of the most successful invasive species in recent history is the house mouse, *Mus musculus* (Pimentel et al. 2000; Genovesi et al. 2012). Originating in Asia on the Indian subcontinent, *Mus musculus* split into three parapatrically distributed subspecies, *Mus musculus castaneus, Mus musculus musculus*, and *Mus musculus domesticus*. *M. m. castaneus* is now found primarily in southeast Asia. *M. m. musculus* is currently distributed throughout Northern Asia and Eastern Europe. *M. m. domesticus* was originally distributed across the Middle East, North Africa, and Western Europe, and has since been introduced worldwide in association with humans (Geraldes et al. 2008; Suzuki et al 2013; Didion & Villena 2013; Phifer-Rixey & Nachman 2015; Morgan et al. 2022).

Commensalism between humans and house mice dates to roughly 12,000 – 15,000 years ago (Cucchi et al. 2005; Weissbrod et al. 2017). As humans developed farming practices and adopted more sedentary lifestyles, house mice were able to adapt to a variety of climates and live in close association with humans, taking

advantage of stored grain as a readily accessible food source. From their origins in southwest Asia, zooarchaeological surveys have identified *M. m. domesticus* in the Near East and Eastern Mediterranean within the Neolithic approximately 10,000 years ago, and within the Western Mediterranean and Western Europe during the Iron Age around 3,000 years ago (Cucchi et al. 2005; Cucchi et al. 2020).

Only within the past 1000 years have house mice migrated out of Western Europe (Rosevear 1969; Gabriel et al. 2010; Jones et al. 2013; Phifer-Rixey and Nachman 2015). House mice were spread throughout the North Atlantic as passengers of Viking ships during the tenth century, with *M. m. domesticus* having reached Iceland, Greenland, and possibly Newfoundland and the Azores, and *M. m. musculus* having more recently colonized Greenland (Searle et al. 2009; Jones et al. 2012; Gabriel et al. 2015). House mice from Western Europe likely similarly invaded the Americas as the unintended passengers of human migrants. The late 15th and early 16th centuries are known to have marked the beginning of significant interactions between the eastern and western hemispheres. The widespread exchange of organisms between the hemispheres played a tremendous role in shaping global biodiversity (Crosby 1972, Knight & Lisk 1991; McCusker 2006; Jones et al. 2013). This event likely facilitated the establishment of *M. m. domesticus* populations in the Americas.

However, the demographic details of this recent colonization history have not been explored. In particular, the colonization of new areas may result in founder effects, the loss of genetic variation associated with population bottlenecks. Reductions in genetic variation in North American populations relative to source populations have been documented for diverse species including starlings (Cabe 1998), horses (Luis et al. 2006), and fruit flies (Arguello et al. 2019). In contrast, multiple colonization events from different source populations may result in admixture within the founding populations, leading to increased levels of genetic variation. Such a pattern is seen in brown rats which appear to have colonized North America from Asia multiple times (Puckett et al. 2016). It is not known whether house mouse populations in North America experienced founder effects and associated reductions in levels of genetic variation or admixture and associated increases in genetic variation. Finally, the timing of the colonization of North America by house mice has not been carefully studied. It remains unclear whether the earliest Europeans brought house mice with them, or whether house mouse populations were not established until much later, for example during the industrial revolution when shipping between Europe and North America increased substantially.

To study the demographic history of house mice in eastern North America, we analyzed whole-exome data of mice from five populations in North America and two populations in Europe, focusing on three primary questions. (1) Was the colonization of eastern North America associated with a population bottleneck? (2) Do North American populations show mixed ancestry, as would be expected if multiple colonization events occurred from different source populations? (3) When did house

mice begin to colonize North America? We discovered reduced genetic diversity and fewer rare alleles in North America than in Europe, consistent with a bottleneck. One population of house mice from Florida showed a strong signature of admixture, consistent with multiple introductions of mice to this region. Demographic modeling revealed recent split times between eastern North American populations, consistent with the timing of human migrants from Europe. These results highlight the close association between the movement of mice and humans and provide perspective into one of the world's most successful invasive species.

Materials and Methods

Sample collection and sequencing dataset

We compiled exome sequences of 66 wild house mice from published data representing five populations from eastern North America (Phifer-Rixey et al. 2018) and two populations from Europe (Harr et al. 2016) (Figure 1). The North American samples consisted of 50 mice from five localities along a latitudinal transect: New Hampshire/Vermont (NHVT); Pennsylvania (PA); Virginia (VA); Georgia (GA); and Florida (FL). In brief, DNA was extracted from samples of mice collected from North America and western Europe via DNAeasy kits (Qiagen, Hilden, Germany) or salt extractions. NimbleGen probes were used to capture exomes which were then sequenced at an average depth of ~15X, as described by Phifer-Rixey et al 2018. Sixteen *M. m. domesticus* samples from two European populations (Cologne-Bonn, Germany and Massif Central, France) and 8 samples of *M. spretus* (Spain) were obtained via the European Nucleotide Archive (ENA: PRJEB9450); these were sequenced with an average depth of ~20X (Harr et al 2016). Reads from North America and western Europe were mapped against the mouse GRCm38/mm10 reference genome using novoalign and bwa-mem, respectively (Mu et al. 2012; Li et al. 2013). In all cases, wild mice were caught more than 500 m from one another to avoid sampling related individuals.

Exomes were extracted from whole genome samples using *bedtools* (Quinlan and Hall 2010). The *samtools* 'mpileup' command was used to call variants (Li 2011), filtering away sites with greater than 50% missing data and/or quality scores less than 30. VCF files were then compressed, indexed, and merged at intersecting positions using *vcftools* (Danecek et al 2011). After this stage of filtering, 955,312 SNPs were retained. Further filtering was done using PLINK to remove Hardy-Weinberg outliers and prune for linkage disequilibrium using a window size of 50kb and a r^2 threshold of 0.5. After LD pruning, 395,879 SNP were passed on for downstream analysis. The final VCF file is available on Dryad (https://doi.org/10.6078/D1CX2R).

Population structure

We estimated phylogenetic relationships among individuals using RAxML (Stamatakis 2014). A maximum likelihood phylogeny was constructed using concatenated SNPs representing all populations under the GTRGAMMA model with rapid bootstrapping, and using *M. spretus* as an outgroup.

We conducted principal component analysis (PCA) using PLINK to examine population structure (Purcell 2007). The eigenvectors and eigenvalues associated with the first 20 principal components were derived using PLINK and the first two principal components were plotted in R (R Core Team 2021). ADMIXTURE was used to infer shared ancestral tracts between sample subpopulations represented in the genomic datasets (Alexander 2009). PLINK was used to generate the input bed file for ADMIXTURE from a filtered VCF file containing the seven *M. m. domesticus* populations included in the PCA and one population of *M. spretus*. *M. spretus* was included in all admixture analyses since introgression between *M. musculus* and *M. spretus* is known to occur in Europe (Liu et al. 2015; Banker et al. 2022).

F_{st} was used to quantify the extent of genetic divergence observed between populations in relation to the amount of variation observed within populations. Cockerman-Weir's weighted F_{st} was computed using vcftools '—weir-fst-pop' flag on filtered exomic SNPs from French, German, and eastern North American samples of *M. m. domesticus*.

Nucleotide diversity (π) and Watterson's theta (θ) are estimators of the neutral mutation parameter, $4N_e\mu$, where N_e is the effective population size and μ is the neutral mutation rate. Tajima's D (Tajima 1989) is the normalized difference between π and θ. These estimators, together with the distribution of allele frequencies for all SNPs (the site frequency spectrum), summarize present levels of variation and can be used to illuminate past demographic processes. We estimated π and θ using ANGSD (Korneliussen et al. 2014). ANGSD generates per-site sample allele frequency likelihoods, as well as the folded site frequency spectrum using realSFS. Log posterior probabilities were computed to derive π, θ, and Tajima's D from the folded SFS.

Demographic inference and analysis

We used f_3 statistics to explore the possibility that multiple source populations contributed to present populations of house mice in eastern North America. This statistic is based on a three-population comparison, and negative values of the test statistic provide evidence of admixture. We used the five North American and two European populations of *M. musculus* as well as the population of *M. spretus,* and we calculated f_3 for each of the 56 possible three-way comparisons among these eight samples. f_3 statistics were computed using the 'threepop' command in TreeMix (Reich

et al. 2009, Patterson et al 2012). We used Admixture-induced Linkage Disequilibrium for Evolutionary Relationships (ALDER) to generate test admixture statistics and to infer the approximate time of mixture (Loh et al. 2013). Reported parameter results correspond to simultaneously computed 2-reference and 1-reference models.

We used the Diffusion Approximation for Demographic Inference (*dadi*, Gutenkunst et al. 2009) and Genetic Algorithm for Demographic Model Analysis (GADMA, Noskova et al. 2020) to explore the demographic history of *M. m. domesticus* populations. We estimated two-dimensional SFS from SNP allele frequencies using ANGSD, requiring a minimum mapping quality of 30, base quality score of 20, and the removal of poorly mapped, duplicate, non-unique, or unpaired reads. To mitigate the effects of selection, we restricted demographic analyses to synonymous SNPs, predicted using Variant Effects Predictor, when generating per-site allele frequencies (McLaren et al. 2016).

Folded joint SFS were prepared as input for *dadi* using realSFS with the 'dadi' flag option. We explored a variety of demographic models taking the general form of an ancestral population that gives rise to two descendant populations, with and without migration between them. Parameters in the models included divergence time, population sizes, and migration rates. *dadi* was used to estimate these parameters in a likelihood framework. We used GADMA to carry out a global heuristic search of parameter space via the genetic algorithm (GA) to infer the best fit demographic model. Parameter values in *dadi* are scaled to $\theta = 4N_{anc}\mu$, the neutral mutation parameter as a function of the ancestral effective population size and mutation rate per generation. We used this relationship to convert the output to standard units (e.g. N_e, t in years, and m) using the per-site mutation rate $\mu = 4 \times 10^{-9}$ and a generation time of 1 year (Geraldes et al 2008). To generate confidence intervals for the parameter estimates, we divided each dataset into non-overlapping 10kb sections and simulated pseudo-replicate datasets by sampling with replacement for bootstrapping under the parameters that generated the maximum likelihood for the best fitting model. The standard deviations for all parameters were derived from uncertainty analysis using the Godambe Information Matrix for composite likelihoods.

Results

Population structure

Phylogenetic reconstruction with RAxML was used to uncover the relationships among mice from Europe and eastern North America (Figure 2a). Using *M. spretus* as an outgroup, mice from France were paraphyletic with respect to the remaining populations in the phylogeny. Mice from Germany formed a monophyletic group, and this was the sister group to mice from North America. Among the samples

from North America, all populations except Florida formed monophyletic groups. Mice from Florida were paraphyletic with respect to the remaining North American populations.

Samples from Germany, France, and eastern North American formed separate clusters when plotted with the first two principal components which together accounted for 29% of the variation (Figure 2b). Among the North American samples, individual localities largely formed distinct clusters, although the positions of these clusters did not correspond with the geographic distance between populations. The Florida population was unique in clustering farther from other North American samples.

To further examine the structure of populations, we used ADMIXTURE to infer shared ancestry proportions. When modeling for distinct ancestral clusters at K = 7 individuals from all populations except Florida showed little to no admixture (Figure 2c). In contrast, all individuals from the Florida population showed high levels of admixture. Notable levels of admixture were inferred in the Florida population across all values of K (Supplementary Figure 1).

The average F_{st} among populations in North America was 0.092 while F_{st} between the two European populations was 0.168, despite the fact that the European populations are separated by ~1,000 km while the most distant North American populations are separated by ~2,000 km. Pairwise F_{st} among all populations ranged from 0.041 to 0.281, with the lowest levels of differentiation seen between Florida and Pennsylvania and the greatest levels of differentiation seen between France and New Hampshire/Vermont (Table 1).

Reduced genetic diversity and the distribution of allele frequencies suggest a bottleneck associated with the founding of M. m. domesticus populations in eastern North America

A reduction in genetic diversity in a derived population relative to the source population is expected following a recent population bottleneck. We estimated per-site nucleotide diversity (π) and Watterson's theta (θ) from allele frequencies computed in ANGSD. The populations from Germany and France had substantially higher levels of nucleotide diversity (average π = 0.28%) than any of the North American populations (average π = 0.17%) (Table 2). Similarly, populations from Germany and France had a higher proportion of segregating sites (average θ = 0.26%) than any of the North American populations (average θ = 0.15%) (Table 2).

Population bottlenecks are also expected to lead to a loss of rare alleles, thus skewing the shape of the allele frequency spectrum in comparison to source populations. The proportion of segregating sites was consistently lower than pairwise nucleotide diversity, leading to a positive Tajima's D for all populations. However, Tajima's D was consistently larger in populations from eastern North America than in European populations (Table 2), indicating a greater scarcity of rare alleles in the

American populations. This pattern can also be seen in the folded site frequency spectra for the individual populations (Supplementary Figure 2). In particular, the North American populations show more intermediate frequency alleles (relative to neutral expectations) compared to the European populations.

Evidence of admixture in house mice from Florida

To test the hypothesis that some North American populations may have arisen from multiple introductions, we used three-population *f* statistics. We computed all 56 (8 choose 3) possible f_3 statistics using the seven population samples of *M. m. domesticus* and the single population sample of *M. spretus*. Of these 56 tests, eight returned a negative value, all involving the Florida population, suggesting that this population is admixed (Table 3). Four comparisons suggest that allele frequencies in Florida are intermediate between those of populations from France and populations from the remaining four North American populations (Table 3). The remaining four comparisons suggest that allele frequencies in Florida are intermediate between those of *M. spretus* and the remaining four North American populations. While it is unclear whether the sampled populations are the direct contributors to the admixture observed in Florida, a negative f_3 statistic implies that a phylogenetic tree for the given populations is a poor fit to the data without admixture along an internal branch. The inference of admixture in the Florida population is consistent with several other observations, including the fact that this population has the highest nucleotide diversity in North America (Table 2), the lowest levels of differentiation from the other sampled populations (Table 1), is phylogenetically paraphyletic with respect to the other North American populations (Figure 2a), and appears admixed in other analyses (Figure 2c).

We further investigated the details of admixture in Florida using ALDER. ALDER estimates the dates of plausible admixture events while also providing a formal 2-reference weighted LD statistic to test if the two potential source populations are the probable contributors to the admixed population. This analysis returned no significant 2-reference weighted LD statistics using any two pairs of populations as the contributing populations for Florida's admixture (Table 3). We did, however, observe significant 1-reference weighted LD statistics using France and New Hampshire/Vermont ($p = 0.012874$ and $p = 0.001035$, respectively). From the decay rate, we estimated a time to admixture from France (or a related population) of 102.08 ± 45.83 years, assuming 1 generation per year. We estimated a time to admixture from New Hampshire/Vermont (or a related population) of 87.25 ± 19.31 years.

Demographic modeling suggests a recent introduction of M. m. domesticus to eastern North America

We used *GADMA* to infer divergence times and effective population sizes under a variety of two-population models for all pairs of populations from Germany, France,

and North America. First, we estimated demographic parameters between Germany and France. Next, we compared each North American population with each European population, constraining the inferred divergence time to the recent history during which house mice are known to have emerged from the Middle East (past 7,500 years). Finally, we compared North American populations with each other, again using a constraint on divergence time of 7,500 years. In all cases, we compared results under a simple divergence model without migration to a divergence model including migration (Figure 3). Since these models are nested, we used a likelihood ratio test (LRT) to determine if the inclusion of migration improved the fit of the models. The standard deviations of the estimated parameters in these models were generally large (Figure 3). Thus, the estimates of parameter values should not be taken as precise; nonetheless several clear patterns emerged.

The estimated divergence time under a simple split model without migration comparing Germany and France was 2,584 ± 11 ya. This is consistent with zooarchaeological evidence which suggests that *M. m. domesticus* spread through western Europe during the Iron Age, which ended inside of 2.5 kya (Cucchi et al 2005; Jones et al. 2013). When migration was incorporated into the model, a similar divergence time was inferred (2,871 ± 1,444 ya) and the inferred migration rates were zero.

The estimated divergence times under two-population split models when comparing Germany and populations from North America were between 1-7 kya (Figure 3a), within the range of the late Neolithic/Copper Age and late Iron Age, and overlapping in confidence interval with the divergence time between Germany and France. Inference of demographic parameters under two-population modeling for France and eastern North American populations failed to converge under the specified time constraints for divergence time. Nonetheless, the deep age estimates between German and North American populations predate the known arrival of European settlers to the Americas by thousands of years, suggesting that the source populations for North American mice were not from France or Germany.

In these models, the estimated effective population sizes for North American populations were generally smaller than those of the European populations (Figure 3a), consistent with a bottleneck during the founding of most North American populations. The one exception to this pattern involved the Florida population, which revealed a higher effective population size when compared to the population from Germany. The higher estimated N_e most likely reflects admixture in Florida, as documented above.

Finally, we considered models with pairs of North American populations. Under a simple split model without migration, inferred divergence times were mostly within the last 500 years, corresponding to the timing of human colonization (Figure 3b). Likelihood ratio tests almost uniformly supported divergence models without migration. The one case where the LRT favored a model with migration was the Virginia- Pennsylvania comparison ($p = 0.0143$) (Supplementary Table 1). In that

case, the model with migration also suggested a divergence time of 300 years ago. Similarly, the LRT narrowly failed to reject a model without migration for New Hampshire-Vermont and Pennsylvania ($p = 0.0566$), and a model with migration suggested a divergence time (856 years ago) whose confidence interval substantially overlaps the divergence times estimated from other populations pairs (Figure 3b, Supplementary Table 1). In sum, these models indicate that pairs of North American populations share ancestry within the time frame of European colonization of the Americas.

Discussion

Here, we connect the invasion history of house mice in the eastern United States to the colonization of the Americas by European settlers which began in the 16th century. Previous genetic studies on the recent colonization history of house mice have mainly used mitochondrial DNA and microsatellites and have focused on the colonization of islands in the Atlantic including Iceland, Greenland, New Foundland, Gough Island, and the Azores (Jones et al. 2012; Gray et al. 2014; Gabriel et al. 2015). These studies revealed close associations between human and mouse colonization histories, but did not address the connections between US and European house mice.

Using genome-wide data, we found that North American populations of house mice underwent a bottleneck compared to European populations. We discovered that one population in Florida shows evidence of considerable admixture, suggesting that house mice may have colonized parts of North America from distinct sources. Finally, we found that divergence times among North American populations of house mice correspond roughly to the timing of colonization by Europeans. Below we discuss each of these in turn.

Evidence for a bottleneck comes from several sources. First, levels of genetic variation are substantially lower in North American populations of house mice than in the European populations. For example, the average nucleotide diversity in North America ($\pi = 0.17\%$) was only 60% of that seen in European populations ($\pi = 0.28\%$). Second, Tajima's D was more positive in North American populations than in European populations, consistent with the loss of rare alleles. Finally, demographic modeling using *dadi* revealed smaller population sizes for most North American populations compared to the German populations (Figure 3b), with the one exception being the population from Florida, as discussed below. In aggregate, these results indicate that the founding of North America was accompanied by a modest bottleneck.

From the 16th century onward, numerous ships from various parts of Western Europe landed on the shores of eastern North America. Many intentionally transported non-native species, such as horses, cattle, pigs, sheep, rock pigeons, house sparrows, and starlings (Robbins 1973; Cabe 1998; Guiffra 2000; Luis et al.

2006; Ajmone-Marsan et al. 2010; Sydney et al. 2012; Bertolini 2018). Several ships also unintentionally harbored invasive species, such as rats, mice, zebra mussels, and fruit flies (Duffy 1951; Astanei et al. 2005; Aplin et al. 2011; Jones et al. 2013; Puckett et al. 2016; Mallez & McCartney 2018; Arguello et al. 2019). In many cases, the resulting founder populations experienced modest bottlenecks through their introductions (Cabe 1998; Luis et al. 2006; Puckett et al. 2016; Bertolini 2018; Arguello et al. 2019).

Evidence of multiple colonization events from distinct sources comes from observations of admixture in the Florida population (Figure 2c, Table 3). Lack of significance among all 2-reference weighted LD statistics using ALDER suggests that the Florida population is not the direct result of admixture from any of the populations included in this study (Table 3). Florida has a complex colonization history involving Spanish and British rule (Hudson 1997; Glanville 2009), and it is possible that both groups of people introduced house mice to the region. The Spanish first explored Florida in 1513, when Juan Ponce de Leon recorded his exploration of the peninsula. The Spanish ruled over Florida until 1763, when the British overtook Spanish rule for 20 years, before the Spanish reestablished sovereignty over Florida in 1783, though many British and American settlements in Florida remained. The second Spanish occupation of Florida lasted until 1821, at which point, Florida became an organized colony of the United States (Knight and Liss 1991; Hudson 1997; McCusker 2006; Glanville 2009). Future studies with better sampling across Europe would help to identify the possible source populations contributing to admixture in Florida.

To ask whether divergence times of mouse populations reflect the timing of human settlements, we estimated divergence times between each pair of populations using *dadi*. Direct comparisons of eastern North American populations with one another largely yielded split times within the past 500 years, consistent with the known period of European colonization in the Americas. Inferred divergence times between Germany and eastern North America fell within the range of the late Neolithic to Iron Age period (5,000 BC – 1000 AD) (Cucchi et al. 2005; 2012; 2020). Similarly, the divergence time inferred between Germany and France was around 2.5 kya. Consistent with zooarchaeological evidence, the late Bronze Age to Iron Age are the time periods when house mice progressed toward western Europe from the Middle East. These observations suggest a scenario in which mice in eastern North America came from an unsampled population in Western Europe. The unsampled population likely diverged from the German and French populations sometime in the last few thousand years (Figure 4).

Human colonization history suggests that source populations for North American mice are likely to have come from northern Europe (Knight and Liss 1991; McCusker 2006). Virginia is known to have included the first enduring British settlement of the Americas in Jamestown, beginning in 1607. Pennsylvania became a British colony in 1681. Prior to the British, the Dutch and Swedish had established colonies in Pennsylvania early in the 17th century. New Hampshire was also one of

the original thirteen colonies, with the first British settlements established around 1623. For most of pre-American colonial history, New York and New Hampshire had split claim over the territory that is now the state of Vermont. Georgia, named after King George II of Great Britain, became the last of the thirteen colonies in 1752. Given the predominance of British rule through much of the recent colonization history of America, England is a reasonable source location for North American mice. Moreover, British migration through the eastern seaboard in establishing the thirteen colonies may have facilitated the spread of house mice from initial sites of introduction.

Despite the strong British influence, the Spanish were the first Europeans to establish consistent transatlantic exchanges with North America and thereby had the earliest opportunities to introduce non-native organisms. A recent study by Morgan et al. (2022) used a SNP genotyping array to assess population structure across Europe and elsewhere and found that mice in Europe consist of a northern clade (including Germany, Denmark, Belgium, and Scotland) and several southern groups (including Greece, Italy, Portugal, and Spain). Their study did not include samples from England, but one interesting possibility is that mice in eastern North America derive primarily from England, but that mice in Florida may represent a mixture of mice from England and mice from southern Europe. This hypothesis could be tested by deep sampling of mice across Europe.

Conclusion

We conclude that house mice in the eastern United States arrived and dispersed largely within the past 500 years. Their arrival was associated with a genetic bottleneck, as US populations harbor only 60% of the average genetic variation seen within European populations. Since the arrival of the initial source population, at least one other source population successfully mixed with an established population in Florida. Notably, Florida was also colonized by different Western European settlers within the past few centuries. The overall similarity in the estimated divergence times of North American populations of house mice with the known timing of human colonization of eastern North America suggests a very close association between the movements of humans and mice and suggests that the earliest European settlers may have brought mice with them.

Figures

Figure 1.1: Locations of house mouse samples from eastern North America and Western Europe. Map showing locations of sampled mice from New Hampshire/Vermont (green), Pennsylvania (purple), Virginia (pink), Georgia (brown), Florida (gold), Germany (red) and France (blue).

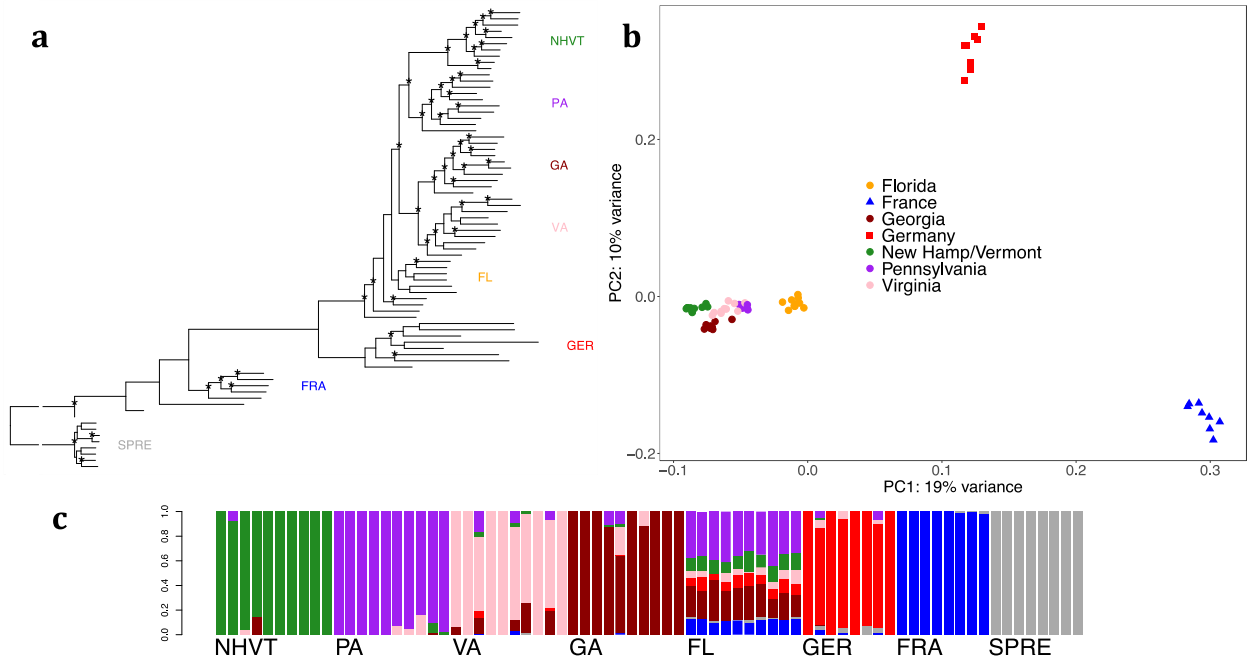
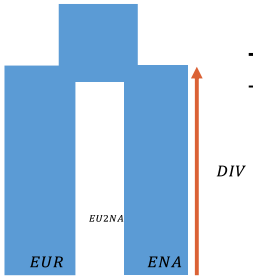


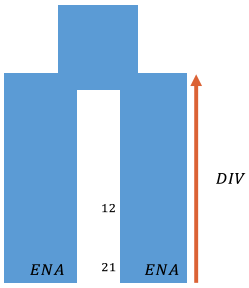
Figure 1.2: Analysis of population structure of eastern North American and Western European house mice. (a) Phylogenetic tree of *M. m. domesticus* constructed using RAxML with *M. spretus* as an outgroup. *M. m. domesticus* are from Florida (FL, n=10), Georgia (GA, n=10), Virginia (VA, n=10), Pennsylvania (PA, n=10), New Hampshire/Vermont (NHVT, n=10), Germany (GER, n=8), and France (FRA, n=8). Asterisks indicate nodes with > 80% bootstrap support. (b) Principal component analysis of 66 wild-caught *M. m. domesticus*. (c) Inferred admixture proportions for K = 7 distinct ancestral subpopulations for eastern North American, German, and French *M. m. domesticus* populations, and an outgroup *M. spretus* population.

a



	N	± sd	N	± sd	T	± sd	m	± sd
GER-NHVT	15,889	5,562	1,352	655	1,365	179	5.84e-05	5.57e-05
GER-PA	12,549	7,189	8,269	6,937	3,412	2,144	0	0
GER-VA	20,432	4,594	15,282	1,913	6,412	5,511	1.24e-05	6.11e-05
GER-GA	15,770	19,832	8,424	12,842	4,415	5,907	0	0
GER-FL	22,137	10,391	36,599	4,469	7,056	3,098	0	0

b



	N	± sd	N	± sd	T	± sd	m	± sd	m	± sd
FL⊟NHVT	123,585	534,609	944	41	397	27	0	0	0	0
FL⊟PA□	85,622	193,396	885	59	189	8	0	0	0	0
FL⊟VA	14,178	5,755	1,368	29	302	30	0	0	0	0
FL⊟GA	140,956	539,165	1,037	61	307	15	0	0	0	0
PA⊟NHVT	5,101	4,792	2,469	2,879	856	610	1.51e-4	1.39e-4	4.63e-7	2.55e-5
PA⊟VA	1,727	4,473	1,714	4,131	330	466	1.28e-6	2.04e-5	3.5e-4	2.76e-4
PA⊟GA	3,247	576	1,887	42	488	62	0	0	0	0
VA⊟NHVT	4,303	245	1,723	248	611	12	0	0	0	0
VA⊟GA	5,459	514	2,480	118	431	56	0	0	0	0
GA⊟NHVT	2,580	125	1,701	154	549	19	0	0	0	0

Figure 1.3: Two-population split modeling using Europe and eastern North American populations. (a). The models depict a split occurring at time T_{DIV} years ago, leading to present-day populations of N_{EUR} and N_{ENA}, allowing for continued unidirectional migration from Europe to eastern North America. (b) The model is similar to (a) but allowing for bidirectional migration, reported as parameters m_{12} and m_{21}. Best fit demographic parameters inferred under these two divergence models are reported to the right.

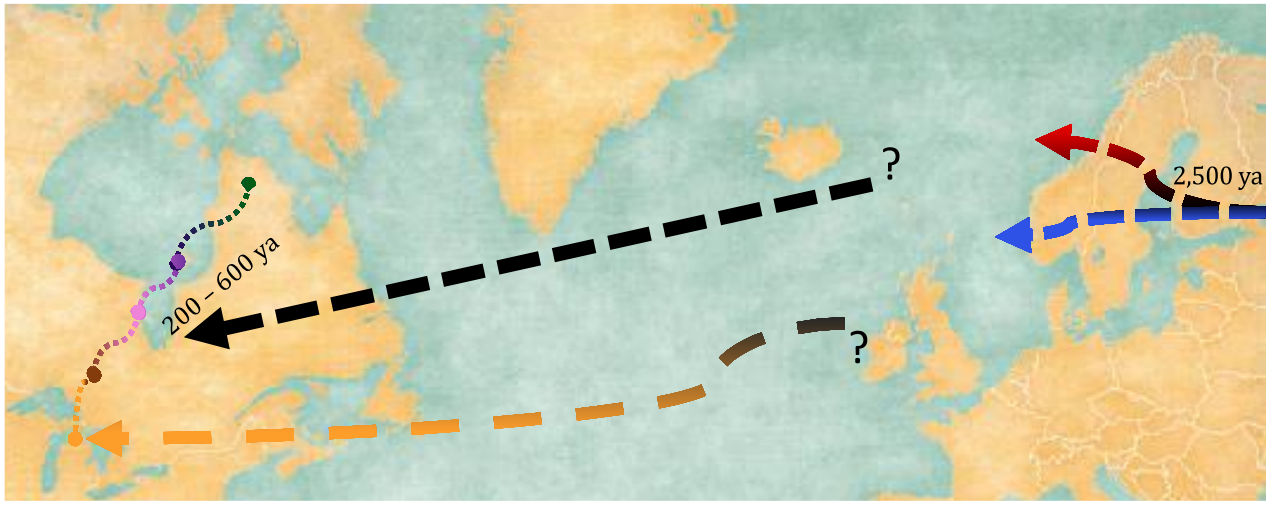


Figure 1.4: Recent progression of house mice from Western Europe to eastern North America. Hypothesized colonization history of house mice in eastern North America showing divergence times as inferred from demographic modeling. Black dashed arrow illustrates migration of house mice from Western Europe to eastern North America. Orange dashed arrow illustrates additional migration from Western Europe to Florida. Lines connecting North American populations illustrate possible spread across eastern coast of North America during founding period. Dashed arrows in Europe depict ancestral split between European populations from Cologne-Bonn, Germany and Massif Central, France. Colors match Figures 1 & 2.

Table 1.1: Pairwise Fst comparisons between two western European population (France and Germany) and five eastern North America population samples.

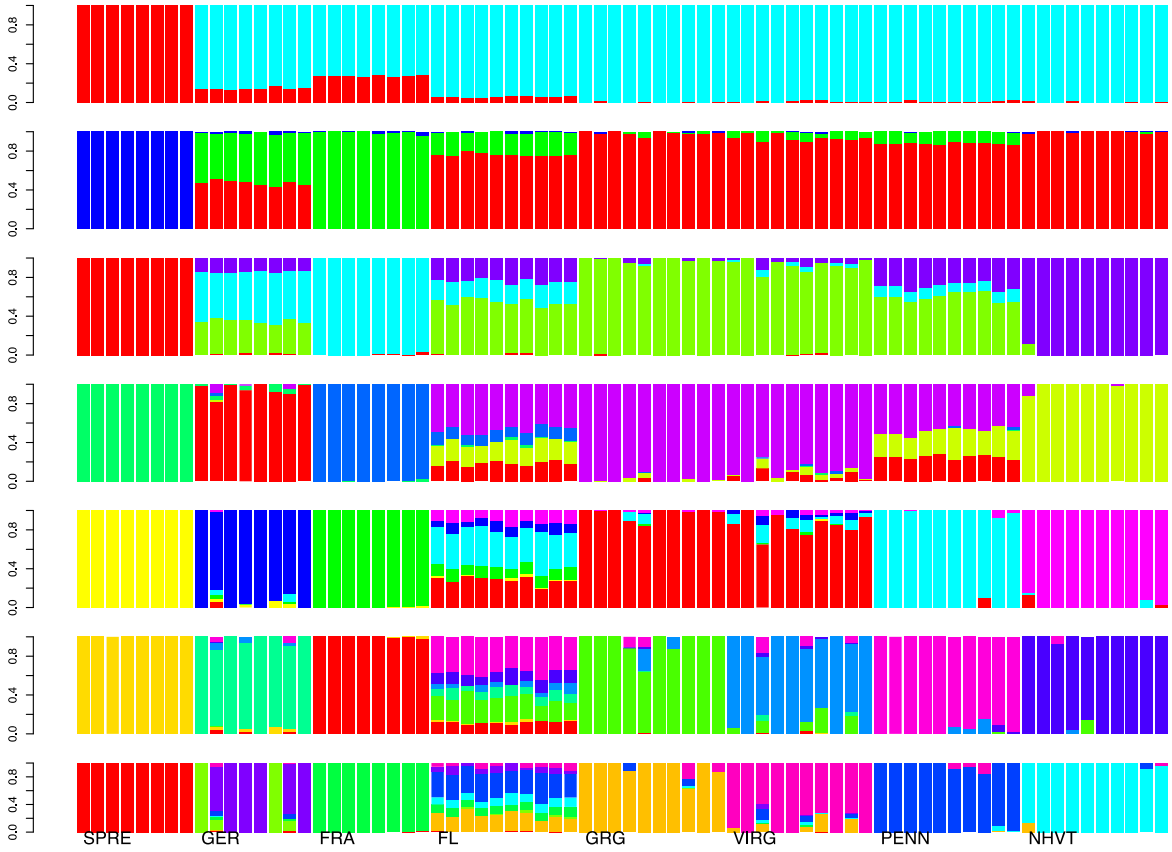
	France	Germany	New Hampshire/ Vermont	Pennsylvania	Virginia	Georgia
Germany	0.168					
New Hampshire/ Vermont	0.281	0.202				
Pennsylvania	0.215	0.143	0.124			
Virginia	0.222	0.146	0.135	0.087		
Georgia	0.237	0.170	0.159	0.106	0.056	
Florida	0.161	0.101	0.096	0.041	0.054	0.063

Table 1.2: Summary of genetic diversity in sampled *M. m. domesticus* populations.

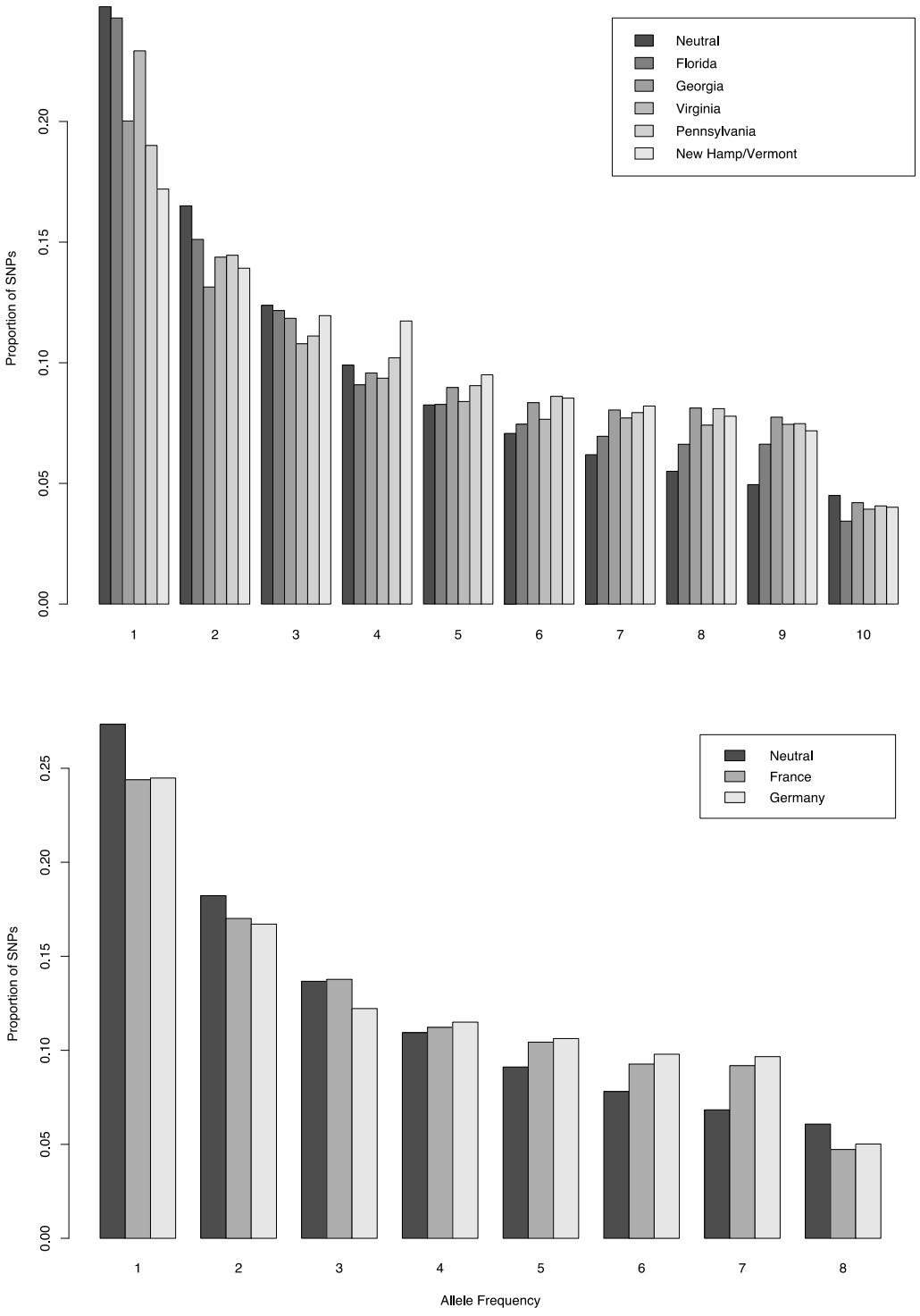
POPULATION	AVG PER SITE WATTERSON'S THETA (%)	AVG PER SITE NUCLEOTIDE DIVERSITY (%)	TAJIMA'S D
France	0.269	0.289	0.438
Germany	0.252	0.271	0.440
New Hampshire/ Vermont	0.121	0.140	0.754
Pennsylvania	0.149	0.169	0.649
Virginia	0.155	0.170	0.472
Georgia	0.133	0.151	0.641
Florida	0.188	0.205	0.453

Table 1.3: Summary and admixture statistics and inferred demographic parameters. Negative f_3 statistic indicating admixture, standard error, and associated z-score were computed with TreeMix. Remaining columns are inferred demographic parameters generated using ALDER.

Population A (Target)	Population B	Population C	F_3 statistic	s.e.	Z-score	Pop B 1-ref decay	Pop B 1-ref amp_exp	Pop B 1-ref z-score	Pop C 1-ref decay	Pop C 1-ref amp_exp	Pop C 1-ref z-score	2-ref decay	2-ref amp_exp	2-ref z-score	2-ref p-value (test status)
Florida	France	New Hampshire/ Vermont	-5.62e-3	6.29e-4	-8.92869	102.08 ± 45.83	3.91e-4 ± 1.65e-4	2.23	87.25 ± 19.31	1.46e-4 ± 4.72e-5	3.08	154.42 ± 106.08	1.59e-3 ± 2.45e-3	0.65	0.52 (failure)
Florida	France	Pennsylvania	-1.66e-3	4.84e-4	-3.43814	102.08 ± 45.83	3.91e-4 ± 1.65e-4	2.23	2.00 ± inf	-3.66e-6 ± inf	0	16.49 ± 57.05	8.97e-5 ± 2.58e-4	0.29	0.77 (failure)
Florida	France	Virginia	-1.01e-3	5.06e-4	-1.98888	102.08 ± 45.83	3.91e-4 ± 1.65e-4	2.23	13.46 ± 28.09	1.66e-5 ± 1.96e-5	0.48	2.67 ± inf	6.76e-5 ± inf	0	1 (failure)
Florida	France	Georgia	-2.3e-3	5.33e-4	-4.31575	102.08 ± 45.83	3.91e-4 ± 1.65e-4	2.23	2.00 ± inf	-3.83e-6 ± inf	0	4.82 ± inf	5.25e-5 ± inf	0	1 (failure)
Florida	*M. spretus*	New Hampshire/ Vermont	-8.57e-3	1.1e-3	-7.79405	2.00 ± inf	1.42e-3 ± inf	0	87.25 ± 19.31	1.46e-4 ± 4.72e-5	3.08	2.00 ± inf	1.01e-3 ± inf	0	1 (failure)
Florida	*M. spretus*	Pennsylvania	-6.14e-3	8.32e-4	-7.38383	2.00 ± inf	1.42e-3 ± inf	0	2.00 ± inf	-3.66e-6 ± inf	0	2.00 ± inf	1.26e-3 ± inf	0	1 (failure)
Florida	*M. spretus*	Virginia	-3.91e-3	8.51e-4	-4.59203	2.00 ± inf	1.42e-3 ± inf	0	13.46 ± 28.09	1.66e-5 ± 1.96e-5	0.48	2.00 ± inf	1.36e-3 ± inf	0	1 (failure)
Florida	*M. spretus*	Georgia	-4.36e-3	8.87e-4	-4.91237	2.00 ± inf	1.42e-3 ± inf	0	2.00 ± inf	-3.83e-6 ± inf	0	2.00 ± inf	1.05e-3 ± inf	0	1 (failure)



Supplementary Figure 1.1: ADMIXTURE analysis for K = 2 through 8.



Supplementary Figure 1.2: Site frequency spectra comparison to expectation under neutral equilibrium. Top: Site frequency spectra for the five eastern North American populations. Bottom: Site frequency spectra for the populations of Germany and France.

Supplementary Table 1.1: Results from likelihood ratio test. The first two columns give the reported maximum log composite likelihood from two divergence models. The third column is the Godambe adjusted D-statistic reported from the likelihood ratio test. The fourth column is the LRT associated p-value.

	MLCL (w/o migration model)	MLCL (w/ migration model)	Adjusted D-statistic	p-value
GER-FRA	-7138	-7171	-0.4551	1.0000
GER-NHVT	-6821	-6783	0.2276	0.3167
GER-PENN	-5519	-5518	0.0609	0.4025
GER-VIRG	-5619	-5614	0.3273	0.2836
GER-GEOR	-6033	-6026	0.1484	0.3500
GER-FL	-4778	-4782	-0.6041	1.0000
FL-NHVT	-3477	-3210	0.8440	0.1791
FL-PENN	-1847	-1592	1.9043	0.0838
FL-VIRG	-1979	-1861	0.8648	0.1762
FL-GRG	-2431	-2131	1.4598	0.1135
VIRG-NHVT	-3555	-3589	-1.2546	1.0000
VIRG-PENN	-2386	-2379	4.7886	0.0143
VIRG-GRG	-1886	-1708	2.2615	0.0663
PENN-NHVT	-3338	-3336	2.5081	0.0566
PENN-GRG	-2856	-2918	-4.1012	1.0000
GRG-NHVT	-3786	-3863	-12.9825	1.0000

Chapter 2

Genetic structure and demographic history of house mice in Western Europe inferred using whole genome sequences

Introduction

Detailing the history of demographic and evolutionary processes that has shaped modern populations is a fundamental goal in evolutionary and ecological studies. The house mouse (*Mus musculus*) is a premier mammalian model organism for biological research that has gained a near global distribution through their commensalism with humans (Phifer-Rixey and Nachman 2015). Despite its standing as a widely used model, important details regarding the genetic structure and demographic history of house mice are still unknown. House mice consist of three major subspecies which diverged from an ancestral population in the northern Indo-Pakistani subcontinent and neighboring regions roughly 200,000–500,000 years ago (Salcedo et al. 2007; Geraldes et al. 2008; Duvaux et al. 2011; Geraldes et al. 2011; Bonhomme and Searle 2012; Suzuki et al. 2013; Phifer-Rixey et al. 2020; Fujiwara et al. 2022): *Mus musculus castaneus* in Southeast Asia, *Mus musculus musculus* in Eastern Europe and Northern Asia, and *Mus musculus domesticus* in the Middle East and Western Europe. In association with humans, *M. m. domesticus* extended its geographic range to Africa, Australia, and the Americas during the last several centuries, presumably as a consequence of human colonialism (Rosevear 1969; Searle et al. 2009; Gabriel et al. 2010; Jones et al. 2013; Suzuki et al. 2013; Phifer-Rixey and Nachman 2015; Lippens et al. 2017; García-Rodríguez et al. 2021).

Much of our current understanding of the history of house mice in Western Europe comes from the zooarchaeological record (Cucchi et al. 2005; Cucchi 2008; Cucchi et al. 2020). This record suggests that human–house mouse commensalism dates to between 12,000 – 15,000 years ago (Cucchi et al. 2005; Weissbrod et al. 2017; Cucchi et al. 2020). Some authors have argued that house mice have since become an anthrodependent species (Hulme-Beaman et al. 2016). The introduction of early farming practices is thought to have enabled the cohabitation of humans and house mice in the Near East during the Neolithic. Despite notable human migration from the Near East to Western Europe during the Neolithic, the spread of house mice was limited to movement from the Near East to the eastern Mediterranean and Anatolia around this time. Following the Neolithic, the next notable period of diffusion occurred in the late Bronze and early Iron Age, with rapid spread across the western

Mediterranean. Increased maritime transport is suspected to have played a major role (Cucchi et al. 2005; Cucchi et al. 2020). The first direct evidence of house mice found on ships was obtained from the shipwreck of a Late Bronze Age vessel found off the coast of Turkey (Cucchi 2008). The most recent zooarchaeological survey dates the arrival of *M. m. domesticus* in Southern Europe around 4,000 years ago, and in Western and Northern Europe around 3,000 years ago (Cucchi et al. 2020).

A small number of studies have used genetic data from contemporary populations to analyze present-day population structure and reconstruct the diffusion history of house mice in Western Europe. For example, previous studies have identified "Mediterranean-like" and "Northern European-like" clusters with datasets based on mitochondrial DNA or microarrays using pre-ascertained single nucleotide polymorphisms (SNPs) (Suzuki et al. 2013; Morgan et al. 2022). A study using ancient mtDNA from the British Isles provided evidence for a Late Bronze to Iron Age colonization (García-Rodríguez et al. 2021). Whole-genome sequences provide an opportunity to more completely survey variation across the genome. They also provide data with which to estimate evolutionary relationships, patterns of historic gene flow, divergence times, and ancestral population sizes, all of which will generate a context for understanding the subsequent spread of house mice worldwide.

Here, we assemble 83 whole genome sequences of wild-caught *M. m. domesticus*, including 59 new whole genome sequences of mice from England, Scotland, Wales, Guernsey (an island in the English Channel just off the coast of northern France), northern France, Italy, Portugal, and Spain. We present relevant metrics on the newly sequenced genomes, and then use these data to address (1) the genetic structure of house mice in Western Europe including the extent to which house mice cluster by geographic location and the genetic similarity of mice within groups versus between groups, (2) the phylogenetic relationships of populations, (3) the amount of gene flow between populations of house mice in Western Europe, (4) historical changes in population size, and (5) divergence times between lineages. We find that mice in Western Europe form genetic clusters that generally match human genetic clusters, and the timing of divergence between these mouse clusters is consistent with human migration history.

Materials and Methods

Samples and molecular methods

We assembled 83 whole genome sequences of *M. m. domesticus* sampled from Western Europe and Iran. This included 24 samples from a previously published dataset representing populations from southern France, Germany, and Iran (Harr et al. 2016). The remaining 59 samples are newly sequenced genomes of wild-caught mice collected in northern France, Italy, northeast Spain (Catalonia), southwest Spain, Portugal, England, Wales, Scotland, and Guernsey. Specimen ID, sex, and

collecting locality for each individual are given in Table 1 and Supplementary Table 1. Sampling of mice took place between 1987–2008, and the majority of mice from these localities were caught more than 500 meters apart to avoid sampling related individuals. Animals were sacrificed in the field and a subset of the mice were prepared as museum specimens (Supplementary Table 1). Liver was collected and stored in ethanol or flash frozen in liquid nitrogen.

DNA was extracted using a MagBind Blood and Tissue DNA extraction kit (Omega Biotek). DNA quantity was assayed on a Qubit Fluorometer v.2 (Thermo Fisher Scientific) and relative fragment size was assessed through agarose gel electrophoresis. A portion of each extraction was diluted to 10 ng/μL at a volume of 110 μL using 10 mM Tris elution buffer, pH 8. The aliquot was sonicated on a qSonica Q800R sonicator at 40% amplitude and 15s on/off pulse for 7.5 minutes of active sonication time for extractions that were fully intact. A double-sided SPRI bead cleaning process was used to size-select fragmented DNA to ~300-500 bp and to concentrate to a volume of 12.5 μL. The enzymatic steps of library preparation followed the Kapa Hyper Prep (Roche Diagnostics) protocol. A universal stub adapter was ligated, and then was extended to full length during amplification with TruSeq-style unique dual-indexing oligos provided by the Functional Genomics Laboratory (University of California, Berkeley). Samples were assessed for sizing on a Bioanalyzer DNA 1000 chip (Agilent Technologies) and quantified using qPCR with Kapa biosystems standards (Roche Diagnostics). At the Vincent J. Coates Genomics Sequencing Lab, 20-24 libraries were pooled together per lane and sequenced on Illumina NovaSeq S4 to collect paired-end 150 bp data (QB3 Genomics, UC Berkeley, Berkeley, CA, RRID:SCR_022170). All sequence data generated in this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive under accession BioProject ID (PRJNA1050608).

SNP discovery

Sequencing reads were processed using a standardized pipeline, following the general procedures and parameters outlined in Harr et al (2016). Raw Illumina sequencing reads were first cleaned and preprocessed using FastP (Chen et al. 2018), then aligned to the mouse reference genome (GRCm38/mm10, RefSeq: GCF_000001635.20) using BWA-MEM (Li 2013). Following GATK4 Best Practices, we marked and filtered PCR duplicates using PicardTools, performed indel realignment to correct SNPs flanking regions containing insertion and deletions, and applied base quality score recalibration to mitigate the propagation of systematic errors from inaccurate base calling. The final recalibrated bams were merged by population using samtools 'mpileup' (Li 2011). All population vcfs were merged using 'bcftools merge' (Danecek et al. 2021), resulting in 81,013,753 total SNPs, and is available on Dryad (https://doi.org/10.5061/dryad.s1rn8pkgh).

Alignment statistics were generated using samtools (Li et al. 2009; Danecek et al. 2021). Specifically, total reads passing quality control and the percent of mapped

reads were computed using 'samtools flagstat', and average per-base autosomal and sex coverage was computed using 'samtools depth'. Prior to downstream analyses, we additionally used VCFtools v0.1.15 (Danecek et al. 2011) to extract autosomes and filter out sites with quality scores lower than 30, variants with greater than 10% missing genotypes, and a minor allele frequency of less than 5%, retaining 3,548,561 sites.

Carriers of *t*-haplotypes were determined by measuring heterozygosity on chromosome 17 using 'vcftools –het' flag. The *t*-haplotypes are associated with inversions on chromosome 17 that suppress recombination and therefore result in high heterozygosity in carriers (i.e. *t*-haplotype heterozygotes) (Keleman and Vicoso 2018; Keleman et al. 2022). Vcftools computes heterozygosity as a function of F, the inbreeding coefficient. A negative F value indicates an excess of heterozygotes relative to expectation. Vcftools' '—relatedness2' flag was also used to identify first degree relatives based on the method of Manichaikul et al. (2010).

Population structure

Phylogenetic relationships of populations were estimated from a maximum likelihood tree constructed using RaxML (Stamatakis 2014). The maximum likelihood tree was constructed first using RaxML under a GTRGAMMA model with bootstrapping and correcting for ascertainment bias using the Lewis model. Analysis of population structure was based on principal component analysis derived from PLINK and proportions of shared ancestry estimated using ADMIXTURE (Purcell 2007; Alexander 2009). The optimal number of distinct ancestral source populations (K) for ADMIXTURE was selected using mean cross-validation error. For phylogenetic analyses and analyses of population structure, autosomal sites were extracted and pruned for pairwise linkage disequilibrium above 0.1 (PLINK v.1.9; –indep-pairwise 50 50 0.1). Following pruning, 659,303 sites were retained.

ADMIXTURE analyses suggested that gene flow had occurred between some populations (see Results). To further study gene flow, we used the topology inferred with RaxML as a newick population tree to compute *f*-branch statistics using D-Suite (Malinksy et al. 2021). To get a better sense of the degree of shared drift between any two populations, we computed outgroup f_3 statistics. Outgroup f_3 statistics measures similarity in two populations by the degree of shared genetic drift relative to an outgroup. Normalized f_3-statistics were computed using ADMIXTOOLS qp3Pop, specifying samples from Iran as the outgroup (Patterson et al. 2012).

Population size and divergence times

We used smc++ (vers. 1.15.2) to estimate the history of effective population size for each European population and divergence times between each pair of populations using default parameters (Terhorst et al. 2017). For each population, biallelic autosomal sites were extracted using vcftools options --min-alleles 2 and --max-alleles

2. Only one individual from each related set were retained for demographic inference. *vcf2smc* was called separately for each chromosome and stored in population specific data files. For each data file, cross validated historical population sizes were estimated using the smc++ *cv* command with the --fold option set to the number of provided population samples. Split times were inferred using the smc++ *split* command on the consensus best k-fold cross-validated parameter estimates provided by smc++ *cv* output. The mutation rate was set to 4×10^{-9} per generation per base pair, and we used a generation time of 1 year (Geraldes et al. 2008). In temperate climates, most house mice breed seasonally, although some commensal populations can breed year-round depending on the availability of food. We used a conservative estimate of one generation per year based on the assumption of seasonal breeding, as done in Gray et al. (2014), Phifer-Rixey et al. (2020), and Fujiwara et al. (2022). Estimates of divergence time are dependent on the estimates of mutation rate and generation time and thus should be interpreted accordingly.

Results

Genome sequences

We sequenced whole genomes of 59 mice from Italy, France, Portugal, Spain, England, Scotland, Wales, and Guernsey to an average depth of 11x and aligned reads to the GRCm38/mm10 reference genome. The average depth of coverage on the autosomes was 11.41x (± 2.21 st dev), ranging from a minimum depth per individual of 6.51x (a mouse from La Roca del Vallès, Barcelona, Spain) to a maximum depth of 18.35x (a mouse from Prado del Rey, Cádiz, Spain). Average per-base depth of coverage for each individual, separated by autosomes, X-chromosome, and Y-chromosome, is provided in Table 1.

The expected depth of coverage for the X-chromosome in males is half of the autosomal coverage, and the expected depth of coverage for the Y-chromosome in females is zero. We used these expectations to confirm the sex of samples recorded in the field and to determine the sex of individuals where sex was not provided. The sex of two individuals was corrected using this approach. Sample PT68 from Portugal was labeled as a female in the field while genomic data suggest that this individual was a male. Sample JBS41 from Scotland was recorded as a male in the field and genomic evidence suggests that this individual was a female. Six other individuals (MWN03-B_B4, MWN03-C_C6, MWN03-D_D2, MWN03-M_En, MWN03-N_JBS74, MWN03-P_JBS10) were previously missing information on sex and had this information provided using genomic coverage.

Identifying t-haplotype carriers using measurements of excess heterozygosity

The *t*-haplotype is associated with inversions along the proximal third of

chromosome 17 in house mice. It exhibits meiotic drive with male *t*-heterozygotes showing a transmission distortion of up to 99% *t*-haplotypes rather than the expected 50% frequency (Bennett 1978; Bennett et al. 1983; Lenington et al. 1988; Hammer et al. 1989; Keleman and Vicoso 2018). Despite this extreme transmission distortion, *t*-haplotypes are typically found at low frequencies in natural populations, in large part because homozygous *t*-haplotype mice suffer reduced fitness often due to embryonic lethality (Huang et al. 2002). Recombination is suppressed by inversions in *t*-heterozygotes, leading to an excess of polymorphic sites and increased heterozygosity along the inverted regions of chromosome 17 (Keleman and Vicoso 2018; Keleman et al. 2022). We looked for increased heterozygosity on the proximal third of chromosome 17 to identify *t*-haplotypes, first in the Harr et al. (2016) samples, where *t*-haplotypes had previously been identified from genotyping with diagnostic markers, and then among the 59 newly sequenced samples (Supplementary Table 2).

We confirmed that four of the 16 European *M. m. domesticus* samples from Harr et al. (2016) were *t*-haplotype carriers (FRA_B2C, FRA_C1, FRA_F1B, and GER_TP51D). Additionally, we found that 17 of the 59 newly sequenced mice were *t*-haplotype carriers (Table 1; Supplementary Table 2). Thus, we observed a total of 21 *t*-haplotype carriers out of 83 individuals (25.3%), a frequency above the average of 6% seen across many North American populations (Ardlie and Silver 1998) but consistent with many previous estimates (Lewontin and Dunn 1960; Bennett 1978; Bennett et al. 1983; Lenington et al. 1988; Ruvinsky et al. 1991; Ardlie and Silver 1996).

Relatedness

We examined pairwise relatedness among the 59 new samples using vcftools 'relatedness2' option. Two pairs and one trio of mice were found to be first degree relatives, having kinship coefficients approximately equal to or greater than 0.25 (Table 2). One pair was from Birmingham, England, one pair was from Policastro, Italy, and the trio was from Sutherland, Scotland. No duplicate samples were detected (phi $\geq$ 0.5), nor were any first-degree relatives detected between different populations.

Population structure

Principal component analysis revealed four distinct groups (Figure 1a). Western house mice from Germany, northern France, Guernsey, England, Scotland, and Wales form a Northern European group. Mice from southern France, Italy, and northeast Spain form a Mediterranean group. Mice from Portugal and southwest Spain form an Atlantic Iberian group. Samples from Iran form their own distinct group.

These groups are also seen in ADMIXTURE analysis. The ADMIXTURE level of K with highest support (by lowest cross-validation error) is K=4 (Figure 1d). At

K=4, we again observed a distinct cluster for Iran, the Northern European samples, the Mediterranean samples, and the Atlantic Iberian samples. ADMIXTURE results also identified ancestral tracts from the Mediterranean group in the populations of northern France, Guernsey, and Germany, suggesting gene flow across continental Europe from the Mediterranean to Northern Europe.

In our analysis of population structure, eight samples from northern Scotland (6 from Sutherland and 2 from Perthshire) show greater affinity to one another than to the nine remaining UK samples (Supplementary Figure 1 and 2). In the following analyses, we consider these samples separately and refer to this group as northern UK. We refer to those nine remaining UK samples as southern UK. However, the southern UK samples are not exclusively sampled south of the northern UK samples. Eight of the nine samples come from various parts of England and Wales, and the Dumfries and East Lothian counties of Scotland, all which are located south of the northern UK samples. The remaining southern UK sample is found in Inverness, Scotland, which is located between Perthshire and Sutherland.

Outgroup f_3 statistics further corroborate the distinctiveness of the Northern European, Mediterranean, and Atlantic Iberian groups. The pairwise outgroup f_3 statistic measures shared drift between each pair of populations relative to a designated outgroup. Higher values indicate greater genetic similarity between the two test populations. We treated each labeled group in Figure 1c as a population and set Iran as the outgroup. Eight of the ten possible pairs of Northern European populations gave the highest values for the outgroup f_3 statistics (Figure 2). It is interesting to note that the southern UK samples (England, Wales, and southern Scotland) and the sample from Guernsey share the most genetic drift of all measured samples. Guernsey is an island in the English Channel geographically close to France but with cultural and political ties with the UK. Other intra-group pairs also generate high values for the outgroup f_3 statistics such that all intra-group pairs provide the highest fourteen values. The pairs of samples between different groups with high f_3 statistic values were between Italy and northern France, Italy and Guernsey, and Italy and Germany. Each of the Northern European populations in those pairings were previously inferred to have high levels of admixture from the Mediterranean group.

Phylogenetic relationships

Samples from Iran were paraphyletic with respect to the samples from Europe. Within Europe, we observed three major monophyletic groups, corresponding to the groups identified by PCA and ADMIXTURE: an Atlantic Iberian clade, a Northern European clade, and a Mediterranean clade (Figure 1c). The Atlantic Iberian clade was the sister group to the other two groups. Within the Atlantic Iberian clade, samples from Portugal were monophyletic as were samples from southwest Spain. Thus, despite their evenly spaced and continuous geographic sampling (Figure 1b), mice from Portugal and southern Spain clustered by geopolitical region. Within the

Mediterranean clade, mice were also grouped by country, with mice from northeastern Spain forming a clade that was the sister group to clades of mice from Italy and southern France. Finally, the Northern European group consisted of clades corresponding to Germany, northern France, northern UK, and Guernsey together with southern UK. Overall, these phylogenetic relationships agree well with PCA and ADMIXTURE analyses. Given that Guernsey was represented by a single individual, due to limitations of the methods used and the possibility of introducing bias from inadequate sampling, Guernsey was not included in the remaining analyses.

Gene flow

The *f*-branch statistic assigns gene flow to specific branches of a phylogeny based on excess allele sharing. This analysis revealed gene flow between the three major clades. The strongest signals of gene flow involved Portugal, northern France, northern UK, and Italy, as depicted through the shaded red cells along the corresponding rows of Figure 3. There are many possible four-taxon D tests in a phylogeny of even modest size. As a result of the phylogenetic non-independence of these tests, gene flow can introduce correlated D statistics in related lineages (Malinsky et al. 2021). For example, the horizontal red bars associated with Portugal probably reflect gene flow involving Portugal and one of the other populations in the North European or Mediterranean clades, with the remaining red squares reflecting the correlated signal in related lineages. The single largest signal was between northern France and Germany, where ADMIXTURE analysis inferred a similar admixture history. Northern France also appears to have a strong signal of gene flow with Portugal, suggesting gene flow along the Atlantic seaboard, a result consistent with previous studies (Gündüz et al. 2001; Förster et al. 2009; Gabriel et al. 2015; García-Rodríguez et al. 2021).

Inferring past population bottlenecks

All populations of house mice experienced a population size decline from 500,000 to roughly 20,000 generations ago, at which point the effective population size in Iran began to stabilize (Figure 4). Using a generation time of one year (Geraldes et al. 2008), twenty thousand years ago coincides with the Last Glacial Maximum. After this time, the Western European populations continued to experience a reduction in population size relative to Iran. The first Western European populations began to stabilize inside of 10,000 generations ago, and all Western European populations stabilized to their present effective population size within the past 1,500–3,000 generations (Figure 4).

Divergence times between lineages

We estimated divergence times between all pairs of populations using smc++. All estimated divergence times were between 1,500 and 5,500 years ago (Table 3).

The estimated divergence times between Iran and Western European mice ranged from 3,700 to 5,500 years ago. Of the samples included in this study, the Mediterranean clade includes mice from populations that are geographically closest to the Levant, the geographic region from where mice in Western Europe are thought to have originated. Divergence times between Iran and the Mediterranean clade were roughly between 4,200 and 4,700 years ago.

Divergence times between Northern Europe and Iran ranged roughly between 3,700 and 5,100 years ago, while divergence times between Northern Europe and the other two Western European clades generally ranged from 2,400 to 4,600 years ago. Divergence times between Atlantic Iberian mice and Mediterranean mice were estimated to be approximately between 3,100 and 4,000 years ago. The wider range of divergence times for comparisons involving Northern European populations might reflect a more complex demographic history within the Northern European clade, as analysis of population structure and *f*-statistics suggests notable admixture in northern France and Germany.

We also estimated divergence times among populations within major clades. These times ranged between 1,500 and 3,300 years ago, and if we only consider populations that geographically border one another, these divergence times are even younger, ranging from 1,500 to 2,500 years ago (Table 3; Figure 5). The populations with the most recent split times were those in Germany and northern France (1,500 years ago), followed by Spain and Portugal (1,700 years ago).

Discussion

Mus musculus domesticus is the most widespread subspecies of *Mus musculus*, and can be found across the Middle East, Western Europe, Africa, the Americas, and Australasia. Western Europe is the historical range of *M. m. domesticus* from which mice have been introduced to other regions of the world within the past few hundred years. Understanding the history of house mice within their more ancestral range lays the groundwork for identifying the source populations for other regions of the world. A detailed understanding of the demographic history of house mice will also provide a useful null model for studies aiming to uncover signatures of selection and to identify the genetic basis of adaptive traits.

We sequenced 59 whole genomes of *M. m. domesticus* from Western Europe and combined these with a previously published dataset of 24 whole genomes. We found that house mice in Western Europe fall into three major groups: a Northern European group, a Mediterranean group, and an Atlantic Iberian group. We found evidence of gene flow among populations within and between these groups, with the highest levels of gene flow being detected between populations in Germany and northern France. Interestingly, all sampled populations appear to have undergone a contraction in population size roughly coincident with the Last Glacial Maximum.

Finally, divergence times between Western European populations were estimated to be between 1,500 and 5,500 years ago, consistent with previous reports (Cucchi et al. 2005; Cucchi et al. 2020; Garcia-Rodriguez et al. 2021). Below we discuss these results in light of human history in Western Europe. However, we emphasize that the dates obtained here should be construed as approximate and interpreted with caution. The exact dates depend on estimates of mutation rate and generation time.

Since house mice are commensal with humans, we might expect patterns of historical relationships and migration in mice to reflect similar patterns in humans. This appears to be largely true. For example, house mice form three major clades in Western Europe (Figure 1). Similarly, human populations from Western Europe form genetically distinct groups in the Iberian Peninsula, Mediterranean, and Northern Europe (Lazaridis et al. 2014; Gilbert et al. 2022). We found that mice from France are divided into those in the north that show clear affiliations with mice in the UK and Germany and those in the south that show clear affiliations with mice from Italy and northeastern Spain, a pattern also found in people from these regions (Lazaridis et al. 2014). The phylogenetic grouping of mice from the Atlantic coast of the Iberian Peninsula matches current geopolitical boundaries, despite the closer proximity of some Spanish mice to some Portuguese mice relative to other mice sampled within the same country. Mice along the Atlantic Iberian coast were sampled in an evenly-spaced fashion, with the closest samples between Portugal and Spain being from Tavira (Portugal) and Aljaraque (Spain) which are only 65 km apart. Yet, mice from Portugal and mice from Spain formed two distinct phylogenetic groups (Figure 1c). On the other hand, we found that mice along the Atlantic Iberian coast grouped separately from mice in northeastern Spain (Catalonia), a pattern that may reflect socio- and ethnolinguistic differences found in Spain, but is not similarly observed in patterns of human genetic variation (Lazaridis et al. 2014; Gilbert et al. 2022).

Some of the patterns documented here using whole genome data are consistent with previous studies of house mice. Searle et al. (2009) and Garcia-Rodriguez et al. (2021) found that British populations of house mice exhibit high mitochondrial DNA variation, and that northern Germany and southern and central British populations contained a different haplogroup from that seen in mice from northern Scotland and Ireland. We observed similar patterns in the distinctness of northern and southern UK populations. These differences are likely due to greater gene flow among populations in southern and central Britain than between these populations and those in northern Scotland, reflecting migration patterns observed in human British populations (Leslie et al. 2015; Antonio et al. 2022).

More generally, gene flow highlights the importance of maritime routes in the spread of house mice. The populations with the most evidence of gene flow were in coastal nations with notable histories of migration across the Mediterranean Sea and the Atlantic Ocean. Evidence of gene flow between Portugal and Northern Europe supports previous reports of house mice in some islands of the North Atlantic having come from Northern Europe, despite greater proximity and historical links to mainland Portugal (Gündüz et al. 2001; Förster et al. 2009; Gabriel et al. 2015).

Studies of ancient DNA have revealed high mobility among human populations across Europe over the past 3,000 years (Antonio et al. 2022). Leslie et al. (2015) estimated that human populations in Britain possess substantial genetic contributions from human populations in northern France and Germany. They suggest that the historical context of this contribution centered around waves of human migration associated with the expansion of the Roman Empire. Mirroring this pattern in humans, admixture analysis revealed that house mice from Germany and northern France experienced gene flow from the UK and the Mediterranean (Figure 1d).

The estimated divergence times between geographically close populations within each of the three major clades were between 1,500 – 2,500 years ago. Cucchi et al. (2005) emphasized the Iron Age as being the time of mass movement of mice by humans, but our molecular dating also would suggest some movement slightly later – in the Roman period. The Roman Republic began about 2,500 years ago and lasted until about 2,000 years ago, and was followed by the Roman Empire which ended approximately 1,500 years ago (Heather 2007; Harper 2017). The estimated divergence time between populations of mice in Italy and northeastern Spain was 2,277 years ago, and the estimated divergence time between populations of mice in Italy and south France was 2,339 years ago. Both of these inferred times fall within the period of the late Iron Age and early Roman Republic.

Roman conquest of the Iberian Peninsula began around 2,200 years ago and occupation lasted until the end of the Roman Empire, at which point the Visigoths, whose invasion began around 1,600 years ago, occupied much of Iberia (Heather 2007, Harper 2017). The inferred divergence time between mice from Portugal and Spain of 1,700 years ago falls within that period, and this human invasion, migration, and occupation history may have played a role in the dispersal of house mice in that period.

The fall of the Roman Empire came with constant invasions from nomadic tribes. The Franks moved across Northern Europe from current Germany to current north France in their invasion of the Roman Empire. This invasion and occupation occurred from 1,700 years ago through to the collapse of the Roman Empire, similar to the Visigoths in Southern Europe (Heather 2007; Harper 2017). The inferred divergence time between mice from Germany and mice from northern France was estimated to be 1,500 years ago, hence around the fall of the Roman Empire. While being largely consistent with zooarchaeological records (Cucchi et al. 2005; Cucchi et al. 2020), estimates of divergence times from smc++ should be interpreted with caution, as molecular dating is sensitive to assumptions of fixed model constraints, such as generation time and mutation rate. Nonetheless, these examples illustrate how patterns in the demographic history of house mice may reflect the movement of the ancient humans that likely facilitated their spread, as has been previously reported with black rats (Yu et al. 2022).

In summary, many of the patterns we observe in house mice match patterns previously reported in human populations or aligned with human history. We also uncovered some phylogeographic patterns in mice that do not align to geopolitical boundaries, but most of these agree with previous genetic findings in house mice. We conclude that house mice in Western Europe dispersed within the past several thousand years, diffusing from east to west, in agreement with previous studies based on the zooarchaeological record. We also introduce genomic resources of house mice from regions of historical relevance that were previously not represented in public datasets. This sampling will enable future researchers to explore the dispersal history of house mice outside of Western Europe and connect potential source populations to extant populations worldwide.

Figures

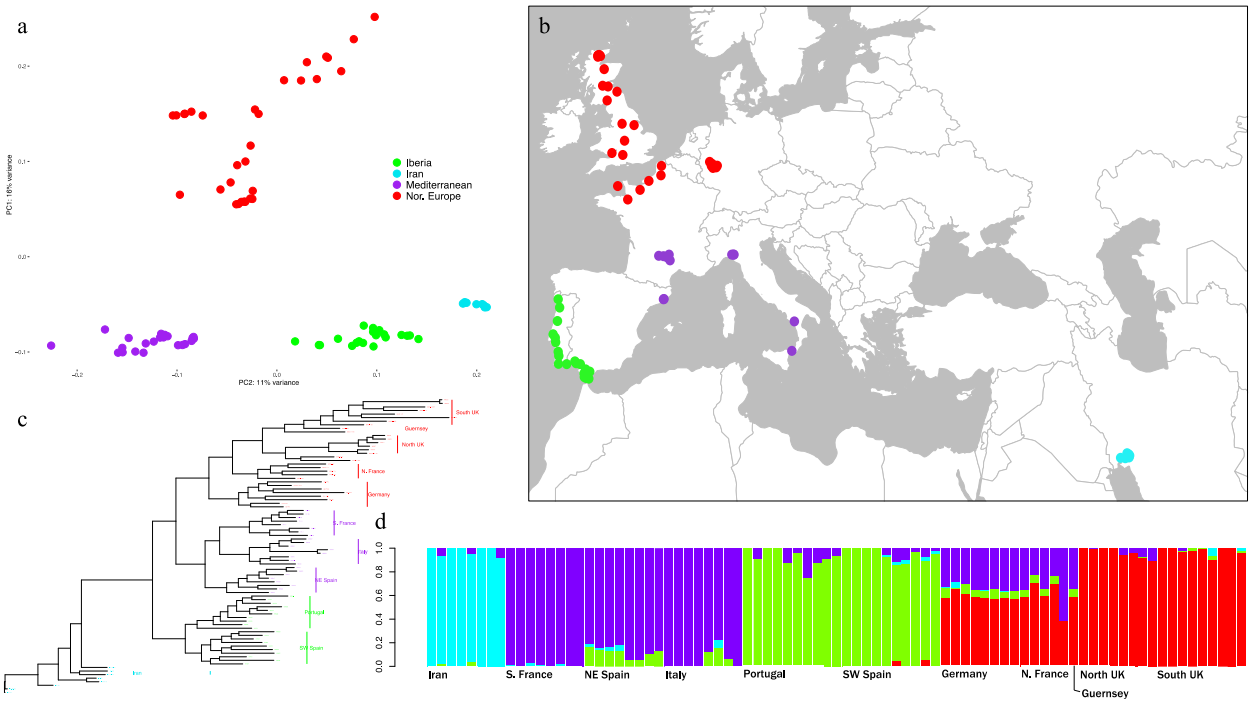
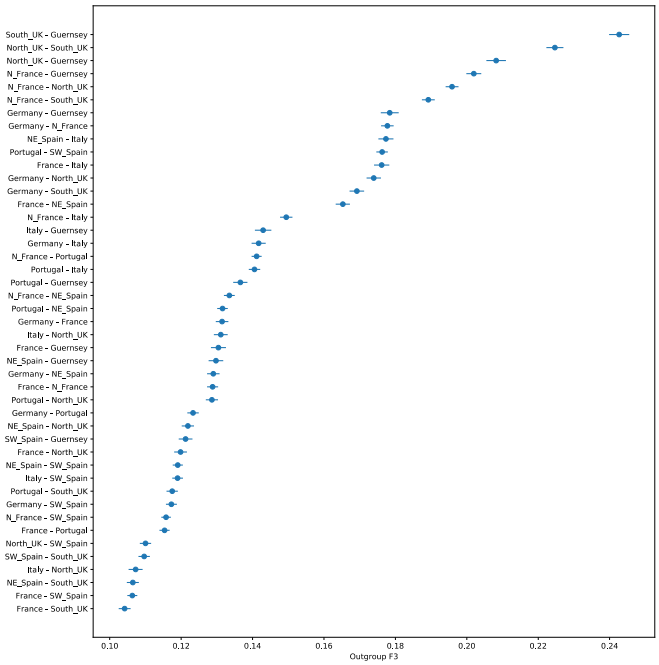


Figure 2.1: Sample distribution and population structure of western house mice. (a) Principal component analysis (PCA). (b) Sample map depicts locations in Portugal (green), southwest Spain (green), southern France (purple), Italy (purple), northeast Spain (purple), England (red), Wales (red), Scotland (red), Guernsey (red), northern France (red), Germany (red), and Iran (cyan) from which western house mouse genomes were obtained or were previously available. (c) Maximum-likelihood phylogeny and (d) ADMIXTURE analysis identifies four distinct groups corresponding to samples from Northern Europe (red), the Mediterranean (purple), the Atlantic coast of Iberia (green), and Iran (cyan).



Pop 1	Pop 2	f3	std err	Z	SNPs
South UK	Guernsey	0.242682	0.002818	86.107	6354312
North UK	South UK	0.224673	0.002385	94.222	7382637
North UK	Guernsey	0.208231	0.00274	75.987	6273338
North France	Guernsey	0.20196	0.002094	96.469	5337671
North France	North UK	0.195873	0.001791	109.362	6246824
North France	South UK	0.18923	0.001796	105.388	6303346
Germany	Guernsey	0.178415	0.00251	71.069	6179775
Germany	North France	0.177757	0.001769	100.485	6192626
NE Spain	Italy	0.17735	0.002111	84.03	6606218
Portugal	SW Spain	0.176282	0.001619	108.868	6968925
South France	Italy	0.176151	0.002146	82.079	6640886
Germany	North UK	0.173944	0.002029	85.728	7218408
Germany	South UK	0.16921	0.002042	82.877	7265949
South France	NE Spain	0.165273	0.001984	83.305	6800688
North France	Italy	0.149435	0.001728	86.488	5883561
Italy	Guernsey	0.142961	0.002305	62.024	5847706
Germany	Italy	0.141708	0.001975	71.758	6817315
North France	Portugal	0.141135	0.001414	99.841	6026903
Portugal	Italy	0.140534	0.001606	87.522	6590308
Portugal	Guernsey	0.13657	0.002	68.281	5995285
North France	NE Spain	0.133494	0.001523	87.669	6035122
Portugal	NE Spain	0.131586	0.001478	89.016	6753093
Germany	South France	0.131457	0.001752	75.033	7017118
Italy	North UK	0.131092	0.001948	67.309	6900855
South France	Guernsey	0.130438	0.002106	61.926	6035862
NE Spain	Guernsey	0.129737	0.002045	63.438	6002266
Germany	NE Spain	0.129004	0.001759	73.33	6980768
South France	North France	0.1288	0.001536	83.86	6070623
Portugal	North UK	0.128605	0.00171	75.229	7050287
Germany	Portugal	0.123317	0.001602	76.991	6959545
NE Spain	North UK	0.121878	0.001727	70.559	7071185
SW Spain	Guernsey	0.121237	0.001936	62.609	6226578
South France	North UK	0.119835	0.001787	67.044	7108825
NE Spain	SW Spain	0.119053	0.001424	83.617	7001165
Italy	SW Spain	0.118991	0.001506	78.988	6838205
Portugal	South UK	0.117493	0.00159	73.913	7125258
Germany	SW Spain	0.117263	0.001539	76.214	7200060
North France	SW Spain	0.115775	0.001344	86.141	6242007
South France	Portugal	0.115355	0.001423	81.079	6796309
North UK	SW Spain	0.110013	0.001587	69.311	7298539
SW Spain	South UK	0.109638	0.001594	68.799	7364350
Italy	South UK	0.107247	0.001963	54.639	6983533
NE Spain	South UK	0.106458	0.001676	63.53	7150872
South France	SW Spain	0.106317	0.001402	75.847	7046361
South France	South UK	0.104172	0.001664	62.588	7185316

Figure 2.2: Outgroup f_3 statistics for pairwise analysis of western house mouse populations. (Left) Biplot of normalized f_3 statistics. (Right) Table of normalized f_3 statistics, with z-score computed using qp3Pop.

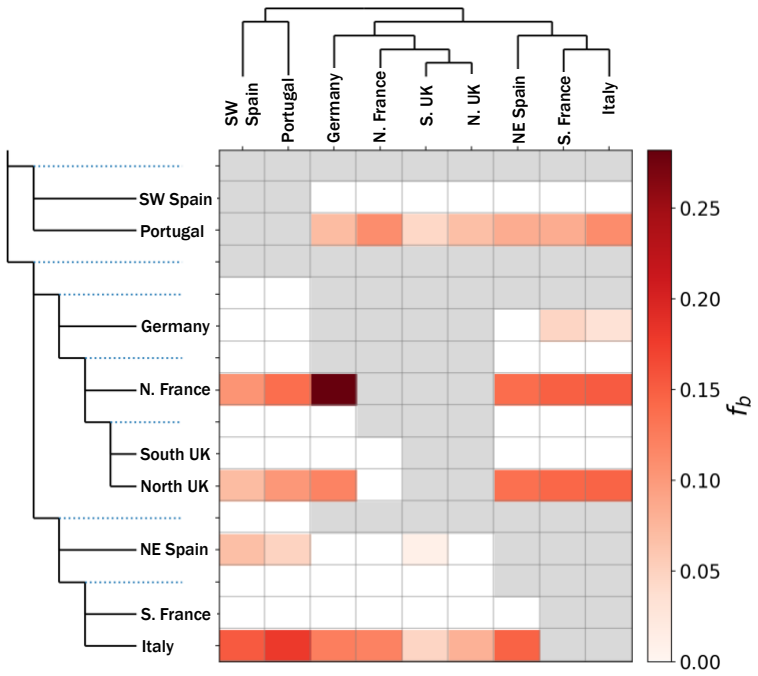


Figure 2.3: Depiction of gene flow between Western European house mouse populations. Colored cells in the matrix highlight excess allele sharing between populations along the x-axis and branch along the expanded phylogeny on y-axis. Cells are shared from white (f_b=0) to dark red (highest value) according to f_b, the *f*-branch statistics, as described in Malinksy et al. (2021). The program sets f_b to 0 for any non-significant tests, since very short internal branches can lead to large but non-significant values even in the absence of gene flow. Grey squares correspond to cells whose value cannot be computed because of the structure of the tree. For example, there is no four-taxon test that enables comparison of the most basal internal lineages (e.g. the top row).

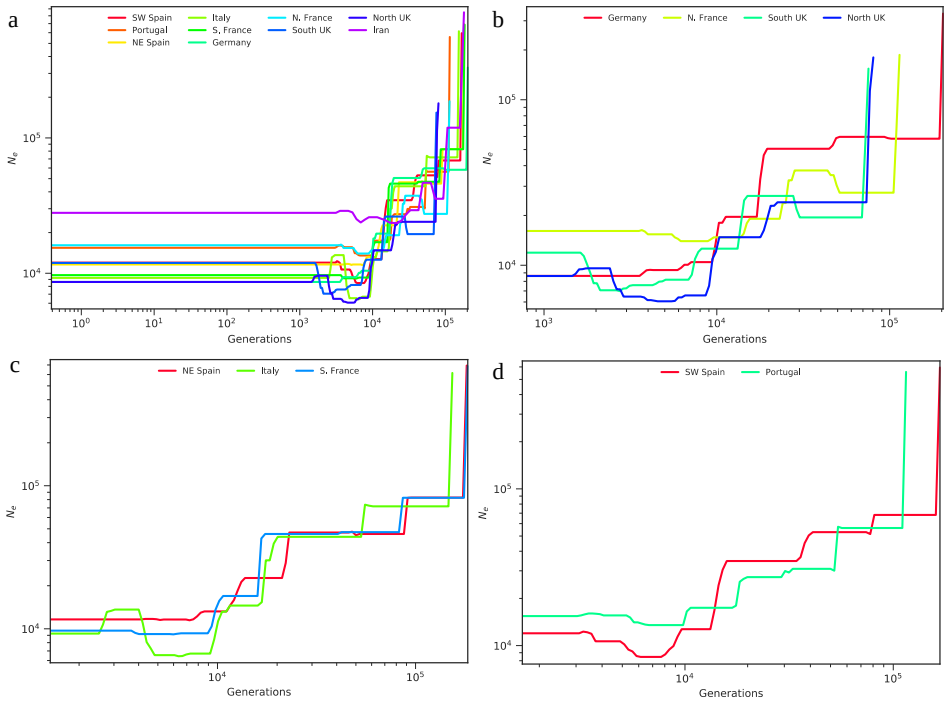


Figure 2.4: Demographic history of western house mice inferred using smc++. Plot of effective population size changes in (a) all populations (b) Northern European populations (c) Mediterranean populations (d) Iberian populations. Plots were generated using a per site mutation rate of 4 x 10^{-9} per generation and a generation time of 1 year.

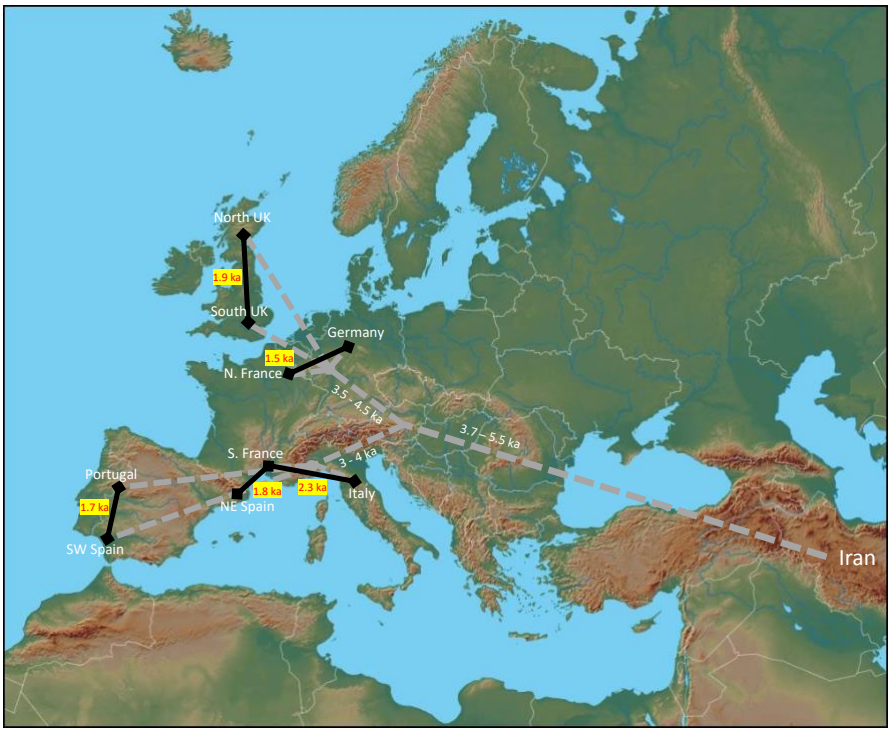


Figure 2.5: Model of diffusion using inferred divergence times of sampled house mouse populations. Each clade comparison found in Table 3 reports inferred split time used to construct the time intervals summarized here in grey. Intra-clade comparisons of neighboring geographic regions connected by the black line are provided in red and highlighted in yellow.

Table 2.1: Summary of WGS data. Detailed sample information and summary statistics from DNA sequencing of 59 wild caught house mice from Western Europe

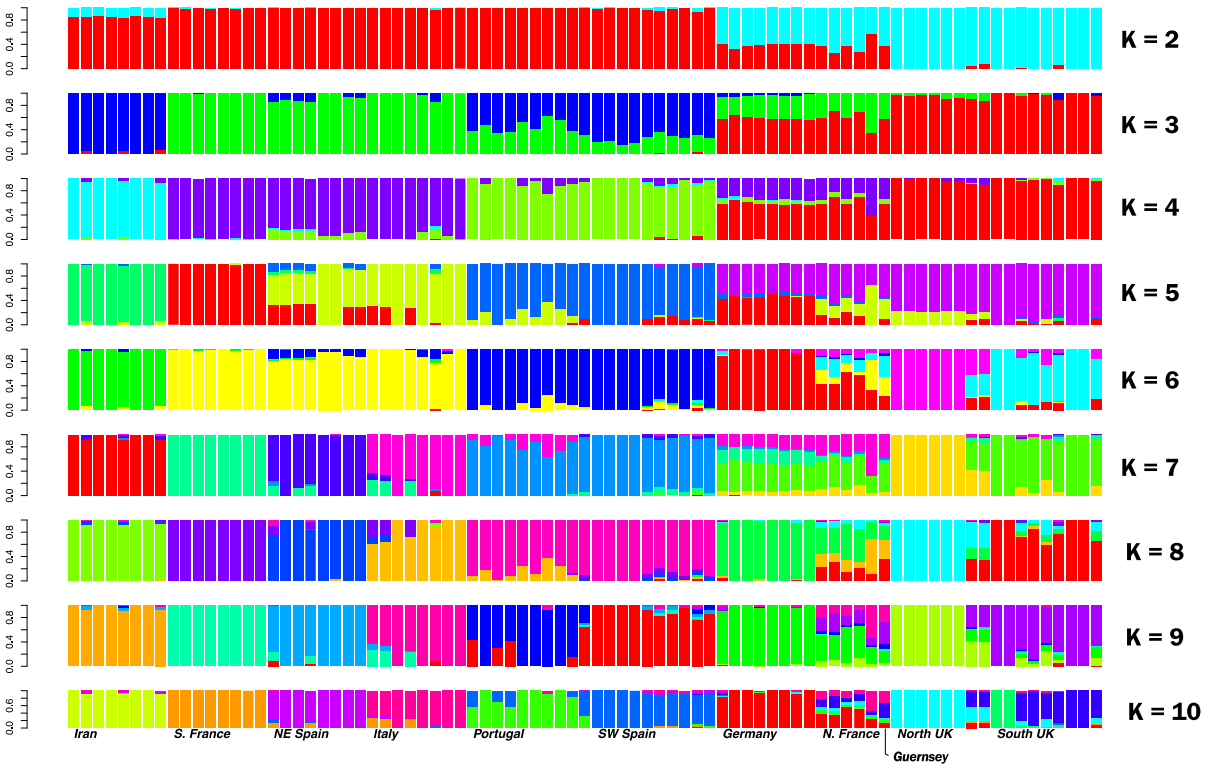
Sample ID	Population location	Sex	Total Reads	% Reads Mapped	Total Depth of Coverage	Autosomal Coverage	Fold Coverage X Chr	Fold Coverage Y Chr	T-haplotype
MWN01A_1285	Northeast Spain	male	189,513,699	99.8	9.72	10.00	5.24	4.06	Yes
MWN01B_1287	Northeast Spain	female	134,627,470	99.73	7.22	7.18	7.10	0.11	Yes
MWN01C_1288	Northeast Spain	male	200,512,740	99.77	10.58	10.92	5.57	4.21	No
MWN01D_1289	Northeast Spain	female	216,349,432	99.79	11.72	11.69	11.12	0.23	No
MWN01E_1290	Northeast Spain	male	207,984,613	99.77	10.65	11.00	5.69	4.19	Yes
MWN01F_1291	Northeast Spain	male	212,381,350	99.77	11.20	11.58	5.94	4.35	No
MWN01G_1292	Northeast Spain	female	189,321,374	99.79	9.80	9.76	9.32	0.18	Yes
MWN01H_1293	Northeast Spain	female	220,887,828	99.78	11.54	11.52	11.11	0.21	Yes
MWN01I_1024	Italy	female	206,681,297	99.76	11.54	11.54	10.46	0.27	Yes
MWN01J_1029	Italy	male	174,661,479	99.81	8.95	9.20	4.64	3.65	No
MWN01K_1032	Italy	female	205,949,996	99.78	11.10	11.07	10.30	0.25	No
MWN01L_1036	Italy	male	209,195,087	99.78	10.89	11.27	5.66	4.09	No
MWN01M_1204	Italy	male	287,987,809	99.79	15.60	16.18	7.80	5.89	Yes
MWN01N_1206	Italy	female	200,076,278	99.77	11.13	11.12	10.42	0.25	No
MWN01O_1208	Italy	male	238,001,809	99.73	12.35	12.74	6.50	4.70	No
MWN01P_1215	Italy	female	282,932,715	99.78	15.32	15.32	13.86	0.36	No
MWN01Q_1316	England	male	193,535,807	99.82	10.29	10.61	5.34	4.01	No
MWN01R_1317	England	male	221,071,876	99.74	11.34	11.71	6.01	4.49	No
MWN01S_1318	Scotland	male	218,210,494	99.8	11.42	11.79	6.04	4.23	Yes
MWN01T_1319	Scotland	male	255,301,409	99.8	13.56	14.04	6.93	4.93	Yes
MWN01U_1320	Scotland	female	224,708,710	99.81	11.87	11.84	11.16	0.31	Yes
MWN01V_1321	Scotland	female	234,234,395	99.82	12.27	12.23	11.58	0.28	Yes
MWN01W_1323	Scotland	female	218,173,627	99.81	11.51	11.49	10.74	0.24	No
MWN01X_1324	Scotland	female	199,761,130	99.78	10.60	10.55	9.94	0.24	No
MWN02A_PT22	Portugal	female	306,588,644	99.82	16.60	16.66	15.70	0.35	No
MWN02B_PT35	Portugal	female	234,459,728	99.82	12.58	12.59	12.01	0.28	Yes
MWN02C_PT49	Portugal	male	286,761,526	99.74	15.25	15.85	7.93	6.13	No
MWN02D_PT51	Portugal	female	250,773,526	99.82	13.65	13.70	12.76	0.28	No
MWN02E_PT57	Portugal	female	218,152,545	99.82	11.76	11.78	11.08	0.26	No
MWN02F_PT60	Portugal	male	232,091,119	99.82	12.13	12.58	6.35	4.45	No
MWN02G_PT64	Portugal	male	283,613,674	99.82	14.99	15.57	7.89	6.13	Yes
MWN02H_PT68	Portugal	male	197,429,085	99.81	10.51	10.87	5.52	4.45	No
MWN02I_PT69	Portugal	male	365,728,682	99.84	18.56	19.32	9.70	7.57	No
MWN02J_PT81	Portugal	male	227,975,847	99.82	12.01	12.45	6.36	4.93	No
MWN02K_SP7	Southwest Spain	female	269,393,735	99.76	14.44	14.47	13.77	0.30	No
MWN02L_SP9	Southwest Spain	male	280,186,973	99.65	14.79	15.36	7.83	6.24	No
MWN02M_SP21	Southwest Spain	male	380,213,172	99.78	19.93	20.74	10.74	7.22	No
MWN02N_SP48	Southwest Spain	female	308,103,830	99.78	16.43	16.51	15.23	0.38	No
MWN02O_SP58	Southwest Spain	female	297,529,564	99.82	15.76	15.78	15.15	0.33	No
MWN02P_SP67	Southwest Spain	male	267,597,321	99.79	13.90	14.41	7.26	5.82	No
MWN02Q_SP72	Southwest Spain	male	331,806,477	99.8	17.20	17.85	9.19	7.82	Yes
MWN02R_SP86	Southwest Spain	female	239,266,872	99.79	12.48	12.50	11.75	0.31	No
MWN02S_SP93	Southwest Spain	male	286,452,314	99.81	14.85	15.40	7.95	6.37	No
MWN02T_SP104	Southwest Spain	female	237,037,450	99.8	12.60	12.62	11.99	0.32	Yes
MWN03-A_A7	North France	female	178,595,662	99.73	13.48	13.47	12.88	0.30	No
MWN03-B_B4	North France	female	181,368,589	99.72	14.28	14.28	13.57	0.29	No
MWN03-C_C6	North France	female	184,891,672	97.63	13.07	13.07	12.30	0.36	No
MWN03-D_D2	North France	female	158,850,311	99.72	12.04	12.00	11.23	0.30	No
MWN03-E_E2	North France	male	176,516,724	99.73	12.99	13.41	6.87	5.49	Yes
MWN03-F_1060	Guernsey	male	196,742,814	99.8	14.33	14.83	7.64	5.21	No
MWN03-G_121	Scotland	female	217,525,681	99.63	15.76	15.77	14.79	0.35	No
MWN03-H_593	Scotland	female	187,302,858	99.77	14.20	14.18	13.53	0.30	No
MWN03-I_670	England	female	227,483,806	99.75	17.29	17.23	16.54	0.33	No
MWN03-J_693	Scotland	male	155,476,084	99.77	11.47	11.85	6.01	4.45	No
MWN03-K_1172	Wales	female	190,638,693	99.74	14.55	14.53	13.67	0.28	No
MWN03-M_En	England	male	226,577,467	99.77	17.03	17.69	8.91	5.95	No
MWN03-N_JBS74	Scotland	male	163,178,483	98.16	11.68	12.04	6.20	4.68	Yes
MWN03-P_JBS10	England	female	224,286,720	99.79	16.14	16.15	15.02	0.36	No
MWN03-Q_JBS41	Scotland	female	228,782,736	99.78	17.29	17.29	16.42	0.38	No

Table 2.2: List of first-degree relatives among all studied house mouse populations. The kinship coefficient, Phi, was computed in vcftools using the method of Manichaikul et al. (2010)

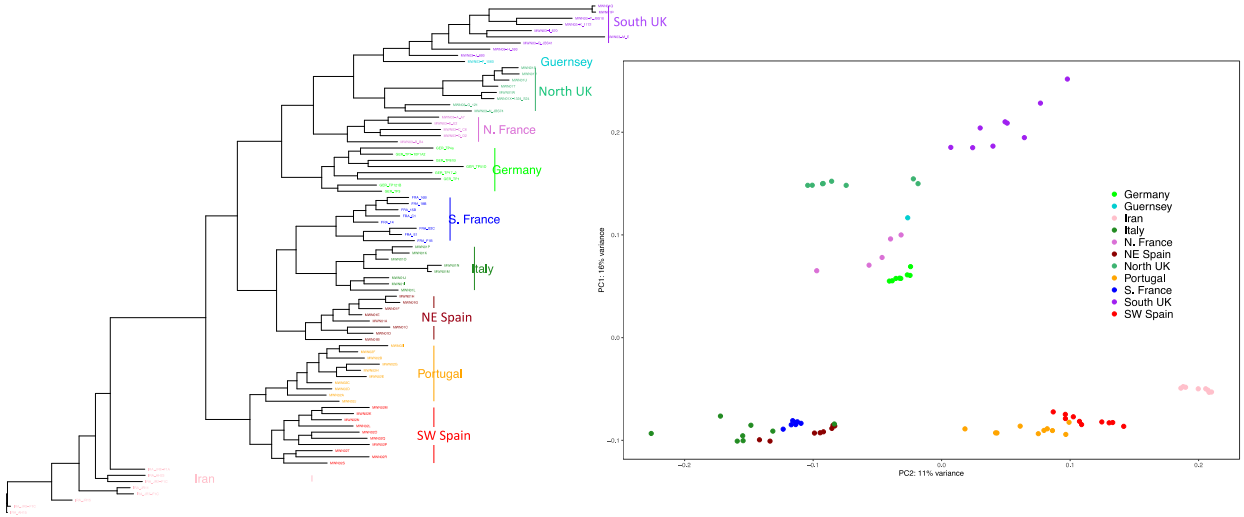
Sample 1	Sex	Sample 2	Sex	Population	N_AaAa	N_AAaa	N1_Aa	N2_Aa	PHI
MWN01Q_1316	male	MWN01R_1317	male	England	324533	7032	495270	475980	0.319659
MWN01T_1319	male	MWN01U_1320	female	Scotland	397330	35534	660960	632492	0.252241
MWN01T_1319	male	MWN01V_1321	female	Scotland	431176	49995	660960	635834	0.255388
MWN01U_1320	female	MWN01V_1321	female	Scotland	387684	40375	632492	635834	0.241999
MWN01M_1204	male	MWN01N_1206	female	Italy	514821	23333	801909	757616	0.300191

Table 2.3: Clade comparison of inferred divergence times between western house mouse populations. For each comparison, the table reports the inferred split time in years into the past.

Paired Clade Comparison	Population 1	Population 2	split time (years ago)
Iran – Western Europe	Iran	South France	4242
	Iran	NE Spain	4745
	Iran	Italy	4418
	Iran	Germany	3745
	Iran	North France	4636
	Iran	South UK	4542
	Iran	North UK	5067
	Iran	SW Spain	4155
	Iran	Portugal	5499
Mediterranean – Atlantic Iberian	NE Spain	SW Spain	3078
	NE Spain	Portugal	3483
	Italy	SW Spain	3583
	Italy	Portugal	3886
	South France	SW Spain	3498
	South France	Portugal	4005
Northern Europe – Atlantic Iberian	South UK	SW Spain	2437
	South UK	Portugal	2895
	Germany	SW Spain	2627
	Germany	Portugal	2962
	North France	SW Spain	3932
	North France	Portugal	4405
	North UK	SW Spain	4299
	North UK	Portugal	4590
Mediterranean – Northern Europe	South France	Germany	2690
	NE Spain	Germany	2788
	Italy	Germany	3018
	South France	North France	3548
	NE Spain	North France	3740
	Italy	North France	3887
	South France	South UK	4158
	NE Spain	South UK	4033
	Italy	South UK	4644
	South France	North UK	4346
	NE Spain	North UK	4431
	Italy	N. Scotland	4551
Intra-clade comparisons	Germany	North France	1503
	SW Spain	Portugal	1700
	South France	NE Spain	1843
	South UK	North UK	1892
	Italy	NE Spain	2277
	Italy	South France	2339
	N. France	South UK	2504
	Germany	South UK	2523
	North UK	Germany	3065
	North UK	N. France	3258



Supplementary Figure 2.1: ADMIXTURE analysis for levels of K = 2 to 10 for European house mice samples based on WGS data. The mean cross-validation error for each level of K are as follows: K2 = 0.457, K3 = 0.449, K4 = 0.444, K5 = 0.457, K6 = 0.456, K7 = 0.467, K8 = 0.480, K9 = 0.498, K10 = 0.523



Supplementary Figure 2.2: Phylogenetic tree and principal component analysis colored by specific sampling location of all analyzed house mouse populations.

Supplementary Table 2.1: Population location and collector information for new European genome samples of western house mice. All mice collected by Anonymous have been deposited as specimens in the University of Michigan Museum of Zoology. Population location information in parentheses match labeling derived from structural analyses later used for admixture and demographic analyses. Parenthetical notes are not intended to indicate exact geographic locations.

Sample ID	Population location	Collector	Location Notes
MWN01A_1285	Northeast Spain	Anonymous	Spain: Catalunya, 1.6 km N of Roca del Valles (Rabbit farm)
MWN01B_1287	Northeast Spain	Anonymous	Spain: Catalunya, 1.6 km N of Roca del Valles (Rabbit farm)
MWN01C_1288	Northeast Spain	Anonymous	Spain: Catalunya, 1.6 km N of Roca del Valles (Cow farm)
MWN01D_1289	Northeast Spain	Anonymous	Spain: Catalunya, 1.6 km N of Roca del Valles (Cow farm)
MWN01E_1290	Northeast Spain	Anonymous	Spain: Catalunya, 3.2 km NE of La Roca del Valles
MWN01F_1291	Northeast Spain	Anonymous	Spain: Catalunya, 3.2 km NE of La Roca del Valles
MWN01G_1292	Northeast Spain	Anonymous	Spain: Catalunya, 3.2 km NE of La Roca del Valles
MWN01H_1293	Northeast Spain	Anonymous	Spain: Catalunya, 3.2 km NE of La Roca del Valles
MWN01I_1024	Italy	Anonymous	Italy: Gremiasco
MWN01J_1029	Italy	Anonymous	Italy: Ronco
MWN01K_1032	Italy	Anonymous	Italy: Menconico
MWN01L_1036	Italy	Anonymous	Italy: Canova
MWN01M_1204	Italy	Anonymous	Italy: Region Campania, Policastro
MWN01N_1206	Italy	Anonymous	Italy: Region Campania, Policastro
MWN01O_1208	Italy	Anonymous	Italy: Sicily, 6 km S of Milazzo, Olivarella
MWN01P_1215	Italy	Anonymous	Italy: Sicily, 6 km S of Milazzo, Olivarella
MWN01Q_1316	England (South UK)	Anonymous	Birmingham, England, UK
MWN01R_1317	England (South UK)	Anonymous	Birmingham, England, UK
MWN01S_1318	Scotland (North UK)	Anonymous	Ribigill, Sutherland, Scotland, UK
MWN01T_1319	Scotland (North UK)	Anonymous	Ribigill, Sutherland, Scotland, UK
MWN01U_1320	Scotland (North UK)	Anonymous	Ribigill, Sutherland, Scotland, UK
MWN01V_1321	Scotland (North UK)	Anonymous	Ribigill, Sutherland, Scotland, UK
MWN01W_1323	Scotland (North UK)	Anonymous	Ribigill, Sutherland, Scotland, UK
MWN01X_1324	Scotland (North UK)	Anonymous	Ribigill, Sutherland, Scotland, UK
MWN02A_PT22	Portugal	Anonymous	Lagos, Portugal
MWN02B_PT35	Portugal	Anonymous	Setúbal, Portugal
MWN02C_PT49	Portugal	Anonymous	Sines, Portugal
MWN02D_PT51	Portugal	Anonymous	Vila Nova de Milfontes, Portugal
MWN02E_PT57	Portugal	Anonymous	Figueira da Foz, Portugal
MWN02F_PT60	Portugal	Anonymous	Vila Franca de Xira, Portugal
MWN02G_PT64	Portugal	Anonymous	Viana do Castelo, Portugal
MWN02H_PT68	Portugal	Anonymous	Porto, Portugal
MWN02I_PT69	Portugal	Anonymous	Peniche, Portugal
MWN02J_PT81	Portugal	Anonymous	Tavira, Portugal
MWN02K_SP7	Southwest Spain	Anonymous	Villamartin, Spain
MWN02L_SP9	Southwest Spain	Anonymous	Jerez de la Frontera, Spain
MWN02M_SP21	Southwest Spain	Anonymous	Prado del Rey, Spain
MWN02N_SP48	Southwest Spain	Anonymous	Arcos de la Frontera, Spain
MWN02O_SP58	Southwest Spain	Anonymous	Vejer de la Frontera, Spain
MWN02P_SP67	Southwest Spain	Anonymous	Algeciras, Spain
MWN02Q_SP72	Southwest Spain	Anonymous	Tarifa, Spain
MWN02R_SP86	Southwest Spain	Anonymous	Palos de la Frontera, Spain
MWN02S_SP93	Southwest Spain	Anonymous	Aljaraque, Spain
MWN02T_SP104	Southwest Spain	Anonymous	El Rocio, Spain
MWN03-A_A7	North France	Anonymous	Caen, France
MWN03-B_B4	North France	Anonymous	Fécamp, Normandy, France
MWN03-C_C6	North France	Anonymous	Abbeville, Somme, France
MWN03-D_D2	North France	Anonymous	Calais, France
MWN03-E_E2	North France	Anonymous	Mont St Michel, France
MWN03-F_1060	Guernsey	Anonymous	Guernsey
MWN03-G_121	Scotland (North UK)	Anonymous	Crook of Devon, Perthshire, Scotland, UK
MWN03-H_593	Scotland (South UK)	Anonymous	Dumfries, Scotland, UK
MWN03-I_670	England (South UK)	Anonymous	Epworth, Lincolnshire, England, UK
MWN03-J_693	Scotland (South UK)	Anonymous	Phantassie, East Lothian, Scotland, UK
MWN03-K_1172	Wales (South UK)	Anonymous	Cardiff, Wales, UK
MWN03-M_En	England (South UK)	Anonymous	Todmorden, Yorkshire, England, UK
MWN03-N_JBS74	Scotland (North UK)	Anonymous	Doune, Perthshire, Scotland, UK
MWN03-P_JBS10	England (South UK)	Anonymous	Steeple Ashton, Wiltshire, England UK
MWN03-Q_JBS41	Scotland (South UK)	Anonymous	Castle Stuart, Inverness, Scotland UK

Supplementary Table 2.2: Heterozygosity of Western European house mice. Heterozygosity as measured by the inbreeding coefficient, F, computed using vcftools '--het'. Specifically, F = (O.HOM-E.HOM)/(N_SITES-E.HOM), where O.HOM is observed homozygotes, E.HOM is expected homozygotes, and N_SITES is the number of sites genotyped

INDV	O.HOM.	E.HOM.	N_SITES	F
FRA_14	438928	381261	603529	0.25945
FRA_15B	443646	386472.7	611186	0.25443
FRA_16B	445174	386365.3	610947	0.26186
FRA_18B	443654	387145.1	612124	0.25117
FRA_B2C	352749	386342.1	610732	-0.14971
FRA_C1	310097	386623.8	611244	-0.34069
FRA_E1	446329	386773.6	611570	0.26493
FRA_F1B	296455	387245.7	612209	-0.40358
GER_TP121B	423331	376670.5	596266	0.21248
GER_TP17-2	520957	371978.4	589261	0.68564
GER_TP1	477053	376748.2	596392	0.45667
GER_TP3	440921	377017.1	596765	0.29081
GER_TP4a	491853	373641.1	591969	0.54144
GER_TP51D	328644	376701.8	596276	-0.21887
GER_TP7-10F1A2	425980	376585.1	596138	0.22498
GER_TP81B	491598	376010	595153	0.52745
MWN01Q_1316	419032	339357.5	531556	0.41454
MWN01R_1317	460902	338390.8	529876	0.63979
MWN01S_1318	240839	339939.8	532441	-0.51481
MWN01T_1319	218598	340448.2	533304	-0.63182
MWN01U_1320	219513	340108.5	532723	-0.6261
MWN01V_1321	244903	340012.5	532542	-0.494
MWN01W_1323	488865	336883.2	527673	0.79659
MWN01X_1324	446035	336118	526576	0.57712
MWN01I_1024	346369	402492.7	635881	-0.24047
MWN01J_1029	497755	399762.4	631845	0.42223
MWN01K_1032	515018	401944.4	635214	0.48473
MWN01L_1036	601692	398137.6	629272	0.88068
MWN01M_1204	307134	403605.6	637655	-0.41218
MWN01N_1206	474278	403272	637103	0.30366
MWN01O_1208	467668	403518.2	637577	0.27408
MWN01P_1215	474613	403748.3	637925	0.30261
MWN01A_1285	284470	388267.2	614134	-0.45955
MWN01B_1287	303410	385863.4	610260	-0.36745
MWN01C_1288	520925	386769.9	611523	0.5969
MWN01D_1289	485979	382938.3	606168	0.46159
MWN01E_1290	265756	388783.8	614984	-0.54389
MWN01F_1291	410307	387842	613597	0.09951
MWN01G_1292	288733	387279.2	612612	-0.43734
MWN01H_1293	273752	387927.8	613677	-0.50576
MWN03-A_A7	455525	337479.9	537790	0.58931
MWN03-B_B4	371896	338500.9	539559	0.1661
MWN03-C_C6	398219	338166.1	538953	0.29909
MWN03-D_D2	387704	333142.8	531572	0.27497
MWN03-E_E2	273664	338411.6	539383	-0.32217
MWN02A_PT22	484518	423118.7	669366	0.24934
MWN02B_PT35	320995	422482.7	668269	-0.41291
MWN02C_PT49	515756	421045.3	666115	0.38646
MWN02D_PT51	505706	422368	668093	0.33915
MWN02E_PT57	499978	422197.6	667825	0.31666
MWN02F_PT60	547510	421165.4	666074	0.51588
MWN02G_PT64	369772	422591.2	668460	-0.21483
MWN02H_PT68	608564	419784.7	663919	0.77326
MWN02I_PT69	601841	422068.8	667662	0.73199
MWN02J_PT81	518873	421714.1	666987	0.39613
MWN02K_SP7	492222	418760.3	662500	0.30139
MWN02L_SP9	620296	415749.7	657641	0.84561
MWN02M_SP21	507119	419034	662973	0.36109
MWN02N_SP48	489091	418953.9	662823	0.2876
MWN02O_SP58	517867	418692.2	662362	0.407
MWN02P_SP67	522680	418292.8	661716	0.42883
MWN02Q_SP72	335979	418905.1	662811	-0.33999
MWN02R_SP86	517314	418152.4	661476	0.40753
MWN02S_SP93	491050	418177	661549	0.29943
MWN02T_SP104	326646	418461.9	661990	-0.37702
MWN03-F_1060	560086	380442.8	604055	0.80337
MWN03-G_121	458376	383129.4	608517	0.33385
MWN03-H_593	469069	383501.7	609157	0.37919
MWN03-I_670	452493	383430.4	609103	0.30603
MWN03-J_693	458970	382555	607604	0.33955
MWN03-K_1172	480041	382110.4	607026	0.43541
MWN03-M_En	545947	382546.7	607597	0.72606
MWN03-N_JBS74	325631	382372.2	607267	-0.2523
MWN03-P_JBS10	464251	383890	609791	0.35574
MWN03-Q_JBS41	454782	383339.9	608979	0.31662

Chapter 3

New country for old mice: On the recent colonization history of *Mus musculus domesticus* in the Americas

Summary

Humans have had an enormous impact on the distribution of species across the globe. The western house mouse, *Mus musculus domesticus*, is a human commensal whose ancestral range extended from the northern Indo-Pakistani subcontinent to the Middle East. Following the development of agriculture, house mice spread with humans into Western Europe several thousand years ago. House mice have since been introduced very widely around the globe, presumably by European colonizers. Using an exomic dataset of 276 mice sampled from Western Europe, North America, and South America, we investigate the recent colonization history of western house mice in the Americas. Admixture graphs reveal that house mice from the Atlantic coast of the Iberian Peninsula are the sister group to all populations of house mice in North and South America. A secondary introduction of house mice from the southern half of the UK to North America is indicated by a migration edge from southern UK to the base of the North America clade. Complementary analysis of genetic structure provides evidence of shared genetic ancestry between the Atlantic coast of the Iberian Peninsula and both American continents, between the UK and North America, between Portugal and Brazil, and between Spain and Argentina. Demographic modeling reveals that American populations diverged roughly within the last 500 years, consistent with the timing of European colonization history in the Americas. Together, these results identify the Western European source populations involved in the recent introduction of house mice to North and South America, and support a connection between human colonization history with the spread of house mice across vast geographic regions.

Result and Discussion

During the Anthropocene, human migration facilitated the global dispersal of many species outside of their native range (Hulme 2009; Pimentel et al. 2000; Genovesi et al. 2012). With the advent of agriculture, house mice developed a close association with humans roughly 12,000 - 15,000 years ago and have consequently migrated with humans to various new geographic areas (Pocock et al. 2005; Cucchi et al. 2005, 2020; Weissbrod et al. 2017; Agwamba and Nachman 2023). In particular, the western house mouse, *Mus musculus domesticus*, initially migrated with humans from the Fertile Crescent to Western Europe over the last few millennia (Cucchi et al. 2020, Agwamba et al. Chapter 2). Most recently, *M. m. domesticus* expanded its geographic range to the Americas, Africa, and Oceania, presumably in association with European colonization, although the details of this recent expansion have not been well studied. Previous studies have provided evidence for the westward movement of house mice in their colonization history of the islands in the North Atlantic (Mathias and Mira 1992; Gündüz et al. 2001; Förster et al. 2009; Gabriel et al. 2010; Jones et al. 2012; Jones et al. 2013; Rando et al. 2014; Gabriel et al. 2015). However, these studies did not trace the colonization history of house mice found throughout continental North and South America, and were limited in both the inferential methods used to reveal demographic history and the datasets used to power those methods.

The Americas have a mixed recent human colonization history which notably involved an outsized colonial presence from the Portuguese, Spanish, and English. We hypothesize that the ancestry contributions of house mice in the Americas will reflect the imprints of this human colonization history. Here, we investigate this hypothesis to trace the origins of house mice in the Americas using genomic data from 276 mice sampled across Western Europe, North and South America.

Globally distributed populations of mice form distinct genetics clusters

We gathered an exomic dataset consisting of 193 North and South American house mice derived from latitudinal transects in North, Central and South America (Phifer-Rixey et al. 2018, Ferris et al. 2021, Gutierrez-Guerrero et al. 2024). We combined those samples with the exomic intersect taken from whole genomes of 83 Western European house mice (Harr et al. 2016, Agwamba et al. Chapter 2) (Figure 1a).

Principal component analysis shows that house mouse samples cluster based on geographic location, as their relative locations in PC space approximately mirror their relative locations on a map (Figure 1, Supplementary Figure 1). Human genetic variation also closely mirrors geography in densely sampled human populations such as in Western Europe (Novembre et al. 2008, Lazaridis et al. 2014, Gilbert et al. 2022).

Phylogenetic analyses reveal a Western Iberian origin for all mice in the Americas and a secondary migration wave into North America from the southern UK

We investigated the genetic relationship between European and American house mice using TreeMix, a model which builds maximum likelihood population trees while accounting for gene flow between diverged populations using migration edges (Pickrell and Pritchard 2012). The TreeMix hill-climbing method starts from an optimal mixture free tree, and sequentially adds migration edges to account for deviations between the model predictions and the observed data. This ensures that the fit of the model improves with the addition of each edge, but it also leaves the TreeMix algorithm susceptible to falling into local optima (Molloy et al. 2021, Nielsen et al. 2023). To ensure that we selected an optimal number of migration edges in the TreeMix analysis, we used OptM, which exploits the second order rate of change in likelihood to select an ideal number of edges across incremental values on the number of edges (Fitak 2021). The number of edges with the greatest second-order rate of change in likelihood was m = 1 (Figure 2a and 2b). In this admixture graph, house mice from the Atlantic coast of the Iberian Peninsula (i.e. Portugal and southwest Spain) are depicted as sister to all populations of house mice from the Americas. The sole migration edge goes from the southern UK to the base of the clade containing all North American house mice.

The next optimal admixture graph supported under OptM contained four edges (i.e. m = 4; Figure 2c). At m = 4, in addition to the migration edge from southern UK to North America, we observed a migration edge from Spain to Mexico, a migration edge from the UK to the base of eastern North America, and a migration edge from Mexico to Arizona.

The results obtained using TreeMix are consistent with the history of human colonization. Following the voyages of Columbus, numerous ships potentially harboring rodents as stowaways crossed the Atlantic Ocean to transport human migrants and cargo to North and South America (Crosby 1972, Jones et al. 2013, Gabriel et al. 2014). Having commissioned the exploration of Columbus, the Spanish were among the first Western European people to have significant exchanges with the Americas, followed by the Portuguese (Knight and Liss 1991). House mice from the Atlantic coast of the Iberian Peninsula being phylogenetically sister to all house mice in the Americas reflects this historical relationship (Figure 2). The UK commissioned voyages to the Americas in the late 16th century, and concentrated their colonization efforts to North America beginning in the 17th century (Knight and Liss 1991, McCusker 2006). The migration edge from the southern UK to North America reflects this secondary wave of migration following the initial introduction of house mice from the Iberian Peninsula.

To provide further support for the migration events inferred using TreeMix, we used D-suite to compute the *f*-branch statistic to highlight populations connected by excess allele sharing (Figure 2d) (Malinsky et al 2021). The strongest signal

indicates that mice from Mexico share excess derived alleles with mice from southwest Spain, and to a lesser degree, Portugal. This draws a direct connection between the Mexican population and the Spanish population, or more broadly, the Atlantic Iberian house mice, that reflects human colonization history and recapitulates the results found in our phylogenetic analysis. The next strongest signal of gene flow suggests that the ancestral lineage leading to the southern UK shares excess derived alleles with Florida. This signal extends to all the North American samples to a lesser extent, suggesting gene flow from the southern UK to North America, as depicted in the TreeMix analysis (Fig. 2a). The strength in the signal to Florida may be due to a more recent gene flow event into Florida or a higher frequency of introgressed alleles. We also observed signals of admixture in Florida in a previous study (Agwamba and Nachman 2023) using the f_3 statistic and populations from the east coast of North America and France. Here, we used the f_3 statistics as a test for admixture and demonstrate that mice from Florida can be explicitly modeled as a mixture of mice from the UK or North America and mice from the Atlantic Iberian Peninsula or South America (Table 1). The third largest *f*-branch signal indicates gene flow between southwestern Spain and the internal branch leading to all South American house mice.

Ancestral contributions of house mice from Atlantic Iberia and southern UK to the Americas

The phylogenetic tree in Figure 2a shows contributions of mice from the Atlantic coast of the Iberian Peninsula and from the southern UK to mice in the Americas. To directly infer the relative genetic contributions of mice in Atlantic Iberia and southern UK to the mice in the Americas, we performed supervised inference of shared ancestral fractions using ADMIXTURE.

In a supervised estimation of Atlantic Iberian and UK shared ancestral tracts in the American populations (Figure 3a), much of South America's ancestral proportions mapped to Atlantic Iberian ancestry (between 92% and 99%). In the Mexican population, between 5% and 24% mapped to UK ancestry. In the North American population, UK ancestry makes up most of the inferred ancestral tracts. However, both Florida (between 14% and 28%) and western North America (between 0.7% and 25%) possess notable levels of ancestry with Atlantic Iberia.

We also observe similar patterns in the distribution of private alleles (Figure 3b). The proportion of UK private alleles in North America is significantly greater than the proportion of UK private alleles in South America (t-test; $p = 1.64e\text{-}69$). Conversely, the proportion of Atlantic Iberian private alleles is significantly greater in South America compared to North America (t-test; $p = 4.61e\text{-}06$). Analyzing the distribution of private alleles also allowed us to distinguish disproportionate genetic contributions from different Atlantic Iberian populations to South America (Supplementary Figure 2). The northern South American populations, all sampled from Brazil, had a higher proportion of Portuguese private alleles (t-test; $p = 2.99e\text{-}$

12). The southern South American populations, consisting of three Argentinian populations and one Brazilian population directly bordering Argentina, had a lower proportion of Portuguese private alleles and a higher proportion of Spanish private alleles (t-test; $p = 5.18e\text{-}09$). This pattern is also reflected in a supervised admixture analysis which infers largely Portuguese shared ancestral tracts in northern South America, and majority Spanish ancestral tracts in southern South America (Supplementary Figure 2). These patterns closely mirror human history. Spain and Portugal famously divided their claims to the land and resources of the Americas with the 1494 Treaty of Tordesillas (Knight and Liss 1991) which established a meridian granting Spain the lands to the West and granting Portugal the lands to the East. Under this treaty, the western portion of South America, including Brazil, was claimed, and subsequently colonized by the Portuguese.

Patterns and timing of divergence and admixture links house mouse introduction to human colonization history

To determine the approximate time of divergence between Iberian and American populations, we estimated split times using *dadi* via GADMA for representative populations from different geographic areas. Given a generation time of 1 year (Geraldes et al. 2008, Gray et al. 2014), the demographic models of the highest composite likelihood estimated divergence times of approximately 250 to 550 years ago. Each of the four models determined a constant population size for the Iberian population following divergence, while modeling exponential growth for the American populations post-divergence (Figure 4). Notably, the estimated divergence time for eastern North America (551 ± 220 years ago) predates the divergence time for western North America (253 ± 313 years ago). This is also the case for northern South America (439 ± 570 years ago) relative to southern South America (315 ± 287 years ago). Inferred divergence times are dependent on the accuracy of fixed model parameters, such as mutation rate and generation time. Provided the assumption of accuracy in these constraints, the relative timing of the arrival of house mice to each region of the Americas aligns well with human colonization history.

We estimated the timing of admixture from the UK to North America using ALDER. Of the North American populations, only Virginia, Arizona and the populations from northern and southern Utah had significant LD statistics (p-values: 0.0143, 0.0257, 0.0058, and 0.0003) when testing for admixture with the UK (Table 2). The estimated admixture dates based on LD decay rate for those four populations were 209.85 ± 85.80, 114.80 ± 27.70, 331.85 ± 148.84, and 265.80 ± 55.74 years ago, respectively. Therefore, when ALDER successfully detected signals of admixture, the estimated admixture dates were within the past few hundred years, consistent with the timing of European colonization of the Americas.

Parallels between humans and house mice

Humans historically have had a tremendous influence on their environment, impacting ecosystems and affecting the ecology of natural populations. House mice, for example, have lived in association with humans since at least 8000 BCE (Cucchi et al. 2005). Since the advent of agriculture, house mice have exploited stored grain and other accessible food storage, and have lived in human dwellings. As different human populations have migrated and colonized different regions of the world, mice have presumably traveled with them. This study highlights this unique connection between house mice and humans within recent history. The western house mouse (*Mus musculus domesticus*) began to spread throughout Western Europe roughly 3000 years ago (Cucchi et al. 2005). The population structure of house mice in Western Europe includes three genetic clusters corresponding to Northern Europe, the Mediterranean, and the Atlantic coast of the Iberian Peninsula (Figure 1). These genetic clusters generally match well with the genetic structure found among human populations in Western Europe (Lazaridis et al. 2014; Gilbert et al. 2022).

Analysis of population structure, phylogenetic relationships, and gene flow supports a connection between American mice and house mice from the Atlantic coast of the Iberian Peninsula, as well as a connection between North American mice with mice from the southern UK populations of the Northern European cluster. These associations agree with human history, as the Spanish and Portuguese were the first to explore and settle in the Americas in the 16th century, followed by the English in the 17th century. Broadly, the timing of their introduction to America can be gathered from inferred divergence and admixture dates (Figure 4), which all land roughly within the past several hundred years.

These three populations were responsible for the majority of Western European colonization and settlements in the Americas. Ships starting in Western Europe and ultimately ending up in the Americas transported material and human cargo and likely carried mice and other rodents as stowaways (Hopkins 1973, Knight and Liss 1991, Thorton 1998, McCusker 2006, Coleman 2018). Our results support this historical context, drawing a connection between these European population's transatlantic activities and their opportunities to facilitate the introduction of house mice to the Americas (Figure 2, Figure 3).

A supervised estimation of ancestry proportions also corroborates the historical associations between Spain and Portugal to South America, and the UK to North America (Figure 3). The larger relative ancestral imprint of the Atlantic Iberian population in North America compared to the UK ancestral proportions in South America may reflect the Atlantic Iberian population's earlier arrival to the American continents, which was eventually overtaken in North America by mice introduced from the UK. Furthermore, analysis of gene flow specifically highlights a connection between Spain and Mexico (Figure 2). The Spanish colonized Mexico in 1521 and maintained a significant presence until Mexico gained independence in

1821. Arizona also shares a genetic connection to Spain, possibly a vestige of its history as a Spanish and Mexican territory.

Altogether, we discovered striking connections between the house mice found throughout North and South America, and their source populations found in the Western Europe. These connections are in agreement with human colonization history, and reflect the impact of human migration on the distribution of a commensal species.

Methods and Materials

We assembled a genomic dataset comprised of 100 North American house mice derived from two latitudinal transects sampled in eastern and western North America (Phifer-Rixey et al. 2018, Ferris et al. 2021), 93 Central and South American house mice (Gutierrez-Guerrero et al. 2024), and 83 Western European house mice (Agwamba et al. Chapter 2) (Figure 1a). All sequences are publicly available from the NCBI Short Read Archive (SRA PRJNA1050608, PRJNA397150, PRJNA718321, PRJNA776897) or European Nucleotide Archive (ENA PRJEB9450). We extracted the exomic intersect using *bedtools* and generated a merged vcf using *vcftools*, retaining sites with a quality score greater than 30 and for which genotypes were successfully called for 95% of sites (Quinlan and Hall 2010, Danecek et al. 2011). Following filtering, 82,704 sites remained for analysis.

Principal components were computed using PLINK to generate a broad scale genetic comparison of these samples based on genetic variation (Purcell et al 2007). To investigate phylogenetic relationships and obtain a better understanding of global gene flow, we used TreeMix to infer a maximum likelihood population tree with migration events (Pickrell and Pritchard 2012). We use OptM to determine the tree with the ideal number of migration edges using the second order rate of change in likelihood (Fitak 2021). To further support evidence of gene flow, f_3 statistics were computed using qpfstat in ADMIXTOOLS and *f*-branch statistics were computed using D-suite (Patterson et al. 2012, Malinsky et al. 2021).

We used ADMIXTURE to perform a supervised estimation of southern UK and Atlantic Iberian ancestry proportions in the American populations by first subsetting the original variant call file to include only UK, Atlantic Iberian, and American populations. Next, we created a .pop file to specify the individual samples of known ancestry from the UK and Atlantic Iberian populations. We then ran ADMIXTURE using the *--supervised* flag for K = 2. We computed the number of southern UK and Atlantic Iberian private alleles shared among all samples, and Portuguese and Spanish private alleles shared among South American samples using a custom script. We defined a UK private allele as an allele segregating in the UK but not in the Atlantic Iberian populations, and an Atlantic Iberian private allele as an allele segregating in the Atlantic Iberian populations but not in the UK. Similarly, we defined Spanish private alleles as alleles segregating in southwest Spain but not in

Portugal, and Portuguese private alleles as alleles segregating in Portugal but not in southwest Spain. Samples MWN02F and MWN02H from Portugal were removed for these analyses for having an abnormally higher proportion of shared private alleles with the UK populations and were inferred to have higher non-Iberian ancestry proportions than the remaining samples in our unsupervised analysis.

Two-dimensional site frequency spectra were obtained for each Iberian-American population pairing using the *realSFS* program in ANGSD. We restricted demographic analyses to synonymous SNPs, predicted using Variant Effects Predictor, to mitigate the effects of selection (McLaren et al. 2016). We maintained the filters used by Agwamba and Nachman (2023), requiring filtered sites to have a mapping quality of 30, base quality score of 20, and removing of poorly mapped, duplicate, non-unique, or unpaired reads.

We inferred divergence times using *dadi* through GADMA (genetic algorithm for demographic analysis). We limited our inference of estimated split times to representative populations from different geographic areas. Mid-latitude populations of Virginia and Provo, Utah were selected to represent eastern and western North America, respectively. Similarly, Manaus was chosen to represent northern South America for being the most central sampled population longitudinally. Chubut was chosen for being most central latitudinally of the southern South America populations.

GADMA uses a global search heuristic in order to return the best-fit demographic model based on composite likelihood scores. We allowed for 50 independent runs of GADMA, restricting population size changes to a single event before and after divergence. For all demographic analysis, we used a generation time of 1 year and a mutation rate of 4×10^{-9} (Geraldes et al. 2008).

To generate confidence intervals for the parameter estimates, we divided each dataset into non-overlapping 10kb sections and simulated pseudo-replicate datasets by sampling with replacement for bootstrapping under the parameters that generated the maximum likelihood for the best fitting model. The standard deviations for all parameters were derived from uncertainty analysis using the Godambe Information Matrix for composite likelihoods.

Figures

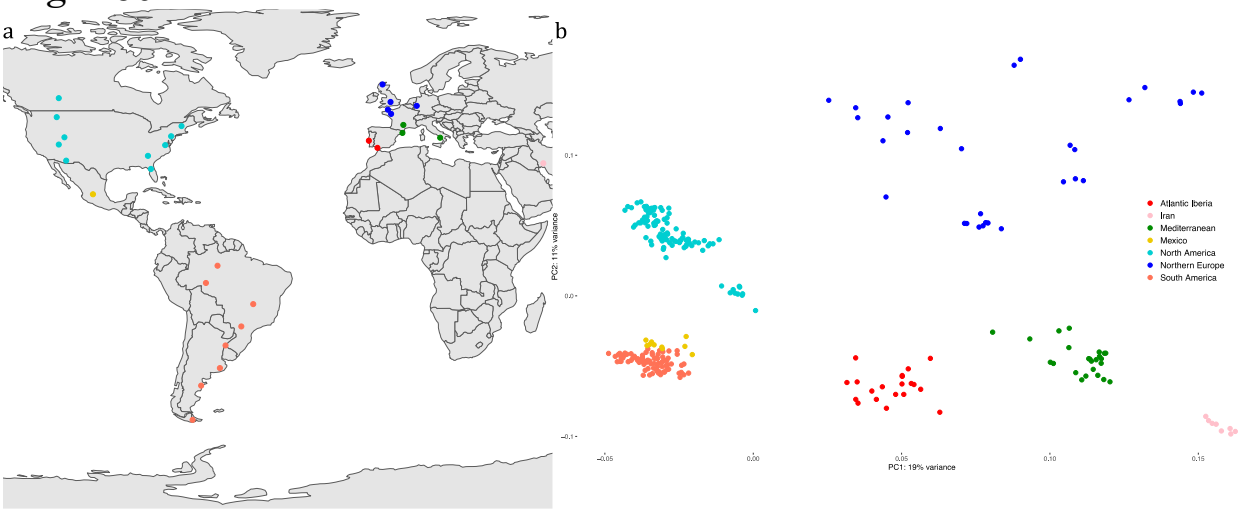


Figure 3.1: Geographic locations and population structure of house mouse samples from Europe, North America, and South America. (a) Map displaying sampling locations of house mice from North America (cyan), Mexico (gold), South America (coral), Atlantic Iberia (red), Mediterranean (forest green), Northern Europe (blue), and Iran (pink) (b) Principal component analysis of house mice from sampled geographic locations based on exomic data

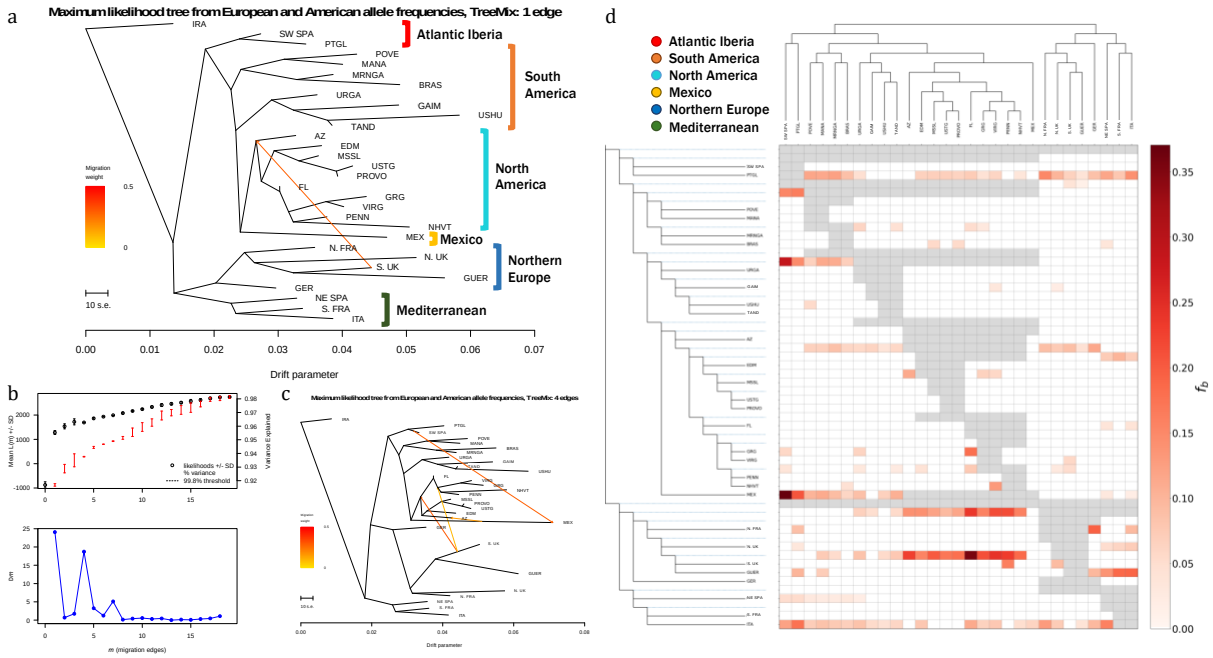


Figure 3.2: Phylogenetic modeling of genetic relationship between Western European and American house mice. (a) Admixture graph of *M. m. domesticus* constructed using TreeMix with one migration edge. The migration edge depicts gene flow from the UK to the base of the North American clade (b) Results from OptM (Fitak 2021). The y-axis of the top figure depicts mean log likelihood and percent of variance explained as a function of the number of migration edges on the x-axis. The bottom graph depicts second order rate of change in likelihood as a function of the number of migration edges. (c) Admixture graph with 4 migration edges. (d) D-suite results depicting *f*-branch statistic in scaled colored matrix. The top axis maps individual populations to columns. The left axis maps internal nodes and individual sampled populations to rows of matrix. Location abbreviations (in alphabetical order): Arizona (AZ), Brasilia (BRAS), Edmonton (EDM), Florida (FL), Gaiman (GAIM), Georgia (GRG), Germany (GER), Guernsey (GUER), Iran (IRA), Italy (ITA), Manaus (MANA), Maringa (MRNGA), Mexico (MEX), Missoula, Montana (MSSL), North France (N. FRA), Northeast Spain (NE SPA), New Hampshire/Vermont (NHVT), Northern UK (N. UK), Pennsylvania (PENN), Porto Velho (POVE), Portugal (PTGL), Provo, Utah (PROVO), South France (S. FRA), Southern UK (S. UK), St. George, Utah (USTG), Southwest Spain (SW SPA), Tandil (TAND), Uruguaiana (URGA), Ushuaia (USHU), Virginia (VIRG)

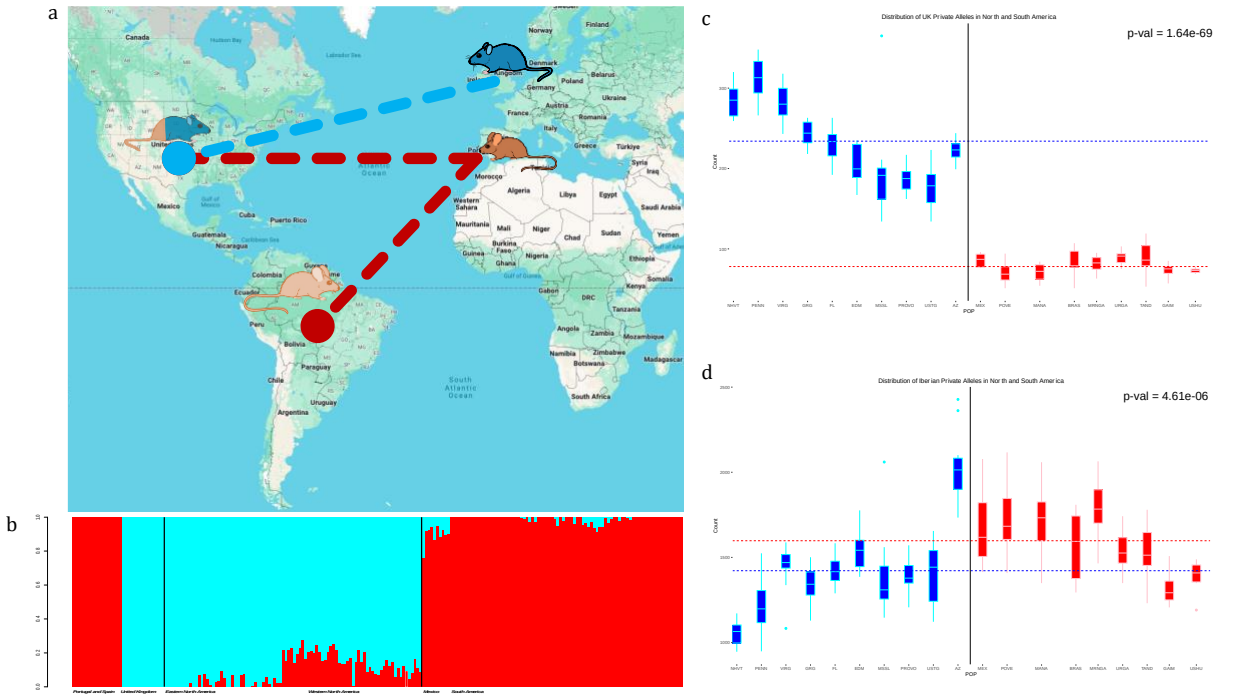
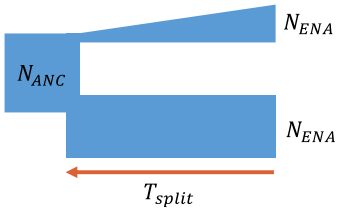


Figure 3.3: Analysis of shared UK and Atlantic Iberian ancestry in North and South America. (a) Schematic detailing the historical relationship between UK and Iberian house mice with recently introduced populations in North and South America. (b) ADMIXTURE plot using supervised analysis mapping ancestral tracts found in North and South America shared with Atlantic Iberian (red) and UK populations (blue) of house mice. (c) Distribution of shared UK private alleles found in North and South America (d) Distribution of shared Atlantic Iberian private alleles found in North and South America. In both figures (c) and (d), the x-axis shows populations with abbreviations as in Fig. 2, and the y-axis shows counts of private alleles. The blue dashed line depicts the mean count for North America. The red dashed line depicts the mean count for South America. The p-value from a one-tailed t-test comparing these means is given in the top right corner of each figure



Population comparisons	T_{split}	± sd
Iberia – Eastern North America (Virginia)	551.05	220.10
Iberia – Western North America (Provo, Utah)	253.03	312.92
Iberia – Northern South America (Manaus, Brazil)	438.95	570.12
Iberia – Southern South America (Gaiman, Argentina)	314.87	286.86

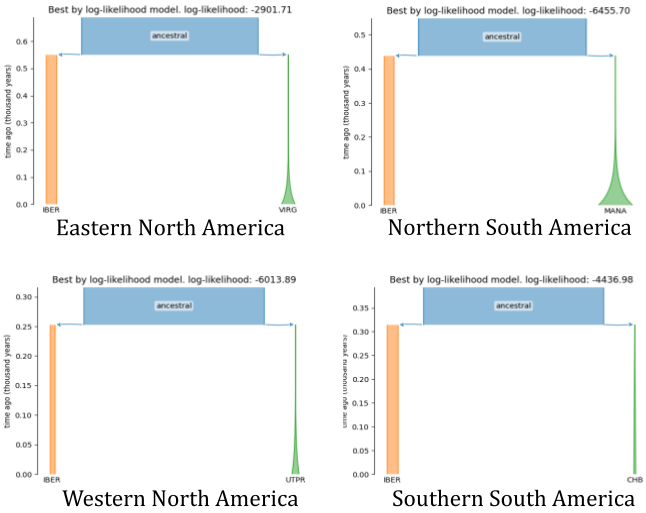


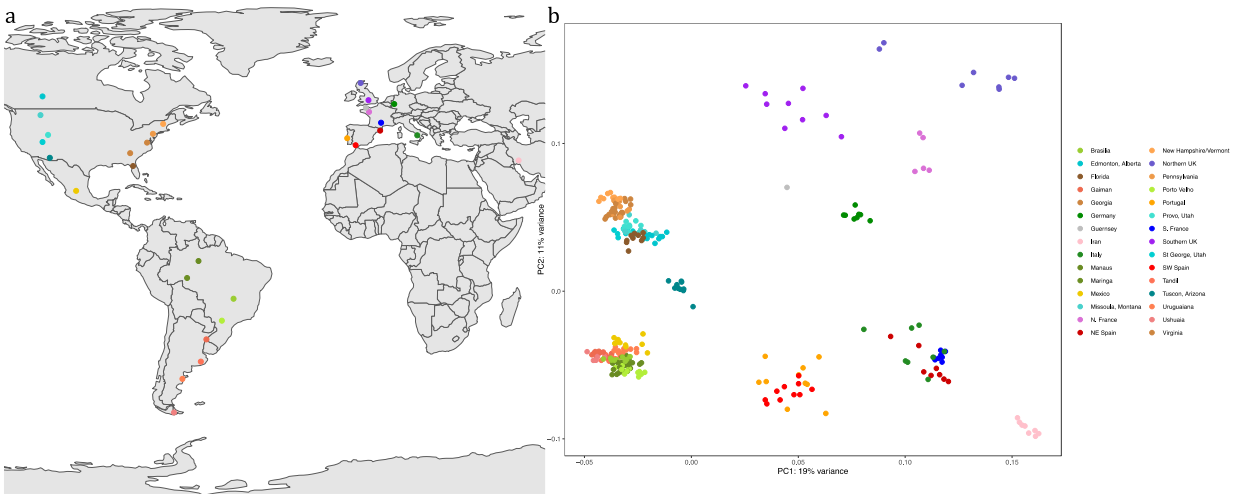
Figure 3.4: Demographic inference of population divergence time parameter between Atlantic Iberia and populations in the Americas. GADMA was used to infer the best demographic model by log-likelihood provided a single population size change before and after divergence. Graphical depiction of GADMA results for Eastern North America, Western North America, Northern South America, and Southern South America are provided on the right. Each of these four figures generated by GADMA depict time on the y-axis and split from an ancestral population leading to relative present-day populations. This demographic model is generically depicted on the top left. The inferred divergence times from each of these models are provided on the bottom left.

Table 3.1: Admixture f_3 statistics for select UK or eastern North American, Atlantic Iberian or South American, and Florida trios. Florida is the only population to confer a negative f_3 statistic as the target population, indicating admixture. This table lists negative f_3 statistics indicating the UK or eastern North American and the Atlantic Iberian or South/Central American as the admixing populations, as first suggested in Agwamba and Nachman (2023).

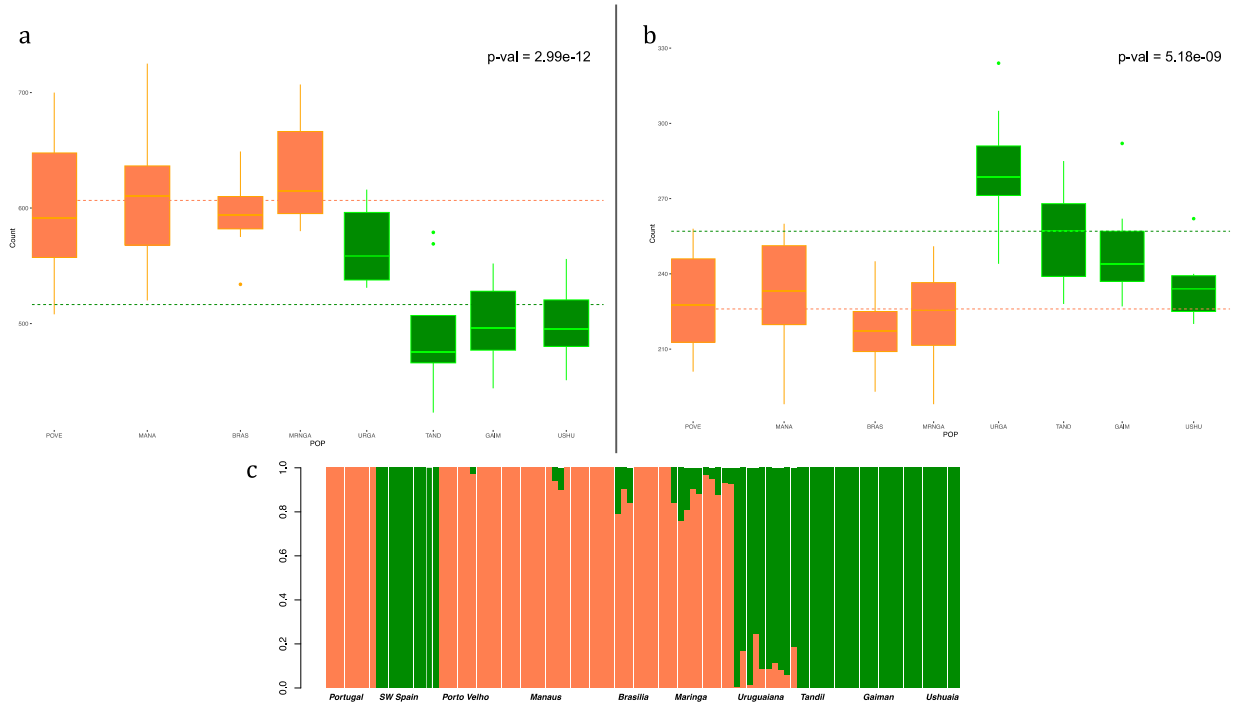
Population A (Target)	Population B	Population C	F_3 statistic	Z-score
Florida	S. UK	MEX	-0.00243433	-3.73913
Florida	S. UK	MANA	-0.0019053	-3.74386
Florida	S. UK	MRNGA	-0.0019812	-3.46088
Florida	S. UK	BRAS	-0.00178675	-2.90069
Florida	GRG	SW SPA	-0.0020653	-4.31123
Florida	VIRG	SW SPA	-0.00194395	-4.26871
Florida	PENN	SW SPA	-0.00158276	-3.98845
Florida	NHVT	SW SPA	-0.00221219	-4.60004
Florida	GRG	PTGL	-0.00199487	-4.15897
Florida	VIRG	PTGL	-0.00121466	-2.42617
Florida	PENN	PTGL	-0.00184533	-4.51525
Florida	NHVT	PTGL	-0.00208428	-3.74194
Florida	GRG	MEX	-0.00236488	-4.44548
Florida	VIRG	MEX	-0.00134646	-2.67603
Florida	PENN	MEX	-0.00162987	-3.86718
Florida	NHVT	MEX	-0.00270788	-4.94425

Table 3.2: ALDER results for all North and Central American populations testing admixture from the UK and southwest Spain. All North and Central American populations have a significant 1-reference LD decay rate with SW Spain. All but the Central American population have a significant 2-referernce z-score. However, ALDER test for admixture minimally requires both 1-reference test to convey significance along with the 2-reference test. This was only accomplished for the North American populations of Virginia; Provo, Utah; St. George, Utah; and Tucson, Arizona.

Population A (Target)	Population B	Population C	Pop B 1-ref decay	Pop B 1-ref amp_exp	Pop B 1-ref z-score	Pop C 1-ref decay	Pop C 1-ref amp_exp	Pop C 1-ref z-score	2-ref decay	2-ref amp_exp	2-ref z-score	2-ref p-value (test status)
New Hampshire/ Vermont	SW Spain	S. UK	130.04 +/- 54.95	0.00152273 +/- 0.00033298	2.37	2.05 +/- inf	0.00019141 +/- inf	0.00	235.09 +/- 72.67	0.00265064 +/- 0.00085381	3.10	0.0019
Pennsylvania	SW Spain	S. UK	285.83 +/- 74.09	0.00150505 +/- 0.00059944	2.51	500.00 +/- inf	0.00231157 +/- inf	0.00	339.33 +/- 90.09	0.00318081 +/- 0.00107273	2.97	0.003
Virginia	SW Spain	S. UK	166.48 +/- 79.53	0.00118894 +/- 0.00044584	2.09	209.85 +/- 85.80	0.00140375 +/- 0.00035540	2.45	280.16 +/- 116.06	0.00255800 +/- 0.00099067	2.41	0.016
Georgia	SW Spain	S. UK	277.06 +/- 123.25	0.00121038 +/- 0.00029587	2.25	457.97 +/- inf	0.00204202 +/- inf	0.00	266.39 +/- 42.37	0.00192032 +/- 0.00054380	3.53	0.00041
Florida	SW Spain	S. UK	368.03 +/- 165.70	0.00085113 +/- 0.00026972	2.22	500.00 +/- inf	0.00199949 +/- inf	0.00	286.47 +/- 57.34	0.00184706 +/- 0.00039434	4.68	2.8e-06
Edmonton	SW Spain	S. UK	301.12 +/- 77.61	0.00133650 +/- 0.00042791	3.12	2.56 +/- inf	0.00022424 +/- inf	0.00	284.24 +/- 100.33	0.00232742 +/- 0.00057484	2.83	0.0046
Missoula, Missouri	SW Spain	S. UK	124.16 +/- 82.26	0.00118150 +/- 0.00036145	1.51	2.00 +/- inf	0.00033930 +/- inf	0.00	220.39 +/- 43.6	0.00213225 +/- 0.00070146	3.04	0.0024
Provo, Utah	SW Spain	S. UK	241.97 +/- 80.04	0.00127714 +/- 0.00029895	3.02	331.85 +/- 148.84	0.00185061 +/- 0.00062207	2.23	213.28 +/- 34.86	0.00188732 +/- 0.00053904	3.50	0.00046
St. George, Utah	SW Spain	S. UK	150.54 +/- 72.57	0.00121719 +/- 0.00029848	2.07	265.80 +/- 55.74	0.00260496 +/- 0.00094279	2.76	194.42 +/- 50.33	0.00178652 +/- 0.00043074	3.86	0.00011
Arizona	SW Spain	S. UK	189.12 +/- 101.67	0.00138588 +/- 0.00028922	1.86	114.80 +/- 27.70	0.00154273 +/- 0.00042549	3.63	334.16 +/- 98.74	0.00169742 +/- 0.00048539	3.38	0.00071
Mexico	SW Spain	S. UK	381.58 +/- 119.82	0.00163260 +/- 0.00047721	3.18	500.00 +/- inf	0.00270287 +/- inf	0.00	500.00 +/- inf	0.00244541 +/- inf	0.00	1



Supplementary Figure 3.1: Detailed sample locations and PCA of exomic data from house mouse samples. This figure is identical to Figure 3.1, but provides the specific localities of each population sample (a) Map of sample locations and (b) PCA of house mouse from Western North Americas (shades of turquoise), Eastern North America (shades of tan), Mexico (gold), northern South America (shades of olive drab green), southern South America (shades of coral), Portugal (orange), SW Spain (red), Italy (forest green), northeast Spain (dark red), southern France (blue), Germany (dark green), northern France (orchid), southern UK (purple), northern UK (slate blue), Guernsey (gray), and Iran (pink)



Supplementary Figure 3.2: Distribution of Atlantic Iberian private alleles in South America. (a) Comparison of Portuguese private alleles found in South America (b) Comparison of Spanish private alleles (c) Supervised inference of Spanish and Portuguese ancestry in South America

Bibliography

Agwamba KD, Nachman MW. 2023. The demographic history of house mice (*Mus musculus domesticus*) in eastern North America. *G3*. 9;13(2): jkac332.

Ajmone-Marsan P, Garcia JF, Lenstra JA. 2010. On the origin of cattle: how aurochs became cattle and colonized the world. *Evolutionary Anthropology: Issues, News, and Reviews*. *19*(4): 148-157.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 19: 1655–1664.

Antonio M , Weiß CL, Gao Z, Sawyer S, Oberreiter V, Moots HM, Spence JP, Cheronet O, Zagorc B, Praxmarer E, Özdoğan KT, Demetz L, Lucci M, Alihodžić T, Amrani S, Avetisyan P, Baillif-Ducros C, Bedić Ž, Bertrand A, Bilić M, Bondioli L, Borówka P, Botte E, Burmaz J, Bužanić D, Candilio F, Cvetko M, De Angelis D, Drnić I, Elschek K, Fantar M, Gaspari A, Gasperetti G, Genchi F, Golubović S, Hukeľová Z, Jankauskas R, Vučković KJ, Jeremić G, Kaić I, Kazek K, Khachatryan H, Khudaverdyan A, Kirchengast S, Korać M, Kozlowski V, Krošláková M, Špalj DK, La Pastina F, Laguardia M, Legrand S, Leleković T, Leskovar T, Lorkiewicz W, Los D, Silva AM, Masaryk R, Matijević V, Cherifi YMS, Meyer N, Mikić I, Miladinović-Radmilović N, Zakić BM, Nacouzi L, Natuniewicz-Sekuła M, Nava A, Neugebauer-Maresch C, Nováček J, Osterholtz A, Paige J, Paraman L, Pieri D, Pieta K, Pop-Lazić S, Ruttkay M, Sanader M, Sołtysiak A, Sperduti A, Pesterac TS, Teschler-Nicola M, Teul I, Tončinić D, Trapp J, Vulović D, Waliszewski T, Walter D, Zivanovic M, el Mostefa Filah M, Čaušević-Bully M, Šlaus M, Boric D, Novak M, Coppa A, Pinhasi R, Pritchard JK. 2022. Stable population structure in Europe since the Iron Age, despite high mobility. *bioRxiv* 2022.05.15.491973.

Aplin KP, Suzuki H, Chinen AA, Chesser RT, ten Have J, Donnellan SC, Austin J, Frost A, Gonzalez JP, Herbreteau V, Catzeflis F, Soubrier J, Fang YP, Robins J, Matisoo-Smith E, Bastos AD, Maryanto I, Sinaga MH, Denys C, Van Den Bussche RA, Conroy C, Rowe K, Cooper A. 2011. Multiple Geographic Origins of Commensalism and Complex Dispersal History of Black Rats. *PLoS ONE* 6(11): e26357.

Ardlie KG, Silver LM. 1996. Low frequency of mouse t haplotypes in wild populations is not explained by modifiers of meiotic drive. *Genetics*. *144*(4): 1787–1797.

Ardlie KG, Silver LM. 1998. Low frequency of t haplotypes in natural populations of house mice (*Mus musculus domesticus*). *Evolution*. 52: 1185-1196.

Arguello JR, Laurent S, Clark AG. 2019. Demographic History of the Human Commensal Drosophila melanogaster. *Genome Biology and Evolution.* 11(3): 844-854.

Astanei I, Gosling E, Wilson J, Powell E. 2005. Genetic variability and phylogeography of the invasive zebra mussel, Dreissena polymorpha (Pallas). *Molecular Ecology*. 14(6): 1655-66.

Banker SE, Bonhomme F, Nachman MW. 2022. Bidirectional introgression between *Mus musculus domesticus* and *Mus spretus*. *Genome Biology and Evolution*. 14(1): evab288.

Belkadi A, Pedergnana V, Cobat A, Itan Y, Vincent QB, Abhyankar A, Shang L, El Baghdadi J, Bousfiha A, Exome/Array Consortium, Alcais A, Boisson B, Casanova JL, Abel L. 2016. Whole-exome sequencing to analyze population structure, parental inbreeding, and familial linkage. *Proceedings of the National Academy of Sciences of the United States of America*. 113(24): 6713-8.

Bennett D. 1978. Population genetics of T/t complex mutations, pp. 615-632. *Origins of Inbred Mice*. Academic Press, New York.

Bennett D, Alton AK, Artzt K. 1983. Genetic analysis of transmission ratio distortion by t-haplotypes in the mouse. *Genetical Research*. 41: 29-45.

Bertolini F, Cardoso TF, Marras G, Nicolazzi EL, Rothschild MF, Amills M; AdaptMap Consortium. 2018. Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution*. 50, 59.

Bonhomme F and Searle JB. 2012. House mouse phylogeography. *In Evolution of the House Mouse* (Cambridge Series in Morphology and Molecules). eds M. Macholán, S.J.E. Baird, P. Munclinger & J. Piálek (Cambridge University Press, Cambridge), pp. 278-296.

Cabe P. 1998. The effects of founding bottlenecks on genetic variation in the European starling (Sturnus vulgaris) in North America. *Heredity*. 80: 519–525.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34: i884–i890.

Coleman, D. 2018. *Henry Smeathman, the Flycatcher: Natural History, Slavery, and Empire in the Late Eighteenth Century*, Romantic Reconfigurations Studies in Literature and Culture 1780, 1850; Liverpool.

Crosby AW. 1972. *The Columbian Exchange; Biological and Cultural Consequences of 1492*. Westport, Conn. Greenwood Pub. Co.

Cucchi T, Vigne JD, Auffray JC. 2005. First occurrence of the house mouse in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences, *Biological Journal of the Linnean Society*. 84 (3): 429–445.

Cucchi T. 2008. Uluburun shipwreck stowaway house mouse: molar shape analysis and indirect clues about the vessel's last journey. *Journal of Archaeological Science*. 35(11): 2953-2959.

Cucchi T, Auffray JC, Vigne JD. 2012. On the origin of the house mouse synanthropy and dispersal in the Near East and Europe: zooarchaeological review and perspectives. In M. Macholan, S. J. E. Baird, P. Munclinger, J. Pialek (Eds.), *Evolution of the House Mouse*. 65-93.

Cucchi T, Papayianni K, Cersoy S, Aznar-Cormano L, Zazzo A, Debruyne R, Berthon R, Bălăşescu A, Simmons A, Valla F, Hamilakis Y, Mavridis F, Mashkour M, Darvish J, Siahsarvi R, Biglari F, Petrie CA, Weeks L, Sardari A, Maziar S, Denys C, Orton D, Jenkins E, Zeder M, Searle JB, Larson G, Bonhomme F, Auffray JC, Vigne JD. 2020. Tracking the Near Eastern origins and European dispersal of the western house mouse. *Scientific Reports*. *10*(1): 8276.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics*. 27(15): 2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*. *10*(2): giab008.

Didion JP, de Villena FP-M. 2013. Deconstructing Mus gemischus: advances in understanding ancestry, structure, and variation in the genome of the laboratory mouse. *Mamm Genome*. 24(0): 1–20.

Doherty TS, Glen AS, Nimmo DG, Ritchie EG, Dickman CR. 2016. Invasive predators and global biodiversity loss. *Proceedings of the National Academy of Sciences of the United States of America*. 113(40): 11261-11265.

Dueñas MA, Ruffhead HJ, Wakefield NH, Roberts PD, Hemming DJ, Diaz-Soltero, H. 2018. The role played by invasive species in interactions with endangered and threatened species in the United States: a systematic review. *Biodiversity and Conservation*. 27: 3171–3183.

Duffy J. 1951. The Passage to the Colonies. *The Mississippi Valley Historical Review*. *38*(1): 21–38.

Duvaux L, Belkhir K, Boulesteix M, Boursot P. 2011. Isolation and gene flow: inferring the speciation history of European house mice. *Molecular Ecology*. *20*(24): 5248-5264.

Early R, Bradley B, Dukes J, Lawler JJ, Olden JD, Blumenthal DM, Gonzalez P, Grosholz ED, Ibañez I, Miller LP, Sorte CJ, Tatem AJ. 2016. Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications*. 7: 12485.

Ewens W. 1972. The sampling theory of selectively neutral alleles, *Theoretical Population Biology*. 3: 87–112.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*. 14: 2611– 2620.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLOS Genetics*. 9(10): e1003905.

Ferris, KG, Chavez AS, Suzuki TA, Beckman EJ, Phifer-Rixey M, Bi K, and Nachman MW. 2021. The genomics of rapid climatic adaptation and parallel evolution in North American house mice. *PLoS Genetics*. 17(4): e10094.

Fitak RR. 2021. *OptM*: estimating the optimal number of migration edges on population trees using *Treemix*. *Biology Methods and Protocols*. *6*(1): bpab017.

Förster DW, Gündüz I, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle, JB. 2009. Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*. 18: 4477-4494.

Fujiwara K, Kawai Y, Takada T, Shiroishi T, Saitou N, Suzuki H, Osada N. 2022. Insights into *Mus musculus* population structure across Eurasia revealed by whole-genome analysis. *Genome Biology and Evolution*. *14*(5): evac068.

Gabriel SI, Jóhannesdóttir F, Jones EP, Searle JB. 2010. Colonization, mouse-style. *BMC Biology*. 8: 131.

Gabriel SI, Mathias ML, Searle JB. 2015. Of mice and the 'Age of Discovery': the complex history of colonization of the Azorean archipelago by the house mouse (Mus

musculus) as revealed by mitochondrial DNA variation, *Journal of Evolutionary Biology*. 28: 130–145.

García-Rodríguez O, Hardouin EA, Hambleton E, Monteith J, Randall C, Richards MB, Edwards CJ, Stewart JR. 2021. Ancient mitochondrial DNA connects house mice in the British Isles to trade across Europe over three millennia. *BMC Ecology and Evolution*. *21*(1): 9.

Genovesi P, Carnevali L, Alonzi A, Scalera R. 2012. Alien mammals in Europe: updated numbers and trends, and assessment of the effects on biodiversity. *Integrative Zoology*. 7(3): 247–253.

Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu H-T, Bulatova N, Ziv Y, Nachman MW. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*. 17(24): 5349–63.

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. *Molecular Ecology*. 20(22): 4722–36.

Gilbert E, Shanmugam A, Cavalleri GL. 2022. Revealing the recent demographic history of Europe via haplotype sharing in the UK Biobank. *Proceedings of the National Academy of Sciences USA*. 119: e2119281119.

Giuffra E, Kijas JMH, Amarger V, Carlborg Ö, Jeon J-T, Andersson L. 2000. The Origin of the Domestic Pig: Independent Domestication and Subsequent Introgression, *Genetics*. 154(4): 1785–1791.

Glanville J. 2009. 16th Century Spanish Invasion of Southwest Virginia. *Historical Society of Western Virginia Journal*. XVII(l): 34-42.

Gray MM, Wegmann D, Haas RJ, White MA, Gabriel SI, Searle JB, Cuthbert RJ, Ryan PG, Payseur BA. 2014. Demographic history of a recent invasion of house mice on the isolated island of Gough. *Molecular Ecology*. 23(8): 1923–1939.

Gündüz İ, Auffray JC, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG, Mathias ML, Searle JB. 2001. Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology* 10(8):2023–9.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics*. 5(10): e1000695.

Gutiérrez-Guerrero YT, Phifer-Rixey M, Nachman MW. Across two continents: the genomic basis of environmental adaptation in house mice (*Mus musculus domesticus*) from the Americas. bioRxiv [Preprint]. 2023 Nov 2:2023.10.30.564674. doi: 10.1101/2023.10.30.564674. PMID: 37961195; PMCID: PMC10634997.

Hammer MF, Schimenti J, Silver LM. 1989. Evolution of mouse chromosome 17 and the origin of inversions associated with t haplotypes. *Proceedings of the National Academy of Sciences USA*, 86(9), 3261-3265.

Harper K. 2017. The Fate of Rome: Climate, Disease, and the End of an Empire. Princeton University Press.

Harr B, Karakoc E, Neme R, Teschke M, Pfeifle C, Pezer Ž, Babiker H, Linnenbrink M, Montero I, Scavetta R, Abai MR, Molins MP, Schlegel M, Ulrich RG, Altmüller J, Franitza M, Büntge A, Künzel S, Tautz D. 2016. Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*. *Sci. Data* 3:160075.

Heather, P. 2007. *The fall of the Roman Empire: a new history of Rome and the Barbarians*. Oxford; New York, Oxford University Press.

Hopkins AG. *An Economic History of West Africa*. 1st ed. Routledge. London: Longman; 1973.

Huang SW, Ardlie KG, Yu HT. 2001. Frequency and distribution of *t*-haplotypes in the Southeast Asian house mouse (*Mus musculus castaneus*) in Taiwan. *Molecular Ecology*. *10*(9): 2349–2354.

Hudson, Charles M. 1997. *Knights of Spain, Warriors of the Sun*. University of Georgia Press.

Hulme PE. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*. 46: 10–18.

Hulme-Beaman A, Dobney K, Cucchi T, Searle JB. 2016. An ecological and evolutionary framework for commensalism in anthropogenic environments. *Trends in Ecology and Evolution*. *31*(8): 633-645.

Jones EP, Eager HM, Gabriel SI, Jóhannesdóttir F, Searle JB. 2013. Genetic tracking of mice and other bioproxies to infer human history. *Trends in Genetics*. 29: 298–308.

Jones, EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB. 2012. Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evol. Biol.* 12, 35.

Kelemen RK, Vicoso B. 2018. Complex history and differentiation patterns of the *t*-haplotype, a mouse meiotic driver. *Genetics*, *208*(1): 365–375.

Kelemen RK, Elkrewi M, Lindholm AK, Vicoso B. 2022. Novel patterns of expression and recruitment of new genes on the *t*-haplotype, a mouse selfish chromosome. *Proceedings of the Royal Society B*. 289: 20211985.

Knight FW, Liss PK. 1991. *Atlantic Port Cities: Economy, Culture, and Society in the Atlantic World, 1650-1850*. 1st Edition. University of Tennessee Press, c1991.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 15(1): 1–13.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prüfer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Rohland N, Delsate D, Francken M, Guinet JM, Wahl J, Ayodo G, Babiker HA, Bailliet G, Balanovska E, Balanovsky O, Barrantes R, Bedoya G, Ben-Ami H, Bene J, Berrada F, Bravi CM, Brisighelli F, Busby GB, Cali F, Churnosov M, Cole DE, Corach D, Damba L, van Driem G, Dryomov S, Dugoujon JM, Fedorova SA, Gallego Romero I, Gubina M, Hammer M, Henn BM, Hervig T, Hodoglugil U, Jha AR, Karachanak-Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kučinskas V, Kushniarevich A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Melegh B, Metspalu E, Molina J, Mountain J, Näkkäläjärvi K, Nesheva D, Nyambo T, Osipova L, Parik J, Platonov F, Posukh O, Romano V, Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekegn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA, Yepiskoposyan L, Zalloua P, Zemunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA, Singh L, Thangaraj K, Villems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, Slatkin M, Pääbo S, Kelso J, Reich D, Krause J. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 513: 409–413.

Lenington S, Franks P, Williams J. 1988. Distribution of *t* haplotypes in natural populations of wild house mice. *Journal of Mammalogy*. 69: 489-499.

Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Hutnik K, Royrvik EC, Cunliffe B; Wellcome Trust Case Control Consortium 2; International Multiple Sclerosis Genetics Consortium; Lawson DJ, Falush D, Freeman C, Pirinen M, Myers S, Robinson M, Donnelly P, Bodmer W. 2015. The fine-scale genetic structure of the British population. *Nature*. 519: 309–314.

Lewontin RC, Dunn LC. 1960. The evolutionary dynamics of a polymorphism in the house mouse. *Genetics.* 45: 705-722.

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

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

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv* https://arxiv.org/abs/1303.3997

Lippens, C., Estoup, A., Hima, MK, Loiseau A, Tatard C, Dalecky A, Bâ K, Kane M, Diallo M, Sow A, Niang Y, Piry S, Berthier K, Leblois R, Duplantier JM, Brouat C. 2017. Genetic structure and invasion history of the house mouse (*Mus musculus domesticus*) in Senegal, West Africa: a legacy of colonial and contemporary times. *Heredity*. 119: 64–75

Liu KJ, Steinberg E, Yozzo A, Nakhleh L. 2015. Interspecific introgressive origin of genomic diversity in the house mouse. *Proceedings of the National Academy of Sciences of the United States of America*. 112(1): 196–201.

Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics*. 193(4): 1233-1254.

Lohmueller KH, Bustamante CD, Clark AG. 2010. The Effect of Recent Admixture on Inference of Ancient Human Population History. *Genetics*. 185(2): 611-622.

Luís C, Bastos-Silveira C, Cothran EG, do Mar Oom M. 2006. Iberian Origins of New World Horse Breeds. *Journal of Heredity*. 97(2): 107–113.

Malinsky, M, Matschiner, M, Svardal, H. 2021. Dsuite - Fast *D*-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* 21: 584– 595.

Mallez S, McCartney M. 2018. Dispersal mechanisms for zebra mussels: population genetics supports clustered invasions over spread from hub lakes in Minnesota. *Biological Invasions*. 20: 2461–2484.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 26(22): 2867–2873.

Mathias ML, Antonio M. 1992. On the origin and colonization of house mice in the Madeira Islands, *Biological Journal of the Linnean Society* 46 (1-2): 13–24

McCusker JJ. 2006. *History of World Trade since 1450*. Macmillan Reference USA, Thomson Gale, c2006.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biology*. 17: 122.

Molloy EK, Durvasula A, Sankararaman S 2021. Advancing admixture graph estimation via maximum likelihood network orientation. *Bioinformatics (Oxford, England). 37* (Suppl_1): i142–i150.

Morgan AP, Hughes JJ, Didion JP, Jolley WJ, Campbell KJ, Threadgill DW, Bonhomme F, Searle JB, de Villena FP. 2022. Population structure and inbreeding in wild house mice (*Mus musculus*) at different geographic scales. *Heredity. 129: 183–194.*

Mu JC, Jiang H, Kiani A, Mohiyuddin M, Asadi NB, Wong WH. 2012. Fast and accurate read alignment for resequencing, *Bioinformatics*. 28(18): 2366-2373.

Nielsen SV, Vaughn AH, Leppälä K, Landis MJ, Mailund T, Nielsen R. 2023. Bayesian inference of admixture graphs on Native American and Arctic populations. *PLoS Genetics*. 19(2): e1010410.

Noskova E, Ulyantsev V, Koepfli K, O'Brien, SJ, Dobrynin P. 2020. GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. *GigaScience*. 9: 1-18.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. 2008. Genes mirror geography within Europe. *Nature.* 456: 98–101.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics*. 192(3): 1065–1093.

Phifer-Rixey M, Bi K, Ferris KG, Sheehan MJ, Lin D, Mack KL, Keeble SM, Suzuki TA, Good JM, Nachman MW. 2018. The genomic basis of environmental adaptation in house mice. *PLoS Genetics*. 14(9): e1007672.

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 Evolutionary Biology*. 20: 120.

Phifer-Rixey M, Nachman MW. 2015. Insights into mammalian biology from the wild house mouse *Mus musculus*. *eLife*. 4: e05959.

Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*. 8(11): e1002967.

Pimentel D, Lach L, Zuniga R, Morrison D. 2000. Environmental and economic costs of nonindigenous species in the United States. *BioScience*. 50(1): 53–65.

Pocock MJO, Hauffe HC, Searle JB. 2005. Dispersal in house mice. *Biological Journal of the Linnean Society*. 84: 565–583.

Puckett EE, Park J, Combs M, Blum MJ, Bryant JE, Caccone A, Costa F, Deinum EE, Esther A, Himsworth CG, Keightley PD, Ko A, Lundkvist Å, McElhinney LM, Morand S, Robins J, Russell J, Strand TM, Suarez O, Yon L, Munshi-South J. 2016. Global population divergence and admixture of the brown rat (Rattus norvegicus). *Proceedings of the Royal Society B: Biological Sciences*. 283: 20161762.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*. 81: 559–575.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. (26)6: 841–842.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rando JC, Pieper H, Alcover JA. 2014. Radiocarbon evidence for the presence of mice on Madeira Island (North Atlantic) one millennium ago. *Proceedings of the Royal Society B: Biological Sciences*. 281: 20133126.

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

Robbins C. 1973. *Introduction, Spread, and Present Abundance of the House Sparrow in North America*. Ornithological Monographs, (14), 3-9.

Rosevear DR. 1969. *Rodents of West Africa*. British Museum of Natural History: London.

Ruvinsky A, Polyakov A, Agulnik A, Tichy H, Figueroa F, Klein J. 1991. Low diversity of *t* haplotypes in the eastern form of the house mouse, *Mus musculus* L. *Genetics*. 127: 161–168.

Salcedo T, Geraldes A, Nachman MW. 2007. Nucleotide variation in wild and inbred mice. *Genetics*. *177*(4): 2277–2291.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*. 46(8): 919–925.

Searle JB, Jones CS, Gündüz I, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry RJ, Giménez MD, Jóhannesdóttir F. 2009. Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society B: Biological Sciences*. 276: 201-207.

Song Y, Endepols S, Klemann N, Richter D, Matuschka FR, Shih CH, Nachman MW, Kohn MH. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*. 21(15): 1296-1301.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9): 1312–1313.

Stringham SA, Mulroy EE, Xing J, Record D, Guernsey MW, Aldenhoven JT, Osborne EJ, Shapiro MD. 2012. Divergence, convergence, and the ancestry of feral populations in the domestic rock pigeon. *Current Biology*. (22)4: 302-308.

Suzuki H, Nunome M, Kinoshita G, Aplin KP, Vogel P, Kryukov AP, Jin ML, Han SH, Maryanto I, Tsuchiya K, Ikeda H, Shiroishi T, Yonekawa H, Moriwaki K. 2013. Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity*. 111(5): 375–390.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*.

Tavaré S, Balding D, Griffith R, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics*. 145: 505–518.

Terhorst J, Kamm JA, and Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*. *49*(2): 303–309.

Thornton J. 1998. *Africa and Africans in the Making of the Atlantic World, 1400–1800*. 2nd ed. Cambridge University Press.

Weissbrod L, Marshall FB, Valla FR, Cucchi T. 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. U.S.A.* 114: 4099–4104.

Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F. 2007. On the subspecific origin of the laboratory mouse. *Nature Genetics*. 39(9): 1100-7.

Yang H, Ding Y, Hutchins LN, Szatkiewicz J, Bell TA, Paigen BJ, Graber JH, de Villena FP, Churchill GA. 2009. A customized and versatile high-density genotyping array for the mouse. *Nature Methods*. 6(9): 663-6.

Yu H, Jamieson A, Hulme-Beaman A, Conroy CJ, Knight B, Speller C, Al-Jarah H, Eager H, Trinks A, Adikari G, Baron H, Böhlendorf-Arslan B, Bohingamuwa W, Crowther A, Cucchi T, Esser K, Fleisher J, Gidney L, Gladilina E, Gol'din P, Goodman SM, Hamilton-Dyer S, Helm R, Hillman JC, Kallala N, Kivikero H, Kovács ZE, Kunst GK, Kyselý R, Linderholm A, Maraoui-Telmini B, Marković N, Morales-Muñiz A, Nabais M, O'Connor T, Oueslati T, Quintana Morales EM, Pasda K, Perera J, Perera N, Radbauer S, Ramon J, Rannamäe E, Sanmartí Grego J, Treasure E, Valenzuela-Lamas S, van der Jagt I, Van Neer W, Vigne JD, Walker T, Wynne-Jones S, Zeiler J, Dobney K, Boivin N, Searle JB, Krause-Kyora B, Krause J, Larson G, Orton D. (2022) Palaeogenomic analysis of black rat (*Rattus rattus*) reveals multiple European introductions associated with human economic history. *Nature Communications* 13, 2399.