Inferring Distributions of Fitness Effects of Wild House Mice From Allele Frequency Spectra

Olivia Fernflores[1], David Castellano[1], Emanuel Fonseca[1], Travis Struck[1], Ryan N. Gutenkunst[1]

[1] Molecular and Cellular Biology, University of Arizona; oliviafernflores@arizona.edu; http://gutengroup.mcb.arizona.edu

Abstract

The distribution of fitness effects (DFE) of new mutations is a key input into the evolutionary process. We aim to infer the 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. Here we present distributions of fitness effects for pairs of populations from Iran and France, France and Germany, and Germany and Heligoland. We find that for all population pairs, the best DFEs are those that infer a high correlation between mutational fitness effects in the two populations.

Joint Distributions of Fitness Effects

The joint distribution of fitness effects quantifies the correlation of mutation fitness effects between two populations. A high correlation will result in more shared high frequency polymorphisms versus a low correlation. On the horizontal and vertical axes, the strength of selection is plotted. S1 indicates the strength of selection for a mutation in population 1 and S2 indicates the strength of selection for a mutation in population 2. Between populations, we expect to see differences in the fitness effect of a mutation due to a combination of differences in environental and genetic context, although we do not yet know the importance of either in determining the overall fitness effect of a mutation (1).

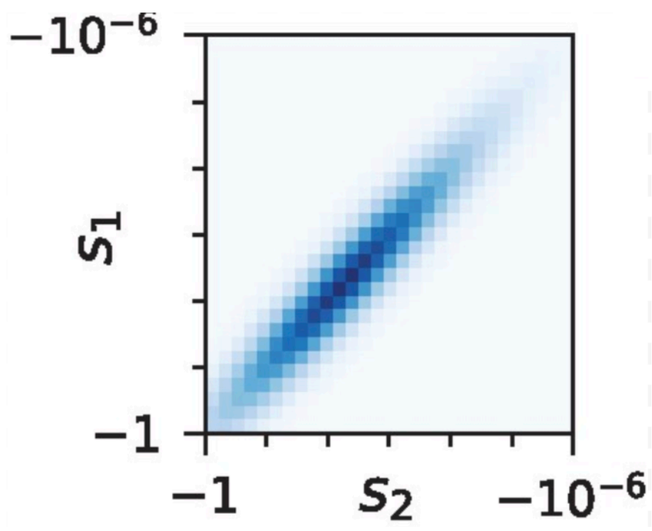


Image from Huang et al. 2021

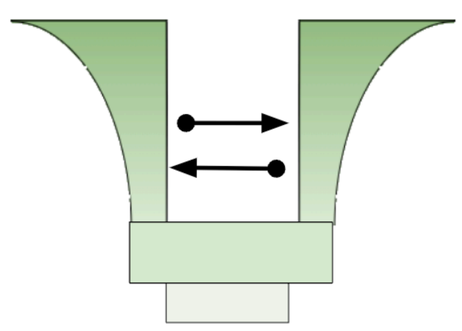
Methods

- We analyzed four populations of *Mus musculus domesticus* (2).

1. France
2. Germany
3. Heligoland
4. Iran

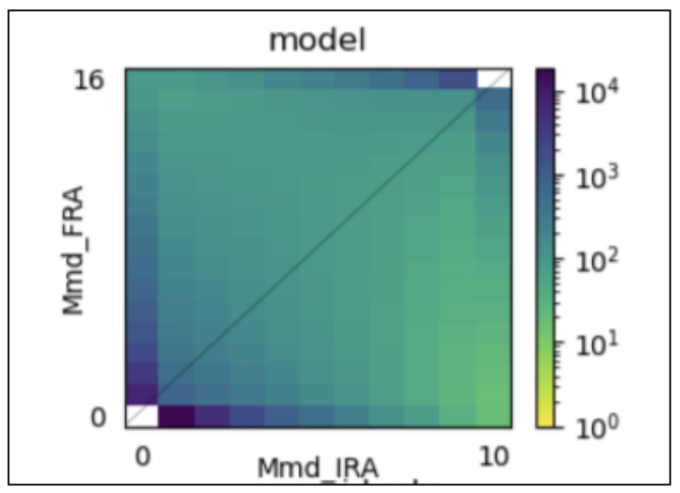
Modified from Harr et al. 2016.

- To process the data, we used the Mus spretus species as an outgroup, allowing us to infer the ancestral state of each allele (3). Sites were marked as synonymous or non-synonymous using Annovar (4).
- We inferred demographic history for synonymous sites using dadi (5). For all three population pairs, the isolation-migration-pre model with inbreeding was the best fit.



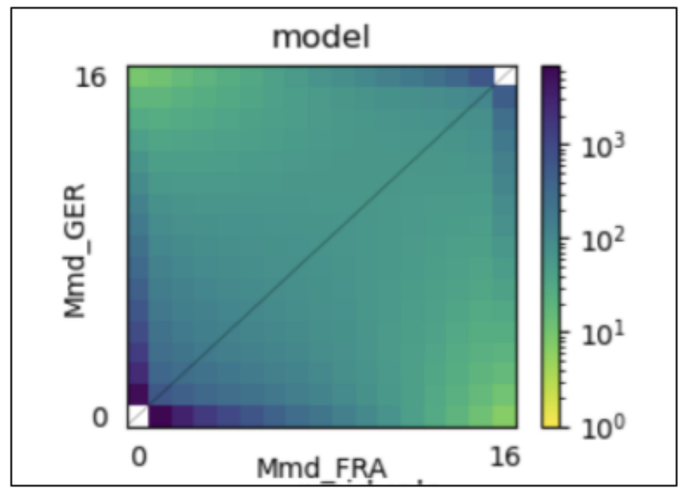
- For each population pair, parameters from the demography inference were used to create a cache of site frequency spectra (SFS) of non-synonymous sites under a range of selection coefficients.
- Cached SFS were used as the input for joint DFE inference, which was completed using dadi-cli (6).

Results



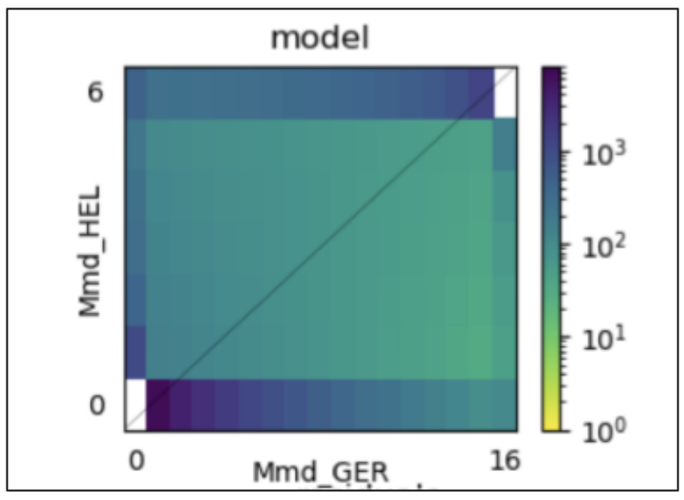
Iran and France

Asymmetric Bivariate Lognormal DFE with a correlation coefficient (rho) of 0.97. The mean and standard deviation of the distribution for each population are not equal.



France and Germany

Lognormal DFE showing a perfect correlation between fitness effects in the two populations (rho = 1). The mean and standard deviation of the distribution for each population are equal.



Germany and Heligoland

Bivariate Lognormal DFE with a correlation coefficient (rho) of 0.98. The mean and standard deviation of the distribution for each population are equal.

Discussion

Here we show that the DFE of each pair of wild mice populations is highly correlated, suggesting similar mutational effects among populations. 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 (7, 8). 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 (9). 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. Huang et al. (2021) Mol Biol Evol 38(10):4588-4602.
2. Harr et al. (2016) Sci Data 3:160075.
3. Agwamba & Nachman (2022) G3 13(2):jkac332.
4. Wang et al. (2010) Nucleic Acids Res
5. Gutenkunst et al. (2009) PLOS Genet:1000695.
6. Huang et al. (2023) bioRxiv:p. 2023.06.15.545182.
7. Phifer-Rixey & Nachman (2015) eLife
8. Ramakers et al. (2018) Nat Ecol Evol 2(7):1093-1103.
9. Yang et al. (2007) Nat Genet 39(9):1100-1107.

Acknowledgements

Thank you to the University of Arizona Undergraduate Biology Research Program and the Arnold and Mabel Beckman Foundation for providing resources and the opportunity to pursue research. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (R01GM127348 and R35GM149235 to RNG).