

“The story of a photo is just out of frame,” said the author John Green about social media, where photos are interpreted to mean everything and anything about the people posting them. The assumption that snapshots are representative of the whole story is not only pervasive in pop culture but also in science.

Background and Intellectual Merit Effective population size (N_e), defined as the number of individuals in a population contributing offspring to the next generation, is a key metric in evolutionary and conservation biology because it is a useful proxy for understanding the adaptive capacity of natural populations. N_e is often based on a genetic snapshot of a population taken at a single point in time and, as such, the conclusions drawn from it may not represent the “whole story” of a population¹. Only recently have studies investigated how adding more genetic snapshots across time can increase our understanding of N_e and thus evolutionary dynamics of populations. One genomic time series of a wild butterfly population identified conflict between the prediction of strong genetic drift from low contemporary N_e and minimal observed allelic change via drift due to high levels of historical gene flow². Most genomic time series have been conducted in short-lived animal systems, tracking multi-generational evolutionary change^{2,3}. Long-lived perennials, such as trees, exist outside of the phenomena described in these systems¹. Trees generally have one reproductive event per year with substantial variability in fecundity across time due to both genetic and environmental factors⁴. This variance could lead to highly variable effective population size on an annual scale. Conversely, theory predicts that locally adapted individuals will exhibit consistently higher reproductive success⁵, reducing N_e values and N_e variability across time⁶. These contrasting dynamics have the potential to drastically change the genetic composition of progeny from year to year, shaping the evolutionary trajectory of populations. Understanding the interaction of local adaptation, variable reproductive output, and allele frequency changes in progeny obtained across multiple time points from a single parental generation will increase our understanding of the evolutionary dynamics of long-lived populations like trees. Therefore, my PhD will ask: **to what extent does temporal variation in parental reproductive output impact progeny’s genetic variation and resulting genetic diversity estimates in a tree species?**

Research Overview I will pair empirical observations using genomic data from maternal and paternal lineages and three consecutive cohorts of progeny of the dioecies species Oregon Ash (*Fraxinus latifolia*) with a simulation-based approach. I will use these two approaches to test how the genetic composition of progeny and frequency of locally adapted alleles within progeny varies on a temporal scale due to variable reproductive output by parental lineages. Parental lineages in two populations, along with individuals from another 8 populations distributed across Oregon will be analyzed using genomic data to identify putative locally adaptive alleles. I will then build a simulation based on Waples *et al.* 2014⁷ to generate multiple cohorts of potential progeny from the sampled parents and calculate estimates of genetic diversity, N_e and the co-ancestry coefficient, from simulated progeny. These diversity metrics have correlated theoretical expectations (described in Aim 3)^{6,8}. I will analyze genomic data from progeny collected over three years and compare the observed diversity metrics to those generated by simulation. I will also evaluate the frequency of locally adaptive alleles in the observed vs simulated progeny. These analyses will elucidate how reproductive output of parental individuals varies on a temporal scale within a generation and how this variability is associated with local adaptation. The genomic analyses described will be computationally intensive, so I will use my extensive experience in BASH and my access to the Pennsylvania State High Powered Computing Cluster, ROAR, to conduct them.

Aim 1: To what degree are *F. latifolia* individuals locally adapted? I will analyze RADseq data from 10 populations of *F. latifolia* across Oregon collected as part of an ongoing study led by the Hamilton Lab at Pennsylvania State University. I will sample 10 maternal and 10 paternal plants for all 10 of these populations. Two focal populations will act as biological replicates from which I will sample an additional 20 maternal plants. I will tag these plants and revisit them annually for progeny sampling over three years, as discussed in Aim 3. Using parental genomic data, I will conduct a multivariate Genotype-Environment Association (GEA) analysis to identify putatively adaptive alleles associated with environmental factors that may influence reproductive output in trees, such as precipitation and temperature averages in the preceding year. I will also calculate metrics of genetic diversity, N_e and co-ancestry, to tune the simulation model, thus ensuring it is representative of realistic population dynamics.

Aim 2: How does the genetic composition of progeny vary temporally if the null hypothesis is true?

I will create a simulation to represent my null hypothesis that there is no effect of local adaptation on the reproductive output of parents. My simulation will run 10,000 times for three reproductive events with parental individuals having a unique, random reproductive output assigned for each reproductive event. The number of offspring produced by each parent during each reproductive event will be highly variable such that it will resemble realistic observed variability in tree populations⁷. Each reproductive event will result in a single cohort of progeny, which will be subsampled to obtain 10–30 individuals per maternal line. From the subsampled progeny, I will calculate N_e and co-ancestry to generate a null distribution of values representing the expectations in the absence of local adaptation of parental reproductive output.

Aim 3: How does the genetic composition of progeny temporally vary in natural populations? I will sample progeny from the two focal parental populations defined in Aim 1. I will collect and analyze genetic material from 10–30 fertilized seeds from each maternal tree every year for three years. From this genetic data I will calculate N_e and co-ancestry as done for the simulated progeny, so the resulting genomic time series will be directly comparable to the distribution of values obtained from the simulation. Under local adaptation, we expect some parental individuals to have consistently higher reproductive output than others. If this is true, the observed N_e obtained from the progeny will be lower⁶, the co-ancestry coefficient will be higher⁸, and there will be lower variability across years compared to the simulated values⁶. Evolutionary theory predicts the parents with higher reproductive output are better adapted to the local environment⁵. To test for this, I will evaluate the frequency of the locally adaptive alleles identified in Aim 1 in both progeny sets. I expect observed progeny to have a higher frequency of locally adaptive alleles than simulated progeny. These analyses will show if, as expected by theory, local adaptation is associated with consistently higher reproductive output over time such that locally adapted parental individuals contribute more to future generations. I plan to collect offspring from the same maternal trees each year, so observed differences in progeny genetic diversity will be due to variable reproductive output of only paternal individuals. This will allow specific inference for paternal genotypes in this system, while providing broader conclusions for reproductive output in trees.

Broader Impacts Oregon Ash is a keystone species that provides essential ecosystem services and is listed as Near Threatened by the IUCN due to imminent invasion of the Emerald Ash Borer (EAB), which has decimated ash throughout the Eastern U.S.⁹. The imminent threat of EAB to Oregon Ash has emphasized the need for genetic conservation for this species to ensure long-term evolutionary potential is maintained. My research will contribute to a genomic monitoring program of *ex situ* collections in Oregon Ash in partnership with the Huntington Botanic Garden, Oregon Department of Forestry, and US Forest Service. This monitoring program will be aided by SNP arrays developed from the candidate neutral and adaptive genetic markers identified during my study which will allow for consistent, efficient, and inexpensive future genotyping of individuals¹⁰. *Ex situ* collections, where seed is preserved outside a species' native habitat, provide an ideal means to conserve genetic resources at risk of extinction. However, as genetic variation can vary across both space and time there is an inherent challenge to ensuring collections meet conservation goals. Temporal sampling of populations is expected to capture adaptive variation more effectively, but it is unknown to what extent this is true¹¹. As such, the sampling and modeling of temporal variation in progeny diversity performed during my project will demonstrate how adaptive potential captured by collections changes temporally. The conclusions generated by my efforts will guide conservation management via *ex situ* collections for *F. latifolia* as well as other long-lived tree species. I plan to disseminate this information through webinars and trainings with conservation biologists, facilitated through collaboration with arboreta and the USFS.

[1] Rellstab *et al.* 2021. *Evol. Appl.* 14(5):1202-1212. [2] Gompert *et al.* 2021. *bioRxiv*. [3] Brüniche-Olsen *et al.* 2016. *PLoS One*. 11(3):e0147875. [4] Clark *et al.* 2004. *Ecol. Monogr.* 74(3):415-442. [5] Darwin C. 1859. *On the Origin of Species by Means of Natural Selection*. London: Murray. [6] Lee *et al.* 2011. *Proc. R. Soc. B.* 278:3303–3312. [7] Waples *et al.* 2014. *Genetics*. 197(2):769-780. [8] Vendrami *et al.* 2021. *Sci. Adv.* 7(37) [9] Semizer-Cuming *et al.* 2019. *Nat. Ecol. Evol.* 3:141-144. [10] Amish *et al.* 2012. *Mol. Ecol. Resour.* 12:653-660. [11] Hoban *et al.* 2020. *Proc. R. Soc. B.* 287:20200102.