

Abstract

This simulation was modeled after a physical experiment conducted by Ozan Kiratli. In that experiment, three *Drosophila melanogaster* populations from Maine, Pennsylvania, and Florida were sampled. Preliminary findings suggest that these populations are equally differentiated from each other (Fig. 2). Migration events were simulated by crossing the PA populations with ME and FL in a 3:1 ratio. The populations were subsequently placed under summer and fall selection and their response was measured. Some hybrid populations were more responsive to the summer/fall conditions, adapting faster than the original populations. The next step, and where this experiment begins, is to simulate the empirical experiment in order to determine the extent of the effect that migration and local adaptation have on population structure and allele frequency.

The first stage of the simulation is to create a null model influenced solely by genetic drift, without selection or migration. Three subpopulations would be allowed to diverge until they had reached similar levels of differentiation as those of the wild populations, measured with F_{ST} values. SLiM (Haller and Messer, 2019) was used to simulate three subpopulations of individuals for a range of generations. Other parameters such as recombination rate, mutation rate, and number of loci were chosen to be as close to the *Drosophila* genome as possible while simultaneously scaled down to keep a reasonable runtime.

It is too early to draw conclusions, but at this stage in the simulation we can say for sure that there are factors other than genetic drift affecting the subpopulations. The next step in creating the simulations would be to apply fitness effects and create migration events to gain a better understanding of whether migration, local adaptation, or both are responsible for the differences among the populations.

Introduction

F_{ST} is a measurement of genetic differentiation between populations. It is dependent on genetic drift, migration, mutation, and selection, and is commonly used in evolutionary biology (Holsinger and Weir, 2009).

D. melanogaster populations at different latitudes are locally adapted and thus exhibit different traits. For example, the FL population has trait that are better adapted to warm climates than ME. However, hybrid populations such as PA-FL can be better adapted when faced with the same selection pressures. There is still the question of what forces are causing this pattern.

The goal of simulating the experiment is to create a comparison for the empirical data. Take, for example, a simulation with migration only where the F_{ST} values are lower than the expected values. We can conclude that another force, such as selection, is causing the rest of the differentiation between populations. The simulation has not reached this point – right now, the null model has been completed and additional parameters may be applied.

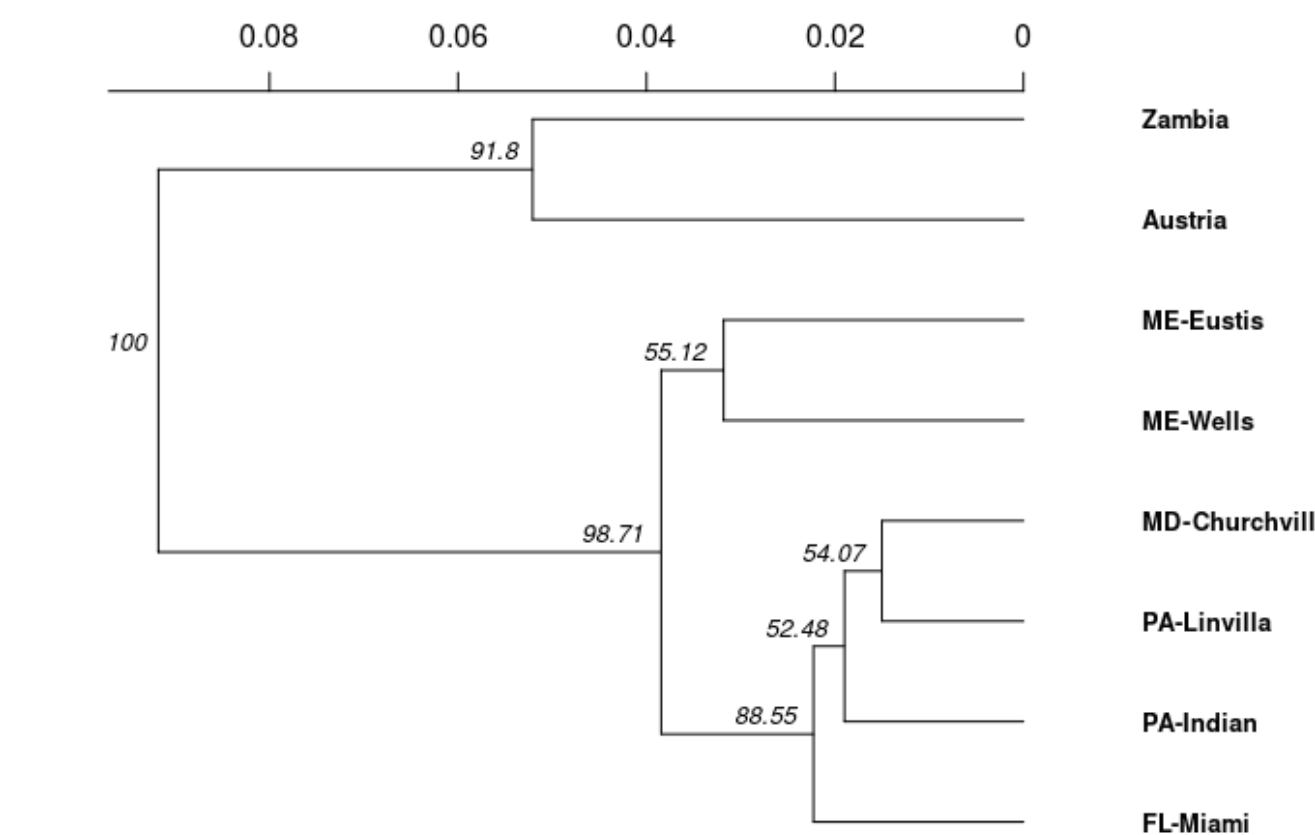
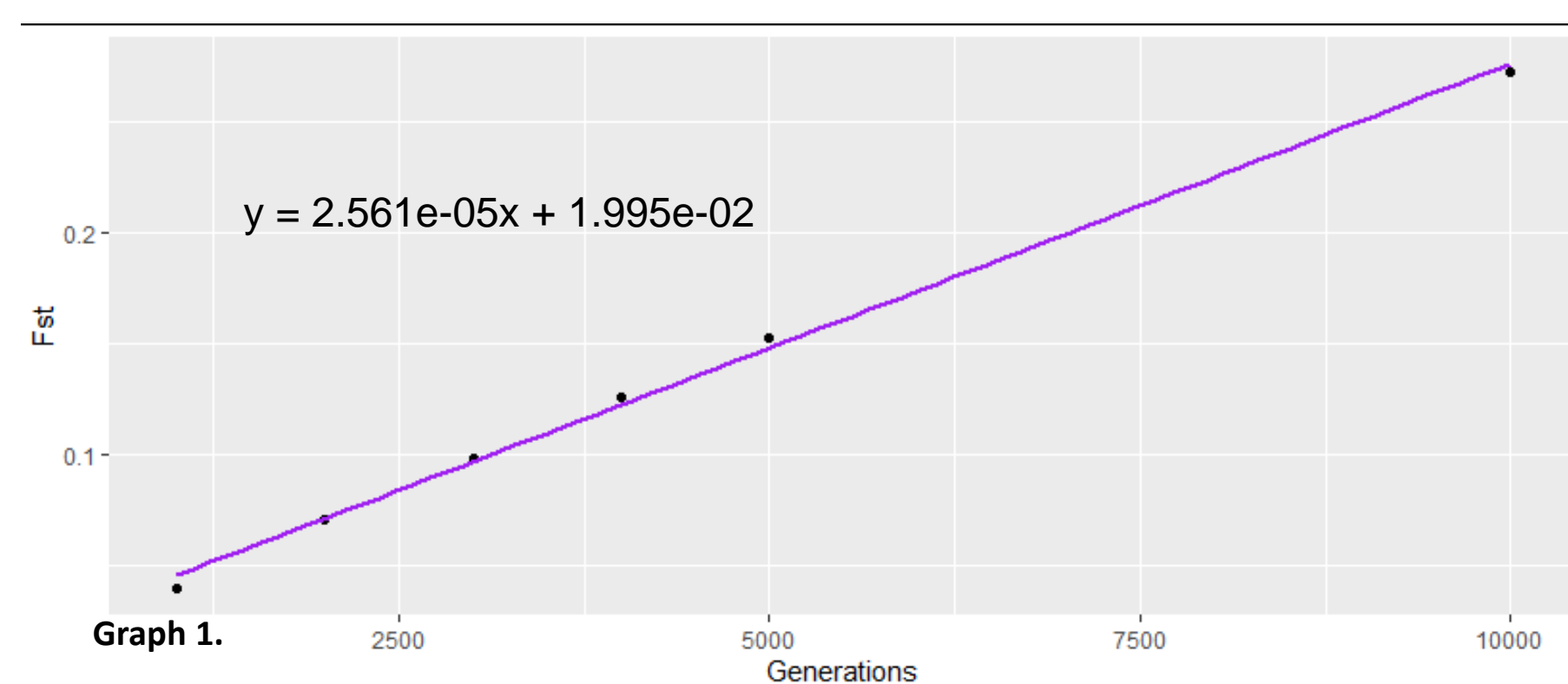


Figure 1. Phylogenetic tree by Ozan Kiratli from the original experiment of several *Drosophila melanogaster* populations.

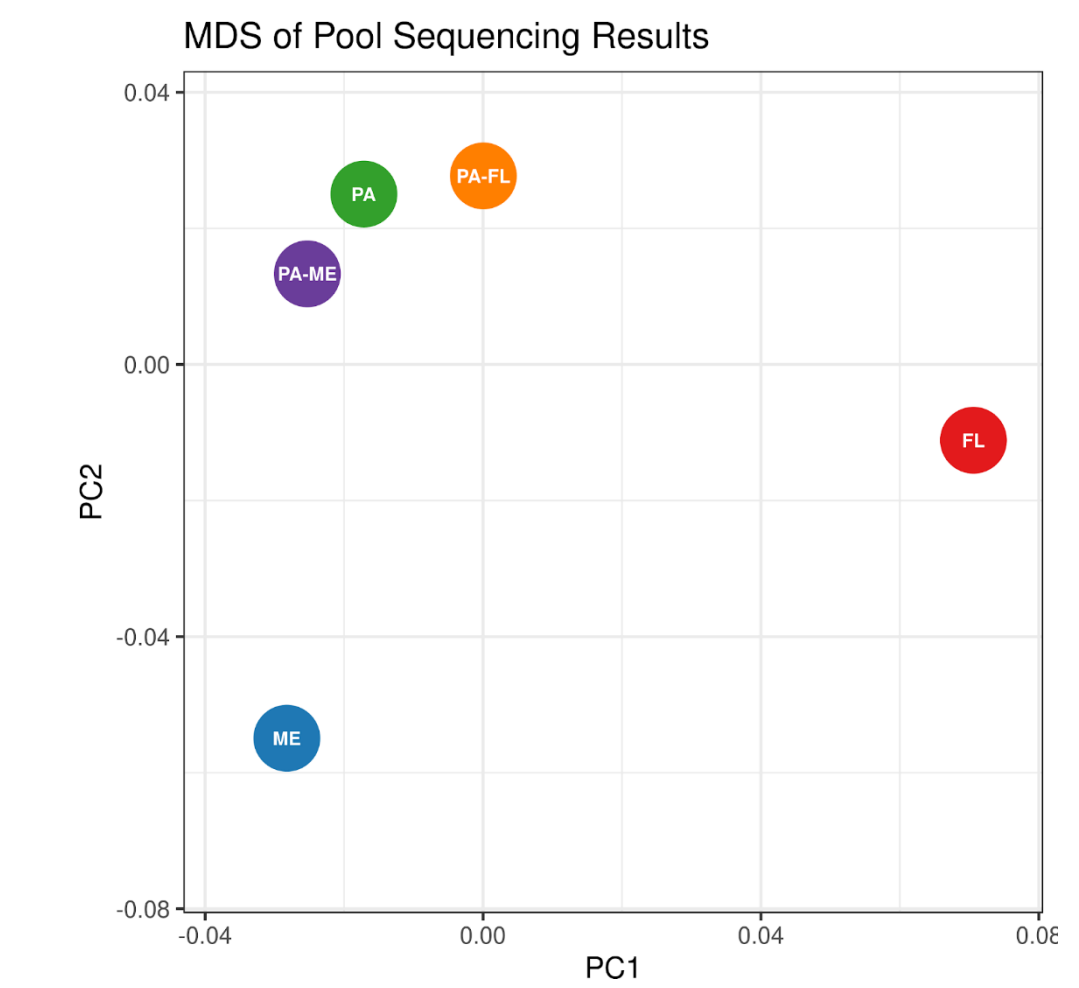


Figure 2. MDS plot by Ozan Kiratli from the original experiment of PA, ME, FL, PA-FL, and PA-ME populations.

Methods and Materials

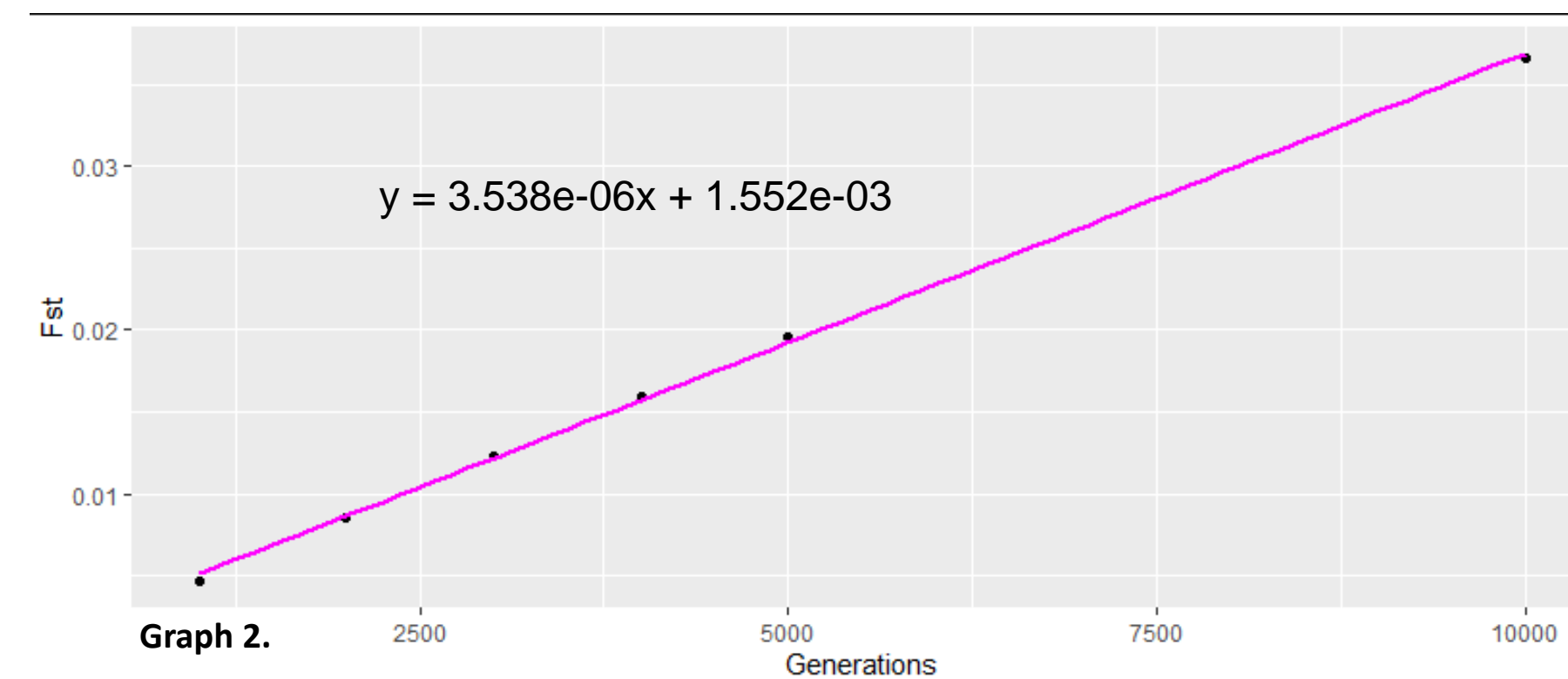
We used a program called SLiM (Haller and Messer, 2019) to set up all the simulations. This program allows the user to manipulate parameters such as subpopulation size, mutation rate, number of generations, and gene number/type, outputting a vcf file that contains a list of loci with segregating alleles. The output was analyzed through R and the vcftools package in bash.

The goal of the first stage of this experiment was to reconcile the time and memory limits of the software with building a realistic model. A realistic model of the *Drosophila* genome of 180 Mb would take weeks to run, several parameters were scaled down. The base of the model was 1M bp, a 1e-6 mutation rate (Keightley et al., 2013), with generations ranging from 1000-10,000 and the size of the three subpopulations being either 1000 or 10,000. The subpopulations all diverged from the same ancestral DNA sequence.

These various combinations were run in the absence of selection and migration, only under the influence of genetic drift. Subsequently, mean pairwise F_{ST} values were calculated with the goal of getting as close as possible to the values recorded in the literature: 0.044 between FL and ME, 0.043 between FL and PA, and 0.027 between PA and ME (Fabian et al., 2012).

Results

Graphs of the average mean pairwise F_{ST} between all three populations versus generation number. Graph 1 is for 1000 – 10000 generations for 1000 individuals per subpopulation, with F_{ST} values in the range of 0.0395 – 0.273. Graph 2 is for 1000 – 10000 generations for 10000 individuals, with F_{ST} values in the range of 0.00463 – 0.0367. Other parameters include a 1e-6 mutation rate, 1 million base pairs, and 10 exons of varying length.



Discussion

According to Fig. 2, there is enough differentiation between the three experimental populations that a migration event would predictably position the hybrid populations on the MDS plot. This can be seen when PA and ME/FL are combined in a 3:1 ratio and the resulting hybrid populations are positioned closer to PA than ME or FL. Thus, genetically, the hybrid populations are differentiated in a predictable way. However, as found in the original experiment, they do not always exhibit intermediate trait values as expected.

There is conflicting evidence in the literature on whether this variation in North American populations is due to spatially varying selection or colonization events of already locally adapted flies from Africa and Europe (Bergland et al., 2015). Preliminary results suggest that North American populations of *D. melanogaster* are more closely related to each other than external populations (Fig. 1). Thus, the high degree of differentiation among North American populations (Fabian et al., 2012) may not be heavily dependent on secondary contact but rather caused by the populations diverging over a significant amount of time.

The most realistic model had 10,000 individuals and 10,000 generations. The wild population would have had millions of individuals and would have been diverging for 1000-5000 generations. Holding all other factors constant, increasing the simulated population size resulted in F_{ST} values that were much lower than those recorded in the literature unless the number of generations was increased. Thus, there are more forces beyond the genetic drift included in the simulation that affect population differentiation.

Conclusions

In summary, to get F_{ST} values that are comparable to empirical values, it is necessary to run the simulation either under a very high number of generations or a very low subpopulation size. The most realistic model has an F_{ST} value of 0.0367, 10,000 generations, and 10,000 individuals per subpopulation. As of now, we can be sure that genetic drift is not the only factor influencing the differentiation between these populations.

Now that the first stage of simulations, run in the absence of selection and migration, is finished, there are three more stages to run: applying summer and fall selection in the absence of migration; migration in the absence of selection; and finally both selection and migration. Selection would be applied both under an omnigenic model and for a few genes of large effect. If the trait in question is omnigenic, it could result in above-average F_{ST} values. Conversely, if the trait is governed by a few genes of large effect, it would have a lower average F_{ST} but a higher per-loci F_{ST} . These additional simulations, once compared to the experimental data, will help determine the effect that gene flow and migration have on *D. melanogaster* population and evolutionary dynamics.

Contact

Daria Zhadanova
Email: dariazh@sas.upenn.edu

References

- Haller, B.C., and Messer, P.W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution* 36(3), 632–637. DOI: <https://doi.org/10.1093/molbev/msy228>
- Fabian, D.K., Kapun, M., Nolte, V., Kofler, R., Schmidt, P.S., Schlotterer, C., Flatt, T. (2012, August 22). Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology*, 21(19), 4748–4769. <https://doi.org/proxy.library.upenn.edu/10.1111/j.1365-294X.2012.05731.x>
- Bergland, A. O., Tobler, R., Gonzalez, J., Schmidt, P., Petrov, D. (2015, November 07). Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. *Molecular Ecology*, 25(5), 1157–1174. <https://doi.org/proxy.library.upenn.edu/10.1111/mec.13455>
- Holsinger, K.E., Weir, B.S. (2009, September). Genetics in geographically structured populations: defining, estimating, and interpreting F_{ST} . *Nature Reviews Genetics*, 10, 639–650. <https://doi.org/proxy.library.upenn.edu/10.1038/nrg2611>
- Keightley, Peter D., Ness, Rob W., Halligan, Daniel L., and Penelope R. Haddrill. "Estimation of the Spontaneous Mutation Rate per Nucleotide Site in a *Drosophila melanogaster* Full-Sib Family." *Genetics* 196, no. 1 (November 2013): 313–320. <https://doi.org/10.1534/genetics.113.158758>

