The fitness impact of immigration in a population of Florida Scrub-Jays

Introduction: Dispersal between populations can increase genetic variation and improve average fitness (Bell et al. 2019). However, these responses are neither guaranteed (Kyriazis et al. 2020) nor is it known how long any genetic or fitness changes would last (Frankham 2016). The new alleles brought by gene flow can limit (Lenormand 2002) or advance (Garant et al. 2007) adaptive evolution depending on factors including the population levels of genetic variation. Studies addressing these unknowns are rare due to the difficulty of acquiring demographic and genetic data on whole populations and dispersers (Slatkin 1985). Yet, untangling the consequences of gene flow will both reveal the workings of contemporary evolution and foretell the long-term efficacy of human-assisted gene flow.

A population of the federally threatened Florida Scrub-Jays (*Aphelocoma coerulescens*, hereafter FSJ) at Archbold Biological Station has been intensively studied since 1969. This study provides decades of census and life history data (Woolfenden and Fitzpatrick 1984), genetic samples of nearly all individuals in the population over time, and a complete pedigree (Chen et al. 2016). Recent work found that most of the recent fledglings in the study tract are the descendants of immigrants (Chen et al. 2019). The genetic consequences of this immigration could be broad. For example, lower levels of inbreeding in immigrant-resident pairs can reduce the chances that their offspring inherit two copies of a recessive deleterious allele, raising fitness (Charlesworth and Willis 2009) (Figure 1a,b). However, as a lineage remains in a population, recurrent pairings with residents can quickly eliminate this hike in fitness. Fitness can also decrease in immigrant lineages if they carry maladaptive alleles (Lynch 1991) (Figure 1c). These disparate outcomes of gene flow paint an unfinished picture of the ability of dispersal to improve population persistence. Understanding the role of gene flow in a natural population such as the FSJs at Archbold Biological Station will fill-in crucial areas and support future work in population genetics and conservation.

I will test how genetic diversity and fitness change in admixed lineages by comparing heterozygosity and individual fitness in different generations of the descendants of immigrants. I predict that continuous immigration into the study tract will maintain genetic diversity and increase fitness of the descendants of immigrants across generations (as in Figure 1b). Such a result would bode well for the use of genetic rescue as a tool to boost the fitness of threatened populations (Bell et al. 2020). I can also follow-up this work to identify which regions of the genome are responsible for long-term fitness differences in admixed lineages and what role the genetic diversity of source populations plays.

Methods: I have access to genotype data for 3,225 individuals censused from 1990-2013. These individuals include residents, new immigrants, and the descendants of immigrants, with >600 individuals three generations removed from their earliest immigrant ancestor. To expand my analysis further than three generations, I will genotype >80% of juveniles born within the study tract from 2013-2020 (800). I will genotype these individuals using a novel method called haplotagging. Haplotagging is a cutting-edge method that provides a high-throughput and cost-effective way to get haplotype information (Meier et al. 2020). I will use Genotype Imputation Given Inheritance (GIGI) (Cheung et al. 2013), a pedigree-based imputation algorithm, for genotype imputation, which allows me to use a reference panel of samples sequenced at greater coverage to genotype most samples at low coverage. I will use the haplotype information, as well as the population-wide pedigree, in the software BEAGLE (Browning and Browning 2013) to detect regions that juveniles have inherited directly from an immigrant ancestor. I will also use the software PLINK (Purcell et al. 2007) to calculate mean

heterozygosity for each individual. I will complete all sequencing and analysis at the University of Rochester. I expect to begin sequencing during the coming spring (April 2021) and finish three months later. I will use the summer to analyze the data and prepare a publication for submission near the end of 2021. I will complete library preparation within the Chen lab, and will use the Rochester Genomics Center (RGC) for sequencing. I consulted the RGC for the cost estimates presented in my budget below.

I will use generalized linear mixed models to estimate the effects of heterozygosity and the proportion of the genome inherited from an immigrant ancestor on three metrics of fitness: survival to one year, annual fecundity, and lifetime reproductive success. Using multiple metrics of fitness will make my results more robust because each metric makes different trade-offs between tracking long-term fitness and sample size (Brommer et al. 2004). FSJ reproduction varies with territory quality, territory size, group size, and annual climate events (Woolfenden and Fitzpatrick 1984). I will thus include fixed effects of territory size and group size as well as random effects of territory identity, natal year, and breeding years in my models. I will also include fixed effects of sex and parental immigrant status and a random effect of parental identity to account for variation in dispersal behavior between males and females, variation in fitness of long-distance dispersers (Suh et al. 2020), and multiple measurements of individuals.

Using these models, I will quantify the role of immigrant ancestry and genetic diversity on individual fitness. I will compare models created using different generations of the descendants of immigrants to detect changes in the fitness consequences of gene flow over time. Overall, my work will answer how contemporary gene flow can act as a long-term evolutionary force in a natural population.
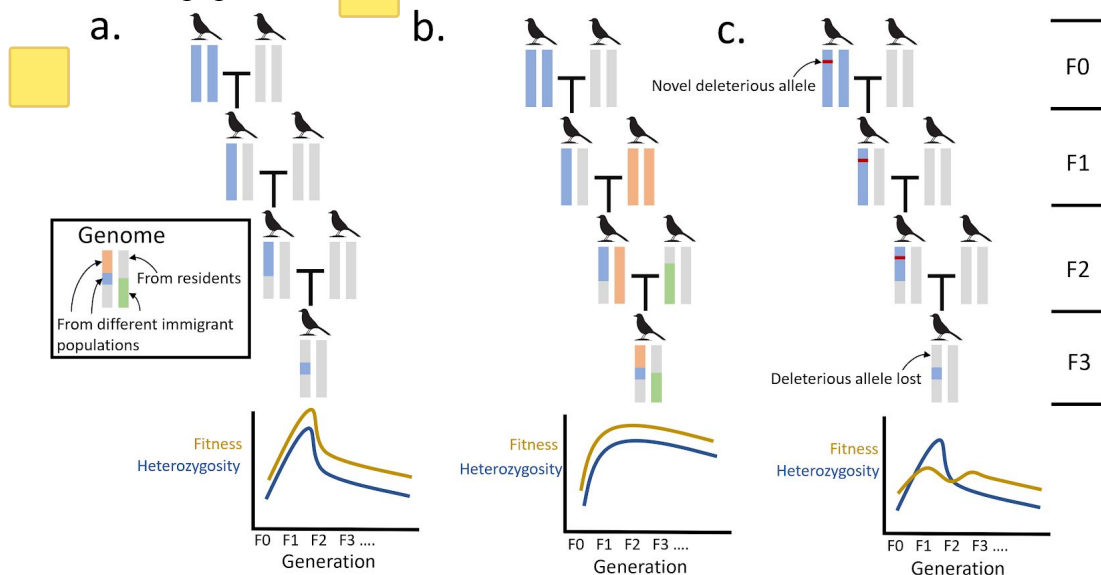

Figure 1. The impact of gene flow on genetic diversity (measured as heterozygosity) and fitness depends on the individuals that the descendants of immigrants mate with. In *a*, recurrent backcrossing to residents causes the heterozygosity of the lineage to quickly return to its previous level. Fitness follows the heterozygosity curve closely because the fitness benefit of outbreeding is assumed to come from masking deleterious recessive alleles (Charlesworth and Willis 2009). In *b*, regular mating with individuals with ancestry from different immigrant populations maintains the initial increase in heterozygosity and fitness. In *c*, a novel deleterious allele (dark red band) from the founding immigrant ancestor decreases fitness in their descendants independent of heterozygosity.

Literature Cited:

Bell, D. A., Z. L. Robinson, W. C. Funk, S. W. Fitzpatrick, F. W. Allendorf, D. A. Tallmon, and A. R. Whiteley. 2019. The exciting potential and remaining uncertainties of genetic rescue. Trends in Ecology and Evolution 34:1070-1079.

Brommer, J. E., L. Gustafsson, H. Pietiäinen, and J. Merilä. 2004. Single-generation estimates of individual fitness as proxies for long-term genetic contribution. The American Naturalist 163:505-517.

Browning, B. L., and S. R. Browning. 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194:459-471.

Charlesworth, D., and J. H. Willis. 2009. The genetics of inbreeding depression. Nature Reviews Genetics 10:783-796.

Chen, N., E. J. Cosgrove, R. Bowman, J. W. Fitzpatrick, and A. G. Clark. 2016. Genomic consequences of population decline in the endangered Florida scrub-jay. Current Biology 26:2974-2979.

Chen, N., I. Juric, E. J. Cosgrove, R. Bowman, J. W. Fitzpatrick, S. J. Schoech , Clark A, Coop G. 2019. Allele frequency dynamics in a pedigreed natural population. Proceedings of the National Academy of Sciences of the United States 116:2158-2164.

Cheung CYK, Thompson EA, Wijsman EM. 2013. GIGI: An approach to effective imputation of dense genotypes on large pedigrees. American Journal of Human Genetics 92:504-516.

Frankham R. 2016. Genetic rescue benefits persist to at least the F3 generation, based on a meta-analysis. Biological Conservation 195:33-36

Garant D, Forde SE, Hendry AP. 2007. The multifarious effects of dispersal and gene flow on contemporary adaptation. Functional Ecology 21:434-443.

Kyriazis CC, Wayne RK, Lohmuller KE. 2020. Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. Evolution Letters.

Lenormand T. 2002. Gene flow and the limits to natural selection. Ecology and Evolution 17:183-189.

Lynch M. 1991. The genetic interpretation of inbreeding depression and outbreeding depression. Evolution 45:622-629.

Meier JI, Salazar PA, Kučka M, Davies RW, Dréau A, et al. Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. bioRxiv.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81:559-575.

Slatkin M. 1985. Gene flow in natural populations. Annual Review of Ecology and Systematics 16:393-430.

Suh YH, Pesendorfer MB, Tringali A, Bowman R, Fitzpatrick JW. 2020. Investigating social and environmental predictors of natal dispersal in a cooperative breeding bird. Behavioral Ecology.

Woolfenden GE, Fitzpatrick JW. 1984. The Florida scrub jay: Demography of a cooperative-breeding bird. Princeton University Press, Princeton, NJ

Budget: Due to the large amount of existing demographic and genomic data, I will only need funding for sequencing and genotyping 800 individuals. With these samples sequenced, I will have sufficient power to detect fitness effects of immigrant ancestry in individuals as many as six generations removed from an admixture event, far greater than in previous studies of outbreeding (Frankham 20126). I require funding for library preparation and sequencing services. The Rochester Genomics Center will sequence a reference panel of 75 samples at 5x coverage, and sequence the remaining 725 samples at 1x coverage. I have divided the total costs of my project by the cohort of juveniles that I will sequence. If I do not receive all the funding I am applying for I can remove the two most recent cohorts from my analysis. Removing these cohorts will reduce the number of generations I can study, but I will still be able to detect fitness effects of immigrant ancestry in individuals as many as four generations removed from an admixture event.

Item	Item Breakdown	Unit Cost	# of Units	Total	Funding Source
Reference Panel	Library Preparation Reagents + Illumina Sequencing (5x)	$1.82 + $30 = $31.82	75	$2386.50	American Ornithological Society Research Awards ($2400) or Rosemary Grant Advanced Award ($2400) (Pending)
2013-2014 Juveniles	Library Preparation Reagents + Illumina Sequencing (1x)	$1.82 + $9 = $10.82	230	$2488.60	Louis Agassiz Fuertes Award ($2,500) (Pending)
2015-2016 Juveniles	Library Preparation Reagents + Illumina Sequencing (1x)	$1.82 + $9 = $10.82	180	$1947.60	**American Society of Naturalists Student Research Award ($2000)** (Pending)
2017-2018 Juveniles	Library Preparation Reagents + Illumina Sequencing (1x)	$1.82 + $9 = $10.82	135	$1460.70	Neotropical Bird Club Conservation Fund ($1500) (Pending)
2019-2020 Juveniles	Library Preparation Reagents + Illumina Sequencing (1x)	$1.82 + $9 = $10.82	180	$1947.60	Frank M. Chapman Memorial Fund ($2000) or Mary J. and William B. Robertson Fellowship Award ($2000) (Pending)