

University of Montana

## ScholarWorks at University of Montana

---

Graduate Student Theses, Dissertations, &  
Professional Papers

Graduate School

---

2022

# GENETIC MONITORING AND RESCUE IN MID-ATLANTIC BROOK TROUT (SALVELINUS FONTINALIS) POULATIONS

Zachary Lorn Robinson

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

**Let us know how access to this document benefits you.**

---

### Recommended Citation

Robinson, Zachary Lorn, "GENETIC MONITORING AND RESCUE IN MID-ATLANTIC BROOK TROUT (SALVELINUS FONTINALIS) POULATIONS" (2022). *Graduate Student Theses, Dissertations, & Professional Papers*. 12032.

<https://scholarworks.umt.edu/etd/12032>

This Dissertation is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact [scholarworks@mso.umt.edu](mailto:scholarworks@mso.umt.edu).

GENETIC MONITORING AND RESCUE IN MID-ATLANTIC BROOK TROUT  
(*SALVELINUS FONTINALIS*) POULATIONS

By

ZACHARY LORN ROBINSON

B.S. Biology, James Madison University, Harrisonburg, Virginia, 2012  
M.S. Environmental Conservation, University of Massachusetts, Amherst, MA, 2015

Dissertation

presented in partial fulfillment of the requirements  
for the degree of

Doctor of Philosophy in  
Fish and Wildlife Biology

The University of Montana  
Missoula, MT

December 2022

Approved by:

Scott Whittenburg,  
Graduate School Dean

Andrew Whiteley, Chair  
Wildlife Biology

Lisa Eby  
Wildlife Biology

Winsor Lowe  
Division of Biological Sciences

Paul Lukacs  
Wildlife Biology

Gordon Luikart  
Flathead Lake Biological Station

Abstract Genetic monitoring and rescue in mid-Atlantic brook trout populations

Chairperson: Dr. Andrew Whiteley

**ABSTRACT**

Brook trout (*Salvelinus fontinalis*) populations have experienced dramatic declines throughout their native range, in part, due to anthropogenic land use and habitat fragmentation. In the mid-Atlantic region, brook trout populations often occupy small, headwater habitat fragments in demographic and genetic isolation, making them vulnerable to inbreeding and genetic drift. My dissertation evaluates different methods for genetic assessment, monitoring, and management of small, isolated brook trout populations. First, I examined the potential value of effective number of breeders ( $N_b$ ) estimates for genetic monitoring by determining whether  $N_b$  estimates were sensitive to habitat characteristics known to affect brook trout populations. Using genetic data from 71 brook trout habitat patches, I found significant evidence that  $N_b$  estimates were positively related to habitat size and base flow index, and negatively related to temperature. These results provide further support for the use of  $N_b$  in genetic assessments and monitoring of isolated salmonid populations. Human-mediated gene flow is a promising approach to reduce extinction risk and alleviate negative fitness effects associated with small effective population size (i.e., genetic rescue). However, there had not been an assessment of the statistical power of commonly used approaches to determine fitness effects of gene flow, despite calls for more widespread use of human-mediated gene flow. I addressed this need by using individual-based simulations of gene flow and found that these monitoring approaches frequently suffered from low statistical power but also identified strategies to improve inference. Finally, I examined the multigenerational effects of genetic rescue in a small, isolated population of brook trout and found consistent evidence of elevated fitness in  $F_1$  hybrids as compared to resident individuals. In contrast, I found a negative relationship between proportion migrant ancestry and lifetime reproductive success in backcrosses ( $F_2$  and later generations). Still, backcrosses with less than 0.48 migrant ancestry had lifetime reproductive success greater than residents, on average. These results highlight that gene flow often introduces beneficial and deleterious variation with the net-effect depending on the efficacy of natural selection, which suggests that ecological conditions affecting demography can play an outsized role in determining the outcome of genetic rescue attempts.

## ACKNOWLEDGEMENTS

I am extremely grateful for the many people and institutions that provided inspiration and opportunities that ultimately led me to pursue a PhD in Wildlife Biology. It has been an immense privilege to spend years thinking about brook trout and the interplay between ecology and evolution that often makes small populations vulnerable to extirpation. First and foremost, I would like to thank Dr. Andrew Whiteley for being a great graduate advisor. His never-ending patience and passion for the conservation genetics of salmonid fishes propelled me forward and were essential to my achievements. I would also like to thank my committee members Dr. Lisa Eby, Winsor Lowe, Paul Lukacs, and Dr. Gordon Luikart for their contributions and for being available when I needed support. I am extremely thankful to Sally Painter and Angela Lodmell at the UM Conservation Genomics Lab for their assistance and attention to detail, which saved me from burning the place down on multiple occasions. Steve Amish and Seth Smith expertly fielded all my bioinformatic questions.

The genetic rescue project would not have been possible without the support of Mark Hudy, Keith Knislow, Ben Letcher, and Jason Coombs. I extend a special thank you to Mark Hudy, my undergraduate advisor, for introducing me to the fisheries field and sharing with me his passion for brook trout. For the  $N_b$  chapter, the Virginia Department of Wildlife Resources, West Virginia Department of Natural Resources, Maryland Department of Natural Resources, Pennsylvania Fish and Boat Commission, and the New York Department of Environmental Conservation graciously provided samples. I'm grateful for the funding I received from The National Science Foundation, The W.A. Franke Graduate Fellowship, The National Fish and Wildlife Foundation, and USDA National Institute of Food And Agriculture McIntire Stennis Grant. I thank the countless field and laboratory staff and volunteers who made this research possible.

I have been privileged to have the support and camaraderie of great family and friends through the years as I worked on this dissertation. My parents Teresa and Spencer Robinson always placed a premium on education and made immense sacrifices to put their sons through college before earning their own college degrees. Grace Barber was a continuous source of love and support—as well as great editor of draft manuscripts. The LD crew: Jeff Strait, Donovan Bell, Seth Smith, Anthony Dangora, and Sam Pannoni.

Regardless of whether we were debating population genetics or “spinning the wheel at Katy’s” you all made graduate school experience a real adventure. Finally, I would like to thank the UM Wildlife Biology Program for providing a stimulating intellectual environment and great community.

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>ii</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>iii</b>
<b>CHAPTER 1: Introduction and Overview</b> .....	<b>1</b>
<b>CHAPTER 2: Estimates of effective number of breeders identify drivers of decline in mid-Atlantic brook trout populations.</b> .....	<b>7</b>
<b>Abstract</b> .....	<b>7</b>
<b>Introduction</b> .....	<b>7</b>
<b>Methods</b> .....	<b>12</b>
Brook trout sampling .....	12
Genotyping .....	13
Genetic Analysis .....	13
Variable Selection.....	14
Statistical Analysis .....	15
Relationships with existing brook trout population assessments .....	17
<b>Results</b> .....	<b>18</b>
Genetic Summary .....	18
Statistical Analysis .....	19
Relationships with existing brook trout population assessments .....	20
<b>Discussion</b> .....	<b>20</b>
Conclusions .....	24
<b>CHAPTER 3: Evaluating the outcomes of genetic rescue attempts</b> .....	<b>30</b>
<b>Abstract</b> .....	<b>30</b>
<b>Introduction</b> .....	<b>31</b>
<b>Methods</b> .....	<b>32</b>
Simulation framework overview .....	32
Genomic parameterization and selection scenarios .....	33
Demographic parameterization .....	34
Metrics for inferring fitness effects of gene flow .....	35
Power analysis.....	35
Habitat-constrained scenario.....	36
Inference based on multiple metrics .....	36
<b>Results</b> .....	<b>37</b>
Effects of immigration on persistence probability.....	37
Trends in genetic metrics following gene flow .....	37
Relative performance of metrics for evaluating genetic-rescue attempts.....	38
Habitat-constrained scenario.....	39
<b>Discussion</b> .....	<b>40</b>
Detecting the fitness effects of gene flow .....	40
Temporal patterns and relative power of metrics.....	41
Statistical power in natural populations .....	42

Assessing genetic rescue with multiple metrics .....	43
Implications for evaluating genetic rescue attempts .....	44
<b>CHAPTER 4: Microhaplotypes increase parentage assignment accuracy for an experimental test of genetic rescue.....</b>	<b>51</b>
<b>Abstract:</b> .....	<b>51</b>
<b>Introduction</b> .....	<b>51</b>
<b>Methods</b> .....	<b>54</b>
Sample collection and RADseq library preparation .....	54
SNP discovery and filtering .....	54
GT-seq primer design, panel optimization, and genotyping .....	56
Testing the GT-seq panel performance .....	57
<b>Results</b> .....	<b>59</b>
Identification of target variants, primer design and optimization. ....	59
GT-seq panel genetic summary.....	60
Parentage Assignment Accuracy.....	61
<b>Discussion</b> .....	<b>62</b>
<b>CHAPTER 5: Evaluating the multigenerational effects of gene flow in an isolated population of brook trout .....</b>	<b>69</b>
<b>Abstract:</b> .....	<b>69</b>
<b>Introduction</b> .....	<b>69</b>
<b>Methods</b> .....	<b>71</b>
Study System and Sampling .....	71
Brook Trout Translocation .....	72
Genetic Analysis .....	73
Genetic mark-recapture and pedigree reconstruction.....	75
Statistical analyses .....	76
<b>Results</b> .....	<b>79</b>
Demographic and Genetic Summary .....	79
Pedigree Reconstruction and Migrant Ancestry.....	79
Effects of Admixture on Individual Fitness and Related Traits .....	80
<b>Discussion</b> .....	<b>82</b>
<b>Conclusions</b> .....	<b>85</b>
<b>REFERENCES</b> .....	<b>94</b>
<b>APPENDIX A. Chapter 2 Supplementary Materials</b> .....	<b>110</b>
<b>APPENDIX B. Chapter 3 Supplementary Materials</b> .....	<b>112</b>
<b>APPENDIX C. Chapter 5 Supplementary Materials</b> .....	<b>127</b>

## **CHAPTER 1: Introduction and Overview**

Global biodiversity is declining at an alarming rate, driven by anthropogenic land-use, over-exploitation, and climate change (Dirzo et al. 2014). The rate of loss of biodiversity in freshwater ecosystems far exceeds that of terrestrial environments (Sala et al. 2000). This is due, in part, to freshwater streams and rivers being particularly vulnerable to habitat fragmentation due to dams and other instream barriers, which often result in absolute barriers to dispersal (Liermann et al. 2012). This characteristic of lotic environments often leads to numerous, isolated populations of freshwater species that are nearly or completely demographically and genetically independent (Brauer and Beheregaray 2020). As a result, freshwater species conservation often involves the challenging task of prioritizing, monitoring, and managing numerous, small populations (Linke et al. 2011).

Stream-dwelling fishes in the Salmonidae family are of high economic and cultural importance (Lynch et al. 2016), and their widespread declines are emblematic of the problems of habitat loss and fragmentation in freshwater ecosystems (Dauwalter et al. 2020). Salmonid fishes often occupy highly fragmented landscapes (e.g., Hudy et al. 2008), and loss of contiguous habitat has been shown to reduce persistence probability of salmonid populations (Dunham et al. 1997, Morita and Yamamoto 2002). Even the loss of connectivity with small tributaries can appreciably affect the persistence probability of an entire metapopulation (Letcher et al. 2007). Habitat size is often positively correlated with genetic variation and effective population size ( $N_e$ ) in salmonid fishes (Peacock and Dochtermann 2012, Fraser et al. 2014), which highlights the vulnerability of habitat-limited salmonid populations to genetic drift and inbreeding.

Small populations are inherently vulnerable to extinction due to demographic and environmental stochasticity (Lande 1993). Over time, small populations can experience declines in individual fitness associated with the accumulation of genetic load due to inbreeding and genetic drift (Keller and Waller 2002, Willi et al. 2006), the effects of which have been shown to increased extinction risk (Saccheri et al. 1998) and reduce population growth rate (Bozzuto et al. 2019). The synergistic effects of demographic stochasticity and the accumulation of genetic load can lead to a positive feedback loop,

termed an extinction vortex (Soulé and Mills 1998). Attempting to identify small populations at risk of extinction due to these factors and to develop methods to increase population persistence probability epitomizes the small-population paradigm in conservation biology (Caughley 1994).

Recently, there have been numerous calls to better incorporate genetic data into management and species status listing criteria at the regional and global level (Laikre 2020, Garner et al. 2020). Genetic monitoring is a powerful and often cost-effective way to gain ecological and evolutionary insights into a population over time and identify populations vulnerable to inbreeding (Schwartz et al. 2007, Luikart et al. 2010). For example, monitoring programs often seek to estimate important evolutionary parameters such as generational  $N_e$  (Wright 1931), which describes expected rates of inbreeding and loss of genetic variation and predicts the efficacy of natural selection. The 50/500 rule is a general guideline in conservation biology and advises that a population size greater than 500 is needed to combat genetic drift and maintain adaptive potential and that populations with  $N_e$  less than 50 are at immediate risk of inbreeding (Jamieson and Allendorf 2012). However,  $N_e$  was originally described for discrete generations and is difficult to estimate in age-structured populations with overlapping generations (Waples et al. 2014). Single-cohort estimates of effective population size, termed effective number of breeders ( $N_b$ ), avoid the assumption of discrete generations, and have a predictable relationship with generational  $N_e$  under equilibrium conditions (Waples et al. 2013)

Human-mediated gene flow into small, inbred populations can alleviate genetic load and increase population persistence probability, termed genetic rescue (Bell et al. 2019). However, salmonid fishes are expected to be locally adapted (Fraser et al. 2011a), and migrants have been shown to have lower fitness in some interconnected metapopulations (Fenster and Galloway 2000; Mobley et al. 2019). This characteristic of salmonid fishes highlights the concern that genetic rescue attempts may instead reduce population persistence probability, through a reduction of individual fitness in hybridized individuals, termed outbreeding depression (Edmands 2007). In addition to disrupting local adaptation, outbreeding depression can be caused by epistatic gene interactions, co-adapted gene complexes, and other genomic incompatibilities (Tallmon et al. 2004). Still, there are few documented cases of outbreeding depression having appreciable effect on

population persistence (Bell et al. 2019), and the benefits of heterosis may offset the risk and consequences of disrupted local adaptation (Vergeer et al. 2004, Fitzpatrick et al. 2020).

Brook trout populations in the mid-Atlantic region of the United States epitomize the threats faced by freshwater vertebrates due to habitat fragmentation and thus provide an opportunity to evaluate both genetic rescue and metrics for genetic monitoring. This species occupies thousands of habitat patches in the mid-Atlantic region that are often small (<2000 ha drainage area) and are expected to have little or no population connectivity to adjacent populations (EBTJV 2016). Brook trout have experienced dramatic declines throughout their native range caused by anthropogenic land-use, competition with non-native salmonids, and climate change (Hudy et al. 2008, Merriam et al. 2019). Additionally, a broad-scale genetic assessment of brook trout populations revealed that many populations had  $N_e$  estimates less than 30 and were vulnerable to inbreeding and genetic drift (Kazyak et al. 2022).

My dissertation evaluates different methods for genetic assessment, monitoring, and management of small, isolated mid-Atlantic brook trout populations. I provide empirical evaluations of  $N_b$  as a tool for population assessment and evaluate the multigenerational effects of gene flow on individual fitness in an inbred population of brook trout. Additionally, I provide a simulation-based assessment of different monitoring strategies for genetic rescue attempts, and develop a powerful, and cost-effective genetic marker panel for brook trout. Please note that throughout my dissertation, I use the first-person plural “we” in recognition of the highly collaborative nature of my research.

In Chapter 2, we evaluated the sensitivity of  $N_b$  estimates to known drivers of population status. Additionally, we introduced relevant population genetic theory and the reproductive ecology of freshwater fishes to put  $N_b$  estimates into an ecological context. We estimated  $N_b$  for 71 brook trout habitat units in mid-Atlantic region and obtained a mean  $N_b$  of 73.2 (range 6.90 – 493). Our modelling approach tested whether  $N_b$  estimates were sensitive to differences in habitat size, presence of non-native salmonids, base flow index, temperature, acid rain deposition, number of road crossings, and percent canopy cover among populations. We found significant support for three of our hypotheses

including the positive influence of available habitat and base flow index, and negative effect of temperature. Our results are consistent with presently observed and predicted future impacts of climate change on populations of this cold-water fish. Importantly, these findings support the use of  $N_b$  in population assessments as an index of relative population status. Finally, we discuss the difficulties in sampling for  $N_b$  in continuously distributed populations, complexity in interpreting the estimates, and directions for future research.

In Chapter 3, we address the lack of guidelines for best monitoring practices for genetic rescue attempts. We used genomically explicit, individual-based simulations to examine the effectiveness of common approaches (i.e., tests for increases in fitness, migrant ancestry, heterozygosity, and abundance) for determining whether genetic rescue or outbreeding depression occurred. Statistical power to detect the effects of gene flow on fitness was high ( $\geq 0.8$ ) when effect sizes were large, a finding consistent with those from previous studies on severely inbred populations. Smaller effects of gene flow on fitness can appreciably affect persistence probability but current evaluation approaches fail to provide results from which reliable inferences can be drawn. The power of the metrics we examined to evaluate genetic rescue attempts depended on the time since gene flow and whether gene flow was beneficial or deleterious. Encouragingly, the use of multiple metrics provided nonredundant information and improved inference reliability, highlighting the importance of intensive monitoring efforts. Further development of best practices for evaluating genetic rescue attempts will be crucial for a responsible transition to increased use of translocations to decrease extinction risk.

In Chapter 4, we provide a description of GT-seq panel design and optimization for use in parentage assignment in an ongoing genetic rescue study on brook trout (*Salvelinus fontinalis*). The optimized GT-seq panel is comprised of 166 amplicons with multiple single nucleotide polymorphisms (SNPs) that can be genotyped collectively as microhaplotypes and 53 amplicons with a single SNP. We used population genetic simulations to evaluate parentage assignment accuracy when genotyping all amplicons as single SNPs or using microhaplotypes when available. We tested the effect of unsampled parents, missing genotypic data, and unknown parental sex on parentage assignment accuracy. Parentage assignment accuracies were high using either genotyping method,

but the panel including microhaplotypes outperformed the SNP panel at high levels (75%) of genotypic missingness. This GT-seq panel had sufficient power for parentage assignment in a genetically depauperate population of brook trout within its native range and may provide a useful tool for future conservation efforts targeting this species.

In Chapter 5, we evaluate the multigenerational effects of a genetic rescue attempt in a small, isolated (~26 generations) population of brook trout intensively sampled from 2010 to 2018. Ten Individuals (5 of each sex) were introduced from an adjacent watershed prior to reproduction in 2011. Following pedigree reconstruction, we estimated lifetime reproductive success (LRS) and found that  $F_1$  hybrids were 2.23 times more likely to successfully contribute progeny and produced 2.26 more offspring on average compared to resident individuals. Additionally,  $F_1$  hybrids had significantly higher juvenile survival relative to residents. However, we found a consistent negative relationship between migrant ancestry and survival and LRS in hybridized individuals of the  $F_2$  and later generations. Still, backcrossed individuals with migrant ancestry less than 0.48 had higher LRS than residents, on average. Gene flow resulted in a 7.60% increase in heterozygosity, and mean population-level migrant ancestry was 0.31, indicating a retention of local ancestry. These results provide an empirical demonstration that a pulse of gene flow can have varied effects on individual fitness that are not revealed until later generations.

The research we present has management implications for mid-Atlantic brook trout populations and the management of small, isolated populations in general. In Chapter 2, we demonstrate that single-sample estimates of  $N_b$  are sensitive to factors known to affect brook trout population dynamics, and thus support the use of  $N_b$  in genetic monitoring and assessment. Our multigenerational rescue study shows that the fitness effects of genetic admixture can be both positive and deleterious and vary temporally. Importantly, the dramatic population growth during the study period likely increased efficacy of natural selection against deleterious genetic variation. Higher abundance also increased the statistical power to detect the negative association of migrant ancestry with fitness in backcrosses, paradoxically, this pattern is more consequential but more difficult to detect when population sizes are smaller, as discussed in Chapter 3. It is worth considering that net effects of a genetic rescue attempt may

depend on temporal variation in ecological drivers of population dynamics as much as the genetic attributes of source and recipient populations.

## **CHAPTER 2: Estimates of effective number of breeders identify drivers of decline in mid-Atlantic brook trout populations.**

### **Abstract**

Brook trout (*Salvelinus fontinalis*) populations have experienced marked declines throughout their native range and are presently threatened due to isolation in small habitat fragments, land use changes, and climate change. The existence of numerous, spatially disparate populations poses substantial challenges for monitoring population status (e.g., abundance, recruitment, or occupancy). Genetic monitoring with estimates of effective number of breeders ( $N_b$ ) provides a potentially powerful metric to complement existing population monitoring, assessment, and prioritization. We estimated  $N_b$  for 71 brook trout habitat units in mid-Atlantic region and obtained a mean  $N_b$  of 73.2 (range 6.90–493). Our modelling approach tested whether  $N_b$  estimates were sensitive to differences in habitat size, presence of non-native salmonids, base flow index, temperature, acid rain deposition, number of road crossings, and percent canopy cover among populations. We found significant support for three of our hypotheses including the positive influence of available habitat and base flow index, and negative effect of temperature. Our results are consistent with presently observed and predicted future impacts of climate change on populations of this cold-water fish. Importantly, these findings support the use of  $N_b$  in population assessments as an index relative population status. Finally, we discuss the difficulties in sampling for  $N_b$  in continuously distributed populations, complexity in interpreting the estimates, and directions for future research.

### **Introduction**

Habitat loss and fragmentation are primary contributors to the elevated extinction rate in freshwater environments (Brauer and Beheregaray 2020). Freshwater streams are vulnerable to complete barriers to migration and dispersal (i.e., dams, culverts, unsuitable habitat sections) due to their dendritic and directional nature. This characteristic of these systems often creates numerous populations that are nearly or completely demographically and genetically independent. Thus, conservation practitioners are often faced with the challenging task of monitoring and managing numerous, rapidly declining,

and spatially disparate populations (Linke et al. 2011, Merriam et al. 2019). Genetic monitoring is increasingly used approach for prioritization and monitoring freshwater systems and can provide valuable insights into rates of inbreeding and loss of genetic variation, connectivity among habitat fragments, and demography (Luikart et al. 2010). Recently, there have been numerous calls to better incorporate genetic data into management and species status listing criteria at the regional and global level (Laikre 2020, Garner et al. 2020).

Generational effective population size ( $N_e$ ) is *the* fundamental evolutionary parameter and describes the genetic properties of a population such as rates of inbreeding and loss of genetic variation, the efficacy of natural selection, and can provide insights into the demographic circumstances of a population. As a result,  $N_e$  is widely regarded as the gold standard of genetic metrics in conservation. However,  $N_e$  is difficult to estimate in natural populations, which, among other challenges, often have long, overlapping generations (Waples et al. 2014). Estimation of generational  $N_e$  for species with overlapping generations requires detailed demographic information (Jorde and Ryman 1995, Waples et al. 2013) or multiple genetic samples spaced apart by multiple generations (Waples and Yokota 2007, Waples et al. 2011). Single-sample estimators of effective population size have clear logistical advantages for conservation practitioners, but also assume discrete generations. When single-sample estimators are applied to mixed-age samples in populations with overlapping generations there is uncertainty about the generational time period to which estimators apply and the magnitude of bias associated with violating assumptions (Waples et al. 2014). This ambiguity can be avoided by estimating the effective size of a single age-class or cohort, referred to as effective number of breeders ( $N_b$ ) (Whiteley et al. 2015a) .

Effective number of breeders is an attractive parameter for genetic monitoring because it can be estimated with single genetic samples using existing, well-tested software (Do et al. 2014) and has a predictable relationship with generational  $N_e$  under equilibrium conditions (Waples et al. 2013). There has been sustained interest in the conservation insights that can be provided by  $N_b$ , mostly motivated by its potential relationship to census size ( $N_c$ ) (Luikart et al. 2010, Whiteley et al. 2015a, Ruzzante et al. 2016, Ferchaud et al. 2016). However, previous research on natural populations of

freshwater fishes has revealed inconsistent relationships between  $N_c$  and  $N_b$ , which is often expressed as temporal variation in the  $N_b/N_c$  ratio (Duong et al. 2013, Yates et al. 2017). Although sampling strategies and statistical power likely contribute to this inconsistent relationship, a more parsimonious explanation is provided by theoretical population genetics and the reproductive ecology of freshwater fishes.

Placing the theoretical description of this parameter in an ecological context aids the interpretation of  $N_b$  estimates and sets expectations of its usefulness within a broader genetic monitoring program. Waples & Waples (2011) demonstrated that  $N_b$  is determined by the number of contributing breeders and the variance of their reproductive contribution:

$$N_b = \frac{2S - 1}{\frac{\sum k_i^2}{2S} - 1} \quad (\text{Waples and Waples 2011})$$

where the vector of  $k_i$  values is the number of progeny of each parent and the number of progeny ( $S$ ) is equal to  $\frac{\sum k_i}{2}$ , given each progeny has two parents. The potential for demographic inference based on  $N_b$  estimates is apparent from the equation above, that is, if variance in reproductive success and the contributing proportion of the adult population are constant,  $N_b$  will increase proportionally with  $N_c$  (i.e., there will be a stable  $N_b/N_c$  ratio). However, there is an abundance of evidence that  $N_b/N_c$  ratios are often unstable in natural populations. Variable  $N_b/N_c$  ratios illustrates that  $N_b$  is not directly dependent on the total number of sexually mature adults in a population, but rather the number of individuals that reproductively contribute and the variance of their contribution to a cohort (Waples and Waples 2011). Therefore, biological or ecological constraints on the number of contributing breeders can be reflected in empirical estimates of  $N_b$ .

Many populations of conservation concern are affected by breeding site quality or quantity (Geist and Dauble 1998; Aitken and Martin 2012; Mottl et al. 2020). For example, many stream-dwelling fish species have point distributions of reproduction (Whiteley et al. 2014a) with a finite number of suitable breeding locations (Beard and

Carline 1991). In such cases, the availability of suitable reproductive habitat effectively limits the potential number of contributing breeders of one or both sexes, which can create a nonlinear relationship between  $N_b$  and  $N_c$ . For example,  $N_b$  can be estimated for each sex and the  $N_b$  of the population can be calculated as follows:

$$\frac{1}{N_b} = \frac{1}{4N_{bm}} + \frac{1}{4N_{bf}}$$

(Wright 1938)

If we assume  $N_b$  of females ( $N_{bf}$ ) is constrained by available breeding sites,  $N_{bf}$  is then equal to the number of available breeding sites (i.e., females are ideal individuals and there is one unique female per breeding site). Ecological constraints such as this place a limit on the maximum population value of  $N_b$ :

$$\lim_{N_{bm} \rightarrow \infty} \frac{1}{4N_{bm}} + \frac{1}{4N_{bf}} = \frac{1}{4N_{bf}}$$

(Modified from Waples and Antao (2014))

In this hypothetical case of extreme breeding site limitation,  $N_b$  of the population will not exceed  $4N_{bf}$ , which would correspond to four times the number of breeding sites, even as the effective number of breeders of males ( $N_{bm}$ ) becomes arbitrarily large. This example, albeit simplistic, illustrates that  $N_b$  can become decoupled from census size for certain taxa and ecological contexts. Therefore, it is important to recognize when  $N_b$  is employed as a genetic monitoring metric, it is simultaneously providing genetic, demographic, and ecological insights into a population, which may strongly correlate with adult abundance, reproductive habitat quantity or quality, or early-life mortality depending on the ecological conditions experienced by the progenitors of a cohort or the progeny themselves (Whiteley et al. 2015a, Bacles et al. 2018). We therefore predict that, when used as a genetic monitoring metric,  $N_b$  will often be related to other metrics used to infer relative population status, such as genetic variation, but also offer additional insights into cohort-specific processes.

The demographic and ecological interpretations of  $N_b$ , illustrated by the

mathematical descriptions above, help interpret previous empirical work on brook trout *Salvelinus fontinalis* populations. Well-studied brook trout populations have demonstrated relationships between  $N_b$  and stream flow during reproduction (Whiteley et al. 2015a) and a relationship between  $N_b$  and adult census size (Ruzzante et al. 2016, Yates et al. 2017). Additionally, brook trout populations in the mid-Atlantic region of the United States are an example of a widely distributed, declining freshwater fish species that exemplify the problems of habitat loss, and fragmentation, and vulnerability to climate change (Hudy et al. 2008, Merriam et al. 2019). This species occupies thousands of habitat patches in the mid-Atlantic region that are often small (< 2000 ha drainage area) and are expected to have little or no population connectivity to adjacent populations (EBTJV 2016). Importantly for estimating  $N_b$ , these relatively small and isolated populations minimize bias that is associated with violated assumptions such as continuous population structure or migration (Waples and England 2011, Neel et al. 2013). Brook trout populations, along with other salmonids, are also strongly affected by access to quality spawning and early rearing habitat (Beard and Carline 1991, Petty et al. 2005, Kanno et al. 2016b), which highlights how both abundance of mature adults and ecological factors (e.g., reproductive habitat) can influence  $N_b$  estimates.

In this paper, our goal is to evaluate the use of  $N_b$  as a genetic monitoring metric by determining if it is influenced by factors known to affect brook trout status, abundance, and occupancy. Our modelling choices and interpretation of our results reflects the theoretical lens we introduce above. Although, one advantage of  $N_b$  as a monitoring metric is that it can reflect annual processes such as reproduction and early juvenile survival, we instead employ it as an index of relative population status, which we predict to be influenced by major drivers of population status at the basin scale. We build models of  $N_b$  that aim to identify relationships with physical, ecological, and chemical aspects of the habitat patch that are known to affect brook trout population status, abundance, or occupancy. These factors include habitat size (Whiteley et al. 2013), temperature (Trumbo et al. 2014), acidification (Cleveland et al. 1986), nonnative species presence (Hitt et al. 2017), stream flow (Kanno et al. 2016b), and anthropogenic influence (Hudy et al. 2008, Merriam et al. 2019). Finally, we describe and discuss the primary drivers of  $N_b$  in brook trout populations within the mid-Atlantic region, describe

the relationship between genetic variation and  $N_b$ , and discuss the monitoring value of  $N_b$  for this species.

## **Methods**

### *Brook trout sampling*

A consortium of state and federal agencies collected 8,121 tissue samples from 71 brook trout habitat patches in the mid-Atlantic region of the United States from 2009 to 2018 (Figure 2-1). A habitat patch is defined as an area of contiguous catchments (seventh level, 14-digit hydrologic unit codes; USGS 2012) within which fluvial habitats are continuously occupied by brook trout (EBTJV 2016). Of the sampled habitat patches, 59 occur in the Chesapeake Bay basin, and 12 occur in eastern Ohio River basin. Biologists and managers within each jurisdiction played a major role in selecting habitat patches of interest for genetic monitoring. We chose habitat patches as the unit of sampling because, in this region, they often represent a discrete, biological population for which genetic parameters can be accurately estimated (Whiteley et al. 2014b), and the scale of management objectives or targets. Importantly, single-cohort samples of brook trout can be obtained because age-0 brook trout can be easily distinguished based upon length during their first summer (Hudy 2000). Upon capture, total length was measured and a small ( $< 1 \text{ cm}^2$ ) caudal fin clip obtained for genetic analysis and the fish was immediately returned to the approximate point of capture. Brook trout were captured using one-pass electrofishing using the sampling protocol provided by Whiteley et al. (2012), which has been shown to produce unbiased estimates and avoid family overrepresentation. Briefly, juvenile (age-0) brook trout were sampled at three equidistant locations within the habitat patch with a target of 25 individuals from each location. In certain cases, sample sizes less than recommended ( $n < 75$ ) were included for analysis (minimum  $n = 17$ ), because it is not unusual for habitat patches in this region to have age-0 abundance much less than 50 (e.g., Robinson et al. 2017). Accurate estimates of  $N_b$  can be obtained with samples less than 75, particularly when true  $N_b$  is less than the sample size (Waples 2006, Ackerman et al. 2017), which is likely in these populations.

### *Genotyping*

DNA was extracted from all age-0 brook trout tissue samples and genotyped using two genotyping methods. Samples collected prior to 2015 (5,397 DNA samples from 45 habitat patches) were genotyped at 8-microsatellite loci (*SfoC113*, *SfoD75*, *SfoC88*, *SfoD100*, *SfoC115*, *SfoC129*, *SfoC24*; King et al., 2012) and *SsaD237* (King et al. 2005). These microsatellite markers have been extensively tested and have not exhibited systematic deviations from HW proportions or LD in mixed-aged samples from hundreds of brook trout populations (e.g., Annett et al., 2012; Kanno et al., 2011; D. C. Kazyak et al., 2022; Robinson et al., 2017). Samples collected from 2016 on (2754 DNA samples from 43 habitat patches) were genotyped using genotyping-in-the-thousands by sequencing (GT-seq; Campbell et al., 2015). The GT-seq marker panel was development by Idaho Department of Fish and Game and included 240 amplicons, each possessing a single nucleotide polymorphism (SNP) targeted for genotyping (M. Campbell, *personal communication*). Sequencing libraries were prepared using the protocol described in Campbell et al. (2015) and sequenced on an Illumina NextSeq instrument. We tested the 240 GT-seq loci for deviations from HW proportions and for LD in program GENEPOP version 4.7.5 (Rousset 2008), after which 167 polymorphic SNP-loci were retained (S2-1).

### *Genetic Analysis*

For each habitat patch sampled, we estimated genetic variation and  $N_b$ . To describe genetic variation, we calculated observed and expected heterozygosity and  $F_{IS}$  using the statistical computing software *R* v4.0.3 (R Core Team 2020) and the *R* package ‘hierfstat’ (Goudet 2014). Effective number of breeders was estimated for each habitat patch using the LD-method in NeEstimator v2.1 (Do et al. 2014). We estimated  $N_b$  assuming a monogamous mating system, based-up on the observation of that 80% of reproductive individuals contribute to a single family in two brook trout populations (Coombs 2010). Importantly for this investigation, the mating system assumption will influence the absolute value of  $N_b$ , but not the relative value among habitat patches (Waples 2006).  $N_b$  estimates were derived using a minimum allele frequency cutoff ( $P_{crit}$ ) of 0.02, which has been shown to provide an adequate balance between precision and bias across sample

sizes (Waples and Do 2008). We report uncertainty in our  $N_b$  estimates with 95% confidence-intervals produced using the jackknife method across individuals within each sample.

Waples et al. (2014) demonstrated bias in genetic estimates of  $N_b$  in population with overlapping generations and age structure such as brook trout. Additionally, they provide a bias correction that can be applied to empirical estimates. This bias correction requires the true  $N_b / N_e$  ratio, which can be calculated with a life table in the program AgeNe (Waples et al. 2011). We chose not to apply this bias correction for two reasons. First, although a life table is available from an intensively studied brook trout population in Massachusetts, USA (Letcher et al. 2014), without estimated vital rates for each habitat patch the correction will have no impact on relative  $N_b$  estimates among patches. Secondly, the estimated bias correction for the same intensively studied Massachusetts population referenced above was minimal at 3.4% (Whiteley et al. 2015a), and is likely insignificant compared to the potential bias generated by sampling (Whiteley et al. 2012).  $N_b$  estimates based upon the two different marker types are not expected to be significantly biased relative to one another and can be directly compared (Waples and Do 2010), however, we do expect higher precision in estimates generated by the GT-seq marker panel due to the higher number of loci (Luikart et al. 2021). An empirical comparison of the marker types used in this dataset supports this expectation (figure S2-1).

### *Variable Selection*

We hypothesized that  $N_b$  will be useful in inferring the population status within brook trout habitat patches based upon previous research in this species. We test this hypothesis, albeit indirectly, by determining if  $N_b$  is related to the factors that are known to affect occupancy, abundance, and status of brook trout populations. We include variables that are intended to represent habitat size, stream flow, temperature, competition with nonnative salmonids, acid rain deposition, and anthropogenic disturbance (Table 2-1). For habitat size, we calculate habitat patch size as the product of NHDplus V2 stream length (km) and habitat patch area (ha) (EBTJV 2016). Base flow index (BFI) was obtained from the U.S. Geological Survey (USGS 2003), and the mean

value within a habitat patch was used. Within a habitat patch, BFI is intended to reflect relative groundwater contribution to stream flow, which has an empirically observed, positive effect on brook trout reproduction and stability of stream flow and temperature (Curry and Noakes 1995, Nuhfer et al. 2017). Brook trout populations have been negatively affected by rising stream temperatures and are expected to decline further with climate change (Trumbo et al. 2014, Bassar et al. 2016). Within the mid-Atlantic region, we predict the range of stream temperatures experienced would produce a negative, approximately linear relationship between  $N_b$  and stream temperature. We used mean maximum annual air temperature from 1991–2020 from the PRISM climate group as an index of stream temperature. The presence or absence of nonnative salmonids was obtained from the Eastern Brook Trout Joint Venture habitat patch layer (EBTJV 2016) and field observations during sampling. This metric was used to account for negative effects of competition with non-native salmonids within the habitat patch (Hitt et al. 2017). Acid rain and low stream pH has had a significant negative effect on brook trout populations (Hudy et al. 2000, Nislow and Lowe 2003). Hence, we used the within-patch average sum of total nitrogen and sulfur deposition, hereafter deposition, using data from 2000–2002 and 2010–2020 as an index of acidification (NADP 2022). We also include the number of road-stream crossings per hectare and percent canopy cover from the 2011 National Land Cover Database as a proxy for human disturbance and land use, which is negatively associated with extirpation of brook trout populations (Merriam et al. 2019). All spatial datasets were summarized using the geographic information software QGIS (QGIS Development Team 2020). All variables except for the presence or absence of non-native salmonids were  $z$ -score standardized prior to modeling.

### *Statistical Analysis*

$\hat{N}_b$  was modelled with a Bayesian mixed effect, generalized linear model with a gaussian error distribution and log link function using weighted observations in the statistical program JAGS version 4.3.0 (Plummer 2003). We constructed and fit the model in the statistical computing program *R* and called JAGS using the package ‘R2jags’ (Su and Masanao 2015). For inclusion in the model, we required that each sample generate a positive, finite point estimate of  $N_b$ . For subsequent modelling, we chose to only retain

positive and finite point estimates of  $N_b$ . Additionally, 26 of the 71 habitat patches had multiple cohorts sampled (i.e., years) and we used a single estimate for modelling purposes generated by the harmonic mean of cohort-specific  $N_b$  estimates. To accommodate differing levels of uncertainty, our  $N_b$  estimates were given relative weights in the model based on the inverse of the mean standardized, 95% jack-knifed confidence interval width corresponding to each  $N_b$  estimate. If the upper bound confidence interval included infinity, the observation was given a weight based on two times the largest finite confidence interval within the dataset. In cases where we used the harmonic mean of multiple  $N_b$  estimates for a habitat patch we used the arithmetic mean of the confidence intervals for weighting. These relative weights were intended to account for differences in precision generated by field sampling and the two different marker panels used. The model was fit with a random intercept term based on the US state that contained the habitat patch. We selected these political boundaries for the random intercept because sampling and site selection was most often conducted by different entities in different states. Additionally, state boundaries roughly correspond to geographical clusters of environmentally similar brook trout populations (Zhang et al. 2008). Covariates with missing values were included and values were drawn from a fitted normal distribution corresponding to each covariate.

The relative importance of each variable in the model was assessed using Bayesian indicator variable selection *sensu* (Kuo and Mallick 1998), so that effects were estimated by sampling from the conditional spike and slab posterior  $\theta_j = I_j \beta_j$  (O'Hara and Sillanpää 2009). The probability of switching from  $I_j = 0$  to  $I_j = 1$  was estimated and was given an uninformative beta distributed prior ( $\alpha=1, \beta=1$ ). Slope coefficients ( $\beta_j$ ) of variables and the random intercept were estimated with uninformative, normally distributed priors centered on zero. We report the posterior probability of inclusion for each variable  $\beta_j$  as the mean value of indicator  $I_j$ . Where applicable, we express statistical significance of effects based on whether 90% credible intervals include a zero-effect size, which is analogous to a one-tailed test at  $\alpha = 0.05$ . We fit the model with 5 chains of 10,000 adaptive phase iterations and 100,000 estimation iterations with a thin rate of 10. Model convergence was assessed by visually inspecting chains with the *R* package 'mcmcplots' (Curtis 2015) and the potential scale reduction factor (diagnostic

values  $< 1.1$  indicate good chain mixing; (Gelman and Rubin 1992). Goodness of fit of the model was assessed by Bayesian  $P$ -value and visual evaluation of residuals and predicted values.

#### *Relationships with existing brook trout population assessments*

Concordance between independent brook trout population assessments and  $N_b$  estimates would provide additional support that  $N_b$  performs well as a genetic monitoring metric for mid-Atlantic brook trout populations. For this exploratory analysis, we used a composite habitat integrity score and future security score generated by the organization Trout Unlimited (TU) (Fesenmyer et al. 2017). Habitat Integrity is a percentile scaled composite score that represents aforementioned factors known to influence brook trout populations including anthropogenic land use (e.g., Riparian forest cover, percent agriculture, stream-road crossings) and acid rain deposition. The future security score used by TU is a percentile scaled estimate of stream temperature. Both future security and habitat integrity scores were estimated using TU habitat units and the average value was taken when multiple units were within an EBTJV habitat patch. We also use predicted probability of occupancy from U.S. Geological Survey's Spatial Hydro-Ecological Data Systems (SHEDS) as a predictor of  $\hat{N}_b$  (Walker et al. 2021). Predicted occupancy was estimated at the catchment scale and the average value was taken from catchments within an EBTJV habitat patch.

We constructed three models using habitat integrity, future security, and probability of occupancy each as a single predictor variable of  $\hat{N}_b$ . All three variable have values that range from zero to one and were not transformed prior to modelling. We used a Bayesian mixed effect, generalized linear model with a gaussian error distribution and log link function using weighted observations in the statistical program JAGS. Consistent with the models above we used US state as a random intercept and used the width of jack-knifed confidence intervals of  $\hat{N}_b$  to weight observations. We fit the model with 5 chains of 1000 adaptive phase iterations and 15,000 estimation iterations with a thin rate of 1. We evaluated model convergence using potential scale reduction factor and goodness of fit using Bayesian  $P$ -value.

## Results

### *Genetic Summary*

All 8,121 brook trout tissue samples from 71 habitat patches were extracted and genotyped with 75% genotyping success. Mean sample size was 76.9 (range 17.0 – 510). As expected, estimates of mean population heterozygosity were consistently lower when estimated with the biallelic, SNP markers compared to the microsatellite marker panel. Forty-five of the habitat patches had at least one microsatellite-based estimate of expected heterozygosity with an average of 0.676 (range 0.376–0.796). Forty-three of the habitat patches had at least one estimate of bi-allelic, SNP-based mean expected heterozygosity with an average of 0.196 (range 0.064–0.311). Estimates of heterozygosity for both marker types were available for 17 habitat patches and exhibited a Pearson's  $r$  of 0.317 ( $t=1.29$ ,  $df=15$ ,  $p=0.215$ ). This correlation between SNP- and microsatellite-based heterozygosity estimates is lower than reported elsewhere (e.g., Pearson's  $r=0.45$ ; (Lemopoulos et al. 2019)), however, our estimates are among cohorts and therefore do not correspond to the same collections. Mean population  $F_{IS}$  was 0.009 (range -0.128–0.134) for microsatellites-based estimates and 0.0260 (range -0.041–0.081) for SNP-based estimates.

We obtained a positive and finite point estimate of  $N_b$  for all 71 habitat patches. We removed one estimate that produced a negative point estimate and retained four estimates that had infinite upper confidence intervals. Mean  $\hat{N}_b$  was 73.2 (range 6.9–493.1) including habitat patches for which the harmonic mean of repeated samples was used (Figure 2-1). In general, repeated estimates of  $\hat{N}_b$  were similar among cohorts. Excluding one habitat patch, the range of  $\hat{N}_b$  estimates for a single patch with multiple sampled cohorts was 45.4 on average (range 1.0–210.8). However, a habitat patch in Maryland, had an estimate of  $\hat{N}_b = 1811.4$  (95% CI 308– $\infty$ ) based on a sample size of 43, compared to two years prior when this habitat patch had an  $\hat{N}_b = 285.2$  (95% CI 134–2715) with a sample size of 75. After taking the harmonic mean of  $\hat{N}_b$  for modelling purposes, we retained an  $\hat{N}_b = 493.1$  for this habitat patch, which lies within the confidence intervals of both estimates and within a biologically plausible range.

### *Statistical Analysis*

Collinearity among independent variables was minimal and the results of the generalized linear model indicate that model performance was adequate to address our hypotheses. The mean, absolute value of all pairwise correlations (Pearson's  $|r|$ ) among the independent variables was 0.180 (range 0.026–0.490). The maximum correlation among variables ( $r = 0.490$ ) occurred between deposition and BFI. The distribution of variables in our sample was, in general, representative of the broader mid-Atlantic region, although we sampled warmer streams with higher percent canopy cover on average (Figure 2-1). In terms of model performance, there was no clear pattern between model residuals and predicted values, and the model residuals were approximately normally distributed. Adequate model performance was further supported by our Bayesian  $p$ -value of 0.582. We observed excellent model convergence with a maximum potential scale reduction factor of 1.03 considering all estimated parameters.

Our generalized linear mixed model supported the hypotheses that quantity of habitat, temperature extremes, and the relative contribution of groundwater to surface flow influenced  $\hat{N}_b$  within a habitat patch. As predicted, we found that habitat patch size had a significant, positive relationship with  $\hat{N}_b$  with the posterior probability of an effect size of zero or less of 0.037 (Figure 2-2). Base flow index also had a significant positive relationship with  $\hat{N}_b$  with a posterior probability of an effect size of zero or less of 0.031. Mean maximum annual temperature had a significant, negative relationship with  $\hat{N}_b$  with a posterior probability of an effect size of zero or more of 0.002. We found little support for effects of deposition, canopy cover, and non-native salmonid presence on  $\hat{N}_b$ . Contrary to our predictions, road crossings per hectare had a significantly positive relationship with  $\hat{N}_b$  (Figure 2-3). In fact, the posterior probability that the road crossings per hectare effect size is zero or negative was 0.003, and represents the lowest uncertainty in any effect size estimate (Figure 2-2). The mean of the indicator vector ( $I_j$ ) for each variable further supported the selection of the four aforementioned variables with significant coefficients (Table 2-2).

### *Relationships with existing brook trout population assessments*

$N_b$  estimates were significantly positively related to two out of the three independent metrics of relative population status. The mean posterior slope coefficient of the TU future security score was 0.648 (CrI 0.260–1.044). The mean posterior slope coefficient of the SHEDS probability of occupancy was 0.880 (CrI 0.478–1.38). Finally, the mean posterior slope coefficient of the TU habitat integrity score was -0.458 (CrI -1.18–0.284). Bayesian P-values indicated adequate goodness of fit for these models with 0.530, 0.616 and 0.476 for future security score, probability of occupancy, and habitat integrity score, respectively. We observed adequate model convergence as indicated by visual inspection of chains and by the maximum potential scale reduction factor ( $< 1.01$ ).

### **Discussion**

Our results demonstrate a relationship between known drivers of population status and single-sample estimates of effective number of breeders in a set of fragmented brook trout populations. Smaller estimates of  $N_b$  were associated with less available habitat, higher temperatures, and less influence of groundwater on surface flow. These relationships emphasize the importance of conservation activities that increase the amount of interconnected brook trout habitat (Wood et al. 2018), mitigate increasing stream temperatures due to climate change, and maintain or improve hyporheic exchange (Weber et al. 2017). Importantly, these results demonstrate that  $N_b$  is sensitive to factors of high relevance to conservation practitioners, is in concordance with independent population assessments, and supports that there is value added to genetic monitoring programs by incorporating  $N_b$  estimates.

Traditional genetic monitoring using estimates of genetic variation (e.g., heterozygosity, allelic richness) and genetic structure (e.g.,  $F_{ST}$  among populations) provide insights into the past, such as longer-term  $N_e$ , migration, and population bottlenecks. In contrast,  $N_b$  is determined by the number of successfully reproducing parents, the environmental conditions under which those parents reproduced, and the conditions affecting early rearing success prior to sampling the focal cohort. The importance of reproductive habitat, juvenile abundance, and survival has been well-

demonstrated in brook trout populations and is particularly important in small habitat fragments (Letcher et al. 2007, Bassar et al. 2016, Kanno et al. 2016b). Gaining insights into these processes have often been difficult, time-intensive, and imprecise (e.g., redd counts; (Dunham et al. 2001)). Another benefit is that with using a few life-history traits  $N_b$  can readily be converted to generational  $N_e$  (Waples et al. 2013) and put in the context of existing conservation frameworks (e.g., 50/500 rule; (Franklin 1980)). Importantly, our work demonstrates that  $N_b$  estimated across populations can identify factors likely influencing reproductive processes across a broad spatial scale, and potentially represents a valuable supplement to existing population prioritization schemes. Future work should aim to better describe the relationship between  $N_b$  and near-term persistence probability in this and other species, and intrapopulation variation in  $N_b$  over time.

The complex evolutionary, ecological, and demographic factors that influence  $N_b$  estimates can complicate their interpretation and application to conservation and management. In the introduction we provide a theoretical example of how  $N_b$  can become decoupled from the absolute number of potential breeders, and how environmental factors (i.e., limited spawning habitat) could strongly influence estimates. An empirical example in Atlantic salmon (*Salmo salar*) illustrates this point, lower  $\hat{N}_b$  is associated with intermediate spawning aggregations, where highly competitive males can dominate reproduction at a few breeding sites. Conversely, when breeding sites are diffuse or abundant (i.e., low spawning aggregation) there is less variance in reproductive success and thus higher  $\hat{N}_b$ . Complicating matters, high density spawning aggregations appear to induce scramble sexual competition resulting in lower variance in reproductive success and elevate  $\hat{N}_b$  in Atlantic salmon (Bacles et al. 2018). Similarly,  $N_b$  did not increase in a brook trout population following removal of non-native competitor, despite an increase in the absolute number of contributing breeders due to an increase in variance of reproductive success (Miller et al. 2019). These examples highlight that interannual variation in  $N_b$  may not be conducive to straightforward interpretation, such as inferring  $\hat{N}_c$  from  $\hat{N}_b$ , due to complex density-dependent and ecological factors (Bernos and Fraser 2016; Yates et al. 2017). Our approach of comparing  $N_b$  among populations is consistent with the observation that  $N_b$  is relatively stable within populations of brook trout and varies dramatically among populations, likely corresponding to reproductive habitat

quantity and quality (Whiteley et al. 2015a, Bernos and Fraser 2016).

Mid-Atlantic brook trout are vulnerable to climate change and the results of our model demonstrate a link between climatic variables and  $\hat{N}_b$  in this cold-water species. During the last 50 years, water temperature in the mid-Atlantic region increased faster than air temperature at a rate of 0.028 °C per year (Rice and Jastram 2015) and is negatively associated with brook trout occupancy (Deweber and Wagner 2014, Merriam et al. 2019). We found a strongly negative effect of mean maximum annual air temperature on  $\hat{N}_b$  in brook trout populations in this region. Additionally, we found that base flow index was positively associated with  $\hat{N}_b$ . High base flow is associated with stability of stream flow and low sensitivity of water temperature to changes in air temperature (Trumbo et al. 2014). High base flow index may also indicate that a habitat patch has higher quality spawning habitat due an abundance of hyporheic exchange (Curry and Noakes 1995), which would also yield higher  $\hat{N}_b$  on average. The importance of temperature on brook trout is further emphasized by our exploratory analysis of two independent population assessments. The positive relationship between SHEDS occupancy model is likely driven by temperature, as it is the most significant effect in the occupancy model (Walker et al. 2021). TU's future security score represents a transformed estimate of stream temperature, and therefore the positive association with  $\hat{N}_b$  observed here is unsurprising, however, these results lend further support to the use of stream temperature as an index of future security.

Contrary to our hypotheses, our model did not support the inclusion of variables meant to represent competition with nonnative salmonids, acidification, and anthropogenic land use. Brook trout and nonnative salmonids have been observed directly competing for spawning habitat (Grant et al. 2002) and nonnative salmonids have displaced or extirpated populations of brook trout (Larson and Moore 1985, Kanno et al. 2016a). Both phenomena provide an explicit mechanism through which nonnative competitors would reduce the absolute number of breeders in a brook trout population. We conclude that presence or absence of nonnative salmonids does not affect  $\hat{N}_b$  in our sample; however, we hypothesize that a metric that accounts for relative abundance of nonnative salmonids would reveal the well-documented, negative effects of competition. Acidification has had a consistent, negative impact on brook trout populations throughout

their range (Hudy et al. 2000, Nislow and Lowe 2003). However, our index of deposition was likely inadequate to capture a biologically relevant summary of the geochemical conditions in a brook trout habitat patch. The population response to acidifying precipitation is complex and depends not only on the quantity of acidifying precipitation but also on buffering capacity of the catchments and aggravating factors (e.g., aluminosilicate deposits; (Schofield and Trojnar 1980)). A large number of studies have also documented negative associations with anthropogenic land use (e.g, agriculture, residential development, and deforestation) and brook trout occupancy at a variety of spatial scales (Deweber and Wagner 2014, Kanno et al. 2015, Merriam et al. 2019). We found no support for a positive effect of percent canopy cover, which was used as an index of anthropogenic disturbance of forests. The difference between prior studies on brook trout distribution and our results may simply be an artifact of  $N_b$  estimates being conditional on occupancy and given that a patch is occupied, relatively less canopy cover does not have a marked, negative impact on  $\hat{N}_b$  in the mid-Atlantic region.

Our results also revealed an unexpected pattern involving the relationship between  $\hat{N}_b$  and road crossing per hectare. Road crossings per hectare had a significant, positive association with  $\hat{N}_b$  and its parameter estimates had the lowest uncertainty and the second highest magnitude of all modelled variables. This result is unexpected because of the negative effects of forestry-based roads on brook trout spawning habitat through the deposition of silt and other fine sediments have been well understood for over half a century (Saunders and Smith 1965), and directly affect the reproductive processes that are reflected in  $\hat{N}_b$  (Hartman and Hakala 2006). We find two potential interpretations of these data compelling. First, the opportunistic, nonrandom selection of habitat patches for sampling may have given rise to a spurious correlation involving road access. Second, it may be that robust brook trout populations are attractive for recreational development and are more likely to have more stream road crossings within that habitat patch. More generally, our approach using nonrandom sampling and a modest sample size ( $n = 71$ ) is certainly a limitation of this work. The strength of our findings is derived from its congruence with previous research on this species, including the significance of climate related variables and habitat quantity, and the ability to illuminate such relationships with single genetic samples.

Sampling with the intent to estimate  $N_b$  is difficult in stream-dwelling brook trout populations due to continuous population structure and the inherent vulnerability of stream systems to fragmentation. Many mid-Atlantic brook trout populations exist in small habitats that can be described as discrete populations that generate unbiased estimates of the population parameter  $N_b$  (Whiteley et al. 2012), however larger populations or metapopulations exist in this region (Huntsman et al. 2016). Even without adding the complexity of metapopulation dynamics or barriers to movement, larger habitat patches have continuous genetic structure that is best described with the concept of genetic neighborhoods (Wright 1946), rather than as single, panmictic populations. Pooling individuals from multiple genetic neighborhoods creates a two-locus Wahlund effect or mixture LD and biases  $\hat{N}_b$  low (Neel et al. 2013, Whiteley et al. 2017). Moreover, incidental sampling above and below an instream barrier can induce the same phenomenon and bias  $\hat{N}_b$  low. There is little doubt that some the habitat patches sampled in this study are large enough that they do not represent a discrete population, such as the Savage River tributaries in Maryland, USA (Kazyak et al. 2016). However, in our data set we do not observe a positive relationship between  $F_{IS}$  and habitat patch size that would be expected when pooling multiple genetic neighborhoods and we found a consistent positive relationship between patch size and  $\hat{N}_b$ . We would expect that at a certain habitat patch size,  $\hat{N}_b$  would no longer increase proportionally with habitat size because the existence of more potential breeders does not apply at the scale of the  $N_b$  estimate. Our results suggest that being part of a larger habitat patch may have benefits that exceed the simple numerical addition of more breeders given more available habitat. Further work is needed to address sampling issues in continuously distributed populations and the proper interpretation of estimates at various spatial scales.

### *Conclusions*

Single-sample estimates of effective number of breeders demonstrated links between population status and temperature, base flow index, and available habitat. Our results provide further support for the use of  $N_b$  for genetic monitoring, not only because it is theoretically meaningful, but it is sensitive to factors that affect brook trout population dynamics and occupancy. Additionally, single cohort genetic samples have other valuable

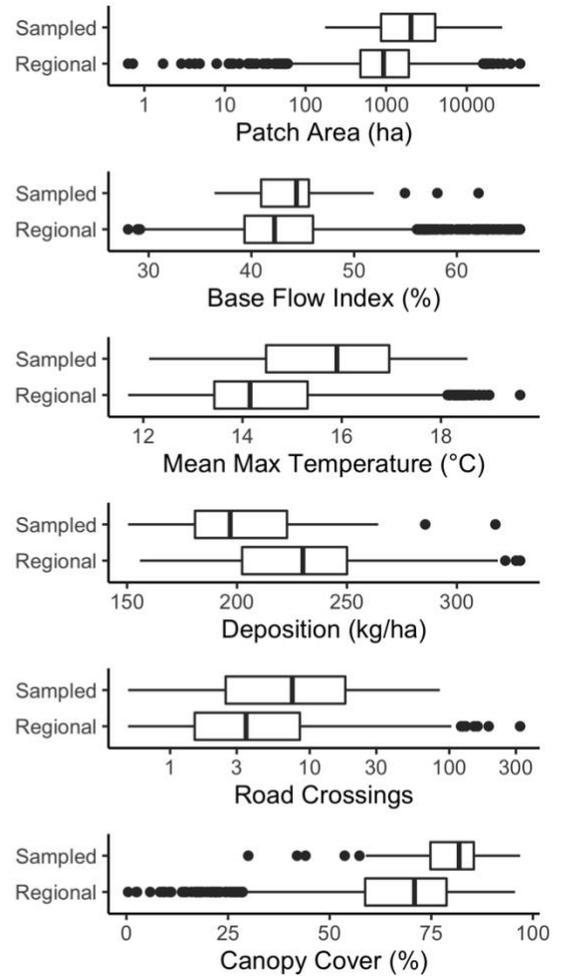
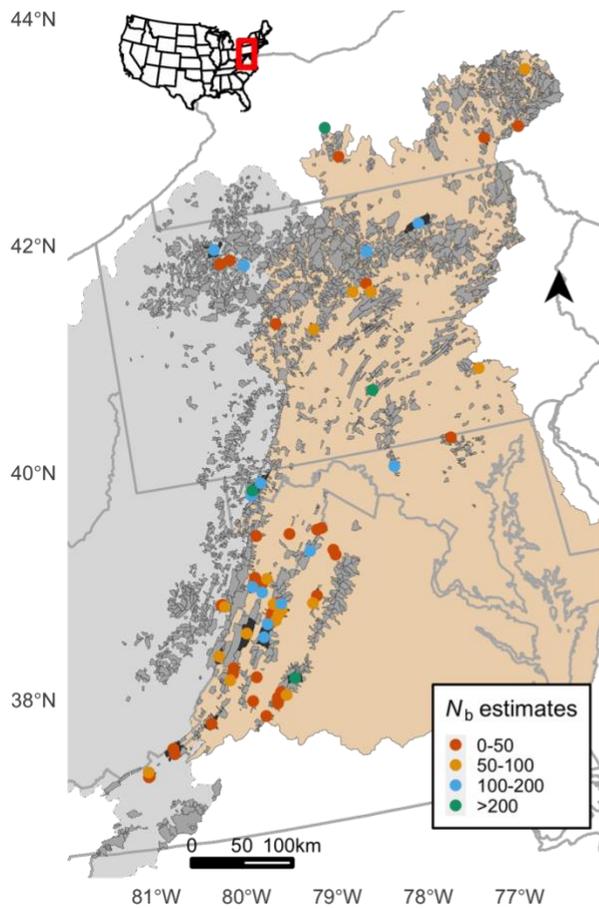
applications such as identifying likely spawning locations using sibship reconstruction (Hudy et al. 2010) or assessing potential instream barriers (Whiteley et al. 2014a). Using the same single-cohort samples, additional summaries can be calculated, such as the number of families captured per unit sampling effort and variance in family size, which can provide closely related and ecologically meaningful information to an audience untrained in genetics. Expanding models such as the one we provide can aid in population prioritization by identifying populations that have positive or negative residuals. The model formulation could be modified explicitly for this purpose, to only account for things outside of the control of conservation management (e.g., habitat size), and identify populations that are relatively robust or imperiled given the physical conditions of the habitat patch. The effective number of breeders and single-cohort samples more generally provide powerful insights into a population and can likely complement many ongoing demographic and genetic monitoring programs in a diverse array of taxa.

**Table 2-1. Description of variables included in models of  $N_b$ .** The time period over which the variable applies, the hypothesized direction of effect on  $N_b$ , and data source are also reported.

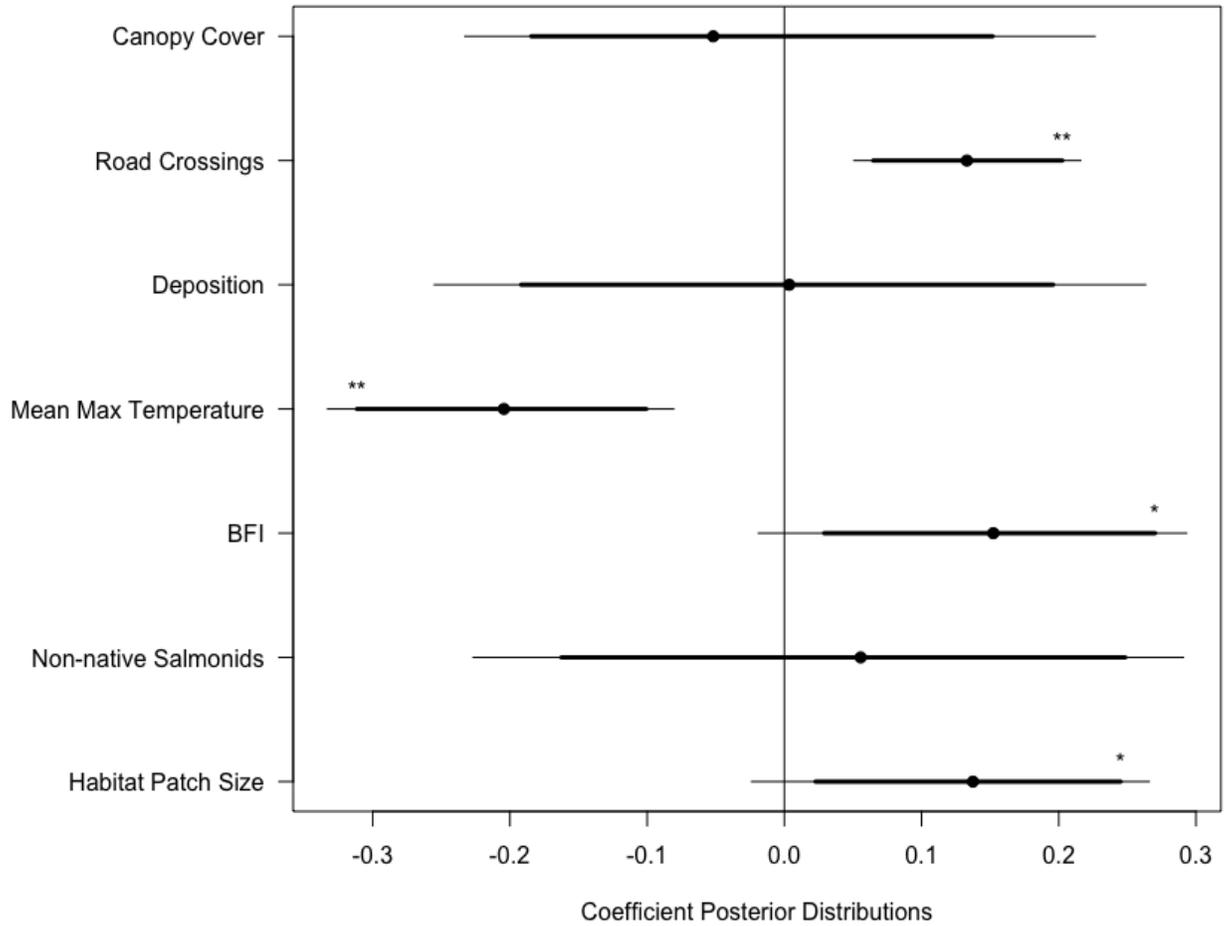
Variable Name	Variable Time Period	Description	Hypothesis	Source
Patch Size	–	Index of Available Habitat. Product of patch area (ha) and stream length (km).	Positive relationship	EBTJV and NHD+V2
Presence of non-native salmonids	–	Index of competition with non-native salmonids	Negative relationship	EBTJV and Field Sampling Reports
Base Flow Index	2003	Index of flow and temperature stability and reproductive habitat.	Positive relationship	U.S. Geological Survey
Mean Max Air Temperature	1991 – 2020	Index of thermal environment	Negative relationship	PRISM Climate Group
Mean Cumulative Sulfur & Nitrogen Deposition	2000-2002 2010-2020	Index of acidifying precipitation	Negative relationship	National Atmospheric Deposition Program
Road Crossings per hectare	2011	Index of human activity and disturbance	Negative relationship	Delorme Road Layer
Percent Canopy Cover	2016	Index of anthropogenic land use and deforestation	Negative relationship	National Landcover Database

**Table 2-2. Model results of Bayesian generalized linear mixed model.** Mean parameter estimates and 95% credible intervals are reported. The proportion of iterations that a parameter was included in the model is also reported.

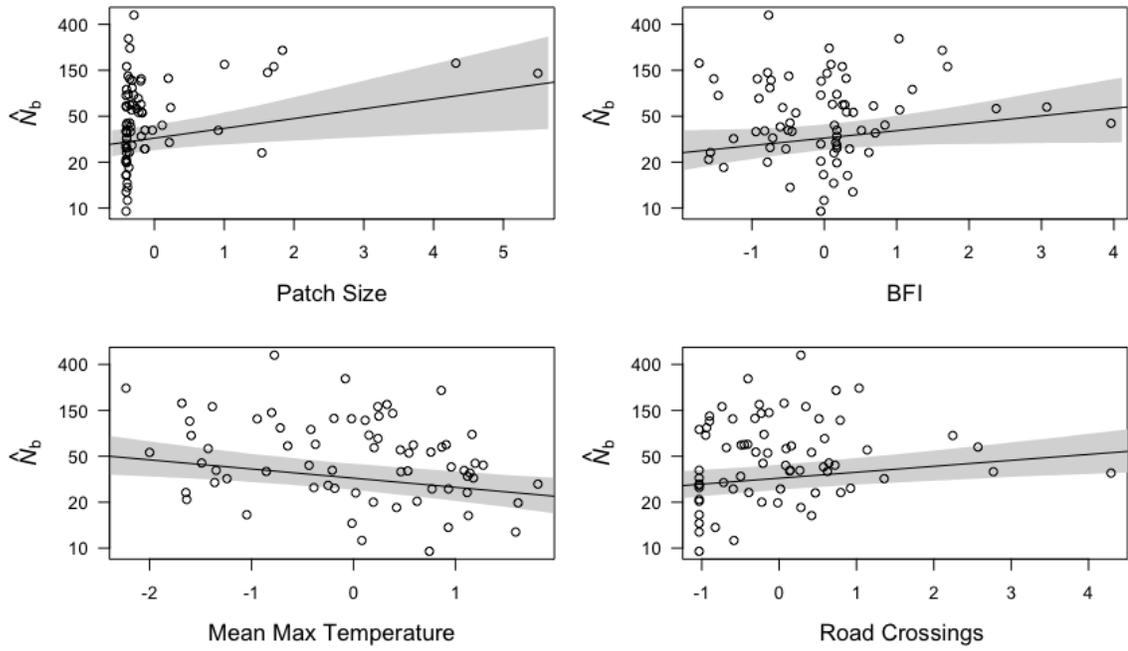
Parameter	Mean Estimate	95% Credible Interval	$P(I_j=1)$
Mean Intercept	3.49	3.22 – 3.82	
Patch Size	0.136	-0.031 – 0.266	0.946
Non-native salmonids	0.057	-0.224 – 0.290	0.702
BFI	0.152	-0.019 – 0.294	0.953
Mean Max Temperature	-0.204	-0.332 – -0.080	0.996
Deposition	0.002	-0.261 – 0.258	0.561
Road Crossings	0.138	0.050 – 0.215	0.993
Canopy Cover	-0.053	-0.233 – 0.221	0.706
$pr(I=1)$	0.715	0.369 – 0.955	



**Figure 2-1. Map of study area in the mid-Atlantic region of the United States.** The brown and gray shaded areas indicate the Chesapeake Bay and Ohio River region, respectively. Each point on the map corresponds to a sampled EBTJV brook trout habitat patch ( $n = 71$ ) colored with respect to the magnitude of  $\hat{N}_b$ . The numerous dark gray drainages areas ( $n = 2773$ ) represent unsampled habitat patches. Box plots present the distribution of the variables used in modeling for the sampled habitat patches (Sampled) compared to all habitat patches in the focal region (Regional).



**Figure 2-2. Posterior distribution of slope coefficients for habitat patch size, non-native salmonid presence, BFI, mean maximum annual temperature, deposition, road crossings, and canopy cover.** The mean estimate (point), 90% (darker line) and 95% (lighter line) credible intervals are reported. Statistical significance ( $\alpha = 0.05$ ) is denoted with a one or two asterisks for one-tailed and two-tailed tests, respectively.



**Figure 2-3. Linear relationships between the four significant covariates and  $N_b$ .** The gray area represents 95% credible intervals of the slope coefficient.

### **CHAPTER 3: Evaluating the outcomes of genetic rescue attempts**

Zachary L. Robinson, Donovan A. Bell, Tashi Dhendup, Gordon Luikart, Andrew R. Whiteley, Marty Kardos

This chapter has been published as:

Robinson, Z.L., Bell, D.A., Dhendup, T., Luikart, G., Whiteley, A.R. and Kardos, M. (2021), Evaluating the outcomes of genetic rescue attempts. *Conservation Biology*, 35: 666-677.

#### **Abstract**

Augmenting gene flow is a powerful tool for the conservation of small, isolated populations. However, genetic rescue attempts have largely been limited to populations at the brink of extinction, in part due to concerns over negative outcomes (e.g., outbreeding depression). Increasing habitat fragmentation may necessitate more proactive genetic management. Broader application of augmented gene flow will, in turn, require rigorous evaluation to increase confidence and identify pitfalls in this approach. To date, there has been no assessment of best monitoring practices for genetic rescue attempts. We used genomically explicit, individual-based simulations to examine the effectiveness of common approaches (i.e., tests for increases in fitness, migrant ancestry, heterozygosity, and abundance) for determining whether genetic rescue or outbreeding depression occurred. Statistical power to detect the effects of gene flow on fitness was high ( $\geq 0.8$ ) when effect sizes were large, a finding consistent with those from previous studies on severely inbred populations. However, smaller effects of gene flow on fitness can appreciably affect persistence probability but current evaluation approaches fail to provide results from which reliable inferences can be drawn. The power of the metrics we examined to evaluate genetic rescue attempts depended on the time since gene flow and whether gene flow was beneficial or deleterious. Encouragingly, the use of multiple metrics provided nonredundant information and improved inference reliability, highlighting the importance of intensive monitoring efforts. Further development of best practices for evaluating genetic rescue attempts will be crucial for a responsible transition to increased use of translocations to decrease extinction risk.

## Introduction

Human-driven habitat loss and fragmentation have produced many small, isolated populations that face increased extinction risk due to interactions between demographic and genetic factors (Gilpin and Soule 1986). Augmenting gene flow into small, isolated populations can alleviate inbreeding depression and increase persistence probability across diverse taxa (i.e., genetic rescue (Frankham 2015, Bell et al. 2019)). The promise of genetic rescue has led to recent calls for increased use of augmented gene flow (Ralls et al. 2018, Chan et al. 2019). However, concerns over negative outcomes have restricted genetic rescue attempts to populations on the brink of extinction (e.g., Johnson et al., 2010; Weeks et al., 2017), yet many populations with less severe inbreeding depression would benefit from gene flow.

A major concern with genetic rescue is that gene flow can instead decrease fitness (i.e., outbreeding depression) (Edmands 2007), potentially increasing extinction risk. Current guidelines to avoid outbreeding depression suggest selecting source and recipient populations that occur in similar habitats and with low population divergence thereby reducing the risk of adaptive differentiation or genomic incompatibilities, respectively (Frankham et al. 2011). When these guidelines are followed and inbreeding depression is severe, fitness benefits are often substantial and outbreeding depression is unlikely (Frankham 2015). However, source populations satisfying these guidelines may be unavailable for threatened species. Although evolutionary theory suggests that gene flow often disrupts local adaptations (Lenormand 2002), severely inbred populations are expected to be less locally adapted. Importantly, the benefits of heterosis may offset the consequences of disrupted local adaptation (Vergeer et al. 2004, Fitzpatrick et al. 2020).

Previous research has made considerable progress in understanding the genetic effects of translocations through intensive monitoring. Although increasing persistence probability is the primary aim of genetic rescue attempts, persistence probability is rarely estimated (but see van de Kerk et al., 2019). Instead, researchers typically use genetic and demographic metrics associated with population viability, including genetic variation, migrant ancestry, vital rates, and population growth rate (e.g., Weeks et al. 2017, Hasselgren et al. 2018). Population growth rate may be the most important metric because it is strongly associated with persistence probability (Bell et al. 2019).

Other support for genetic rescue comes from monitoring of vital rates that strongly affect population growth rate (Hedrick and Fredrickson 2010), but collecting fitness data is expensive and time consuming. An important question is whether genetic monitoring approaches can reliably be used to infer the outcomes of translocations with lower costs and sampling effort. Hedrick et al. (2011) suggest that proportion of migrant ancestry in the population is a useful surrogate of hybrid fitness. An increase in migrant ancestry by more than neutral expectations suggests higher fitness of hybrids and immigrants relative to residents. Alternatively, individual heterozygosity may be an effective metric due to its strong association with individual inbreeding (Kardos et al. 2016b) and its positive relationship with fitness (Chapman et al. 2009). However, heterozygosity almost always increases following gene flow and is negatively correlated with fitness when outbreeding depression occurs (Bell et al. 2019). As with migrant ancestry, interpreting changes in heterozygosity requires a comparison with a null expectation.

Despite the need for careful evaluation of genetic rescue attempts, there has not been an assessment of best monitoring practices. Two questions that are crucial to address before broader-scale implementation of genetic rescue are under what fitness effects of gene flow and demographic conditions can the outcomes of augmenting gene flow be reliably predicted and what are the best genetic and demographic approaches for evaluating genetic rescue attempts? We used individual-based, genomically explicit simulations to evaluate the extent to which different fitness effects of gene flow change persistence probability. We also examined the performance of genetic and demographic metrics to identify the outcomes of augmented gene flow. Based on our results, we recommend best practices for evaluating the outcome of genetic rescue attempts.

## **Methods**

### *Simulation framework overview*

We used stochastic, individual-based, genomically explicit simulations to evaluate patterns and statistical power of common genetic and fitness-related metrics following gene flow. This simulation framework is a modified parameterization of previously

published simulations (Kardos et al. 2015, Hedrick et al. 2016) and is freely available ([https://github.com/zakrobinson/GR\\_Eval](https://github.com/zakrobinson/GR_Eval)). The simulations and subsequent analyses were implemented in the statistical computing program R, version 3.4.1 (R Core Team 2017).

The simulations included three phases (figure S3-1). First, a large population of constant size evolved for 20 generations. Second, habitat fragmentation was mimicked by splitting the population into two subpopulations. We designated one population as the small, recipient population ( $n$  varied) and the other as the large, migrant source population ( $N = 300$ ) to mirror common genetic management translocations. The two populations evolved with constant size and without migration until  $F_{ST} \approx 0.2$  (range 0.192–0.208), which occurred in 52 generations on average. We selected an  $F_{ST}$  of 0.2 because it is widely used in conservation planning and is theoretically equivalent to one migrant per generation (Mills and Allendorf 1996). The fitness effects of gene flow were explicitly parameterized and were unaffected by  $F_{ST}$ . The recipient population then evolved for 10 generations with ceiling-type density-dependent population growth (Lande 1993), which provided realistic levels of demographic stochasticity and genetic drift. Finally, a single migration event occurred with a varying number of individuals from the source population to the recipient population. The recipient population then evolved for 30 generations after gene flow, which permitted persistence probability to be calculated for a conservation-relevant time frame. We evaluated the influence of two carrying capacities, two levels of immigration, and seven different levels of immigrant fitness. We defined each simulation scenario as a unique combination of these factors. Considering all parameter combinations, we examined 28 simulation scenarios, each with 500 replicate simulations (table S3-1).

#### *Genomic parameterization and selection scenarios*

We simulated 400 Mb genomes with four equal-size chromosomes. The recombination rate was 1 cM/Mb, which is similar to humans (Dumont and Payseur 2008), resulting in a genetic map length of 400 cM. Each chromosome was randomly assigned 100 polymorphic, equally spaced, diallelic loci (i.e., 1 SNP/Mb). Allele frequencies were initially beta distributed ( $\alpha = 0.5$ ;  $\beta = 0.5$ ) prior to differentiation of source and recipient

populations, resulting in a typical U-shaped allele frequency distribution. Population diagnostic loci (i.e., where immigrants were fixed for one allele and residents for another) were superimposed on each variable genomic location to calculate the exact genome-wide migrant ancestry proportions following gene flow (i.e., each allele's population of origin was known).

The fitness effects of gene flow were parameterized as differences among residents, immigrants, and hybrids in fecundity and survival probability. We specified the strength and direction of selection as the signed proportional fitness of pure immigrants relative to pure residents ( $w_I$ ). We included seven different levels of relative fitness, including  $w_I = -0.8, -0.4, -0.1, 0, 0.1, 0.4, \text{ and } 0.8$ , which were applied to both survival and fecundity. For example,  $w_I = 0.8$  meant that a pure immigrant had 80% higher survival and 80% higher fecundity relative to a pure resident. The fitness effects were equally distributed across all loci and represented outbreeding depression (negative  $w_I$ ), demographic rescue ( $w_I = 0$ ), and genetic rescue (positive  $w_I$ ).

We defined a hybrid as an individual with both migrant and resident ancestry, and hybrid fitness was determined by the number of migrant alleles at each of the 400 diagnostic loci. In beneficial gene-flow scenarios, resident alleles were completely recessive to migrant alleles because inbreeding depression appears to be primarily caused by deleterious recessive alleles (Charlesworth and Willis 2009). In deleterious gene-flow scenarios, immigrant alleles had purely additive effects because outbreeding depression is likely caused by disruption of local adaptation and local adaptation is often due to quantitative traits governed by additive genetic variation (Pritchard and di Rienzo 2010).

#### *Demographic parameterization*

The simulated populations were sexually-reproducing, non-selfing, and randomly-mating with discrete generations. Density-dependent population growth was dictated by mean survival and fecundity with ceiling-type density dependence (Lande 1993). We used two carrying capacities ( $K = 50$  and  $K = 150$ ). Abundance was initialized at  $0.5K$  ( $N_0 = 25$  and  $75$ ). Prior to gene flow, the individual survival probability was  $S = 0.5$  and the number of offspring per female was drawn from a Poisson distribution with a mean of 3.92 when  $N_e \leq K$ . The selected fecundity and survival probabilities mimic  $K$ -selected species, which

have historically been the focus of genetic rescue attempts (e.g., (Johnson et al. 2010, Poirier et al. 2019)). This combination of survival and fecundity resulted in a slightly negative population growth rate ( $\lambda = 0.98$ ) in a population composed of only resident individuals. Thus, population dynamics after immigration were an emergent property of introgression, natural selection, genetic drift, and demographic stochasticity.

Immigrants were randomly selected from the source population and mate pairing was random. We tested 4 and 12 immigrants for simulations with  $K = 50$  and increased immigration proportionally to carrying capacity for simulations with  $K = 150$ , where 12 and 36 immigrants were introduced. Additionally, we simulated populations without immigration to characterize the effect of gene flow on population viability.

#### *Metrics for inferring fitness effects of gene flow*

We evaluated metrics that are often estimated to infer the fitness effects of gene flow, including heterozygosity, migrant ancestry, abundance (i.e., census size), survival, and lifetime reproductive success (LRS) (table S3-2). Mean observed population-level heterozygosity was calculated each generation, and individual-level heterozygosity was calculated for every individual. Given the genetic marker density in our simulated genomes, individual heterozygosity is strongly correlated with individual inbreeding (Kardos et al. 2016b). Migrant ancestry was calculated as the proportion of migrant alleles and was estimated at the population and individual level during each generation. Survival information was retained for all individuals and abundance was recorded every generation. We calculated LRS as the number of surviving offspring per individual. Additionally, persistence probability was calculated as the proportion of the 500 simulation repetitions for each scenario in which the population persisted from the onset of gene-flow to the end of the simulation (30 generations).

#### *Power analysis*

We estimated the statistical power of tests based on observed heterozygosity ( $H_o$ ), migrant ancestry, abundance, survival, and LRS to identify the fitness effects of gene flow. Mean generational migrant ancestry, observed heterozygosity, and abundance were compared with the simulated neutral gene flow scenario. Percent change in

heterozygosity relative to the baseline before gene flow (generation 9) was calculated for each generation after gene flow. We then calculated the probability of each observed value of generational percent change in heterozygosity under the neutral gene flow distribution (500 neutral observations per simulation scenario). For each non-neutral simulation scenario, we reported power as the proportion of repetitions with a statistically significant probability ( $\alpha = 0.05$ ) under the simulated neutral distribution given it was in the parametrized direction of fitness effects (i.e., single-tailed test). We used the same approach for the power analyses of migrant ancestry and abundance, except observed values, rather than percent change, were compared with the neutral distributions.

We used generalized linear models (GLMs) to test for the effect of migrant ancestry on survival and LRS separately for every generation after gene flow. The survival GLM had a binomial distribution and a logit link, and the LRS GLM had a Poisson distribution with a log link. We did not model LRS or survival when fewer than 10 individuals remained in the population. We measured power as the proportion of the simulation repetitions with a statistically significant regression coefficient in the parametrized direction of fitness effects (i.e., single-tailed test with  $\alpha = 0.05$ ).

#### *Habitat-constrained scenario*

Populations that are not declining but are severely constrained by habitat limitations could still benefit from augmented gene flow. We hypothesized that populations whose abundance is closer to  $K$  pose unique challenges for evaluation with common approaches. This was assessed by initializing the population at carrying capacity ( $K = 50$ ) with mean survival ( $S$ ) of 0.5 and mean fecundity of 4.08 (Poisson distributed), which resulted in a positive population growth rate ( $\lambda = 1.02$ ) in a population composed of only residents.

#### *Inference based on multiple metrics*

Reporting multiple demographic and genetic metrics is common in genetic-rescue studies, and an implicit weight of evidence approach is often used (e.g., Hasselgren et al. 2018). To date, there has been no assessment of whether additional metrics are redundant or improve reliability of inferences. We used a binomial GLM to evaluate which set of estimated parameters were best at distinguishing between simulated genetic rescue and

outbreeding depression. The dependent variable was the parameterized fitness effect and each non-neutral ( $w_I \neq 0$ ) simulation repetition was categorized as genetic rescue ( $w_I > 0$ ) or outbreeding depression ( $w_I < 0$ ). The predictor variables were the estimated effect sizes of survival, LRS, migrant ancestry, and abundance. For effect sizes of survival and LRS, we used the regression coefficients from the GLMs with migrant ancestry as a predictor. The effect sizes for migrant ancestry and abundance were their probability under the neutral gene-flow distribution (described above). Heterozygosity was excluded from the model because it had high collinearity with migrant ancestry (Pearson's  $r > 0.7$ ). We conducted this analysis for opposite parameterized fitness effects ( $w_I = -0.1$  and  $0.1$ ) with the 500 repetitions from the  $K = 150$ ,  $N_0 = 75$ , and 36 immigrants scenario. The analysis included data from the second generation after gene flow because genetic rescue studies are often limited to one or two generations. We used Akaike information criterion (AIC) model selection in AICmodavg (Mazerolle 2019).

## **Results**

### *Effects of immigration on persistence probability*

All positive and negative fitness effects of gene flow significantly altered population persistence probability. Genetic rescue scenarios increased persistence probability by 21.7–101.3%, whereas outbreeding depression scenarios decreased persistence probability by 10.3–78.7%. Demographic rescue alone (i.e., neutral gene flow) increased persistence probability by 7.4–18.3%. Higher immigration rates had larger effects on persistence (Figure 3-1). Changes in persistence probability appeared to be mediated by deterministic change in abundance and altered demographic stochasticity (Figure 3-2). Mean intergenerational variance in abundance, an index of demographic stochasticity, was 137% higher in neutral and deleterious gene flow scenarios relative to beneficial scenarios on average.

### *Trends in genetic metrics following gene flow*

Migrant ancestry and heterozygosity initially increased following immigration in all scenarios and were strongly influenced by genetic drift. Within a few generations after

gene flow, the influence of different fitness effects on migrant ancestry became increasingly apparent with migrant ancestry declining in deleterious gene flow scenarios and increasing in beneficial gene flow scenarios (Figure 3-2). Migrant ancestry and heterozygosity were highly correlated during the first and second generations after the immigration event (Pearson's  $r = 0.80$ ) but were less correlated over the entire period after gene flow (Pearson's  $r = 0.66$ ).

Importantly, heterozygosity remained elevated above its before-gene-flow trajectory for 10 generations following highly deleterious gene flow ( $w_I = -0.8$ ) and for the full duration of simulations with minorly deleterious gene flow ( $w_I = -0.1$ ) (Figure 3-2). However, the rate of loss of heterozygosity was greater in deleterious scenarios relative to neutral and positive gene flow scenarios (i.e., decreased effective population size (figures S3-2–S3-6). In beneficial gene-flow scenarios, heterozygosity had a greater increase and a slower rate of loss relative to neutral gene flow. The variance in abundance, migrant ancestry, and heterozygosity was high among iterations of the simulations with many beneficial and deleterious gene flow scenarios having overlapping distributions (Figure 3-2). These patterns suggest it may be difficult to determine whether gene flow had beneficial, neutral, or deleterious effects on population viability.

#### *Relative performance of metrics for evaluating genetic-rescue attempts*

The most powerful metric for inferring the fitness effects of gene flow changed depending on the time since immigration and the magnitude and direction of fitness effects (Figures 3-3 & 3-4). Statistical power for survival and LRS was maximal immediately following gene flow and declined over time (Figure 3-3). This temporal decline in the statistical power appeared to be caused by a rapid loss of variance in individual migrant ancestry. For example, greater than 50% of the variance in migrant ancestry was lost within 3 generations on average (figures S3-7–S3-10). Conversely, population-level migrant ancestry, heterozygosity, and abundance gained statistical power over time. In strongly deleterious gene-flow scenarios, genetic metrics outperformed survival and LRS for the entire period after gene flow (Figure 3-3). Higher abundance and immigration increased power of all metrics on average (Figure 3-4).

Despite census sampling and dramatic influences on persistence probability

(Figure 3-1), the statistical power of a single metric rarely exceeded the recommended threshold ( $P = 0.8$ ) (Cohen 1992) in the 10 generations following gene flow in the  $K=150$  scenarios (Figure 3-3). Generally, statistical power was low ( $P < 0.8$ ), except for the most extreme fitness effects in scenarios with high immigration or carrying capacity (Figure 3-4). Regression coefficients of survival and LRS were often in the opposite direction of the parameterized fitness effect when the fitness effects of gene flow were low. For example, in the second generation after immigration in the  $K=50$  scenario with minor deleterious gene flow ( $w_I = -0.1$ ), the regression coefficient was in the wrong direction 44.8% and 41.9% of the time and was statistically significant in the wrong direction 8.0% and 0.7% of the time for LRS and survival, respectively (Figure 3-5). Population trajectory provided little inferential value without at least 15 generations of monitoring after gene flow (figure S3-11 & S3-12).

The regression analysis we used to predict whether gene flow had minor deleterious ( $w_I = -0.1$ ) or beneficial ( $w_I = 0.1$ ) effects demonstrated that inference can be improved by combining information from multiple metrics (Table 3-1). The global model, including all metrics considered, was overwhelmingly supported ( $\Delta AIC_C > 10$ ). This result supports the idea that each metric considered provides unique information that can be used to predict the fitness consequences of gene flow.

#### *Habitat-constrained scenario*

The habitat-constrained scenario resulted in significantly increased persistence probability for all non-neutral, high-immigration-rate scenarios (Figure 3-1). Despite an initial abundance at carrying capacity ( $K = 50$ ) and a slightly positive population growth rate, the baseline (i.e., without gene flow) 30-generation persistence probability was 0.84 due to demographic stochasticity. The statistical power of all metrics was reduced by 19.7% on average relative to the declining populations ( $K = 50$ ) with an initial abundance of  $0.5K$ . Compared with scenarios with declining population sizes, genetic metrics had increased statistical power relative to individual, fitness-related metrics in the habitat-constrained scenario. Abundance had little inferential value in the habitat-constrained scenario (Figure 3-4).

## Discussion

Genetic rescue has emerged as a powerful strategy to conserve small, isolated populations (Frankham 2015). However, genetic rescue is an inherently complex ecoevolutionary process requiring careful monitoring and further study to increase confidence in its broader application. Our results represent the first assessment of commonly used metrics for evaluating genetic rescue attempts under realistic levels of demographic stochasticity and genetic drift. We found that smaller changes in survival and fecundity than have been previously reported from empirical genetic rescue studies can strongly affect population viability over a conservation-relevant time frame. We also found that common approaches for monitoring genetic rescue attempts have limited statistical power to identify these smaller fitness effects despite their large consequences for near-term persistence probability (Figure 3-1). Our results emphasize the need for long-term monitoring of multiple metrics and innovative modeling approaches for reliable inference.

### *Detecting the fitness effects of gene flow*

Recent calls have been made for broader application of genetic rescue based on its apparent success in highly inbred populations (Ralls et al. 2018, Chan et al. 2019). Genetic rescue studies have documented substantial increases in survival and fecundity (Frankham 2015). For example, an inbred Artic fox (*Vulpes lagopus*) population had a 190% increase in juvenile survival and a 130% increase in breeding success immediately following gene flow (Hasselgren et al. 2018). In our most extreme beneficial gene flow scenario, immigrants and their first-generation decedents showed an 80% increase in survival and fecundity relative to residents. In general, we found that the fitness effects of this magnitude are well documented with individual-based fitness metrics in the first few generations after gene flow (Figure 3-3). Although empirical studies rarely estimate persistence probability, our results are consistent with findings that the fitness effects of genetic rescue can strongly increase population viability (e.g., [van de Kerk et al. 2019]). However, despite substantial impacts on persistence probability, we found that small beneficial or deleterious effects of gene flow (e.g.,  $w_I = 0.1$  or  $w_I = -0.1$ ) were difficult to detect (Figure 3-3).

Genetic rescue and outbreeding depression were easier to detect in large recipient populations and when the number of immigrants introduced was high. This is likely due to an increase in the sample size of both residents and migrants and a decrease in genetic drift and demographic stochasticity in large populations (Gilpin and Soule 1986). However, population size and the fitness effects of gene flow are not independent in wild populations. Small populations are more likely to experience substantially reduced fitness due to inbreeding depression (Caughley 1994) and thus exhibit the largest genetic rescue effects. This creates a paradox in which populations need to be relatively large for one to detect the fitness effects of gene flow, whereas small populations are more likely to exhibit severe inbreeding depression. It may appear that our results suggest high immigration rates are advantageous, but we also found that high immigration rates increased extinction risk when outbreeding depression occurred (Figure 3-1). Further, high immigration rates may be undesirable due to concerns of genetic swamping (Harris et al. 2019) and biotic homogenization (Olden et al. 2005, Derry et al. 2019) or impractical due to financial or sociopolitical limitations.

#### *Temporal patterns and relative power of metrics*

A better understanding of the temporal patterns in relative power of metrics will improve the design of monitoring programs and lead to a better allocation of conservation funds. The precipitous decline in statistical power of individually based fitness metrics was due to a rapid loss of variance in migrant ancestry, which in turn resulted in reduced genetic variance in survival and fecundity. The ephemeral capacity to detect the fitness effects of gene flow with individual fitness metrics has implications in other conservation contexts (e.g., invasive hybridization). Survival had higher statistical power than LRS on average, which is likely due to LRS of an individual also depending on its mate's ancestry. In general, individual-fitness metrics performed better in beneficial gene-flow scenarios due to the positive selection creating high variation in migrant ancestry and thus individual fitness. Similarly, positive assortative mating increased statistical power of individual-fitness metrics by altering the magnitude and rate of loss of variance in migrant ancestry.

The population-level metrics increased in statistical power over time (Figure 3-3), which is the opposite temporal pattern of the individual-based metrics. These results

suggest that long-term monitoring of heterozygosity, migrant ancestry, and abundance can provide a powerful framework for inference. However, the duration of genetic rescue studies are often only a few reproductive cycles, which is when population-level metrics had the lowest power. In practice, generation intervals vary greatly and genetic monitoring for several generations may be prohibitively costly for some species. It is also crucial that these metrics be compared with a null expectation (i.e., neutral gene flow) that accounts for uncertainty arising from demographic stochasticity and genetic drift. An increase or decrease in genetic variation alone provided little inferential value for detecting genetic rescue or outbreeding depression; however, increasing genetic variation may increase adaptive potential, perhaps reaching conservation objectives even when local adaptation is temporarily disrupted (Derry et al. 2019).

The fundamental goal of genetic rescue attempts is to prevent population extinction, and arguably the best documentation is reversal of population declines (Bell et al. 2019). However, genetic rescue attempts are often restricted to a single imperiled population (e.g., Johnson et al. 2010) and thus have no replication or control populations (Tallmon 2017). This often-unavoidable limitation leaves environmental conditions as a possible explanation for increases in abundance. For example, Robinson et al. (2017) documented > 1000% increases in juvenile abundance in brook trout (*Salvelinus fontinalis*) populations following a translocation, but similar increases also occurred in the control population. These results demonstrate the need to employ additional strategies to attribute changes in abundance to the genetic contribution of immigrants (Tallmon et al. 2004). Additionally, gene flow can still decrease extinction risk in habitat-constrained populations, which greatly reduces the value of trends in abundance for evaluating genetic rescue attempts. Researchers will be unaware of the focal population's relationship to carrying capacity without long-term monitoring prior to translocations, and, despite its importance, detecting trends in abundance is often difficult (e.g., Dauwalter et al. (2009).

#### *Statistical power in natural populations*

Our results provide novel insights into monitoring genetic rescue attempts. However, our simulations provide optimistic estimates of statistical power. We used census sampling,

and our simulations did not include environmental variation. Importantly, we calculated individual fitness through lifetime survival and LRS, whereas many species' life histories (e.g., long-lived species) may only permit measuring stage-specific fitness components, such as juvenile survival or annual breeding success. Measuring stage-specific fitness components provides significant logistical advantages, but they may be weakly related to population growth rate and persistence (Wisdom et al. 2000). The two implicit assumptions of using migrant ancestry as a predictor of individual fitness are that fitness is linearly related to migrant ancestry (i.e., no intermediate optimum) and that fitness effects are distributed genome wide. Although these assumptions were met in our models, they will likely be violated in natural populations.

We determined whether migrant ancestry, heterozygosity, and abundance were significantly different from the simulated neutral gene flow scenario. It is noteworthy that our neutral distribution was parameterized identically to the scenarios that were being compared, which will not be the case in conservation practice. In natural populations, migrants will often introduce deleterious genetic variation regardless of the net effect on persistence probability, which may affect the temporal dynamics of genetic rescue, particularly in very small populations (Hedrick and Garcia-Dorado 2016). Although we provide optimistic estimates of statistical power, the relative performance of these metrics and their temporal patterns can help inform the design and monitoring of genetic rescue attempts.

#### *Assessing genetic rescue with multiple metrics*

We provide the first demonstration that migrant ancestry, vital rates, and abundance can provide non-redundant information and that monitoring multiple metrics can increase the ability to detect genetic rescue or outbreeding depression. This finding supports the implicit weight-of-evidence approach taken by many genetic rescue researchers. An important path forward is finding innovative ways to incorporate multiple metrics into a single modeling framework, such as an extension of integrated population models (Oppel et al. 2014). Estimating key demographic and life-history parameters among migrants, residents, and hybrids will permit population viability analyses to describe the translocations effect on persistence (e.g., van de Kerk et al. 2019, Poirier et al. 2019).

This approach is critical when experimental procedures (e.g., controls and replicates) are impractical due to the conservation status of the species or sociopolitical constraints.

#### *Implications for evaluating genetic rescue attempts*

Our results provide insights into how genetic rescue attempts are best monitored. Individual-based fitness metrics provide the highest statistical power for detecting genetic rescue in the first few generations after gene flow, which is the duration of most monitoring efforts. However, we found that these metrics rapidly lost statistical power over time. Conversely, population-level migrant ancestry and heterozygosity were the most powerful metrics for monitoring in later generations, but long-term monitoring and generating realistic null expectations pose significant challenges. Despite the clear relationship of population growth to persistence probability, making reliable inference based on trends in abundance requires a long time series and accounting for environmental influences on demography is difficult. However, monitoring multiple metrics can provide non-redundant information and enable population modelling and emerged as the best option in our study.

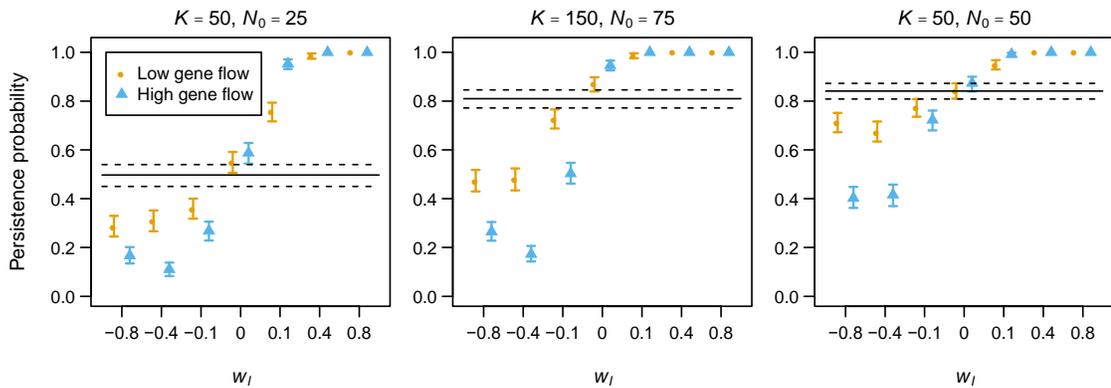
Together, our results provide a positive but cautionary message about more proactive augmentation of gene flow. We demonstrated that smaller fitness effects than historically reported in genetic rescue studies can have consequential effects on persistence probability. This supports the notion that augmented gene flow could be useful for many small, inbred populations before they are at the brink of extinction (Ralls et al. 2018). However, we also found that even intensive monitoring efforts may provide ambiguous results in these scenarios due to limited statistical power. Monitoring genetic rescue attempts is expensive, but the ecoevolutionary nature of genetic rescue provides ample opportunity for collaboration among population geneticists, ecologists, and conservation practitioners. Low statistical power and limited monitoring budgets mean that verifying genetic rescue will be impractical in many management contexts. This highlights the importance of long-term experimental research to improve understanding of the generality of successful genetic rescue attempts to date and increase confidence in the broader use of augmented gene flow as a tool for global biodiversity conservation.

**Table 3-1. Akaike information criterion (AIC) model selection for the top 5 binomial generalized linear models for distinguishing between minor deleterious and minor beneficial ( $w_I = -0.1, 0.1$ ) fitness effects in the second-generation following gene flow. McFadden's Pseudo  $R^2$  is shown.**

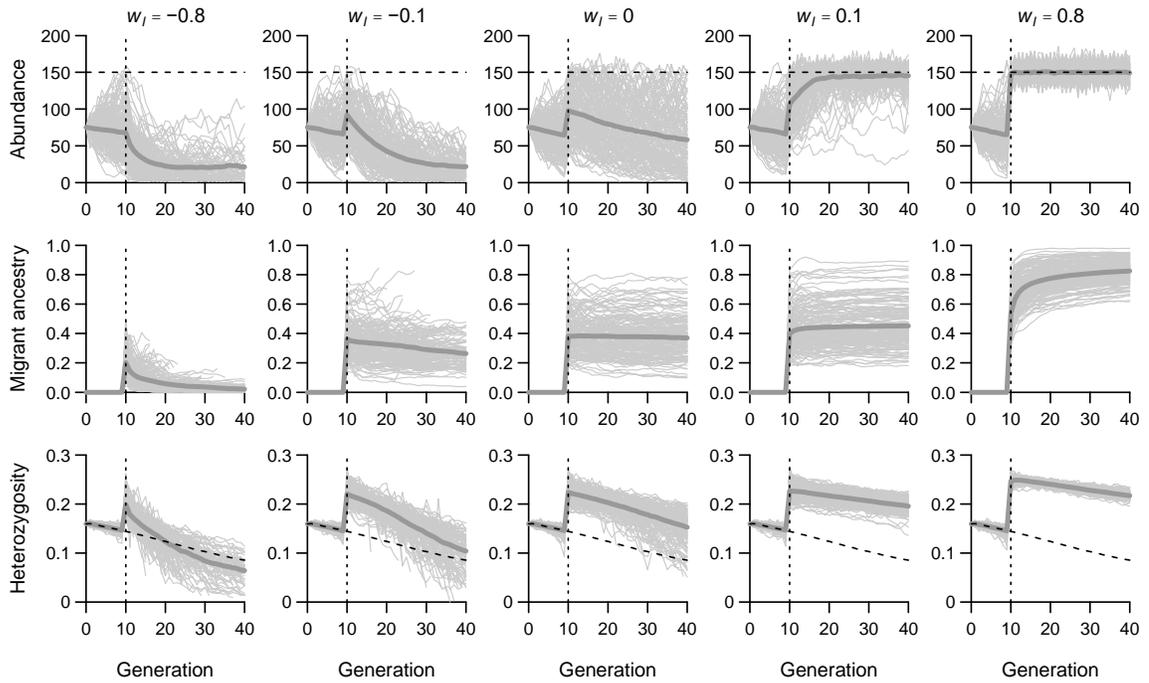
Models <sup>a</sup>	$\Delta AIC_c$	$R^{2b}$
MA + LRS + $N_c$ + survival	0.0	0.53
MA + LRS + $N_c$	10.2	0.52
MA+ $N_c$ + Survival	19.1	0.51
MA+ $N_c$	28.3	0.51
LRS + $N_c$ + Survival	384.0	0.25

<sup>a</sup>Abbreviations: MA, Migrant ancestry; LRS, Life-time reproductive success;  $N_c$ , Abundance

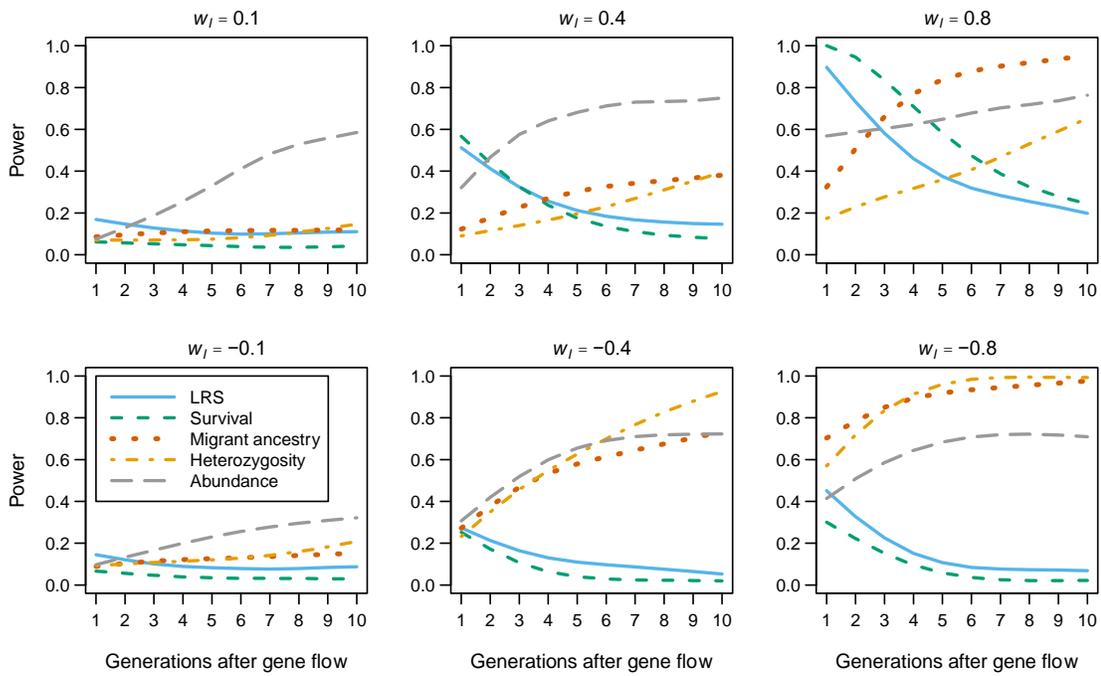
<sup>b</sup>McFadden's pseudo  $R^2$ .



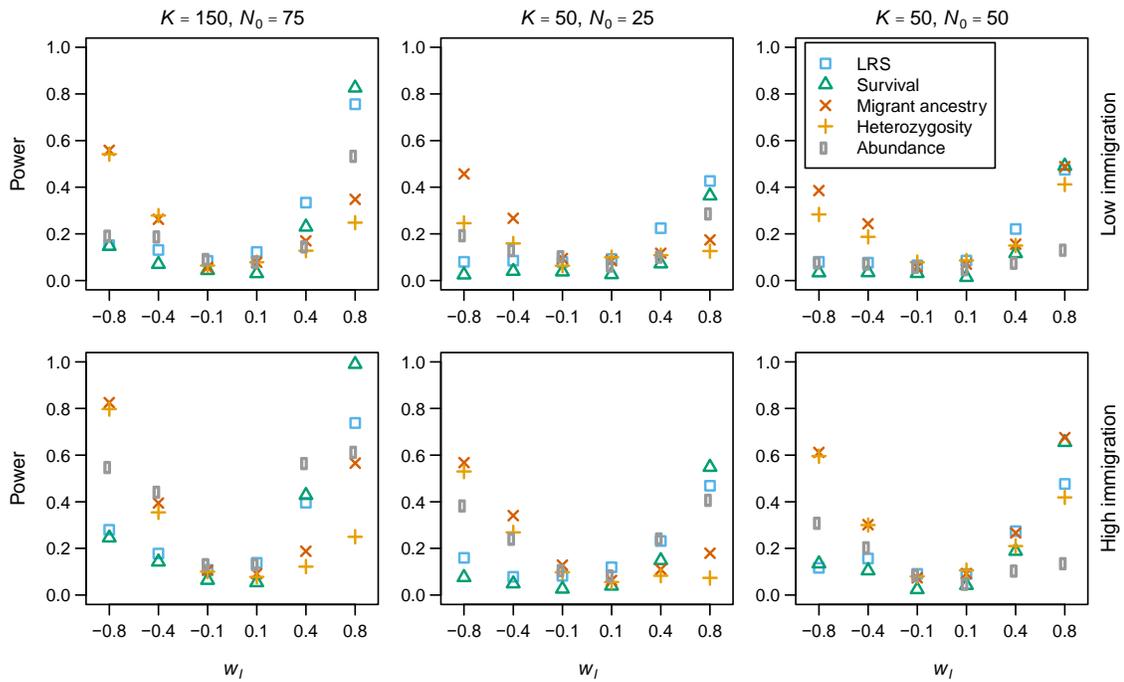
**Figure 3-1. The 30-generation persistence probability after gene flow of populations with differing carrying capacities, fitness effects, and levels of gene flow (points, mean persistence probability; bars, 95% bootstrap CIs; solid and dashed lines, mean and 95% bootstrapped CIs for mean population persistence probability without gene flow). High and low gene flow correspond to 4 and 12 individuals introduced when  $K$  (carrying capacity) = 50 and to 12 and 36 individuals introduced when  $K = 150$ , respectively. The initial abundance ( $N_0$ ) is reported at the top of each graph. The fitness effect of gene flow ( $w_I$ ) is expressed as the signed proportional change in survival and fecundity of pure immigrants versus pure residents.**



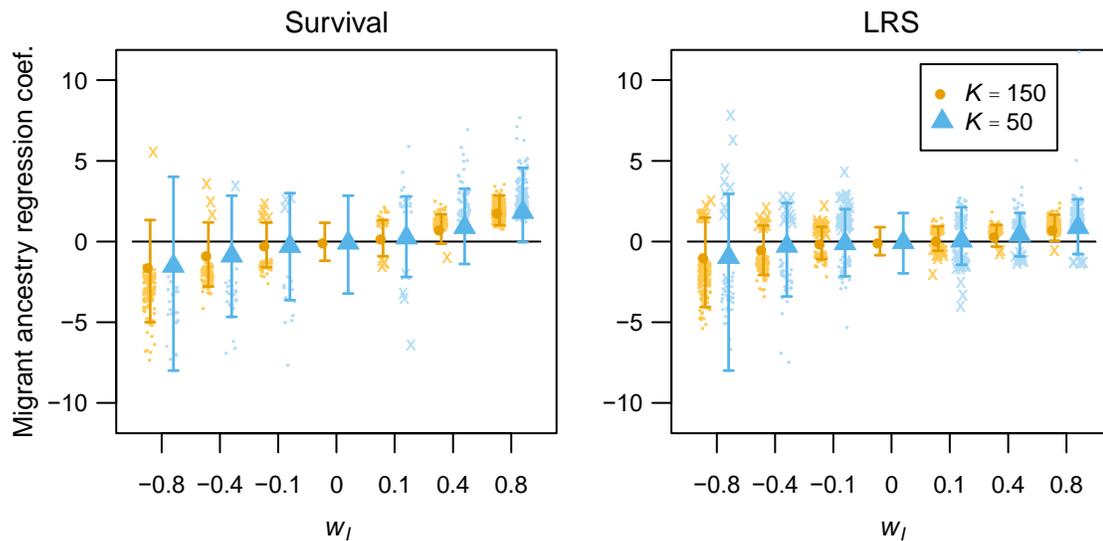
**Figure 3-2. Trends in abundance, migrant ancestry, and observed heterozygosity among different effects of gene flow on fitness ( $w_I$ ).** The simulation scenario of  $K$  (carrying capacity) = 150 and 36 immigrants is shown (dark gray lines, mean; light gray lines, 100 randomly selected iterations of the simulation; dotted vertical lines at year 10, single immigration event; dashed horizontal lines in abundance graphs,  $K$ ; dashed trend lines in heterozygosity graphs, trajectory before gene flow).



**Figure 3-3. Statistical power to detect the effect of gene flow on fitness using genetic and demographic metrics (LRS, lifetime reproductive success; Migrant ancestry, proportion migrant alleles, observed Heterozygosity; Abundance). The baseline scenario with a  $K$  of 150 and 36 immigrants is shown. Smoothing curves are shown to represent the general trends in power.**



**Figure 3-4. Statistical power to correctly infer the fitness effect of gene flow using genetic (Migrant ancestry, proportion migrant alleles;  $H_o$ , observed heterozygosity) and demographic metrics (LRS, lifetime reproductive success; Survival; Abundance). Six different simulated fitness effects of gene flow ( $w_I$ , signed proportional change in survival and fecundity of pure immigrants versus pure residents) are shown in the second generation after gene flow. Each column of graphs corresponds to the simulation scenarios' carrying capacity ( $K$ ) and initial abundance ( $N_0$ ).**



**Figure 3-5. Regression coefficients from generalized linear models of migrant ancestry as a predictor of survival and lifetime reproductive success (LRS) in the second generation after gene flow.** We present data for each simulated effects of gene flow on fitness ( $w_I$ , signed proportional change in survival and fecundity of pure immigrants versus pure residents). The panels correspond to  $K$  (carrying capacity) = 150 and  $K = 50$  under the high immigration scenarios (small dots, significant regression coefficients in the correct direction; x's, significant regression coefficients in the wrong direction). Means and 95% CIs are shown.

## **CHAPTER 4: Microhaplotypes increase parentage assignment accuracy for an experimental test of genetic rescue.**

### **Abstract:**

Next generation sequencing technology has dramatically increased the availability of molecular resources for non-model organisms. Genotyping-in-Thousands by sequencing (GT-seq) is an amplicon sequencing approach that provides a cost-effective method to genotype hundreds of amplicons for thousands of individuals on a single next generation sequencing lane. We provide a description of GT-seq panel design and optimization for use in parentage assignment in an ongoing genetic rescue study on brook trout (*Salvelinus fontinalis*). The optimized GT-seq panel is comprised of 166 amplicons with multiple single nucleotide polymorphisms (SNPs) that can be genotyped collectively as microhaplotypes and 53 amplicons with a single SNP. We used population genetic simulations to evaluate parentage assignment accuracy when genotyping all amplicons as single SNPs or using microhaplotypes when available. We tested the effect of unsampled parents, missing genotypic data, and unknown parental sex on parentage assignment accuracy. Parentage assignment accuracies were high using either genotyping method, but the panel including microhaplotypes outperformed the SNP panel at high levels (75%) of genotypic missingness. This GT-seq panel had sufficient power for parentage assignment in a genetically depauperate population of brook trout within its native range and may provide a useful tool for future conservation efforts targeting this species.

### **Introduction**

Genotyping-by-sequencing has led to recent advances in conservation genetics by enabling the widespread genomic analysis of non-model organisms at an increasingly low cost. High-coverage genomic data has improved our estimates of inbreeding coefficients in natural populations (Kardos et al. 2016a) and enabled the identification of genomic regions responsible for important life-history variation in species of conservation concern (Prince et al. 2017). However, many analytical methods with conservation relevance are adequately performed with a few hundred genetic markers. These methods include pedigree reconstruction (Huisman 2017), genetic mark-recapture (Mills et al. 2000), and

parentage-based tagging (Steele et al. 2013). Next-generation sequencing technology (NGS) and library preparation techniques such as restriction site-associated sequencing (RADseq) have dramatically reduced the cost and increased the efficiency of identifying thousands of candidate variants for more targeted sequencing (Andrews et al. 2016). Importantly, NGS can accommodate diverse library preparation techniques allowing researchers to design cost-effective genetic marker panels tailored to their research questions.

Genotyping-in-Thousands by sequencing (GT-seq) is a powerful, cost-effective approach to genotype hundreds of amplicons for thousands of individuals on a single high-throughput sequencing lane. Briefly, a GT-seq library is prepared using a large multiplex polymerase chain reaction (PCR) to amplify hundreds of target regions followed by a second PCR step to attach sequencing adapters and barcodes and, finally, pooling thousands of individuals for NGS (Campbell et al. 2015). This approach has been effective in large-scale parentage-based tagging and genetic stock identification for pacific salmon fisheries involving thousands of individuals (Beacham et al. 2018, Hargrove et al. 2021). Each amplicon in a GT-seq genetic marker panel can be used to determine the genotype at one or more single-nucleotide polymorphism (SNP) individually or as a microhaplotype. Compared with an equivalent number of SNPs, microhaplotypes can increase statistical power dramatically by increasing the number of alleles per locus, similar to microsatellite genetic markers (Baetscher et al. 2018). This potential for increased power is only realized in cases where alleles at adjacent loci generate multiple allelic combinations (i.e., microhaplotypes). Importantly for conservation applications, the comparative advantage of microhaplotypes over SNPs will likely be less in populations with small effective population sizes due to lower allelic richness (Allendorf 1986).

Genetic parentage assignment can provide profound ecological and evolutionary insights into natural populations and is made practical due to non-invasive genetic sampling techniques and the increasing availability of powerful genetic marker panels for non-model organisms. Parentage assignment provides the foundational links for pedigree reconstruction and has been used to describe mating strategies, natural and sexual selection, and trait heritability, as well as to estimate individual fitness (Kruuk and Hill

2008) and census size (Bravington et al. 2016).

Estimating individual-fitness in natural populations can be highly valuable for estimating inbreeding depression (Stoffel et al. 2021) or to describe the fitness effects of conspecific or interspecific admixture of divergent individuals (Muhlfeld et al. 2009, Hasselgren et al. 2018). The standard measure of individual fitness is lifetime reproductive success (LRS), which relies on counting the number of recruiting offspring per individual during their lifespan (Clutton-Brock 1988). For long-lived organisms, this requires years of sampling and, likely, thousands of individuals in even modest sized populations (e.g., < 500). In most species and contexts, the only way to build pedigrees is through genetic parentage assignment (Pemberton 2008).

The two primary analytical approaches for genetic parentage assignment are simple exclusion and likelihood-based estimators. Likelihood-based methods have been demonstrated to have greater statistical power than exclusion-based approaches (Anderson and Garza 2006), and allow for the simultaneous implementation of sibship clustering which can dramatically improve parentage assignment accuracy (Wang 2007, 2012). Among the factors known to affect parentage assignment accuracy in simulated and natural populations are unsampled parents, unknown sex of potential parents, genotyping errors, and missing genotypes (Oliehoek and Bijma 2009). Although many parentage assignment error rates become trivial with a few hundred markers (Flanagan and Jones 2019) avoiding them is still worth consideration. Incorrect parentage links within pedigrees has been demonstrated to downwardly bias estimates of inbreeding depression (Reid et al. 2014) and trait heritability (Béréanos et al. 2014), which may lead to erroneous inferences highly relevant to species conservation.

In this paper, we design and optimize a GT-seq genetic marker panel for an ongoing genetic rescue study in brook trout (*Salvelinus fontinalis*) and then evaluate panel performance for parentage assignment using population genetic simulations. Robinson et al. (2017) used eight microsatellite markers to evaluate differences in reproductive success between residents and translocated brook trout. Although Robinson et al. (2017) observed high parentage assignment rates for the relatively genetically diverse translocated individuals, they had low parentage assignment rates for resident parents due to low genetic variation and a limited marker panel. This pattern was most

pronounced in the genetically depauperate population occupying Dry Run, Virginia. Accurate parentage assignments are crucial for accurately describing the influence of introgression on individual fitness and ultimately determining if genetic rescue occurred (Bell et al. 2019).

Here, we design and evaluate a GT-seq panel using both SNPs and microhaplotypes to determine the accuracy of parentage assignments for resident progeny in Dry Run prior to the introduction of migrants. Additionally, we test the effects of unsampled parents, missing genotypes, and the inclusion and exclusion of parental sex on assignment accuracy. Finally, we describe the adequacy of the GT-seq panel to evaluate the outcomes of the genetic rescue attempt in Dry Run.

## **Methods**

### *Sample collection and RADseq library preparation*

We used 238 brook trout tissue samples collected from five streams in Virginia, three in Massachusetts, and one in Maine for the initial variant discovery using restriction site-associated DNA sequencing (RADseq). Detailed descriptions of the sample locations and methods can be obtained for each sampling location in Whiteley et al. (2013), Kazyak et al. (2014), and Whiteley et al. (2015). The tissue samples were extracted following a salt-based tissue digestion and isopropanol precipitation protocol (Aljanabi and Martinez 1997). RADseq libraries were prepared using the restriction enzyme *SbfI* by the Conservation Genomics Laboratory at Michigan State University following the protocol in Ali et al. (2016), with the modifications that 100 ng of DNA was used per individual and the restriction enzyme digestion and ligation reactions were incubated for 12 hours (Mamoozadeh et al., In Revision). The pooled RADseq libraries included 96 individuals per run and were sequenced on two lanes of an Illumina HiSeq 4000 NGS platform for 300 cycles of paired-end sequencing by the Michigan State University Research Technology Support Facility.

### *SNP discovery and filtering*

We received individual FASTQ files from the Conservation Genomics Laboratory that

had been demultiplexed using the ‘process\_radtags’ module in *STACKS* v2.4 (Rochette et al. 2019) with settings to remove low-quality reads and to allow for one barcode mismatch (Mamoozadeh et al., In Revision). At present, there is not a genome assembly available for brook trout and we instead used a chromosome-level genome assembly for a closely related species in the same genus, Dolly Varden (*Salvelinus malma*) as our reference genome (Christensen et al. 2018). We aligned reads to the reference genome using the program *BWA* (Li 2013), specifying the *BWA-MEM* algorithm and using a minimum alignment quality score of 30. We used the program *SAMtools* (Li et al. 2009) to remove unmapped reads, supplementary, and secondary alignments. We then submitting the filtered alignment files to ‘gstacks’ module using standard parameters and exported a VCF file with the ‘populations’ module. We used the option to output consensus sequences for RAD loci in the program *STACKS*.

The primary goal of variant filtering was to identify approximately 800 highly informative target regions for primer design and avoid paralogous regions that are common in organisms with ancestral genome duplication events such as salmonids (Macqueen and Johnston 2014). We began by retaining only single-nucleotide polymorphisms (SNPs) with 50% or fewer missing genotypes and variants with a minor allele frequency (MAF) > 0.05 and a mean read depth > 8 using *VCFtools* v0.1.15 (Danecek et al. 2011). Next, we attempted to filter out paralogous loci using the rationale that these loci would deviate from Hardy-Wienberg proportions (HWP) and have allelic ratios in heterozygotes that deviate from 1:1 more than singleton loci, on average. In the statistical computing program *R* (R Core Team 2017) we calculated average allele balance for heterozygotes for each locus weighted by the total number of reads at that locus for each individual. Using per locus  $F_{IS}$  averaged across populations and the weighted-average heterozygous allele balance per locus we calculated the 2-dimensional Mahalanobis distance of each locus from expected values of 0 and 0.5 for  $F_{IS}$  and allele balance, respectively. We then removed loci with a  $D^2$  values greater than 1.5. After this step was completed, we became aware that McKinney et al. (2017) had written the program *HDplot* and tested the efficacy of a similar approach for identifying paralogous loci, which we recommend for future use. We then used *VCFtools* to identify and remove regions with variant densities exceeding 26 variants per kb in 300 base-pair windows. At

this point, we retained 102 individuals from the Virginia sampling locations for the identification of highly informative GT-seq target amplicons to be applied to the genetic rescue study.

We identified candidate microhaplotype markers using custom *R* scripts, thinned candidates so that they were equally spaced through the reference genome, and did not significantly deviate from HWP, or were not in significant linkage disequilibrium with other candidate microhaplotypes. We first used *VCFtools* to remove sites that had minimum allele frequencies  $< 0.05$  or were ungenotyped in more than 30% of the individuals considering all Virginia populations. Using custom *R* scripts, we identified regions that had two or more adjacent SNPs within 150 base pairs. We then removed candidate microhaplotypes that had mean MAF less than 0.1 considering all adjacent SNPs within the potential microhaplotype. Given that primer design was our goal, we removed candidate microhaplotypes whose corresponding consensus sequences contained more than 5% of ambiguous and target SNPs within 30 bp from the beginning or end of the consensus sequence. We ensured that each candidate microhaplotype was not fixed in any population (i.e.,  $MAF > 0$ ). We then selected candidate microhaplotypes so they were at least 1 Mb from other candidates on the same chromosome or contiguous sequence. Finally, we iteratively removed candidate microhaplotypes that had linkage disequilibrium values ( $R^2$ ) that exceeded 95% distribution of interchromosomal  $R^2$  values among candidate microhaplotypes until all candidates fell within this distribution. In addition to the set of candidate microhaplotypes, we also identified a set of single SNPs for primer design sampling regions at least 5 Mb away from candidate microhaplotypes and repeated the filter steps described above.

#### *GT-seq primer design, panel optimization, and genotyping*

We designed primers for candidate targets and optimized the panel by iteratively removing primer pairs that failed to amplify, produced off-target amplification, or produced a disproportionately high number of amplicons. We used BatchPrimer3 using consensus sequences provided by *STACKS* to design generic primers and used default settings with the following modifications. We specified product sizes of 80–160 bp with an optimum of 115, primer sequence lengths of 15–25 with an optimum of 20, a primer

annealing temperature of 55–60 °C with an optimum of 57 °C, and a GC content of 30–70. We screened the primer sequences by aligning them to the reference genome with the program *BLAST* v2.2.31+ (Camacho et al. 2009) and removed primers that had multiple high-quality alignments. Following Campbell et al. (2015), we then added Illumina primer sequence tags to the to the forward and reverse primers to allow for the addition of dual index barcodes. In addition to the primer pairs designed here, we also included on the panel a sex marker targeting the *Sdy* sex determining region in salmonids (Yano et al. 2013). Primers were obtained from Integrated DNA Technologies Inc. and we prepared GT-seq libraries following the protocol provided by Campbell et al. (2015).

Our initial GT-seq optimization libraries were processed by the University of Montana Genomics Core on an Illumina MiSeq for 300 cycles of paired-end sequencing. After receiving demultiplexed FASTQ files, we used the program *Trimmomatic* v0.39 (Bolger et al. 2014) to remove Illumina adapter sequences in paired-end and palindrome mode and with a minimum sequence length of 35 bp after trimming. We used the program *FLASH* v1.2.11 (Magoc and Salzberg 2011) to join paired-end sequences to generate contiguous sequences of each amplicon. We then used a custom script in Python v3.7.6 to count the number forward and reverse primer sequences present in each read. Sequencing reads were subsequently aligned to the consensus sequences used to design primers using *BWA-MEM*. We calculated the off-target proportion by dividing the forward primer counts by the number of correct alignments. We removed primer pairs from future optimization runs by removing the top 10% represented in primer counts and those with high levels of off-target amplification. We subsequently called variants using *STACKS* and exported a VCF as described above. We ensured that each amplicon indeed possessed a callable variant by using the program ‘microhaplot’ (Ng 2019). We repeated this process for four optimization runs. On the fifth optimization run, after achieving a desirable read count distribution among amplicons and on-target percentage, we designed allele-specific probes for count-based genotyping, similar to that described by Campbell et al. (2015).

#### *Testing the GT-seq panel performance*

We genotyped individuals from the most genetically depauperate population sampled as

part of a genetic rescue study on brook trout (Robinson et al. 2017) to test the adequacy of this panel to study the population into future generations. We genotyped 74 individuals from Dry Run, Virginia, sampled in 2011, prior to the introduction of immigrants described in Robinson et al. (2017). We calculated observed heterozygosity,  $F_{IS}$ , and estimated allele frequencies using the *R* package ‘hierfstat’ (Goudet 2005) for both microhaplotypes and SNP genotypes from the newly designed GT-seq panel. The probability of identity ( $P_{ID}$ ) and the probability of sibling identity ( $P_{SIB}$ ) were also calculated following Waits et al. (2001) using the *R* package ‘PopGenUtils’ (Tourvas 2022). The 40 translocated individuals used in the Robinson et al. (2017) genetic rescue study were of known sex and we genotyped them on the panel to evaluate the accuracy of the *Sdy*-derived sex diagnostic marker. We also evaluated genotyping error rate for 96 individuals collected from Dry Run in 2014 that we sequenced twice on two separate sequencing runs.

We used a population genetic simulation to evaluate the accuracy of parentage assignment on this marker panel using estimated allele frequencies for SNPs and microhaplotypes in 2011 samples from Dry Run. Using the estimated allele frequencies, we simulated 250 founding individuals with a sex ratio of 1:1. Each generation’s parents were drawn at random and had a family size drawn from a Poisson distribution ( $\lambda = 5$ ). We selected a mean full-sibling family size of 5 because the average family size in Dry Run in the years 2010–2012 was 5.76 (Robinson et al. 2017). We continued to generate families until we achieved 250 progeny. If the final family generated caused the progeny pool to exceed 250 individuals, we truncated the last family drawn. Our simulation was parameterized with discrete generations and constant population size; therefore, each individual had the opportunity to contribute gametes to one generation of 250 progeny. Every progeny’s genotype was generated by drawing gametes from its parents randomly and independently of other full siblings. The sex of progeny was determined by a Bernoulli trial ( $P = 0.5$ ). This process continued for 15 discrete generations. We discarded the first five generations and retained the last 10 generations for parentage analysis. We ran this simulation parameterized with allele frequencies from both the microhaplotype and SNP genotyping methods.

Parentage assignment accuracy is known to be affected by unsampled parents,

missing genotypes, and unknown sex of parents. We tested the effect of genotypic missingness by censoring genotypes in the simulation output randomly, in both parents and progeny, at a rate of 0.05, 0.10, 0.25, 0.5, and 0.75. To test the effect of unsampled parents, we randomly selected 0.05, 0.10, 0.25, 0.5, and 0.75 of parents to not be included in the parentage assignment program. Lastly, we evaluated the effect of including or excluding sex information for parents. We used *Colony* V2.0.6.8 (Jones and Wang 2010) for parentage assignment and submitted each generation of the simulation as an independent run. This allowed for each scenario of unsampled parents, missing genotypes, and unknown sex to have 10 replicate generations of parentage analysis. *Colony* was ran using the full-likelihood and pair-wise likelihood combined method specifying high-precision and medium run length. The run was conducted without updating or specifying allele frequencies and without the presence of inbreeding. We specified polygamy and diploidy. We provided *Colony* with mean genotyping error rates estimated for each of the genotyping methods.

We categorized two types of parentage assignment accuracy—total accuracy and assignment accuracy—and the over-assignment proportion. Total accuracy represents the proportion of correct parentage assignments and the number of unassigned parents given they were available to assign. Assignment accuracy is the proportion of total parental assignments that were correct without respect to unassigned parents. The over-assignment proportion is the proportion of falsely assigned parents when the true parents were unavailable for assignment. We calculated confidence intervals for total accuracy, assignment accuracy, and over-assignment proportion using 1000 bootstrapped samples of the ten generations of parentage assignment.

## **Results**

### *Identification of target variants, primer design and optimization.*

We used *SbfI* digested RAD sequence data to identify candidate microhaplotypes and SNPs for inclusion on a GT-seq panel. After preliminary filtering on MAF, mean read depth, and genotypic missingness, we retained 46,399 variants for consideration. After filtering putative paralogous loci, we retained 33,679 variant sites (Figure 4-1). After

selecting the individuals from the Virginia populations included in the genetic rescue study and removing variants based on MAF and genotypic missingness, we retained 18,968 variant sites. We then identified 3904 regions that contained two or more SNPs within 150 bp as candidate microhaplotype markers. After filtering based on the suitability of the consensus sequence for primer design, physical distance > 1 Mb, HWP, and linkage disequilibrium we retained 542 for designing primers. We additionally selected 58 SNPs that were located at least 5 Mb from any candidate microhaplotype and were in HWP and did not exhibit significant linkage disequilibrium with other included candidate markers.

We successfully designed primers pairs for 387 candidate targets and an additional 5 primer pairs including the sex marker targeting the *Sdy* region from an existing GT-seq panel (M. Campbell, Idaho Department of Fish and Game, personal communication). The initial set of candidate primers used in the optimization libraries included 316 targets with two or more SNPs (i.e., potential microhaplotypes) and 76 single SNP targets (Figure 4-2A). Following five rounds of library preparation and sequencing, while dropping primer pairs that were overrepresented or produced a high percentage of off-target amplicons, we retained 219 markers that represented 176 potential microhaplotypes and 43 SNPs (Figure 4-2B). After designing allelic probes for *in silico* count-based genotyping, we retained 166 of the potential microhaplotypes, and the remaining 53 markers were genotyped as SNPs. The ten potential microhaplotypes that were genotyped as SNPs had insertion or deletion variation in the sequence between adjacent SNPs, which impaired allelic probe design. We genotyped each individual twice by either genotyping only one SNP per amplicon with the highest MAF or calling microhaplotypes. We hereafter refer to these as the SNP and microhaplotype panels, respectively. The optimized library pool on which the brook trout sampled from Dry Run in 2011 were sequenced and had 74.4% of total raw reads on-target following demultiplexing. After trimming and flashing paired-end reads, we observed an on-target percentage of 88.4%.

#### *GT-seq panel genetic summary*

We observed higher levels of genetic variation when calling genotypes with the

microhaplotype panel versus the SNP panel. Mean observed heterozygosity for the SNP panel was 0.287 and 0.379 for the microhaplotype panel. Mean  $F_{IS}$  considering all loci was the same for both panels ( $F_{IS} = 0.043$ ). The mean number of alleles per locus was 1.81 and 2.21 for the SNP and microhaplotype panel, respectively (Figure 4-3). Estimated  $P_{ID}$  and  $P_{SIB}$  suggest that both panels are adequate for genetic mark-recapture. Considering all loci,  $P_{ID}$  was  $3.95 \times 10^{-88}$  for the SNP panel and  $3.52 \times 10^{-126}$  for the microhaplotype panel.  $P_{SIB}$  was  $1.39 \times 10^{-42}$  and  $2.13 \times 10^{-56}$  for the SNP and microhaplotype panel, respectively. Individual genotyping success was 89.8% and 89.6% for the SNP and microhaplotype panel, respectively. We sequenced 96 individuals twice to estimate genotypic mismatch rate, which was low for both panels with 0.083% and 0.096% for SNPs and microhaplotypes, respectively. The *Sdy*-derived sex marker had an accuracy rate of 92.7% based on 40 individuals of field-observed sex. The three errors in genotypic sex calls included two field-observed males and one female.

#### *Parentage Assignment Accuracy*

High parentage assignment accuracy rates were observed for both genotyping methods and assignment errors only occurred at high rates of genotypic missingness. We observed no assignment errors for any proportion of unsampled parents and as a result the over-assignment proportion was zero for all scenarios considered. When parental sex information was not provided to the program *Colony*, assignment accuracy was 100% and total accuracy was 97.3%, on average, and did not differ significantly between genotyping methods (Table 4-1). The difference between total accuracy and assignment accuracy when parental sex was not provided was due to approximately 2.7% of parents being unassigned by *Colony* although no incorrect assignments were made. Parentage assignment errors were not observed for either genotyping method when the proportion of missing genotypes was 0.05, 0.10, or 0.25. However, parentage assignment errors did occur at proportions of missing genotypes of 0.5 and 0.75 for both genotyping methods, but parentage assignment accuracies were significantly higher for the microhaplotype genetic marker panel (Figure 4-4).

## Discussion

We designed and optimized a GT-seq genetic marker panel for the purpose of obtaining accurate parentage assignment in an ongoing genetic rescue study in brook trout. Based upon the parentage assignment performance using population genetic simulations, the genetic marker panel designed herein provides sufficient accuracy for our purpose. Whiteley et al. (2012) reports mean full-sibling reconstruction accuracy of 91.2% for the eight microsatellite markers originally used in the genetic rescue study, based upon simulations parameterized with allele frequencies from the most genetically diverse populations in the study. In contrast, we parameterized a population genetic simulation using allele frequencies from the most genetically depauperate population in the genetic rescue study and achieved 100% parentage accuracy, which is equivalent to 100% full-sibling reconstruction accuracy, excluding scenarios with 50% missing genotypes or greater. The GT-seq genetic marker panel that we provide is a precise and cost-effective tool for genotyping thousands of individuals for future generations of this genetic rescue study.

Molecular resources such as the GT-seq panel that we provide can improve estimates of population genetic parameters and contribute to conservation decision making (Bernos et al. 2020). Eastern brook trout have experienced large-scale population declines and extirpation, and presently occupy a fraction of their historical range (Hudy et al. 2008, Merriam et al. 2019). Climate change and human land-use changes have resulted in widespread reduction of suitable habitat in river mainstems and is expected to restrict many salmonid populations to thermal refugia in headwaters (Isaak et al. 2016). As a result, small salmonid populations occupying restricted headwater habitats are vulnerable to genetic drift and inbreeding (Gilpin and Soule 1986) and environmental and demographic stochasticity (Lande 1993). The GT-seq panel we provide may be well-suited to address many conservation-relevant questions in eastern brook trout and strike a balance between statistical power and cost. However, ascertainment bias in genetic marker panels has been shown to lead to erroneous inferences of population structure (Dokan et al. 2021) and estimates of genetic variation (Geibel et al. 2021). Future work should address the potential ascertainment bias of our GT-seq panel given that it was tailored for use within five brook trout populations in the mid-Atlantic region.

We observed high parentage assignment accuracies for both SNP and microhaplotype genotyping methods. However, the microhaplotype panel outperformed the SNP panel at extremely high rates of genotypic missingness. This is a promising result for two reasons: 1) Following guidelines from Huisman (2017), we will not perform parentage analysis for individuals with more than 50% missing genotypes, 2) Genotypes were censored at random in the simulation and in practice missing genotypes occur in certain loci more than others, resulting in more genotypes in common among parents and progeny. Additionally, we expect the attempted genetic rescue in Dry Run to introduce new microhaplotype alleles that increase the comparative advantage of microhaplotypes relative to SNPs. In contrast to our results, Baetscher et al. (2018) found that microhaplotypes improved relationship inference by orders of magnitude compared to SNPs. However, they performed their analysis using allele frequencies from a marine fish with large effective population size and detected up to 12 distinct microhaplotype alleles per locus. In comparison, we discovered a maximum of four microhaplotype alleles per locus, which we attribute the small effective size in Dry Run (Allendorf 1986, Whiteley et al. 2013).

Parentage assignment accuracy was unaffected by the absence of true parents from the candidate parent pool and unknown parental sex reduced the number of parental assignments without inducing assignment errors. Although parentage analysis with unsampled parents can induce errors where sampled parents are erroneously assigned when the true parent is absent (Wilson and Ferguson 2002), this is uncommon when there are a few hundred genetic markers (*reviewed in* Flanagan & Jones, 2019). Our GT-seq panel is consistent with this expectation. The 219 genetic markers appear to be robust to missing parents in the absence of other factors such as missing genotypes.

Parental sex information is known to improve the accuracy and assignment rates of parentage analysis programs (Wang and Santure 2009), making a sex marker a valuable addition to our GT-seq panel. Unfortunately, the sex marker that we selected exhibited an undesirable error rate. However, we found that not including sex information did not result in falsely assigning parents and sex-naïve parentage assignment may prove useful in refining the sex calls in our genetic rescue study.

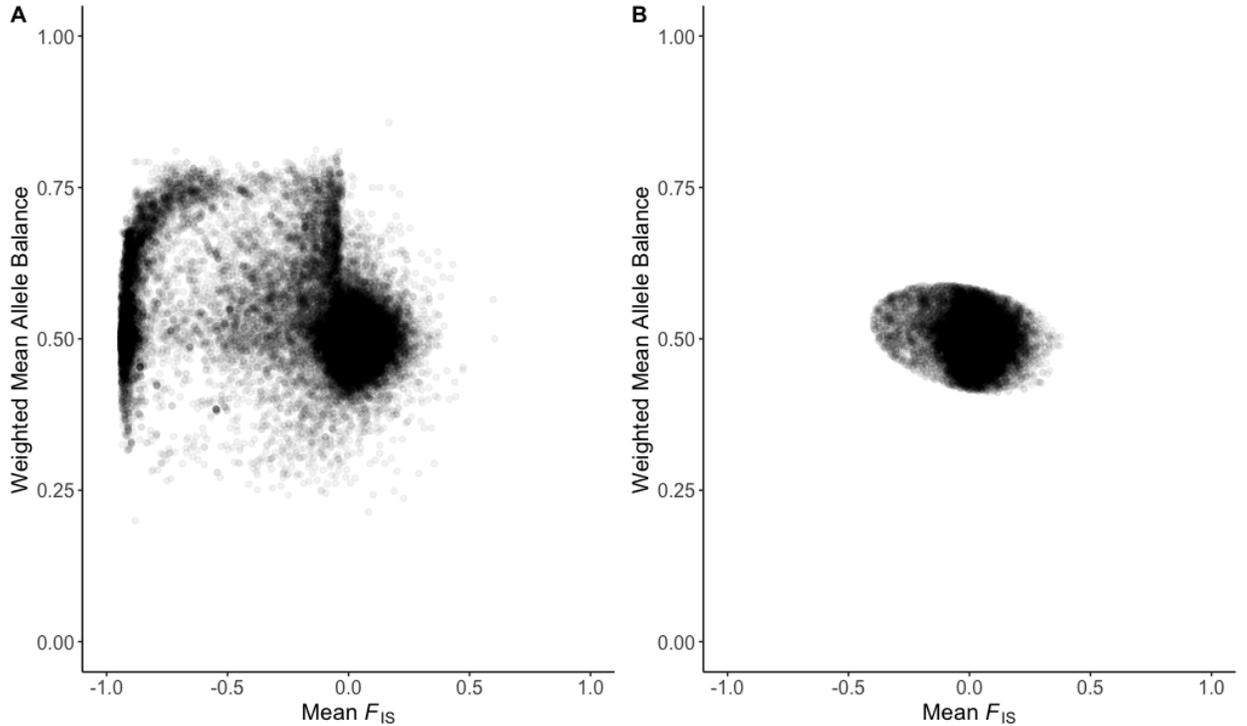
We parameterized our parentage simulation to have full-sibling family structure,

similar to that observed in stream-dwelling populations of brook trout, which likely increased our estimated parentage assignment accuracy. Brook trout in this region have been observed to have considerable full-sibling family structure (Hudy et al. 2010) and it has been observed that over 80% of successful parents contributed to a single family (Coombs 2010). Although realistic, our simulations were unlikely to produce many half-sibling relationships, which would have posed a greater challenge for parentage assignment (Jones et al. 2010). The simultaneous implementation of sibship clustering and parentage assignment has been shown to increase accuracy in pedigree reconstruction, and this effect is most pronounced in the presence of large full-sibling families (Wang 2007, 2012). As a result of our simulation parameterization, the estimated parentage assignment accuracies for this GT-seq panel are most informative for intensively sampled brook trout populations with many large full-sibling families.

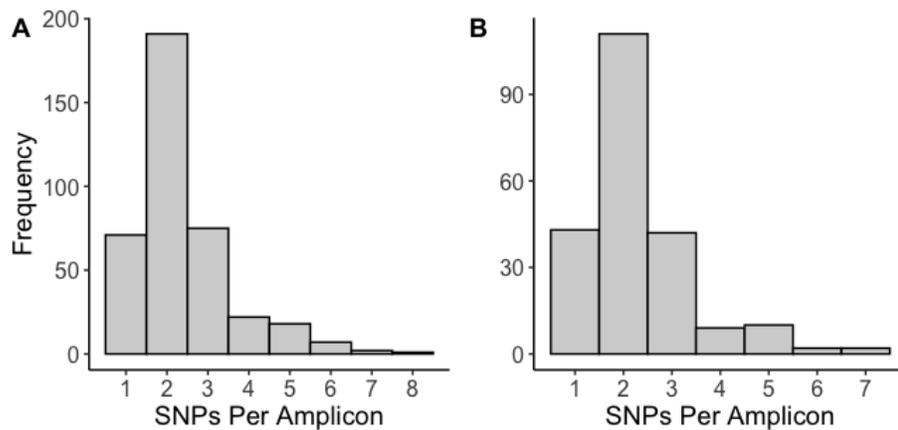
The efficiency and flexibility of genotyping-by-sequencing has led to rapid advances in the field of conservation genetics. The improvement in our ability to answer conservation-relevant questions with these methods is well demonstrated by our parentage assignment results. Parentage assignment provides the fundamental links in pedigrees and enables the study of ecological and evolutionary aspects of wild populations (Pemberton 2008). The GT-seq panel that we designed is sufficiently powerful to confidently conduct parentage analysis in an ongoing genetic rescue study and may provide a useful tool for the conservation of eastern brook trout moving forward.

**Table 4-1. Total and assignment parentage assignment accuracies for the twelve simulation scenarios evaluated for the SNP and microhaplotype genotyping methods (MHAP).** Over-assignment proportion was zero for all scenarios considered and omitted from the table. Bootstrapped confidence intervals are provided when errors were documented in any of the ten replicates of parentage analysis.

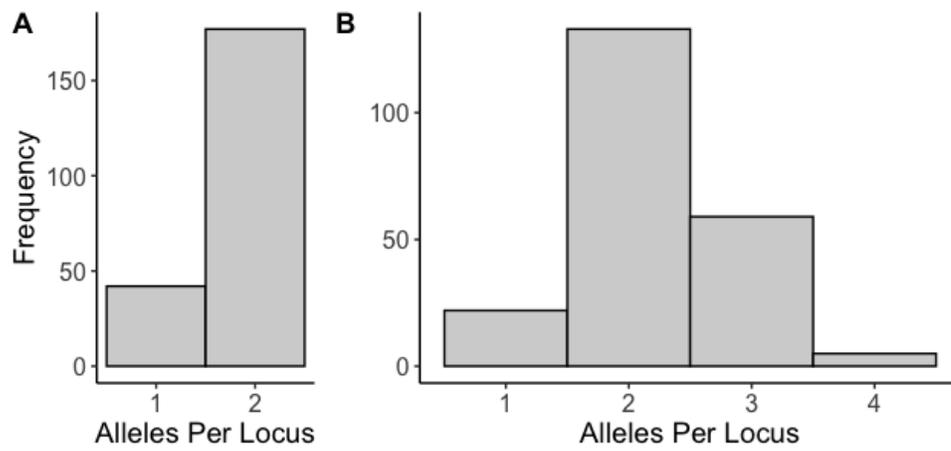
		Total Accuracy		Assignment Accuracy	
		SNP	MHAP	SNP	MHAP
Full Simulation		100%	100%	100%	100%
Without Sex Information		97.0% (95.7, 98.3)	97.6% (96.7, 98.6)	100%	100%
Proportion of Unsampled Parents	0.05	100%	100%	100%	100%
	0.1	100%	100%	100%	100%
	0.25	100%	100%	100%	100%
	0.5	100%	100%	100%	100%
	0.75	100%	100%	100%	100%
Proportion of Missing Genotypes	0.05	100%	100%	100%	100%
	0.1	100%	100%	100%	100%
	0.25	100%	100%	100%	100%
	0.5	99.6% (99.5, 99.8)	99.9% (99.8, 100)	99.7% (99.5, 99.8)	99.9% (99.8, 100)
	0.75	42.7% (37.3, 47.7)	75.9% (72.7, 79.6)	66.0% (62.6, 69.2)	86.8% (84.9, 88.9)



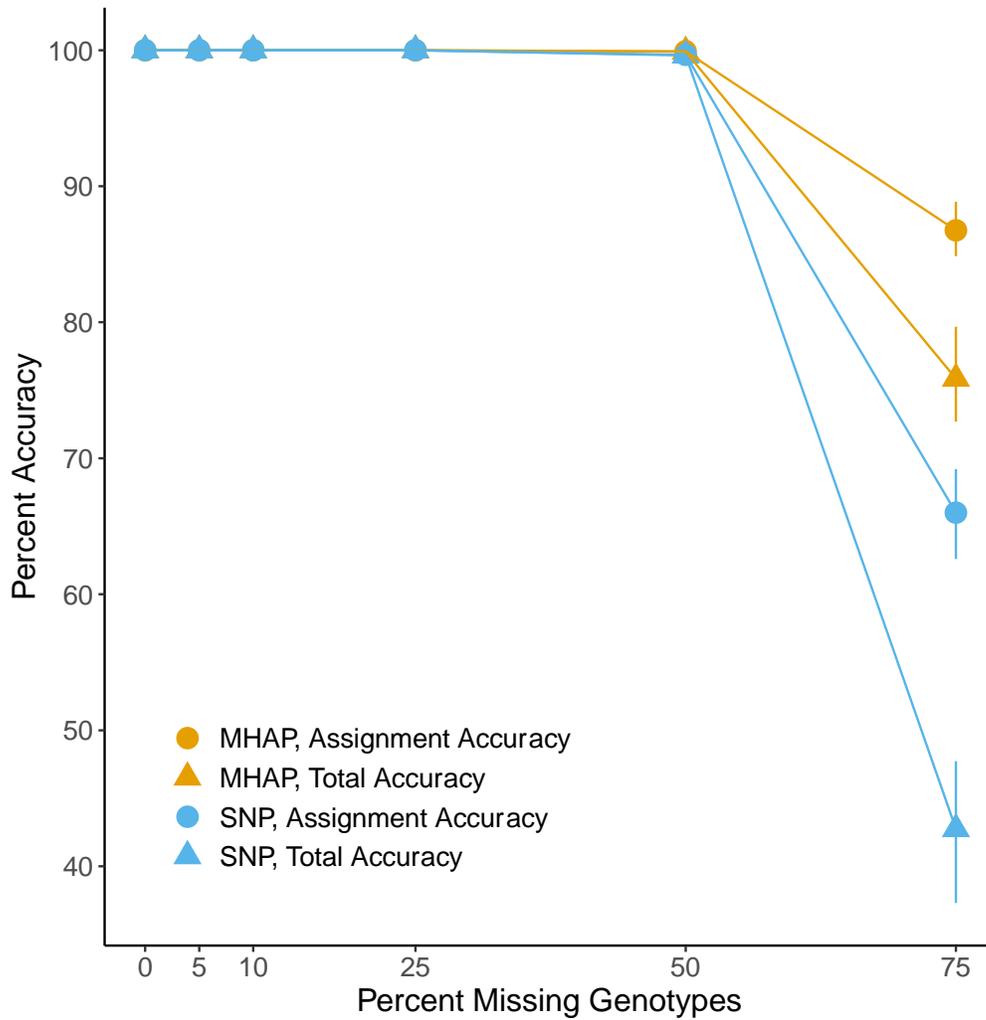
**Figure 4-1. Scatterplot of per-locus mean  $F_{IS}$  and weighted mean allele balance for heterozygous genotypes before filtering (panel A) and after filtering putative paralogous loci based on Mahalanobis distance (panel B).**



**Figure 4-2. Histogram representing the number of single nucleotide polymorphism per amplicon before (Panel A) and after optimization (Panel B) of the GT-seq panel.**



**Figure 4-3. Histogram of allele counts per locus detected in Dry Run, Virginia in 2011. Allele counts are provided for the SNP (Panel A) and microhaplotype (Panel B) genotyping methods.**



**Figure 4-4. The effect of missing genotypic data on parentage assignment accuracy.** Total accuracy and assignment accuracy are reported for both the single nucleotide polymorphism (SNP; blue) and microhaplotype (MHAP; orange) genotyping methods.

## **CHAPTER 5: Evaluating the multigenerational effects of gene flow in an isolated population of brook trout**

### **Abstract:**

Conservation practitioners are increasingly faced with the challenge of managing numerous, small populations in demographic and genetic isolation. Small, isolated populations experience an increased risk of extirpation due to genetic drift and inbreeding. Human-mediated gene flow can alleviate genetic load and result in genetic rescue. Here, we evaluate the multigenerational effects of a genetic rescue attempt in a small, isolated (~26 generations) population of brook trout intensively sampled from 2010 to 2018. Ten Individuals (5 of each sex) were introduced from an adjacent watershed immediately before reproduction in 2011. We estimated lifetime reproductive success (LRS) based on a reconstructed pedigree and found that  $F_1$  hybrids were 2.23 times more likely to successfully contribute progeny and produced 2.26 more offspring on average compared to resident individuals. Additionally,  $F_1$  hybrids had significantly higher juvenile survival compared to residents. However, we found a consistent negative relationship between migrant ancestry and survival and LRS in hybridized individuals of the  $F_2$  and later generations. Still, backcrossed individuals with migrant ancestry less than 0.48 had higher LRS than residents, on average. Gene flow resulted in a 7.60% increase in heterozygosity, and mean population-level migrant ancestry was 0.31, indicating a retention of local ancestry. These results provide an empirical demonstration that a pulse of gene flow can have varied effects on individual fitness that are not revealed until later generations.

### **Introduction**

Habitat fragmentation is a global threat to biodiversity worldwide (Haddad et al. 2015), and has resulted in demographic and genetic isolation of numerous populations (Frankham et al. 2017). Isolated populations occupying small habitat fragments are vulnerable to demographic stochasticity and experience high levels of genetic drift (Lande 1993, Jamieson and Allendorf 2012). Over time, small populations can experience declines in fitness associated with accumulation of genetic load and

inbreeding depression (Willi et al. 2006). Importantly, these deleterious effects have been shown to reduce population growth rate (Bozzuto et al. 2019) and increase extinction risk (Saccheri et al. 1998, O'Grady et al. 2006). Fortunately, human-mediated gene flow can induce genetic rescue, which is an increase in population persistence probability due to beneficial effects of gene flow (Bell et al. 2019). The potential for natural or human-mediated gene flow to alleviate genetic load and increase individual fitness has been widely documented in a variety of taxa (Madsen et al. 1999, Hogg et al. 2006, Johnson et al. 2010, Hasselgren et al. 2018).

Many species of conservation concern are widely distributed but comprised of many disconnected populations in small or low-quality habitats. Though these species may not meet conservation listing requirements, each population could have high vulnerability to extirpation. These circumstances are common among species with low dispersal and those dependent on threatened environments (Cushman 2006, Brauer and Beheregaray 2020). As a result, there have been calls for more widespread application of human-mediated gene flow to mitigate the negative genetic effects of small population size and isolation (Ralls et al. 2018, Chan et al. 2019). Although genetic rescue attempts are primarily focused on improving near-term persistence probability, it is worth noting that human-mediated gene flow can accomplish other important conservation objectives. For example, human-mediated gene flow can restore historical population connectivity and genetic variation (Hedrick 2005), and potentially facilitate adaptation to future climatic conditions (Kelly and Phillips 2019).

Evolutionary theory predicts that populations will adapt to local environmental conditions and that gene flow among populations will be regulated by migration, natural selection, and local genetic load (Whitlock et al. 2000). In fact, instances of reduced individual fitness due to increased gene flow have been widely documented (Fenster and Galloway 2000, Mobley et al. 2019). This naturally raises the concern that genetic rescue attempts may instead reduce population persistence probability through a reduction of individual fitness in hybridized individuals, termed outbreeding depression (Edmands 2007). In addition to disrupting local adaptation, outbreeding depression can be caused by epistatic gene interactions, co-adapted gene complexes, and other genomic incompatibilities (Whiteley et al. 2015b). However, there are few documented cases of

outbreeding depression having appreciable effect on population persistence (Bell et al. 2019), and there are existing frameworks for predicting the probability of outbreeding depression for potential genetic rescue attempts (Frankham et al. 2011, 2017).

To address concerns of outbreeding depression, the multigenerational dynamics of genetic rescue attempts need to be more thoroughly explored (Tallmon 2017). The majority of studies of genetic rescue evaluate a few reproductive bouts when the fitness benefit should be maximal due to heterosis (Bell et al. 2019). One of the few multigenerational, experimental tests of genetic rescue found that gene flow from an adaptively divergent source population increased abundance and that hybridized individuals had elevated fitness in replicate, wild populations of Trinidadian guppies (Fitzpatrick et al. 2016). Importantly, this study showed that natural selection was sufficient to resist introgression of maladaptive alleles, while benefiting from the alleviation of genetic load. However, more work is needed to determine how well these results generalize to other taxa and ecological contexts.

In this chapter, we examine the multigenerational effects of a genetic rescue attempt in a small, isolated population of brook trout that occurred in 2011. Robinson et al. (2017) presented the initial results in the first filial ( $F_1$ ) generation produced in 2012 and found that translocated individuals had significantly higher reproductive success and that their progeny had significantly larger body sizes compared to residents. Using data collected from 2010 to 2018, we reconstructed the pedigree for this population, estimated lifetime reproductive success, and survival to characterize the effects of admixture on individual fitness beyond the  $F_2$  generation. Brook trout populations in the mid-Atlantic region are often isolated in small headwater streams (Hudy et al. 2008), and experience high levels of genetic drift (Kazyak et al. 2022). Our findings will aid in the conservation and management of stream-dwelling brook trout, a taxon of general conservation concern, and of high ecological, social, and economic value.

## **Methods**

### *Study System and Sampling*

This genetic rescue study was conducted in a small, isolated, stream-dwelling population

of brook trout that occupies Dry Run, Virginia within the Chesapeake Bay drainage basin, USA (Figure 5-1). Dry Run is a 8.06 km-long tributary of Dry River immediately upstream of a flood control reservoir constructed in 1968 in the George Washington National Forest (Whiteley et al. 2013). This study system is in a heavily forested watershed with a seasonally intermittent flow-regime, which restricts the brook trout population to residual pools during mid-June to late-August annually (Courtwright and May 2013). The brook trout population was exhaustively sampled each year from 2010–2018 in July and August using one-pass electrofishing surveys in 50-meter sections. Brook trout were never detected further than 3.85 km from the inflow of Dry Run reservoir. Upon capture, we recorded individual total length (nearest mm) and location (corresponding 50-meter section), and a tissue sample was taken as a source of genetic material and to serve as a mark for mark-recapture. In total, 6084 brook trout were captured during the nine years of sampling, of which, 4368 were determined to be age-0 and 1716 were age-1 or older.

Sampling during late-summer allowed age-0 brook trout to become large enough to be captured efficiently while still enabling year-class differentiation based upon length (Hudy et al. 2000). Empirical length frequency histograms, constructed for each sample year, were highly bimodal as expected, and age-0 fish were easily distinguished from age-1 and older individuals, hereafter adults. During the years 2012 to 2017, Dry Run was resampled approximately two weeks after the one-pass electrofishing sampling event to determine the proportion of marked and unmarked age-0 and adult brook trout. For years with a second sampling event, abundance was estimated for both age groups using the Lincoln-Peterson estimator (Otis et al. 1978). For the years without a second sampling event, we estimated abundance and confidence intervals by dividing the number of marked individuals by 1000 bootstrapped samples of observed capture probabilities corresponding to each age group.

### *Brook Trout Translocation*

Previous work in this system demonstrated that the brook trout population in Dry Run exhibited strong initial indications of genetic rescue following the translocation of individuals from a nearby, genetically diverse population (Robinson et al. 2017). Dry

Run was selected for a genetic rescue attempt because of evidence of substantial genetic drift and likely inbreeding. For example, microsatellite-based heterozygosity was 49.5% lower than mainstem Dry River, with an estimated effective number of breeders of 4.9 (3.8–8.7) in 2010 (Whiteley et al. 2013). In October of 2011, ten adult brook trout (five of each sex) were introduced from Dry River immediately downstream from the flood control dam that isolates Dry Run (Figure 5-1). Ten resident adults (five of each sex) from Dry Run were subsequently removed to control for the demographic contribution of immigrants and maintain the adult population size of Dry Run. The translocation coincided with the earliest expected onset of spawning in brook trout and improved the chances of successful reproduction of translocated individuals (Jenkins and Burkhead 1994). At the time of translocation, a reliable genetic sex marker was unavailable (A. Whiteley, unpublished data), and the sex of translocated and removed fish was determined based on physiological differences (e.g., expressing milt, gravid females with swollen abdomen, and head shape). Robinson et al. (2017) tested for and found no evidence of body size differences between the fish introduced and the resident adult population in Dry Run. The translocated individuals had dramatically high reproductive success in Dry Run as demonstrated by their progeny comprising more than 50% of the 2012 cohort.  $F_1$  hybrid progeny (defined as individuals produced by admixture between resident and translocated parents) had significantly larger body sizes and were members of larger full-sibling families on average, which was concluded to be consistent with successful genetic rescue (Robinson et al. 2017).

### *Genetic Analysis*

All captured individuals from Dry Run and 40 individuals from Dry River, including the ten translocated individuals, were genotyped using a targeted amplicon sequencing approach known as GT-seq (Campbell et al. 2015). The GT-seq genetic marker panel used was designed specifically for pedigree reconstruction in this population and included 219 amplicons each containing one or more single-nucleotide polymorphisms (SNPs). All individuals were genotyped using two methods described in detail in Chapter 4. For the first method, we genotyped a single SNP per amplicon and for the second method, we genotyped multiple SNPs together as a microhaplotype (e.g., Baetscher et al.

2018). In this chapter, we used the microhaplotype panel only for identifying unique individuals and building capture histories (described below). For all other analyses and genetic summary statistics we used SNP genotypes. In addition to polymorphic markers, the GT-seq panel includes a sex marker targeting the *Sdy* sex determining region shared among many salmonid species (Yano et al. 2013). Previous work has shown that this sex marker is 92.7% accurate based on 40 individual brook trout of field called sex from Dry River (see Chapter 4). We retained loci and individuals with at least a 50% genotyping success rates, which is consistent with existing guidelines for parentage analysis (Huisman 2017) and the findings in Chapter 4.

Given the demographic and genetic conditions within Dry Run, we predicted that Dry Run would show strong deviations from Hardy-Weinberg proportions (HWP), and loci would show significant linkage disequilibrium (LD) for several reasons. First, in the years following the introduction of immigrants, we observed significant excess of heterozygotes (i.e., negative  $F_{IS}$ ; Robinson et al. 2017) and genetic admixture is known to cause deviations from HWP and LD (Slate and Pemberton 2007). Additionally, Dry Run has strong full-sibling family structure and appears to have experienced a population bottleneck in 2011 (Whiteley et al. 2013), both conditions can cause significant LD and deviations from HWP (Rodriguez-Ramilo and Wang 2012). As a result, we did not perform traditional hypothesis testing for HWP or LD, and instead removed loci that produced significant deviations from HWP in multiple sample years and had a consistent positive or negative  $F_{IS}$ , which could indicate a null alleles or paralogous loci, respectively. We calculated locus-specific  $F_{IS}$  and tested for deviations from HWP with the program *GENEPOP* v4.7.5 (Rousset 2008). To summarize the effect of the translocation on genetic variation we calculated heterozygosity for Dry Run each year and Dry River from a 2011 sample of 40 individuals. We also calculated Nei's unbiased estimator of  $F_{ST}$  (Nei 1987) between Dry River and Dry Run. We provide confidence intervals for heterozygosity by performing 1000 bootstraps across individuals. All statistical analyses in this paper were performed in the statistical computing program *R* (R Core Team 2020), and heterozygosity and  $F_{ST}$  were estimated using the *R* package 'hierfstat' (Goudet 2014).

### *Genetic mark-recapture and pedigree reconstruction*

Genetic recaptures were determined using the number of genotypic mismatches between all pairwise comparisons of genotyped individuals. The microhaplotype genetic marker panel exhibited low  $P_{ID}$  and  $P_{SIB}$  (Waits et al. 2001) values in this population, which indicates that genetic mark-recapture was a valid approach unlikely to produce spurious recaptures (see Chapter 4). A threshold of two genotypic mismatches was used to identify a recaptured individual, which was determined based on genotyping error rate and histograms of genotypic mismatches (figure S3-1). After identifying all unique individuals, we performed parentage assignment and sibship clustering in the program *Sequoia* v2.20 (Huisman 2017), which reconstructs multigenerational pedigrees using age and basic life-history information. We specified complex, overlapping generations, an empirically derived genotyping error rate of 0.086%, a max sibship size of 400, and max sibship clustering iterations of 60. We specified that *Sequoia* use age information and provided a flat age prior with a max age of seven. In addition, we provided individual genotypic sex calls when available and birth year when an individual was captured at age-0. We estimated birth year for individuals not captured initially at age-0 by providing a range of birth years conservatively estimated based on knowledge of the system and the individual's length at first capture. Individuals first captured with body size less than 150 mm were specified as one or two years old, less than 200 mm as one to three years old, less 250 mm as one to five years old, and greater than 250 mm as two to six years old.

Pedigree reconstruction results were used to estimate lifetime reproductive success and to assign individuals into ancestry classes for modeling the effects of admixture on individual fitness or fitness-related traits. Following pedigree reconstruction, we estimated lifetime reproductive success (LRS) by summing the total number of progeny assigned to each individual over the sample period (Clutton-Brock 1988). We used the *R* package 'kinship2' (Sinnwell et al. 2014) to calculate pedigree-based migrant ancestry for each individual. Briefly, migrant ancestry is an estimate of the proportion of an individual's genome that is identical by descent to the translocated individuals (i.e., migrants). We hypothesized that first filial ( $F_1$ ) hybrids produced by non-hybridized individuals from Dry Run and Dry River would exhibit distinctive characteristics due to hybrid vigor. Additionally, we predicted that progeny produced by

two parents of common origin would also have distinctive qualities relative to progeny produced by backcrossing, which allows for recombination between the genomes of Dry Run and Dry River origin individuals. As a result, we adopted a modelling strategy that separately estimated parameters for individuals in each ancestry class (described below). We used the following nomenclature for ancestry classes which defined an individual based on whether they were the progeny of (1) Dry Run resident-by-resident (RR) parents, (2) Resident-by-transplant (RT) parents, (3) Dry River transplant-by-transplant (TT) parents, and 4) Backcrosses (BC), were defined as any mating involving a hybridized individual. Note that these classifications were made based on parental identity and not migrant ancestry, given that individuals with equivalent migrant ancestry can be RT hybrids or backcrosses.

### *Statistical analyses*

The effects of introgression on individual fitness were described by modelling length at age, survival, and LRS as a function of ancestry classes and differing levels of migrant ancestry. All models were fit in a Bayesian framework using the Markov Chain Monte Carlo simulation program *JAGS* v4.3.1 (Plummer 2003) through the *R* package ‘jagsUI’ (Kellner 2021). The parameters of interest in each analysis were estimated within generalized linear mixed-model (GLM) framework, which permitted the use of an identical model formulation, albeit with different link functions and random effects. The generic model formulation was as follows:

$$Response \sim (\beta_{0 [Ancestry\ Class[i]]} \times I_i) + (\beta_{0\ MA} + \beta_{1\ MA} \times I_i \equiv 0) + \varepsilon_{time} \quad (1)$$

Each of the ancestry classes RR, RT, TT had an independently estimated mean (i.e., intercept) for each response of interest. We used an indicator variable ( $I_i$ ) to identify back-crossed individuals whose parameters were instead estimated as a linear function of migrant ancestry (MA) with an intercept estimated independently of the other ancestry classes. The random effect varied among models, but generally was used to minimize the influence of temporal changes across the sampling period (e.g., density, birth year). We explored alternative functional relationships between migrant ancestry and responses of

interest (e.g., quadratic) and found that linear relationships were most supported by the data for body size at age, survival, and lifetime reproductive success. The statistical significance of a parameter estimate or the difference among estimates was calculated by using the probability of direction ( $pd$ ), which indicates the probability that an estimate overlaps zero, which is analogous to a frequentist one-sided  $p$ -value. Unless otherwise stated, models were fit with five chains of up to 10,000 adaptive phase iterations, 5,000 burn-in iterations and 10,000 estimation iterations with a thin rate of ten. Model convergence was assessed by visually inspecting chains with the *R* package ‘mcmcplots’(Curtis 2015) and the potential scale reduction factor (diagnostic values < 1.1) indicated good chain mixing (Gelman and Rubin 1992).

We tested for effects of different ancestry classes on total length at age because body size could be a source of individual fitness differences. Body size is strongly related to early juvenile survival (Hunt 1969) and fecundity (Wydoski and Cooper 1966) in brook trout and other fishes. Using only known-age individuals, we modelled size at age with the model formulation presented in Equation 1 for each age-class with a Gaussian error distribution and variance  $\sigma_{age}^2$  estimated for each age class. Given the age-structure of this population we only considered ages 0–3 (discussed below). We used four random effects for sample year estimated for each age and fitted as a zero-mean Gaussian distribution with variance  $\sigma_{year[age]}^2$ . In addition to body size at age, we estimated mean age at successful reproduction. To be included in the model, individuals needed to be of known age and have captured progeny of known age. Unlike other models, we found no support for a linear relationship between migrant ancestry and mean age at reproduction and thus compared means among ancestry classes (RR, RT, TT, and BC) in a Bayesian one-way ANOVA with unequal variance. Although age at successful reproduction takes the form of count data, we used a Gaussian error distribution rather than Poisson because of under-dispersion (Schaub and Kery 2021).

The effect of ancestry on survival was estimated using a multistate Cormack-Jolly-Seber (CJS) model (Lebreton et al. 2009) using capture histories from genetic mark-recapture. We used three states based on body size to control for the effect of age and size-class on survival. The first state corresponded with age-0 individuals and the body size threshold varied slightly among years (described above). The second and third

states corresponded with body size  $< 200$  mm and  $\geq 200$  mm, respectively. Only forward transitions or remaining in the same state were allowed, except for the transition from state one to state two, which was forced due to its correspondence with age ( $\psi_{1,1} = 0$ ). Apparent survival in each state ( $\phi_i$ ) and the transitions ( $\psi_{1,3}, \psi_{2,3}$ ) were modelled following equation 1 with logit link functions. The other possible transition probabilities were derived from the complement of  $\psi_{1,3}$  and  $\psi_{2,3}$ . We estimated mean probability of detection ( $p$ ) for state two and state three, given that state one is the first possible capture,  $p_1$  is not estimable. We used three random effects based on sample year for survival probabilities, transition probabilities, and probability of detection fitted as zero-mean Gaussian distributions with variance terms  $\sigma_\phi^2$ ,  $\sigma_\psi^2$ , and  $\sigma_p^2$ , respectively. We fit the multistate CJS model using 30,000 burn-in iterations and 30,000 estimation iterations with a thin rate of 20.

Differences in lifetime reproductive success (individual fitness) based on ancestry class would be among the best evidence for evaluating the outcome of this genetic rescue attempt (Bell et al. 2019). However, individual reproductive output in salmonids is notoriously over-dispersed (Araki et al. 2007, Koch and Narum 2021) and zero-inflated when unsuccessful parents are known. As a result, we used an approach similar to a hurdle model (Welsh et al. 1996), and modeled binary reproductive success and non-zero progeny counts separately. Only individuals with a true or estimated birth year prior to 2016 were included in either model to minimize the potential bias caused by individuals reproducing after the sampling period. We modeled the probability of lifetime contribution (PLC) following the model formulation in Equation 1 with a logit link function and Bernoulli error distribution. LRS of individuals with non-zero progeny counts (LRS<sub>ZT</sub>) were again modeled following Equation 1 with a log link function using a zero-truncated negative binomial error distribution. We used a random effect in both PLC and LRS<sub>ZT</sub> models for birth year fitted as a zero-mean Gaussian distribution with variance  $\sigma_{BirthYear}^2$ . We constructed *post hoc* distributions of composite LRS, hereafter LRS, that combined results of the PLC and LRS<sub>ZT</sub> parameter estimates. This was performed by taking 10,000 random samples from the posterior distributions of parameters and taking the product of back-transformed values (inverse link function)

from each model. For linear models of migrant ancestry, we performed the same process but over the observed range of migrant ancestry values in 0.001 increments.

## **Results**

### *Demographic and Genetic Summary*

Following the genetic rescue attempt, we observed positive changes in abundance and genetic variation in the Dry Run brook trout population. Adult abundance increased by 344.47% from an estimated 62.60 individuals in 2011 compared to the post-translocation average of 278.40. The abundance of age-0 individuals increased by an order of magnitude (1992.12%) from a low of 50.85 estimated in 2011 to the post-translocation average of 1063.89 (Figure 5-2). Of the 6082 captured brook trout, the threshold of 50% genotyping success rate was achieved for 5670 individuals, and the retained individuals had a mean genotyping success rate of 94.20%. We removed 15 loci that had genotyping success rates less than 50% and four that exhibited consistent deviations from HWP among sample years. Using the 198 remaining SNP loci, observed heterozygosity increased by 7.60% from 0.30 in 2011 compared to the post-translocation average ( $H_O=0.32$ ) (Figure 5-3). Additionally, genetic differentiation declined between Dry Run and Dry River from an initial  $F_{ST} = 0.13$  compared to the post-translocation average of  $F_{ST} = 0.07$  (range 0.06–0.08).

### *Pedigree Reconstruction and Migrant Ancestry*

Pedigree reconstruction was successful in assigning parents to a majority of the captured individuals and revealed the extent of introgression of migrant alleles into the Dry Run brook trout population. Following genetic mark-recapture, we identified 4629 unique individuals, 87.72% of which were captured at age-0, so were of known age. The capture histories revealed that 852 individuals were captured more than once, 149 more than twice, and 17 individuals were captured more than three times. A single known-age individual reached age five, which is the maximum observed age in this population. Of the possible parental assignments ( $2n$ ) for the 4629 uniquely captured individuals, 91.18% were successfully assigned to captured individuals. The remaining individuals

were assigned inferred parents, which formed the remaining pedigree links and permitted the calculation of pedigree-based migrant ancestry. The mean post-translocation migrant ancestry was 0.31 (range 0.25–0.37) considering all captured individuals (Figure 5-3). The cohort-specific distributions of migrant ancestry reveal that the majority of age-0 individuals captured in 2014 or later were backcrosses (Figure 5-4), which corresponds with the  $F_1$  RT hybrids reaching age-1 at the time of reproduction in autumn 2013.

The translocated individuals made large contributions to the 2012 cohort with one individual contributing to the 2013 cohort. We detected progeny from eight out of ten of translocated individuals with a mean  $LRS_{ZT}$  of 63.8 (range 4–199) compared to a mean of 22.1 (range 1–223) considering all individuals hatched prior to 2015. All five translocated females contributed progeny compared to three of the translocated males. Despite the differences in reproductive success rate, the sexes contributed similar numbers of progeny, with females contributing 260 compared to 250 by males. We recaptured two translocated individuals, both in 2012, one of which successfully reproduced in the autumn of 2012. A single full-sibling family ( $n = 22$ ) produced by two translocated parents (TT) was detected in the 2012 cohort (Figure 5-4). Members of this family were also reproductively successful with 45.5% producing offspring with a mean  $LRS_{ZT}$  of 32.8 range (4 – 153). It is worth noting that the TT ancestry class represents the fewest individuals and includes translocated individuals and a single full-sibling family.

#### *Effects of Admixture on Individual Fitness and Related Traits*

The progeny of translocated individuals and their descendants showed a consistent pattern of larger size at age and the ancestry classes exhibited differences in mean age of reproduction. Considering only individuals with known birth year, we observed 23 individuals that reached age four and older but omitted them from analyses due to sample size. All other age classes were represented by more than 100 observations. Individuals in the RT and TT ancestry classes were significantly larger ( $pd < 0.05$ ) than RR at age-0, age-1, and age-2 (Table 5-1). Interestingly, RR individuals were larger on average than both RT and TT at age-3. We observed a significantly positive relationship with migrant ancestry and body size in back-crossed individuals, except in age-2 fish (Figure 5-5). In general, we found that mean age at reproduction was similar among the ancestry classes

with a mean of 1.48 years, however, RT individuals reproduced significantly earlier, on average, than those of RR ancestry (Table 5-2).

Survival is an important component of individual fitness and comparisons among ancestry classes revealed significant differences. Our multistate CJS survival model results indicate that RT individuals had consistently higher apparent survival compared RR individuals, which was statistically significant in state one (age-0) and at a state three (length  $\geq 200$  mm). Individuals in the ancestry class TT had higher survival on average in states one and two compared to both RR and RT, but survival estimates similar to RR in state 3 (Table 5-3). Interestingly, we found consistently negative, mean slope coefficients for the relationship between migrant ancestry and apparent survival in backcrossed individuals. This negative relationship was statistically significant in state two ( $pd = 0.01$ ). However, the mean, estimated y-intercept, corresponding with zero migrant ancestry, was greater than the mean estimate of survival for RR in all states (Figure 5-6). Transition probabilities among states are not reported because they were consistent, and therefore redundant, with the size at age model results, which suggest faster growth rates in RT and TT and a positive relationship with migrant ancestry.

The models of lifetime reproductive success revealed similar patterns to the apparent survival results with elevated fitness in the RT ancestry class compared to RR and a negative association with migrant ancestry in back-crossed individuals. Individuals in the RT and TT ancestry classes had significantly higher estimates of probability of lifetime contribution (PLC) compared to RR with odd ratios of 2.23 and 3.31, respectively (Table 5-4). The relationship between PLC and migrant ancestry was negative and highly significant ( $pd < 0.001$ ), with a mean, estimated y-intercept higher than that of the mean estimate for RR individuals (Figure 5-7). The results of modelling zero-truncated progeny counts ( $LRS_{ZT}$ ) among the ancestry classes were directionally consistent with the PLC model but had high uncertainty due to the over-dispersed nature of the data. Despite the high uncertainty, we found  $LRS_{ZT}$  was significantly higher in TT compared to RT and RR ancestry classes. LRS was derived *post hoc* from the product of PLC and  $LRS_{ZT}$  posterior distributions and showed that TT individuals had significantly higher reproductive success than RT and TT individuals (Table 5-4). On average, RT individuals had 2.52 more progeny than RR. Despite a negative relationship between

LRS and migrant ancestry, backcrossed individuals had higher LRS, on average, than RR when migrant ancestry was 0.487 or less (Figure 5-8).

## **Discussion**

Our results provide a demonstration of the complex interplay between demographic and evolutionary factors during genetic rescue attempts in natural populations. We observe consistent support for elevated fitness of  $F_1$  hybrids relative to resident individuals in body size, survival, and lifetime reproductive success. In contrast, we found a gradient of fitness with only low levels of migrant ancestry conveying a fitness advantage relative to resident individuals in the  $F_2$  generation and later. These results appear to represent the simultaneous beneficial and deleterious effects of gene flow (Garant et al. 2007). These findings are consistent with an alleviation of genetic load in the recipient population, and subsequent selection to maintain local ancestry, as a likely result of local adaptation. Our study is among few multigenerational genetic rescue studies to examine differences in individual fitness across the full lifespan in natural populations (but see Fitzpatrick et al. 2020). To our knowledge, this is the first multigenerational study to separately estimate the fitness of parental types,  $F_1$  hybrids, and backcrosses, which was critical to reveal the multifaceted fitness effects of gene flow in this population. Our findings provide important, conservation-relevant insights into genetic rescue, supporting that gene flow can elevate fitness in genetically depauperate populations without swamping local ancestry, but also that the process is mediated by fitness trade-offs that may be difficult to anticipate.

Following gene flow, we observed dramatic increases in adult and juvenile brook trout abundance. An increase in population growth rate is strong evidence for genetic rescue, as long as it can be attributed to fitness benefits of gene flow (Tallmon et al. 2004). Robinson et al. (2017) monitored a nearby population that did not receive migrants and observed similar increases in abundance in 2012 and 2013. The observed increase in abundance over the sample period is likely due to the Moran effect (Moran 1953), or spatial population synchrony, often caused by regional climatic patterns. Regional synchrony in population dynamics is well documented in multiple species of salmonid fishes (Zorn and Nuhfer 2007, Kovach et al. 2018), including in mid-Atlantic

brook trout populations (Kanno et al. 2016b). We therefore cannot attribute the initial increase in abundance in Dry Run to the fitness effects of gene flow. However, it remains a possibility that the elevated reproductive success of the  $F_1$  hybrids and backcrosses with low levels of migrant ancestry resulted a higher population growth rate than would have otherwise occurred (e.g., Johnson et al. 2010). Future work will evaluate this possibility through simulation or integrated population models that appropriately accounts for density-dependence due to its importance in this species.

Consistent positive relationships with migrant ancestry and fitness following gene flow is likely a naïve expectation. A multigenerational study of genetic rescue in Trinidadian guppies showed higher fitness with intermediate levels of migrant ancestry and resistance to introgression of maladaptive alleles (Fitzpatrick et al. 2016). This trade-off between gene flow and local adaptation is predicted by population genetic theory and represents a balance between migration, selection, and drift load in metapopulations (Whitlock et al. 2000). The pulse of gene flow into Dry Run compressed an otherwise gradual process of migration and selection into a few short generations. Salmonid fishes are widely considered to exhibit strong local adaptation, with a meta-analysis finding local adaptation occurring within 6–30 generations (Fraser et al. 2011b). Although we interpret our results as being most likely mediated by local adaptation, we did not identify a causal mechanism. The possible contribution of disrupted epistatic gene interactions (Hansen 2013) or genomic incompatibilities such as chromosomal inversions cannot be ruled out (Stenløkk et al. 2022).

Differing stream temperature and flow regimes between the source and recipient population provides a potential driver of local adaptation. The source population experiences altered temperature and flow-regimes due to the numerous dams in the watershed, while the recipient population experiences unregulated, seasonally intermittent flow. Both stream temperature and flow can regulate important life-history traits (Warren et al. 2012) that are known to be locally adapted in salmonid fishes (Fraser et al. 2011b). We examined fitness-related traits, size at age, and mean age at reproduction among individuals of different ancestry. We found positive association between migrant ancestry and size at age in backcrossed individuals. Additionally, we observed a nonsignificant pattern of earlier mean age of reproduction in backcrosses

compared to both parental types. However, these differences were even more pronounced in  $F_1$  hybrids, which exhibited significantly lower mean age of reproduction and larger size at age than residents. As a result, differences in these traits do not illuminate a potential mechanism for the decreased fitness of highly admixed backcrosses. However, it is possible that the consequences of earlier mean age at reproduction and larger size were masked by heterosis in  $F_1$  hybrids (Bell et al. 2019). In contrast, the consistent larger juvenile body size of admixed individuals is more likely directly related to fitness because of its association with early juvenile survival (Hunt 1969, Elliott 1993).

Genetic rescue remains a controversial management strategy, in part, due to concerns over negative outcomes. The potential for more proactive use of assisted gene flow to mitigate the effects of widespread habitat fragmentation is suggested to outweigh the risks due to the lack of evidence for consequential effects of outbreeding depression on population persistence when using existing guidelines for genetic rescue attempts (Frankham et al. 2017). However, others are less sanguine about the potential risks of genetic rescue (Edmands 2007, Waller 2015). Our results suggest that the brook trout population in Dry Run was indeed experiencing reduced fitness due to inbreeding and genetic drift, which were alleviated following gene flow. However, we also observed evidence for negative fitness consequences in some individuals, namely those with migrant ancestry  $> 0.48$ . The potential for negative outcomes appears to have been present and such outcomes were minimized due, in part, to regional climatic patterns influencing abundance. It is likely that the large abundance increases that coincided with gene flow allowed for natural selection to be effective in promoting beneficial phenotypes and reducing deleterious ones, similar to the process documented in Fitzpatrick et al. (2020). If instead, the translocation coincided with unfavorable environmental conditions or if more individuals were translocated, the net effect of introgression may have been clearly negative or resulted in the loss of local ancestry (e.g., Hedrick et al. 2019). Genetic rescue is fundamentally an eco-evolutionary process which is highly context specific (Lowe et al. 2017), and risk assessments for genetic rescue attempts should consider that temporal variation in ecological drivers of population dynamics may be as important as the genetic attributes of source and recipient populations.

## Conclusions

This work contributes to a growing body of literature indicating that human-mediated gene flow can positively affect individual fitness in small, isolated populations experiencing high levels of inbreeding and genetic drift. It also provides an important demonstration that gene flow into small, inbred populations will often involve trade-offs, that play out over multiple generations, between the costs of disrupting local adaptation and alleviation of genetic load. The net effect on population persistence probability should determine whether genetic rescue occurred. The population dynamics of our focal species is highly density-dependent (Letcher et al. 2007, Huntsman and Petty 2014) and, as a result, the observed increases in relative fitness of admixed individuals may not translate to an increase in mean absolute fitness (Bell et al. 2021). Although, our results are consistent with genetic rescue, there is insufficient evidence to conclude that gene flow had a consequential effect on population growth rate. Nonetheless, the attempted genetic rescue restored genetic connectivity within a historical metapopulation, increased standing genetic variation, and appeared to alleviate the deleterious effects of inbreeding and genetic drift. Assisted gene flow is a powerful management strategy and, despite difficulties in conclusive documentation, it may be the most cost-effective way to increase the persistence probability for countless populations on an increasingly fragmented landscape.

Table 5-1. Estimates of mean length at age for each ancestry class. The mean estimate and credible interval of the posterior distribution for the slope coefficient for migrant ancestry are provided for backcrosses (BC). The posterior distributions of the ancestry classes representing resident (RR) and transplant (TT) parental types, and  $F_1$  hybrids (RT) were compared *post hoc* and probability of direction (*pd*) is reported.

Length at Age-0				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	68.5 (63.5, 72.7)	RT-RR	4.59 (3.75, 5.45)	<0.001
RT	73.1 (68.1, 77.3)	RT-TT	0.374 (-2.29,3.15)	0.393
TT	72.7 (67.2, 77.6)	TT-RR	4.23 (1.49,6.88)	0.001
BC	3.33 (1.95, 4.78)	–	–	<0.001
Length at Age-1				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	142 (136, 149)	RT-RR	18.9 (15.1, 22.6)	<0.001
RT	161 (154, 169)	RT-TT	6.16 (-4.70, 17.1)	0.138
TT	155 (142, 168)	TT-RR	12.8 (1.77, 23.7)	0.010
BC	23.7 (15.2,32.3)	–	–	<0.001
Length at Age-2				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	177 (170, 184)	RT-RR	18.6 (12.7, 24.3)	<0.001
RT	195 (187, 203)	RT-TT	3.24 (-8.42, 14.4)	0.290
TT	192 (180, 205)	TT-RR	15.4 (4.23, 26.6)	0.005
BC	-5.12 (-16.6, 6.22)	–	–	0.194
Length at Age-3				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	213 (204, 222)	RT-RR	-1.68 (-11.9, 8.21)	0.633
RT	212 (202, 222)	RT-TT	12.8 (-7.22, 33.6)	0.097
TT	199 (178, 219)	TT-RR	-14.6 (-35.5, 5.18)	0.075
BC	27.6 (0.803, 54.9)	–	–	0.021

Table 5-2. Estimates of mean age at reproduction in each ancestry class. The posterior distributions of the ancestry classes representing resident (RR) and transplant (TT) parental types,  $F_1$  hybrids (RT), and backcrosses (BC) were compared *post hoc* and probability of direction (*pd*) is reported.

Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	1.59 (1.45, 1.74)	RR-RT	0.249 (0.058, 0.439)	0.004
		RR-TT	0.094 (-0.481, 0.653)	0.365
		RR-BC	0.119 (-0.110, 0.350)	0.149
RT	1.34 (1.22, 1.47)	RT-TT	-0.155 (-0.734, 0.397)	0.279
		RT-BC	-0.128 (-0.340, 0.088)	0.127
TT	1.50 (0.98, 2.06)	TT-BC	0.031 (-0.529, 0.598)	0.548
BC	1.47 (1.29, 1.65)	—	—	—

Table 5-3. Estimates of mean survival for each ancestry class in three states corresponding to age and total length. The mean estimate and credible interval of the posterior distribution for the slope coefficient for migrant ancestry are provided for backcrosses (BC). The posterior distributions of the ancestry classes representing resident (RR) and transplant (TT) parental types, and  $F_1$  hybrids (RT) were compared *post hoc* and probability of direction (*pd*) is reported.

Age-0 Survival				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	0.277 (0.136, 0.488)	RT-RR	0.067 (-0.001, 0.145)	0.027
RT	0.346 (0.167, 0.578)	RT-TT	-0.106(-0.393,0.103)	0.183
TT	0.467 (0.186, -0.792)	TT-RR	0.171 (-0.035,0.461)	0.059
BC	-0.086 (-0.795, 0.650)	—	—	0.400
< 200 mm Survival				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	0.236 (0.109, 0.447)	RT-RR	0.0420 (-0.039, 0.153)	0.155
RT	0.281 (0.127, 0.520)	RT-TT	-0.426(-0.752, -0.111)	0.002
TT	0.801 (0.337, 0.997)	TT-RR	0.471 (0.155, 0.790)	<0.001
BC	-1.52 (-3.01, -0.159)	—	—	0.013
≥ 200 mm Survival				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	0.178 (0.070, 0.372)	RT-RR	0.127 (-0.009, 0.299)	0.033
RT	0.311 (0.136, 0.570)	RT-TT	0.112 (-0.158, 0.353)	0.179
TT	0.182 (0.035, 0.528)	TT-RR	0.009 (-0.177, 0.284)	0.465
BC	-0.265 (-5.48, 4.81)	—	—	0.461

Table 5-4. Estimates of mean probability of lifetime contribution, zero-truncated lifetime reproductive success, and the derived lifetime reproductive success for each ancestry class. Lifetime reproductive success is generated *post hoc* by the product of the posterior distributions of the two former models. The mean estimate and credible interval of the posterior distribution for the slope coefficient for migrant ancestry are provided for backcrosses (BC). The posterior distributions of the ancestry classes representing resident (RR) and transplant (TT) parental types, and  $F_1$  hybrids (RT) were compared *post hoc* and probability of direction (*pd*) is reported.

Probability of Lifetime Contribution				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	0.223 (0.102, 0.454)	RT-RR	0.158 (0.077, 0.248)	<0.001
RT	0.390 (0.195, 0.656)	RT-TT	-0.093 (-0.260, 0.076)	0.144
TT	0.487 (0.228, 0.763)	TT-RR	0.250 (0.080, 0.422)	<0.001
BC	-3.21 (-4.32, -2.08)	–	–	<0.001
LRS   Progeny Detected				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	11.3 (4.80, 24.9)	RT-RR	1.42 (-4.68, 11.2)	0.304
RT	12.8 (5.27, 30.7)	RT-TT	-17.8 (-101, 1.10)	0.037
TT	32.1 (9.41, 121)	TT-RR	19.6 (0.714, 102)	0.017
BC	-0.036 (-2.22, 2.24)	–	–	0.477
LRS				
Ancestry Class	Estimate	Comparison	Difference	<i>pd</i>
RR	2.52 (0.787, 7.35)	RT-RR	2.26 (-3.35, 11.2)	0.193
RT	4.91 (1.60, 13.8)	RT-TT	-9.69 (-56.1, 4.31)	0.099
TT	14.9 (3.68, 61.5)	TT-RR	12.2 (0.127, 58.6)	0.023

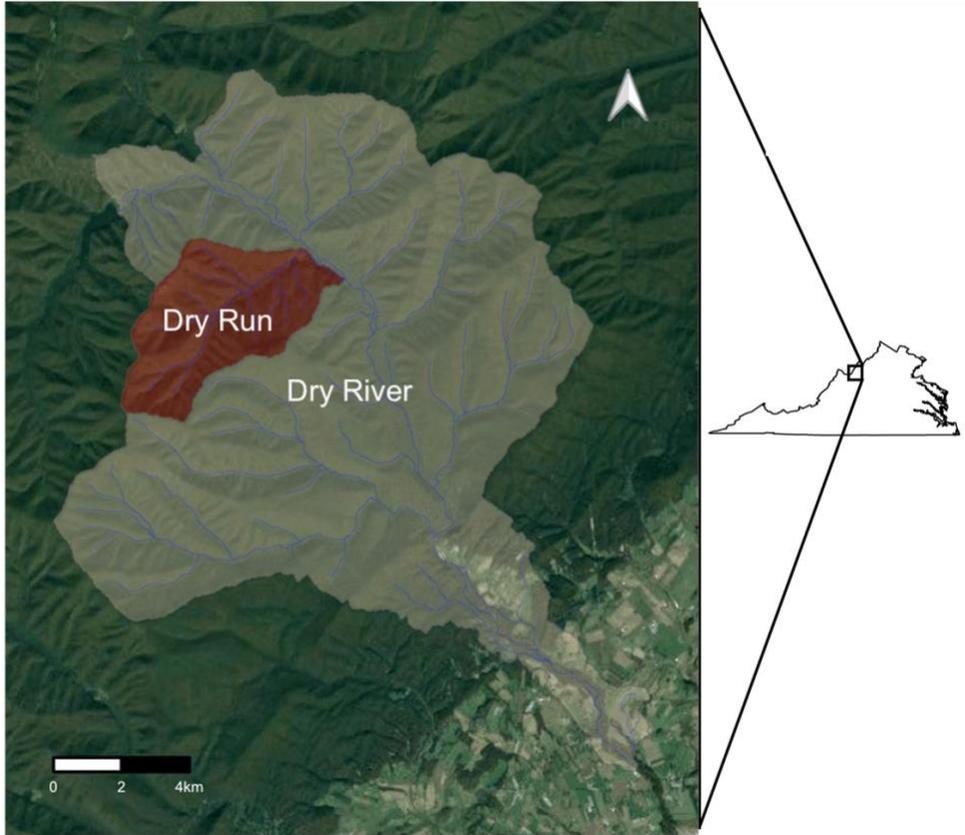


Figure 5-1. Map of the Dry River and Dry Run watersheds in Virginia, USA. Dry Run was isolated from the broader Dry River watershed in 1968 by a flood control dam and was the recipient of translocated individuals from Dry River to test for successful genetic rescue.

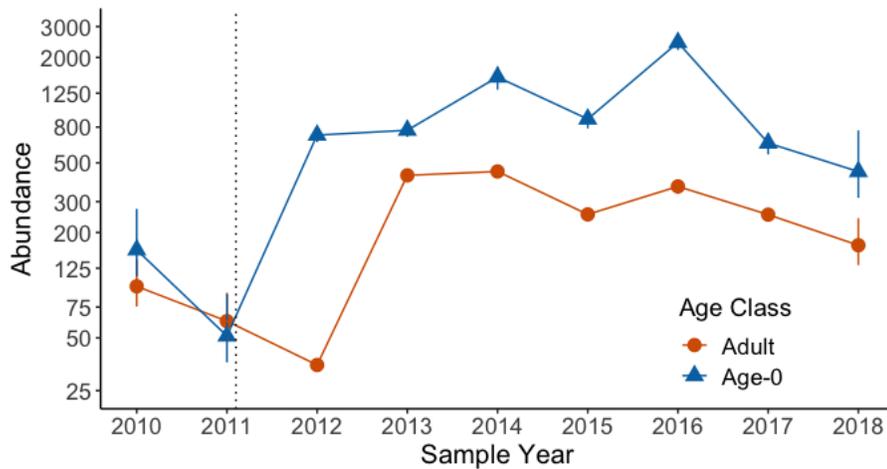


Figure 5-2. Trends in abundance for age-0 and adult brook trout over the study period. The vertical, dotted line represents the date of introduction of ten translocated individuals. Abundance was estimated using mark-recapture and the Lincoln-Peterson estimator for sample years 2012–2017. For remaining years, we estimated abundance and confidence intervals by dividing the number of marked fish by 1000 bootstrapped samples of observed probability of detection, corresponding to age class, in years with recapture events.

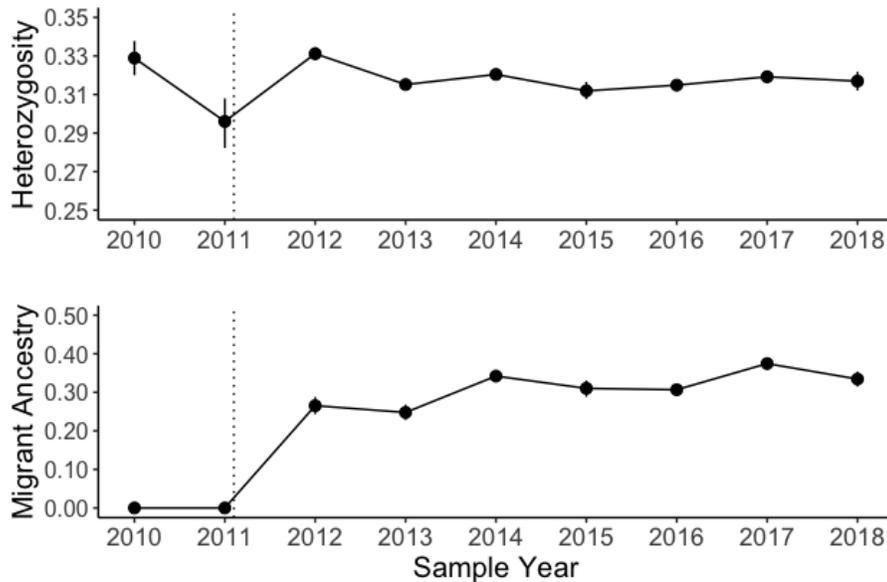


Figure 5-3. Trends in mean heterozygosity and migrant ancestry over the sampling period. The vertical, dotted line represents the date of introduction of ten translocated individuals. Confidence intervals were generated using 1000 bootstrapped samples across individuals.

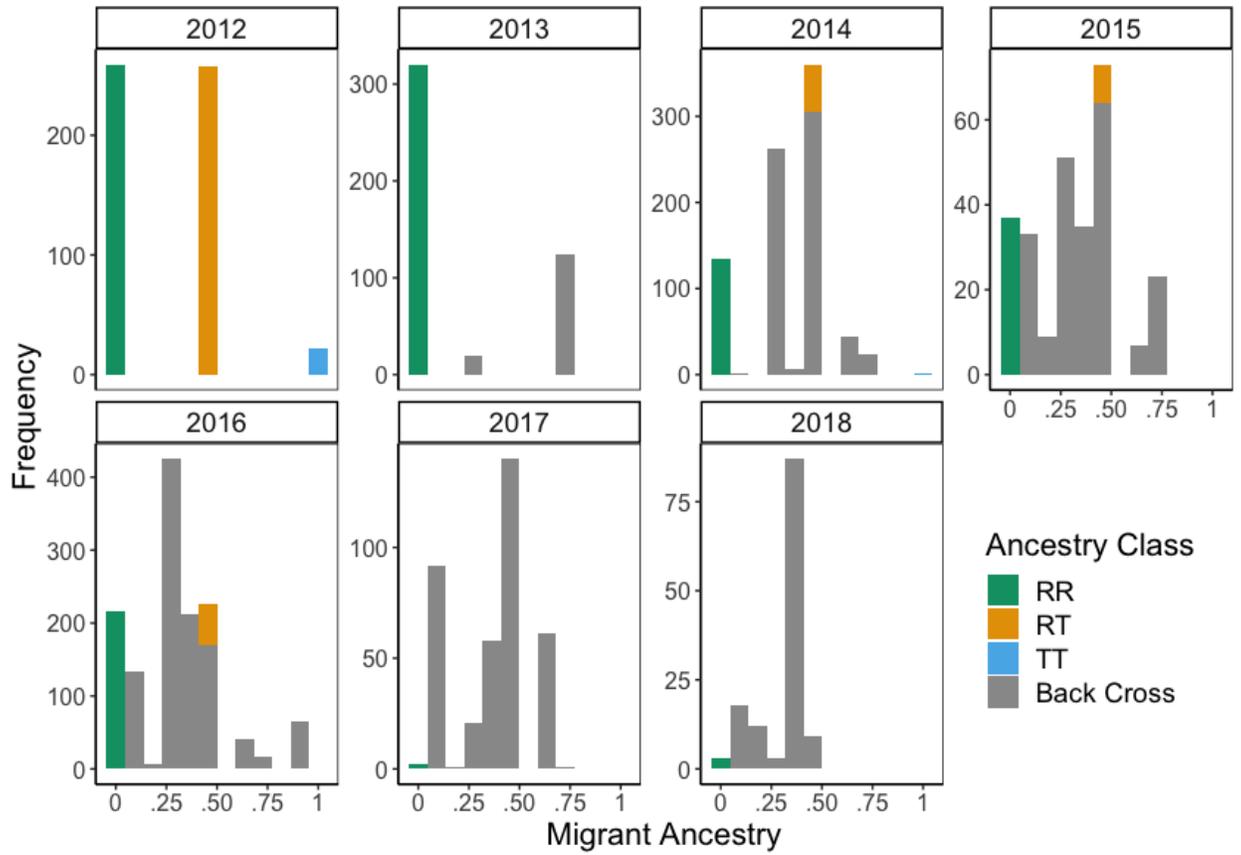


Figure 5-4. Distributions of cohort-specific migrant ancestry over the sampling period. The ancestry classes representing resident (RR; green) and transplant (TT; blue,) parental types, and  $F_1$  hybrids (RT; orange) and backcrosses (grey) are presented.

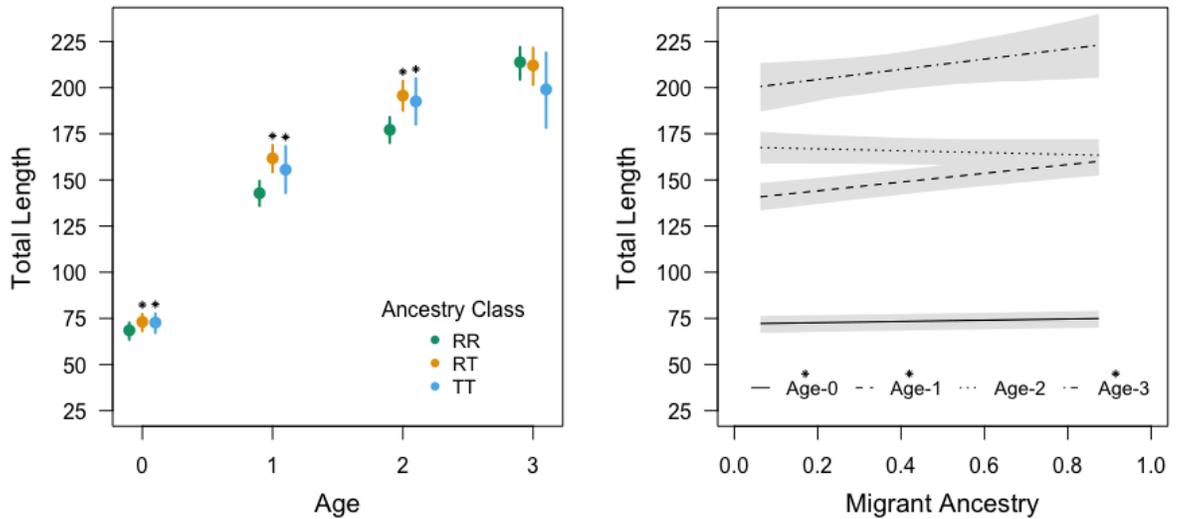


Figure 5-5. The effects of ancestry on length at age. The left panel presents the mean estimate and credible intervals of the posterior distribution for each ancestry class. Statistical significance ( $pd < 0.05$ ) was only observed relative to resident ancestry (RR) and asterisks denote significant estimates. The right panel presents the linear effect of migrant ancestry on length at age in backcrosses. The 95% credible intervals for the linear models are shown in grey and significant slope coefficients are denoted by asterisks.

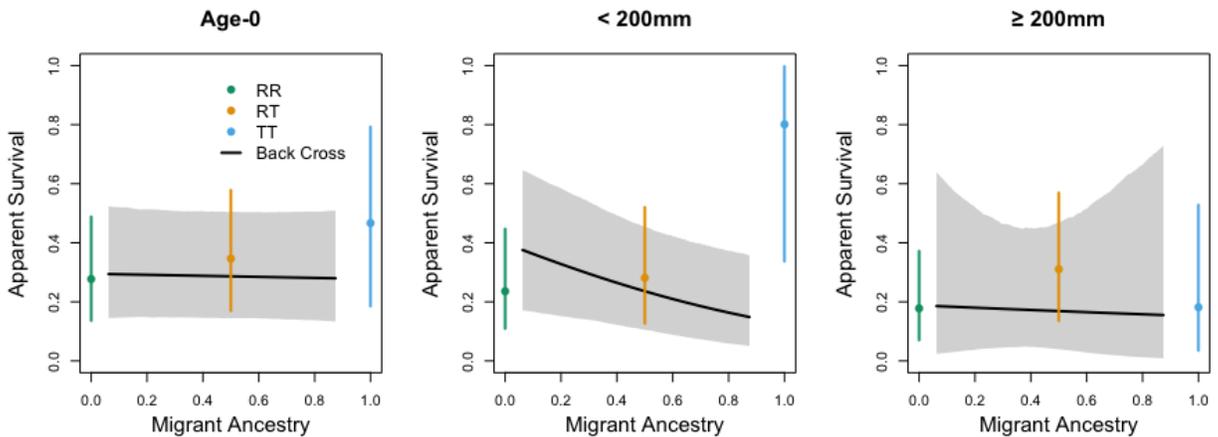


Figure 5-6. The effects of ancestry on mean survival in three states corresponding to age and total length. In each panel, the mean estimate and credible intervals of the posterior distribution for each ancestry class, and the linear relationship of migrant ancestry on survival in backcrosses are presented. The linear model results are presented over the observed range of migrant ancestry. The 95% credible intervals for the linear models are shown in grey.

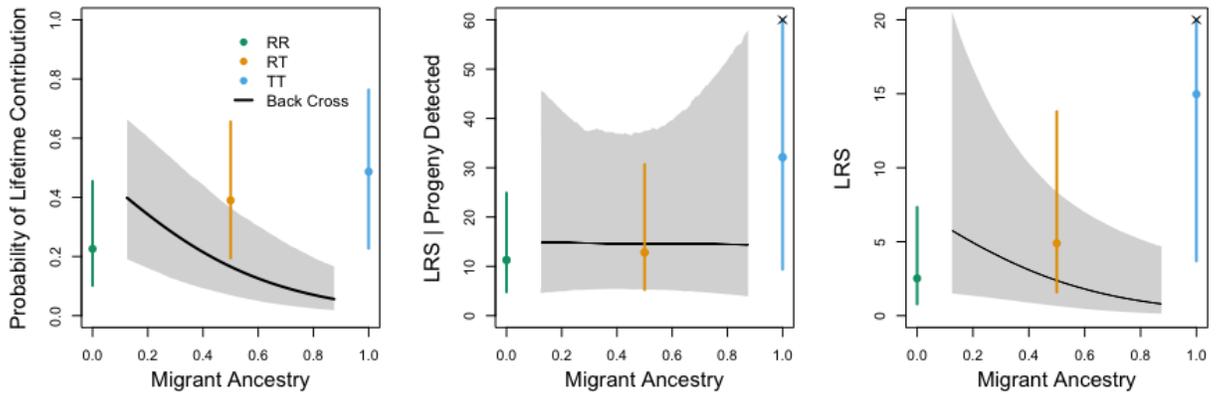


Figure 5-7. The effects of ancestry on probability of lifetime contribution, zero-truncated lifetime reproductive success, and lifetime reproductive success. Lifetime reproductive success is generated *post hoc* by the product of the posterior distributions of the two former models. In each panel, the mean estimate and credible intervals of the posterior distribution for each ancestry class, and the linear relationship of migrant ancestry on survival in backcrosses. The linear model results are presented over the observed range of migrant ancestry. The 95% credible intervals for the linear models are shown in grey.

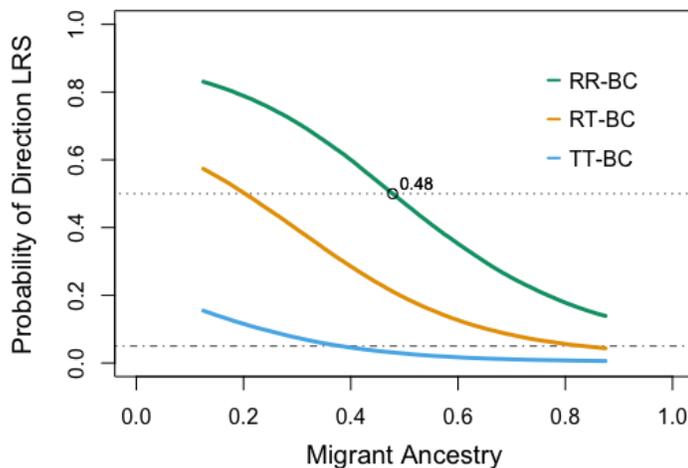


Figure 5-8. The probability of direction ( $pd$ ) comparing the mean lifetime reproductive success (LRS) of ancestry classes that received mean estimates (RR, RT, TT) to backcrosses (BC) across the range of observed migrant ancestry. BC had higher fitness, on average, when  $pd$  is greater than 0.50, which is represented by a dotted line. The comparison of RR to BC intersects  $pd=0.50$  at a migrant ancestry value of 0.48, which is the point at which BC have lower fitness than RR, on average. The compared ancestry classes achieve significantly higher reproductive success than backcrosses at ( $pd < 0.05$ ), which is represented by a dashed line.

## REFERENCES

- Ackerman, M. W., B. K. Hand, R. K. Waples, G. Luikart, R. S. Waples, C. A. Steele, B. A. Garner, J. McCane, and M. R. Campbell. 2017. Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. *Evolutionary Applications* 10:146–160.
- Aitken, K. E. H., and K. Martin. 2012. Experimental test of nest-site limitation in mature mixed forests of central British Columbia, Canada. *The Journal of Wildlife Management* 76:557–565.
- Åkesson, M., O. Liberg, H. Sand, P. Wabakken, S. Bensch, and Ø. Flagstad. 2016. Genetic rescue in a severely inbred wolf population. *Molecular ecology* 25:4745–56.
- Ali, O. A., S. M. O'Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffres, and M. R. Miller. 2016. RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* 202:389–400.
- Aljanabi, S. M., and I. Martinez. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* 25:4692–4693.
- Allendorf, F. W. 1986. Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* 5:181–190.
- Anderson, E. C., and J. C. Garza. 2006. The Power of Single-Nucleotide Polymorphisms for Large-Scale Parentage Inference. *Genetics* 172:2567–2582.
- Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet advance on*:81–92.
- Annett, B., G. Gerlach, T. L. King, and A. R. Whiteley. 2012. Conservation genetics of remnant coastal brook trout populations at the southern limit of their distribution: Population structure and effects of stocking. *Transactions of the American Fisheries Society* 141:1399–1410.
- Araki, H., W. R. Ardren, E. M. Olsen, B. Cooper, and M. S. Blouin. 2007. Reproductive success of captive-bred steelhead trout in the wild: evaluation of three hatchery programs in the Hood River. *Conservation Biology* 21:181–190.
- Bacles, C. F. E., C. Bouchard, F. Lange, A. Manicki, C. Tentelier, and O. Lepais. 2018. Estimating the effective number of breeders from single parr samples for conservation monitoring of wild populations of Atlantic salmon *Salmo salar*. *Journal of Fish Biology* 92:699–726.
- Baetscher, D. S., A. J. Clemento, T. C. Ng, E. C. Anderson, and J. C. Garza. 2018. Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources* 18:296–305.
- Bassar, R. D., B. H. Letcher, K. H. Nislow, and A. R. Whiteley. 2016. Changes in seasonal climate outpace compensatory density-dependence in eastern brook trout. *Global Change Biology* 22.
- Beacham, T. D., C. Wallace, C. Macconnachie, K. Jonsen, B. McIntosh, J. R. Candy, and R. E. Withler. 2018. Population and individual identification of chinook salmon in British Columbia through parentage-based tagging and genetic stock identification with single nucleotide polymorphisms. *Canadian Journal of Fisheries and Aquatic Sciences* 75:1096–1105.

- Beard, T. D., and R. F. Carline. 1991. Influence of spawning and other stream habitat features on spatial variability of wild brown trout. *Transactions of the American Fisheries Society* 120:711–722.
- Bell, D. A., R. P. Kovach, Z. L. Robinson, A. R. Whiteley, and T. E. Reed. 2021. The ecological causes and consequences of hard and soft selection. *Ecology Letters* 24:1505–1521.
- Bell, D. A., Z. L. Robinson, C. W. 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 & Evolution* 34:1070–1079.
- Béréanos, C., P. A. Ellis, J. G. Pilkington, and J. M. Pemberton. 2014. Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular Ecology* 23:3434–3451.
- Bernos, T. A., and D. J. Fraser. 2016. Spatiotemporal relationship between adult census size and genetic population size across a wide population size gradient. *Molecular ecology* 25:4472–4487.
- Bernos, T. A., K. M. Jeffries, and N. E. Mandrak. 2020, December 1. Linking genomics and fish conservation decision making: a review. Springer Science and Business Media Deutschland GmbH.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bossuyt, B. 2007. Genetic rescue in an isolated metapopulation of a naturally fragmented plant species, *Parnassia palustris*. *Conservation biology* 21:832–41.
- Bozzuto, C., I. Biebach, S. Muff, A. R. Ives, and L. F. Keller. 2019. Inbreeding reduces long-term growth of Alpine ibex populations. *Nature Ecology and Evolution* 3:1359–1364.
- Brauer, C. J., and L. B. Beheregaray. 2020. Recent and rapid anthropogenic habitat fragmentation increases extinction risk for freshwater biodiversity. *Evolutionary Applications* 13:2857–2869.
- Bravington, M. v., P. M. Grewe, and C. R. Davies. 2016. Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications* 7:1–8.
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10:1–9.
- Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources* 15:855–867.
- Caughley, G. 1994. Directions in Conservation Biology. *The Journal of Animal Ecology* 63:215.
- Chan, W. Y., A. A. Hoffmann, and M. J. H. Oppen. 2019. Hybridization as a conservation management tool. *Conservation Letters* 12:e12652.
- Chapman, J. R., S. Nakagawa, D. W. Coltman, J. Slaters, and B. C. Sheldon. 2009. A quantitative review of heterozygosity – fitness correlations in animal populations. *Molecular Ecology* 18:2746–2765.
- Charlesworth, D., and J. H. Willis. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* 10:783–796.

- Christensen, K. A., E. B. Rondeau, D. R. Minkley, J. S. Leong, C. M. Nugent, R. G. Danzmann, M. M. Ferguson, A. Stadnik, R. H. Devlin, R. Muzzerall, M. Edwards, W. S. Davidson, and B. F. Koop. 2018. The Arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLOS ONE* 13:e0204076.
- Cleveland, L., E. E. Little, S. J. Hamilton, D. R. Buckler, and J. B. Hunn. 1986. Interactive toxicity of aluminum and acidity to early life stages of brook trout. *Transactions of the American Fisheries Society* 115:610–620.
- Clutton-Brock, T. H. 1988. Reproductive success: studies of individual variation in contrasting breeding systems. University of Chicago Press.
- Cohen, J. 1992. A power primer. *Psychological Bulletin* 112:155–159.
- Coombs, J. A. 2010. Reproduction in the wild: the effect of individual life history strategies on population dynamics and persistence. University of Massachusetts Amherst.
- Courtwright, J., and C. L. May. 2013. Importance of terrestrial subsidies for native brook trout in Appalachian intermittent streams. *Freshwater Biology*.
- Curry, R. A., and D. L. G. Noakes. 1995. Groundwater and the selection of spawning sites by brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences* 52:1733–1740.
- Curtis, S. M. 2015. mcmcplots: Create plots from MCMC output. <https://cran.r-project.org/package=mcmcplots>.
- Cushman, S. A. 2006. Effects of habitat loss and fragmentation on amphibians: a review and prospectus. *Biological Conservation* 128:231–240.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
- Dauwalter, D. C., A. Duchi, J. Epifanio, A. Gandolfi, R. Gresswell, F. Juanes, J. Kershner, J. Lobón-Cerviá, P. McGinnity, A. Meraner, P. Mikheev, K. Morita, C. C. Muhlfeld, K. Pinter, J. R. Post, G. Unfer, L. A. Vøllestad, and J. E. Williams. 2020. A call for global action to conserve native trout in the 21st century and beyond. *Ecology of Freshwater Fish* 29:429–432.
- Dauwalter, D. C., F. J. Rahel, and K. G. Gerow. 2009. Temporal variation in trout populations: implications for monitoring and trend detection. *Transactions of the American Fisheries Society* 138:38–51.
- Derry, A. M., D. J. Fraser, S. P. Brady, L. Astorg, E. R. Lawrence, G. K. Martin, J. Matte, J. O. Negrín Dastis, A. Paccard, R. D. H. Barrett, L. J. Chapman, J. E. Lane, C. G. Ballas, M. Close, and E. Crispo. 2019. Conservation through the lens of (mal)adaptation: concepts and meta-analysis. *Evolutionary Applications* 12:1287–1304.
- Deweber, J. T., and T. Wagner. 2014. Predicting brook trout occurrence in stream reaches throughout their native range in the Eastern United States.
- Dirzo, R., H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen. 2014. Defaunation in the Anthropocene. *Science* 345.
- Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size ( $N_e$ ) from genetic data. *Molecular Ecology Resources* 14.

- Dokan, K., S. Kawamura, and K. M. Teshima. 2021. Effects of single nucleotide polymorphism ascertainment on population structure inferences. *G3 Genes|Genomes|Genetics* 11.
- Dumont, B. L., and B. A. Payseur. 2008. Evolution of the genomic rate of recombination in mammals. *Evolution* 62:276–294.
- Dunham, J. B., G. L. Vinyard, and B. E. Rieman. 1997. Habitat fragmentation and extinction risk of lahontan cutthroat trout. *North American Journal of Fisheries Management* 17:1126–1133.
- Dunham, J., B. Rieman, and K. Davis. 2001. Sources and magnitude of sampling error in redd counts for bull trout. *North American Journal of Fisheries Management* 21:343–352.
- Duong, T. Y., K. T. Scribner, P. S. Forsythe, J. A. Crossman, and E. A. Baker. 2013. Interannual variation in effective number of breeders and estimation of effective population size in long-lived iteroparous lake sturgeon (*Acipenser fulvescens*). *Molecular ecology* 22:1282–1294.
- EBTJV. 2016. Range-wide assessment of brook trout at the catchment scale: a summary of findings.
- Edmunds, S. 2007. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology* 16:463–475.
- Elliott, J. M. 1993. The pattern of natural mortality throughout the life cycle in contrasting populations of brown trout. *Fisheries Research* 17:123–136.
- Fenster, C. B., and L. F. Galloway. 2000. Inbreeding and outbreeding depression in natural populations of *Chamaecrista fasciculata* (*Fabaceae*). *Conservation Biology* 14:1406–1412.
- Ferchaud, A.-L., C. Perrier, J. April, C. Hernandez, M. Dionne, and L. Bernatchez. 2016. Making sense of the relationships between  $N_e$ ,  $N_b$  and  $N_c$  towards defining conservation thresholds in Atlantic salmon (*Salmo salar*). *Heredity* 117:268–278.
- Fesenmyer, K. A., A. L. Haak, S. M. Rummel, M. Mayfield, S. L. McFall, and J. E. Williams. 2017. Eastern brook trout conservation portfolio, range-wide habitat integrity and future security assessment, and focal area risk and opportunity analysis. Arlington, Virginia.
- Finger, A., C. J. Kettle, C. N. Kaiser-bunbury, T. Valentin, D. Doudee, D. Matatikin, and J. Ghazoul. 2011. Back from the brink: potential for genetic rescue in a critically endangered tree. *Molecular Ecology* 20:3773–3784.
- Fitzpatrick, S. W., G. S. Bradburd, C. T. Kremer, P. E. Salerno, L. M. Angeloni, and W. C. Funk. 2020. Genomic and fitness consequences of genetic rescue in wild populations. *Current Biology* 30:517-522.e5.
- Fitzpatrick, S. W., J. C. Gerberich, L. M. Angeloni, L. L. Bailey, E. D. Broder, J. Torres-Dowdall, C. A. Handelsman, A. López-Sepulcre, D. N. Reznick, C. K. Ghalambor, and W. Chris Funk. 2016. Gene flow from an adaptively divergent source causes rescue through genetic and demographic factors in two wild populations of Trinidadian guppies. *Evolutionary Applications* 9:879–891.
- Flanagan, S. P., and A. G. Jones. 2019. The future of parentage analysis: from microsatellites to SNPs and beyond. *Molecular Ecology* 28:544–567.

- Frankham, R. 2015. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology* 24:2610–2618.
- Frankham, R., J. D. Ballou, M. D. B. Eldridge, R. C. Lacy, K. Ralls, M. R. Dudash, and C. B. Fenster. 2011. Predicting the probability of outbreeding depression. *Conservation Biology* 25:465–475.
- Frankham, R., J. D. Ballou, K. Ralls, M. D. B. Eldridge, M. R. Dudash, C. B. Fenster, R. C. Lacy, and P. Sunnucks. 2017. Genetic management of fragmented animal and plant populations. Oxford University Press, Oxford, United Kingdom.
- Franklin, I. R. 1980. Evolutionary change in small populations. Pages 135–150 *in* M. E. Soule and B. A. Wilcox, editors. *Conservation Biology: an Evolutionary–Ecological Perspective*. Sinauer Associates.
- Fraser, D. J., P. V Debes, L. Bernatchez, and J. A. Hutchings. 2014. Population size, habitat fragmentation, and the nature of adaptive variation in a stream fish. *Proceedings of the Royal Society Series B* 281:20140370.
- Fraser, D. J., L. K. Weir, L. Bernatchez, M. M. Hansen, and E. B. Taylor. 2011a. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity* 106:404–420.
- Fraser, D. J., L. K. Weir, L. Bernatchez, M. M. Hansen, and E. B. Taylor. 2011b. Extent and scale of local adaptation in salmonid fishes: Review and meta-analysis. *Heredity* 106:404–420.
- Garant, D., S. E. Forde, and A. P. Hendry. 2007, June. The multifarious effects of dispersal and gene flow on contemporary adaptation.
- Garner, B. A., S. Hoban, and G. Luikart. 2020. IUCN Red List and the value of integrating genetics. *Conservation Genetics* 21:795–801.
- Geibel, J., C. Reimer, S. Weigend, A. Weigend, T. Pook, and H. Simianer. 2021. How array design creates SNP ascertainment bias. *PLoS ONE* 16:e0245178.
- Geist, D. R., and D. D. Dauble. 1998. Redd site selection and spawning habitat use by fall chinook salmon: the importance of geomorphic features in large rivers. *Environmental Management* 22:655–669.
- Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457–472.
- Gilpin, M. E., and M. E. Soule. 1986. Minimum viable populations: the processes of species extinction. Pages 19–34 *in* M. E. Soule, editor. *Conservation biology: the science of scarcity and diversity*. Sinauer, Sunderland, MA.
- Goudet, J. 2005. hierfstat, a package for *R* to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5:184–186.
- Goudet, J. 2014. hierfstat: estimation and tests of hierarchical F-statistics.
- Grant, G. C., B. Vondracek, and P. W. Sorensen. 2002. Spawning interactions between sympatric brown and brook trout may contribute to species replacement. *Transactions of the American Fisheries Society* 131:569–576.
- Haddad, N. M., L. A. Brudvig, J. Clobert, K. F. Davies, A. Gonzalez, R. D. Holt, T. E. Lovejoy, J. O. Sexton, M. P. Austin, C. D. Collins, W. M. Cook, E. I. Damschen, R. M. Ewers, B. L. Foster, C. N. Jenkins, A. J. King, W. F. Laurance, D. J. Levey, C. R. Margules, B. A. Melbourne, A. O. Nicholls, J. L. Orrock, D.-X. Song, and J. R. Townshend. 2015. Habitat fragmentation and its lasting impact on Earth’s ecosystems. *Science Advances* 1:e1500052.

- Hansen, T. F. 2013. Why epistasis is important for selection and adaptation. *Evolution* 67:3501–3511.
- Hargrove, J. S., C. A. Camacho, W. C. Schrader, J. H. Powell, T. A. Delomas, J. E. Hess, S. R. Narum, and M. R. Campbell. 2021. Parentage-based tagging improves escapement estimates for ESA-listed adult chinook salmon and steelhead in the Snake River basin. *Canadian Journal of Fisheries and Aquatic Sciences* 78:349–360.
- Harris, K., Y. Zhang, and R. Nielsen. 2019. Genetic rescue and the maintenance of native ancestry. *Conservation Genetics* 20:59–64.
- Hartman, K. J., and J. P. Hakala. 2006. Relationships between fine sediment and brook trout recruitment in forested headwater streams. *Journal of Freshwater Ecology* 21:215–230.
- Hasselgren, M., A. Angerbjörn, N. E. Eide, R. Erlandsson, Ø. Flagstad, A. Landa, J. Wallén, and K. Norén. 2018. Genetic rescue in an inbred Arctic fox (*Vulpes lagopus*) population. *Proceedings of the Royal Society B: Biological Sciences* 285:20172814.
- Heber, S., A. Varsani, S. Kuhn, A. Girg, B. Kempnaers, and J. Briskie. 2013. The genetic rescue of two bottlenecked South Island robin populations using translocations of inbred donors. *Proceedings of the Royal Society B Biological Sciences* 280:20122228.
- Hedrick, P. 2005, March 1. “Genetic restoration:” a more comprehensive perspective than “genetic rescue.” Elsevier Ltd.
- Hedrick, P. W., J. R. Adams, and J. A. Vucetich. 2011. Reevaluating and broadening the definition of genetic rescue. *Conservation Biology* 25:1069–1070.
- Hedrick, P. W., and R. Fredrickson. 2010. Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation Genetics* 11:615–626.
- Hedrick, P. W., and A. Garcia-Dorado. 2016. Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology and Evolution* 31:940–952.
- Hedrick, P. W., M. Kardos, R. O. Peterson, and J. A. Vucetich. 2016. Genomic variation of inbreeding and ancestry in the remaining two Isle Royale wolves. *Journal of Heredity* 108:esw083.
- Hedrick, P. W., J. A. Robinson, R. O. Peterson, and J. A. Vucetich. 2019. Genetics and extinction and the example of Isle Royale wolves. *Animal Conservation*:1–8.
- Hitt, N. P., E. L. Snook, and D. L. Massie. 2017. Brook trout use of thermal refugia and foraging habitat influenced by brown trout. *Canadian Journal of Fisheries and Aquatic Sciences* 74:406–418.
- Hogg, J. T., S. H. Forbes, B. M. Steele, and G. Luikart. 2006. Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society Series B* 273:1491–1499.
- Hudy, M., J. A. Coombs, K. H. Nislow, and B. H. Letcher. 2010. Dispersal and within-stream spatial population structure of brook trout revealed by pedigree reconstruction analysis. *Transactions of the American Fisheries Society* 139:1276–1287.
- Hudy, M., D. M. Downey, and D. W. Bowman. 2000. Successful restoration of an acidified native brook trout stream through mitigation with limestone sand. *North American Journal of Fisheries Management* 20:453–466.

- Hudy, M., T. M. Thieling, N. Gillespie, E. P. Smith, T. M. Theiling, N. Gillespie, and E. P. Smith. 2008. Distribution, status, and land use characteristics of subwatersheds within the native range of brook trout in the eastern United States. *North American Journal of Fisheries Management* 28:1069–1085.
- Huisman, J. 2017. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources* 17:1009–1024.
- Hunt, R. L. 1969. Overwinter survival of wild fingerling brook trout in Lawrence Creek, Wisconsin. *Journal of the Fisheries Research Board of Canada* 26:1473–1483.
- Huntsman, B. M., and J. T. Petty. 2014. Density-dependent regulation of brook trout population dynamics along a core-periphery distribution gradient in a central Appalachian watershed. *PLoS ONE* 9:e91673.
- Huntsman, B. M., J. T. Petty, S. Sharma, and E. R. Merriam. 2016. More than a corridor: use of a main stem stream as supplemental foraging habitat by a brook trout metapopulation. *Oecologia* 182:463–473.
- Isaak, D. J., M. K. Young, C. H. Luce, S. W. Hostetler, S. J. Wenger, E. E. Peterson, J. M. ver Hoef, M. C. Groce, D. L. Horan, and D. E. Nagel. 2016. Slow climate velocities of mountain streams portend their role as refugia for cold-water biodiversity. *Proceedings of the National Academy of Sciences* 113:4374–4379.
- Jamieson, I. G., and F. W. Allendorf. 2012. How does the 50/500 rule apply to MVPs? *Trends in Ecology and Evolution* 27:578–584.
- Jenkins, R. E., and N. M. Burkhead. 1994. *Freshwater fishes of Virginia*. American Fisheries Society, Bethesda, Maryland.
- Johnson, W. E., D. P. Onorato, M. E. Roelke, E. D. Land, M. Cunningham, R. C. Belden, R. McBride, D. Jansen, M. Lotz, D. Shindle, J. Howard, D. E. Wildt, L. M. Penfold, J. a Hostetler, M. K. Oli, and S. J. O'Brien. 2010. Genetic restoration of the Florida panther. *Science* 329:1641–1645.
- Jones, A. G., C. M. Small, K. A. Paczolt, and N. L. Ratterman. 2010. A practical guide to methods of parentage analysis. *Molecular Ecology Resources* 10:6–30.
- Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* 10:551–555.
- Jorde, P. E., and N. Ryman. 1995. Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics* 139.
- Kanno, Y., M. A. Kulp, and S. E. Moore. 2016a. Recovery of native brook trout populations following the eradication of nonnative rainbow trout in southern Appalachian Mountains streams. *North American Journal of Fisheries Management* 36:1325–1335.
- Kanno, Y., B. H. Letcher, N. P. Hitt, D. A. Boughton, J. E. B. Wofford, and E. F. Zipkin. 2015. Seasonal weather patterns drive population vital rates and persistence in a stream fish. *Global Change Biology* 21:1856–1870.
- Kanno, Y., K. C. Pregler, N. P. Hitt, B. H. Letcher, D. J. Hocking, and J. E. B. B. Wofford. 2016b. Seasonal temperature and precipitation regulate brook trout young-of-the-year abundance and population dynamics. *Freshwater Biology* 61:88–99.
- Kanno, Y., J. C. Vokoun, and B. H. Letcher. 2011. Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. *Molecular Ecology* 20:3711–3729.

- Kardos, M., A. Husby, S. E. McFarlane, A. Qvarnström, and H. Ellegren. 2016a. Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources* 16:727–741.
- Kardos, M., G. Luikart, and F. W. Allendorf. 2015. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity* 115:63–72.
- Kardos, M., H. R. Taylor, H. Ellegren, G. Luikart, and F. W. Allendorf. 2016b. Genomics advances the study of inbreeding depression in the wild. *Evolutionary applications* 9:1205–1218.
- Kazyak, D. C., R. H. Hilderbrand, T. L. King, S. R. Keller, and V. E. Chhatre. 2016. Hiding in plain sight: a case for cryptic metapopulations in brook trout (*Salvelinus fontinalis*). *PLoS ONE* 11:1–18.
- Kazyak, D. C., B. A. Lubinski, M. A. Kulp, K. C. Pregler, A. R. Whiteley, E. Hallerman, J. A. Coombs, Y. Kanno, J. M. Rash, R. P. Morgan, J. Habera, J. Henegar, T. C. Weathers, M. T. Sell, A. Rabern, D. Rankin, and T. L. King. 2022. Population genetics of brook trout in the southern Appalachian Mountains. *Transactions of the American Fisheries Society* 151:127–149.
- Kazyak, D., B. H. Letcher, J. Zydlewski, and M. J. O’Donnell. 2014. Growth variability of brook charr (*Salvelinus fontinalis*) in coastal Maine. *Ecology of Freshwater Fish* 23:516–526.
- Keller, L. F., and D. M. Waller. 2002. Inbreeding effects in wild populations. *Trends in Ecology & Evolution* 17:230–241.
- Kellner, K. 2021. jagsUI: A Wrapper Around “rjags” to Streamline “JAGS” Analyses.
- Kelly, E., and B. Phillips. 2019. How many and when? Optimising targeted gene flow for a step change in the environment. *Ecology Letters* 22:447–457.
- van de Kerk, M., D. P. Onorato, J. A. Hostetler, B. M. Bolker, and M. K. Oli. 2019. Dynamics, persistence, and genetic management of the endangered Florida panther population. *Wildlife Monographs* 203:3–35.
- King, T. L., M. S. Eackles, and B. H. Letcher. 2005. Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. *Molecular Ecology Notes* 5:130–132.
- King, T. L., B. A. Lubinski, M. K. Burnham-Curtis, W. Stott, and R. P. Morgan II. 2012. Tools for the management and conservation of genetic diversity in brook trout (*Salvelinus fontinalis*): tri- and tetranucleotide microsatellite markers for the assessment of genetic diversity, phylogeography, and historical demographics. *Conservation Genetics Resources* 4:539–543.
- Koch, I. J., and S. R. Narum. 2021, August 1. An evaluation of the potential factors affecting lifetime reproductive success in salmonids. John Wiley and Sons Inc.
- Kovach, R. P., J. B. Armstrong, D. A. Schmetterling, R. Al-Chokhachy, and C. C. Muhlfeld. 2018. Long-term population dynamics and conservation risk of migratory bull trout in the upper Columbia River basin. *Canadian Journal of Fisheries and Aquatic Sciences* 75:1960–1968.
- Kruuk, L. E. B., and W. G. Hill. 2008. Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. *Proceedings of the Royal Society B: Biological Sciences* 275:593–596.

- Kuo, L., and B. Mallick. 1998. Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 60:65–81.
- Laikre, L. 2020. Post-2020 goals overlook genetic diversity. *Science* 367:1083–1085.
- Lande, R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *American Naturalist* 142:911–927.
- Larson, G. L., and S. E. Moore. 1985. Encroachment of exotic rainbow trout into stream populations of native brook trout in the southern Appalachian Mountains. *Transactions of the American Fisheries Society* 114:195–203.
- Lebreton, J. D., J. D. Nichols, R. J. Barker, R. Pradel, and J. A. Spendelov. 2009, January 1. Chapter 3 modeling individual animal histories with multistate capture-recapture models. Academic Press Inc.
- Lemopoulos, A., J. M. Prokkola, S. Uusi-Heikkilä, A. Vasemägi, A. Huusko, P. Hyvärinen, M. L. Koljonen, J. Koskiniemi, and A. Vainikka. 2019. Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. *Ecology and Evolution* 9:2106–2120.
- Lenormand, T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* 17:183–189.
- Letcher, B. H., K. H. Nislow, J. A. Coombs, M. J. O’Donnell, T. L. Dubreuil, and E. Svensson. 2007. Population response to habitat fragmentation in a stream-dwelling brook trout population. *PLoS ONE* 2:e1139.
- Letcher, B. H., P. Schueller, R. D. Bassar, K. H. Nislow, J. A. Coombs, K. Sakrejda, M. Morrissey, D. B. Sigourney, A. R. Whiteley, M. J. O’Donnell, and T. L. Dubreuil. 2014. Robust estimates of environmental effects on population vital rates: an integrated capture–recapture model of seasonal brook trout growth, survival and movement in a stream network. *Journal of Animal Ecology* 84:337–352.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Liermann, C. R., C. Nilsson, J. Robertson, and R. Y. Ng. 2012. Implications of Dam Obstruction for Global Freshwater Fish Diversity. *BioScience* 62:539–548.
- Linke, S., E. Turak, and J. Nel. 2011. Freshwater conservation planning: the case for systematic approaches. *Freshwater Biology* 56:6–20.
- Lowe, W. H., R. P. Kovach, and F. W. Allendorf. 2017. Population genetics and demography unite ecology and evolution. *Trends in Ecology & Evolution* 32:141–152.
- Luikart, G., T. Antao, B. K. Hand, C. C. Muhlfeld, M. C. Boyer, T. Cosart, B. Trethewey, R. Al-Chockhachy, and R. S. Waples. 2021. Detecting population declines via monitoring the effective number of breeders ( $N_b$ ). *Molecular Ecology Resources* 21:379–393.
- Luikart, G., N. Ryman, D. A. Tallmon, M. K. Schwartz, and F. W. Allendorf. 2010. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches.
- Lynch, A. J., S. J. Cooke, A. M. Deines, S. D. Bower, D. B. Bunnell, I. G. Cowx, V. M. Nguyen, J. Nohner, K. Phouthavong, B. Riley, M. W. Rogers, W. W. Taylor, W.

- Woelmer, S. J. Youn, and T. D. Beard. 2016. The social, economic, and environmental importance of inland fish and fisheries. Canadian Science Publishing.
- Macqueen, D. J., and I. A. Johnston. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences* 281.
- Madsen, T., R. Shine, M. Olsson, and H. Wittzell. 1999. Restoration of an inbred adder population. *Nature* 402:34–35.
- Magoc, T., and S. L. Salzberg. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963.
- Mamoozadeh, N., A. R. Whiteley, B. H. Letcher, D. C. Kazyak, C. Tarsa, and M. H. Meek. (n.d.). Evaluating genomic relationships across spatial and temporal scales to guide conservation and management of imperiled species. *Molecular Ecology Resources*.
- Mazerolle, M. J. 2019. AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). <https://cran.r-project.org/package=AICcmodavg>.
- McKinney, G. J., R. K. Waples, L. W. Seeb, and J. E. Seeb. 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources* 17:656–669.
- Merriam, E. R., J. T. Petty, and J. Clingerman. 2019. Conservation planning at the intersection of landscape and climate change: brook trout in the Chesapeake Bay watershed. *Ecosphere* 10:e02585.
- Miller, L. M., D. J. Dieterman, and R. J. H. Hoxmeier. 2019. Reproductive dynamics of a native brook trout population following removal of non-native brown trout from a stream in Minnesota, north-central USA. *Hydrobiologia* 840:49–61.
- Mills, L. S., and F. W. Allendorf. 1996. The one-migrant-per-generation rule in conservation and management. *Conservation Biology* 10:1509–1518.
- Mills, L. S., J. J. Citta, K. P. Lair, M. K. Schwartz, D. A. Tallmon, M. K. Schwarz, and D. A. Tallmon. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. *Ecological Applications* 10:283–294.
- Mobley, K. B., H. Granroth-Wilding, M. Ellmen, J. P. Vähä, T. Aykanat, S. E. Johnston, P. Orell, J. Erkinaro, and C. R. Primmer. 2019. Home ground advantage: local Atlantic salmon have higher reproductive fitness than dispersers in the wild. *Science Advances* 5:eaav1112.
- Moran, P. A. P. 1953. The statistical analysis of the Canadian Lynx Cycle: II. Synchronization and meteorology. *Australian Journal of Zoology* 1:291–298.
- Morita, K., and S. Yamamoto. 2002. Effects of habitat fragmentation by damming on the persistence of stream-dwelling charr populations. *Conservation Biology* 16:1318–1323.
- Mottl, O., J. Yombai, T. M. Fayle, V. Novotný, and P. Klimeš. 2020. Experiments with artificial nests provide evidence for ant community stratification and nest site limitation in a tropical forest. *Biotropica* 52:277–287.
- Muhlfeld, C. C., S. T. Kalinowski, T. E. McMahon, M. L. Taper, S. Painter, R. F. Leary, and F. W. Allendorf. 2009. Hybridization rapidly reduces fitness of a native trout in the wild. *Conservation Biology* 5:328–331.

- NADP. 2022. National Atmospheric Deposition Program.  
<https://nadp.slh.wisc.edu/committees/tdep/>.
- Neel, M. C., K. McKelvey, N. Ryman, M. W. Lloyd, R. Short Bull, F. W. Allendorf, M. K. Schwartz, and R. S. Waples. 2013. Estimation of effective population size in continuously distributed populations: there goes the neighborhood. *Heredity* 111:189–199.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia university press.
- Ng, T. 2019. *microhaplot: microhaplotype constructor and visualizer*.
- Nislow, K. H., and W. H. Lowe. 2003. Influences of logging history and stream pH on brook trout abundance in first-order streams in New Hampshire. *Transactions of the American Fisheries Society* 132:166–171.
- Nuhfer, A. J., T. G. Zorn, and T. C. Wills. 2017. Effects of reduced summer flows on the brook trout population and temperatures of a groundwater-influenced stream. *Ecology of Freshwater Fish* 26:108–119.
- O’Grady, J. J., B. W. Brook, D. H. Reed, J. D. Ballou, D. W. Tonkyn, and R. Frankham. 2006. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation* 133:42–51.
- O’Hara, R. B., and M. J. Sillanpää. 2009. A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis* 4:85–118.
- Olden, J. D., M. E. Douglas, and M. R. Douglas. 2005. The human dimensions of biotic homogenization. *Conservation Biology* 19:2036–2038.
- Oliehoek, P. A., and P. Bijma. 2009. Effects of pedigree errors on the efficiency of conservation decisions. *Genetics Selection Evolution* 41:9.
- Oppel, S., G. Hilton, N. Ratcliffe, C. Fenton, J. Daley, G. Gray, J. Vickery, and D. Gibbons. 2014. Assessing population viability while accounting for demographic and environmental uncertainty. *Ecology* 95:1809–1818.
- Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs*:3–135.
- Peacock, M. M., and N. A. Dochtermann. 2012. Evolutionary potential but not extinction risk of Lahontan cutthroat trout (*Oncorhynchus clarkii henshawi*) is associated with stream characteristics. *Canadian Journal of Fisheries and Aquatic Sciences* 69:615–626.
- Pemberton, J. M. 2008, March 22. *Wild pedigrees: the way forward*. Royal Society.
- Petty, J. T., P. J. Lamothe, and P. M. Mazik. 2005. Spatial and seasonal dynamics of brook trout populations inhabiting a central Appalachian watershed. *Transactions of the American Fisheries Society* 134:572–587.
- Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Page Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.
- Poirier, M., D. W. Coltman, F. Pelletier, J. Jorgenson, and M. Festa-Bianchet. 2019. Genetic decline, restoration and rescue of an isolated ungulate population. *Evolutionary Applications* 12:1318–1328.
- Prince, D. J., S. M. O’Rourke, T. Q. Thompson, O. A. Ali, H. S. Lyman, I. K. Saglam, T. J. Hotaling, A. P. Spidle, and M. R. Miller. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances* 3:e1603198.

- Pritchard, J. K., and A. di Rienzo. 2010. Adaptation - Not by sweeps alone. *Nature Reviews Genetics* 11:665–667.
- QGIS Development Team. 2020. QGIS.
- R Core Team. 2017. *R: a language and environment for statistical computing*. Vienna, Austria.
- R Core Team. 2020. *R: a language and environment for statistical computing*. Vienna, Austria.
- Ralls, K., J. D. Ballou, M. R. Dudash, M. D. B. Eldridge, C. B. Fenster, R. C. Lacy, P. Sunnucks, and R. Frankham. 2018. Call for a paradigm shift in the genetic management of fragmented populations. *Conservation Letters* 11:e12412.
- Reid, J. M., L. F. Keller, A. B. Marr, P. Nietlisbach, R. J. Sardell, and P. Arcese. 2014. Pedigree error due to extra-pair reproduction substantially biases estimates of inbreeding depression. *Evolution* 68:802–815.
- Rice, K. C., and J. D. Jastram. 2015. Rising air and stream-water temperatures in Chesapeake Bay region, USA. *Climatic Change* 128:127–138.
- Robinson, Z. L., J. A. Coombs, M. Hudy, K. H. Nislow, B. H. Letcher, and A. R. Whiteley. 2017. Experimental test of genetic rescue in isolated populations of brook trout. *Molecular Ecology* 26:4418–4433.
- Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* 28:4737–4754.
- Rodriguez-Ramilo, S. T., and J. Wang. 2012. The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. *Molecular Ecology Resources* 12:873–884.
- Rouseset, F. 2008. GENEPOP '007 : a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology* 8:103–106.
- Ruzzante, D. E., G. R. McCracken, S. Parmelee, K. Hill, A. Corrigan, J. MacMillan, and S. J. Walde. 2016. Effective number of breeders, effective population size and their relationship with census size in an iteroparous species, *salvelinus fontinalis*. *Proceedings of the Royal Society B: Biological Sciences* 283.
- Saccheri, I., M. Kuussaari, M. Kankare, P. Vikman, W. Fortelius, and I. Hanski. 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature* 392:491–494.
- Sala, O. E., F. S. Chapin, J. J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L. F. Huenneke, R. B. Jackson, A. Kinzig, R. Leemans, D. M. Lodge, H. A. Mooney, M. Oesterheld, N. L. R. Poff, M. T. Sykes, B. H. Walker, M. Walker, and D. H. Wall. 2000, March 10. Global biodiversity scenarios for the year 2100. American Association for the Advancement of Science.
- Saunders, J. W., and M. W. Smith. 1965. Changes in a stream population of trout associated with increased silt. *Journal of the Fisheries Research Board of Canada* 22:395–404.
- Schaub, M., and M. Kery. 2021. *Integrated population models: theory and ecological applications with R and JAGS*. Academic Press.
- Schofield, C. L., and J. R. Trojnar. 1980. Aluminum toxicity to brook trout (*Salvelinus fontinalis*) in acidified waters. Pages 341–366 *Polluted Rain*. Springer US.
- Schwartz, M. K., G. Luikart, and R. S. Waples. 2007, January 1. Genetic monitoring as a promising tool for conservation and management. Elsevier *Current Trends*.

- Sinnwell, J. P., T. M. Therneau, and D. J. Schaid. 2014. The kinship2 R package for pedigree data. *Human Heredity* 78:91–93.
- Slate, J., and J. M. Pemberton. 2007. Admixture and patterns of linkage disequilibrium in a free-living vertebrate population. *Journal of Evolutionary Biology* 20:1415–1427.
- Steele, C. A., E. C. Anderson, M. W. Ackerman, M. A. Hess, N. R. Campbell, S. R. Narum, and M. R. Campbell. 2013. A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. *Canadian Journal of Fisheries and Aquatic Sciences* 70:1046–1054.
- Stenlkk, K., M. Saitou, L. Rud-Johansen, T. Nome, M. Moser, M. rnyasi, M. Kent, N. J. Barson, and S. Lien. 2022. The emergence of supergenes from inversions in Atlantic salmon. *Philosophical Transactions of the Royal Society B: Biological Sciences* 377.
- Stoffel, M. A., S. E. Johnston, J. G. Pilkington, and J. M. Pemberton. 2021. Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. *Nature Communications* 12:1–10.
- Su, Y., and Y. Masanao. 2015. R2jags: using R to run “JAGS.”
- Tallmon, D. A. 2017. Get a move on: the value of rescue. *Animal Conservation* 20:16–17.
- Tallmon, D. A., G. Luikart, and R. S. Waples. 2004. The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution* 19:489–496.
- Tourvas, N. 2022. PopGenUtils: a collection of useful functions to deal with genetic data in R.
- Trumbo, B. A., K. H. Nislow, J. Stallings, M. Hudy, E. P. Smith, D.-Y. Kim, B. Wiggins, and C. A. Dolloff. 2014. Ranking site vulnerability to increasing temperatures in southern Appalachian brook trout streams in Virginia: an exposure-sensitivity approach. *Transactions of the American Fisheries Society* 143:173–187.
- USGS. 2012. National Hydrography Dataset Plus. U.S. Dept. of the Interior, U.S. Geological Survey, Reston, VA.
- Vergeer, P., E. Sonderen, and N. J. Ouborg. 2004. Introduction strategies put to the test: local adaptation versus heterosis. *Conservation Biology* 18:812–821.
- Vil, C., A.-K. Sundqvist, . Flagstad, J. Seddon, S. B. Rnerfeldt, I. Kojola, A. Casulli, H. Sand, P. Wabakken, H. Ellegren, S. Bjrnerfeldt, I. Kojola, A. Casulli, H. Sand, P. Wabakken, and H. Ellegren. 2003. Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270:91–97.
- Waits, L. P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology* 10:249–256.
- Walker, J. D., B. H. Letcher, and D. J. Hocking. 2021. SHEDS Northeast brook trout occupancy model.
- Waller, D. M. 2015. Genetic rescue: a safe or risky bet? *Molecular Ecology* 24:2595–2597.
- Wang, J. 2007. Parentage and sibship exclusions: higher statistical power with more family members. *Heredity* 99:205–217.
- Wang, J. 2012. Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics* 191:183–194.

- Wang, J., and A. W. Santure. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics* 181:1579–1594.
- Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics* 7:167–184.
- Waples, R. S., and F. Allendorf. 2015. Testing for hardy-weinberg proportions: have we lost the plot?
- Waples, R. S., and T. Antao. 2014. Intermittent breeding and constraints on litter size: consequences for effective population size per generation ( $N_e$ ) and per reproductive cycle ( $N_b$ ). *Evolution* 68:1722–1734.
- Waples, R. S., T. Antao, and G. Luikart. 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics* 197:769–780.
- Waples, R. S., and C. Do. 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources* 8:753–756.
- Waples, R. S., and C. Do. 2010. Linkage disequilibrium estimates of contemporary  $N_e$  using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications* 3:244–262.
- Waples, R. S., C. Do, and J. Chopelet. 2011. Calculating  $N_e$  and  $N_e/N$  in age-structured populations: a hybrid Felsenstein-Hill approach. *Ecology*.
- Waples, R. S., and P. R. England. 2011. Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics* 189:633–644.
- Waples, R. S., G. Luikart, J. R. Faulkner, and D. A. Tallmon. 2013. Simple life-history traits explain key effective population size ratios across diverse taxa. *Proceedings of the Royal Society Series B* 280:20131339.
- Waples, R. S., and R. K. Waples. 2011. Inbreeding effective population size and parentage analysis without parents. *Molecular Ecology Resources* 11:162–171.
- Waples, R. S., and M. Yokota. 2007. Temporal estimates of effective population size in species with overlapping generations. *Genetics*.
- Warren, D. R., J. M. Robinson, D. C. Josephson, D. R. Sheldon, and C. E. Kraft. 2012. Elevated summer temperatures delay spawning and reduce redd construction for resident brook trout (*Salvelinus fontinalis*). *Global Change Biology* 18:1804–1811.
- Weber, N., N. Bouwes, M. M. Pollock, C. Volk, J. M. Wheaton, G. Wathen, J. Wirtz, and C. E. Jordan. 2017. Alteration of stream temperature by natural and artificial beaver dams. *PLOS ONE* 12:e0176313.
- Weeks, A. R., D. Heinze, L. Perrin, J. Stoklosa, A. A. Hoffmann, A. van Rooyen, T. Kelly, and I. Mansergh. 2017. Genetic rescue increases fitness and aids rapid recovery of an endangered marsupial population. *Nature Communications* 8:1071.
- Welsh, A. H., R. B. Cunningham, C. F. Donnelly, and D. B. Lindenmayer. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling* 88:297–308.
- Westemeier, R. L., J. D. Brawn, S. A. Simpson, T. L. Esker, R. W. Jansen, J. W. Walk, E. L. Kershner, J. L. Bouzat, and K. N. Paige. 1998. Tracking the long-term decline and recovery of an isolated population. *Science* 282:1695–1698.
- Whiteley, A. R., J. A. Coombs, M. Cembrola, M. J. O'Donnell, M. Hudy, K. H. Nislow, and B. H. Letcher. 2015a. Effective number of breeders provides a link between

- interannual variation in stream flow and individual reproductive contribution in a stream salmonid. *Molecular Ecology* 24:3585–3602.
- Whiteley, A. R., J. A. Coombs, M. H. Hudy, Z. L. Robinson, K. H. Nislow, and B. H. Letcher. 2012. Sampling strategies for estimating brook trout effective population size. *Conservation Genetics* 13:625–637.
- Whiteley, A. R., J. A. Coombs, M. Hudy, Z. Robinson, A. R. Colton, K. H. Nislow, and B. H. Letcher. 2013. Fragmentation and patch size shape genetic structure of brook trout populations. *Canadian Journal of Fisheries and Aquatic Sciences* 70:678–688.
- Whiteley, A. R., J. A. Coombs, B. H. Letcher, and K. H. Nislow. 2014a. Simulation and empirical analysis of novel sibship-based genetic determination of fish passage. *Canadian Journal of Fisheries and Aquatic Sciences* 71:1667–1679.
- Whiteley, A. R., J. A. Coombs, M. J. O'Donnell, K. H. Nislow, and B. H. Letcher. 2017. Keeping things local: subpopulation  $N_b$  and  $N_e$  in a stream network with partial barriers to fish migration. *Evolutionary Applications* 10:348–365.
- Whiteley, A. R., S. W. Fitzpatrick, W. C. Funk, and D. A. Tallmon. 2015b. Genetic rescue to the rescue. *Trends in Ecology & Evolution* 30:42–49.
- Whiteley, A. R., M. Hudy, Z. L. Robinson, J. A. Coombs, and K. H. Nislow. 2014b. Patch-based metrics: a cost effective method for short- and long-term monitoring of EBTJV wild brook trout populations? Pages 170–179 in R. F. Carline and C. LoSapio, editors. *Proceedings of the Wild Trout XI Symposium*. West Yellowstone, Montana.
- Whitlock, M. C., P. K. Ingvarsson, and T. Hatfield. 2000. Local drift load and the heterosis of interconnected populations. *Heredity* 84:452.
- Willi, Y., J. Van Buskirk, A. A. Hoffmann, J. Van Buskirk, and A. A. Hoffmann. 2006. Limits to the adaptive potential of small populations. *Annual Review of Ecology Evolution and Systematics* 37:433–458.
- Wilson, A. J., and M. M. Ferguson. 2002. *Molecular pedigree analysis in natural populations of fishes: approaches, applications, and practical considerations*. NRC Research Press Ottawa, Canada .
- Wisdom, M. J., L. S. Mills, and D. F. Doak. 2000. Life stage simulation analysis: estimating vital-rate effects on population growth for conservation. *Ecology* 81:628–641.
- Wood, D. M., A. B. Welsh, and J. Todd Petty. 2018. Genetic assignment of brook trout reveals rapid success of culvert restoration in headwater streams. *North American Journal of Fisheries Management* 38:991–1003.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.
- Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science* 87.
- Wright, Sewall. 1946. Isolation by distance under diverse systems of mating. *Genetics* 31:39–59.
- Wydoski, R. S., and E. L. Cooper. 1966. Maturation and fecundity of brook trout from infertile streams. *Journal of the Fisheries Research Board of Canada* 23:623–649.
- Yano, A., B. Nicol, E. Jouanno, E. Quillet, A. Fostier, R. Guyomard, and Y. Guiguen. 2013. The sexually dimorphic on the Y-chromosome gene (*sdY*) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications* 6:486–496.

- Yates, M. C., T. A. Bernos, and D. J. Fraser. 2017. A critical assessment of estimating census population size from genetic population size (or vice versa) in three fishes. *Evolutionary Applications* 10:935–945.
- Zhang, H., T. Thieling, S. C. B. Prins, E. P. Smith, and M. Hudy. 2008. Model-based clustering in a brook trout classification study within the eastern United States. *Transactions of the American Fisheries Society*. 137:841.
- Zorn, T. G., and A. J. Nuhfer. 2007. Regional synchrony of brown trout and brook trout population dynamics among Michigan rivers. *Transactions of the American Fisheries Society* 136:706–717.

## APPENDIX A. Chapter 2 Supplementary Materials

### Testing for Hardy-Weinberg proportions and Linkage disequilibrium

We tested all populations for which the GTseq genetic marker panel was used for deviations from Hardy-Weinberg proportions (HWP) and for Linkage disequilibrium (LD) in the program GENEPOP version 4.7.5. Single cohort samples, such as the ones used in this study, have been shown to be out of global HWP and exhibit LD due to strong family structure in brook trout populations (Whiteley et al. 2013). As a result, we reject the null hypothesis these samples are in HWP or without linkage disequilibrium a priori, and instead use these tests to remove problematic loci. It's worth highlighting, that we are using the LD signal of family structure and genetic drift to estimate  $N_b$ , and with sufficient statistical power this will inevitably generate significant tests for LD. We filtered the marker panel using a binomial distribution approach for multiple testing across populations and markers similar to the approach demonstrated in Waples and Allendorf (2015). The goal of this testing is to remove loci that are likely physically linked, have null alleles present, or represent paralogous loci.

First, we performed tests for HW proportions for each of the 240 loci in  $n = 57$  populations. In total, we conducted 9249 exact tests for HWP and 1034 (11%) were significant at  $\alpha = 0.05$ . We filtered based on an expected probability of significant tests of  $p = 0.25$ , rather than the traditional Type-1 error rate of 0.05 to account for elevated number of expected significant tests due to family structure. Assuming a locus can be tested in all populations (i.e., polymorphic in all populations), we expect on average to observe 14 significant tests ( $n \times p$ ), and we would expect less than 5% of all loci to exhibit more than 20 significant tests. We removed all loci that had more significant tests more than the 5% threshold (20 significant tests) under the assumption of  $p = 0.25$ . Note that we based the threshold of significant tests upon the actual number of tests conducted rather than maximum number of tests possible (i.e., 57).

Of the possible 1634 760 LD tests assuming all loci were polymorphic, 917,633 LD tests were conducted. 160,264 (8.7%) of the tests were statistically significant at  $\alpha = 0.05$ . We again filtered based on an expected probability of significant tests of  $p = 0.25$ , which would correspond to an average of 17 expected significant tests per locus-pair

comparison assuming the test was possible in each of the 57 populations. Under the expectation of  $p = 0.25$  for significant tests, less than 5% of locus-pair comparisons will exceed 20 significant tests out of 57 total tests. We iteratively dropped loci until all locus-pair comparisons had a proportion of significant tests no more than 20/57. In total, we removed 63 loci from consideration due to LD testing and 10 due to HWP testing. Following filtering based upon HWP and LD, we retained 167 SNP-loci for subsequent analysis.

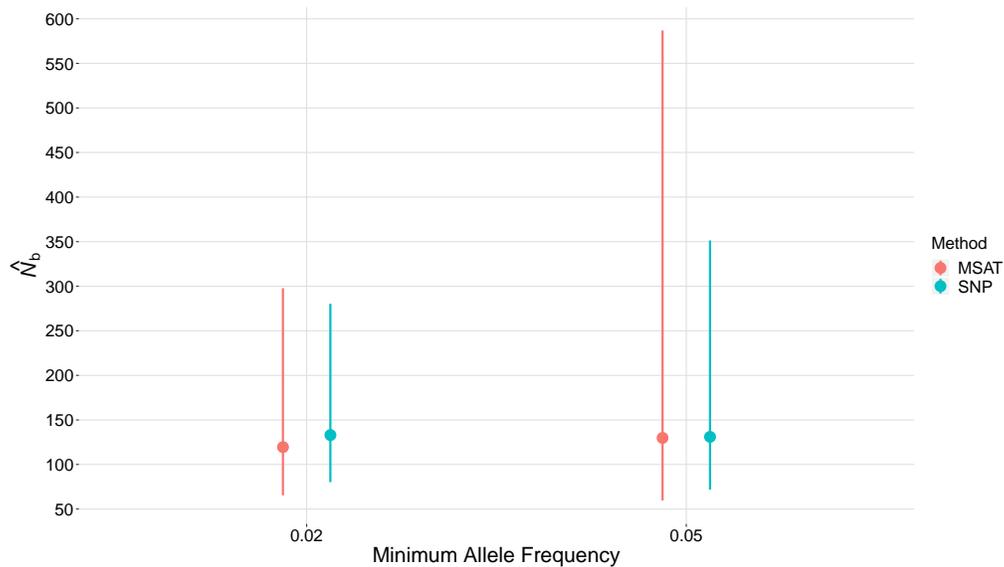


Fig. S2-1. Comparison of effective number of breeder estimates based on the GT-seq marker panel (SNP), and those based on microsatellites (MSAT) for North River, Virginia brook trout habitat patch. Two different critical allele frequencies (0.02 and 0.05) were used within the program NeEstimator V2.1.

## APPENDIX B. Chapter 3 Supplementary Materials

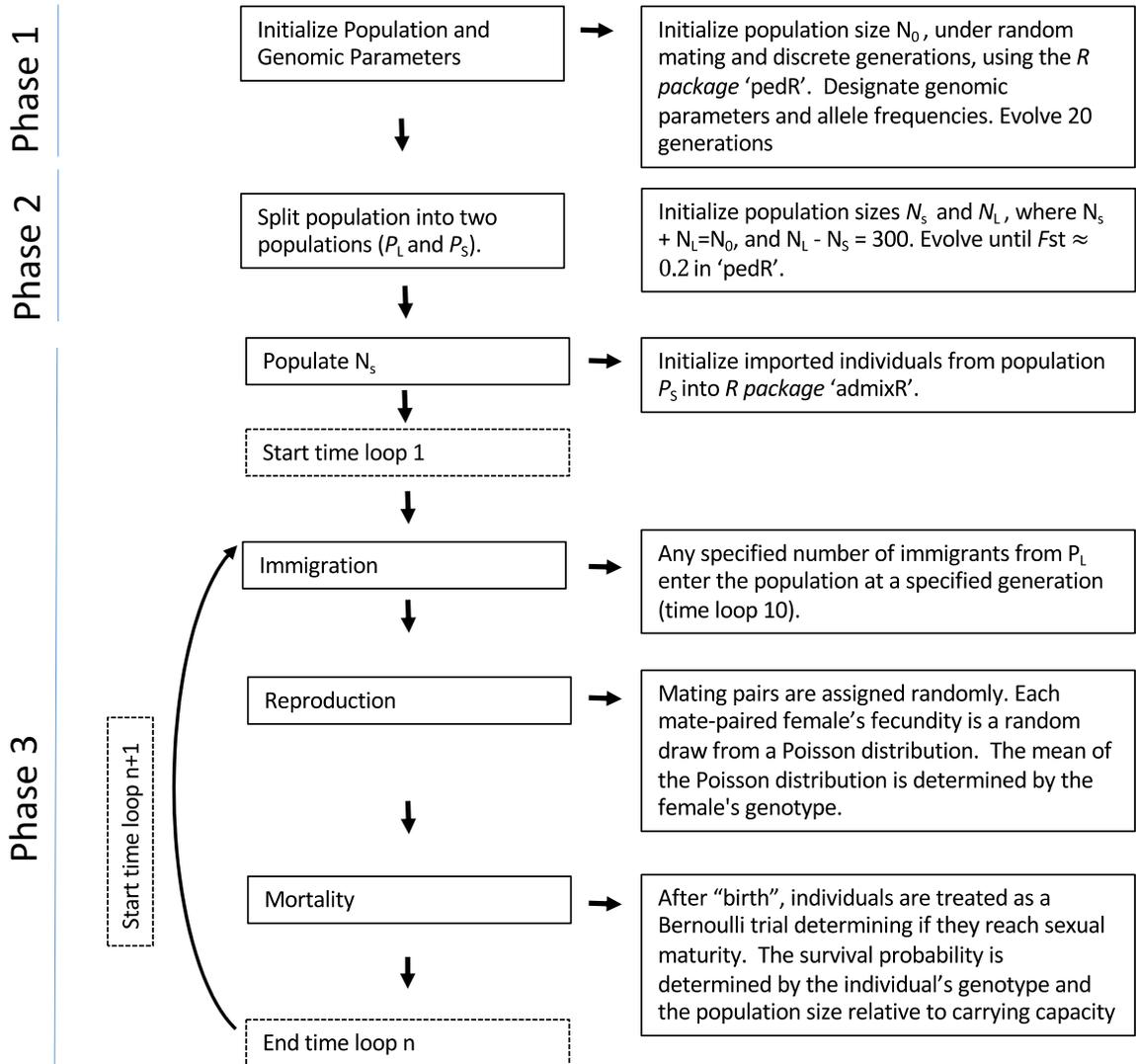
Table S3-1. Table of descriptions of all simulations conducted. The number of repetitions (reps) shown reflect the number of simulation repetitions for each combination of the parameters reported in each row of the table. Carrying Capacity ( $K$ ) and initial abundance ( $N_0$ ) are reported. The additional scenarios are described above.

$K$	$N_0$	Gene Action	Fitness Effects	Intrinsic Growth Rate	Immigration Rate	reps	Additional Scenario
50	25	Additive	-0.8, -0.4, -0.1, 0	0.98	0, 4, 12	500	–
50	25	Dominance	0.1, 0.4, 0.8	0.98	0, 4, 12	500	–
50	50	Additive	-0.8, -0.4, -0.1, 0	1.02	0, 4, 12	500	–
50	50	Dominance	0.1, 0.4, 0.8	1.02	0, 4, 12	500	–
150	75	Additive	-0.8, -0.4, -0.1, 0, 0.1, 0.4, 0.8	0.98	0, 12, 36	500	–
150	75	Dominance	-0.8, -0.4, -0.1, 0, 0.1, 0.4, 0.8	0.98	0, 12, 36	500	–
150	75	Additive	-0.8, -0.4, -0.1	0.98	36	250	Heterosis
150	75	Additive	-0.8, -0.4, -0.1	0.98	36	250	Single Locus
150	75	Dominance	0.1, 0.4, 0.8	0.98	36	250	Single Locus
150	75	Additive	-0.8, -0.4, -0.1	0.98	36	250	Assortative Mating
150	75	Dominance	0.1, 0.4, 0.8	0.98	36	250	Assortative Mating
150	75	Additive	-0.8, -0.4, -0.1	0.98	36	250	Denser Markers
150	75	Dominance	0.1, 0.4, 0.8	0.98	36	250	Denser Markers

Table S3-2. Table of published genetic rescue effect sizes. Genetic rescue publications using similar methods to identify the fitness effects of gene flow. This table provides examples of effect sizes and is not intended to be comprehensive. See (Frankham 2015) for a more comprehensive treatment of the magnitude of genetic rescue effects.

Citation	Species	Metrics	Effect sizes
(Westemeier et al. 1998)	<i>Tympanuchus cupido</i>	Early survival (hatch rate)	23.68% increase in hatching rate
(Madsen et al. 1999)	<i>Vipera berus</i>	Male $N_c$	700.0% increase in male $N_c$
(Vilà et al. 2003)	<i>Canis lupus</i>	$N_c$	~ 800% increase in $N_c$
(Hogg et al. 2006)	<i>Ovis canadensis</i>	Annual Reproductive Success	220% increase in female annual reproductive success
(Bossuyt 2007)	<i>Parnassia palustris</i>	Reproductive failure	71.42% increase in viable seeds in inter-metapopulation crosses
(Johnson et al. 2010)	<i>Puma concolor coryi</i>	Juvenile Survival, $N_c$	113.2% increase in kitten survival, 14% increase in abundance per year
(Finger et al. 2011)	<i>Medusagyne oppositifolia</i>	Seed viability	21.5% increase in between population crosses
(Heber et al. 2013)	<i>Petroica australis</i>	Juvenile survival, recruitment	162.1% increase in outbred offspring, 61.0% increase in outbred offspring
(Åkesson et al. 2016)	<i>Canis lupus</i>	Breeding Success	150.0% increase in outbred offspring
(Weeks et al. 2017)	<i>Burramys parvus</i>	Female survival	55.5 % increase in female survival
(Hasselgren et al. 2018)	<i>Vulpes lagopus</i>	Juvenile Survival, Breeding Success	190% increase in juvenile survival and an 130% increase in breeding success
(Poirier et al. 2019)	<i>Ovis canadensis</i>	Juvenile survival	28.3% increase

Fig S3-1. Flow diagram of simulation model.



Plots of temporal patterns in migrant ancestry, heterozygosity, and abundance for scenarios not presented in Figure 3-2

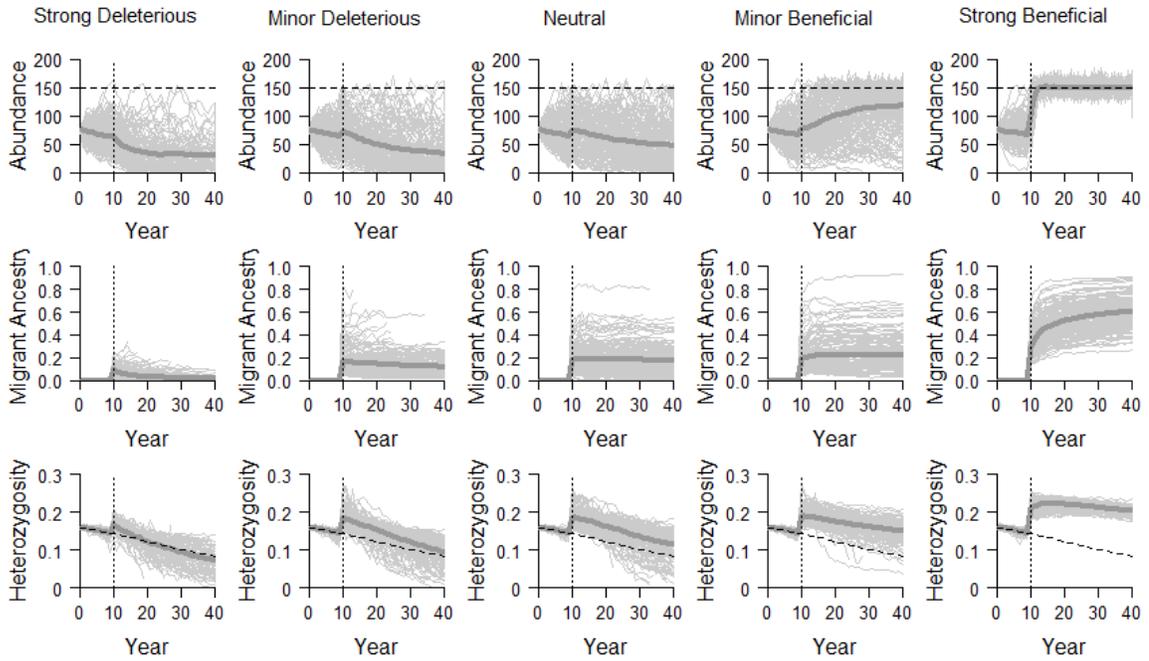


Fig. S3-2. Trends in abundance, migrant ancestry and observed heterozygosity with differing fitness effects of gene flow. The baseline scenario with  $K = 150$ ,  $N_0 = 75$ , and 12 immigrants is shown. Columns represent fitness effects of  $-0.8$ ,  $-0.1$ ,  $0$ ,  $0.1$ , and  $0.8$ . Dark bold grey lines are the mean, and light thin lines are 100 randomly selected iterations of the simulation. The dotted vertical line at year 10 represents the single gene flow event. The dashed lines represent  $K$  in the abundance figures, and the dashed trend line in the heterozygosity figures represents the pre-gene flow trajectory.

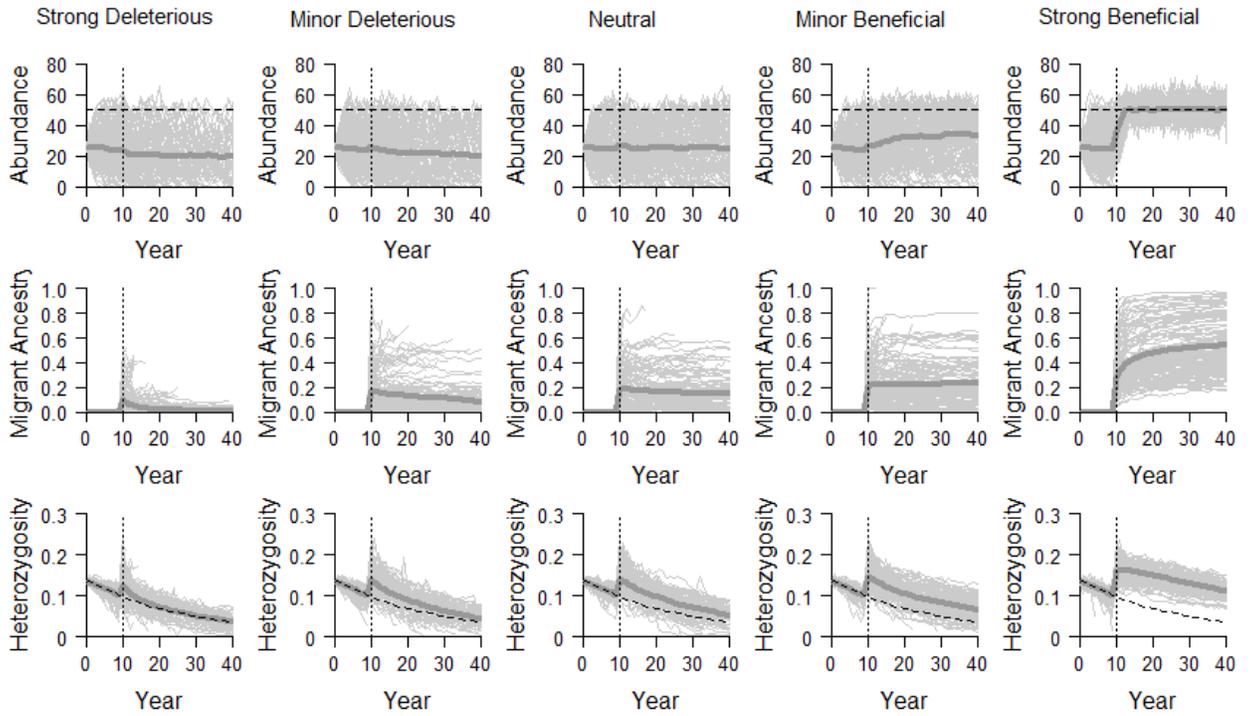


Fig. S3-3. Trends in abundance, migrant ancestry and observed heterozygosity with differing fitness effects of gene flow. The baseline scenario with  $K = 50$ ,  $N_0 = 25$ , and 4 immigrants is shown. Columns represent fitness effects of  $-0.8$ ,  $-0.1$ ,  $0$ ,  $0.1$ , and  $0.8$ . Dark bold grey lines are the mean, and light thin lines are 100 randomly selected iterations of the simulation. The dotted vertical line at year 10 represents the single gene flow event. The dashed lines represent  $K$  in the abundance figures, and the dashed trend line in the heterozygosity figures represents the pre-gene flow trajectory.

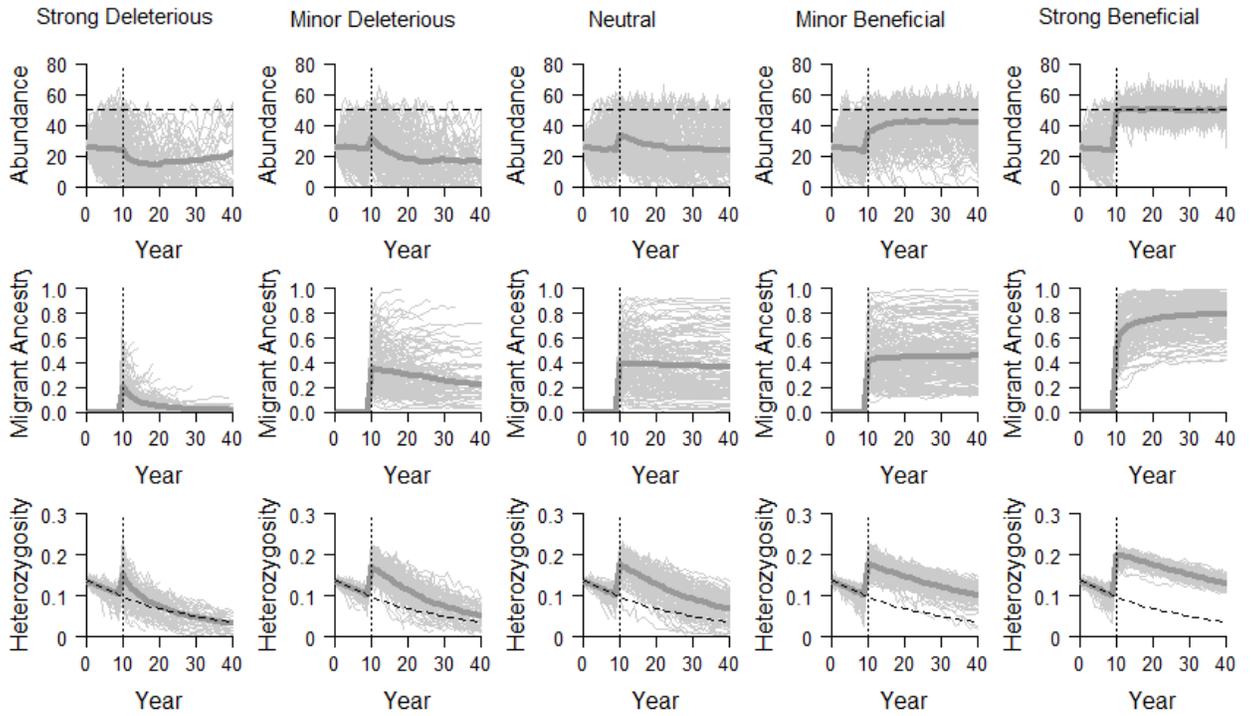


Fig. S3-4. Trends in abundance, migrant ancestry and observed heterozygosity with differing fitness effects of gene flow. The baseline scenario with  $K = 50$ ,  $N_0 = 25$ , and 12 immigrants is shown. Columns represent fitness effects of  $-0.8$ ,  $-0.1$ ,  $0$ ,  $0.1$ , and  $0.8$ . Dark bold grey lines are the mean, and light thin lines are 100 randomly selected iterations of the simulation. The dotted vertical line at year 10 represents the single gene flow event. The dashed lines represent  $K$  in the abundance figures, and the dashed trend line in the heterozygosity figures represents the pre-gene flow trajectory.

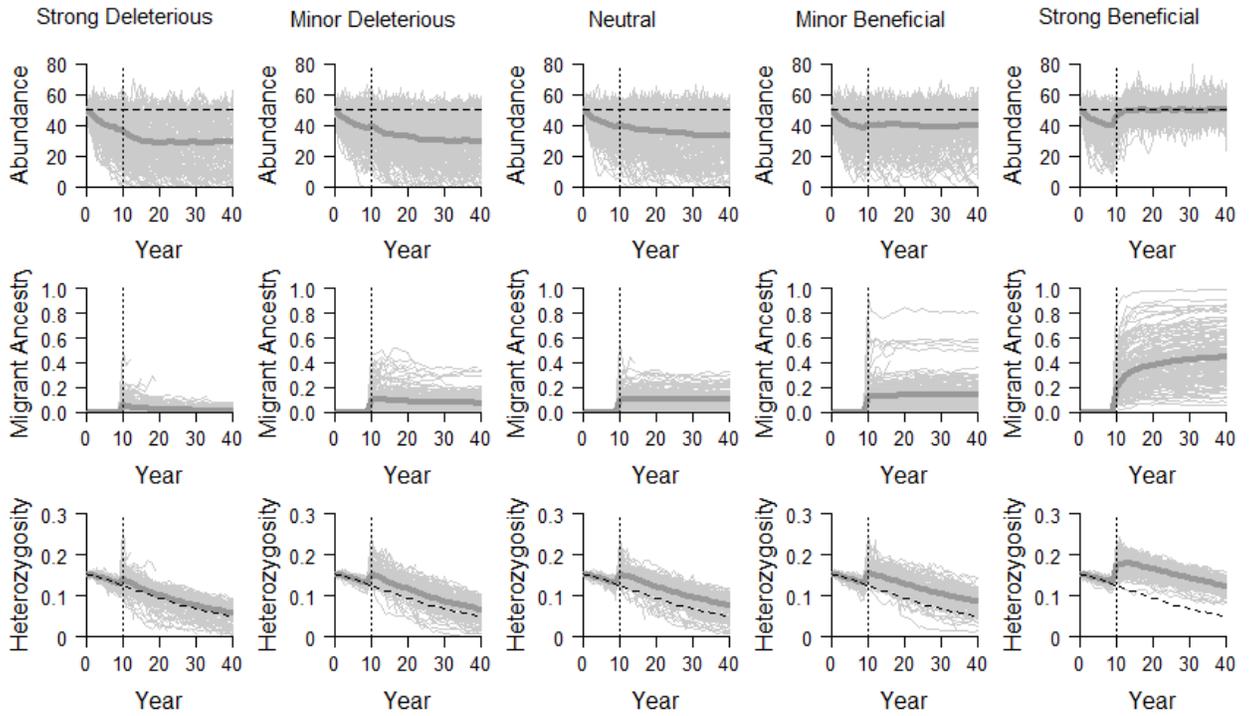


Fig. S3-5. Trends in abundance, migrant ancestry and observed heterozygosity with differing fitness effects of gene flow. The baseline scenario with  $K = 50$ ,  $N_0 = 50$ , and 4 immigrants is shown. Columns represent fitness effects of -0.8, -0.1, 0, 0.1, and 0.8. Dark bold grey lines are the mean, and light thin lines are 100 randomly selected iterations of the simulation. The dotted vertical line at year 10 represents the single gene flow event. The dashed lines represent  $K$  in the abundance figures, and the dashed trend line in the heterozygosity figures represents the pre-gene flow trajectory.

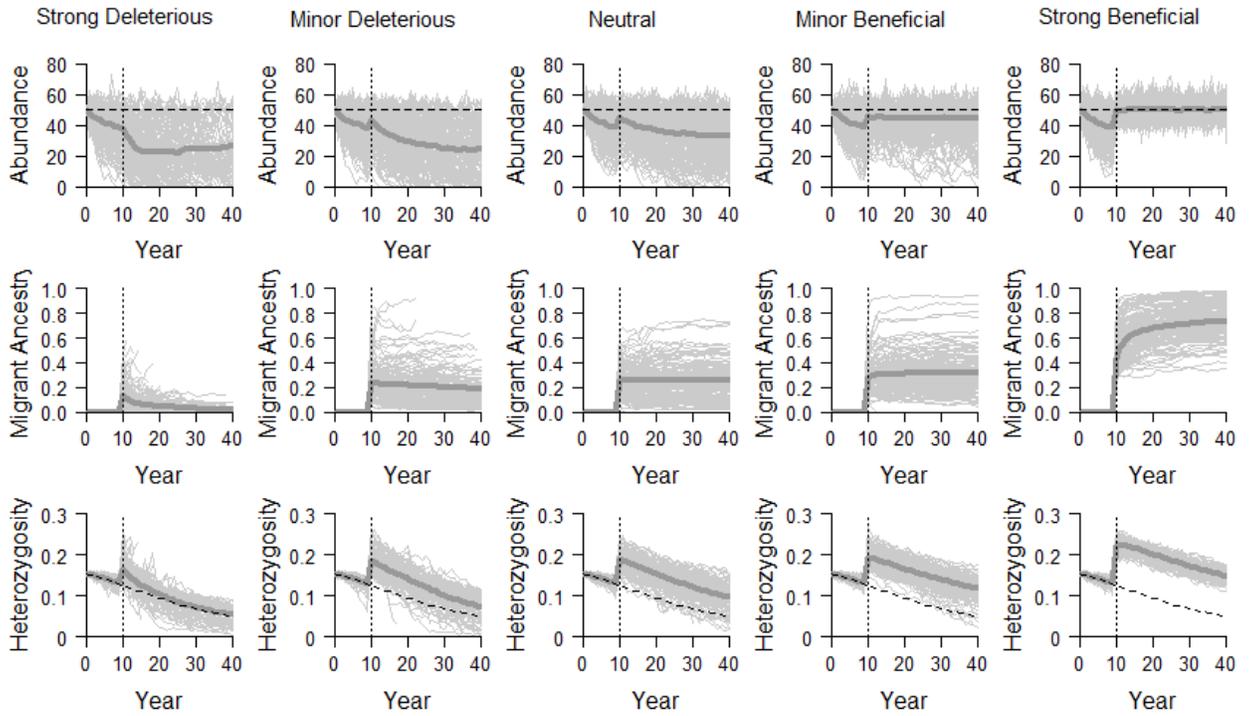


Fig. S3-6. Trends in abundance, migrant ancestry and observed heterozygosity with differing fitness effects of gene flow. The baseline scenario with  $K = 50$ ,  $N_0 = 25$ , and 12 immigrants is shown. Columns represent fitness effects of  $-0.8$ ,  $-0.1$ ,  $0$ ,  $0.1$ , and  $0.8$ . Dark bold grey lines are the mean, and light thin lines are 100 randomly selected iterations of the simulation. The dotted vertical line at year 10 represents the single gene flow event. The dashed lines represent  $K$  in the abundance figures, and the dashed trend line in the heterozygosity figures represents the pre-gene flow trajectory.

Plots of temporal trends in statistical power for scenarios not presented in Figure 3.

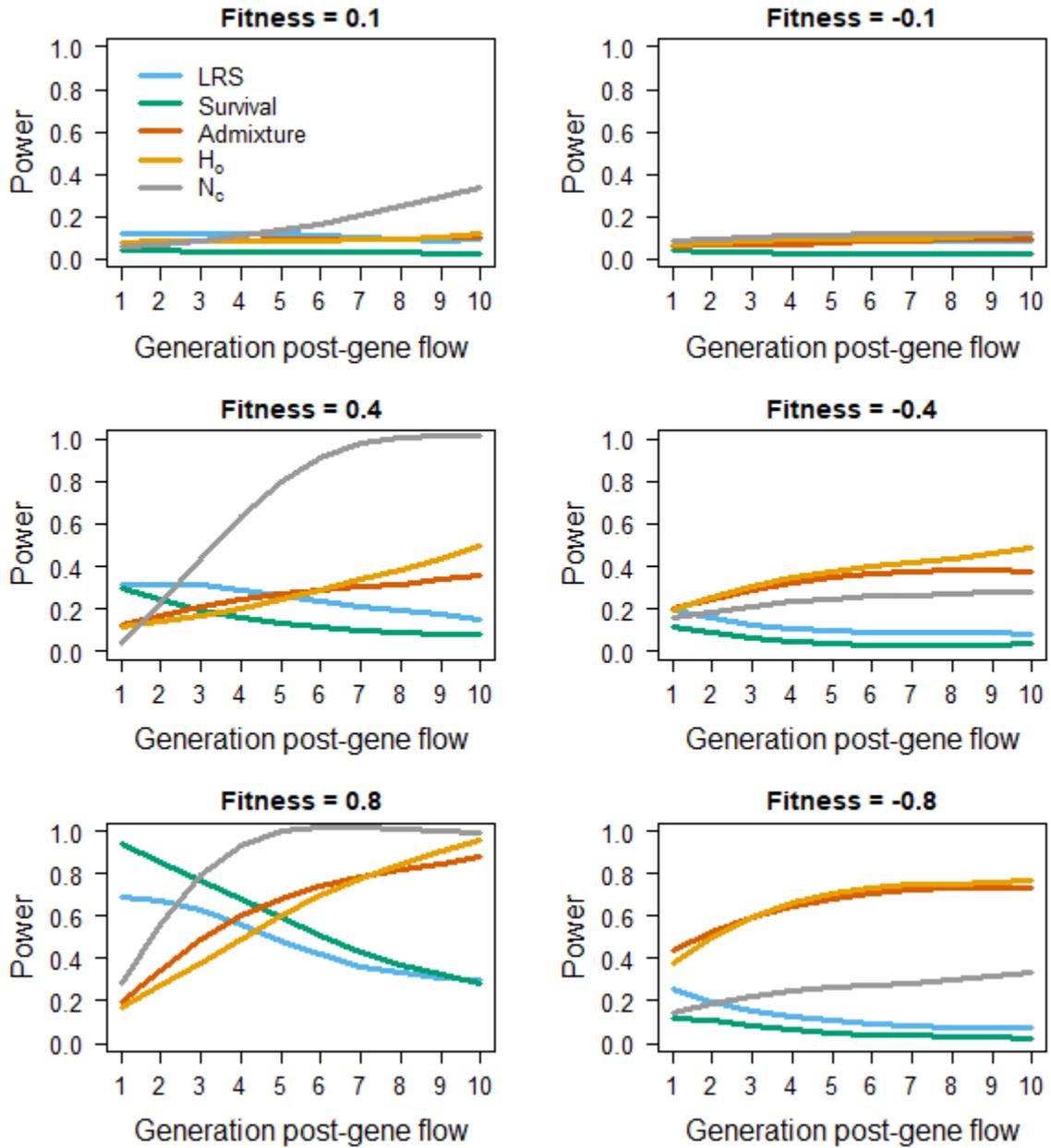


Fig. S3-7. Statistical power to correctly infer the fitness effect of gene flow for genetic and demographic metrics given different simulated fitness effects of gene flow. The scenario with a  $K=150$ ,  $N_0=75$ , and 12 immigrants is shown. Smoothing curves are shown to represent the general trends in power.

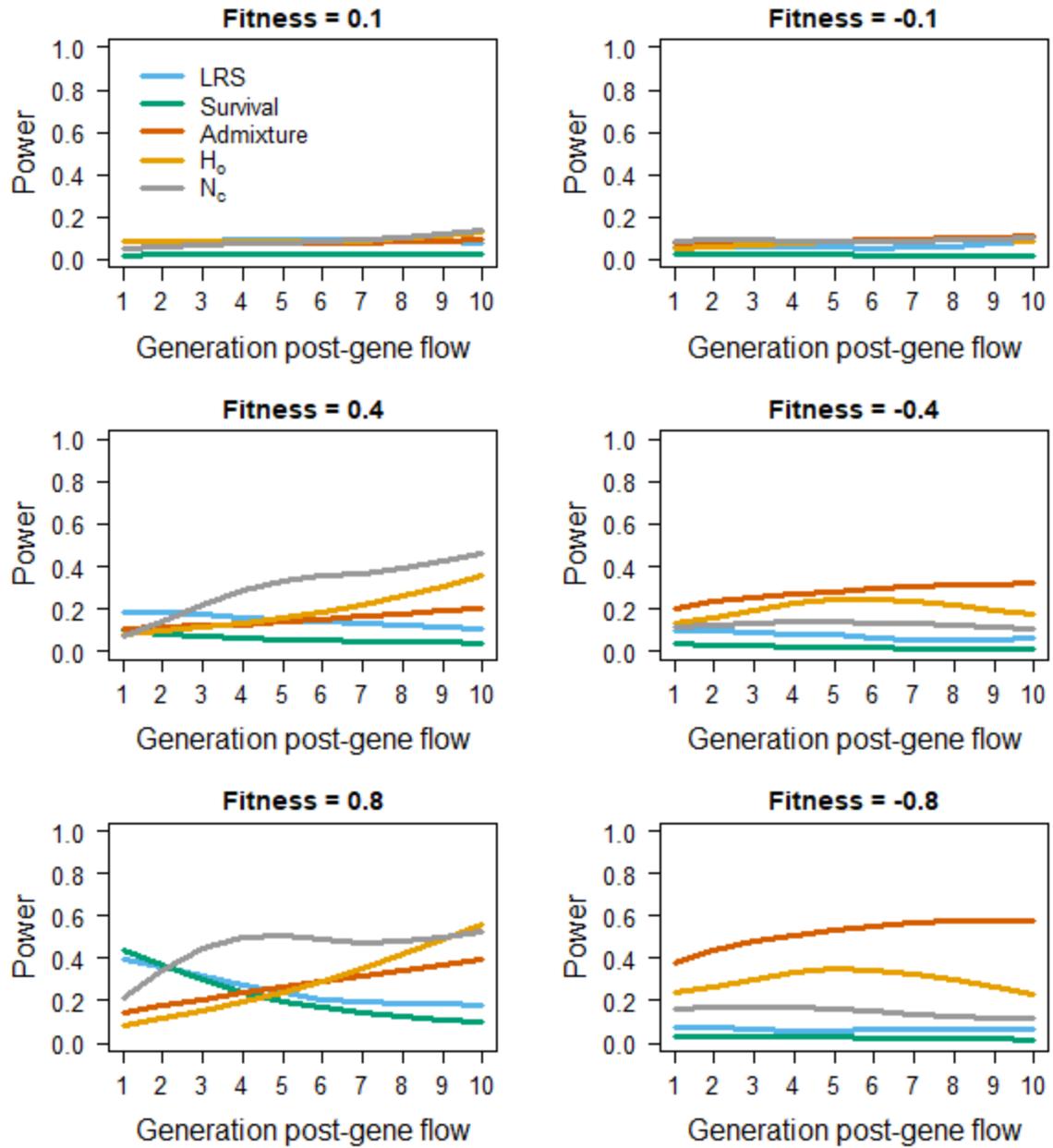


Fig. S3-8. Statistical power to correctly infer the fitness effect of gene flow for genetic and demographic metrics given different simulated fitness effects of gene flow. The scenario with a  $K = 50$ ,  $N_0 = 25$ , and 4 immigrants is shown. Smoothing curves are shown to represent the general trends in power.

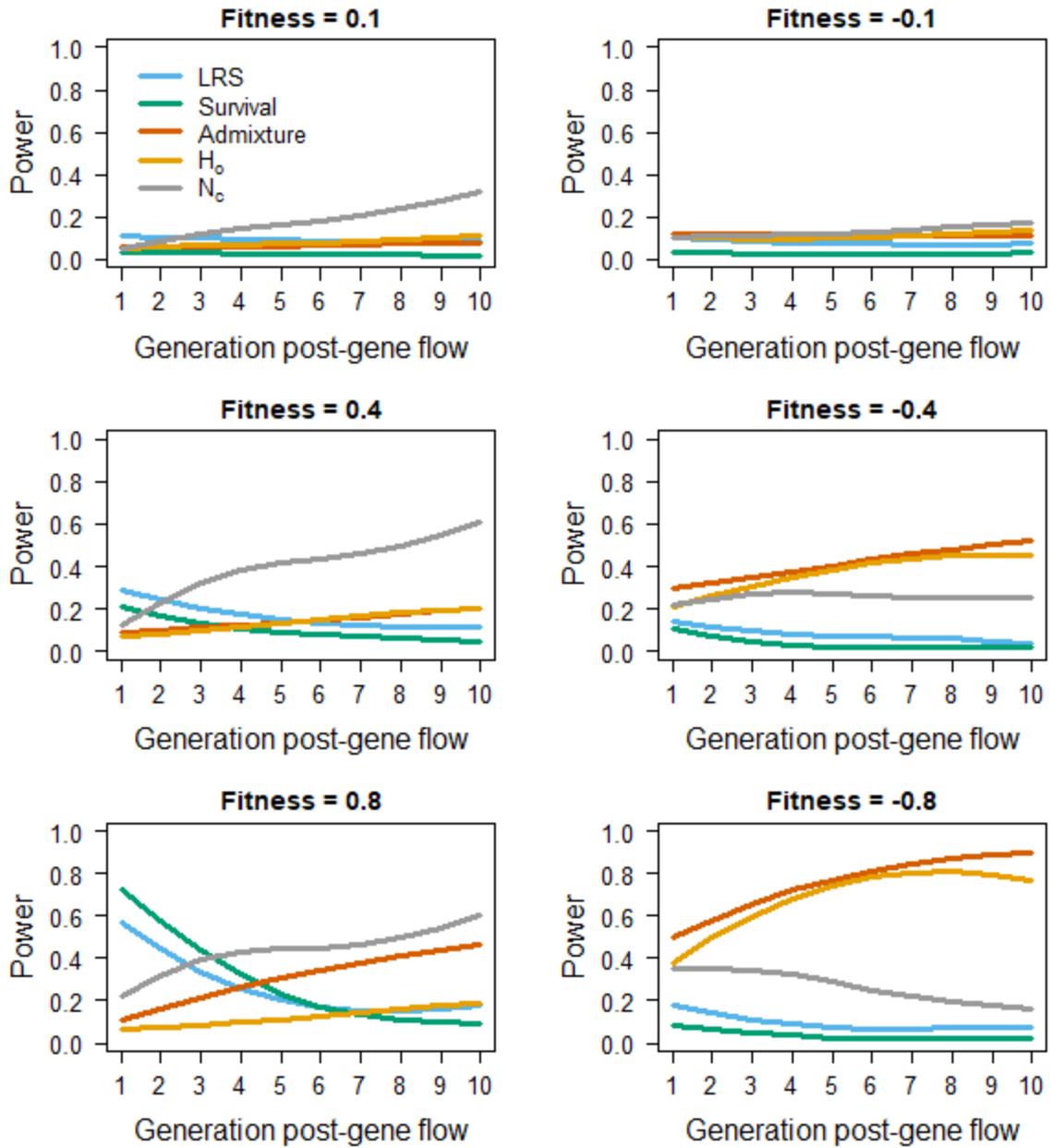


Fig. S3-9. Statistical power to correctly infer the fitness effect of gene flow for genetic and demographic metrics given different simulated fitness effects of gene flow. The scenario with a  $K = 50$ ,  $N_0 = 25$ , and 12 immigrants is shown. Smoothing curves are shown to represent the general trends in power.

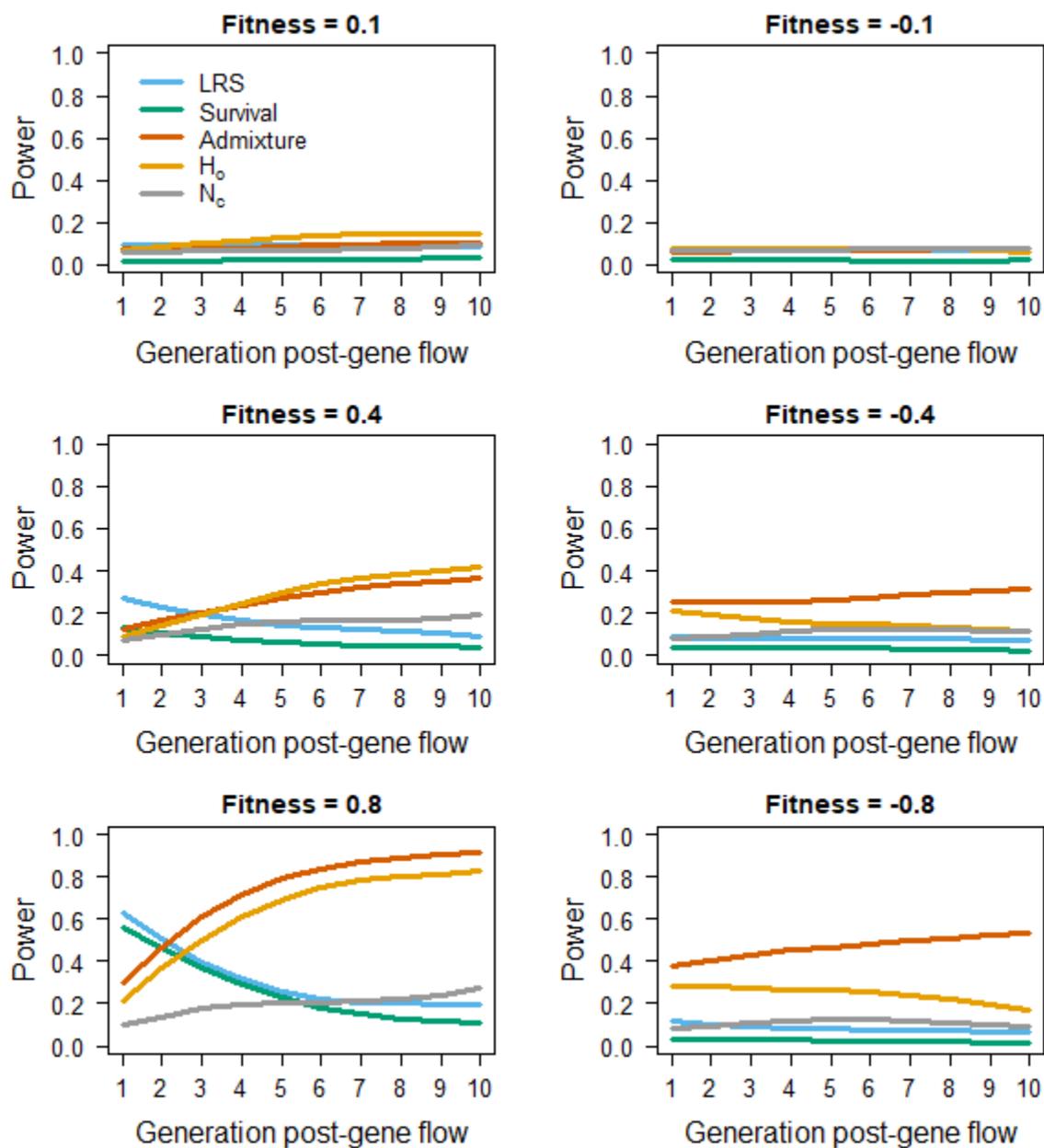


Fig. S3-9. Statistical power to correctly infer the fitness effect of gene flow for genetic and demographic metrics given different simulated fitness effects of gene flow. The scenario with a  $K = 50$ ,  $N_0 = 50$ , and 4 immigrants is shown. Smoothing curves are shown to represent the general trends in power.

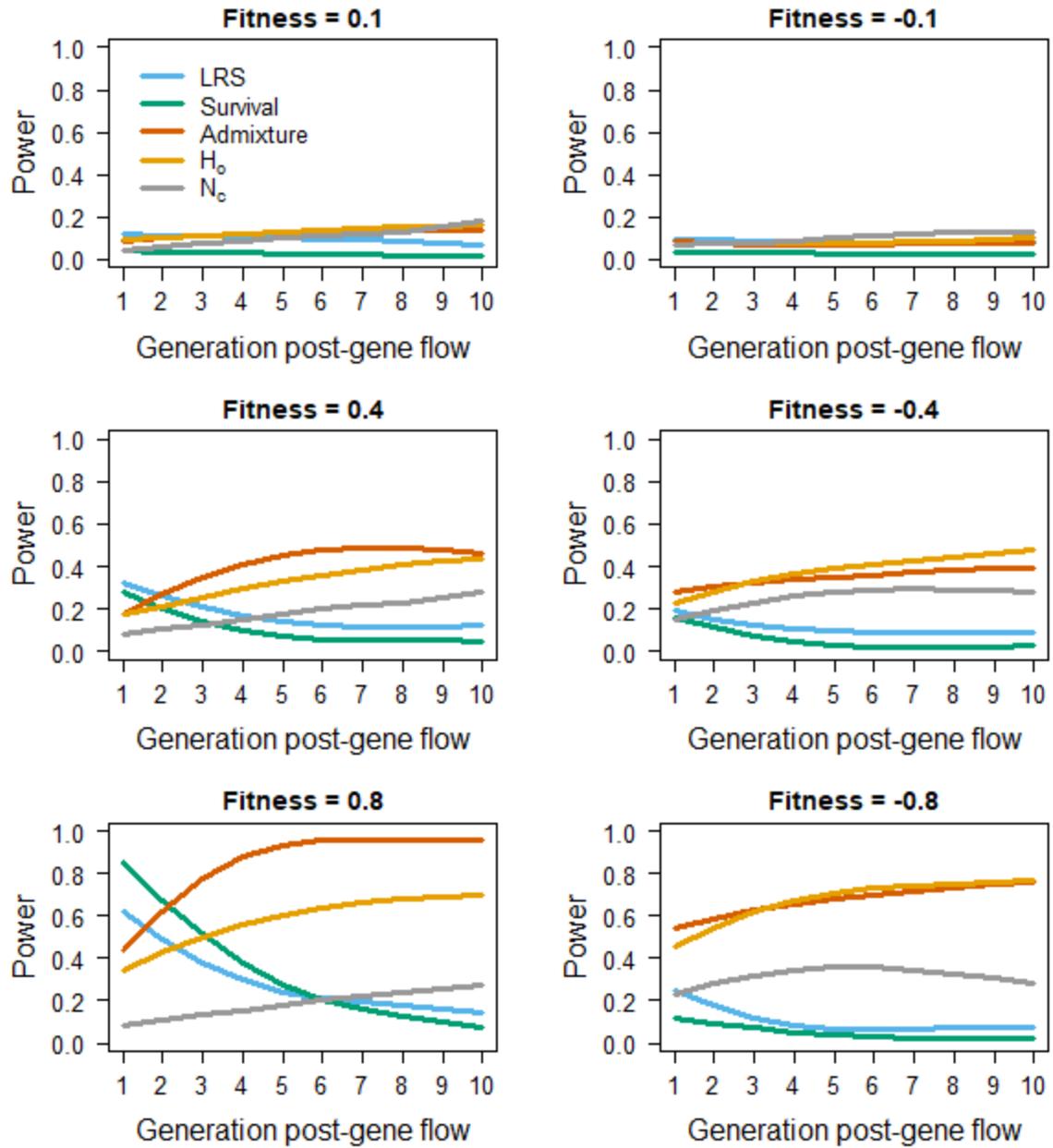


Fig. S3-10. Figure 5. Statistical power to correctly infer the fitness effect of gene flow for genetic and demographic metrics given different simulated fitness effects of gene flow. The scenario with a  $K = 50$ ,  $N_0 = 50$ , and 12 immigrants is shown. Smoothing curves are shown to represent the general trends in power.

Making inference on direction of effect in abundance, heterozygosity, and migrant ancestry.

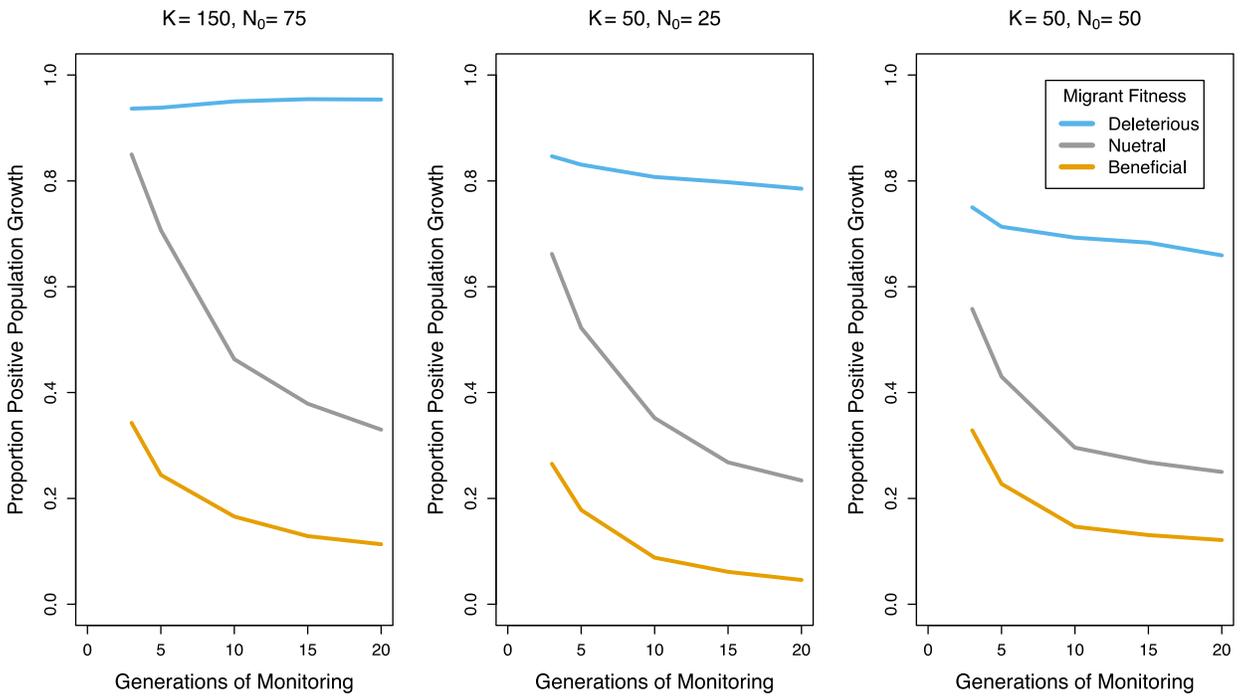


Fig. S3-11. The proportion of simulation repetitions where abundance has a positive linear slope coefficient under beneficial, neutral, and deleterious gene flow. Each panel shows a different demographic scenario. The figure demonstrates that many of the neutral and deleterious repetitions of the simulations produce positive slopes in abundance with a few generations of monitoring, regardless of the fitness effect of gene flow. This demonstrates the need for long-term monitoring of abundance and accounting for environmental and demographic stochasticity when possible.

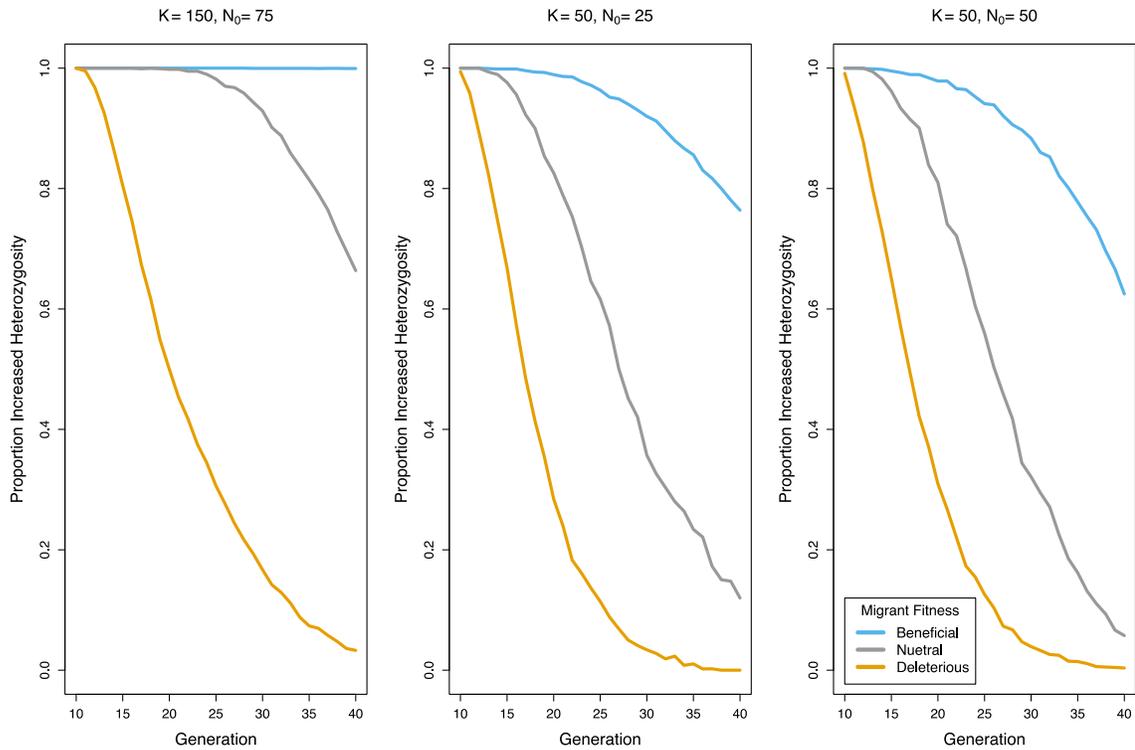


Fig S3-12. The proportion of simulation repetitions where heterozygosity is elevated above the pre-gene flow level under beneficial, neutral, and deleterious gene flow. Each panel shows a different demographic scenario. The figure demonstrates that nearly all repetitions of the simulation observe an increase in heterozygosity initially and depending on population size eventually decline to pre-gene flow levels of heterozygosity, regardless of the fitness effect of gene flow. Which suggests that making inference on the direction of effect in heterozygosity, even with long-term monitoring, without accounting for effective size and demographic stochasticity will often lead to incorrect inference. A similar plot is not presented for migrant ancestry because it always increases when there is introgression, and thus must be compared to a neutral distribution (see main document).

## APPENDIX C. Chapter 5 Supplementary Materials

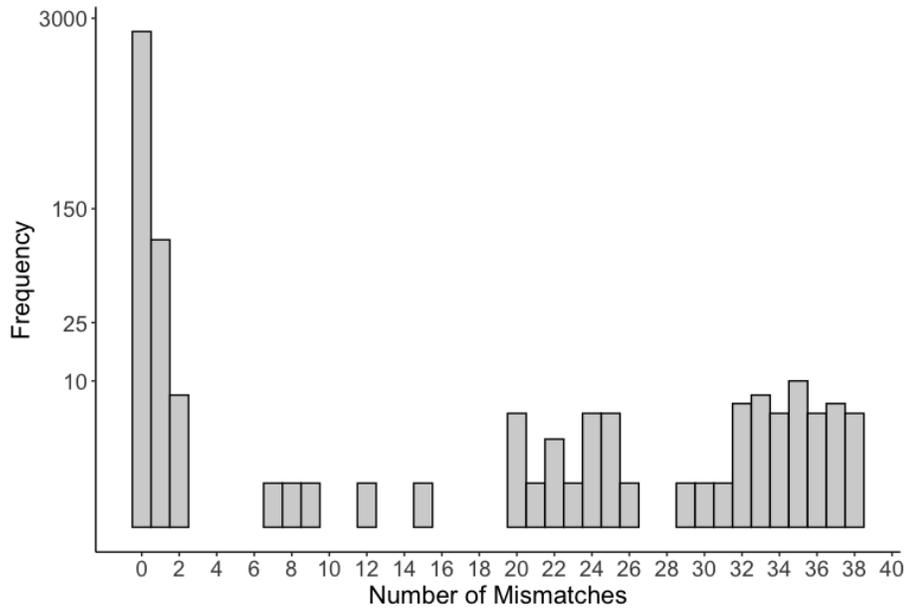


Fig. S3-1. The number of multilocus genotypic mismatches among pairwise comparisons of individuals with greater than 80% genotypic similarity. We used a threshold two or fewer mismatches to identify recaptures. Note that the y-axis is on the log scale and that the majority of pairwise comparisons were below 80% similarity