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 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. Here we present preliminary demographic history inferences and distributions of fitness effects for populations from Iran and France. We find the best demographic models are those that include a distinct split between the two populations and account for inbreeding between closely related individuals. The best DFEs seem to be those that have an asymmetric two dimensional bivariate lognormal distribution.

Demography Inferences

We explored demographic history models using the software dadi. For the populations of *Mus musculus domestics* from Iran and France, there were four models that fit reasonably well: an isolation migration model, an isolation migration pre model, a split asymmetric migration model, and a split delay migration model, all with the inclusion of inbreeding. The four models had similar likelihoods with the isolation migration model being the best. Each model estimated an ancestral population size of about 87000 individuals, which matches previous estimates of the ancestral population size for *Mus musculus domesticus* (1). All four models also estimated low - moderate levels of inbreeding in both populations.
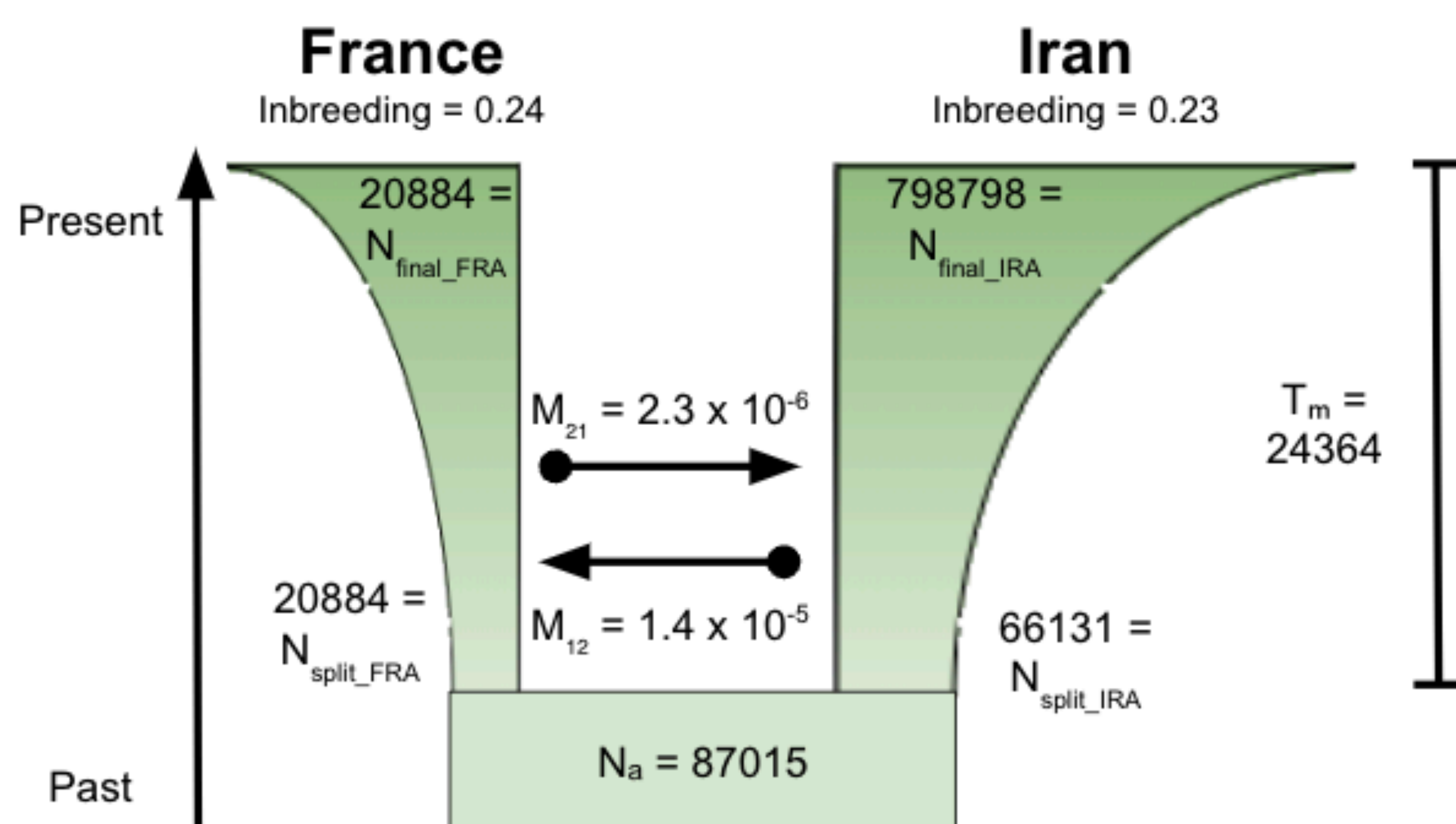
Joint Distributions of Fitness Effects

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 (2).
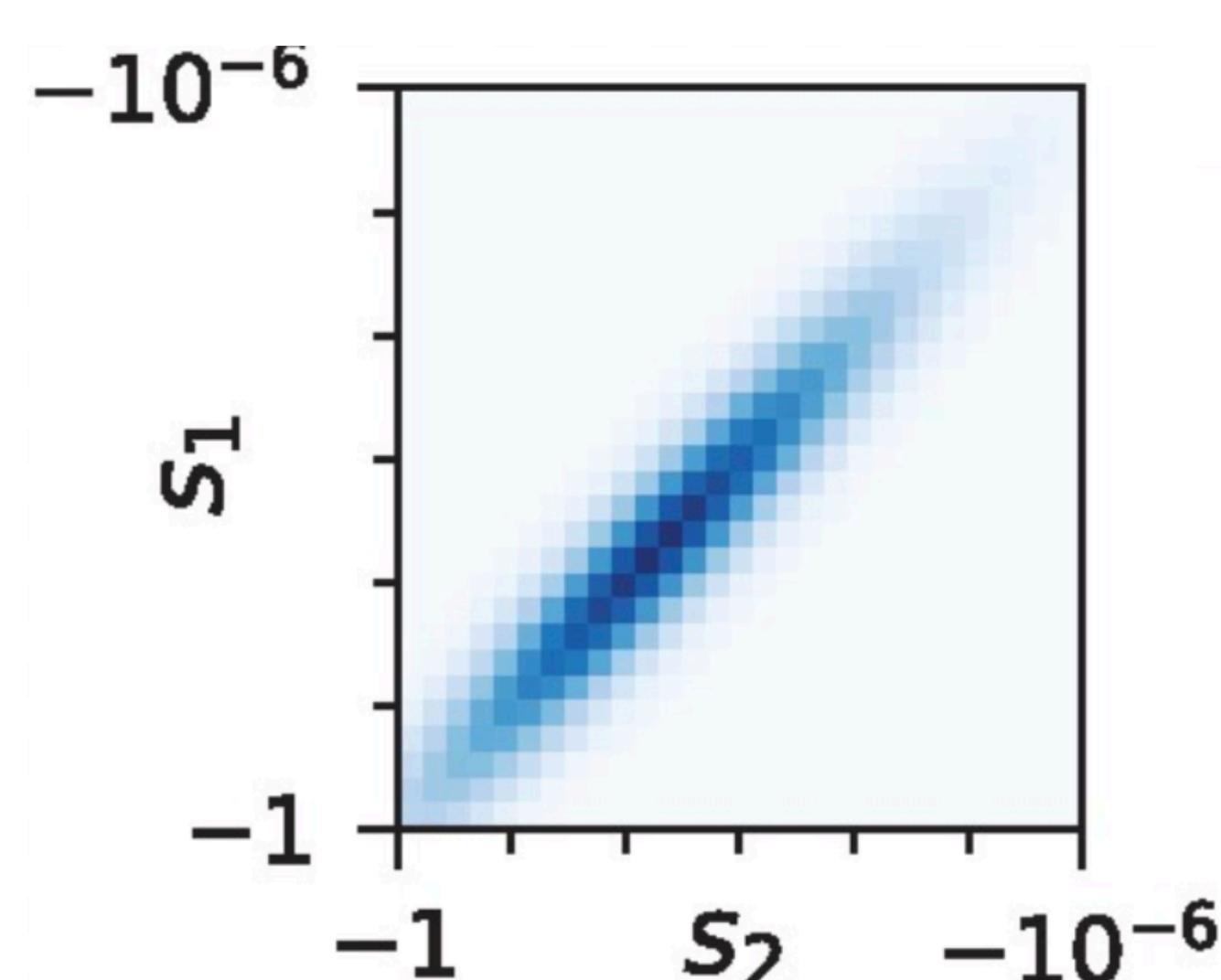

Image from Huang et. al 2021

Distributions of Fitness Effects

We find that the best distribution was a two dimensional bivariate lognormal distribution that inferred asymmetric selection coefficients for mutation in both populations. For the French population, the distribution has a mean of 4.71 with a standard deviation of 6.23. The distribution for the Iranian population has a mean of 4.88 with a standard deviation of 6.97. The correlation between the two distributions is 0.97.
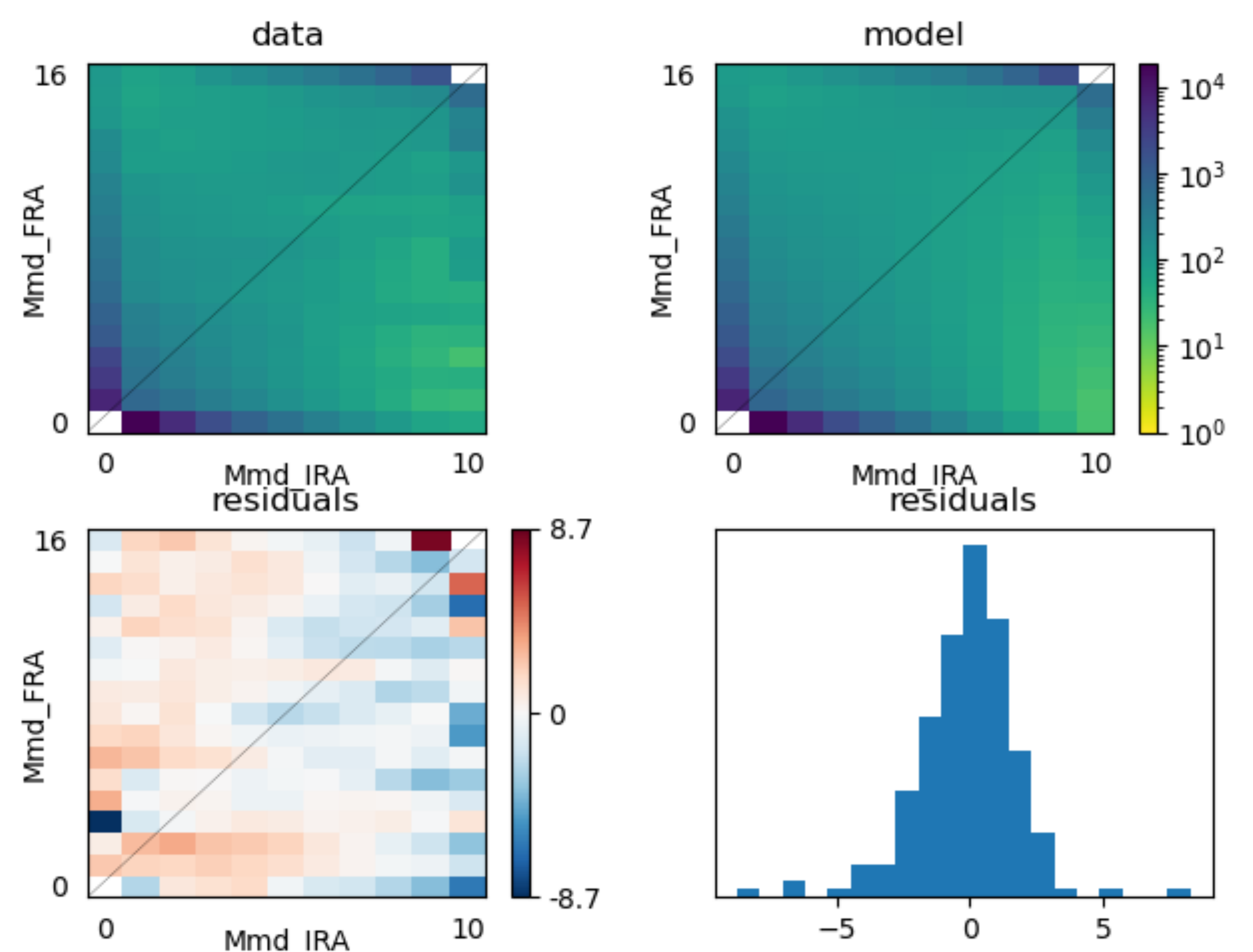
Discussion

Our hope is that this work will help us gain a better understanding of the biological basis of the DFE by tapping into extensive mouse biology knowledge derived from laboratory mice, whose genomic origin is, on average, 92% *Mus musculus domesticus* (3). The DFE demonstrates the similarity of mutation effects in the two populations, suggesting that deleterious mutations in one population are probably also deleterious in the second population. Our next steps are to analyze the sites of non synonymous mutations and investigate the biological function of these genes.

References

1. Paul D Blischak *et al.* Inferring the Demographic History of Inbred Species from Genome-Wide SNP Frequency Data, *Molecular Biology and Evolution*, Volume 37, Issue 7, July 2020, Pages 2124–2136, https://doi.org/10.1093/molbev/msaa042

2. Xin Huang, Alyssa Lyn Fortier, Alec J Coffman, Travis J Struck, Megan N Irby, Jennifer E James, José E León-Burguete, Aaron P Ragsdale, Ryan N Gutenkunst, Inferring Genome-Wide Correlations of Mutation Fitness Effects between Populations, *Molecular Biology and Evolution*, Volume 38, Issue 10, October 2021, Pages 4588–4602, https://doi.org/10.1093/molbev/msab162

3. Yang, H., Bell, T., Churchill, G. *et al.* On the subspecific origin of the laboratory mouse. *Nat Genet* **39**, 1100–1107 (2007). https://doi.org/10.1038/ng2087

Acknowledgements