Inferring Demographic Histories and Distributions of Fitness Effects of Wild House Mice from Allele Frequency Spectra

Olivia Fernflores, David Castellano, Emanuel Fonseca, Travis J. Struck, Ryan N. Gutenkunst

Molecular and Cellular Biology, University of Arizona; oliviafernflores@arizona.edu, https://gutengroup.mcb.arizona.edu/

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. The DFE provides information about what proportions of mutations in nonsynonymous sites are deleterious, neutral, and advantageous, which can provide key input into the evolutionary process. Here we present demographic histories and DFEs for pairs of populations of *Mus musculus domesticus* from Iran and France, France and Germany, and Germany and Heligoland. 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.
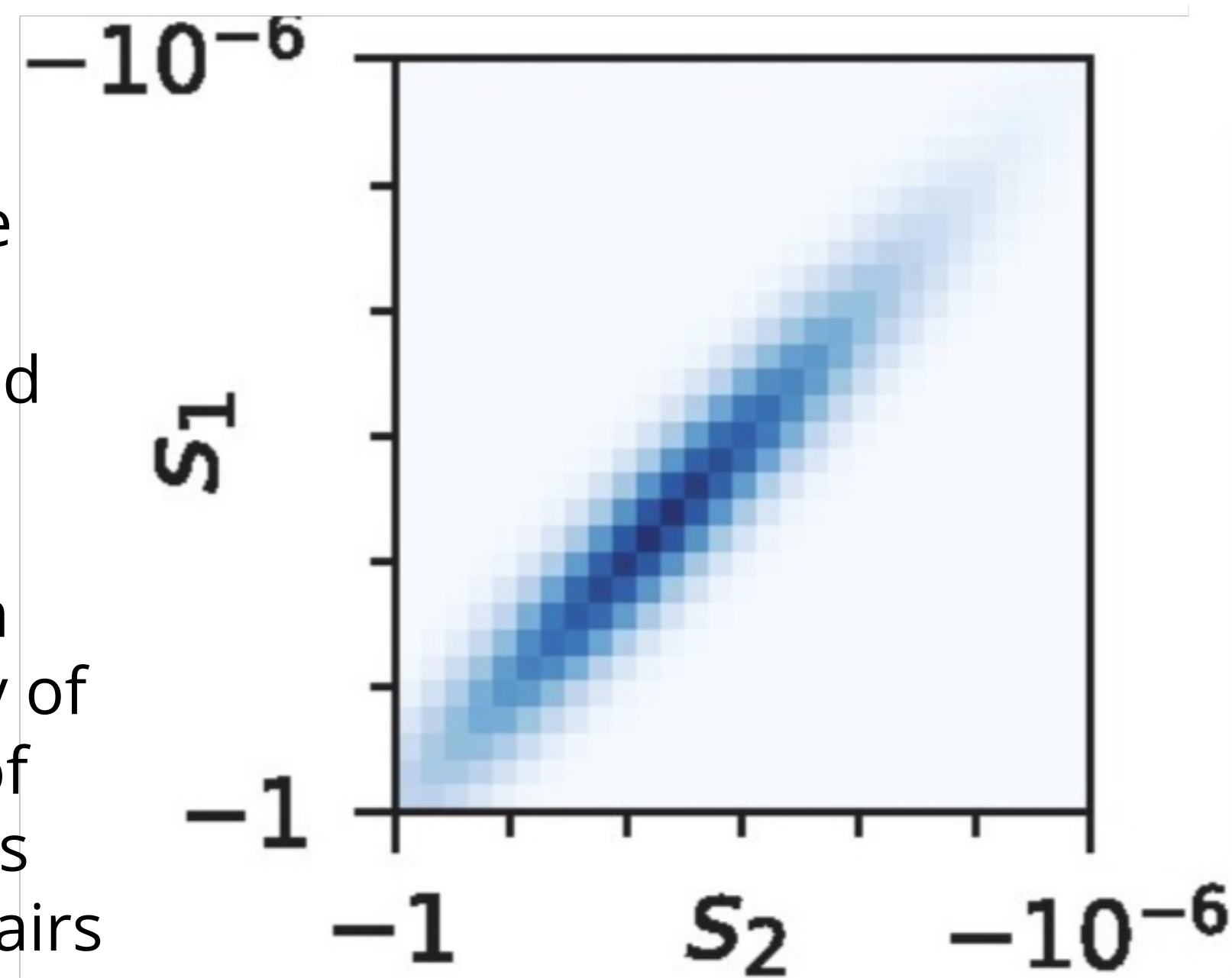
Demographic History

Inferring demographic history is a common goal in population genetics and can done computationally using genomic data. A demographic history describes a population or set of populations over time and seeks to explain their relationship through migrations, splits, and size changes. This project uses a site frequency spectra (SFS) based approach to infer demographic histories for population pairs using a diffusion method, which is implemented in the software *dadi* (1). The parameters from demographic inference describe a population or pair of populations over time, which helps complement archaeological evidence of population history and guides exploration of selection.
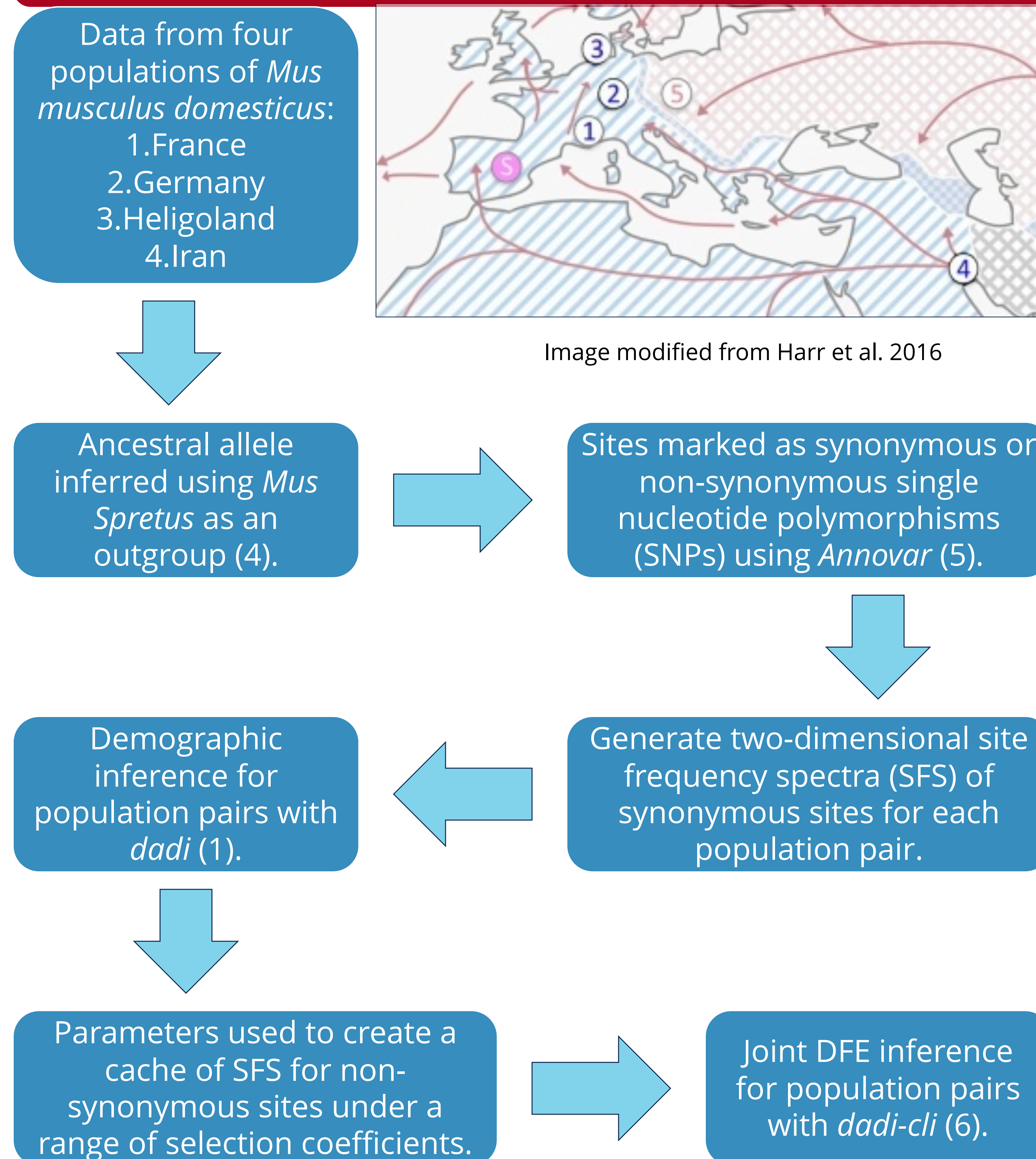
Joint Distribution of Fitness Effects

Another key question in population genetics is understanding what selective pressures contribute to certain mutations reaching fixation or extinction. Each mutation exists on a spectrum of how deleterious it is, which yields a distributions of fitness effects (DFE). Using a joint DFE, we can quantify the correlation between mutation fitness effects in two populations. A high correlation results in more shared high frequency polymorphisms compared to a low correlation.

On the horizontal and vertical axes, we plot the strength of selection with S1 indicating the strength of selection for a mutation in population one and S2 indicating the strength of selection for a mutation in population two (modified from Huang et al. 2021). The density of color represents the number of mutations for both populations that have the corresponding pairs of selection coefficients.
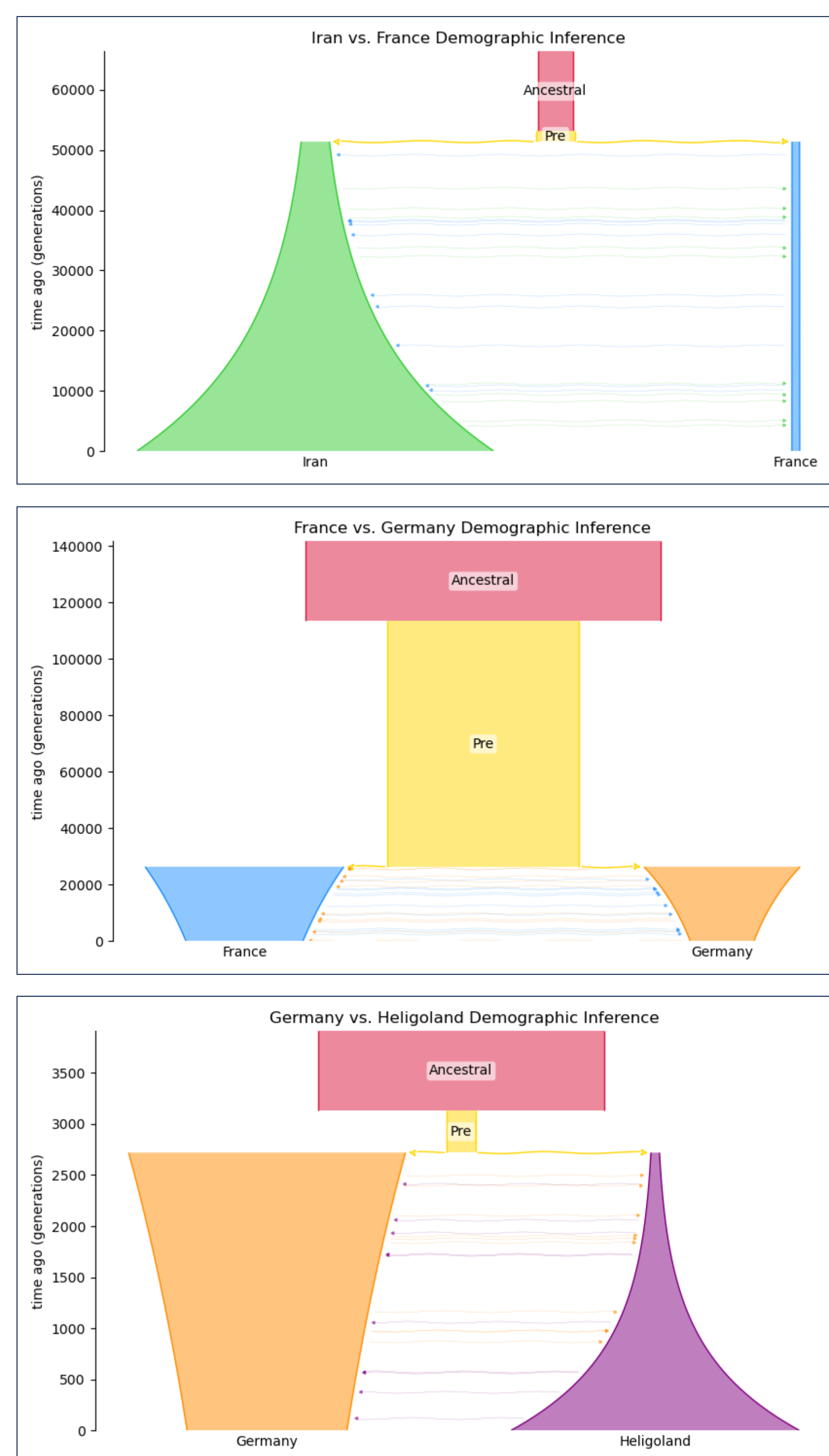

Between populations, we expect to see differences in the fitness effect of a mutation due to a combination of environmental and genetic context, although we do not yet know the importance of either in determining the overall fitness effect of a mutation.

Methods

Data from four populations of *Mus musculus domesticus*:
1. France
2. Germany
3. Heligoland
4. Iran

Image modified from Harr et al. 2016

Ancestral allele inferred using *Mus Spretus* as an outgroup (4). → Sites marked as synonymous or non-synonymous single nucleotide polymorphisms (SNPs) using *Annovar* (5). ↓

Demographic inference for population pairs with *dadi* (1). ← Generate two-dimensional site frequency spectra (SFS) of synonymous sites for each population pair. ↓

Parameters used to create a cache of SFS for non-synonymous sites under a range of selection coefficients. → Joint DFE inference for population pairs with *dadi-cli* (6).

Demographic History Results

For all three population pairs, the best-fitting demographic model is the **isolation-migration-pre model** with inbreeding.

Inferences for all three population pairs estimated an ancestral population size of 60-100 thousand individuals.

According to these models, the populations from Iran and Heligoland have grown since their split from the ancestral population, whereas the German and French populations have shrunk.

Inbreeding Estimates

Iran	0.07
France	0.227
Germany	0.045
Heligoland	0.11

Joint DFE Results
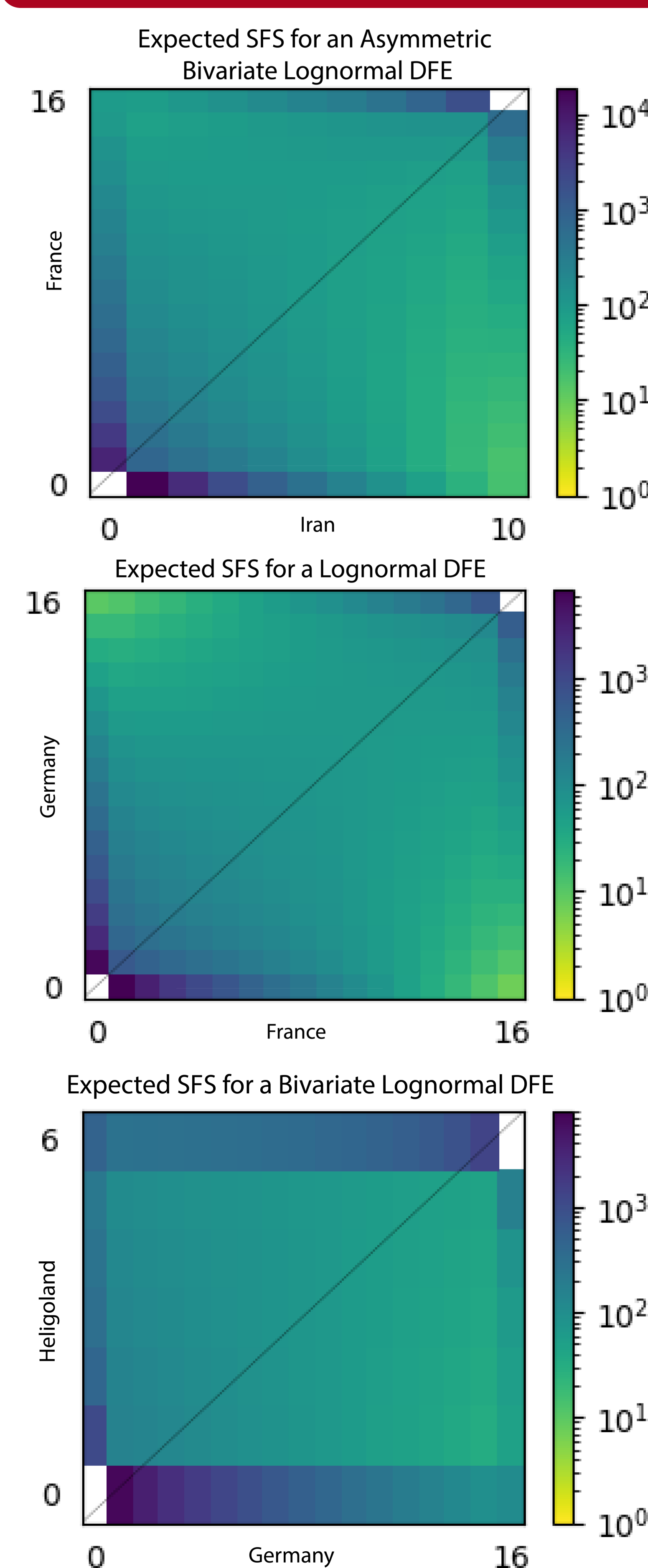
Iran and France

The best-fitting DFE for the Iranian and French populations has a correlation coefficient (rho) of 0.97. In this model, the mean and standard deviation of the distribution for each population are not equal.

Germany and France

The best-fitting DFE for the German and French populations has perfect correlation between fitness effects in the two populations. The mean and standard deviation of the distribution for each population are equal.

Heligoland and Germany

The best-fitting DFE for the populations from Germany and Heligoland has a correlation coefficient (rho) of 0.98. In this model, the mean and standard deviation of the distribution for each population are equal.

Discussion

The demographic inferences presented here describe three population pairs of *Mus musculus domesticus* and each pair is best represented by an isolation-migration-pre model. Our results match known migration patterns of *Mus musculus domesticus* and our estimates of ancestral population sizes and inbreeding levels are consistent with literature(3, 7, 8). Having parameters to describe paired populations allows us to construct a combined demographic history for all four populations, which could, when combined with archeological evidence, help us further understand the demographic background of the species.

The DFE results show high correlation between mutational fitness effects in population pairs. One explanation for this result is that each of the populations in a pair are living in similar environments, and thus subject to similar environmental effects on selection, driving a highly correlated DFE (8, 9). We hope this work will help us understand the biological basis of the DFE by leveraging the extensive knowledge of the common laboratory mouse, whose genomic origin is, on average, 92% *Mus musculus domesticus* (10). To continue this, we plan to look at joint-DFEs for the same population pairs for a subset of mutations located in genes annotated to specific gene ontology terms.

References

1. Gutenkunst et al. (2009) PLOS Genet:1000695.
2. Huang et al. (2021) Mol Biol Evol 38(10):4588-4602.
3. Harr et al. (2016) Sci Data 3:160075.
4. Agwamba & Nachman (2022) G3 13(2):jkac332.
5. Wang et al. (2010) Nucleic Acids Res
6. Huang et al. (2023) bioRxiv:p. 2023.06.15.545182.
7. Morgan et al. (2022) Heredity:129:183
8. Phifer-Rixey & Nachman (2015) eLife
9. Ramakers et al. (2018) Nat Ecol Evol 2(7):1093-1103.
10. Yang et al. (2007) Nat Genet 39(9):1100-1107.

Acknowledgements