

Inferring the Demographic History and Joint Distribution of Fitness Effects  
in the Wild House Mouse, *Mus musculus domesticus*

By

Olivia Fernflores

---

A Thesis Submitted to The W.A. Franke Honors College

In Partial Fulfillment of the bachelor's degree  
With Honors in

Molecular and Cellular Biology

THE UNIVERSITY OF ARIZONA

M A Y 2 0 2 4

Approved by:

Professor Ryan Gutenkunst  
Molecular and Cellular Biology

# Table of Contents

<b>Abstract</b>	<b>2</b>
<b>Introduction</b>	<b>3</b>
Site Frequency Spectra	3
Demographic History	4
Distribution of Fitness Effects (DFE)	4
<b>Results</b>	<b>6</b>
Demography	6
Modeling	6
Iran and France	6
France and Germany	9
Germany and Heligoland	11
DFE	13
<b>Discussion</b>	<b>15</b>
<b>Methods</b>	<b>16</b>
Data Processing	16
Demography Inference	16
DFE Inference	17
<b>References</b>	<b>17</b>

# Abstract

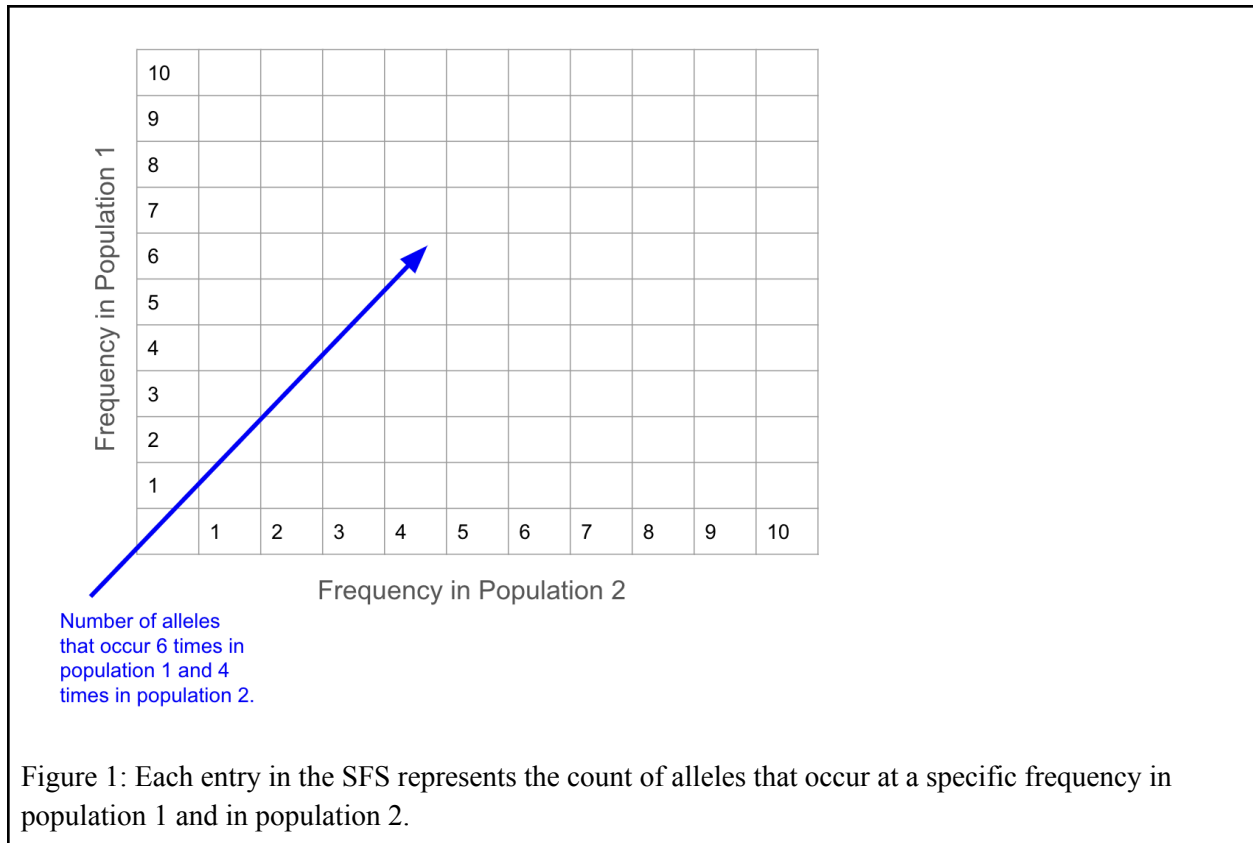
Much can be learned about a species' recent evolutionary past by fitting models to contemporary patterns of genetic variation. We aim to infer the distribution of mutation fitness effects (DFE) among multiple populations of wild house mice, so that the extensive knowledge of mouse molecular biology can be leveraged to understand the biological basis of the DFE. To infer the DFE, we first use synonymous mutations to infer a model of demographic history. Inferring a demographic history can be done for a single population or for two populations and helps us learn about population size(s), divergence time(s), migration rate(s), and level(s) of inbreeding. We then use the demographic model describing two populations to create a set of frequency spectra for nonsynonymous sites under a range of strengths of selection (selection coefficients), which allows us to infer the DFE. A distribution of fitness effects provides information about what proportions of mutations in nonsynonymous sites are deleterious, neutral, and advantageous, which can provide key input into the evolutionary process. We inferred demographic histories and DFEs for pairs of populations of *Mus musculus domesticus* from Iran and France, France and Germany, and Germany and Heligoland and found that in all population pairs, the best demographic models are those that include migration between populations following a distinct split into two populations and that account for potential inbreeding with populations. We also found that distributions of fitness effects had very high to perfect correlations for each population pair.

## Introduction

The mouse is a commonly used model system in genetics and biomedical research, but much of this work is done in classic inbred lines. These classic laboratory strains have reduced genetic diversity in comparison to their wild counterparts, with some aspects like the mitochondrial lineage being traceable to a single *Mus musculus domesticus* ancestor (Salcedo et al., 2007), (Goios et al., 2007). Because the wild *Mus musculus domesticus* shares about 92% of their genome with classic laboratory strains for whom we have a significant base of genetic knowledge (Yang et al., 2007), wild populations of mice are particularly interesting to study. In this paper we seek to infer joint demographic histories and joint distributions of fitness effects of three pairs of four populations *Mus musculus domesticus*.

## Site Frequency Spectra

The site frequency spectrum (SFS, also known as allele frequency spectrum) is a common method in population genetics for summarizing genomic data within a dataset (Sawyer & Hartl, 1992). Because we are interested in looking at joint demographic histories and joint DFEs, this project uses two-dimensional site frequency spectra, which display the frequencies of how many times each allele occurs in each population. For two diploid populations of five individuals each, there is a maximum of ten occurrences of each allele in the population. To create a 2D SFS, counts of how many times each allele occurs in population one and population two are compiled into a grid as shown in Figure 1. This grid is commonly visualized as a heat map showing the density of alleles occurring at each combination of frequencies in the two populations.



## Demographic History

In population genetics, inferring the demographic history of a population or pair of populations is often a goal. A demographic history describes a population or set of populations over time and seeks to explain their relationship through migrations, splits, and size changes. Thanks to advances in genetics and computing, we can infer demographic history computationally using genomic data. There are a variety of ways that demographic history can be inferred, including hidden Markov models, approximate bayesian computation methods, and site frequency spectra based approaches (Marchi et al., 2021). In this project, I use an approach based on site frequency spectra using a diffusion method as part of the software dadi (Gutenkunst et al., 2009). The demographic inferences provide information about parameters describing population size(s), divergence time(s), migration rate(s), and level(s) of inbreeding. These parameters help complement archaeological evidence regarding population history and guide exploration of selection in the genome across a population.

## Distribution of Fitness Effects (DFE)

A key question in population genetics is understanding what selective pressures lead certain mutations to fixation or extinction. Across a genome there are many mutations, most of which are deleterious or neutral, with only a few being beneficial. These mutations exist on a spectrum of how deleterious or beneficial they are, with some being strongly deleterious, some being strongly beneficial, and most falling

somewhere in the middle. Where they fall on this spectrum is termed their fitness effect, and this distribution of all the fitness effects of mutations is known as the DFE (Eyre-Walker & Keightley, 2007). For a single population, this distribution is one dimensional and tends to take the shape seen in Figure 2.1. When you are interested in comparing the distribution of fitness effects of new mutations across two populations, you must use a joint distribution of fitness effects of nonsynonymous mutations (Huang et al., 2021). This becomes a three dimensional graph where one axis is the selection coefficient in population one, one axis is the selection coefficient in population two, and the density/height is the combined frequency. The diagonal of this graph outlined in Figure 2.2 shows the distribution when fitness effects between new mutations in the two populations are perfectly correlated. Looking at a joint distribution of fitness effects is a novel method that allows you to compare fitness effects in the two different environments where the two populations live.

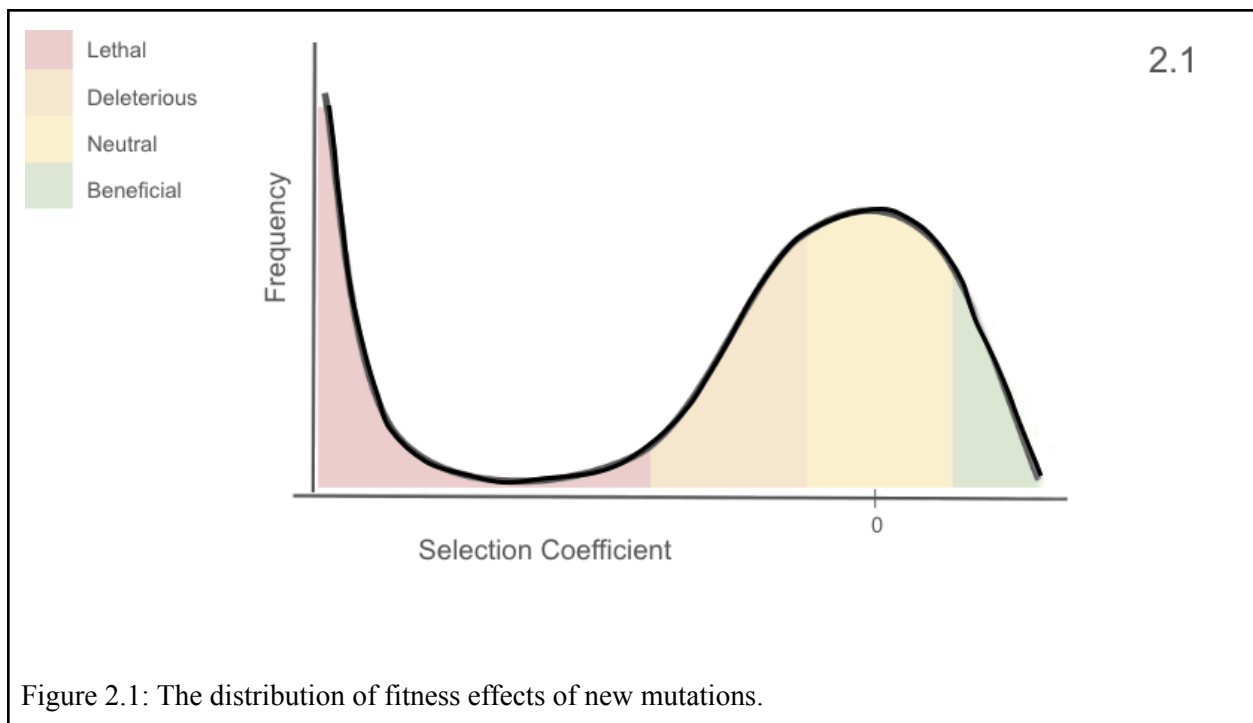


Figure 2.1: The distribution of fitness effects of new mutations.

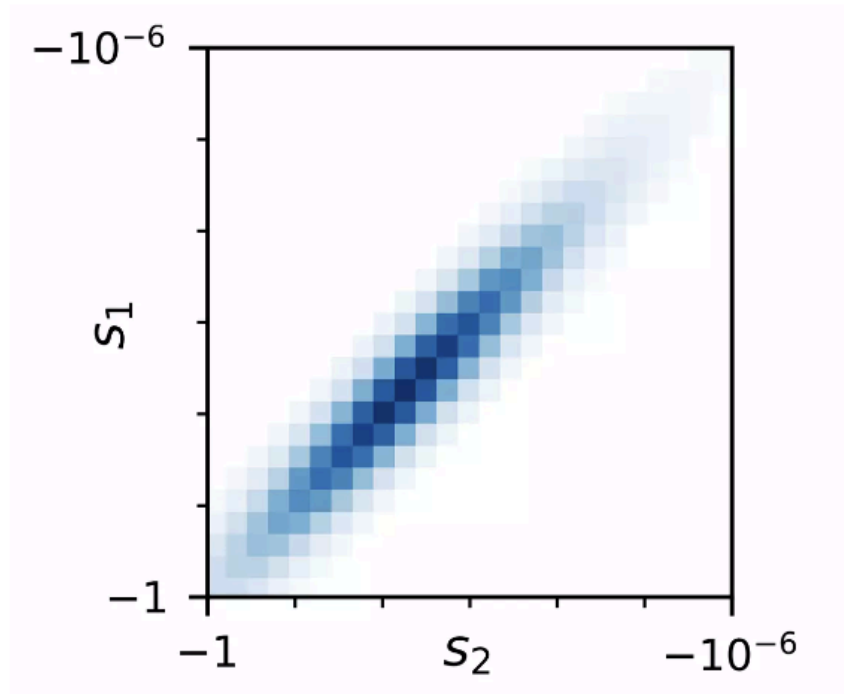


Figure 2.2: The joint distribution of fitness effects with perfect correlation between the two populations. Obtained from (Huang et al., 2021).

## Results

### Demography

#### Modeling

When fitting models, we used a likelihood maximization optimization method to select the best model for each population pair. Demographic models were fit using all synonymous mutations because these are putatively neutral sites that, when combined with mutation rate and generation time, can function as a molecular clock, thus allowing us to infer divergence times (Kimura & Ohta, 1969). We started by fitting the null model for molecular evolution, which is called the standard neutral model. Under this model, we assume that population size is constant, differences between species at the molecular level are due to neutral single nucleotide polymorphisms (SNPs, neutral meaning they have no effect on fitness), and that alternative alleles within a species are equal in fitness effect and vary in number due to a mutation/drift balance (Duret, 2008; Hahn, 2018). After finding that this model didn't fit any of the population pairs very well, we increased the complexity of the models we fit until finding a model that numerically fit well and was biologically reasonable. More information about the specifics of each model can be found in the methods.

## Iran and France

The first two populations that we looked at were sampled from Iran and France (Harr et al., 2016). The Iranian population was sampled in Ahvaz, Iran in 2006 and included eight individuals, three of which were removed from the dataset because they were known to be related to other individuals in the same dataset. The French population was sampled in Massif Central, France in 2005 and includes eight individuals who are not believed to be related to other individuals in the same dataset.

To infer the joint demographic history between these two populations, we ran through a process that started with the simplest demographic history model, the standard neutral model, and increased model complexity until arriving at a well-fitting model. In fitting each model, we looked to maximize the log-likelihood while keeping our parameter estimates in a biologically reasonable range. Unsurprisingly, the standard neutral model did not fit our data well, so we moved to a bottleneck growth model with a split, which infers an ancestral population that underwent an instantaneous size change some time in the past followed by exponential growth and a split into two present day populations. The log likelihood of the bottleneck growth with split model converged at -14306 after 100 optimizations. Because we suspected migration between the two populations after their split, we added an additional parameter to represent the migration rate between the two populations. This improved our fit slightly, with a log likelihood converging at -13995 after 100 optimizations.

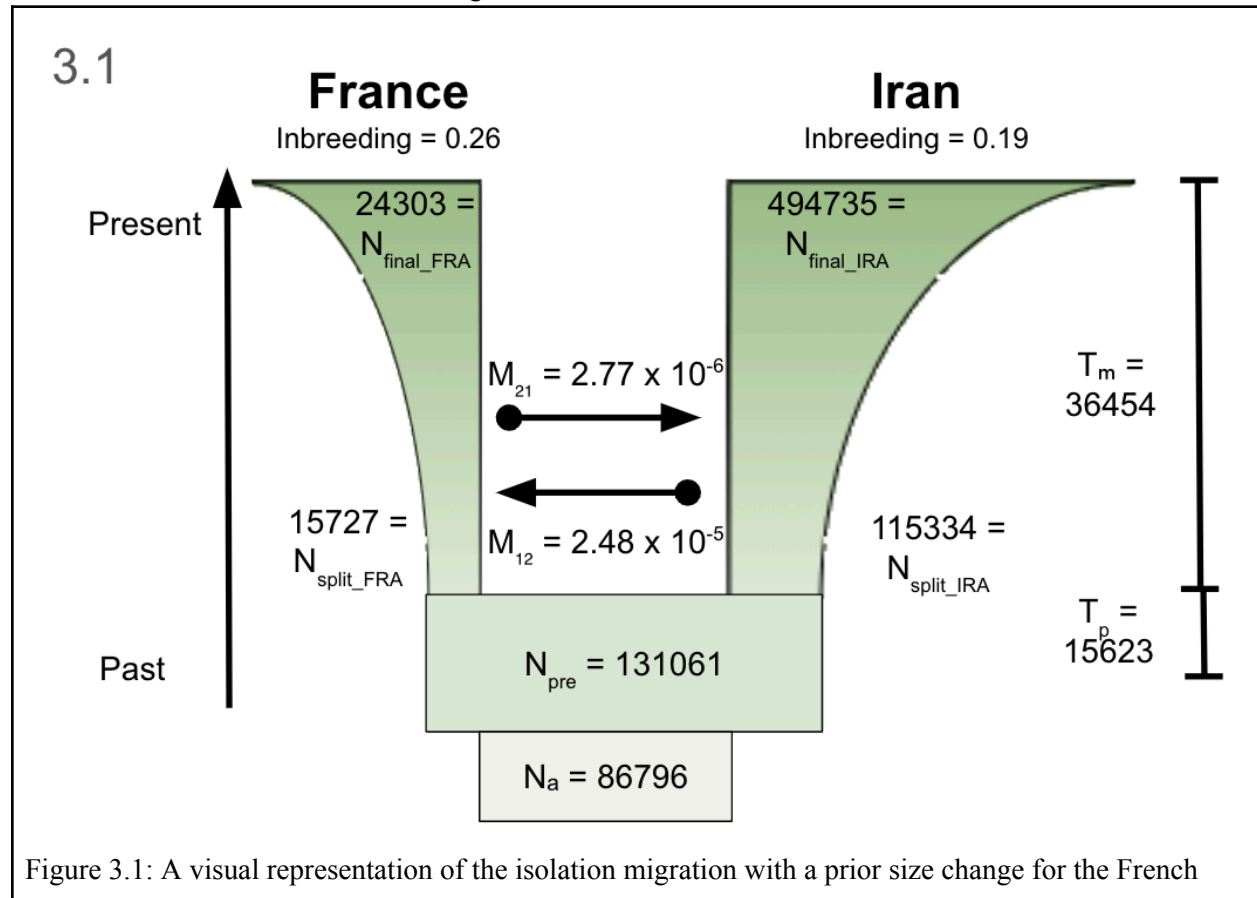
At this stage we wanted to test a model that just included a split of the ancestral populations into two present day populations of fixed size and migration between the two populations at an equal rate in both directions. This split migration model allowed us to test the presence of migration and split into two populations without having the population undergo a bottleneck and then experience exponential growth. The model converged after 100 optimizations at a log likelihood of -1588, which is significantly improved from the log likelihood from the two versions of the bottleneck growth models we tested. To see if we could improve our model even more, we tried two slightly more complex models that are based on the split migration model. One of those, the split asymmetric migration model, allows for asymmetric migration rates between populations, actually made our model fit the data a little bit worse, decreasing the log likelihood to -1679. This model infers two migration rates, one for going from Iran to France and the other for going from France to Iran. In the split asymmetric migration model, the inferred migration rates were 0.28 going from Iran to France and 0.08 going from France to Iran. This differs from the simple split migration model, which inferred a migration rate between populations of 0.43. The other model based on the split migration model that we tried is the split delay migration model, which describes an ancestral population splitting into two populations, then some amount of time passing before asymmetric migration occurs between the two populations. This model also ended up being worse than the simple split migration model, with a log likelihood of -1641.

Related to the three split migration models, we also tried an isolation with migration model, which describes an ancestral population splitting into two populations that undergo exponential growth and exchange migrants at a non-uniform rate. From a biological standpoint, we expected this model to fit better than the split migration models that didn't include exponential growth. When we tested the model, we found an improvement in converged log likelihood to -1465, making the isolation migration model the

best fitting so far. We had one more additional parameter to add to the isolation with migration model, which is a pre-split size change, which improved the log likelihood to -1303.

In wild populations of *Mus musculus domesticus*, there is significant evidence of inbreeding (Morgan et al., 2022). Because inbreeding leads to an overall reduction in heterozygosity (*Population Genetics and Statistics for Forensic Analysts*, n.d.), the underlying assumption in demographic modeling that mating is random is no longer valid. As a result, recent work shows that demographic history inference of inbred populations requires accounting for levels of inbreeding for accurate estimates of demographic history parameters (Blischak et al., 2020). To do this, we tested all of the models we previously tested and included inbreeding coefficients for both the French and Iranian populations as parameters to optimize with each model. For all models, including inbreeding coefficients for each population led to an increase in the log likelihood by at least 200 units, indicating significant improvement.

Once inbreeding was accounted for, our best model was still the isolation migration model with a prehistoric size change and it had a converged log likelihood of -1045 after 100 optimizations. This model inferred a divergence time of about 36000 years ago with about 0.0025% of the individuals in the French population and about 0.00028% of the individuals in the Iranian population being migrants each generation. The present day Iranian population is estimated to be about 20 times larger than the present day French population. Inbreeding coefficients are 0.26 and 0.19 for the French and Iranian populations respectively. A visual representation of the model can be seen in Figure 3.1, and a comparison of the model to the raw data can be seen in Figure 3.2.





and Iranian populations with the best fitting set of parameters. All units for N are the number of individuals and units for time are years.

### 3.2

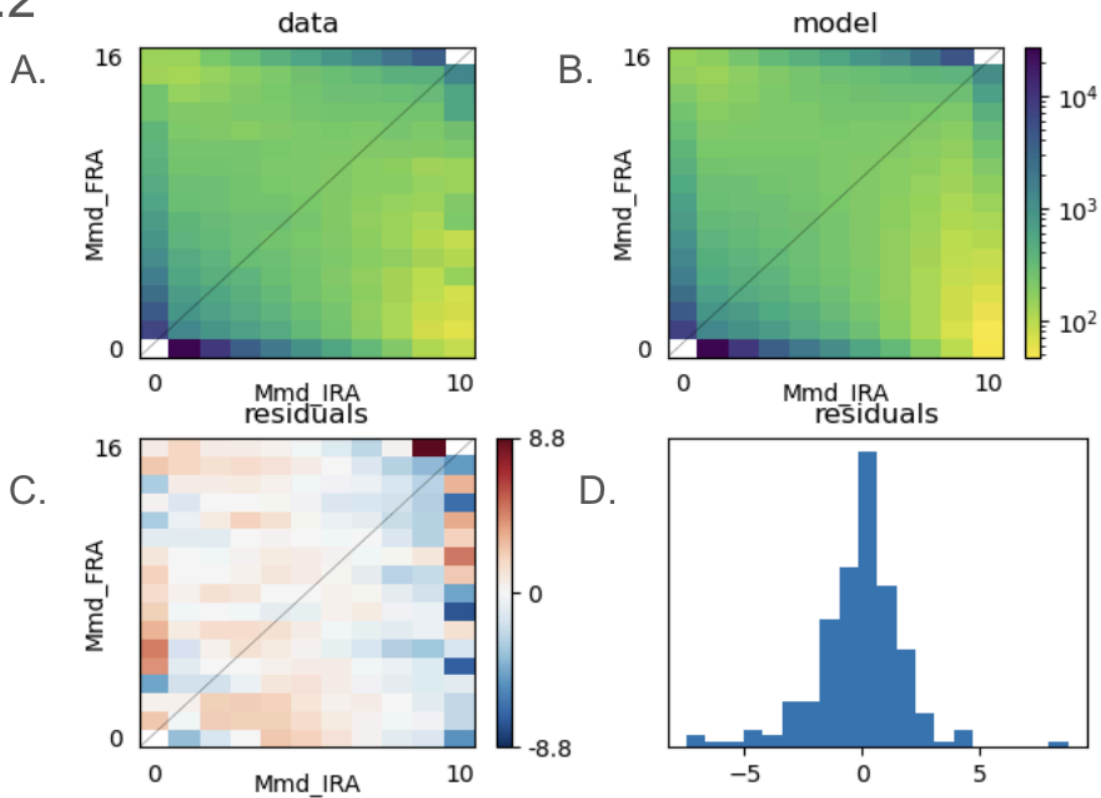


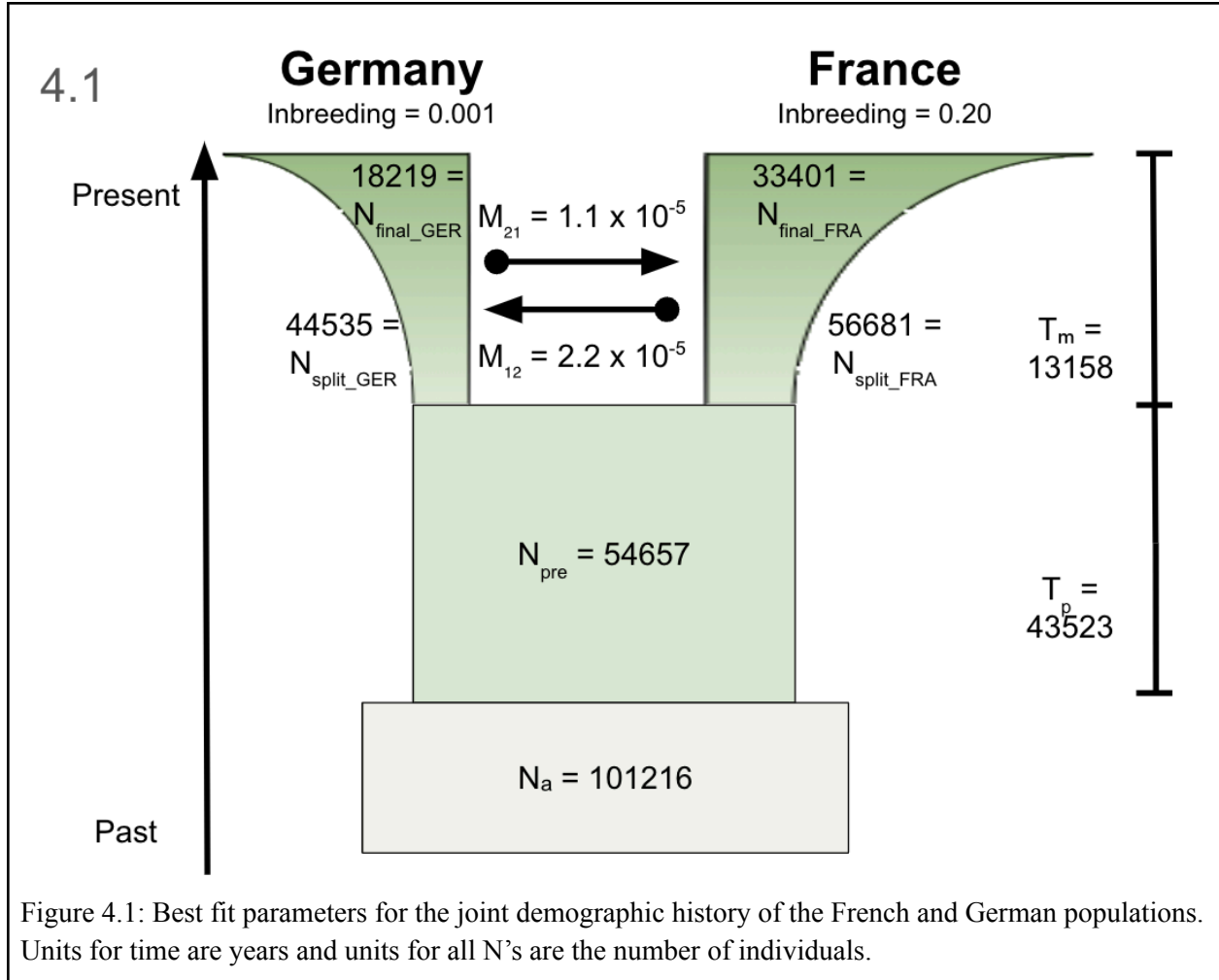
Figure 3.2: The SFS from the raw data (A) and the expected SFS under an isolation migration model with an ancestral size change and inbreeding (B). Residuals for each site in the SFS (C) and distribution of residuals (D).

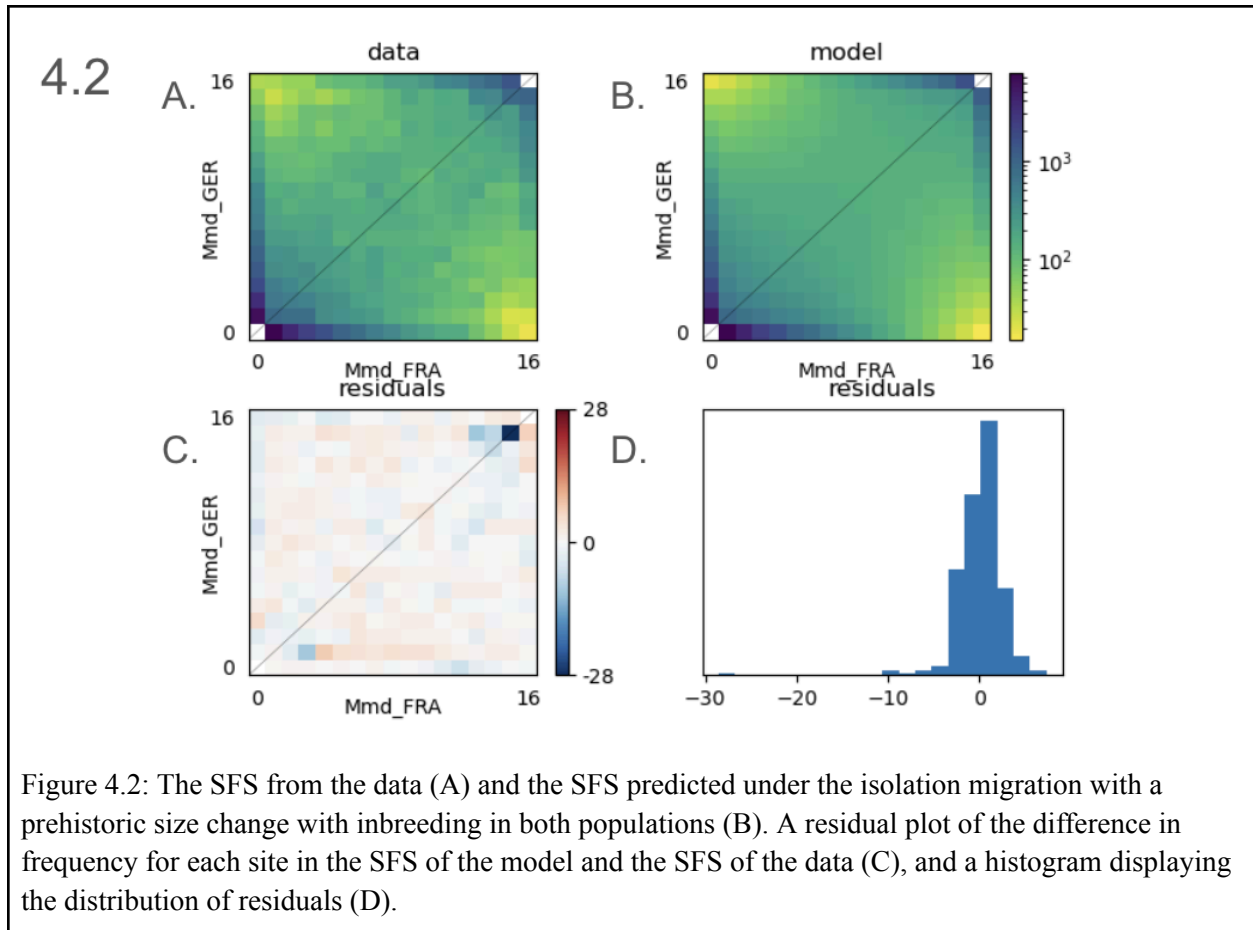
## France and Germany

After looking at the demographic histories of the French and Iranian populations, we decided to analyze the French and German populations, as there is evidence of a migration route of *Mus musculus domesticus* from the Middle East through Western Europe before moving north (Harr et al., 2016). Because of this known migration route, we expected the German and French populations to have diverged more recently than the Iranian and French populations. The German population used for this analysis was sampled in Cologne-Bonn, Germany in 2006 and contains eight individuals, none of whom are known to be related to other individuals in the dataset.

To infer the demographic history of these two populations, we followed the same progression as with the Iranian and French population, starting with the simplest model possible and building complexity to find a model that's a numerically good fit and biologically reasonable. The best model is the isolation migration model with a prehistoric instantaneous size change and inbreeding in both populations. Inbreeding coefficients for the French and German populations are 0.2 and 0.001 respectively. The divergence time

between the French and German populations is estimated to be about 13000 years, with the instantaneous size change happening about 44000 years before the split. Figure 4.1 shows a visual representation of the estimated parameters and 4.2 shows a comparison between the raw SFS for the two populations and the SFS under our best fit parameters.





## Germany and Heligoland

The final pair of populations we investigated were the German population and a population from Heligoland, an island off the coast of Germany. The samples from Heligoland were collected from three individuals in 2012. Because the Heligoland population is an island population, it presents an interesting evolutionary case to explore. We know that there is isolation between the two populations, with one living on an island, but we did not know how much, if any, migration there is between the two populations. As mentioned before when discussing the German and the French populations, the expected migration route across Europe would suggest that Heligoland is one of the last places *Mus musculus domesticus* would reach, causing us to expect the most recent divergence time of all population pairs between the German and Heligoland populations.

To complete the demographic inference, we followed the same process as the other two populations and found the best model to be an isolation migration model with inbreeding and a past instantaneous size change. Under this model, the two populations diverged about 1300 years ago with a prior ancestral population size change about 200 years before the split. Inbreeding is estimated to be higher in the Heligoland population than the German population, with inbreeding coefficients of 0.11 and 0.087 respectively. The model and all parameter estimates are shown in Figure 5.1 and a comparison of the SFS of the raw data and the SFS expected by the model is shown in Figure 5.2.

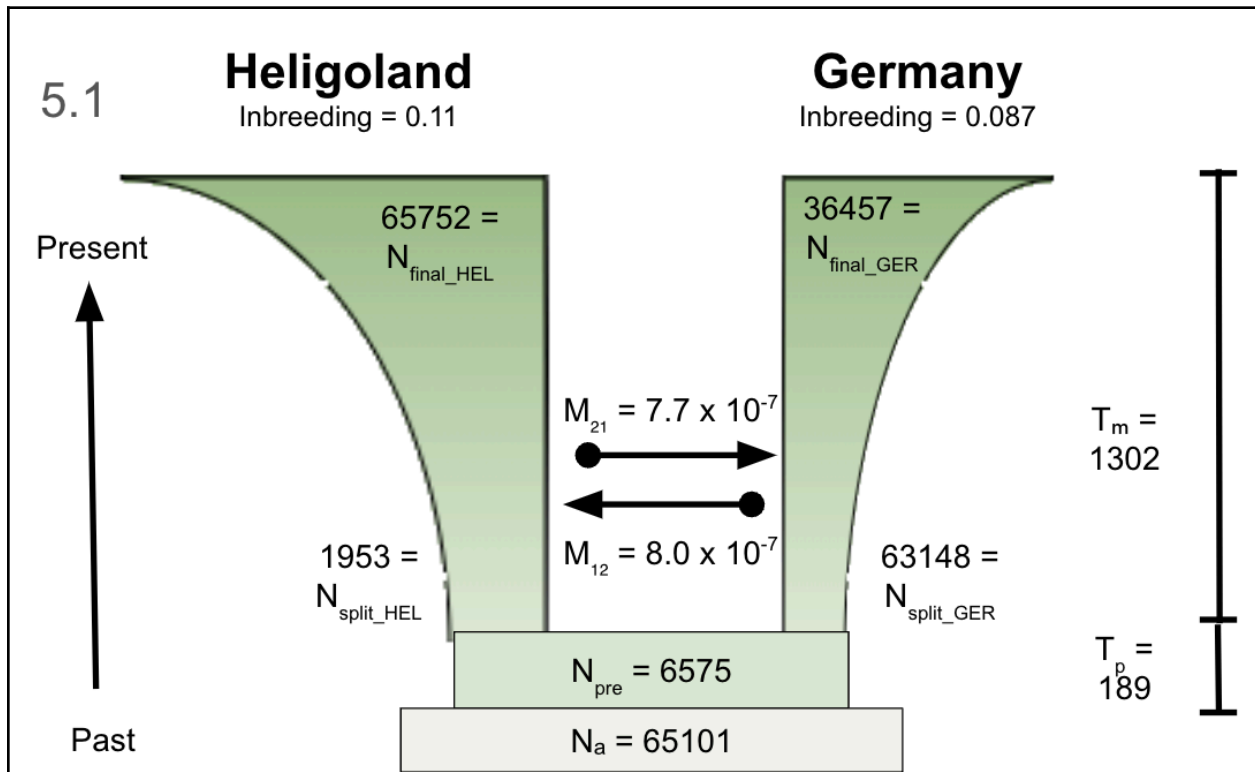


Figure 5.1: The isolation with migration model including inbreeding coefficients for the Heligoland and Germany populations of *Mus musculus domesticus*.

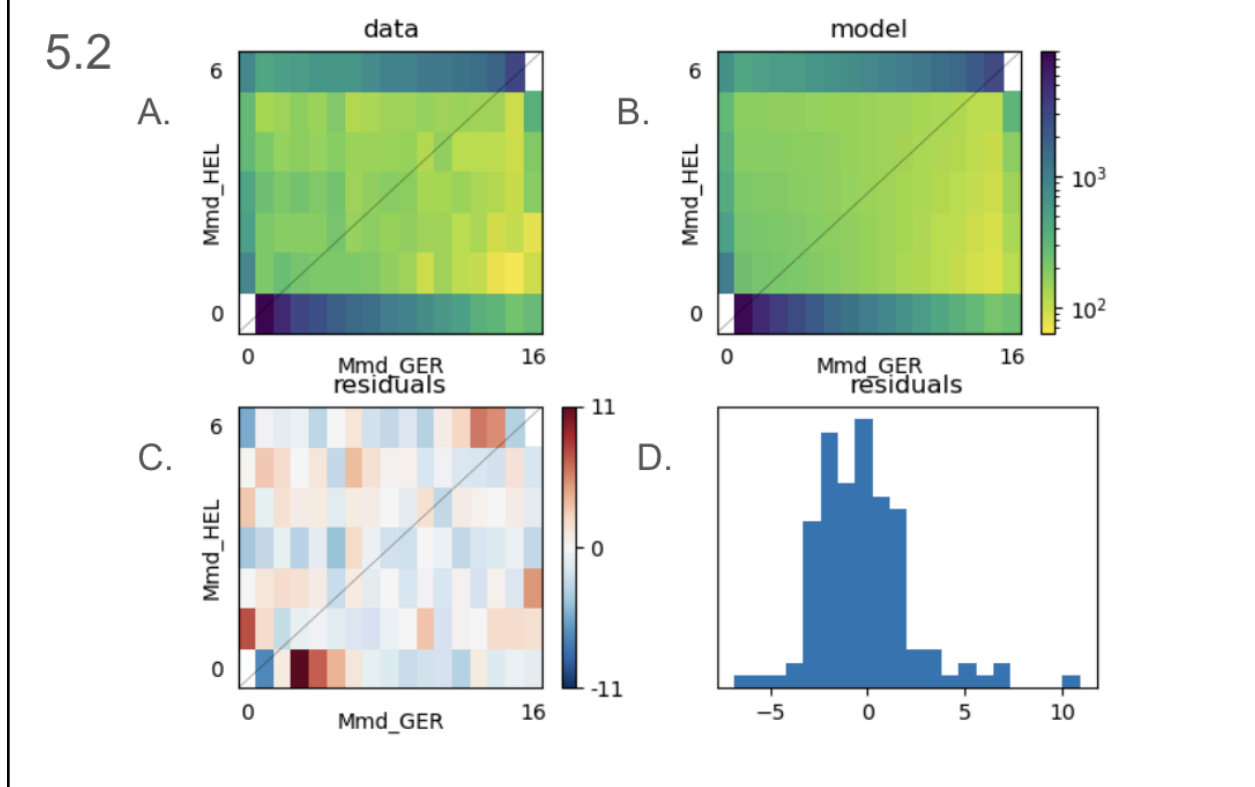
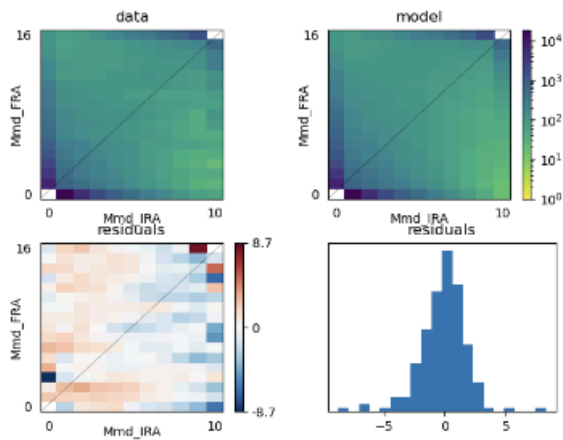


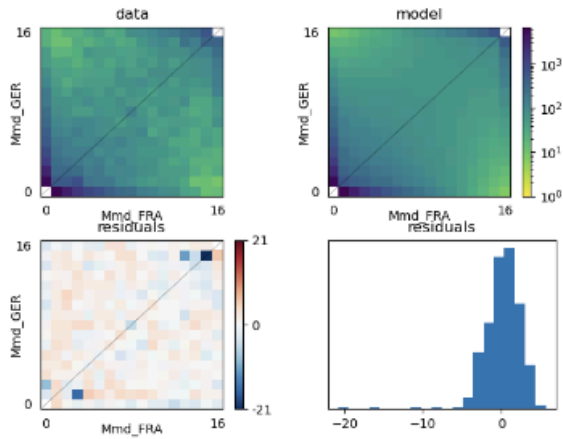
Figure 5.2: Comparison of the SFS from the raw data (A) and the SFS under the model parameters (B). A plot of the difference between the model and the data (residuals) at each site in the SFS (C), and a plot showing the distribution of the residuals (D).

## DFE

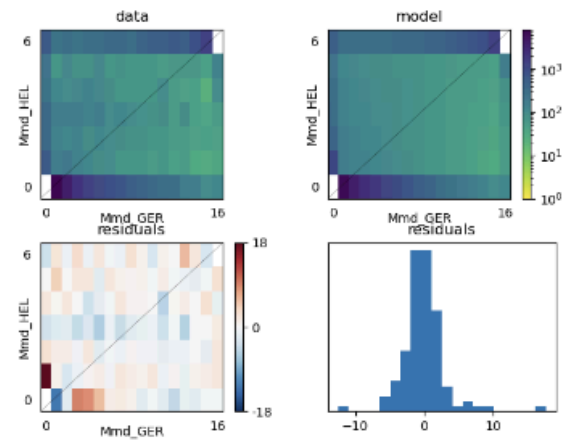
The other main goal of this project was to infer the joint distribution fitness effects for each of the three population pairs. For each pair of populations, we fit lognormal, symmetric bivariate lognormal, and asymmetric bivariate lognormal joint distributions of fitness effects to nonsynonymous mutations. For the joint DFE of the French and Iranian populations, we found a correlation of 0.97 under an asymmetric bivariate lognormal distribution, which had a log likelihood of -954, making it the best of all three models. Because the correlation coefficient is quite close to one for this population pair, we can infer that mutations in the two populations have similar fitness effects. When it comes to the German and French populations, the best joint DFE was the lognormal distribution, which inferred a perfect correlation between fitness effects in the two populations. This model had a log likelihood of -1871. Although the log likelihoods of the symmetric and asymmetric bivariate lognormal distributions were similar (-1870), to avoid overfitting the best model is the lognormal distribution because the result is almost identical and the model is the simplest. With the final population pair, the Germany and Heligoland populations, the best joint-DFE was the symmetric bivariate lognormal distribution, which had a log likelihood of -947 and a correlation of 0.98. Although the asymmetric bivariate lognormal distribution did have a set of optimized parameters with a lower log likelihood (-756), the other parameter estimates under this distribution were unreasonably large. These DFE models can be seen in Figure 6, which shows comparisons of the SFS of nonsynonymous mutations from the raw data and the expected SFS under the best fitting DFE.



6.1: Iran vs. France  
Asymmetric Bivariate  
Lognormal DFE



6.2: France vs. Germany  
Lognormal DFE



6.3: Germany vs.  
Heligoland Symmetric  
Bivariate Lognormal DFE

Figure 6: Plots for the best fitting joint DFE for each population pair.

# Discussion

In this project we sought to infer joint demographic histories and joint distributions of fitness effects for three population pairs of four different populations of *Mus musculus domesticus*. We found that the joint demographic history of the Iranian and French populations as well as the joint demographic history of the French and German populations are best modeled by an isolation migration model with a prior size change that accounts for inbreeding in each population. Our third population pair, the Germany and Heligoland populations, had a joint demographic history best represented by an isolation migration model with inbreeding in both populations. All of our demographic inferences suggest an ancestral population size for *Mus musculus domesticus* of between 65000 and 100000, which matches previous estimates of the ancestral population size for the species (Phifer-Rixey et al., 2020). We found that the best joint-DFE differed slightly for each population pair, with the French and Iranian pair having an asymmetric bivariate lognormal distribution, the German and French pair having a lognormal distribution, and the Heligoland and German pair having a symmetric bivariate lognormal distribution. Our DFE results suggest that in the French and Iranian populations, there is a very high correlation between the deleteriousness of each mutation. This suggests that mutations have similar effects between the two populations. Between the French, German, and Heligoland populations, mutations seem to have the same effect.

Prior research has shown that *Mus musculus domesticus* underwent a species-wide bottleneck as the species migrated from the Middle East and Northern Africa to Western Europe (Phifer-Rixey et al., 2020). Throughout this path of migration, *Mus musculus domesticus* have developed a commensal relationship with humans, which is estimated to have started 12-15 thousand years ago (Agwamba & Nachman, 2022). Within this time period, our estimates of demographic history models show the continuation of existing migration between populations in Iran and France that dates back to about 35 thousand years ago and the beginning of migration between populations in France and Germany. Even in present times, *Mus musculus domesticus* tend to live in close proximity to humans, with many found in residential areas, agricultural land, and commercial establishments (Phifer-Rixey & Nachman, 2015). For the wild populations, like the sources of data for this project, most live in underground burrows or tunnels and that seems to hold true across the wide range of their present day habitats that includes much of North and South America, Australia, Africa, the Middle East, and Western Europe (Phifer-Rixey & Nachman, 2015). This suggests that despite having a complex demographic history involving a commensal relationship with humans, the environment that the populations from Iran, France, Germany, and Heligoland were living in are actually quite similar. For two populations living in similar environments, we would expect little variation at the molecular level because the environmental effect on selection and genetic variance will be similar (Ramakers et al., 2018). Our DFE results suggest high correlation between the fitness effects of mutations in the two populations, which could be driven by the similar environments that all four populations live in. That being said, this is an area where further research is needed to determine if the similar environment is truly driving this correlation or if this correlation would be expected between any two populations from the same species even if they are living in the same environment. Recent work shows that within-species variation in the DFE is very rare, but between-species variation in the DFE is statistically significant even between two closely related species (James et al., 2023). If this is true, then similarities we are seeing among the four populations of *Mus musculus domesticus* are not driven by local adaptations to similar environments, but rather driven by their speciation.

Moving forward, further investigation of the DFE could provide new insights into understanding the biological basis of the DFE. One way to look into this is to do a functional annotation of the sequences we have for each individual in each of the four populations. We can then compare DFEs for different types of genes to see if certain types of genes are more likely to have high correlations of distributions of fitness effects.

## Methods

### Data Processing

This project used genomic sequencing data published by Harr et. al (Harr et al., 2016). Raw data was downloaded in VCF format and was processed using dadi (Gutenkunst et al., 2009). Data processing involved removing relatives from the dataset and writing the ancestral state allele, which was done using *Mus spretus* and *Mus musculus castaneus* as outgroups. Sites were marked as synonymous single nucleotide polymorphisms (SNPs) or nonsynonymous SNPs using Annovar (Wang et al., 2010). This information was used to generate joint allele frequency spectra for synonymous sites and nonsynonymous sites for the population pairs of Iran and France, France and Germany, and Germany and Heligoland.

### Demography Inference

Joint allele frequency spectra for synonymous sites in each population pair were used for demographic history inferences. For each population pair, dadi was used to run 100 optimizations of the following demographic models with and without inbreeding: bottlegrowth split, bottlegrowth split migration, isolation migration, isolation migration with a prior size change, split migration, split asymmetric migration, and split delay migration. Model descriptions are shown in Table 1 below.

Model	Description
Bottlegrowth Split	An instantaneous size change some time in the past, exponential population growth, and a split into two populations.
Bottlegrowth Split Migration	An instantaneous size change some time in the past, exponential population growth, and a split into two populations with migration between the populations.
Isolation Migration	An ancestral population splits into two populations that are isolated from each other and grow exponentially. Migration occurs between the two populations.



Isolation Migration with a Prior Size Change	An instantaneous size change some time in the past followed by a split into two isolated populations that grow exponentially. Migration occurs between the two populations.
Split Migration	An ancestral population splits into two distinct populations of fixed size. Migration occurs between the two populations at a uniform rate.
Split Asymmetric Migration	An ancestral population splits into two distinct populations of fixed size. Migration occurs between the two populations at non-uniform rates in each direction.
Split Delay Migration	An ancestral population splits into two distinct populations of fixed size. Some time in the past after the split, migration occurs between the two populations.

## DFE Inference

Once the best demographic model for each population pair was established, the parameters from the demographic model were used to generate a cache of expected frequency spectra of nonsynonymous mutations under a range of selection coefficients from  $1E-6$  to 2000. These cached frequency spectra were used to infer lognormal, symmetric bivariate lognormal, and asymmetric bivariate lognormal distributions of fitness effects for each population pair. DFE inferences were done in dadi-cli over 15 optimizations, which was always enough optimizations for the results to converge (Huang et al., 2023).

# References

- Agwamba, K. D., & Nachman, M. W. (2022). The demographic history of house mice (*Mus musculus* domesticus) in eastern North America. *G3 Genes|Genomes|Genetics*, *13*(2), jkac332.
- Blischak, P. D., Barker, M. S., & Gutenkunst, R. N. (2020). Inferring the Demographic History of Inbred Species from Genome-Wide SNP Frequency Data. *Molecular Biology and Evolution*, *37*(7), 2124–2136.
- Duret, L. (2008). Neutral Theory: The Null Hypothesis of Molecular Evolution. *Nature Education*, *1*(1), 218.
- Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews. Genetics*, *8*(8). <https://doi.org/10.1038/nrg2146>
- Goios, A., Pereira, L., Bogue, M., Macaulay, V., & Amorim, A. (2007). mtDNA phylogeny and evolution of laboratory mouse strains. *Genome Research*, *17*(3), 293–298.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics*. <https://doi.org/10.1371/journal.pgen.1000695>
- Hahn, M. W. (2018). *Molecular Population Genetics*. Sinauer Associates, Incorporated.
- Harr, B., Karakoc, E., Neme, R., Teschke, M., Pfeifle, C., Pezer, Ž., Babiker, H., Linnenbrink, M., Montero, I., Scavetta, R., Abai, M. R., Molins, M. P., Schlegel, M., Ulrich, R. G., 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*. *Scientific Data*, *3*(1), 1–14.
- Huang, X., Fortier, A. L., Coffman, A. J., Struck, T. J., Irby, M. N., James, J. E., León-Burguete, J. E., Ragsdale, A. P., & Gutenkunst, R. N. (2021). Inferring Genome-Wide Correlations of Mutation Fitness Effects between Populations. *Molecular Biology and Evolution*, *38*(10), 4588–4602.
- Huang, X., Struck, T. J., Davey, S. W., & Gutenkunst, R. N. (2023). dadi-cli: Automated and distributed population genetic model inference from allele frequency spectra. In *bioRxiv* (p.

2023.06.15.545182). <https://doi.org/10.1101/2023.06.15.545182>

- James, J., Kastally, C., Budde, K. B., González-Martínez, S. C., Milesi, P., Pyhäjärvi, T., & Lascoux, M. (2023). Between but Not Within-Species Variation in the Distribution of Fitness Effects. *Molecular Biology and Evolution*, *40*(11), msad228.
- Kimura, M., & Ohta, T. (1969). THE AVERAGE NUMBER OF GENERATIONS UNTIL FIXATION OF A MUTANT GENE IN A FINITE POPULATION. *Genetics*, *61*(3), 763–771.
- Marchi, N., Schlichta, F., & Excoffier, L. (2021). Demographic inference. *Current Biology: CB*, *31*(6), R276–R279.
- Morgan, A. P., Hughes, J. J., Didion, J. P., Jolley, W. J., Campbell, K. J., Threadgill, D. W., Bonhomme, F., Searle, J. B., & de Villena, F. P.-M. (2022). Population structure and inbreeding in wild house mice (*Mus musculus*) at different geographic scales. *Heredity*, *129*(3), 183–194.
- 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*(1), 1–9.
- Phifer-Rixey, M., & Nachman, M. W. (2015). *The Natural History of Model Organisms: Insights into mammalian biology from the wild house mouse Mus musculus*. <https://doi.org/10.7554/eLife.05959>
- Population Genetics and Statistics for Forensic Analysts*. (n.d.). National Institute of Justice. Retrieved April 11, 2024, from <https://nij.ojp.gov/nij-hosted-online-training-courses/population-genetics-and-statistics-forensic-analysts/population-theory/hardy-weinberg-principle/inbreeding>
- Ramakers, J. J. C., Culina, A., Visser, M. E., & Gienapp, P. (2018). Environmental coupling of heritability and selection is rare and of minor evolutionary significance in wild populations. *Nature Ecology & Evolution*, *2*(7), 1093.
- Salcedo, T., Gernaldes, A., & Nachman, M. W. (2007). Nucleotide Variation in Wild and Inbred Mice. *Genetics*, *177*(4), 2277–2291.
- Sawyer, S. A., & Hartl, D. L. (1992). Population genetics of polymorphism and divergence. *Genetics*,

132(4), 1161–1176.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164.

Yang, H., Bell, T. A., Churchill, G. A., & Pardo-Manuel de Villena, F. (2007). On the subspecific origin of the laboratory mouse. *Nature Genetics*, 39(9), 1100–1107.