# EVALUATING GENOMIC ESTIMATES AND RECONSTRUCTED PEDIGREES AS ASSESSMENT TECHNIQUES FOR SEA LAMPREY POPULATIONS 

By
Ellen M. Weise

## A THESIS

Submitted to
Michigan State University in partial fulfillment of the requirements
for the degree of
Fisheries and Wildlife - Master of Science

# PUBLIC ABSTRACT <br> EVALUATING GENOMIC ESTIMATES AND RECONSTRUCTED PEDIGREES AS ASSESSMENT TECHNIQUES FOR SEA LAMPREY POPULATIONS <br> By <br> Ellen M. Weise 

Sea lamprey (Petromyzon marinus) are an extremely harmful invasive species in the Great Lakes. The species decimated native fish populations, causing harm to the ecosystem. To aggressively respond to the invasion, a bi-national program has been dedicated to reducing sea lamprey numbers. Control of lamprey populations includes physical barriers to prevent spawning adults from entering streams, and applications of lampricide (3-trifluormethlyl-4-nitrophenol or TFM) to kill larvae living in stream substrates. Annual assessments of adult sea lamprey are conducted, but are limited to a small number of streams. This study generated genetic data for sea lamprey larvae to reconstruct parental genotypes and estimate effective size of spawning populations. In Chapter 1, we use this information to evaluate the magnitude of barrier failures in three streams. In Chapter 2, we genotyped larvae from 18 streams with different physical characteristics across the Great Lakes and examined the effects of different factors that could affect spawning populations. Additionally, we generated simulated sea lamprey populations to evaluate the effects of sample size, number of genotypes, and true effective population size on the accuracy and precision of genetic estimates. Our simulations showed that a sample size of at least 100 individuals, along with maximization of SNP set size, allows for accurate estimates for all effective population sizes tested. Our work demonstrates that pedigree-based inferences can be effectively used as a management tool to characterize sea lamprey spawning abundance, poorly understood aspects of the species mating system, and relationships between adult reproductive success and associated stream characteristics.

# ABSTRACT <br> EVALUATING GENOMIC ESTIMATES AND RECONSTRUCTED PEDIGREES AS ASSESSMENT TECHNIQUES FOR SEA LAMPREY POPULATIONS 

By

Ellen M. Weise

Sea lamprey (Petromyzon marinus) are an invasive species in the Great Lakes. Their invasion resulted in the decimation of native fish populations, and a large control program has been dedicated to reducing lamprey populations. Control measures are mainly based on the construction of barriers to limit access to spawning habitat and the use of lampricides, such as 3-trifluormethlyl-4-nitrophenol, to kill developing larvae in stream sediments. Current assessment techniques in Great Lakes tributaries include mark-recapture estimation of census size of sea lamprey adult populations. We expanded traditional assessment techniques by generating reconstructed pedigrees and estimates of effective breeding size ( $N_{b}$ ) and minimum spawning size $\left(N_{s}\right)$ of sea lamprey populations using single nucleotide polymorphism (SNP) genotypes of larval sea lamprey. In Chapter 1, we evaluated efficacy of barriers to adult upstream passage in three streams using population genomic data. In Chapter 2, we elucidated the effects of several sampling and environmental factors on $N_{b}$ and $N_{s}$ estimates from 18 streams across the Great Lakes. Additional analyses were conducted to examine the effects of sample size, number of SNP loci, and true $N_{b}$ on estimated $N_{b}$ and $N_{s}$ using simulated sea lamprey populations. As true $N_{b}$ increased, different methods of estimating $N_{b}$ and $N_{s}$ showed different types and levels of bias, highlighting the need for multiple methods of estimating these parameters, as well as sufficient sample sizes and numbers of SNP loci. Overall, the analyses conducted provided unique insight into sea lamprey spawning populations and have potential as annual assessment techniques for evaluating both current and future sea lamprey control efforts.

## ACKNOWLEDGEMENTS

This project was funded by the Great Lakes Fishery Commission (2019_ROB_540840). Special thanks to researchers at the Hammond Bay Biological Station, U.S. Fish and Wildlife Service, and Fisheries and Oceans Canada for their sample collection work. This project was partially supported by computational resources at Institute for Cyber-Enabled Research and the Genomics Core of the Research Technology Support Facility at Michigan State University. Special thanks to committee members Dr. John Robinson, Dr. Kim Scribner, Dr. Cheryl Murphy, and Dr. Juan Pedro Steibel. Thanks to project contributors Dr. Jean Adams, Dr. Nick Johnson, Dr. Aaron Jubar, Dr. Gale Bravener, Olivia Boeberitz, Dr. Nick Sard, Seth Smith, and the Scribner/Robinson Lab.

## TABLE OF CONTENTS

LIST OF TABLES ..... vi
LIST OF FIGURES ..... viii
THESIS INTRODUCTION ..... 1
Sea Lamprey in the Great Lakes ..... 1
Sea Lamprey Control and Assessment ..... 2
Genetic Population Assessment ..... 5
Objectives ..... 10
LITERATURE CITED ..... 11
CHAPTER 1: PEDIGREE ANALYSIS AND ESTIMATES OF EFFECTIVE BREEDING SIZE CHARACTERIZE SEA LAMPREY REPRODUCTIVE BIOLOGY ..... 21
ABSTRACT ..... 21
INTRODUCTION ..... 22
METHODS ..... 28
Study System and Sample Collection ..... 28
RAD-capture Sequencing ..... 32
Genotyping Analysis ..... 33
Gaussian Mixture Analyses ..... 36
Reconstructed Pedigrees ..... 39
$\mathrm{N}_{\mathrm{b}}, \mathrm{N}_{\mathrm{s}}$, and $\widehat{N}_{\mathrm{s}}$ estimates ..... 40
RESULTS ..... 42
Genotyping Analysis ..... 42
Mixture Analyses and Reconstructed Pedigrees ..... 42
$N_{b}$ and $N_{s}$ calculations ..... 48
DISCUSSION ..... 51
$N_{b}$ and $N_{s}$ estimates ..... 51
Cohort identification ..... 53
Importance of Sample Size in $N_{b}$ and $N_{s}$ estimates ..... 54
Application of Results ..... 54
LITERATURE CITED ..... 58
CHAPTER 2: THE EFFECTS OF SAMPLING, BIOTIC, AND ENVIRONMENTAL VARIABLES ON ESTIMATES OF SEA LAMPREY EFFECTIVE BREEDING SIZE AND MINIMUM NUMBER OF SPAWNERS IN GREAT LAKES TRIBUTARES ..... 70
ABSTRACT ..... 70
INTRODUCTION ..... 72
METHODS ..... 77
Sample Collection ..... 77
Sequencing Library Preparation ..... 79
Bioinformatic Analysis ..... 80
Cohort-determining models ..... 82
$N_{b}$ and $N_{s}$ estimates ..... 83
Statistical Analyses ..... 84
Effects of Sample Size, SNP set size, and stream $N_{b}$ on genetic estimates ..... 85
RESULTS ..... 87
Read Processing ..... 87
Mixture Models ..... 89
$N_{b}, N_{s}$, and $\widehat{N}_{\mathrm{s}}$ estimates ..... 94
Correlations and Linear Modeling ..... 98
Effects of Sample Size, SNP set size, and stream $N_{b}$ on genetic estimates ..... 108
DISCUSSION ..... 122
Simulation Recommendations ..... 123
Reconstructed Pedigrees and Genetic Estimates ..... 124
$N_{b}$ and $N_{c}$ relationship and Sampling Effects ..... 125
Applications in Management ..... 127
APPENDIX ..... 129
LITERATURE CITED ..... 131
CONCLUSIONS ..... 140

## LIST OF TABLES

Table 1.1. Summary of results for identifying the optimal number of clusters (K) in the mixture analysis for sea lamprey. Analyses were performed for each larval collection with a range of $\mathrm{K}=1-4$ clusters. $\mathrm{R} \& \mathrm{M}$ criteria and Bmixture shows the estimated probability of each K value from the Rousseau and Mengersen (2011) criteria and Birth Death Marcov Chain Monte Carlo (BD-MCMC; Mohammadi, Salehi-Rad, \& Wit, 2013), respectively. The optimal number of clusters from each method is bolded.

Table 1.2. Estimates of the effective number of breeding adults $\left(N_{b}\right)$ and the number of unique inferred parental genotypes in the inferred pedigree $\left(N_{s}\right)$ for each stream and sea lamprey cohort. Locations are shown in Figure 1. N is the number of larval sea lamprey sampled for a stream and year. $V_{k}$ and $\bar{k}$ represent the inferred variance in reproductive success and mean number of offspring per adult in the population, respectively. LD refers to $N_{b}$ estimates derived from the linkage disequilibrium method. SF refers to $N_{b}$ estimates from the sibship frequency method. PwoP refers to $N_{b}$ estimates from the parentage-without-parents method. $N_{s}-$ Chao and Jackknife represent accumulated $N_{s}$ estimates using the Chao and the Jackknife methods, respectively

Table 2.1. Table showing the SNP set size, as well as the average MAF and percent of individuals genotyped, for each SNP set. SNPs refers to the size of the SNP set, pGT refers to the average percent of individuals genotyped across SNPs in the SNP set, and MAF refers to the average minor allele frequency across SNPs in the SNP set.87

Table 2.2. Summary of results for identifying the optimal number of clusters $(\mathrm{K})$ in the mixture analysis for sea lamprey. Analyses were performed for each larval collection with a range of $\mathrm{K}=1-4$ clusters. $\mathrm{R} \& \mathrm{M}$ criteria and Bmixture shows the estimated probability of each K value from the Rousseau and Mengersen (2011) criteria and BD-MCMC, respectively. The optimal number of clusters from each method is bolded. If the two methods disagree, the method with the higher probability is used. An asterisk indicates that neither method of cluster-determining model had a high probability assigned to a single K value.

Table 2.3. $N_{b}$ and $N_{s}$ estimates and population-based information. N indicates the number of sequenced offspring for the cohort, $N_{c}$ is the census-size estimate based on mark-recapture during the spawning year. linkage disequilibrium (LD), parentage without parents (PwoP) and sibship frequency (SF) columns are $N_{b}$ point estimates with corresponding uncertainty. LD: $N_{c}$ and SF: $N_{c}$ refer to the ratios between the LD and SF method of estimating $N_{b}$ and the mark-recapture $N_{c}$ estimates. $\bar{k}$ and $V_{k}$ are the mean and variance in reproductive success inferred from the reconstructed pedigree. $N_{s}$ is the number of reconstructed parent genotypes for each cohort, and Chao and Jackknife are the extrapolated $N_{s}$ estimates and their corresponding $95 \%$ confidence intervals........................................................................................................... 94

Table 2.4. Environmental, biotic, and sampling linear models. Significant p-values are denoted by an asterix. Treatment year refers to the most recent TFM treatment that occurred in the
stream, $N_{c}$ is the census-size estimate based on mark-recapture for the years 2016, 2017, and 2018, and 'Trap efficiency 2018' refers to the trap efficiency of the values used to generate $N_{c}$. Drainage refers to the drainage area of the stream (hectares), larval potential is a variable that refers to the level of larval habitat, trap to mouth distance refers to the distance in km between the mouth of the river and the traps used for $N_{c}$ estimates. Sampling sites refers to the number of collection locations for the larval collections, years since treatments is the number of years between the last TFM treatment and the collection year. Sampling distance refers to the approximate distance sampled in each stream. If only one site was sample 0.2 km was used based on the standard transect distance for backpack electrofishing.

Table 2.5. Table for environmental, biotic, and sampling linear models. Significant variables are bolded along with the corresponding coefficient and p-value. In Table 2.5A, the global model consists of the following variables: years since TFM treatment, drainage area, number of sampling sites and sample size. In Table 2.5B, the global model consists of the following variables: years since TFM treatment, drainage area, number of sampling sites, sample size, and distance from the mouth of the river to the mark-recapture trap site. In Table 2.5C, the global model consists of the following variables: years since TFM treatment, drainage area, number of sampling sites, sample size, and $N_{c}$ estimates. In Table 2.5 C , only $N_{s}$ - Chao and $V_{k}$ were considered as response variables.

## LIST OF FIGURES

Figure 1.1. Map of the study area where larval sea lamprey were collected. The Black Mallard River is separated into upper and lower sections by Black Mallard Lake. The top-right inset shows the location of the sampled river systems in the Great Lakes region. River lines in black denote sampling locations of the river systems, blue lines denote all other rivers in the region.. 31

Figure S.1. Visualization of principal component analysis (PCA) used to compare sea lamprey larval individuals from two native lamprey species (Lethenteron appendix, Ichthyomyzon fossor). Purple dots labeled P.marinus represent sequenced individuals, green dots labeled I. fosser represent known Northern brook lamprey, and blue dots labeled L. appendix represent known American brook lamprey. Supplemental Figure 1A shows individuals collected in the Black Mallard River, Supplemental Figure 1B shows individuals collected in the Cheboygan River, and Supplemental Figure 1C shows individuals collected in the Ocqueoc River

Figure 1.2. A flow chart describing how inferred cohort assignments from the Gaussian mixture models are combined with information in the reconstructed pedigrees

Figure 1.3. Length frequency distributions for larval sea lamprey from all rivers and collection years, fill colors represent individual cluster assignment from the Gaussian mixture analysis. If mixture models were not completed due to small sample size, length histograms are included and shaded as a single cohort45

Figure 1.4. Boxplots of length distributions for each sea lamprey Colony cluster from the Lower Black Mallard River (A) and the Ocqueoc River (B). Colony clusters are defined as groups of offspring in the pedigree that are connected by parentage, but are not necessarily full- or halfsiblings. Plots are separated by collection. The probability that the Colony cluster cannot be split is represented by a continuous shading scale for both subplots (red clusters have a lower likelihood, white clusters have a higher likelihood).

Figure 1.5. Visualization of reconstructed sea lamprey pedigrees. The center represents genotyped individuals, and dots represent inferred parents. Lines connect each reconstructed parent to sequenced offspring in the pedigree. Black boxes represent cohorts inferred by the mixture method. Note: Since parents were not sequenced, and due to the lack of known sexdetermining genes for sea lamprey, the sex of reconstructed parents cannot be determined. Parent 1 and Parent 2 are used instead.47

Figure 1.6. The estimated number of unique parental genotypes in the pedigree ( $\widehat{N}_{s}$ ) characterized using pedigree accumulation curves for all three stream systems. For all locations, boxplot distributions for each step size overlay a line plot with a grey background for $+/$ - one standard error, and labeled horizontal lines represent $\widehat{N}_{s}$ estimates from the jackknife and chao methods. Due to the large number of individuals, the Ocqueoc River boxplots are plotted in step sizes of 5 sampled individuals and the Lower Black Mallard River boxplots are shown for sample sizes increasing by 10 individuals. The boxplots for all other locations are plotted for a
step size of 1 sampled individual ..... 50

Figure 2.1. Map showing all sampled streams with their location in the Great Lakes system. Each dot represents a stream system

Figure 2.2. Length frequency distributions for larval sea lamprey from all rivers and collection years, fill colors represent individual cluster assignment from the Gaussian mixture analysis. If mixture models were not completed due to small sample size, length histograms are included and shaded as yellow.

Figure 2.3. Boxplots showing the length distributions of each cluster of sequenced offspring. Boxes are shaded by the cluster likelihood, where lower likelihoods are shaded towards red and higher likelihoods are shaded towards white. Boxplots are limited to clusters with 3 or more individuals. The East Au Gres and the Muskegon River are not shown because they do not have any clusters larger than 3 individuals

Figure 2.4. Diagrams of reconstructed pedigrees for all stream systems. The offspring are in the center of the diagram and are connected to their reconstructed parents by grey lines. The offspring are sorted first by parent 1 sibling groups, then parent 2 sibling groups

Figure 2.5. Minimum number of spawning adults ( $N_{s}$ ) accumulation curves showing the increase in unique parent genotypes as the number of sequenced offspring increased for each cohort. The dark red lines in each plot represent the chao asymptote estimates (Chao, 1987a), and the dark blue lines represent the jackknife asymptote estimates (Heltshe \& Forrester, 2009)

Figure 2.6. Scatterplots of effective breeding size $\left(N_{b}\right)$, minimum number of spawning adults $\left(N_{s}\right)$, and census size from mark-recapture $\left(N_{c}\right)$ estimates. $N_{c}$ is shown on the x-axis, the $N_{b}$ or $N_{s}$ estimate is shown on the $y$-axis. No lines of best fit were included due to the lack of significant correlation between variables in the plots.

Figure 2.7. Plots of significant predictors of $N_{b}$ and $N_{s}$ estimates based on the results of the environmental models

Figure 2.8. A figure that visualizes the ratio between estimated $N_{b}$ and the true estimate from each simulation. The sample size parameter is on the $x$-axis, the SNP set size is separated by color, and the plots are subset by the effective breeding size parameter. Figure 2.8 A is the sibship frequency method, figure 2.8 B is the linkage disequilibrium method, and figure 2.8 C is the parents without parents methods.

Figure 2.9. Plots (plot 2.9A-2.9E) to show accuracy of point estimates for simulated populations. The x -axes are $\log _{10}$ of the parameter effective breeding size $\left(N_{b}\right)$, and the $y$-axes are $\log _{10}$ of the estimated $N_{b}$ or the minimum number of spawning adults $\left(N_{s}\right)$. The plots are subset by SNP set size and sample size, and figures are separated by each method. Figure 2.9A shows results from the sibship frequency estimates, figure 2.9B shows results from the linkage disequilibrium estimates, figure 2.9 C shows results from the parentage without parents estimates, figure 2.9D
shows results from the chao estimates, figure 2.9E shows results from the jackknife estimates 111

Figure 2.10. Root mean squared error (RMSE) plots (plots 2.10A-2.10E) for each type of estimate to show the variance in point estimates for simulated populations. RMSE ( y -axis) is plotted versus the sample size ( x -axis). The line colors are the SNP set size, where yellow corresponds to $\mathrm{SNPs}=100$, dark blue corresponds to $\mathrm{SNPs}=500$, and green-grey corresponds to SNPs $=1000$. The plots are subset by parameter effective breeding size $\left(N_{b}\right)$, and the figures are separated by $N_{b}$ and the minimum number of spawning adults ( $N_{s}$ ) estimate method. Figure 2.10A shows results from the sibship frequency estimates, figure 2.10 B shows results from the linkage disequilibrium estimates, figure 2.10 C shows results from the parentage without parents estimates, figure 2.10D shows results from the chao estimates, figure 2.10 E shows results from the jackknife estimates. 116

## THESIS INTRODUCTION

## Sea Lamprey in the Great Lakes

Sea lamprey (Petromyzon marinus) arrived in the Great Lakes following the expansion of the Welland Canal in 1919, and became a destructive invasive species across the ecosystem (Lawrie, 1970). Native fish parasitized by lamprey experienced a subsequent crash in population size, with annual catch rates significantly reduced compared to periods prior to the arrival of sea lamprey (Heinrich et al., 2003; Koonce, Eshenroder, \& Christie, 1993; Lawrie, 1970). By the 1950s, a large annual control and assessment program was established to control sea lamprey population size in the Great Lakes.

Sea lamprey have a multi-stage life cycle that takes place over several years (Applegate, 1950; Manion \& Smith, 1978; Morkert, Swink, \& Seelye, 1998). The larval phase takes place in the stream beds where the sea lamprey spawned (Dawson, Quintella, Almeida, Treble, \& Jolley, 2015). Larvae embed in soft sections of substrate and filter feed for 3-7 years (Manion \& Smith, 1978; Morkert et al., 1998). Over these years, these larvae can occasionally drift downstream, particularly if their current substrate environment becomes poor filter feeding ground (Hardisty \& Potter, 1971; Potter, 1980). Due to the variable length of the larval development period, larvae present in the sediment represent multiple age classes. Length distributions are used to estimate larval age in the stream. Age-0 and age-1 individuals can be separated from larger/older age groups using these data, but age $2+$ individuals have overlapping size ranges that can be difficult to separate (Dawson, Jones, Scribner, \& Gilmore, 2009). Additionally, there is evidence that the quality of the river environment influences larval size and growth rates, particularly for older lamprey (Dawson, Higgins-weier, Steeves, \& Johnson, 2020). Once larvae reach a certain size, approximately $130-145 \mathrm{~mm}$, they metamorphosize and migrate into the Great Lakes to begin the
parasitic phase of their life cycle (Griffiths, Beamish, Morrison, \& Barker, 2001; Henson, Bergstedt, \& Adams, 2003).

As parasitic juveniles, sea lamprey feed on several types of medium to large fish species, including lake trout (Salvelinus namaycush; Harvey, Ebener, \& White, 2008; Pycha \& King, 1975), Chinook salmon (Oncorhynchus tshawytscha; Adams \& Jones, 2020), and lake whitefish (Coregonus clupeaformis; Ebener, Brenden, \& Jones, 2010; McLeod, Cottrill, \& Morbey, 2011). Sea lamprey attach to a fish and bore a hole into the scales to feed on the blood of their host, where each lamprey can cause between 5 and 20 kg of fish mortality during their feeding phase (Swink, 2003). Sea lamprey can travel large distances over their year as a parasitic juvenile as their host fish migrate around the lakes, leading to dispersal of lamprey across the Great Lakes system (Waldman, Grunwald, \& Wirgin, 2008).

The spawning season for sea lamprey occurs in the spring, when adults reenter streams to spawn. Adult sea lamprey do not home to natal streams to spawn (Bergstedt \& Seelye, 1995a), instead sea lamprey respond to pheromone cues produced by developing larvae, implying the existence of large larval populations and implicitly, good spawning habitat (M. B. Twohey et al., 2003). Once the adults enter the stream system, male sea lamprey make nests in rocky substrate and female sea lamprey visit several nests to spawn, leading to a polygamous mating structure (Applegate, 1950; Dhamelincourt, Buoro, Rives, Sebihi, \& Tentelier, 2020; Johnson, Buchinger, $\& L i, 2015)$.

## Sea Lamprey Control and Assessment

The life cycle of sea lamprey is used by management agencies to target sea lamprey in streams for control and assessment efforts. Annual control efforts are undertaken primarily
through the use of barriers and 3-trifluormethlyl-4-nitrophenol (TFM), a selective lampricide (Applegate, 1950; McDonald \& Kolar, 2007; Smith \& Tibbles, 1980). Physical barriers prevent spawning adults from entering stream systems and limit available spawning habitat (Lavis, Hallett, Koon, \& McAuley, 2003; McLaughlin, Marsden, \& Hayes, 2003). The first lampreyspecific barriers expanded on dams already present in large rivers in the Great Lakes. Recently, year-round barriers are slowly being removed due to their effects on the stream ecosystem, but electric barriers and seasonal barriers are increasingly common (Jensen \& Jones, 2018a;

McLaughlin, Hallett, Pratt, O'Connor, \& McDonald, 2007).
TFM is a lampricide applied on a three to four year cycle in streams with prevalent larval populations to eliminate most larvae before they metamorphosize into parasitic juveniles. TFM was designed to be lamprey specific, but it has been shown to be detrimental to native lamprey larvae as well as some native fish species, particularly juvenile sturgeon (Boogaard, Bills, \& Johnson, 2003; Pratt et al., 2020; Weisser et al., 2003). TFM targets the nervous system by creating a mismatch between ATP generation and consumption, leading to a drop in glycogen and eventual death (Birceanu, McClelland, Wang, \& Wilkie, 2009). Lamprey appear to be particularly sensitive to TFM as a lethal agent, when compared to other fish species like bluegill (Lepomis macrochirus) or catfish (Ictalurus punctatus), which are largely unaffected by TFM (Lawrence et al., 2021; Lech \& Statham, 1975).

Annual assessments across the Great Lakes region are used to estimate sea lamprey prevalence as larvae, juveniles, and spawning adults. Larval surveys are used to prioritize streams for TFM treatments each year based on the number of large larvae in the stream that are expected to metamorphosize into the parasitic life stage (Hansen et al., 2003). Lake trout caught in gill nets are used to estimate the number of actively parasitizing juveniles in each lake and
assess damage to commercial fisheries (Jones, 2007). Finally, mark-recapture efforts are used to estimate the abundance of spawning adults using trapping in an index group of streams in each Great Lake (Harper et al., 2018a).

Adult mark-recapture is used in a small number of streams to evaluate the prevalence of spawning adults in annual lamprey-producing streams. In 2018, the Peterson method of markrecapture (Peterson \& Cederholm, 1984) became the primary model used for adult assessment (Barber \& Steeves, 2019). The estimated abundance of spawning adults in the stream is considers the total number of adults, the number of marked individuals, and the number of marked individuals recaptured. The abundance index is estimated using a model that incorporates trapping efficiency for each stream, as well as previous data on lamprey abundance across streams (Barber \& Steeves, 2019). Prior to 2015, models using mark-recapture estimates of abundance as well as drainage area, time since TFM treatment, and other environmental variables were used to predict lake-wide adult sea lamprey abundance (Mullett et al., 2003). Since 2015, the sum of annual mark-recapture estimates across streams within lakes is used to generate an index of adult abundance based on the group of streams where trapping occurs (Adams, Barber, Bravener, \& Lewandoski, 2021; Sullivan, Adair, \& Woldt, 2016).

Sea lamprey populations are currently much smaller than at their historical peak in the 1950s (K. F. Robinson, Miehls, \& Siefkes, 2021), indicating that control efforts have been successful at reducing sea lamprey abundance, but there is a need for additional control and assessment techniques. Some of the largest streams across the Great Lakes are not currently index streams measured for adult assessment, meaning that potentially large spawning populations of sea lamprey are not currently assessed. Additionally, the trapping techniques required for generating mark-recapture estimates are not possible due to environmental
conditions in some streams, limiting the group of streams that can be used for adult assessment. For situations like barrier failure, spawning populations often cannot be assessed since the failure was not discovered until larval assessments in subsequent years, and trapping cannot be performed retroactively. For all of these situations, alternative assessment techniques for spawning populations are required.

In addition to alternate assessment techniques, there are several limitations to control techniques, indicating a need for supplemental control. There has long been concern about sea lamprey developing resistance to TFM, although to date there is no evidence of that resistance in Great Lakes populations (Dunlop et al., 2018). However, genetic models indicate that resistance could start to develop in the near future, and alternative controls would become increasingly necessary (M. R. Christie, Sepúlveda, \& Dunlop, 2019). Additionally, the removal of barriers from many streams could increase the amount of sea lamprey spawning habitat in the system, and supplemental control techniques will be needed to prevent an increase in sea lamprey population numbers. As sea lamprey control becomes more complex, additional assessment techniques will be necessary to evaluate the effectiveness of control methods. Particularly, there is a lack of assessment on the number of successfully spawning adults, and assessment of adults in a larger number of stream systems will become necessary to evaluate new control efforts in those streams.

## Genetic Population Assessment

Genetic assessment is an increasingly common tool in management as genotyping costs decrease and sequencing efficiency increases, making genomic sequencing as a tool for widespread annual assessment possible (e.g. Ovenden et al. 2016; Hunter et al. 2020).

Additionally, sea lamprey are an emerging model species for genomic analysis due to the recent completion of genomic resources. A somatic genome was sequenced in 2013 (Smith et al., 2013), followed by a germline genome in 2018 (Smith et al., 2018) and a chromosome-level genome assembly in 2020 through the vertebrate genome project (Rhie et al., 2020). Additionally, genomic resources that facilitate efficient reduced-representation genomic sequencing have been developed and recently published by Sard et al (2020).

Recent development of genomic resources facilitates the use of population genomic methods for assessment. Valuable information on families can be obtained through populationlevel genotyping and pedigree reconstruction. This can be done with a combination of parent and offspring genotypes (parentage analysis), or with exclusively offspring genotypes (pedigree reconstruction) (Wang, 2004). If only offspring genotypes are available, parental genotypes can be reconstructed from offspring genotypes (Blouin, 2003; Wang, 2004). Reconstructed pedigrees provide information on family relationships that can be utilized to assess populations for either conservation or control purposes. Reconstructed pedigrees have previously been used to evaluate reproductive success of spawning individuals, examine rates of inbreeding, evaluate the potential for inbreeding depression, quantify genetic diversity, and estimate effective population size (De Barba et al., 2010; Keogh, Webb, \& Shine, 2007).

Genetic data and reconstructed pedigrees can be used to generate estimates and metrics that serve as tools for population assessment. Effective population size $\left(N_{e}\right)$ estimates the size of an idealized population consistent with levels of genetic diversity, inbreeding, and genetic drift in the sampled population (Wright, 1931). Differences in fecundity, variance in reproductive success, fluctuation in population size over time, and skewed sex ratios among spawning
individuals all reduce effective population size compared to the census size of the population (Waples, Luikart, Faulkner, \& Tallmon, 2013).
$N_{e}$ is used as a benchmark estimate in conservation genetics for detecting inbreeding depression and the potential for an extinction spiral in a population (Frankham, Bradshaw, \& Brook, 2014). The values generally used to evaluate a population are that a population with an $N_{e}$ of under 50 is at short-term risk of extinction (Soule, 1980), and a population with an $N_{e}$ of under 500 will lose genetic diversity and is at long-term risk of extinction (Franklin, 1980; Franklin \& Frankham, 1998). These metrics have been used to evaluate species for extinction risk in a management context (Mace et al., 2008). However, there is debate about whether the $50 / 500$ numbers are too low, and if higher values like 100/1000 should be used instead (Frankham et al., 2014). Regardless of the debate on the specific benchmarks that should be used, the incorporation of $N_{e}$ in addition to census size $\left(N_{c}\right)$ as an assessment metric is important to evaluate populations for their extinction risk beyond low population numbers (Garner et al., 2016; Hoban et al., 2021). The $N_{e}$ of a population influences many indicators of extinction, such as high levels of inbreeding (Armbruster \& Reed, 2005) and loss of genetic diversity (Blomqvist, Pauliny, Larsson, \& Flodin, 2010). Additionally, $N_{e}$ can be used to evaluate the success of management actions like reintroduction (Anderson et al., 2014; Cochran-Biederman, Wyman, French, \& Loppnow, 2015; Evans et al., 2015; N. M. Sard et al., 2020) and genetic rescue (Fitzpatrick et al., 2016; Frankham, 2015; Heber, Briskie, \& Apiolaza, 2012) that are used for declining populations. Outside of conservation, $N_{e}$ estimates are used to examine the effects of stocking on fished populations (Gossieaux, Bernatchez, Sirois, \& Garant, 2019; Petereit et al., 2018a).

Effective population size $\left(N_{e}\right)$ is calculated on a generation scale rather than a spawning event scale. Effective breeding size $\left(N_{b}\right)$ is a similar metric that estimates the effective population size for a single cohort of offspring rather than a generation (Waples, Antao, \& Luikart, 2014). In species with a semelparous life history, $N_{e}=g * N_{b}$, where g is the generation time for the species (Waples, 1990). Depending on the life history of the organism, $N_{b}$ can be a more appropriate assessment metric than per-generation effective population size. Effective population size is complicated by overlapping generations (Waples et al., 2013), and may require multiple sampling periods or the sampling of multiple cohorts (Waples et al., 2014). Sea lamprey are a semelparous organism, so spawning adults will only be represented in one cohort of offspring. However, due to the varied length of time that larvae spend in substrate sea lamprey have overlapping generations. Due to these two factors, $N_{b}$ is a more appropriate metric for estimating sea lamprey spawning abundance in streams.
$N_{b}$ and $N_{e}$ can be calculated using similar methods. $N_{b}$ can be estimated using a single larval sampling event with a variety of approaches, including linkage disequilibrium (Hill, 1981a), sibship frequency (Wang, 2009), and parentage without parents (Waples \& Waples, 2011) methods. In the sibship frequency method, the rate of full and half-siblings present in sampled offspring is used to estimate $N_{b}$ (Wang \& Santure, 2009; Wang, 2009). Similarly, $N_{b}$ estimates from the parentage without parents method are based on the variance in family size rather than the frequency of sibship (Waples \& Waples, 2011). The linkage disequilibrium method (Hill, 1981a) quantifies the level of non-random association of alleles in genotypes at multiple loci, which is generally due to physical proximity in the genome, sample size, or the linkage that occurs from finite population size. By eliminating physical linkage as a source, and
accounting for influences of sample size with correction factors, the linkage from finite populations size can be used to estimate effective population size (Waples \& Do, 2010).

In addition to $N_{b}$, reconstructed pedigrees can be used to estimate the minimum number of spawning adults $\left(N_{s}\right)$ in a reproductive event. $N_{s}$ is calculated by estimating the number of parental genotypes required to produce the sampled offspring genotypes. $N_{s}$ is obtained directly from the number of unique parental genotypes required to produce sampled offspring genotypes, thus it is limited to twice the sample size of offspring. This can be a large source of bias in the metric, particularly if the sample size is small. However, if there is some presence of sibship within the sampled offspring, the number of parental genotypes can be extrapolated to estimate the minimum number of parents for the population represented by the sampled offspring. The total number of spawning adults can be estimated using a technique similar to a species accumulation curve in community ecology, where unique species accumulate as the number of sampled sites increases (Sard et al., in press). As the total number of species is approached, the number of new species per site decreases, leading to an asymptote at the true number of species. A pedigree rarefaction curve works in a similar way, where the number of unique parental genotypes is accumulated as the number of sampled offspring increases (Israel \& May, 2010; Rawding, Sharpe, \& Blankenship, 2014; Sard et al., in press). Eventually the number of unique parental genotypes will approach the total number of parents in the population, and that asymptote can be estimated $\left(\widehat{N_{s}}\right)$. This extrapolation decreases the bias in $N_{s}$ from limited sample sizes.

## Objectives

$N_{b}$ and $N_{s}$ have significant potential as assessment metrics for the estimation of sea lamprey spawning population size, but they need to be further validated prior to incorporation into management and control efforts. Genetic estimates of spawning abundance were utilized to assess barrier efficacy in three streams in Chapter 1, and across a larger number of streams to quantify associations between $N_{b}, N_{s}, N_{c}$ and the stream's management history (e.g., lampricide treatment interval) and environmental characteristics in Chapter 2. Environmental, biotic, and sampling variables were examined for associations with $N_{b}$ and $N_{s}$ estimates in these systems. Additionally, a subset of sequenced streams are also index streams for mark-recapture census size estimates for adult populations, and the potential correlations between $N_{b}, N_{s}$, and $N_{c}$ were examined. To evaluate the sampling and genotyping effort required to effectively estimate $N_{b}$ and $N_{s}$ across stream systems, simulations were conducted for a variety of population sizes to compare the accuracy and precision of $N_{b}$ and $N_{s}$ estimates as sample size and the number of SNP loci increased.

LITERATURE CITED

## LITERATURE CITED

Adams, J. V., Barber, J. M., Bravener, G. A., \& Lewandoski, S. A. (2021). Quantifying Great Lakes sea lamprey populations using an index of adults. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2021.04.009

Adams, J. V., \& Jones, M. L. (2020). Evidence of host switching: Sea lampreys disproportionately attack Chinook salmon when lake trout abundance is low in Lake Ontario. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2020.03.003

Anderson, J. H., Pess, G. R., Carmichael, R. W., Ford, M. J., Cooney, T. D., Baldwin, C. M., \& McClure, M. M. (2014). Planning Pacific Salmon and Steelhead Reintroductions Aimed at Long-Term Viability and Recovery. North American Journal of Fisheries Management, 34(1), 72-93. https://doi.org/10.1080/02755947.2013.847875

Applegate, V. C. (1950). Natural history of the sea lamprey, Petromyzon marinus in Michigan. Spec Sci Rep US Fish Wildl Serv, 55, 1-237. Retrieved from http://ci.nii.ac.jp/naid/10010684036/en/

Armbruster, P., \& Reed, D. H. (2005). Inbreeding depression in benign and stressful environments. Heredity, 95(3), 235-242. https://doi.org/10.1038/sj.hdy. 6800721

Barber, J., \& Steeves, M. (2019). Sea lamprey control the Great Lakes 2018: Annual report to the Great Lakes Fishery Commission. Detroit, Michigan. Retrieved from http://www.glfc.org/pubs/slcp/annual reports/ANNUAL_REPORT 2018.pdf

Bergstedt, R. A., \& Seelye, J. G. (1995). Evidence for Lack of Homing by Sea Lampreys. Transactions of the American Fisheries Society, 124(2), 235-239. https://doi.org/doi:10.1577/1548-8659(1995) $124<0235: E F L O H B>2.3 . C O ; 2$

Birceanu, O., McClelland, G. B., Wang, Y. S., \& Wilkie, M. P. (2009). Failure of ATP supply to match ATP demand: The mechanism of toxicity of the lampricide, 3-trifluoromethyl-4nitrophenol (TFM), used to control sea lamprey (Petromyzon marinus) populations in the Great Lakes. Aquatic Toxicology, 94(4), 265-274. https://doi.org/10.1016/j.aquatox.2009.07.012

Blomqvist, D., Pauliny, A., Larsson, M., \& Flodin, L. Å. (2010). Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. BMC Evolutionary Biology, 10(1), 1-9. https://doi.org/10.1186/1471-2148-10-33

Blouin, M. S. (2003). DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends in Ecology and Evolution, 18(10), 503-511. https://doi.org/10.1016/S0169-5347(03)00225-8

Boogaard, M. A., Bills, T. D., \& Johnson, D. A. (2003). Acute Toxicity of TFM and a

TFM/Niclosamide Mixture to Selected Species of Fish, Including Lake Sturgeon (Acipenser fulvescens) and Mudpuppies (Necturus maculosus), in Laboratory and Field Exposures. Journal of Great Lakes Research, 29, 529-541. https://doi.org/https://doi.org/10.1016/S0380-1330(03)70514-0

Christie, M. R., Sepúlveda, M. S., \& Dunlop, E. S. (2019). Rapid resistance to pesticide control is predicted to evolve in an invasive fish. Scientific Reports, 9(1), 1-13. https://doi.org/10.1038/s41598-019-54260-5

Cochran-Biederman, J. L., Wyman, K. E., French, W. E., \& Loppnow, G. L. (2015). Identifying correlates of success and failure of native freshwater fish reintroductions. Conservation Biology, 29(1), 175-186. https://doi.org/10.1111/cobi. 12374

Dawson, H. A., Higgins-weier, C. E., Steeves, T. B., \& Johnson, N. S. (2020). Estimating age and growth of invasive sea lamprey : A review of approaches and investigation of a new method. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2020.06.002

Dawson, H. A., Jones, M. L., Scribner, K. T., \& Gilmore, S. A. (2009). An Assessment of Age Determination Methods for Great Lakes Larval Sea Lampreys. North American Journal of Fisheries Management, 29(4), 914-927. https://doi.org/10.1577/m08-139.1

Dawson, H. A., Quintella, B. R., Almeida, P. R., Treble, A. J., \& Jolley, J. C. (2015). The Ecology of Larval and Metamorphosing Lampreys. In Lampreys: Biology, Conservation and Control (Vol. 1, pp. 75-137). https://doi.org/10.1007/978-94-017-9306-3

De Barba, M., Waits, L. P., Garton, E. O., Genovesi, P., Randi, E., Mustoni, A., \& Groff, C. (2010). The power of genetic monitoring for studying demography, ecology and genetics of a reintroduced brown bear population. Molecular Ecology, 19(18), 3938-3951. https://doi.org/10.1111/j.1365-294X.2010.04791.x

Dhamelincourt, M., Buoro, M., Rives, J., Sebihi, S., \& Tentelier, C. (2020). Individual and group characteristics affecting nest building in sea lamprey (Petromyzon marinus L. 1758). Journal of Fish Biology, (October 2020), 557-565. https://doi.org/10.1111/jfb. 14601

Dunlop, E. S., McLaughlin, R., Adams, J. V., Jones, M., Birceanu, O., Christie, M. R., ... Wilkie, M. P. (2018). Rapid evolution meets invasive species control: the potential for pesticide resistance in sea lamprey. Canadian Journal of Fisheries and Aquatic Sciences, 75(1), 152-168. https://doi.org/10.1139/cjfas-2017-0015

Ebener, M. P., Brenden, T. O., \& Jones, M. L. (2010). Estimates of fishing and natural mortality rates for four lake whitefish stocks in Northern Lakes Huron and Michigan. Journal of Great Lakes Research, 36(SUPPL. 1), 110-120. https://doi.org/10.1016/j.jglr.2009.06.003

England, P. R., Cornuet, J. M., Berthier, P., Tallmon, D. A., \& Luikart, G. (2006). Estimating effective population size from linkage disequilibrium: Severe bias in small samples. Conservation Genetics, 7(2), 303-308. https://doi.org/10.1007/s10592-005-9103-8

Evans, M. L., Johnson, M. A., Jacobson, D., Wang, J., Hogansen, M., \& O’Malley, K. G. (2015). Evaluating a multi-generational reintroduction program for threatened salmon using genetic parentage analysis. Canadian Journal of Fisheries and Aquatic Sciences, 73(5), 844-852. https://doi.org/10.1139/cjfas-2015-0317

Fitzpatrick, S. W., Gerberich, J. C., Angeloni, L. M., Bailey, L. L., Broder, E. D., TorresDowdall, J., ... Funk, C. W. (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(7), 879-891. https://doi.org/10.1111/eva. 12356

Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation, 143(9), 1919-1927. https://doi.org/10.1016/J.BIOCON.2010.05.011

Frankham, R. (2015). Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. Molecular Ecology, 24(11), 2610-2618. https://doi.org/10.1111/mec. 13139

Frankham, R., Bradshaw, C. J. A., \& Brook, B. W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation, 170, 56-63. https://doi.org/10.1016/j.biocon.2013.12.036

Franklin, I. R. (1980). Evolutionary Change in Small Populations. In Conservation Biology. An evolutionary-ecological perspective (pp. 135-150).

Franklin, I. R., \& Frankham, R. (1998). How large must populations be to retain evolutionary potential? Animal Conservation, 1, 69-73.

Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M., ... Luikart, G. (2016). Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application. Trends in Ecology and Evolution, 31(2), 81-83.
https://doi.org/10.1016/j.tree.2015.10.009
Gossieaux, P., Bernatchez, L., Sirois, P., \& Garant, D. (2019). Impacts of stocking and its intensity on effective population size in Brook Charr (Salvelinus fontinalis) populations. Conservation Genetics, 20(4), 729-742. https://doi.org/10.1007/s10592-019-01168-2

Griffiths, R. W., Beamish, F. W. H., Morrison, B. J., \& Barker, L. A. (2001). Factors Affecting Larval Sea Lamprey Growth and Length at Metamorphosis in Lampricide-Treated Streams. Transactions of the American Fisheries Society, 130(2), 289-306. https://doi.org/10.1577/1548-8659(2001)130<0289:falslg>2.0.co;2

Hansen, M. J., Adams, J. V, Cuddy, D. W., Richards, J. M., Fodale, M. F., Larson, G. L., ... Zerrenner, A. (2003). Optimizing Larval Assessment to Support Sea Lamprey Control in the Great Lakes. Journal of Great Lakes Research, 29, 766-782.
https://doi.org/https://doi.org/10.1016/S0380-1330(03)70530-9
Hardisty, M. W., \& Potter, I. C. (1971). The Biology Of Lampreys. In Academic Press (Vol. 1). New York. https://doi.org/10.1126/science.176.4042.1409

Harper, D. L. M., Horrocks, J., Barber, J., Bravener, G. A., Schwarz, C. J., \& McLaughlin, R. L. (2018). An evaluation of statistical methods for estimating abundances of migrating adult sea lamprey. Journal of Great Lakes Research, 44(6), 1362-1372.
https://doi.org/10.1016/j.jglr.2018.08.004
Harvey, C. J., Ebener, M. P., \& White, C. K. (2008). Spatial and ontogenetic variability of sea lamprey diets in Lake Superior. Journal of Great Lakes Research, 34(3), 434-449. https://doi.org/10.3394/0380-1330(2008)34[434:SAOVOS]2.0.CO;2

Heber, S., Briskie, J. V., \& Apiolaza, L. A. (2012). A test of the "genetic rescue" technique using bottlenecked donor populations of Drosophila melanogaster. PLoS ONE, 7(8). https://doi.org/10.1371/journal.pone. 0043113

Heinrich, J. W., Mullett, K. M., Hansen, M. J., Adams, J. V, Klar, G. T., Johnson, D. A., ... Young, R. J. (2003). Sea Lamprey Abundance and Management in Lake Superior, 1957 to 1999. Journal of Great Lakes Research, 29, 566-583. Retrieved from https://www.sciencedirect.com/science/article/pii/S0380133003705176

Henson, M. P., Bergstedt, R. A., \& Adams, J. V. (2003). Comparison of spring measures of length, weight, and condition factor for predicting metamorphosis in two populations of sea lamprey (Petromyzon marinus) larvae. Journal of Great Lakes Research, 29(SUPPL. 1), 204-213. https://doi.org/10.1016/S0380-1330(03)70489-4

Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. Genetical Research, 38(03), 209. https://doi.org/10.1017/S0016672300020553

Hoban, S., Bruford, M. W., Funk, W. C., Galbusera, P., Griffith, M. P., Grueber, C. E., ... Vernesi, C. (2021). Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible. BioScience, XX(X), 1-13. https://doi.org/10.1093/biosci/biab054

Hunter, R. D., Roseman, E. F., Sard, N. M., DeBruyne, R. L., Wang, J., \& Scribner, K. T. (2020). Genetic Family Reconstruction Characterizes Lake Sturgeon Use of Newly Constructed Spawning Habitat and Larval Dispersal. Transactions of the American Fisheries Society, 266-283. https://doi.org/10.1002/tafs. 10225

Israel, J. A., \& May, B. (2010). Indirect genetic estimates of breeding population size in the polyploid green sturgeon (Acipenser medirostris). Molecular Ecology, 19(5), 1058-1070. https://doi.org/10.1111/j.1365-294X.2010.04533.x

Jensen, A. J., \& Jones, M. L. (2018). Forecasting the response of Great Lakes sea lamprey
(Petromyzon marinus) to barrier removals. Canadian Journal of Fisheries and Aquatic Science, 75(9), 1415-1426. https://doi.org/10.1139/cjfas-2017-0243

Johnson, N. S., Buchinger, T. J., \& Li, W. (2015). Reproductive Ecology of Lampreys BT Lampreys: Biology, Conservation and Control: Volume 1. In M. F. Docker (Ed.) (pp. 265303). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9306-3_6

Jones, M. L. (2007). Toward Improved Assessment of Sea Lamprey Population Dynamics in Support of Cost-effective Sea Lamprey Management. Journal of Great Lakes Research, 33(2), 35-47. https://doi.org/10.3394/0380-1330(2007)33

Keogh, J. S., Webb, J. K., \& Shine, R. (2007). Spatial genetic analysis and long-term markrecapture data demonstrate male-biased dispersal in a snake. Biology Letters, 3(1), 33-35. https://doi.org/10.1098/rsbl.2006.0570

Koonce, J. F., Eshenroder, R. L., \& Christie, G. C. (1993). An Economic Injury Level Approach to Establishing the Intensity of Sea Lamprey Control in the Great Lakes. North American Journal of Fisheries Management, 13(1), 1-14. https://doi.org/10.1577/1548$\underline{8675(1993) 013<0001: A E I L A T>2.3 . C O ; 2}$

Lavis, D. S., Hallett, A., Koon, E. M., \& McAuley, T. C. (2003). History of and advances in barriers as an alternative method to suppress sea lampreys in the Great Lakes. Journal of Great Lakes Research, 29(SUPPL. 1), 362-372. https://doi.org/10.1016/S0380-1330(03)70500-0

Lawrence, M. J., Mitrovic, D., Foubister, D., Bragg, L. M., Sutherby, J., Docker, M. F., ... Jeffries, K. M. (2021). Contrasting physiological responses between invasive sea lamprey and non-target bluegill in response to acute lampricide exposure. Aquatic Toxicology, 237, 105848. https://doi.org/10.1016/j.aquatox.2021.105848

Lawrie, A. H. (1970). The Sea Lamprey in the Great Lakes. Transactions of the American Fisheries Society, 99(4), 766-775. https://doi.org/10.1577/15488659(1970)99<766:TSLITG>2.0.CO;2

Lech, J. J., \& Statham, C. N. (1975). Role of glucuronide formation in the selective toxicity of 3-trifluoromethyl-4-nitrophenol (TFM) for the sea lamprey: comparative aspects of TFM uptake and conjugation in sea lamprey and rainbow trout. Toxicology and Applied Pharmacology, 31(1), 150-158. https://doi.org/10.1016/0041-008X(75)90063-0

Mace, G. M., Collar, N. J., Gaston, K. J., Hilton-Taylor, C., Akçakaya, H. R., Leader-Williams, N., ... Stuart, S. N. (2008). Quantification of extinction risk: IUCN's system for classifying threatened species. Conservation Biology, 22(6), 1424-1442.
https://doi.org/10.1111/j.1523-1739.2008.01044.x
Manion, P. J., \& Smith, B. R. (1978). Biology of larval and metamorphosing sea lampreys, Petromyzon marinus, of the 1960 year class in the Big Garlic River, Michigan, Part II,

1966-1972. Great Lakes Fishery Commission Technical Report, 30, 1-37.
McDonald, D. G., \& Kolar, C. S. (2007). Research to Guide the Use of Lampricides for Controlling Sea Lamprey. Journal of Great Lakes Research, 33, 20-34. https://doi.org/https://doi.org/10.3394/0380-1330(2007)33[20:RTGTUO]2.0.CO;2

McLaughlin, R. L., Ellen Marsden, J., \& Hayes, D. B. (2003). Achieving the Benefits of Sea Lamprey Control While Minimizing Effects on Nontarget Species: Conceptual Synthesis and Proposed Policy. Journal of Great Lakes Research, 29, 755-765. https://doi.org/https://doi.org/10.1016/S0380-1330(03)70529-2

McLaughlin, R. L., Hallett, A., Pratt, T. C., O’Connor, L. M., \& McDonald, D. G. (2007). Research to Guide Use of Barriers, Traps, and Fishways to Control Sea Lamprey. Journal of Great Lakes Research, 33(2), 7-19. https://doi.org/10.3394/0380-1330(2007)33

McLeod, D. V., Cottrill, R. A., \& Morbey, Y. E. (2011). Sea lamprey wounding in Canadian waters of Lake Huron from 2000 to 2009: Temporal changes differ among regions. Journal of Great Lakes Research, 37(4), 601-608. https://doi.org/10.1016/j.jglr.2011.08.003

Morkert, S. B., Swink, W. D., \& Seelye, J. G. (1998). Evidence for Early Metamorphosis of Sea Lampreys in the Chippewa River, Michigan. North American Journal of Fisheries Management, 18(4), 966-971. https://doi.org/10.1577/15488675(1998)018<0966:efemos>2.0.co;2

Mullett, K. M., Heinrich, J. W., Adams, J. V., Young, R. J., Henson, M. P., McDonald, R. B., \& Fodale, M. F. (2003). Estimating Lake-wide Abundance of Spawning-phase Sea Lampreys (Petromyzon marinus) in the Great Lakes: Extrapolating from Sampled Streams Using Regression Models. Journal of Great Lakes Research, 29, 240-252.
https://doi.org/10.1016/S0380-1330(03)70492-4
Ovenden, J. R., Leigh, G. M., Blower, D. C., Jones, A. T., Moore, A., Bustamante, C., ... Dudgeon, C. L. (2016). Can estimates of genetic effective population size contribute to fisheries stock assessments? Journal of Fish Biology, 89(6), 2505-2518. https://doi.org/10.1111/jfb. 13129

Petereit, C., Bekkevold, D., Nickel, S., Dierking, J., Hantke, H., Hahn, A., ... Puebla, O. (2018). Population genetic structure after 125 years of stocking in sea trout (Salmo trutta L.). Conservation Genetics, 19(5), 1123-1136. https://doi.org/10.1007/s10592-018-1083-6

Peterson, N. P., \& Cederholm, C. J. (1984). A Comparison of the Removal and Mark-Recapture Methods of Population Estimation for Juvenile Coho Salmon in a Small Stream. North American Journal of Fisheries Management, 4(1), 99-102. https://doi.org/10.1577/15488659(1984)4<99:acotra>2.0.co;2

Potter, I. C. (1980). Ecology of Larval and Metamorphosing Lampreys. Canadian Journal of Fisheries and Aquatic Sciences, 37(11), 1641-1657. https://doi.org/10.1139/f80-212

Pratt, T. C., Morrison, B. J., Quinlan, H. R., Elliott, R. F., Grunder, S. A., Chiotti, J. A., \& Young, B. A. (2020). Implications of the sea lamprey control program for lake sturgeon conservation and rehabilitation efforts. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2020.06.014

Pycha, R. L., \& King, G. R. (1975). Changes in the Lake Trout Population of Southern Lake Superior in Relation to the Fishery, the Sea Lamprey, and Stocking, 1950-1970. Great Lakes Fishery Commission Technical Report.

Rawding, D. J., Sharpe, C. S., \& Blankenship, S. M. (2014). Genetic-Based Estimates of Adult Chinook Salmon Spawner Abundance from Carcass Surveys and Juvenile Out-Migrant Traps. Transactions of the American Fisheries Society, 143(1), 55-67. https://doi.org/10.1080/00028487.2013.829122

Rhie, A., Mccarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., London, S. E., ... Friedrich, S. R. (2020). Towards complete and error-free genome assemblies of all vertebrate species, $1-$ 56.
https://doi.org/10.1101/2020.05.22.110833
Robinson, K. F., Miehls, S. M., \& Siefkes, M. J. (2021). Understanding sea lamprey abundances in the Great Lakes prior to broad implementation of sea lamprey control. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2021.04.002

Sard, N. M., Hunter, R. D., Roseman, E. F., Hayes, D. B., DeBruyne, R. L., \& Scribner, K. T. (n.d.). Extending non-parametric species richness estimators to genetic pedigree rarefaction for breeding adult estimation. In press.

Sard, N. M., Smith, S. R., Homola, J. J., Kanefsky, J., Bravener, G., Adams, J. V., ... Scribner, K. T. (2020). RAPTURE (RAD capture) panel facilitates analyses characterizing sea lamprey reproductive ecology and movement dynamics. Ecology and Evolution, (December 2019), 1-20. https://doi.org/10.1002/ece3.6001

Smith, B. R., \& Tibbles, J. J. (1980). Sea Lamprey (Petromyzon marinus) in Lakes Huron, Michigan, and Superior: History of Invasion and Control, 1936-78. Canadian Journal of Fisheries and Aquatic Sciences, 37(37), 1780-1801. Retrieved from www.nrcresearchpress.com

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., ... Li, W. (2013). Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nature Genetics, 45(4), 415-421. https://doi.org/10.1038/ng. 2568

Smith, J. J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M. C., Parker, H. J., ... Amemiya, C. T. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. Nature Genetics, 50(2), 270-277. https://doi.org/10.1038/s41588-017-0036-1

Soule, M. (1980). Thresholds for survival: maintaining fitness and evolutionary potential. In Conservation Biology. An evolutionary-ecological perspective (Vol. 1st, pp. 151-169). Sinauer Associates. Retrieved from http://ci.nii.ac.jp/naid/10018127538/en/

Sullivan, P. W., Adair, R., \& Woldt, A. (2016). Sea Lamprey Control in the Great Lakes 2015: Annual Report to the Great Lakes Fishery Commission. Ottowa, Ontario.

Swink, W. D. (2003). Host Selection and Lethality of Attacks by Sea Lampreys (Petromyzon marinus) in Laboratory Studies. Journal of Great Lakes Research, 29, 307-319. https://doi.org/10.1016/S0380-1330(03)70496-1

Twohey, M. B., Heinrich, J. W., Seelye, J. G., Fredricks, K. T., Bergstedt, R. A., Kaye, C. A., ... Christie, G. C. (2003). The Sterile-Male-Release Technique in Great Lakes Sea Lamprey Management. Journal of Great Lakes Research, 29, 410-423. https://doi.org/https://doi.org/10.1016/S0380-1330(03)70504-8

Waldman, J., Grunwald, C., \& Wirgin, I. (2008). Sea lamprey (Petromyzon marinus): an exception to the rule of homing in anadromous fishes. Biology Letters, 4(6), 659-662. https://doi.org/10.1098/rsbl.2008.0341

Wang, J., \& Santure, A. W. (2009). Parentage and sibship inference from multilocus genotype data under polygamy. Genetics, 181(4), 1579-1594.
https://doi.org/10.1534/genetics.108.100214
Wang, Jinliang. (2004). Sibship reconstruction from genetic data with typing errors. Genetics, 166(4), 1963-1979.

Wang, Jinliang. (2009). A new method for estimating effective population sizes from a single sample of multilocus genotypes. Molecular Ecology, 18(10), 2148-2164.
https://doi.org/10.1111/j.1365-294X.2009.04175.x
Waples, R S. (1990). Conservation Genetics of Pacific Salmon. II. Effective Population Size and the Rate of Loss of Genetic Variability. Journal of Heredity. Retrieved from https://academic.oup.com/jhered/article/81/4/267/808935

Waples, R. S., Antao, T., \& Luikart, G. (2014). Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics, 197(2), 769-780. https://doi.org/10.1534/genetics.114.164822

Waples, R. S., \& Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. Evolutionary Applications, 3(3), 244-262. https://doi.org/10.1111/j.17524571.2009.00104.x

Waples, R. S., Luikart, G., Faulkner, J. R., \& Tallmon, D. A. (2013). Simple life-history traits explain key effective population size ratios across diverse taxa. Proceedings of the Royal

Society B: Biological Sciences, 280(1768), 20131339.
https://doi.org/10.1098/rspb.2013.1339
Waples, R. S., \& Waples, R. K. (2011). Inbreeding effective population size and parentage analysis without parents. Molecular Ecology Resources, 11(1), 162-171.
https://doi.org/10.1111/j.1755-0998.2010.02942.x
Weisser, J. W., Adams, J. V., Schuldt, R. J., Baldwin, G. A., Lavis, D. S., Slade, J. W., \& Heinrich, J. W. (2003). Effects of Repeated TFM Applications on Riffle Macroinvertebrate Communities in Four Great Lakes Tributaries. Journal of Great Lakes Research, 29, 552565. https://doi.org/https://doi.org/10.1016/S0380-1330(03)70516-4

Wright, S. (1931). Evolution in Mendelian populations. Genetics, 16, 97-159. https://doi.org/10.1007/BF02459575

# CHAPTER 1: PEDIGREE ANALYSIS AND ESTIMATES OF EFFECTIVE BREEDING SIZE CHARACTERIZE SEA LAMPREY REPRODUCTIVE BIOLOGY 


#### Abstract

The sea lamprey (Petromyzon marinus) is an invasive species in the Great Lakes and the focus of a large control and assessment program. Current assessment methods provide information on the census size of spawning adult sea lamprey in a small number of streams, but information characterizing reproductive success of spawning adults is rarely available. We used RAD-capture sequencing to genotype single nucleotide polymorphism (SNP) loci for $\sim 1600$ sea lamprey larvae collected from three streams in northern Michigan (Black Mallard, Pigeon, and Ocqueoc Rivers). Larval genotypes were used to reconstruct family pedigrees, which were combined with Gaussian mixture analyses to identify larval age classes for estimation of spawning population size. Three complementary estimates of effective breeding size $\left(N_{b}\right)$, as well as the extrapolated minimum number of spawners $\left(N_{s}\right)$, were also generated for each cohort. Reconstructed pedigrees highlighted inaccuracies of cohort assignments from traditionally used mixture analyses. However, combining genotype-based pedigree information with length-at-age assignment of cohort membership greatly improved cohort identification accuracy. Population estimates across all three streams sampled in this study indicate a small number of successfully spawning adults when barriers were in operation, implying that barriers limited adult spawning numbers but were not completely effective at blocking access to spawning habitats. Thus, the large numbers of larvae present in sampled systems were a poor indicator of spawning adult abundance. Overall, pedigree-based $N_{b}$ and $N_{s}$ estimates provide a promising and rapid assessment tool for sea lamprey and other species.


## INTRODUCTION

Invasive species are a substantial threat to biodiversity and management intervention is often required to mitigate their effects on the ecosystem. Annual control programs to reduce the population size of widespread invasive species (Prior, Adams, Klepzig, \& Hulcr, 2018) often include strategies to reduce recruitment and spread, like barriers that limit access to spawning habitat (Sharov \& Liebhold, 1998). More recently, genetic control techniques like the release of sterile individuals or gene drive have been developed as additional options for control (Bajer et al., 2019).

Genetic technology, used in combination with field techniques, allow managers opportunities to efficiently and cost-effectively sample large areas to quantify the presence of species, community composition, and species biomass and abundance. Environmental DNA was used as an early detection tool for specific invasive species like American Bullfrogs (Lithobates catesbeianus) and invasive shellfish species, allowing for rapid response after the invasion (Dejean et al., 2012; Leblanc et al., 2020). To evaluate widespread invasions, demographic modeling has been used to track the spread of invasive species across a system to determine the introduction point and generate hypotheses for the mechanism of introduction (Blakeslee et al., 2017; Sherpa et al., 2019). Additionally, determining the founding effective size of an invasive population can provide insight into the mechanism of invasion and the severity of the bottleneck present in an introduced species (Sard, Robinson, Kanefsky, Herbst, \& Scribner, 2019). Genetic parentage assessment and effective size estimates can be used to evaluate the size and diversity of spawning populations as an annual assessment tool for managed populations (Taylor, Bangs, \& Long, 2021). This can be used to evaluate the success of control efforts for an invasive species.

Sea lamprey (Petromyzon marinus) are a widespread invasive species in the Laurentian Great Lakes (McGeoch et al., 2010). The expansion of the Welland Canal in 1919 allowed sea lamprey to spread from Lake Ontario to the rest of the Great Lakes by 1938 (Lawrie, 1970). Sea lamprey contributed to major declines in commercially valuable fish species like lake trout (Salvelinus namaycush) and lake whitefish (Coregonus clupeaformis) throughout the Great Lakes basin (Heinrich et al., 2003; Koonce et al., 1993; Lawrie, 1970). As a result of the ecological and economic impacts of the invasion, an annual control and assessment program was implemented in the 1950s to reduce sea lamprey abundance and assist recovery of native fish populations (Smith \& Tibbles, 1980).

The primary methods of sea lamprey control since the 1950s have been physical barriers that block adults from reaching spawning habitat and application of the selective lampricide 3-trifluormethlyl-4-nitrophenol (TFM) to kill larvae (Applegate, 1950; McDonald and Kolar, 2007; Smith and Tibbles, 1980). Several barrier designs have been implemented since the beginning of the control program to reduce migration of sea lamprey into streams (Lavis et al., 2003; McLaughlin et al., 2007). However, these barriers also impede the movement of numerous ecologically and culturally important native fish species (Jensen \& Jones, 2018b). Adjustments and alternative barrier designs have been used to reduce effects on native fish (Katopodis et al., 2009), such as seasonal electric barriers or the addition of a fish ladder (Lavis et al., 2003; Zielinski et al., 2019). Many barriers have been removed altogether, resulting in an increase in spawning habitat for sea lamprey throughout the Great Lakes. Additionally, sea lamprey larvae are occasionally found upstream in systems with barriers. In these cases, managers want to know when and how many adult sea lamprey escaped upstream of the barrier, but given uncertainty in stock-recruitment relationships and a limited ability to age larvae, these questions are largely
unanswered (Dawson, Jones, Scribner, \& Gilmore, 2009; Jones, 2007). Population genetic data can address these questions by estimating the number of successfully spawning adults that contributed to a year class of larvae and tracking the movements of individuals from each year class over several years (Ovenden et al., 2016; Sard et al., 2020).

Sea lamprey have a multistage anadromous life history that can span up to 9 years (Applegate, 1950). Adults migrate upstream, spawn in spring and summer, and die afterward (Johnson et al., 2015). Larvae reside in streams and lentic areas near streams and feed on algae and detritus while burrowed into soft sediment (Dawson et al., 2015). After two (Morkert et al., 1998) to seven years (Manion \& Smith, 1978) in the larval stage, larvae undergo metamorphosis, migrate to the Great Lakes, and feed on fishes for 12-18 months. Adult sea lamprey do not return to natal streams to spawn (Bergstedt \& Seelye, 1995b), but instead stream selection is guided by chemosensory cues released by larval sea lamprey (Fissette et al., 2021). Therefore, population structure of sea lamprey is weak relative to other homing fishes (Bryan et al., 2005). Key uncertainties regarding sea lamprey demographics include stock-recruitment relationships (Dawson \& Jones, 2009), larval survival (Jones et al., 2009), and age at metamorphosis (Griffiths et al., 2001; Treble, Jones, \& Steeves, 2008) in part, because of difficulty aging larvae (Dawson, Higgins-Weier, Steeves, \& Johnson, 2020).

Recent developments in sequencing technologies, the declining costs of high-throughput sequencing, and expanding genomic resources for sea lamprey (Smith et al., 2013, 2018; Sard et al., 2020) present an opportunity to incorporate population genomic methods and data analysis into invasive species assessment efforts. Reduced representation sequencing technologies such as Restriction-site Associated DNA (RAD) sequencing (Baird et al., 2008) and locus-targeted RAD-Capture (Ali et al., 2016) allow for the collection of genome-scale data from large
population-level sample sizes. The use of genomic data to study invasive species populations offers numerous applications to assist managers in assessing sea lamprey reproductive ecology in natural stream settings. These data also provide a means to evaluate the effectiveness of experimental barriers and gain additional insight into sea lamprey reproductive ecology in Great Lakes tributaries.

Several parameters are routinely estimated based on genetic data to quantify spawning adult abundance and reproductive success (e.g., Sard et al. 2020 for sea lamprey). Effective population size $\left(N_{e}\right)$ is the size of an idealized population that experiences the same amount of genetic drift, inbreeding, or loss of diversity as the population in question (Wright, 1931). $N_{e}$ has been used in assessments of populations and as an indicator of potential for future declines in abundance (Antao, Pérez-Figueroa, \& Luikart, 2011). Low $N_{e}$ can also be an indicator of low levels of genetic diversity in a population (Frankham, 2010). In many species, multiple generations produce offspring simultaneously, resulting in overlapping generations (Waples, Antao, \& Luikart, 2014). In this situation, the effective number of breeding individuals contributing to a spawning event $\left(N_{b}\right)$ can also be estimated using samples from a single year class (Robinson \& Moyer, 2013; Waples et al., 2014; Waples \& Do, 2010). $N_{e}$ can be reduced relative to census size by several factors, including skewed sex ratios and variation in reproductive success (Waples, 2010). The ratio of $N_{b}$ to $N_{e}$ has been shown to be strongly associated with life history traits such as time to sexual maturity and adult lifespan (Waples, Luikart, Faulkner, \& Tallmon, 2013). In addition to $N_{b}$, the minimum number of spawning adults $\left(N_{s}\right)$ can also be calculated from reconstructed pedigrees as the minimum number of parental genotypes required to produce the sampled offspring genotypes. Using approaches to estimate total species richness from the field of community ecology (Chao 1987; Heltshe and Forrester
2009), information on the contribution of inferred parental genotypes to sampled larvae can provide estimates of the total number of parents contributing to a cohort (Hunter et al., 2020), including asymptotic estimates of total spawning adult numbers (Sard et al., in press).
$N_{b}$ can be estimated from population genetic or genomic data using several methods. Here we apply three approaches to estimate sea lamprey effective breeding size: linkage disequilibrium (LD; Waples and Do 2010), sibship frequency (SF; Wang 2009), and parentage-without-parents (PwoP; Waples and Waples 2011). The LD method uses non-random associations of alleles across loci that result from finite population size or physical linkage (Hill 1981a,b). If chromosomal locations of loci can be established and effects of physical linkage can be removed, LD resulting from finite breeding population size can be estimated to characterize effective breeding size (Waples, Larson, \& Waples, 2016). In contrast, SF and PwoP both use reconstructed pedigrees, where sampled offspring are used to reconstruct unsampled parental genotypes (Bravington, Skaug, \& Anderson, 2016; De Barba et al., 2010; Keogh et al., 2007). SF uses the frequency of sibling relationships identified in the pedigree to infer $N_{b}$ (Wang, 2009), while the PwoP method uses the mean and variance in reproductive success of parents reconstructed from the sampled individuals in the pedigree (Waples \& Waples, 2011). Notably, both the SF and PwoP methods rely on reconstructed pedigrees for the sampled offspring and they are known to provide equivalent, but not identical, estimates of $N_{b}$ (Ackerman et al., 2017).

In this study, our objective was to estimate effective breeding size and minimum number of spawners for larval sea lamprey cohorts collected from streams above barriers to upstream migration in three locations in the northern Lower Peninsula of Michigan: the Black Mallard, Pigeon, and Ocqueoc Rivers. In all three locations, the presence of larvae upstream of barrier locations raised concerns about barrier failure to impede spawning migrations. We used the
estimates above to evaluate barrier efficacy in all three systems. Furthermore, we used reconstructed pedigrees of each collection along with Gaussian mixture analysis to estimate the number of larval age classes present in each system. We discuss possible explanations for barrier failure in these systems, highlight the utility of population genomic data for rapid assessment of spawning populations and how genetic data can be integrated into monitoring and control efforts for invasive species.

## METHODS

## Study System and Sample Collection

Sampling of larval sea lamprey was conducted in the Black Mallard, Ocqueoc, and Pigeon Rivers, which are located in the northern Lower Peninsula of Michigan, USA (Figure 1). In all three systems, larval sea lamprey were collected above barriers designed to preclude access to spawning habitat. The spatial extent of sampling was extensive in all rivers to define the distribution of the larval sea lamprey infestations and to obtain a comprehensive spatial representation of larvae produced from all family groups.

The Black Mallard River had an electric barrier installed in 2016 following a lampricide treatment that occurred in June 2015. In September 2017, larvae in the section of the Black Mallard River downstream from Black Mallard Lake were collected using backpack electrofishing $(\mathrm{n}=387)$. Sea lamprey were sampled from habitat spanning 500 m upstream and downstream of Ocqueoc Lake Road and U.S. Highway 23. These two sampling points represented the furthest upstream and downstream extent of the lower river with stream substrate suitable for larval sea lamprey, and covered about $50 \%$ of the available larval habitat in the lower river. Lampricide treatment of the Black Mallard River downstream of Black Mallard Lake occurred in July 2018, and dead sea lamprey larvae were collected post treatment by two staff that walked the entire stream length from Ocqueoc Lake Road to U.S. 23 ( $\mathrm{n}=667$ ). These collections will be referred to hereafter as the 'Lower Black Mallard River.' Variation in larval length in the samples raised concerns that larvae might include individuals from multiple age classes that would indicate that the barrier had failed repeatedly. Larvae were also collected upstream of Black Mallard Lake in May 2019 when lampricide was applied. Two staff walked 2 km downstream and 2 km upstream from Elah Road, and covered the entire known distribution
of larval sea lamprey in the upper river. Surveys were also conducted upstream and downstream of Elah Road post lampricide treatment but no sea lampreys were found. This collection will be referred to hereafter as the 'Upper Black Mallard River.'

The Ocqueoc River has had an electric barrier in place since 1951 (Smith \& Tibbles, 1980), with a permanent barrier installed since 1999. The area upstream of the barrier is the site of annual experiments that involve the release of thousands of adult female sea lamprey (Buchinger et al., 2020; Johnson, Thompson, Holbrook, \& Tix, 2014; Wagner, Hanson, Meckley, Johnson, \& Bals, 2018). Adult males are not included in experimental releases, so no successful spawning was expected in the system. However, a population of larvae was found above the barrier in 2018 and surveys conducted throughout the river identified a roughly 5 km infested reach downstream of Ocqueoc Falls. Lampricide was subsequently applied in the stream in September 2018 and larvae were collected during treatment using dip nets and drift nets by four staff that walked the entire infested area $(\mathrm{n}=396)$. Surveys for dead sea lamprey were also conducted at Pomranke Road (5 km downstream of infested area) and in Silver Creek (tributary to Ocqueoc River), but no sea lampreys were found.

The Cheboygan River system has a dam at the mouth of the river, but has small sea lamprey populations which complete the juvenile parasitic phase of their life cycle in several upstream lake and stream systems; the Pigeon River is one such tributary (Johnson et al., 2020). To depress or eradicate these populations, releases of sterile males have been used as a supplemental control technique to limit successful female reproduction (Johnson et al., 2020; Kaye et al., 2003; M. Twohey, 2016). During these efforts, a small number of larvae ( $\mathrm{n}=29$ ) were found at Webb Road in the Pigeon River in September 2018. Ten other locations spanning
a 55 km section of the Pigeon river were also sampled in 2018 (some upstream and some downstream), but no sea lamprey were collected at those other sites.

Sea lamprey collected from all systems were euthanized, preserved in $95 \%$ ethanol and returned to the lab. Length and weight were measured for each individual sampled, to estimate age class. A tissue sample was taken for genetic analysis.


Figure 1.1. Map of the study area where larval sea lamprey were collected. The Black Mallard River is separated into upper and lower sections by Black Mallard Lake. The top-right inset shows the location of the sampled river systems in the Great Lakes region. River lines in black denote sampling locations of the river systems, blue lines denote all other rivers in the region.

## RAD-capture Sequencing

DNA was extracted from each larva using DNeasy blood and tissue kits (QIAGEN, Carlsbad, CA). DNA concentrations were initially quantified using a Nanodrop ND-1000 Spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts) and Quant-iT ${ }^{\mathrm{TM}}$ PicoGreen ${ }^{\text {TM }}$ dsDNA Assay Kits (ThermoFisher Scientific, Waltham, Massachusetts) on a QuantStudio 6 Flex Real-Time PCR system (Thermo Fisher Scientific Inc., Waltham, Massachusetts). Samples were standardized to a concentration of $10 \mathrm{ng} / 1$ for RAD sequencing. RAD library preparation was performed on 100 ng of DNA per individual using a modified version of the BestRAD protocol (Ali et al., 2016). DNA was digested using an $\operatorname{SbfI}$ restriction enzyme, and a biotinylated BestRAD adaptor was ligated to the DNA, which functioned as an individual barcode. DNA from groups of 96 barcoded individuals was pooled, concentrated, and sheared using a Covaris m220 focused-ultrasonicator (Covaris, Woburn, Massachusetts) using manufacturer recommended settings for a fragment size of 325 bp . Next, a streptavidin bead binding assay was used to select DNA fragments with RAD tags attached and a size selection was used to select only the target size fragments for sequencing. Size selection was done using Ampure beads with a 22:50 ratio to select long fragments and a 13:72 ratio to separate target size fragments from shorter fragments. Finally, NEBNext Kits (New England BioLabs Inc, Ipswich, Massachusetts) were used to ligate plate-specific Illumina adaptors and a universal adaptor for sequencing.

Library concentrations were quantified using a Picogreen assay, and the quality of the library was assessed via Tapestation (Agilent, Santa Clara, California) analysis. Libraries were pooled in groups of four to be enriched for a set of 3446 RAD loci that are known to be variable in sea lamprey populations (Sard et al., 2020). Loci were targeted using the RAD-capture
approach (Ali et al., 2016) with a custom MyBaits hybridization capture kit (Arbor Biosciences, Ann Arbor, MI) following the manufacturer recommended protocol. Eleven cycles were used in the final amplification step in the capture kit. Libraries were sequenced on four Illumina HighSeq X lanes at Novogene (Chula Vista, CA) using paired-end 150 base pair sequencing. Sequencing data for the project are available on the NCBI sequence read archive (Accession \#: will be provided prior to publication).

## Genotyping Analysis

Raw sequence data were processed using a bioinformatic pipeline described in Sard et al. (2020). Prior to the pipeline, a quality control report was constructed for each library using FastQC (Andrews, 2010) and evaluated. First, sequences from the HighSeq X run were oriented using the custom perl function bRAD_flip_trim.pl (originally developed by Paul Hohenlohe, University of Idaho, and modified by Brian Hand and Seth Smith, University of Montana) and demultiplexed using the Stacks 2.0 (Catchen, Hohenlohe, Bassham, Amores, \& Cresko, 2013) module 'process_radtags'. PCR duplicates were removed using 'clone_filter'. Next, sequences were quality trimmed using Trimmomatic with a minimum length of 50 , a sliding window of 4 bases, and a minimum quality score of 15 (Bolger, Lohse, \& Usadel, 2014). Sequences were then mapped to the sea lamprey reference genome (Smith et al., 2018), and indexed using bwa and bwa-mem (Li, 2013; Li \& Durbin, 2010). Samtools (version 1.9) was used to sort reads with default settings (Li et al., 2009). Genotypes were called using the Stacks 2.4 (Catchen et al., 2013) module 'gstacks', and the module 'populations' was used generate a .vcf file containing genotypes for all individuals. To avoid the inclusion of paralagous loci in the data set, the software HDplot (McKinney, Waples, Seeb, \& Seeb, 2017) was used to identify and exclude
potential paralogs. Loci were removed if observed heterozygosity was $>0.6$ or the read ratio deviation statistic (D; McKinney et al., 2017) in heterozygotes was greater than 7 in absolute magnitude. Individuals with more than $80 \%$ missing SNPs in the set were removed from analysis to minimize missing data. Each SNP set was checked for significant deviance from HardyWeinberg equilibrium across populations using the output from the Stacks 2.0 'populations' function prior to use in downstream analyses. Final genotype calls were filtered to exclude samples with $<8 \mathrm{X}$ coverage. To determine which SNPs were located on the sections of the genome targeted by the RAD-capture baits, the position of each SNP were compared to the genome position ranges for each RAD-capture tag (Sard et al., 2020).

To ensure that all individuals were sea lamprey samples rather than misidentified native Northern or American brook lamprey (Ichthyomyzon fossor; Lampetra appendix), comparative analyses were conducted. RAD-capture sequences of known American and Northern brook lamprey were aligned to the sea lamprey genome along with sampled individuals. A principal component analysis (PCA) was conducted for both native lamprey species and sampled individuals to identify clusters of individuals based on genotypes. All sampled individuals were compared to look for individuals that were identified as lamprey but clustered with native species, and none were found (Figure S.1). Additionally, neighbor-joining phylogenetic trees were constructed using SNP differences as an additional check for misidentified individuals, and all trees separated along species lines with no sampled individuals sorted with either native lamprey species.


Figure S.1. Visualization of principal component analysis (PCA) used to compare sea lamprey larval individuals from two native lamprey species (Lethenteron appendix, Ichthyomyzon fossor). Purple dots labeled P.marinus represent sequenced individuals, green dots labeled I. fossor represent known Northern brook lamprey, and blue dots labeled L. appendix represent known American brook lamprey.

## Gaussian Mixture Analyses

Offspring from sea lamprey and other fish species often exist in mixtures of individuals of different ages (cohorts), and these age classes need to be separated for estimation of $N_{b}$ and $N_{s}$. We developed a novel extension of Gaussian mixture methods by combining mixture models with reconstructed pedigrees (Figure 1.2). Given the semelparous life history of sea lamprey, full and half-sibling relationships should not span different cohorts; therefore, all individuals connected in the pedigree were assumed to be from the same age class. Aging methods like statolith aging have been found to be unreliable (Dawson et al., 2015), and length-based aging methods have been primarily used by management agencies for sea lamprey (Hardisty \& Potter, 1971; Sethi, Gerken, \& Ashline, 2017; Slade et al., 2003). Lengths of sea lamprey larvae were used in Gaussian mixture analyses to classify individuals into putative age-classes prior to estimation of effective breeding size $\left(N_{b}\right)$ and the minimum number of spawners $\left(N_{s}\right)$. Mixture analyses were conducted separately for each stream and each collection year due to variation in larval length between streams and collection years.

Mixture models were constructed using the R packages BayesMix (Grün \& Leisch, 2010) and bmixture (Mohammadi et al., 2013) to infer the number of age classes (K) and generate individual assignments to those cohorts. We used two different approaches to assess the number of cohorts represented by a sample of sea lamprey larvae. Birth-death MCMC treats K as a model parameter that is allowed to increase or decrease in successive steps of the MCMC chain to provide posterior probabilities for each potential K value (Mohammadi et al., 2013; Stephens, 2009). Rousseau and Mengersen (2011) proposed a cluster determining method that involves fitting a mixture model with a large K value and eliminating clusters with membership proportions below a certain cutoff (between 0.01 and 0.05 ; Nasserinejad, Rosmalen, De Kort, \&

Lesaffre, 2017). For this project, a cutoff of 0.035 and a K of 10 was used. The consensus from birth-death MCMC and the Rousseau and Mengerson (2011) approaches was used as the K value in a BayesMix model to determine individual assignments to clusters. If consensus was not reached, the output with a higher likelihood was used as the K value. All analyses were conducted in R (version 3.6.2). All scripts, data, and documentation for these analyses are available at https://github.com/weiseell/NbdLamprey.


Figure 1.2. A flow chart describing how inferred cohort assignments from the Gaussian mixture models are combined with information in the reconstructed pedigrees.

## Reconstructed Pedigrees

SNP genotype data were used to reconstruct pedigrees for larvae sampled from all locations. SNP loci were selected from the filtered group of SNPs for each population using the following criteria: minimum separation of adjacent SNP loci of 1MB to reduce the influences of physical linkage, variant position with the highest minor allele frequency ( $\mathrm{MAF}>0.05$ ), and highest percent of individuals genotyped with a minimum criteria of $80 \%$. If two or more SNPs met all three criteria equally, a random SNP was selected from that group. For each stream system, pedigree analysis was conducted in Colony version 2.0.6.6 (Jones \& Wang, 2010) using the full-likelihood approach. Due to differences in sample size among systems, a medium length run was used for the Lower Black Mallard and Ocqueoc Rivers, and a long run was used for the Pigeon and Upper Black Mallard Rivers. Other input parameters included unknown allele frequencies, polygamous mating, and no sibship scaling or prior sibship reported. All other parameters were kept at default settings.

Colony clusters from the reconstructed pedigree were compared to cohorts determined by the Gaussian mixture analysis to check for discrepancies between clusters of related individuals in the pedigree and cohorts assigned by the mixture analysis. A family cluster from Colony is defined as a group of offspring that are connected in the pedigree through parentage, but are not necessarily full- or half-siblings. For example, if offspring 1 and offspring 2 are half-siblings, and offspring 2 and offspring 3 are half-siblings, then offspring 1 and offspring 3 are considered to be in the same Colony cluster due to their connection in the pedigree through offspring 2. For each collection with multiple inferred cohorts from the Gaussian mixture analysis, individuals were evaluated for the level of family overlap between inferred cohorts. If there was no overlap of Colony cluster groups between inferred cohorts, they were left separate for subsequent
analysis. If individuals in the inferred cohorts were related (as full- or half-siblings), these individuals were combined into a single cohort for subsequent analyses. If there were multiple sample collections from the same location, the comparison was repeated to determine which cohorts should be combined across collections, and to approximate growth between collections to help separate year classes. Length histograms from previous studies (Dawson et al., 2020), as well as information on barrier installation and TFM treatment years, were used to estimate the cohort year classes. A flow chart of the decision-making process is shown in Figure 2. Additionally, the same decision-making tree was used with full-sibling groups and produced the same cohort groups as the Colony cluster groups.

## $\mathbf{N}_{\mathrm{b}}, \mathbf{N}_{\mathrm{s}}$, and $\widehat{\boldsymbol{N}}_{\mathrm{s}}$ estimates

Colony was used to estimate $N_{b}$ using the SF method (Wang and Santure, 2009), and the family information from Colony was used to estimate $N_{b}$ using the PwoP method (Waples and Waples, 2011). Additionally, mean $(\bar{k})$ and variance $\left(V_{k}\right)$ of adult reproductive success (number of offspring assigned based on the pedigree produced from the full-likelihood implementation in Colony) were calculated for the contributing individuals in the reconstructed parental populations. To generate confidence intervals for the PwoP method, a method based on Wang (2009) confidence intervals used for the SF method, which combines uncertainty in the pedigree reconstruction with uncertainty from sampling. The variance in $1 / 2 N_{b}$ was calculated for archived configurations from Colony to evaluate uncertainty in the pedigree, and the variance in $1 / 2 N_{b}$ for 1000 simulation populations with equal sex ratio with the same $N_{b}$ as the empirical data set. These two variances were summed and the square root of the summed variance was used to calculate $95 \%$ confidence intervals. $N_{s}$ was generated using the number of inferred
parents represented in each cohort. $N_{s}$ was extrapolated using a 'parentage accumulation curve,' which is akin to a species accumulation curve (Colwell, Chang, \& Chang, 2004; Israel \& May, 2010; Rawding et al., 2014), to count the number of unique parental genotypes as the number of offspring genotyped in the sample increases (Hunter, 2018; Sard et al. in press) Briefly, the specaccum function from the R package vegan (Oksanen et al., 2019) was used to generate pedigree accumulation curves and the total number of parental genotypes contributing to each cohort $\left(\widehat{N}_{\mathrm{s}}\right)$ was estimated using the Chao (Chao, 1987a) and jackknife (Heltshe \& Forrester, 2009) methods in the vegan function specpool (Oksanen et al., 2019).

The SNP panel used for estimates of $N_{b}$ from the LD method (LD) was selected with a separate set of criteria due to inherent differences in the estimation methods. SNPs were selected to only include loci in regions of the genome targeted by the Sard et al. (2020) Rapture panel. Within those RAD tags, SNPs with the highest percentage of individuals genotyped were selected and ties were broken with a random variable. NeEstimator (Do et al., 2013) was used for each cohort and stream sample with only the LD method selected, no comparisons within chromosomes were allowed (to avoid LD due to physical linkage of SNP markers; Waples et al. 2016). SNPs with a MAF $<0.05$ were removed to avoid potential upward bias from lowfrequency alleles (Robin S. Waples \& Do, 2010). Estimates were generated using an allele frequency inclusion criterion of $\mathrm{p}_{\text {crit }}=0.05$, and jackknife confidence intervals produced by NeEstimator were used (A. T. Jones, Ovenden, \& Wang, 2016). All analyses for $N_{b}, N_{s}$, and $\widehat{N}_{\mathrm{s}}$ estimates, with the exception of the Colony and NeEstimator programs, were conducted in R (version 3.6.2; R Core Team, 2019), all scripts and documentation for these analyses are available at https://github.com/weiseell/NbdLamprey.

## RESULTS

## Genotyping Analysis

Sequencing generated more than 3 billion total reads with an average of approximately 2 million reads for each individual (range: ~2000-12 million reads). After removal of PCR duplicates and quality filtering, reads were mapped to the sea lamprey reference genome (Smith et al., 2018). Of the filtered mapped reads, $88 \%$ were from sections of the genome targeted by the Rapture panel developed by Sard et al. (2020). Average sequencing depth in targeted regions was 34X. The SNPs targeted by the Rapture panel also had a higher proportion of loci with MAF $>0.05(0.25)$ when compared to non-targeted SNPs ( 0.177 ), and a higher mean proportion of individuals genotyped per SNP (on-target $=0.56$, off-target $=0.20$ ).

## Mixture Analyses and Reconstructed Pedigrees

In the Lower Black Mallard River, two age-classes were identified based on clusterdetermining methods for both collection years, shown in the histograms in Figure 1.3. The number of cohorts was determined by consensus for the 2018 collection, and for the 2017 collection the Rousseau and Mengersen (2011) criteria had a higher likelihood (Table 1.1). Length distributions among the inferred age-classes overlapped, with the exception of a small group in the Lower Black Mallard River 2017 collection, as shown by the boxplots in Figure 1.4. The reconstructed pedigree had 104 full-sibling families and 14 Colony clusters. Figure 1.5 visualizes the family structure across both collections compared to the inferred cohorts from the mixture analysis. The largest Colony cluster contained 755 (75\%) of the sampled offspring. The individuals in this cluster were present in both length-inferred age classes for the 2018 collection and the larger age class in the 2017 collection. The small age class in the 2017 collection likely
comprises offspring from spawning in 2016 (Figure 1.3), whereas the rest of the sampled individuals represent the 2015 cohort. No sibling relationships were inferred between the Lower and Upper Black Mallard River collections, indicating that larvae in these two areas were produced by different sets of spawning adults. Due to the small sample size from the Upper Black Mallard River population, the cluster determining models did not converge, and the mixture analysis was not conducted.

The mixture models for the Ocqueoc River indicated that one age-class of individuals had been collected (Table 1.1, Figure 1.3). The pedigree reconstruction contained 17 clusters and 18 full-sibling families. The pedigree reconstruction contained two half-sibling families that contributed $89 \%$ of sampled offspring (Figure 1.5). All of the individuals from those families were collapsed into the same Colony cluster (Figure 1.4).

Cluster likelihood (the likelihood that a Colony cluster cannot be split) was inconsistent for pedigrees derived from the Ocqueoc and the Lower Black Mallard Rivers. The cluster likelihood for the largest cluster in both systems was $<0.5$. Small clusters in each location had higher probability (Figure 1.4) suggesting some assignment uncertainty for a small group of individuals in each sample.

The reconstructed pedigree in the Pigeon River had six small full-sibling families that were mostly unrelated to each other. The sample size from the Pigeon River was too small to quantitatively compare inferred cohorts and the family structure from the reconstructed pedigree or run mixture models.

Table 1.1. Summary of results for identifying the optimal number of clusters $(\mathrm{K})$ in the mixture analysis for sea lamprey. Analyses were performed for each larval collection with a range of $\mathrm{K}=1-4$ clusters. $\mathrm{R} \& \mathrm{M}$ criteria and $\mathrm{BD}-\mathrm{MCMC}$ shows the estimated probability of each K value from the Rousseau and Mengersen (2011) criteria and Birth Death Markov Chain Monte Carlo (BD-MCMC; Mohammadi, Salehi-Rad, \& Wit, 2013), respectively. The optimal number of clusters from each method is bolded.

| Lower Black Mallard River - 2017 Collection (n = 386) |  |  |
| :---: | :---: | :---: |
| $\mathbf{K}$ | R\&M Criteria | BD-MCMC |
| 1 | 0.074 | 0.008 |
| 2 | $\mathbf{0 . 9 1 2}$ | 0.067 |
| 3 | 0.013 | 0.385 |
| 4 | 0.000 | $\mathbf{0 . 5 4 0}$ |
| Lower Black Mallard River - 2018 Collection (n = 614) |  |  |
| K | R\&M Criteria | BD-MCMC |
| 1 | 0.008 | 0.112 |
| 2 | $\mathbf{0 . 8 2 7}$ | $\mathbf{0 . 4 7 8}$ |
| 3 | 0.164 | 0.319 |
| 4 | 0.000 | 0.091 |
| Ocqueoc River-2018 Collection (n = 396) |  |  |
| $\mathbf{K}$ | R\&M Criteria | BD-MCMC |
| 1 | $\mathbf{0 . 9 9 8}$ | 0.143 |
| 2 | 0.002 | $\mathbf{0 . 5 3 8}$ |
| 3 | 0.000 | 0.277 |
| 4 | 0.000 | 0.042 |



Figure 1.3. Length frequency distributions for larval sea lamprey from all rivers and collection years, fill colors represent individual cluster assignment from the Gaussian mixture analysis. If mixture models were not completed due to small sample size, length histograms are included and shaded as a single cohort.


Figure 1.4. Boxplots of length distributions for each sea lamprey Colony cluster from the Lower Black Mallard River (A) and the Ocqueoc River (B). Colony clusters are defined as groups of offspring in the pedigree that are connected by parentage, but are not necessarily full- or halfsiblings. Plots are separated by collection. The probability that the Colony cluster cannot be split is represented by a continuous shading scale for both subplots (red clusters have a lower likelihood, white clusters have a higher likelihood).


Figure 1.5. Visualization of reconstructed sea lamprey pedigrees. The center represents genotyped individuals, and dots represent inferred parents. Lines connect each reconstructed parent to sequenced offspring in the pedigree. Black boxes represent cohorts inferred by the mixture method. Note: Since parents were not sequenced, and due to the lack of known sexdetermining genes for sea lamprey, the sex of reconstructed parents cannot be determined. Parent 1 and Parent 2 are used instead.

## $N_{b}$ and $N_{s}$ calculations

$N_{b}$ and $N_{s}$ estimates for all cohorts are summarized in Table 1.2, and $\widehat{N}_{\mathrm{s}}$ accumulation curves are shown in Figure 1.6. For the Lower Black Mallard River, the $N_{b}$ estimates for the 2015 cohort ranged from 24 to 32 (Table 1.2), and accumulated $N_{s}$ ranged from 108 to 110 (Table 1.2). The 2016 cohort had $N_{b}$ estimates that ranged from 5 to 29 (Table 1.2) and $N_{s}$ estimates that ranged from 19 to 24 (Table 1.2, Figure 1.6). For the Upper Black Mallard River collection, $N_{b}$ estimates ranged from 3 to 8 (Table 1.2) and $N_{s}$ estimates ranged from 15 to 16 (Table 1.2, Figure 1.6). In the Ocqueoc River, $N_{b}$ estimates ranged from 9 to 50 (Table 1.2) and $N_{s}$ estimates ranged from 90 to 98 (Table 1.2, Figure 1.6). Confidence intervals were small, partially due to the large numbers of loci used in the estimates. $N_{b}$ estimates for the Pigeon River collections ranged from 6 to 8 (Table 1.2), while Chao and jackknife estimates of $N_{s}$ ranged from 14 to 18 (Table 1.2, Figure 1.6).

Table 1.2. Estimates of the effective number of breeding adults $\left(N_{b}\right)$ and the number of unique inferred parental genotypes in the inferred pedigree $\left(N_{s}\right)$ for each stream and sea lamprey cohort. Locations are shown in Figure 1. N is the number of larval sea lamprey sampled for a stream and year. $V_{k}$ and $\bar{k}$ represent the inferred variance in reproductive success and mean number of offspring per adult in the population, respectively. LD refers to $N_{b}$ estimates derived from the linkage disequilibrium method. SF refers to $N_{b}$ estimates from the sibship frequency method. PwoP refers to $N_{b}$ estimates from the parentage-without-parents method $\widehat{N_{s}}-$ Chao and $\widehat{N_{s}}$ - Jackknife represent accumulated $N_{s}$ estimates using the Chao and the Jackknife methods, respectively.

| Location ${ }^{1}$ | Fullsibs | Clusters | Cohort | n | $\bar{k}$ | $V_{k}$ | LD | SF | PwoP | $N_{s}$ | $\widehat{N}_{s}-$ <br> Chao | $\widehat{N}_{s}$ Jackknife |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Lower Black <br> Mallard River (A) | 104 | 14 | 2015 | 1024 | 21.78 | 945.70 | $\begin{gathered} 24.5 \\ (22.7-26.5) \\ \hline \end{gathered}$ | $\begin{gathered} 32 \\ (20-50) \\ \hline \end{gathered}$ | $\begin{gathered} 31.88 \\ (31.68-32.14) \\ \hline \end{gathered}$ | 94 | $\begin{aligned} & \hline 110.06 \\ & \pm 11.02 \\ & \hline \end{aligned}$ | $\begin{gathered} 108.99 \\ \pm 3.87 \\ \hline \end{gathered}$ |
| Lower Black <br> Mallard River (A) |  |  | 2016 | 16 | 1.78 | 0.51 | $\begin{gathered} 4.7 \\ (2.2-13.4) \end{gathered}$ | $\begin{gathered} 13 \\ (7-30) \\ \hline \end{gathered}$ | $\begin{gathered} 29.18 \\ (16.99-103.25) \end{gathered}$ | 18 | $\begin{gathered} 19.53 \pm \\ 1.82 \\ \hline \end{gathered}$ | $\begin{gathered} 23.625 \pm \\ 2.30 \\ \hline \end{gathered}$ |
| Upper Black Mallard River (A) | 9 | 4 |  | 35 | 5.23 | 24.02 | $\begin{gathered} 3.1 \\ (2.4-5.9) \\ \hline \end{gathered}$ | $\begin{gathered} 7 \\ (4-21) \\ \hline \end{gathered}$ | $\begin{gathered} 7.59 \\ (6.95-8.35) \\ \hline \end{gathered}$ | 13 | $\begin{gathered} 15.18 \pm \\ 3.30 \\ \hline \end{gathered}$ | $\begin{gathered} 15.91 \pm \\ 1.68 \\ \hline \end{gathered}$ |
| Ocqueoc River <br> (B) | 87 | 17 |  | 396 | 10.24 | 799.50 | $\begin{gathered} 50.2 \\ (45.6-55.2) \end{gathered}$ | $\begin{gathered} 9 \\ (5-24) \\ \hline \end{gathered}$ | $\begin{gathered} 8.90 \\ (8.81-8.94) \end{gathered}$ | 76 | $\begin{gathered} 90.66 \pm \\ 8.00 \end{gathered}$ | $\begin{gathered} 98.94 \pm \\ 5.55 \end{gathered}$ |
| Pigeon River (C) | 6 | 3 |  | 16 | 4.22 | 13.51 | $\begin{gathered} 7.6 \\ (2.8-21.7) \\ \hline \end{gathered}$ | $\begin{gathered} 10 \\ (5-28) \\ \hline \end{gathered}$ | $\begin{gathered} 5.76 \\ (5.04-6.71) \\ \hline \end{gathered}$ | 9 | $\begin{gathered} 13.26 \pm \\ 6.83 \\ \hline \end{gathered}$ | $\begin{gathered} 11.84 \pm \\ 2.14 \\ \hline \end{gathered}$ |



Figure 1.6. The estimated number of unique parental genotypes in the pedigree ( $\widehat{N}_{s}$ ) characterized using pedigree accumulation curves for all three stream systems. For all locations, boxplot distributions for each step size overlay a line plot with a grey background for $+/$ - one standard error, and labeled horizontal lines represent $\widehat{N}_{s}$ estimates from the jackknife and chao methods. Due to the large number of individuals, the Ocqueoc River boxplots are plotted in step sizes of 5 sampled individuals and the Lower Black Mallard River boxplots are shown for sample sizes increasing by 10 individuals. The boxplots for all other locations are plotted for a step size of 1 sampled individuals.

## DISCUSSION

## $N_{b}$ and $N_{s}$ estimates

Genetic estimates of $N_{b}$ and $\widehat{N}_{\text {s }}$ allow inferences pertaining to the number of successful spawning adults in a Great Lakes tributary. $N_{b}$ and $\widehat{N}_{\mathrm{s}}$ estimates both provide information about spawning populations that can be used to make inferences in management and conservation contexts. In the sampled systems, $N_{b}$ estimates and reconstructed pedigrees indicated skewed sex ratios in the Ocqueoc River. The $\widehat{N}_{s}$ estimates provided insight into the number of successfully spawning adults upstream of control barriers. Despite the small to moderate effective breeding sizes estimated for each sampled cohort, larvae were abundant in all systems (estimates range from approximately 3500 larvae in the Upper Black Mallard River in 2017 to 124,000 larvae in the Pigeon River in 2019; unpublished data, A. Jubar, USFWS). In all systems, the vast majority of individuals had half- and full-siblings within the areas sampled. In the Ocqueoc, $89 \%$ of individuals were in two half-sibling families. In the Black Mallard River, $75 \%$ of individuals were in a single Colony cluster, and $97.3 \%$ of the individuals were determined to be in a single cohort from 2015, prior to the barrier construction.

Increasing sample size and the number of loci analyzed improves $N_{b}$ estimates based on all three methods (England, Cornuet, Berthier, Tallmon, \& Luikart, 2006; Wang, 2016; Waples, 2016). Based on simulations conducted by Sard et al. (2020), a high degree of accuracy in the pedigree assignments from Colony is expected given the expected spawning adult population size for these systems and the number of SNP loci used for the analysis. The large number of SNP loci used for pedigree reconstruction and $N_{b}$ estimation resulted in high confidence in inferred relationships and confidence intervals that were substantially smaller than those for typical microsatellite datasets (Flanagan \& Jones, 2019; J. D. Robinson \& Moyer, 2013). For the

LD estimates, confidence intervals can be artificially narrowed by large numbers of loci, although the corrected jackknife confidence interval approach reduces this effect (Waples, Grewe, Bravington, Hillary, \& Feutry, 2018). Additionally, the high cluster probabilities for large Colony clusters in the Black Mallard and Ocqueoc Rivers bolster confidence in the family relationships identified by Colony. However, individual misassignment could stem from several potential sources. Pedigree reconstructions for the Black Mallard and Ocqueoc Rivers also contain a small group of individuals that were unrelated to any large family groups. These outlier groups are most likely unrelated individuals, but they could be the result of Colony assignment error (Butler et al., 2004). Outlier groups were confirmed to be sea lamprey from a PCA, rather than misidentified native lamprey (Lethenteron appendix, Ichthyomyzon fossor).

Our results provide an empirical application of $\widehat{N}_{\mathrm{s}}$, a comparatively new method of quantifying spawning adults. Previous work has used accumulation curves to evaluate spawner abundance in green sturgeon (Israel \& May, 2010) and chinook salmon (Rawding et al., 2014). $\widehat{N}_{\text {s }}$ has been used for lake sturgeon (Acipenser fulvescens) previously to estimate the number of adults recruited to a spawning site (Hunter, 2018; Sard et al., in press). Given sufficient sample sizes, this method can be used to estimate the number of adults contributing to a cohort (Figure 6). $N_{s}$ estimates without an accumulation method have direct dependence on sample size since they are calculated as the number of unique reconstructed parental genotypes for a set of offspring and are thus limited by sample size. By applying methods designed to estimate total species richness to reconstructed pedigrees, that dependence is reduced.

## Cohort identification

Mixture analysis in sea lamprey has several sources of uncertainty. Techniques rely on the presence of several large cohorts in a stream sample to provide accurate cohort assignments and are expected to be most effective for age-0 and age-1 individuals where length distributions are more distinct from older cohorts (Dawson et al., 2009). Additionally, environmental conditions affect the growth rate of larvae. Variables such as growing degree days, stream temperature and larval sea lamprey density are all significant predictors of larval growth in streams (Dawson et al., 2020).
$N_{b}$ and $N_{s}$ are both estimates generated for a single spawning year, meaning that the ability to separate offspring into cohorts is vital for accurate estimates. Combining Gaussian mixture models with reconstructed pedigree data allows for the identification of potentially misidentified cohorts from the length data alone, minimizing error in cohort identification. Including individuals from multiple cohorts in $N_{b}$ and $N_{s}$ calculations generated from the reconstructed pedigree would upwardly bias estimates due to the inclusion of parents from multiple spawning events (Wang, Santiago, \& Caballero, 2016). For the linkage disequilibrium estimates, linkage that arise from two separate spawning groups are included, leading to an downward bias (Waples \& England, 2011).

Uncertainty in the cohort assignments from the mixture analysis was evident in the Black Mallard River samples. Larvae were separated into multiple cohorts with overlap between length distributions for individuals assigned to older cohorts. Additionally, variability in growth within age classes was greater than previously assumed (Figure 1.4), potentially contributing to the over-splitting of larval cohorts observed in both streams. Incorporating family pedigree information further supported the conclusion that number of cohorts was overestimated by the
mixture analysis, as several sibling groups spanned multiple inferred cohorts. In the Black Mallard River, length-based mixture analysis divided members of the largest family cluster into three cohorts, again indicating over-splitting. In semelparous species like the sea lamprey, family structure present in reconstructed pedigrees can be combined with length data as complementary information to verify cohort assignments.

## Importance of Sample Size in $N_{b}$ and $N_{s}$ estimates

Estimates of $N_{b}$ and $N_{s}$ can be sensitive to small sample size. For estimates generated using reconstructed pedigrees, small sample size can lead to missing family groups. If estimated levels of LD among a set of polymorphic loci can be explained by sample size alone, no signal remains to estimate effective size of the sampled population (Waples \& Do, 2010). Increasing sample size and the number of markers thus increases the power for estimation of effective population size (Waples, 2016). Non-representative sampling of a population can lead to downward bias in $N_{b}$ and $N_{s}$ estimates due to missing diversity in sampled individuals (Whiteley et al., 2012). Minimizing costs for a project while still obtaining accurate point estimates is partially balanced by selecting appropriate sample sizes for a given study system. If the $N_{b}$ of the stream is expected to be small (like the Black Mallard River), then sample size could be proportionally smaller to minimize field and sequencing costs necessary for the project. However, if $N_{b}$ is unknown or expected to be large, a larger sample size is necessary to ensure accurate estimates (Wang, 2016; Waples, Grewe, Bravington, Hillary, \& Feutry, 2018).

## Application of Results

Population estimates across all three streams sampled in this study imply that barriers
limited adult spawning numbers but were not completely effective at blocking access to spawning habitats. Thus, the large numbers of larvae present in sampled systems were a poor indicator of spawning adult abundance, which is an important finding for managers. Another important finding was that members of full- and half-sibling families were identified in multiple year cohorts, which is impossible due to the species' semelparous life history. Cohort assignments identified by mixture models (i.e., in the absence of confirmatory genetic data) showed that length-based analysis alone does not provide accurate cohort assignments. Our analyses illustrate the potential to improve cohort assignments by incorporating population genomic data and pedigree analysis for sampled sea lamprey larvae. Collectively, effective size, minimum spawning size estimates, and reconstructed pedigrees based on larval sequencing were successfully used to make inferences about spawning adult populations in three streams. Recent developments, including the availability of a reference genome (Smith et al., 2013, 2018) and the RAD-Capture marker panel (Sard et al., 2020) employed in this study, position Great Lakes sea lamprey as an emerging model system for the study of species invasions.

Population genomic data were used to infer aspects of sea lamprey biology that contribute valuable information for sea lamprey assessment. Results from the Lower Black Mallard River indicated that the majority of individuals originated from a single cohort due to the existence of full-sibling relationships between inferred cohorts from the mixture analysis. These data are consistent with the expectation that a moderate number of adult sea lamprey spawned in the Black Mallard River in 2015 after lampricide treatment, but prior to the electric barrier installation in 2016. Collectively, our data suggest that the electric barrier in the Black Mallard River was effective at reducing sea lamprey migration upstream, as $N_{b}$ of the 2016 cohort was much smaller than $N_{b}$ of the 2015 cohort, and a 2017 cohort was not confidently
identified by our mixture analyses for the Lower Black Mallard River collections. There are alternative explanations for small $N_{b}$, such as high variance in reproductive success and strongly skewed sex ratios, as seen in the Ocqueoc River estimates. Additionally, the lack of family relationships between the Upper and Lower Black Mallard River implies two separate subsets of spawning adults. In the Ocqueoc River, $89 \%$ of larvae were from two half-sibling families, indicating that a small group of fertile males were present above the barrier along with the females released for research experiments. Estimates from samples collected in the Pigeon River indicated that both $N_{b}$ and $N_{s}$ were small, which is consistent with the expectation that releases of sterile males decreased the number of successful spawning adults in the system.

Although sea lamprey are invasive in the Great lakes, they are endangered in parts of Europe, and conservation efforts are underway to protect declining populations (Hansen et al., 2016). Many of the same questions related to management of invasive Great Lakes populations also apply to threatened marine sea lamprey populations spawning in North American and European tributaries of the Atlantic Ocean. Estimates of $N_{b}$ and the per-generation effective population size ( $N_{e}$ ) can provide important information on patterns of relatedness, the rate of diversity loss due to genetic drift and inbreeding, and the species' potential for adaptation.

Population genomic data, including estimates of effective size, have been used as a monitoring tool in many conservation and management situations for other species, such as translocations and reintroductions (Hess et al., 2015; Roques, Berrebi, Rochard, \& Acolas, 2018; Whitlock, Schultz, Schreck, \& Hess, 2017), quantifying genetic diversity to prevent extinctions (Faulks, Kerezsy, Unmack, Johnson, \& Hughes, 2017), and identifying ecologically significant units (Blower, Pandolfi, Bruce, Gomez-Cabrera, \& Ovenden, 2012). Parentage has been used to evaluate the size of invading populations in species like the Asian Swamp Eel (Monopterus
albus) (Taylor et al., 2021). Genetic data were used in all of the above situations to evaluate the population or assess the success of a management action, and this type of assessment is increasingly needed among managed populations (Hoban et al., 2021). Thus, population genomic data and estimation of effective population sizes could be used to assess the efficacy and level of success of management actions related to invasive species, endangered populations, species of conservation concern, and managed species (Nunziata \& Weisrock, 2018).

LITERATURE CITED

## LITERATURE CITED

Ackerman, M. W., Hand, B. K., Waples, R. K., Luikart, G., Waples, R. S., Steele, C. A., ... Campbell, M. R. (2017). Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. Evolutionary Applications, 10(2), 146160. https://doi.org/10.1111/eva. 12433

Ali, O. A., O’Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., \& Miller, M. R. (2016). Rad capture (Rapture): Flexible and efficient sequence-based genotyping. Genetics, 202(2), 389-400. https://doi.org/10.1534/genetics.115.183665

Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data.
Antao, T., Pérez-Figueroa, A., \& Luikart, G. (2011). Early detection of population declines: High power of genetic monitoring using effective population size estimators. Evolutionary Applications, 4(1), 144-154. https://doi.org/10.1111/j.1752-4571.2010.00150.x

Applegate, V. C. (1950). Natural history of the sea lamprey (Petromyzon marinus) in Michigan. Spec Sci Rep US Fish Wildl Serv, 55, 1-237. Retrieved from http://ci.nii.ac.jp/naid/10010684036/en/

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE, 3(10), e3376. https://doi.org/10.1371/journal.pone. 0003376

Bajer, P. G., Ghosal, R., Maselko, M., Smanski, M. J., Lechelt, J. D., Hansen, G., \& Kornis, M. S. (2019). Biological control of invasive fish and aquatic invertebrates: A brief review with case studies. Management of Biological Invasions, 10(2), 227-254. https://doi.org/10.3391/mbi.2019.10.2.02

Bergstedt, R. A., \& Seelye, J. G. (1995). Evidence for Lack of Homing by Sea Lampreys. Transactions of the American Fisheries Society, 124(2), 235-239. https://doi.org/10.1577/1548-8659(1995)124<0235:eflohb $>2.3 . c o ; 2$

Blakeslee, A. M. H., Kamakukra, Y., Onufrey, J., Makino, W., Urabe, J., Park, S., ... Miura, O. (2017). Reconstructing the Invasion History of the Asian shorecrab, Hemigrapsus sanguineus (De Haan 1835) in the Western Atlantic. Marine Biology, 164(3), 1-19. https://doi.org/10.1007/s00227-017-3069-1

Blower, D. C., Pandolfi, J. M., Bruce, B. D., Gomez-Cabrera, M. D. C., \& Ovenden, J. R. (2012). Population genetics of Australian white sharks reveals fine-scale spatial structure, transoceanic dispersal events and low effective population sizes. Marine Ecology Progress Series, 455, 229-244. https://doi.org/10.3354/meps09659

Bolger, A. M., Lohse, M., \& Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.
https://doi.org/10.1093/bioinformatics/btu170
Bravington, M. V., Skaug, H. J., \& Anderson, E. C. (2016). Close-Kin Mark-Recapture. Statistical Science, 31(2), 259-274. https://doi.org/10.1214/16-sts552

Bryan, M. B., Zalinski, D., Filcek, K. B., Libants, S., Li, W., \& Scribner, K. T. (2005). Patterns of invasion and colonization of the sea lamprey (Petromyzon marinus) in North America as revealed by microsatellite genotypes. Molecular Ecology, 14(12), 3757-3773. https://doi.org/10.1111/j.1365-294X.2005.02716.x

Buchinger, T. J., Scott, A. M., Fissette, S. D., Brant, C. O., Huertas, M., Li, K., ... Li, W. (2020). A pheromone antagonist liberates female sea lamprey from a sensory trap to enable reliable communication. Proceedings of the National Academy of Sciences of the United States of America, 117(13), 7284-7289. https://doi.org/10.1073/pnas. 1921394117

Catchen, J. M., Hohenlohe, P. A., Bassham, S., Amores, A., \& Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology, 22(11), 3124-3140. https://doi.org/10.1111/mec. 12354

Chao, A. (1987). Estimating the Population Size for Capture-Recapture Data with Unequal Catchability. Biometrics, 43(4), 783-791. https://doi.org/10.4081/cp.2017.979

Colwell, R. K., Chang, X. M., \& Chang, J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. Ecology, 85(10), 2717-2727. https://doi.org/10.1890/03-0557

Dawson, H. A., Higgins-weier, C. E., Steeves, T. B., \& Johnson, N. S. (2020). Estimating age and growth of invasive sea lamprey : A review of approaches and investigation of a new method. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2020.06.002

Dawson, H. A., \& Jones, M. L. (2009). Factors affecting recruitment dynamics of Great Lakes sea lamprey (Petromyzon marinus) populations. Journal of Great Lakes Research, 35(3), 353-360. https://doi.org/10.1016/j.jglr.2009.03.003

Dawson, H. A., Jones, M. L., Scribner, K. T., \& Gilmore, S. A. (2009). An Assessment of Age Determination Methods for Great Lakes Larval Sea Lampreys. North American Journal of Fisheries Management, 29(4), 914-927. https://doi.org/10.1577/m08-139.1

Dawson, H. A., Quintella, B. R., Almeida, P. R., Treble, A. J., \& Jolley, J. C. (2015). The Ecology of Larval and Metamorphosing Lampreys. In Lampreys: Biology, Conservation and Control (Vol. 1, pp. 75-137). https://doi.org/10.1007/978-94-017-9306-3

De Barba, M., Waits, L. P., Garton, E. O., Genovesi, P., Randi, E., Mustoni, A., \& Groff, C. (2010). The power of genetic monitoring for studying demography, ecology and genetics
of a reintroduced brown bear population. Molecular Ecology, 19(18), 3938-3951. https://doi.org/10.1111/j.1365-294X.2010.04791.x

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., \& Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog Lithobates catesbeianus. Journal of Applied Ecology, 49(4), 953-959. https://doi.org/10.1111/j.1365-2664.2012.02171.x

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, J. B., \& Ovenden, J. R. (2013). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size $\left(\mathrm{N}_{\mathrm{e}}\right)$ from genetic data. Molecular Ecology Resources, 14(1), 209-214. https://doi.org/doi:10.1111/1755-0998.12157

England, P. R., Cornuet, J. M., Berthier, P., Tallmon, D. A., \& Luikart, G. (2006). Estimating effective population size from linkage disequilibrium: Severe bias in small samples. Conservation Genetics, 7(2), 303-308. https://doi.org/10.1007/s10592-005-9103-8

Faulks, L. K., Kerezsy, A., Unmack, P. J., Johnson, J. B., \& Hughes, J. M. (2017). Going, going, gone? Loss of genetic diversity in two critically endangered Australian freshwater fishes, Scaturiginichthys vermeilipinnis and Chlamydogobius squamigenus, from Great Artesian Basin springs at Edgbaston, Queensland, Australia. Aquatic Conservation: Marine and Freshwater Ecosystems, 27(1), 39-50. https://doi.org/10.1002/aqc. 2684

Fissette, S. D., Buchinger, T. J., Wagner, C. M., Johnson, N. S., Scott, A. M., \& Li, W. (2021). Progress towards integrating an understanding of chemical ecology into sea lamprey control. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2021.02.008

Flanagan, S. P., \& Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. Molecular Ecology, 28(3), 544-567.
https://doi.org/10.1111/mec. 14988
Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation, 143(9), 1919-1927. https://doi.org/10.1016/J.BIOCON.2010.05.011

Griffiths, R. W., Beamish, F. W. H., Morrison, B. J., \& Barker, L. A. (2001). Factors Affecting Larval Sea Lamprey Growth and Length at Metamorphosis in Lampricide-Treated Streams. Transactions of the American Fisheries Society, 130(2), 289-306. https://doi.org/10.1577/1548-8659(2001)130<0289:falslg>2.0.co;2

Grün, B., \& Leisch, F. (2010). BayesMix: an R package for Bayesian mixture modeling. Technique Report, 1-11.

Hansen, M. J., Madenjian, C. P., Slade, J. W., Steeves, T. B., Almeida, P. R., \& Quintella, B. R. (2016). Population ecology of the sea lamprey (Petromyzon marinus) as an invasive
species in the Laurentian Great Lakes and an imperiled species in Europe. Reviews in Fish Biology and Fisheries, 26(3), 509-535. https://doi.org/10.1007/s11160-016-9440-3

Hardisty, M. W., \& Potter, I. C. (1971). The Biology Of Lampreys. In Academic Press (Vol. 1). New York. https://doi.org/10.1126/science.176.4042.1409

Heinrich, J. W., Mullett, K. M., Hansen, M. J., Adams, J. V, Klar, G. T., Johnson, D. A., ... Young, R. J. (2003). Sea Lamprey Abundance and Management in Lake Superior, 1957 to 1999. Journal of Great Lakes Research, 29, 566-583. Retrieved from https://www.sciencedirect.com/science/article/pii/S0380133003705176

Heltshe, J. F., \& Forrester, N. E. (2009). Estimating Species Richness Using the Jackknife Procedure Published by : International Biometric Society Stable URL: http://www.jstor.org/stable/2530802. Society, 39(1), 1-11.

Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., ... Narum, S. R. (2015). Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. Molecular Ecology Resources, 15(1), 187-202. https://doi.org/10.1111/1755-0998.12283

Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. Genetical Research, 38(03), 209. https://doi.org/10.1017/S0016672300020553

Hoban, S., Bruford, M. W., Funk, W. C., Galbusera, P., Griffith, M. P., Grueber, C. E., ... Vernesi, C. (2021). Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible. BioScience, XX(X), 1-13. https://doi.org/10.1093/biosci/biab054

Hunter, R. D. (2018). Assessing reproductive success of Lake Sturgeon (Acipenser fulvescens) associated with natural and constructed spawning reefs in a large river system using pedigree analysis.

Hunter, R. D., Roseman, E. F., Sard, N. M., DeBruyne, R. L., Wang, J., \& Scribner, K. T. (2020). Genetic Family Reconstruction Characterizes Lake Sturgeon Use of Newly Constructed Spawning Habitat and Larval Dispersal. Transactions of the American Fisheries Society, 266-283. https://doi.org/10.1002/tafs. 10225

Israel, J. A., \& May, B. (2010). Indirect genetic estimates of breeding population size in the polyploid green sturgeon (Acipenser medirostris). Molecular Ecology, 19(5), 1058-1070. https://doi.org/10.1111/j.1365-294X.2010.04533.x

Jensen, A. J., \& Jones, M. L. (2018). Forecasting the response of Great Lakes sea lamprey (Petromyzon marinus) to barrier removals. Canadian Journal of Fisheries and Aquatic Science, 75(9), 1415-1426. https://doi.org/10.1139/cjfas-2017-0243

Johnson, N. S., Buchinger, T. J., \& Li, W. (2015). Reproductive Ecology of Lampreys BT Lampreys: Biology, Conservation and Control: Volume 1. In M. F. Docker (Ed.) (pp. 265-303). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9306-3-6

Johnson, N. S., Jubar, A. K., Keffer, D. A., Hrodey, P. J., Bravener, G. A., Freitas, L. E., ... Siefkes, M. J. (2020). A case study of sea lamprey (Petromyzon marinus) control and ecology in a microcosm of the Great Lakes. Journal of Great Lakes Research https://doi.org/10.1016/j.jglr.2020.09.006

Johnson, N. S., Thompson, H. T., Holbrook, C., \& Tix, J. A. (2014). Blocking and guiding adult sea lamprey with pulsed direct current from vertical electrodes. Fisheries Research, 150, 38-48. https://doi.org/10.1016/j.fishres.2013.10.006

Jones, A. T., Ovenden, J. R., \& Wang, Y. G. (2016). Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. Heredity, 117(4), 217-223. https://doi.org/10.1038/hdy.2016.19

Jones, M. L. (2007). Toward Improved Assessment of Sea Lamprey Population Dynamics in Support of Cost-effective Sea Lamprey Management. Journal of Great Lakes Research, 33(2), 35-47. https://doi.org/10.3394/0380-1330(2007)33

Jones, M. L., Irwin, B. J., Hansen, G. J. A., Dawson, H. A., Treble, A. J., Liu, W., ... Bence, J. R. (2009). An Operating Model for the Integrated Pest Management of Great Lakes Sea Lampreys. The Open Fish Science Journal, 2(1), 59-73. https://doi.org/10.2174/1874401x00902010059

Jones, O. R., \& Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources, 10(3), 551-555. https://doi.org/doi:10.1111/j.1755-0998.2009.02787.x

Katopodis, C., Bergstedt, R. A., O’Connor, L. M., McLaughlin, R. L., Pratt, T. C., Hayes, D. B., \& Hallett, A. G. (2009). Balancing Aquatic Habitat Fragmentation and Control of Invasive Species: Enhancing Selective Fish Passage at Sea Lamprey Control Barriers. Transactions of the American Fisheries Society, 138(3), 652-665. https://doi.org/10.1577/t08-118.1

Kaye, C. A., Heinrich, J. W., Hanson, L. H., McDonald, R. B., Slade, J. W., Genovese, J. H., \& Swink, W. D. (2003). Evaluation of Strategies for the Release of Male Sea Lampreys (Petromyzon marinus) in Lake Superior for a Proposed Sterile-Male-Release Program. Journal of Great Lakes Research, 29, 424-434.
https://doi.org/https://doi.org/10.1016/S0380-1330(03)70505-X
Keogh, J. S., Webb, J. K., \& Shine, R. (2007). Spatial genetic analysis and long-term markrecapture data demonstrate male-biased dispersal in a snake. Biology Letters, 3(1), 3335. https://doi.org/10.1098/rsbl.2006.0570

Koonce, J. F., Eshenroder, R. L., \& Christie, G. C. (1993). An Economic Injury Level Approach to Establishing the Intensity of Sea Lamprey Control in the Great Lakes. North American Journal of Fisheries Management, 13(1), 1-14. https://doi.org/10.1577/15488675(1993)013<0001:AEILAT>2.3.CO;2

Lavis, D. S., Hallett, A., Koon, E. M., \& McAuley, T. C. (2003). History of and advances in barriers as an alternative method to suppress sea lampreys in the Great Lakes. Journal of Great Lakes Research, 29(SUPPL. 1), 362-372. https://doi.org/10.1016/S0380-1330(03)70500-0

Lawrie, A. H. (1970). The Sea Lamprey in the Great Lakes. Transactions of the American Fisheries Society, 99(4), 766-775. https://doi.org/10.1577/1548$\underline{8659(1970) 99<766: T S L I T G>2.0 . C O ; 2}$

Leblanc, F., Belliveau, V., Watson, E., Coomber, C., Simard, N., Di Bacco, C., ... Gagné, N. (2020). Environmental DNA (eDNA) detection of marine aquatic invasive species (AIS) in Eastern Canada using a targeted species-specific qPCR approach. Management of Biological Invasions, 11(2), 201-217. https://doi.org/10.3391/mbi.2020.11.2.03

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 00(00), 1-3. Retrieved from http://arxiv.org/abs/1303.3997

Li, H., \& Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics, 26(5), 589-595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Manion, P. J., \& Smith, B. R. (1978). Biology of larval and metamorphosing sea lampreys, Petromyzon marinus, of the 1960 year class in the Big Garlic River, Michigan, Part II, 1966-1972. Great Lakes Fishery Commission Technical Report, 30, 1-37.

McDonald, D. G., \& Kolar, C. S. (2007). Research to Guide the Use of Lampricides for Controlling Sea Lamprey. Journal of Great Lakes Research, 33, 20-34. https://doi.org/https://doi.org/10.3394/0380-1330(2007)33[20:RTGTUO]2.0.CO;2

McGeoch, M. A., Butchart, S. H. M., Spear, D., Marais, E., Kleynhans, E. J., Symes, A., ... Hoffmann, M. (2010). Global indicators of biological invasion: Species numbers, biodiversity impact and policy responses. Diversity and Distributions, 16(1), 95-108. https://doi.org/10.1111/j.1472-4642.2009.00633.x

McKinney, G. J., Waples, R. K., Seeb, L. W., \& Seeb, J. E. (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(4), 656-669. https://doi.org/10.1111/1755-0998.12613

McLaughlin, R. L., Hallett, A., Pratt, T. C., O’Connor, L. M., \& McDonald, D. G. (2007). Research to Guide Use of Barriers, Traps, and Fishways to Control Sea Lamprey. Journal of Great Lakes Research, 33(2), 7-19. https://doi.org/10.3394/0380-1330(2007)33

Mohammadi, A., Salehi-Rad, M. R., \& Wit, E. C. (2013). Using mixture of Gamma distributions for Bayesian analysis in an M/G/1 queue with optional second service. Computational Statistics, 28(2), 683-700. https://doi.org/10.1007/s00180-012-0323-3

Morkert, S. B., Swink, W. D., \& Seelye, J. G. (1998). Evidence for Early Metamorphosis of Sea Lampreys in the Chippewa River, Michigan. North American Journal of Fisheries Management, 18(4), 966-971. https://doi.org/10.1577/1548$\underline{8675(1998) 018<0966: \text { efemos }>2.0 . c o ; 2}$

Nasserinejad, K., Rosmalen, J. Van, De Kort, W., \& Lesaffre, E. (2017). Comparison of criteria for choosing the number of classes in Bayesian finite mixture models. PLoS ONE, 12(1), 1-23. https://doi.org/10.1371/journal.pone. 0168838

Nunziata, S. O., \& Weisrock, D. W. (2018). Estimation of contemporary effective population size and population declines using RAD sequence data. Heredity, 120(3), 196-207. https://doi.org/10.1038/s41437-017-0037-y

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., ... Szoecs, E. (2019). Package ' vegan .'

Ovenden, J. R., Leigh, G. M., Blower, D. C., Jones, A. T., Moore, A., Bustamante, C., ... Dudgeon, C. L. (2016). Can estimates of genetic effective population size contribute to fisheries stock assessments? Journal of Fish Biology, 89(6), 2505-2518. https://doi.org/10.1111/jfb. 13129

Prior, K. M., Adams, D. C., Klepzig, K. D., \& Hulcr, J. (2018). When does invasive species removal lead to ecological recovery? Implications for management success. Biological Invasions, 20(2), 267-283. https://doi.org/10.1007/s10530-017-1542-x

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R foundation for Statistical Computing.

Rawding, D. J., Sharpe, C. S., \& Blankenship, S. M. (2014). Genetic-Based Estimates of Adult Chinook Salmon Spawner Abundance from Carcass Surveys and Juvenile Out-Migrant Traps. Transactions of the American Fisheries Society, 143(1), 55-67. https://doi.org/10.1080/00028487.2013.829122

Robinson, J. D., \& Moyer, G. R. (2013). Linkage disequilibrium and effective population size when generations overlap. Evolutionary Applications, 6(2), 290-302.
https://doi.org/10.1111/j.1752-4571.2012.00289.x

Roques, S., Berrebi, P., Rochard, E., \& Acolas, M. L. (2018). Genetic monitoring for the successful re-stocking of a critically endangered diadromous fish with low diversity. Biological Conservation, 221(March), 91-102.
https://doi.org/10.1016/j.biocon.2018.02.032
Rousseau, J., \& Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 73(5), 689-710. https://doi.org/10.1111/j.1467-9868.2011.00781.x

Sard, N. H., Hunter, R., Roseman, E. F., Hayes, D. B., DeBruyne, R. L., \& Scribner, K. T. (n.d.). Pedigree Rarification: Combining Methods from Community Ecology and Population Genetics to Estimate Spawning Adult Numbers. Methods in Ecology and Evolution.

Sard, N. M., Hunter, R. D., Roseman, E. F., Hayes, D. B., DeBruyne, R. L., \& Scribner, K. T. (n.d.). Extending non-parametric species richness estimators to genetic pedigree rarefaction for breeding adult estimation. In press.
Sard, N. M., Smith, S. R., Homola, J. J., Kanefsky, J., Bravener, G., Adams, J. V., ... Scribner, K. T. (2020). RAPTURE (RAD capture) panel facilitates analyses characterizing sea lamprey reproductive ecology and movement dynamics. Ecology and Evolution, (December 2019), 1-20. https://doi.org/10.1002/ece3.6001

Sard, N.M., Robinson, J., Kanefsky, J., Herbst, S., \& Scribner, K. (2019). Coalescent models characterize sources and demographic history of recent round goby colonization of Great Lakes and inland waters. Evolutionary Applications, 12(5), 1034-1049. https://doi.org/10.1111/eva. 12779

Sethi, S. A., Gerken, J., \& Ashline, J. (2017). Accurate aging of juvenile salmonids using fork lengths. Fisheries Research, 185, 161-168. https://doi.org/10.1016/j.fishres.2016.09.012

Sharov, A. A., \& Liebhold, A. M. (1998). Bioeconomics of Managing the Spread of. Ecological Applications, 8(3), 833-845.

Sherpa, S., Blum, M. G. B., Capblancq, T., Cumer, T., Rioux, D., \& Després, L. (2019). Unravelling the invasion history of the Asian tiger mosquito in Europe. Molecular Ecology, 28(9), 2360-2377. https://doi.org/10.1111/mec. 15071

Slade, J. W., Adams, J. V., Cuddy, D. W., Neave, F. B., Sullivan, W. P., Young, R. J., ... Jones, M. L. (2003). Relative contributions of sampling effort, measuring, and weighing to precision of larval sea lamprey biomass estimates. Journal of Great Lakes Research, 29(SUPPL. 1), 130-136. https://doi.org/10.1016/S0380-1330(03)70482-1

Smith, B. R., \& Tibbles, J. J. (1980). Sea Lamprey (Petromyzon marinus) in Lakes Huron, Michigan, and Superior: History of Invasion and Control, 1936-78. Canadian Journal of Fisheries and Aquatic Sciences, 37(37), 1780-1801. Retrieved from www.nrcresearchpress.com

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., ... Li, W. (2013). Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nature Genetics, 45(4), 415-421. https://doi.org/10.1038/ng. 2568

Smith, J. J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M. C., Parker, H. J., ... Amemiya, C. T. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. Nature Genetics, 50(2), 270-277. https://doi.org/10.1038/s41588-017-0036-1

Stephens, M. (2009). Bayesian Analysis of Mixture Models with an Unknown Number of Components- An Alternative to Reversible Jump Methods. Annals of Statistics, 28(1), 40-74.

Taylor, A. T., Bangs, M. R., \& Long, J. M. (2021). Sibship reconstruction with SNPs illuminates the scope of a cryptic invasion of Asian Swamp Eels (Monopterus albus) in Georgia, USA. Biological Invasions, 23(2), 569-580. https://doi.org/10.1007/s10530-020-02384-5

Treble, A. J., Jones, M. L., \& Steeves, T. B. (2008). Development and evaluation of a new predictive model for metamorphosis of Great Lakes larval sea lamprey (Petromyzon marinus) populations. Journal of Great Lakes Research, 34(3), 404-417. https://doi.org/10.3394/0380-1330(2008)34[404:DAEOAN]2.0.CO;2

Twohey, M. (2016). Evaluation of a Sterile-Male Release Technique: A Case Study of Invasive Sea Lamprey Control in a Tributary of the Laurentian Great Lakes AU - Bravener, Gale. North American Journal of Fisheries Management, 36(5), 1125-1138. https://doi.org/10.1080/02755947.2016.1204389

Wagner, C. M., Hanson, J. E., Meckley, T. D., Johnson, N. S., \& Bals, J. D. (2018). A simple , cost-effective emitter for controlled release of fish pheromones : Development , testing , and application to management of the invasive sea lamprey, 1-17. https://doi.org/10.5061/dryad.1rq65qn

Wang, J., Santiago, E., \& Caballero, A. (2016). Prediction and estimation of effective population size. Heredity, 117(4), 193-206. https://doi.org/10.1038/hdy.2016.43

Wang, J., \& Santure, A. W. (2009). Parentage and sibship inference from multilocus genotype data under polygamy. Genetics, 181(4), 1579-1594.
https://doi.org/10.1534/genetics.108.100214
Wang, Jinliang. (2009). A new method for estimating effective population sizes from a single sample of multilocus genotypes. Molecular Ecology, 18(10), 2148-2164.
https://doi.org/10.1111/j.1365-294X.2009.04175.x

Wang, Jinliang. (2016). A comparison of single-sample estimators of effective population sizes from genetic marker data. Molecular Ecology, 25(19), 4692-4711. https://doi.org/10.1111/mec. 13725

Waples, R. K., Larson, W. A., \& Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity, 117(4), 233-240. https://doi.org/10.1038/hdy.2016.60

Waples, R. S. (2010). Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. Molecular Ecology Resources, 10(5), 785-796. https://doi.org/10.1111/j.1755-0998.2010.02876.x

Waples, R. S. (2016). Making sense of genetic estimates of effective population size. Molecular Ecology, 25(19), 4689-4691. https://doi.org/10.1111/mec. 13814

Waples, R. S., Antao, T., \& Luikart, G. (2014). Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics, 197(2), 769-780. https://doi.org/10.1534/genetics.114.164822

Waples, R. S., \& Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. Evolutionary Applications, 3(3), 244-262. https://doi.org/10.1111/j.17524571.2009.00104.x

Waples, R. S., \& England, P. R. (2011). Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. Genetics, 189(2), 633-644. https://doi.org/10.1534/genetics.111.132233

Waples, R. S., Grewe, P. M., Bravington, M. W., Hillary, R., \& Feutry, P. (2018). Robust estimates of a high $\mathrm{Ne} / \mathrm{N}$ ratio in a top marine predator, southern bluefin tuna. Science Advances, 4(July).

Waples, R. S., Luikart, G., Faulkner, J. R., \& Tallmon, D. A. (2013). Simple life-history traits explain key effective population size ratios across diverse taxa. Proceedings of the Royal Society B: Biological Sciences, 280(1768), 20131339. https://doi.org/10.1098/rspb.2013.1339

Waples, R. S., \& Waples, R. K. (2011). Inbreeding effective population size and parentage analysis without parents. Molecular Ecology Resources, 11(1), 162-171. https://doi.org/10.1111/j.1755-0998.2010.02942.x

Whiteley, A. R., Coombs, J. A., Hudy, M., Robinson, Z., Nislow, K. H., \& Letcher, B. H. (2012). Sampling strategies for estimating brook trout effective population size. Conservation Genetics, 13(3), 625-637. https://doi.org/10.1007/s10592-011-0313-y

Whitlock, S. L., Schultz, L. D., Schreck, C. B., \& Hess, J. E. (2017). Using genetic pedigree reconstruction to estimate effective Spawner abundance from Redd Surveys: An example involving pacific lamprey (Entosphenus tridentatus). Canadian Journal of Fisheries and Aquatic Sciences, 74(10), 1646-1653. https://doi.org/10.1139/cjfas-2016-0154

Wright, S. (1931). Evolution in Mendelian populations. Genetics, 16, 97-159. https://doi.org/10.1007/BF02459575

Zielinski, D. P., McLaughlin, R., Castro-Santos, T., Paudel, B., Hrodey, P., \& Muir, A. (2019). Alternative Sea Lamprey Barrier Technologies: History as a Control Tool. Reviews in Fisheries Science and Aquaculture, 27(4), 438-457. https://doi.org/10.1080/23308249.2019.1625300

# CHAPTER 2: THE EFFECTS OF SAMPLING, BIOTIC, AND ENVIRONMENTAL VARIABLES ON ESTIMATES OF SEA LAMPREY EFFECTIVE BREEDING SIZE AND MINIMUM NUMBER OF SPAWNERS IN GREAT LAKES TRIBUTARES 


#### Abstract

Sea lamprey (Petromyzon marinus) are a harmful invasive species in the Great Lakes, and a large annual control and assessment program is dedicated to reducing their population size. Sea lamprey assessment is performed using larval electrofishing surveys as well as adult markrecapture estimates for estimating the number of adult sea lamprey entering streams to spawn. This assessment data are used to estimate the abundance of adult and larval sea lamprey stream populations over time to evaluate the effectiveness of control efforts. The number of successfully spawning adults is not currently assessed. Mark-recapture estimates are performed in a small number of index streams compared to the number of streams where larval assessment is conducted. Effective breeding size and minimum number of spawners were estimated for 18 larval stream populations using SNPs generated from RAD-capture sequencing. To evaluate the effects of environmental, biotic, and sampling factors on effective breeding size $\left(N_{b}\right)$ and the minimum number of spawning adults $\left(N_{s}\right)$, generalized linear models were constructed. Associations between mark-recapture estimates and $N_{b}$ and $N_{s}$ estimates were evaluated. Simulations were conducted to evaluate the potential biases of $N_{b}$ and $N_{s}$ estimates as sample size, number of SNPs, and true $N_{b}$ in the population increased. We found that sample size collected and genotyped, a sampling factor, was a significant predictor of empirical $N_{b}$ estimates; however, no correlation between mark-recapture estimates and $N_{b}$ or $N_{s}$ estimates was found. Simulations indicated that sample size and a sufficient number of SNPs become increasingly important as true $N_{b}$ increases. Additionally, the different methods of estimating $N_{b}$ have


different biases. The Chao method of calculating $N_{s}$ has less bias than the jackknife method when true $N_{b}$ is large. Overall, our results highlight the utility of $N_{b}$ and $N_{s}$ by providing insight into sea lamprey spawning populations, further demonstrate the complicated relationship between $N_{b}$ and census size, and highlight the importance of representative sampling in empirical data sets.

## INTRODUCTION

Sea lamprey (Petromyzon marinus) are a destructive invasive species in the Great Lakes region. Sea lamprey arrived in the system after the expansion of the Welland Canal in 1919, and their subsequent expansion was partially responsible for the population crash of several native fish species, including lake trout (Salvelinus namaycush) (Lawrie, 1970). Sea lamprey have a multi-stage life cycle that spans several years (Applegate, 1950). Larval lamprey spend three (Morkert et al., 1998) to seven years (Manion \& Smith, 1978) filter feeding on algae and detritus in the soft sediment sections of stream beds (Dawson et al., 2015), before metamorphosing into parasitic juveniles and migrating out of the streams. Once the individuals are in the lakes, they parasitize fish species for 12-18 months. In the following spawning year, they travel back into streams to spawn. Lamprey have a semelparous life history, only spawning once in their life cycle (Renaud, 2011). Lamprey do not return to natal streams to spawn, instead they enter streams based on temperature (Binder \& McDonald, 2008) and pheromone cues from larvae in the stream (Wagner, Twohey, \& Fine, 2009).

To control lamprey population numbers, a large control and assessment program was implemented in the region (Smith \& Tibbles, 1980). The primary methods of control since the start of the program are the selective lampricide 3-triflouromethyl-4-nitrophenol (TFM) and the use of several types of barriers to reduce upstream passage. TFM is applied in streams to kill larval lamprey in the substrate (Applegate, 1950; McDonald \& Kolar, 2007; Smith \& Tibbles, 1980), and barriers reduce the migration of adult lamprey into streams to spawn (Lavis et al., 2003; McLaughlin et al., 2007). Emerging technologies like sterile male release control and trapping techniques are utilized in a small number of streams, but are not widely applied across the region (Hume et al., 2015; Kaye et al., 2003).

The success of sea lamprey control methods is evaluated by assessment techniques such as mark-recapture estimates generated for adult spawning populations. Mark-recapture is conducted through trapping annually in a small group of index streams to estimate the number of spawning adults entering the stream system (Steeves \& Barber, 2020). Prior to 2015, these markrecapture estimates were combined with environmental data such as drainage area and years since TFM treatment to produce models of lake-wide abundance (Mullett et al., 2003).

Recently, mark-recapture estimates have been summed to provide an index of abundance for each lake that can be tracked across years, rather than an estimate of total abundance (Sullivan et al., 2016). Mark-recapture is an effective technique for estimating the number of potentially spawning adults in the stream, but mark-recapture cannot be conducted in many streams due to the environmental conditions in those systems, and the cost associated with evaluating a larger number of streams. Additionally, violation of assumptions of mark-recapture models like repeated capture of individuals, trap avoidance, and small recapture rates can complicate estimates in systems where mark-recapture is conducted annually (Bravener \& McLaughlin, 2013).

Larval surveys are an assessment technique to evaluate larval presence, the number of cohorts, cohort abundance, and distribution to prioritize streams for TFM treatments (Christie et al., 2003). Surveys use backpack electrofishing to check for the presence of larval populations in streams. Length is used to determine the number of potential transformers in the system (Christie \& Goddard, 2003; Hardisty \& Potter, 1971). Length can also be used to determine the number of cohorts in the stream, but separating age classes with length alone is difficult, particularly for larger larvae (Dawson et al., 2009). The addition of information about the stream environment can improve cohort determination using length (Dawson et al., 2020). Larval surveys are
conducted in a larger number of systems than mark-recapture surveys, but annual assessment of larval population numbers is uncommon, and parent populations cannot be evaluated with current larval assessment techniques. By generating genomic data sets for larval populations, spawning populations can be assessed.

Using genomic data, several types of parameters can be estimated and inferences based on pedigrees can be made that are useful for managing populations. Effective population size $\left(N_{e}\right)$ is a parameter describing the size of an idealized population that experiences drift or inbreeding at the same rate as the sampled population. $N_{e}$ is affected by skewed sex ratios, and the level of drift present in the population. $N_{e}$ estimation can be further complicated by large census size and highly dispersed species (Waples, Grewe, Bravington, Hillary, \& Feutry, 2018). Nonetheless, $N_{e}$ is a common and informative metric used in management contexts to evaluate species of conservation and invasive concern, and there are several common methods of estimation. The linkage disequilibrium (LD; Waples \& Do, 2010) method uses nonrandom associations between alleles in a set of loci to estimate $N_{e}$. Linkage between two loci can be caused by physical linkage through proximity on the genome or from finite population size. Thus, correlations in allele frequencies between loci that are not physically linked can provide an estimate of $N_{e}$. Sibship frequency (SF; Wang, 2009) and parentage without parents (PwoP; Waples \& Waples, 2011) methods both use reconstructed pedigrees to estimate $N_{e}$. SF utilizes the rates of full- and half-siblings present in the sequenced offspring (Wang, 2009), while PwoP uses the variation in family size to provide estimates of $N_{e}$ (Waples \& Waples, 2011). $N_{e}$ is generally calculated on a generational basis (Waples, 2016; Waples, Antao, \& Luikart, 2014), but a related measure, the effective breeding size $\left(N_{b}\right)$, can be calculated for individual cohorts with appropriate sampling (Wang, 2009; Waples, 2005; Waples \& Antao, 2014).

Another per-cohort parameter to describe adult spawning population size using genotyped offspring is the minimum number of spawning adults $\left(N_{s}\right)$. Parental genotypes reconstructed from sequenced offspring can be used to estimate the minimum number of adults required to produce the genotypes present in the offspring. However, each offspring only has at most two unique parent genotypes to contribute to the total number of spawning adults so $N_{s}$ is inherently limited by sample size, and can be reduced if the sample is not representative of the total stream population. However, these estimates can be extrapolated to the full stream population, minimizing the sample size limitation, by estimating the asymptote of an accumulation curve of unique parental genotypes (Israel \& May, 2010; Rawding et al., 2014). Like a species accumulation curve, unique parent genotypes are accumulated like unique species occurrences as the number of offspring increases. The asymptote can be calculated using a Chao or a Jackknife method, and is an estimate of the total number of successfully spawning adults in a stream system (Sard et al., in press).

In addition to genetic factors, $N_{b}$ and $N_{s}$ can also be influenced by a variety of environmental and sampling factors. Mark-recapture estimates of sea lamprey are influenced by drainage area and years since TFM treatment, and these factors could also impact $N_{b}$ and $N_{s}$. Previous work on salmonids indicated that environmental variables like stream flow and other environmental factors influence $N_{b}$ estimates, the ratio between $N_{b}$ and $N_{c}$, and the variance of parental success $\left(V_{k}\right)$ (Whiteley et al., 2015). Additionally, adequate sample size and representative sampling is vital for obtaining accurate $N_{b}$ and $N_{s}$ estimates that represent the full system (Whiteley et al., 2012). If there is a relationship between $N_{b}$ and $N_{s}$ and sampling factors that could imply that sample size was too small or that sampling was not representative of the full stream population.

Due to recent advances in genotyping technology and resources, population assessment with large sample size and SNP sets has become more feasible. Developing genomic technologies like restriction site associated DNA (RAD; Baird et al. 2008) and RAD-capture (Ali et al., 2016) sequencing allow for efficient parallel sequencing, increasing the number of sequenced individuals. A chromosome-level and a germline genome have been assembled for sea lamprey (Smith et al., 2013, 2018). Additionally, a RAD-capture panel was recently published for sea lamprey (Sard et al., 2020), allowing for the sequencing of a large number of individuals at a specific known group of variable sequences present in the sea lamprey genome.

Due to annual collections during larval assessment, there is an opportunity to use population genetics methods to assess adult spawning populations using larval collections. Additionally, assessments of $N_{c}$ in a group of index streams allows for comparison between census size and genetic estimates. $N_{b}$ and $N_{s}$ need to be assessed for utility in sea lamprey both through empirical data sets with sequenced offspring and compared through simulated populations to examine the relative performance of all estimate types as sample size, SNP set, and true $N_{b}$ of the system changes. Additional testing to evaluate the effects of environmental, biotic, and sampling factors on $N_{b}$ and $N_{s}$ and the correlation between $N_{b}, N_{s}$, and mark-recapture census size estimates is needed. In this study we estimated $N_{b}$ and $N_{s}$ in a series of Great Lakes tributaries to assess the utility of these estimates for sea lamprey assessment and evaluate the influences of environmental, biotic, and sampling variables on effective breeding size and the minimum number of spawners.

## METHODS

## Sample Collection

Sea lamprey larvae ( $\mathrm{n}=1,877$ ) were collected via backpack electrofishing during larval assessment surveys in 18 streams across the Great Lakes system by collaborators at US Fish and Wildlife Service, United States Geological Survey, and Fisheries and Oceans Canada (Figure 2.1). Collections occurred in the summer and fall of 2019, with the exception of the Middle River, where collections occurred in the summer of 2017. Stream systems ranged from large rivers like the Muskegon River (Drainage $=7,327 \mathrm{ha}$ ) to small streams like Swan Creek (Drainage $=5 h a$ ). All systems are annual sea lamprey producing streams, where TFM treatments were conducted within 10 years of sample collection. At each collection site, larvae were identified to species, anesthetized with MS-222, and euthanized with 95\% ethanol.


Figure 2.1. Map showing all sampled streams with their location in the Great Lakes system. Each dot represents a stream system.

## Sequencing Library Preparation

A tissue sample from each larva was preserved in $95 \%$ ethanol for subsequent DNA extraction and sequencing. DNA extractions were performed using Qiagen DNeasy blood and tissue kits (QIAGEN, Carlsbad, CA) and all DNA concentrations were quantified with a Nanodrop ND-1000 Spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts) and Quant-iT ${ }^{\mathrm{TM}}$ PicoGreen ${ }^{\mathrm{TM}}$ dsDNA Assay Kits (Thermo Fisher Scientific Inc., Waltham, Massachusetts) with a QuantStudio 6 Flex Real-Time PCR system (Thermo Fisher Scientific Inc., Waltham, Massachusetts). DNA was standardized to concentrations below $100 \mathrm{ng} / \mu \mathrm{l}$ for RAD library constructions. Stream populations were randomly distributed across libraries to minimize library effects.

Reduced representation libraries were constructed using a modified version of the BestRAD protocol from Ali et al. (2016). DNA was digested with an SbfI restriction enzyme and biotinylated BestRAD adapters were ligated to samples to serve as individual barcodes. The barcoded DNA was pooled, concentrated with Ampure beads (Beckman Coulter, Indianapolis, Indiana), and sheared to 325 bp using a Covaris m220 focused-ultrasonicator (Covaris, Woburn, Massachusetts). DNA fragments with attached bestRAD tags were selected using a streptavidin bead binding assay, and size selection was used to select target size fragments. A 22:50 ratio of Ampure beads was used to select long fragments and a 13:72 ratio was used to separate target size fragments from short fragments. NEBNext kits (New England BioLabs Inc, Ipswich, Massachusetts) were used to ligate plate-specific Illumina adaptors and an Illumina universal adapter was used to prepare the library for sequencing. $\sim 3400$ SNPs across individuals were targeted for sequencing using a custom MyBaits hybridization capture kit (Arbor Biosciences, Ann Arbor, MI), designed by Sard et al. (2020), with the manufacturer protocol and eleven PCR
cycles in the final amplification step. All libraries were sequenced on an Illumina HighSeq X at Novogene (Chula Vista, CA) with paired-end 150 base pair sequencing.

## Bioinformatic Analysis

A bioinformatic pipeline based on Sard et al. (2020) was used to process read data. First, reads were oriented with a custom perl script bRAD_flip (originally developed by Paul Hohenlohe, University of Idaho, and modified by Brian Hand and Seth Smith, University of Montana) and demultiplexed from library to individual data with the Stacks 2.0 function process_radtags (Catchen et al., 2013). Cloned reads were removed from each individual with Stacks 2.0 function cloneFilter (Catchen et al., 2013), and trimmed and quality filtered with Trimmomatic (Bolger et al., 2014) with a minimum length of 50 , a sliding window of 4 bases, and a minimum quality score of 15 . BWA-mem (Li, 2013; Li \& Durbin, 2010) was then used to map all reads to the sea lamprey chromosome-level reference genome
(https://genomes.stowers.org/sealamprey). SAMtools (version 1.9; Li et al., 2009) was used to sort mapped reads. The sorted reads were genotyped using the stacks function gStacks (Catchen et al., 2013), and a sorted VCF file was generated along with population-level statistics using the Stacks function populations. SNP data were initially filtered for 8X depth for final genotype calls. For each population, HDplot was run to filter potentially paralogous loci from the data set. If observed heterozygosity was greater than 0.6 or the absolute value of the read ratio deviation statistic (D) was greater than 7, the locus was removed from the data set (McKinney et al., 2017). Additionally, SNPs were checked for deviance from Hardy-Weinburg equilibrium across populations using the output from populations (Catchen et al., 2013). No SNPs were found to be deviant from Hardy-Weinburg equilibrium across populations.

Estimation methods for generating $N_{b}$ estimates and reconstructed pedigrees have different data requirements to run optimally. Thus, two SNP sets were generated for each population with different filtering parameters: one for linkage disequilibrium $N_{b}$ estimates, and the other to generate a reconstructed pedigree. For both datasets, SNPs were filtered to exclude loci that were not targeted by the RAD-capture bait panel, and loci where fewer than $80 \%$ of individuals were genotyped. The linkage disequilibrium dataset was limited to one SNP per RAD-capture tag, where the selected SNP had the highest percentage of individuals genotyped among the SNPs on the tag. SNP loci in the dataset used for pedigree reconstruction in Colony (Jones and Wang 2010) were selected using a sliding window of 1 MB to minimize linkage among SNPs, with selection biased towards high minor allele frequency (minimum value of 0.05 ) and high percent genotyped to maximize information content of the dataset. Colony version 2.0.6.6 (Jones \& Wang, 2010) was run for each stream population to reconstruct the pedigrees of each system for the cohort-determining models described below. The full-likelihood approach with a medium-length run was used for all streams. Other input parameters changed from default settings were unknown allele frequencies, polygamous mating, and no sibship scaling or prior sibship reported.

The accidental inclusion of native lamprey in genomic estimates of sea lamprey could cause bias and $N_{b}$ and $N_{s}$ estimates, thus potential native lamprey samples need to be identified and excluded from subsequent analysis. To identify sampled lamprey to species, 10 individuals from two native lamprey samples (L. appendix and I. fossor) were sequenced using the same library preparation and SNPs were identified using the techniques as the sampled individuals. A PCA was conducted with known native lamprey and sampled individuals for each stream population to identify outliers that sort with native lamprey samples instead of the other stream
samples. No outliers were identified, indicating that all samples that were sequenced were sea lamprey samples.

## Cohort-determining models

Length measurements and blotted weight were taken for all collected individuals for mixture models first introduced in Chapter 1. A combination of Gaussian mixture models and reconstructed pedigree data were used to determine the cohort assignments of larval offspring (Figure 1.2). First, inferred cohort groupings were generated from length data using the R packages bmixture (Mohammadi et al., 2013) and BayesMix (Grün \& Leisch, 2010). The number of inferred cohorts (K) was determined by using the birth-death MCMC algorithm implemented in bmixture (Mohammadi et al., 2013) with a maximum K of 4 and 500,000 iterations. The posterior probability of K values was estimated using the proportion of MCMC iterations with a given K value. BayesMix was also used to select K with criteria proposed by Rousseau and Mengerson (2011), which involves fitting a model with a large K value ( $\mathrm{n}=10$ ) and identifying the number of non-empty clusters (less than 0.035 sorted into the given K value) after 500,000 iterations (Nasserinejad et al., 2017; Rousseau \& Mengersen, 2011). Locations with a sample size of less than 80 were excluded from mixture models based on recommendations from Rousseau and Mengerson (2011), which indicated that mixture models with small sample size produced variable results. The posterior probability of K values was based on the proportion of steps with K non-empty clusters. BayesMix was rerun with the optimal K value for each stream to generate individual cohort assignments based on length.

For locations with multiple length cohorts, the level of colony clusters across inferred cohorts was assessed and used to check the cohort assignments with the decision-making chart
defined in Chapter 1 (Weise et al in review). Colony clusters refer to groups of offspring that connected in the pedigree but are not necessarily full- or half-siblings. For example, if offspring 1 and offspring 2 are maternal half-sibling, and offspring 2 and offspring 3 are paternal halfsiblings, offspring 1 and 3 are unrelated but still connected in the pedigree through offspring 2 . All three of those individuals would be in a single colony cluster. All analyses were conducted in R (R Core Team, 2019).

## $N_{b}$ and $N_{s}$ estimates

For all cohorts, three estimates of $N_{b}$ and two estimates of $N_{s}$ were generated. Estimates from the linkage disequilibrium method (LD; Waples and Chi 2008) were calculated using NeEstimator (Do et al., 2013). A MAF cutoff of 0.05 was specified and locus pairs within chromosomes were excluded from the calculation of correlation in allele frequency to reduce the effects of linkage from proximity in the genome (Waples, Larson, \& Waples, 2016). Confidence intervals were estimated using a jackknife method with a correction to minimize effects of large SNP sets on confidence intervals (Jones, Ovenden, \& Wang, 2016). Colony was run with the same parameters as full stream populations to calculate $N_{b}$ with the sibship frequency method (SF; Wang and Santure 2009). Estimates of $N_{b}$ from the parentage without parents method were calculated using a custom R script based on equations in Waples and Waples (2011), and the uncertainty of those estimates was generated based on equations from Wang 2009. Specifically, variance for $1 / 2 N_{b}$ was calculated for archived configurations of the reconstructed pedigree generated by Colony and for simulated populations with equal sex ratio and the same $N_{b}$ as the empirical estimates. These sources of variance were summed and converted into confidence intervals that incorporate sampling uncertainty and uncertainty associated with the construction
of the reconstructed pedigree. Additionally, the mean $(\bar{k})$ and variance $\left(V_{k}\right)$ of adult reproductive success for contributing adults were calculated for each reconstructed pedigree. $N_{s}$ was estimated by counting the number of unique parent genotypes in the reconstructed pedigree, then extrapolated to the minimum number of parents in the stream system $\left(\widehat{N_{s}}\right)$ using both the Chao (Chao, 1987) and jackknife (Heltshe \& Forrester, 2009) methods with the function specpool from the R package vegan (Oksanen et al., 2019).

## Statistical Analyses

Sampling, biotic, and environmental factors may influence estimates of $N_{b}$ and $\widehat{N_{s}}$ in the sampled systems. For instance, if sampling sizes are too small or if sampling is not representative, there could be a linear relationship between $N_{b}$ and $\widehat{N_{s}}$ estimates and sampling size or the number of sampling sites. Linear models were used to assess the influence of several factors on $N_{b}, \widehat{N_{s}}$, and $V_{k}$ for the sampled systems. We evaluated a global model that included several stream characteristics that can affect the $N_{b}$ to $N_{c}$ ratio or $N_{b}$ and $\widehat{N_{s}}$ estimates: representations of sampling structure (sample size, the linear distance of sampling sites for each stream), population size of spawning adults (census size for index streams, drainage area of the stream), and factors that could lead to bias in mark-recapture estimates (trapping distance from the mouth of the river for index streams). Publicly available reports on current sea lamprey control and assessment methods were used to collect information on TFM treatment years and census size estimates for the adult population that was assumed to have produced the sequenced larvae (Barber \& Steeves, 2019; Mullett \& Sullivan, 2017). Personal communications with USFWS collaborators and unpublished data from co-authors were used to obtain data on drainage area for each stream (J. Adams, pers. comm.) and the distance of trap sites from the
mouth of the river for mark-recapture streams (G. Bravener, pers. comm.). The number of larval collection sites and approximate distance of larval collection were also included in a separate model that did not include $N_{c}$ estimates or trapping distance to minimize missing data from both models.

Generalized linear models with the above environmental variables as factors were generated for the estimate of $N_{b}$ based on both LD and SF methods, Chao estimates of $\widehat{N_{s}}$, as well as $V_{k}$. Model selection was conducted using Akaike Information Criteria (AIC; Akaike, 1974) values with a sample size correction (AICc; Hurvich and Tsai 1989). Akaike weights were used to sort models from most to least probable, and the confidence set of models was defined as the best-supported models with a cumulative Akaike weight of 0.9 (Akaike, 1978). Relationships between mark-recapture estimates of census size (Barber \& Steeves, 2019; Mullett \& Sullivan, 2017) and estimates of $N_{b}$ and $N_{s}$ across streams were evaluated via Pearson product-moment correlation tests.

## Effects of Sample Size, SNP set size, and stream $\boldsymbol{N}_{\boldsymbol{b}}$ on genetic estimates

Factors like sample size, the number of SNP loci, and $N_{b}$ of the sampled population can affect the accuracy and precision of $N_{b}$ and $\widehat{N_{s}}$ estimates. To specifically examine the effects of these factors, we estimated $N_{b}$ and $\widehat{N_{s}}$ (using each the methods described above) in simulated genetic datasets. The individual-based simulation model implemented in the R package rmetasim (Strand, 2002) was used to generate 50 replicate populations of six different sizes $\left(N_{b}=50,100\right.$, $500,1000,5000$, and 10000) by initializing a landscape with a carrying capacity of the desired $N_{b}$, polygamous mating for both males and females, and no survivorship after mating to represent semelparous life history. Allele frequencies in simulated populations were initialized using
empirical allele frequencies from the individuals sequenced in Chapter 1 for 1,200 diploid SNP loci from the data set. After 50 generations of burn-in, 200 generations were run where both genotypes and heterozygosity were tracked. Fixed SNPs were removed and all individual genotypes were output for each simulation. True $N_{b}$ was calculated using $V_{k}$ and $\bar{k}$ estimates generated from each simulated population, and the number of parents in the full population was used as true $N_{s}$. These replicates were subsampled for five sample size values $(25,50,100,150$, and 200 individuals) and three SNP set sizes ( 100,500 , and 1000 loci). $N_{b}$ and $\widehat{N_{s}}$ for all simulated datasets were estimated using the programs NeEstimator (Do et al., 2013) and Colony (Jones \& Wang, 2010) with similar input parameters as used for empirical datasets. In NeEstimator, no reference genome was used since the simulated SNPs were modeled as unlinked, independent loci. For the Colony pedigree reconstruction analyses, the FPLS (combined full-likelihood and pairwise-likelihood score) method was used instead of full likelihood for computational efficiency.
$N_{b}$ and $\widehat{N_{s}}$ were estimated using the same methods described for the empirical data. Harmonic mean and median point estimates from each scenario were compared to the true $N_{b}$ and $N_{s}$ value from full simulated populations to examine the accuracy of estimated $N_{b}$ and $\widehat{N_{s}}$ for each parameter set. Root-mean-square error (RMSE) were calculated for each parameter set. All scripts and summary files are available on Github (https://github.com/weiseell/OliviaSims).

## RESULTS

## Read Processing

Larval sea lamprey were sequenced in 20 libraries across 5 sequencing lanes, with an average read count of $234,404,375$ reads per library. The average number of demultiplexed reads was 2,085,919 per individual. The average depth per individual was 26X across all SNPs. After applying a depth filter of $8 \mathrm{X}, 200,190$ identified SNPs remained in the data set. $64.48 \%$ of those identified SNPs matched to the Rapture panel, and an average of $16.1 \%$ of individuals were genotyped per SNP. The SNP sets for the linkage disequilibrium estimates contained between 2,659 and 3,018 SNPs with an average $97.7 \%$ percent of individuals genotyped and 0.12 average MAF (Table 2.1). The SNP sets for pedigree reconstruction contained between 627 and 683 SNPs with $92.5 \%$ percent of individuals genotyped and 0.24 average MAF (Table 2.1). Between populations, there was an average of $69.5 \%$ overlap between SNPs selected for the Colony sets, and $21.0 \%$ overlap among the NeEstimator SNPs. However, there was $98 \%$ overlap among targeted loci selected for the NeEstimator SNP sets.

Table 2.1. Table showing the SNP set size, as well as the average MAF and percent of individuals genotyped, for each SNP set. SNPs refers to the size of the SNP set, pGT refers to the proportion of individuals genotyped across SNPs in the SNP set, and MAF refers to the average minor allele frequency across SNPs in the SNP set.

|  |  |  | NeEstimator |  |  | Colony |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Lake | Stream | Pop | SNPs | pGT | MAF | SNPs | pGT | MAF |
| Superior | Bad River | BAD | 2810 | 0.953 | 0.19 | 658 | 0.929 | 0.23 |
| Michigan | Betsie River | BEI | 3012 | 0.991 | 0.06 | 683 | 0.924 | 0.23 |
| Superior | Betsy River | BET | 3008 | 0.985 | 0.11 | 682 | 0.924 | 0.24 |
| Superior | Brule River | BRL | 2937 | 0.986 | 0.17 | 675 | 0.925 | 0.24 |
| Erie | Cattaraugus River | CAT | 3010 | 0.98 | 0.09 | 679 | 0.925 | 0.24 |
| Huron | East Au Gres River | EAG | 2990 | 0.977 | 0.12 | 678 | 0.92 | 0.23 |
| Michigan | Ford River | FOR | 3015 | 0.995 | 0.05 | 680 | 0.925 | 0.23 |
| Michigan | Manistee River | MAN | 3016 | 0.996 | 0.03 | 683 | 0.924 | 0.24 |
| Michigan | Manistique River | MAI | 2883 | 0.949 | 0.18 | 682 | 0.925 | 0.23 |
| Superior | Middle River | MIR | 3018 | 0.98 | 0.07 | 679 | 0.926 | 0.24 |
| Superior | Misery River | MIS | 2914 | 0.977 | 0.17 | 670 | 0.924 | 0.23 |
| Michigan | Muskegon River | MUS | 2985 | 0.989 | 0.13 | 677 | 0.925 | 0.24 |
| Huron | Ocqueoc River | OCQ | 2999 | 0.987 | 0.1 | 680 | 0.926 | 0.23 |
| Huron | Pigeon River | CHE | 2872 | 0.98 | 0.16 | 627 | 0.926 | 0.28 |
| Ontario | Sterling River | STE | 2985 | 0.978 | 0.13 | 679 | 0.921 | 0.22 |
| Michigan | Swan Creek | SWN | 2659 | 0.92 | 0.21 | 661 | 0.929 | 0.23 |
| Superior | Tahquamenon River | TAQ | 2982 | 0.972 | 0.12 | 678 | 0.924 | 0.24 |
| Superior | Two-Hearted River | TWO | 2946 | 0.984 | 0.16 | 676 | 0.925 | 0.24 |

## Mixture Models

Across 18 locations, ten had a sample size sufficient for cohort-assignment models ( $n>$ 80 individuals). Of those locations, four had multiple inferred cohorts based on the mixture analysis results (Table 2.2, Figure 2.2). When the two methods did not agree, the method that yielded a higher posterior probability was used to determine the K used for the Gaussian mixture models. Additionally, the Middle River and the Manistee River were the only locations with sequenced individuals sorted into multiple cohorts based on both the mixture analysis results (Table 2.2, Figure 2.2) as well as the cohort assignments as visualized in the boxplots (Figure 2.3). For other locations with multiple inferred cohorts, only inferred age-1 individuals were sequenced.

All cohorts of sequenced offspring had full- or half-sibling families present in the sample (Figure 2.4). Cohorts identified in most streams have a mixture of full- and half-sibling families. However, Swan Creek and Bad River have a small number of full-sibling families, and comparatively few half-sibling families (Figure 2.4). Locations like the Ocqueoc River and the Muskegon River are represented by mostly unrelated individuals (Figure 2.4).

Table 2.2. Summary of results for identifying the optimal number of clusters $(\mathrm{K})$ in the mixture analysis for sea lamprey. Analyses were performed for each larval collection with a range of $\mathrm{K}=1-4$ clusters. $\mathrm{R} \& \mathrm{M}$ criteria and $\mathrm{BD}-\mathrm{MCMC}$ shows the estimated probability of each K value from the Rousseau and Mengersen (2011) criteria and BD-MCMC, respectively. The optimal number of clusters from each method is bolded. If the two methods disagree, the method with the higher probability is used.

|  |  |  | K - BD-MCMC |  |  |  | K - R\&M Criteria |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Lake | Stream | Pop | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | SelectK |
| Michigan | Betsie | BEI | 0.048 | 0.223 | 0.469 | 0.26 | 0.673 | 0.278 | 0.044 | 0.004 | 1 |
| Superior | Betsy | BET | 0.018 | 0.116 | 0.407 | 0.459 | 0.954 | 0.046 | 0.001 | 0 | 1 |
| Erie | Cattaraugus | CAT | 0.013 | 0.094 | 0.398 | 0.495 | 0.461 | 0.539 | 0 | 0 | 2 |
| Michigan | Ford | FOR | 0.188 | 0.693 | 0.113 | 0.006 | 0.918 | 0.079 | 0.003 | 0 | 1 |
| Michigan | Manistee | MAN | 0.027 | 0.144 | 0.406 | 0.423 | 0 | 0.996 | 0.004 | 0 | 2 |
| Superior | Middle | MIR | 0.001 | 0.027 | 0.346 | 0.626 | 0 | 0.899 | 0.101 | 0 | 2 |
| Huron | Ocqueoc | OCQ | 0.012 | 0.088 | 0.416 | 0.484 | 0.915 | 0.082 | 0.003 | 0 | 1 |
| Ontario | Sterling | STE | 0.162 | 0.608 | 0.208 | 0.022 | 0.918 | 0.08 | 0.003 | 0 | 1 |
| Superior | Tahquamenon | TAQ | 0.054 | 0.253 | 0.45 | 0.243 | 0.876 | 0.112 | 0.006 | 0 | 1 |



Figure 2.2. Length frequency distributions for larval sea lamprey from all rivers and collection years, fill colors represent individual cluster assignment from the Gaussian mixture analysis. If mixture models were not completed due to small sample size ( $\mathrm{n} \leq 80$ ), length histograms are included and shaded as purple.


Figure 2.3. Boxplots showing the length distributions of each Colony cluster of sequenced offspring. Boxes are shaded by the cluster likelihood, where lower likelihoods are shaded towards red and higher likelihoods are shaded towards white. Boxplots are limited to clusters with 3 or more individuals. The East Au Gres and the Muskegon River are not shown because they do not have any clusters larger than 3 individuals.


Figure 2.4. Diagrams of reconstructed pedigrees for all stream systems. The offspring are in the center of the diagram and are connected to their reconstructed parents by grey lines. The offspring are sorted first by parent 1 sibling groups, then parent 2 sibling groups.

## $N_{b}, N_{s}$, and $\widehat{N_{s}}$ estimates

In most systems, $N_{b}$ estimates were of similar magnitude across the three methods. Estimates from the PwoP method and SF method matched more closely with each other than with estimates from the linkage disequilibrium method (Table 2.3), which was expected based on previous work comparing the two estimators (Ackerman et al., 2017). In systems where the LD method did not agree with the PwoP and SF methods, LD was generally lower than the $N_{b}$ estimates from the reconstructed pedigrees (Table 2.3). The largest estimates occurred in the Middle River ( $N_{b}=230-350$ and the Muskegon River ( $N_{b}=255-309$, and the smallest estimates occurred in Swan creek $\left(N_{b}=6-7\right)$, the Bad River $\left(N_{b}=3-6\right)$, and the East Au Gres River ( $N_{b}=$ $0.2-0.3)$. Five locations had $N_{b}$ estimates under 10 across all three methods. Variance in individual reproductive success $\left(V_{k}\right)$ was less than 50 across most systems, the highest variance occurred in the Sterling River ( $V_{k}=347.92$ ).

The Chao and Jackknife extrapolated estimates of $\widehat{N_{s}}$ were similar in most sampled systems, and were generally higher than the $N_{b}$ estimates for each cohort. Some systems had a small sample size (less than 50 individuals), but finite estimates were still calculable due to the fact that all samples contains some related individuals (Figure 2.5). Confidence intervals were potentially artificially narrow due to the large number of SNPs used in the analyses (Waples, Waples, \& Ward, 2020), although the corrected jackknife estimates used should reduce that bias. All seven locations with small sample size had LD $N_{b}$ estimates of less than 100. Of these locations, three had mark-recapture estimates of over 10,000 and $N_{s}$ estimates of less than 100 (Table 2.3).

Table 2.3. $N_{b}$ and $N_{s}$ estimates and population-based information. N indicates the number of sequenced offspring for the cohort, $N_{c}$ is the census-size estimate based on mark-recapture during the spawning year. linkage disequilibrium (LD), parentage without parents (PwoP) and sibship frequency (SF) columns are $N_{b}$ point estimates with corresponding uncertainty. LD: $N_{c}$ and SF : $N_{c}$ refer to the ratios between the LD and SF method of estimating $N_{b}$ and the mark-recapture $N_{c}$ estimates. $\bar{k}$ and $V_{k}$ are the mean and variance in reproductive success inferred from the reconstructed pedigree. $N_{s}$ is the number of reconstructed parent genotypes for each cohort, and Chao and Jackknife are the $\widehat{N_{S}}$ estimates and their corresponding 95\% confidence intervals.

| Lake | Stream | Pop | N | $N_{c}$ | LD | PwoP | SF | LD: $N_{c}$ | SF: $\boldsymbol{N}_{c}$ | $\overline{\boldsymbol{k}}$ | $V_{k}$ | $N_{s}$ | Chao | Jackknife |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Superior | Bad River | BAD | 37 | 11301 | $\begin{gathered} \hline 2.7 \\ (2.3- \\ 3.3) \end{gathered}$ | $\begin{gathered} 6.22 \\ (5.78- \\ 6.74) \\ \hline \end{gathered}$ | $\begin{gathered} 6 \\ (2- \\ 20) \end{gathered}$ | 0.0002 | 0.0005 | 8.22 | 37.06 | 9 | $\begin{gathered} 9.97 \pm \\ 2.2 \end{gathered}$ | $\begin{gathered} 10.95 \pm \\ 1.96 \end{gathered}$ |
| Michigan | Betsie River | BEI | 160 | 1654 | $\begin{gathered} 62.1 \\ (48.7- \\ 82.1) \\ \hline \end{gathered}$ | $\begin{gathered} 80.51 \\ (68.58- \\ 95.5) \end{gathered}$ | $\begin{gathered} 80 \\ (58- \\ 113) \\ \hline \end{gathered}$ | 0.0375 | 0.0484 | 2.03 | 1.92 | 78 | $\begin{gathered} 97.41 \pm \\ 8.32 \end{gathered}$ | $\begin{gathered} 112.58 \pm \\ 7.41 \end{gathered}$ |
| Superior | Betsy River | BET | 246 | 1097 | $\begin{gathered} 58.6 \\ (46.2- \\ 75.8) \\ \hline \end{gathered}$ | $\begin{gathered} \hline 78.78 \\ (71.06- \\ 88.13) \\ \hline \end{gathered}$ | $\begin{array}{r} 78 \\ (57- \\ 107) \\ \hline \end{array}$ | 0.0534 | 0.0711 | 2.34 | 4.15 | 105 | $\begin{gathered} 188.34 \pm \\ 31.22 \end{gathered}$ | $\begin{gathered} 159.55 \pm \\ 9.58 \end{gathered}$ |
| Superior | Brule River | BRL | 33 | 36558 | $\begin{gathered} \hline 73.2 \\ (40.5- \\ 230.5) \\ \hline \end{gathered}$ | $\begin{aligned} & \hline 112.89 \\ & (77.29- \\ & 209.29) \\ & \hline \end{aligned}$ | $\begin{aligned} & \hline 111 \\ & (72- \\ & 203) \\ & \hline \end{aligned}$ | 0.0020 | 0.0030 | 1.29 | 0.36 | 51 | $\begin{gathered} 161.82 \pm \\ 55.62 \end{gathered}$ | $\begin{gathered} 89.79 \pm \\ 8.52 \end{gathered}$ |
| Erie | Cattaraugus River | CAT | 241 | 1637 | $\begin{gathered} 33.3 \\ (29.3- \\ 37.7) \end{gathered}$ | $\begin{gathered} \hline 39.81 \\ (38.38- \\ 41.35) \end{gathered}$ | $\begin{gathered} \hline 40 \\ (28- \\ 62) \end{gathered}$ | 0.0203 | 0.0244 | 6.89 | 42.67 | 70 | $\begin{gathered} 76.02 \pm \\ 4.78 \end{gathered}$ | $\begin{gathered} 80.95 \pm \\ 4.11 \end{gathered}$ |
| Huron | Pigeon River | CHE | 51 | NA | $\begin{gathered} 8.0 \\ (6.3- \\ 9.7) \\ \hline \end{gathered}$ | $\begin{gathered} 3.90 \\ (3.74- \\ 4.07) \\ \hline \end{gathered}$ | $\begin{gathered} 4 \\ (2- \\ 12) \\ \hline \end{gathered}$ | NA | NA | 9.27 | 163.65 | 11 | $\begin{gathered} 11.98 \pm \\ 1.84 \end{gathered}$ | $\begin{gathered} 12.96 \pm \\ 1.39 \end{gathered}$ |
| Huron | East Au Gres River | EAG | 21 | 2124 | $\begin{gathered} \hline 0.3 \\ (0.2- \\ 0.3) \end{gathered}$ | $\begin{gathered} 172.2 \\ (90.93- \\ 1619.39) \\ \hline \end{gathered}$ | $\begin{array}{r} 168 \\ (84- \\ 1201) \\ \hline \end{array}$ | 0.0001 | 0.0791 | 1.14 | 0.12 | 37 | $\begin{gathered} 134.52 \pm \\ 56.47 \end{gathered}$ | $\begin{gathered} 67.48 \pm \\ 7.59 \end{gathered}$ |
| Michigan | Ford River | FOR | 122 | NA | $\begin{gathered} \hline 44.2 \\ (33.1- \\ 60.4) \\ \hline \end{gathered}$ | $\begin{gathered} \hline 40.33 \\ (37.68- \\ 43.4) \\ \hline \end{gathered}$ | $\begin{gathered} \hline 40 \\ (27- \\ 62) \end{gathered}$ | NA | NA | 2.18 | 10.56 | 112 | $\begin{gathered} 188.76 \pm \\ 25.13 \end{gathered}$ | $\begin{gathered} 178.45 \pm \\ 10.08 \end{gathered}$ |
| Michigan | Manistique River | MAI | 30 | 10420 | $\begin{gathered} 6.7 \\ (3.7- \\ 10) \\ \hline \end{gathered}$ | $\begin{gathered} 12.04 \\ (10.44- \\ 14.22) \\ \hline \end{gathered}$ | $\begin{aligned} & 11 \\ & (6- \\ & 26) \\ & \hline \end{aligned}$ | 0.0006 | 0.0011 | 4.00 | 7.6 | 15 | $\begin{gathered} 22.73 \pm \\ 11.28 \end{gathered}$ | $\begin{gathered} 18.87 \pm \\ 1.93 \end{gathered}$ |
| Michigan | Manistee River | MAN | 185 | 7219 | $\begin{gathered} \hline 126.0 \\ (100.8- \\ 162.4) \\ \hline \end{gathered}$ | $\begin{gathered} 178.32 \\ (160.94- \\ 199.91) \\ \hline \end{gathered}$ | $\begin{gathered} 178 \\ (141- \\ 227) \\ \hline \end{gathered}$ | 0.0175 | 0.0247 | 1.90 | 1.92 | 180 | $\begin{gathered} 312.60 \pm \\ 35.68 \end{gathered}$ | $\begin{gathered} 281.40 \pm \\ 13.21 \end{gathered}$ |

Table 2.3 (cont'd).

| Superior | Middle River | MIR | 447 | 4705 | $\begin{gathered} \hline 230.9 \\ (204.6- \\ 262.4) \\ \hline \end{gathered}$ | $\begin{gathered} \hline 350.01 \\ (330.07- \\ 372.50) \\ \hline \end{gathered}$ | $\begin{gathered} 350 \\ (293- \\ 413) \\ \hline \end{gathered}$ | 0.0744 | 0.0491 | 2.17 | 2.86 | 401 | $\begin{gathered} 567.75 \pm \\ 31.77 \end{gathered}$ | $\begin{gathered} 590.56 \pm \\ 16.58 \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Superior | Misery River | MIS | 37 | NA | $\begin{gathered} \hline 9.8 \\ (7.9- \\ 12.0) \end{gathered}$ | $\begin{gathered} 9.58 \\ (8.70- \\ 10.65) \\ \hline \end{gathered}$ | $\begin{gathered} 9 \\ (5- \\ 24) \\ \hline \end{gathered}$ | NA | NA | 5.29 | 17.63 | 14 | $\begin{gathered} 14.24 \pm \\ 0.71 \end{gathered}$ | $\begin{gathered} 14.97 \pm \\ 0.97 \end{gathered}$ |
| Michigan | Muskegon River | MUS | 53 | NA | $\begin{gathered} \hline 255.0 \\ (172.8- \\ 467.0) \\ \hline \end{gathered}$ | $\begin{gathered} 309.17 \\ (209.62- \\ 588.8) \end{gathered}$ | $\begin{gathered} 306 \\ (202- \\ 594) \end{gathered}$ | NA | NA | 1.19 | 0.18 | 89 | $\begin{gathered} 263.28 \pm \\ 62.16 \end{gathered}$ | $\begin{gathered} 160.62 \pm \\ 11.01 \end{gathered}$ |
| Huron | Ocqueoc River | OCQ | 121 | 4813 | 134.3 (105.0179.9) | 184.56 (158.87220.17) | $\begin{gathered} 184 \\ (142- \\ 237) \end{gathered}$ | 0.0279 | 0.0382 | 1.09 | 1.09 | 147 | $\begin{gathered} 248.05 \pm \\ 28.88 \end{gathered}$ | $\begin{gathered} 234.27 \pm \\ 11.01 \end{gathered}$ |
| Ontario | Sterling River | STE | 105 | 2868 | $\begin{gathered} \hline 6.6 \\ (5.2- \\ 7.9) \end{gathered}$ | $\begin{gathered} 5.74 \\ (5.60- \\ 5.90) \\ \hline \end{gathered}$ | $\begin{gathered} 6 \\ (3- \\ 21) \end{gathered}$ | 0.0023 | 0.0382 | 17.50 | 347.92 | 12 | $12 \pm 0.47$ | $\begin{gathered} 12.99 \pm \\ 0.99 \end{gathered}$ |
| Michigan | Swan Creek | SWN | 39 | NA | $\begin{gathered} 1.9 \\ (1.7- \\ 2.0) \\ \hline \end{gathered}$ | $\begin{gathered} 4.41 \\ (4.09- \\ 4.78) \\ \hline \end{gathered}$ | $\begin{gathered} 4 \\ (2- \\ 12) \\ \hline \end{gathered}$ | NA | NA | 11.14 | 81.55 | 7 | $\begin{gathered} 9.92 \pm \\ 4.33 \end{gathered}$ | $9.92 \pm 2.19$ |
| Superior | Tahquamenon River | TAQ | 94 | 3974 | $\begin{gathered} 52.0 \\ (39.0- \\ 72.3) \\ \hline \end{gathered}$ | $\begin{gathered} \hline 68.93 \\ (61.49- \\ 78.43) \\ \hline \end{gathered}$ | $\begin{gathered} 69 \\ (49- \\ 98) \\ \hline \end{gathered}$ | 0.0131 | 0.0174 | 2.29 | 3.26 | 82 | $\begin{gathered} 125.86 \pm \\ 1467 \end{gathered}$ | $\begin{gathered} 118.61 \pm \\ 7.35 \end{gathered}$ |
| Superior | Two-Hearted River | TWO | 43 | NA | $\begin{gathered} 25.6 \\ (17.2- \\ 40.7) \end{gathered}$ | $\begin{gathered} \hline 30.21 \\ (25.95- \\ 36.14) \\ \hline \end{gathered}$ | $\begin{gathered} 30 \\ (19- \\ 51) \\ \hline \end{gathered}$ | NA | NA | 2.05 | 3.62 | 42 | $\begin{gathered} 77.16 \pm \\ 19.89 \end{gathered}$ | $\begin{gathered} 65.44 \pm \\ 6.51 \end{gathered}$ |



Figure 2.5. Minimum number of spawning adults ( $\widehat{N_{s}}$ ) accumulation curves showing the increase in unique parent genotypes as the number of sequenced offspring increased for each cohort. The dark red lines in each plot represent the chao asymptote estimates (Chao, 1987a), and the dark blue lines represent the jackknife asymptote estimates (Heltshe \& Forrester, 2009).

## Correlations and Linear Modeling

Correlations between $N_{c}, N_{b}$, and $\widehat{N_{s}}$ were not significant with either a Pearson productmoment correlation or a non-parametric Spearman rank order test (Figure 2.6). $N_{b}$ and $\widehat{N_{s}}$ estimates were highly correlated ( SF and $\widehat{N_{s}}:$ corr $=0.954(\mathrm{p}<0.001)$, LD and $\widehat{N_{s}}:$ corr $=0.951(\mathrm{p}$ $<0.001)$ ), indicating consistency across the two types of genetic estimates. Additionally, a version of the model run with an $N_{c}$ estimate corrected for the number of lamprey removed from the stream during mark-recapture was run. This version was run to correct for the fact that removed lamprey are less likely to be among the contributing parents in the stream, and thus may not be represented in $N_{b}$ and $\widehat{N_{s}}$ estimates. However, the corrected version of the $N_{c}$ estimates also had no correlation with $N_{b}$ and $N_{c}$. Additionally, when a version of the model that corrected for small sample size was evaluated there was still no relationship with $N_{b}$ and $N_{c}$.

The linear models with subsets of environmental, biotic and sampling variables (Table 2.4) consistently found that sampling variables were significant predictors for both $N_{b}$ and $\widehat{N_{s}}$ estimates. In order to minimize missing data, three subsets of models were run. A version was run using variables with no missing data across streams: years since TFM treatment, drainage area, the number of sampling sites, and sample size $(\mathrm{n}=18)$. An additional set added the distance between the river mouth and traps, which only applied to some streams ( $n=12$ ). Finally, a third set of models was run for $\widehat{N_{s}}$ and $V_{k}$ with $N_{c}$ estimates ( $\mathrm{n}=13$ ).

For the LD models, sample size was in the confidence set for both subsets of response variables (Table 2.5 A and 2.5 B ). For the version of the model with only response variables collected for all stream locations, the global model, the number of sampling sites and drainage area were also included in the confidence set (Table 2.5B). Sample size (coefficient $=0.40$ and 0.46 ) and drainage area (coefficient $=0.023$ ) were the only significant predictors across the
models (Figure 2.7). Both were positively associated with LD-based estimates of $N_{b}$. The SF models found that sample size and the number of sampling sites were in the confidence set for both subsets of response variables (Table 2.5A and 2.5B), with sample size (coefficients $=0.54$ and 0.59 ) and the number of sampling sites (coefficient $=37.2$ and 37.92 ) as significant predictors (Figure 2.7), both were positively associated with SF estimates of $N_{b}$. Sample size and the number of sampling sites were both included in the confidence set of all three model subsets tested for the Chao models (Table 2.5A, 2.5B, and 2.5C). Sample size was a significant predictor in all three model subsets (coefficients $=1.05,1.07$, and 1.05 ), and the number of sampling sites was a significant predictor (coefficient $=62.73$ ) in the model subset with the $N_{c}$ estimates (Figure 2.7). The $V_{k}$ models had more variance in confidence sets across model subsets. In the version of the model with only response variables collected for all stream locations, drainage area, the number of sampling sites, and the years since TFM treatment were all included in the confidence set (Table 2.5A). When the distance between mark-recapture traps and the mouth of the river is included in the model, it is included in the confidence set along with drainage area and years since TFM treatment (Table 2.5B). When $N_{c}$ is included as a variable, it is included in the confidence set along with the years since TFM treatment, drainage area, and the number of sampling sites (Table 2.5C). However, none of the models in the confidence set contain significant coefficients.


Figure 2.6. Scatterplots of effective breeding size $\left(N_{b}\right)$, minimum number of spawning adults ( $\widehat{N_{s}}$ ), and census size from mark-recapture $\left(N_{c}\right)$ estimates. $N_{c}$ is shown on the x-axis, the $N_{b}$ or $N_{s}$ estimate is shown on the $y$-axis. No lines of best fit were included due to the lack of significant correlation between variables in the plots.

Table 2.4. Environmental, biotic, and sampling variables used in linear models. Treatment year refers to the most recent TFM treatment that occurred in the stream, $N_{c}$ is the census-size estimate based on mark-recapture for the years 2016, 2017, and 2018, and 'Trap efficiency 2018 ' refers to estimated trap efficiency in 2018 (used to generate $N_{c}$ ). Drainage refers to the drainage area of the stream (in hectares), larval potential is a categorical variable that refers to the history of larval production and TFM treatments in the stream, trap to mouth distance refers to the distance in km between the mouth of the river and the traps used for $N_{c}$ estimates.
Sampling sites refers to the number of collection locations for the larval collections, years since treatments is the number of years between the last TFM treatment and the collection year. Sampling distance refers to the approximate distance sampled in each stream. If only one site was sample 0.2 km was used based on the standard transect distance for backpack electrofishing.

| Lake | Stream | Pop | Treatment Year | Treatment <br> Month | $\begin{aligned} & \mathrm{Nc} \\ & 2016 \end{aligned}$ | $\begin{aligned} & \mathrm{Nc} \\ & 2017 \end{aligned}$ | $\begin{aligned} & \mathrm{Nc} \\ & 2018 \end{aligned}$ | Trap <br> Efficiency <br> 2018 <br> 6 | Drainage | Larval potential | Trap To Mouth Distance | Sampling <br> Sites | Year Since Treat | Sampling <br> Distance |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Superior | Bad | BAD | 2017 | September | 1605 | 5878 | 11301 | 6 | 2270 | 1 | 23 | 1 | 2 | 0.2 |
| Michigan | Betsie | BEI | 2017 | June | 1259 | 984 | 1654 | 61 | 590 | 1 | 14 | 1 | 2 | 0.2 |
| Superior | Betsy | BET | 2017 | July | 396 | 3778 | 1097 | 25 | 230 | 1 | 9 | 2 | 2 | 0.2 |
| Superior | Brule | BRL | 2018 | June | 3194 | 21024 | 36558 | 9 | 408 | 1 | 6 | 1 | 1 | 0.2 |
| Erie | Cattaraugus | CAT | 2016 | May | NA | 5901 | 1637 | 10 | 1129 | 1 | 33 | 4 | 2 | 11.0 |
| Huron | Pigeon | CHE | NA | NA | NA | NA | NA | NA | 1550 | 1 | 2 | 3 | NA | 6.0 |
| Huron | East Au Gres | EAG | 2018 | June | 1846 | 1542 | 2124 | 27 | 653 | 1 | 17 | 2 | 1 | 3.0 |
| Michigan | Ford | FOR | 2017 | May | NA | NA | NA | NA | 1216 | 1 | NA | 1 | 2 | 0.2 |
| Michigan | Manistique | MAI | 2016 | September | 8191 | 6549 | 10420 | 54 | 3631 | 1 | 1 | 1 | 2 | 0.2 |
| Michigan | Manistee | MAN | 2016 | August | 2486 | 2972 | 7219 | 6 | 546 | 1 | 32 | 1 | 2 | 0.2 |
| Superior | Middle | MIR | 2013 | June | 4705 | 4519 | 3113 | 0 | 142 | 1 | 5 | 7 | 3 | 6.0 |
| Superior | Misery | MIS | 2018 | August | 18 | 18 | NA | NA | 102 | 1 | 2 | 1 | 1 | 0.2 |
| Michigan | Muskegon | MUS | 2017 | September | NA | NA | NA | NA | 7327 | 1 | NA | 1 | 2 | 0.2 |
| Huron | Ocqueoc | OCQ | 2016 | July | 6016 | 2539 | 4813 | 70 | 363 | 1 | 4 | 2 | 3 | 3.0 |
| Ontario | Sterling | STE | 2018 | May | NA | 1891 | 2868 | 21 | 80 | 1 | NA | 1 | 1 | 0.2 |
| Michigan | Swan | SWN | 2013 | July | NA | NA | NA | NA | 5 | 2 | NA | 1 | 6 | 0.2 |
| Superior | Tahquamenon | TAQ | 2015 | October | 9465 | 10549 | 3974 | 24 | 2176 | 1 | 16 | 1 | 4 | 0.2 |
| Superior | Two-Hearted | TWO | 2016 | August | NA | NA | NA | NA | 521 | 1 | NA | 1 | 3 | 0.2 |

Table 2.5A. Table for environmental, biotic, and sampling linear models. Significant variables are bolded along with the corresponding coefficient and p-value. In Table 2.5A, the global model consists of the following variables: years since TFM treatment, drainage area, number of sampling sites and sample size.

| Response | Explanatory Variables | Coefficient | p-value | AICc | Akaike Weight |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 易 | Sample Size | 0.4034 | 0.0264 | 196.4321 | 0.3712363 |
|  | Years Since TFM Treatment \| Drainage | Sample Sites | Sample Size | $\begin{gathered} 0.672 \mid \mathbf{0 . 0 2 3} \\ \|-7.741\| \\ 0.605 \end{gathered}$ | $\begin{gathered} 0.9543 \\ \mathbf{0 . 0 1 2 2} \\ 0.7179 \\ 0.0826 \\ \hline \end{gathered}$ | 197.6451 | 0.20241474 |
|  | Sample Sites | 22.4 | 0.0689 | 198.3243 | 0.14413415 |
|  | Drainage | 0.017 | 0.105 | 199.1191 | 0.09686757 |
|  | Sample Size \| Sample Sites |  |  | 199.3055 | 0.08824683 |
|  | Intercept |  |  | 199.6083 | 0.07584604 |
|  | Years Since TFM <br> Treatment |  |  | 202.1526 | 0.02125438 |
|  | Sample Sites | 37.2 | 0.0216 | 206.4399 | 0.42189059 |
|  | Sample Size | 0.5414 | 0.0287 | 207.011 | 0.31709707 |
|  | $\begin{array}{\|c} \hline \text { Sample Sites \| Sample } \\ \text { Size } \\ \hline \end{array}$ | $\begin{gathered} 26.502 \mid \\ 0.1805 \end{gathered}$ | 0.45 \| 0.732 | 209.2815 | 0.1018986 |
|  | Intercept |  |  | 210.023 | 0.07032968 |
|  | Drainage |  |  | 211.169 | 0.0396551 |
|  | Years Since TFM <br> Treatment \| Drainage | <br> Sample Sites \| Sample <br> Size |  |  | 211.7466 | 0.02970713 |
|  | YearSinceTreat |  |  | 212.5966 | 0.01942183 |

Table 2.5A (cont'd).

| Response | Explanatory Variables | Coefficient | p-value | AICc | Aikike Weight |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Sample Size | 1.047 | 0.0005 | 209.0347 | 0.724473908 |
|  | Sample Size \| Sample <br> Sites | 1.072 \|-1.783 | $\begin{gathered} 0.076 \\ 0.962 \end{gathered}$ | 212.0209 | 0.162772636 |
|  | Sample Sites |  |  | 212.9807 | 0.100731972 |
|  | Years Since TFM Treatment $\mid$ Drainage $\mid$ Sample Sites \| Sample Size |  |  | 217.9207 | 0.00852031 |
|  | Intercept |  |  | 220.5784 | 0.002256035 |
|  | Years Since TFM Treatment |  |  | 223.1463 | 0.000624782 |
|  | Drainage |  |  | 223.1605 | 0.000620357 |
| $\pm$ | Intercept | 33.3 | 0.122 | 202.0964 | 0.421149264 |
|  | Drainage | -0.009 | 0.433 | 203.9984 | 0.162716566 |
|  | Sample Sites | -6.272 | 0.653 | 204.4507 | 0.129781287 |
|  | Years Since TFM Treatment | -7.511 | 0.666 | 204.4682 | 0.12865016 |
|  | Sample Size |  |  | 204.6794 | 0.11575345 |
|  | $\begin{gathered} \text { Sample Size \| Sample } \\ \text { Sites } \\ \hline \end{gathered}$ |  |  | 206.7859 | 0.040375772 |
|  | Years Since TFM Treatment $\mid$ Drainage $\mid$ Sample Sites $\mid$ Sample Size |  |  | 213.2758 | 0.001573502 |

Table 2.5B. Table for environmental, biotic, and sampling glm models. Significant variables are bolded along with the corresponding coefficient and p-value. In Table 2.5B, the global model consists of the following variables: years since TFM treatment, drainage area, number of sampling sites, sample size, and distance from the mouth of the river to the mark-recapture trap site.

| Response | Explanatory Variables | Coefficient | p-value | AICc | Akaike Weight |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 首 | Sample Size | 0.4597 | 0.001 | 129.3633 | 0.704144855 |
|  | Sample Size \| Sample Sites |  |  | 131.5453 | 0.236509564 |
|  | Sample Sites |  |  | 135.2397 | 0.037293206 |
|  | Drainage |  |  | 138.8017 | 0.006282662 |
|  | Years Since TFM <br> Treatment |  |  | 138.8339 | 0.006182494 |
|  | intercept |  |  | 138.9071 | 0.005960196 |
|  | Trap To Mouth Distance |  |  | 141.6403 | 0.001519708 |
|  | environmental |  |  | 142.2926 | 0.001096781 |
|  | Years Since TFM Treatment \| Drainage | Sample Sites | Sample Size | Trap to River Mouth Distance |  |  | 142.4564 | 0.001010534 |
|  | Sample Size | 0.5866 | 0.0106 | 142.6609 | 0.4623211 |
|  | Sample Sites | 37.92 | 0.0138 | 143.256 | 0.3433462 |
|  | Sample Size \| Sample Sites | $\begin{gathered} 0.3903 \\ 14.28 \end{gathered}$ | $\begin{gathered} 0.434 \\ 0.662 \end{gathered}$ | 146.0585 | 0.08456122 |
|  | Drainage |  |  | 147.1802 | 0.04826085 |
|  | Intercept |  |  | 147.9442 | 0.03293873 |
|  | Years Since TFM Treatment |  |  | 149.9014 | 0.01237913 |
|  | environmental |  |  | 150.7518 | 0.008091478 |
|  | Trap To Mouth Distance |  |  | 150.7556 | 0.008076147 |
|  | Years Since TFM Treatment \| Drainage | Sample Sites | Sample Size | Trap to River Mouth Distance |  |  | 162.295 | 2.52032E-05 |

Table 2.5B (cont'd).

| Response | Explanatory Variables | Coefficient | p-value | AICc | Aikike Weight |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Sample Size | 1.0693 | 0.001 | 209.0347 | 0.724473908 |
|  | Sample Size \| Sample Sites | $\begin{aligned} & 1.2398 \mid \\ & 12.4086 \end{aligned}$ | $\begin{gathered} 0.0764 \mid \\ 0.7697 \\ \hline \end{gathered}$ | 212.0209 | 0.162772636 |
|  | Sample Sites |  |  | 153.1286 | 0.08926541 |
|  | Drainage |  |  | 158.44 | 0.006270724 |
|  | Intercept |  |  | 158.895 | 0.004994732 |
|  | Years Since TFM Treatment |  |  | 160.0882 | 0.002750579 |
|  | Trap To Mouth Distance |  |  | 161.7535 | 0.001196181 |
|  | environmental |  |  | 162.0578 | 0.00102735 |
|  | Years Since TFM Treatment \| Drainage | Sample Sites | Sample Size | Trap to River Mouth Distance |  |  | 167.5806 | $6.49327 \mathrm{E}-05$ |
| $\pm$ | Trap To Mouth Distance | 1.3573 | 0.11 | 101.7188 | 0.3189623 |
|  | Intercept | 10.053 | 0.0378 | 102.0076 | 0.2760846 |
|  | Drainage | 0.0037 | 0.379 | 103.9653 | 0.1037338 |
|  | environmental | $\begin{gathered} 0.0034 \mid \\ 0.6279 \\ \hline \end{gathered}$ | $\begin{gathered} 0.381 \\ 0.124 \\ \hline \end{gathered}$ | 104.3042 | 0.08756584 |
|  | Years Since TFM Treatment | -1.984 | 0.708 | 104.7636 | 0.06959273 |
|  | Sample Sites |  |  | 104.8159 | 0.06779892 |
|  | Sample Size |  |  | 104.9146 | 0.06453247 |
|  | Sample Size \| Sample Sites |  |  | 108.3277 | 0.01171205 |
|  | Years Since TFM Treatment \| Drainage | Sample Sites | Sample Size | Trap to River Mouth Distance |  |  | 121.3673 | $1.72637 \mathrm{E}-05$ |

Table 2.5C. Table for environmental, biotic, and sampling glm models. Significant variables are bolded along with the corresponding coefficient and p-value. In Table 2.5C, the global model consists of the following variables: years since TFM treatment, drainage area, number of sampling sites, sample size, and Nc estimates. In Table 2.5C, only $N_{s}$ - Chao and $V_{k}$ were considered as response variables.

| Response | Explanatory Variables | Coefficient | p-value | AICe | Akaike Weight |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Sample Size | 1.0516 | 0.002 | 150.5441 | 0.6762677 |
|  | Sample Sites | 62.73 | 0.008 | 153.1635 | 0.1825183 |
|  | Sample Size \| Sample Sites |  |  | 154.1869 | 0.1094195 |
|  | Drainage |  |  | 158.5344 | 0.01244596 |
|  | Intercept |  |  | 158.9242 | 0.010242 |
|  | Years Since TFM Treatment |  |  | 160.1061 | 0.00567209 |
|  | $N_{c}$ Estimate |  |  | 161.8385 | 0.002385413 |
|  | environmental |  |  | 163.6739 | 0.0009528 |
|  | Years Since TFM Treatment \| Drainage | Sample Sites | Sample Size | Nc estimates |  |  | 168.2597 | 9.62108E-05 |
| $\pm$ | Intercept | 37.58 | 0.214 | 147.6474 | 0.3422417 |
|  | Years Since TFM Treatment | -42.11 | 0.218 | 148.6694 | 0.2053103 |
|  | Drainage | 0.019 | 0.498 | 150.0009 | 0.1055069 |
|  | $N_{c}$ Estimate | -0.002 | 0.622 | 150.2749 | 0.09200045 |
|  | Sample Sites | -8.344 | 0.635 | 150.2974 | 0.09097165 |
|  | Sample Size |  |  | 150.5627 | 0.07967152 |
|  | environmental |  |  | 151.0046 | 0.06387491 |
|  | Sample Size \| Sample Sites |  |  | 153.2862 | 0.02041256 |
|  | Years Since TFM Treatment \| Drainage $\mid$ Sample Sites \| Sample Size $\mid$ Nc estimates |  |  | 168.5276 | $1.0006 \mathrm{E}-05$ |



Figure 2.7. Plots of significant predictors of $N_{b}$ and $\widehat{N_{s}}$ estimates based on the results of the environmental models.

## Effects of Sample Size, SNP set size, and stream $N_{b}$ on genetic estimates

Rmetasim simulated populations with true $N_{b}$ values within $7 \%$ of the input $N_{b}$ value after 50 generations of burn-in. True $N_{s}$, or the total number of parents in the population, were within $6 \%$ of other replicates.

Point estimates of $N_{b}$ across methods were accurate for populations with small true $N_{b}$ ( $N_{b}$ $<1000$ ), but accuracy varied when $N_{b}$ was large. Particularly when sample size or SNP size were small, the PwoP and SF $N_{b}$ estimates had a downward bias compared to true $N_{b}$, where the estimated Nb values do not increase as true Nb increases (Figure 2.8A,2.8C). Conversely, the LD estimate had an upward bias at some scenarios, although the bias is smaller than the other two estimates (Figure 2.8B). The LD method had more variation in point estimates across replicates than the other two $N_{b}$ estimation methods (Figure 2.9A-C). RMSE values were generally higher when sample size was small across methods (Figure $2.10 \mathrm{~A}-\mathrm{C}$ ), indicating that small sample size decreased precision in estimates across $N_{b}$ methods.

When $N_{b}$ is small, the Chao and Jackknife method estimates performed similarly when compared to true $N_{s}$ (Figure 2.9D-E). However, as $N_{b}$ increased, the Jackknife method had a larger downward bias compared to the Chao method across sample size and SNP groups (Figure 2.9D-E). The Chao method shows a downward bias as well when SNP set size or sample size are small for large true $N_{b}$ populations (Figure 2.9D). For both the Chao and the Jackknife method, RMSE increased with true $N_{b}$ of populations, indicating that variation in estimated $\widehat{N_{s}}$ increases as the true $N_{b}$ of the population increases (Figure 2.10D-E). When $N_{b}$ is greater than 1000, RMSE values are similarly high across SNP size and sample size for the Jackknife method, but in the Chao method increasing the number of SNPs decreases RMSE (Figure 2.10D-E).


Figure 2.8A. A figure that visualizes the ratio between estimated $N_{b}$ and the true $N_{b}$ estimate from each simulation. The sample size parameter is on the $x$-axis, the SNP set size is separated by color, and the plots are subset by the effective breeding size parameter. Figure 2.8 A is the sibship frequency method.


Figure 2.8B. A figure that visualizes the ratio between estimated $N_{b}$ and the true $N_{b}$ estimate from each simulation. The sample size parameter is on the x -axis, the SNP set size is separated by color, and the plots are subset by the effective breeding size parameter. Figure 8B is the linkage disequilibrium method.


Figure 2.8 C . A figure that visualizes the ratio between estimated $N_{b}$ and the true $N_{b}$ estimate from each simulation. The sample size parameter is on the $x$-axis, the SNP set size is separated by color, and the plots are subset by the effective breeding size parameter. Figure 2.8 C is the parents without parents methods.


Figure 2.9A. Plots (plot 2.9A-2.9E) to show accuracy of point estimates for simulated populations. The x-axes are $\log _{10}$ of the parameter effective breeding size $\left(N_{b}\right)$, and the y -axes are $\log _{10}$ of the estimated $N_{b}$ or the minimum number of spawning adults $\left(N_{s}\right)$. The plots are subset by SNP set size and sample size, and figures are separated by each method. Figure 2.9A shows results from the sibship frequency estimates.


Figure 2.9B. Plots (plot 2.9A-2.9E) to show accuracy of point estimates for simulated populations. The x-axes are $\log _{10}$ of the parameter effective breeding size $\left(N_{b}\right)$, and the y -axes are $\log _{10}$ of the estimated $N_{b}$ or the minimum number of spawning adults $\left(N_{s}\right)$. The plots are subset by SNP set size and sample size, and figures are separated by each method. Figure 2.9B shows results from the linkage disequilibrium estimates.


Figure 2.9C. Plots (plot 2.9A-2.9E) to show accuracy of point estimates for simulated populations. The x -axes are $\log _{10}$ of the parameter effective breeding size $\left(N_{b}\right)$, and the y -axes are $\log _{10}$ of the estimated $N_{b}$ or the minimum number of spawning adults $\left(N_{s}\right)$. The plots are subset by SNP set size and sample size, and figures are separated by each method. Figure 2.9C shows results from the parentage without parents estimates.


$$
S N P=500
$$















Figure 2.9D. Plots (plot 2.9A-2.9E) to show accuracy of point estimates for simulated populations. The x -axes are $\log _{10}$ of the parameter effective breeding size $\left(N_{b}\right)$, and the y -axes are $\log _{10}$ of the estimated $N_{b}$ or the minimum number of spawning adults $\left(N_{s}\right)$. The plots are subset by SNP set size and sample size, and figures are separated by each method. Figure 2.9D shows results from the chao estimates.


Figure 2.9E. Plots (plot 2.9A-2.9E) to show accuracy of point estimates for simulated populations. The x-axes are $\log _{10}$ of the parameter effective breeding size $\left(N_{b}\right)$, and the y -axes are $\log _{10}$ of the estimated $N_{b}$ or the minimum number of spawning adults $\left(N_{s}\right)$. The plots are subset by SNP set size and sample size, and figures are separated by each method. Figure 2.9E shows results from the jackknife estimates.


Figure 2.10A. Root mean squared error (RMSE) plots (plots 2.10A-2.10E) for each type of estimate to show the variance in point estimates for simulated populations. RMSE (y-axis) is plotted versus the sample size (x-axis). The line colors are the SNP set size, where yellow corresponds to $\mathrm{SNPs}=100$, dark blue corresponds to $\mathrm{SNPs}=500$, and green-grey corresponds to SNPs $=1000$. The plots are subset by parameter effective breeding size ( $N_{b}$ ), and the figures are separated by $N_{b}$ and the minimum number of spawning adults $\left(N_{s}\right)$ estimate method. Figure 2.10A shows results from the sibship frequency estimates.


Figure 2.10B. Root mean squared error (RMSE) plots (plots 2.10A-2.10E) for each type of estimate to show the variance in point estimates for simulated populations. RMSE (y-axis) is plotted versus the sample size ( x -axis). The line colors are the SNP set size, where yellow corresponds to SNPs $=100$, dark blue corresponds to $\mathrm{SNPs}=500$, and green-grey corresponds to SNPs $=1000$. The plots are subset by parameter effective breeding size ( $N_{b}$ ), and the figures are separated by $N_{b}$ and the minimum number of spawning adults $\left(N_{s}\right)$ estimate method. Figure 2.10B shows results from the linkage disequilibrium estimates.


Figure 2.10C. Root mean squared error (RMSE) plots (plots 2.10A-2.10E) for each type of estimate to show the variance in point estimates for simulated populations. RMSE (y-axis) is plotted versus the sample size ( x -axis). The line colors are the SNP set size, where yellow corresponds to SNPs $=100$, dark blue corresponds to $\mathrm{SNPs}=500$, and green-grey corresponds to SNPs $=1000$. The plots are subset by parameter effective breeding size ( $N_{b}$ ), and the figures are separated by $N_{b}$ and the minimum number of spawning adults $\left(N_{s}\right)$ estimate method. Figure 2.10 C shows results from the parentage without parents estimates.


Figure 2.10D. Root mean squared error (RMSE) plots (plots 2.10A-2.10E) for each type of estimate to show the variance in point estimates for simulated populations. RMSE (y-axis) is plotted versus the sample size ( x -axis). The line colors are the SNP set size, where yellow corresponds to SNPs $=100$, dark blue corresponds to $\mathrm{SNPs}=500$, and green-grey corresponds to SNPs $=1000$. The plots are subset by parameter effective breeding size ( $N_{b}$ ), and the figures are separated by $N_{b}$ and the minimum number of spawning adults ( $N_{s}$ ) estimate method. Figure 2.10D shows results from the chao estimates.


Figure 2.10E. Root mean squared error (RMSE) plots (plots 2.10A-2.10E) for each type of estimate to show the variance in point estimates for simulated populations. RMSE (y-axis) is plotted versus the sample size ( x -axis). The line colors are the SNP set size, where yellow corresponds to SNPs $=100$, dark blue corresponds to $\mathrm{SNPs}=500$, and green-grey corresponds to SNPs $=1000$. The plots are subset by parameter effective breeding size ( $N_{b}$ ), and the figures are separated by $N_{b}$ and the minimum number of spawning adults $\left(N_{s}\right)$ estimate method. Figure 2.10E shows results from the jackknife estimates.

## DISCUSSION

Genetic assessment using $N_{b}$ and $N_{s}$ estimates provides unique insights into sea lamprey systems that cannot be obtained from other types of adult assessment. $N_{b}$ and $N_{s}$ provide information on the number of successfully spawning adults in streams, and reconstructed pedigrees show the variation in reproductive success and family size for spawning populations. $N_{b}$ and $N_{s}$ are informative particularly for streams with potential barrier failure or streams where trapping is difficult. Understanding the influences of environmental, biotic, and sampling factors on $N_{b}$ and $N_{s}$ is important but the specific factors influencing these estimates in a given stream system can be difficult to predict. Simulated sea lamprey populations showed that $N_{b}$ and $N_{s}$ estimates can be obtained for ecologically realistic sea lamprey stream populations if SNP sets are sufficiently large (greater than 500 SNPs), which can be obtained using a RAD-capture panel. Additionally, obtaining a sample size of 100 individuals or greater will help to minimize bias in estimates.

The use of $N_{b}$ as a genetic assessment technique has been well-documented in the literature. $N_{b}$ has previously been used to evaluate rates of inbreeding and genetic diversity for threatened and endangered species (Duong, Scribner, Forsythe, Crossman, \& Baker, 2013; Waller \& Keller, 2002). It has also been used to assess conservation management actions like stocking (Kazyak, Rash, Lubinski, \& King, 2018; Petereit et al., 2018b) and genetic rescue (Hedrick, Peterson, Vucetich, Adams, \& Vucetich, 2014). Genetic estimates are used both as a primary assessment technique as well as a supplementary tool paired with assessment techniques like mark-recapture estimates of adult abundance. In this study, the utility of $N_{b}$ and $N_{s}$ as an assessment tool is illustrated through their estimation in eighteen great lakes tributaries.

However, obtaining representative and sufficient samples in a stream, as well as accurately separating those samples into their respective cohorts, is crucial for obtaining accurate estimates.

## Simulation Recommendations

Simulated populations showed that genomic data sets provide the power necessary to estimate $N_{b}$ and $N_{s}$ even when true $N_{b}$ is large. However, estimates based on the reconstructed pedigrees underestimate $N_{b}$ and $N_{s}$ when true $N_{b}>500$, particularly if the SNP set used is too small $(\mathrm{n}=100)$ regardless of sample size (Figure 2.9A,C). In contrast the average LD estimate is closer to true $N_{b}$ when SNP size is low compared to methods generated with a reconstructed pedigree, but the variation in estimates is much greater (Figure 2.9B). The results of our simulation study are consistent with previous simulations studying the biases of LD that showed a small upward bias with large true $N_{e}$ (Waples, 2016), that SF can have a downward bias especially when true $N_{e}$ is large (Wang, 2016), and other $N_{b}$ method comparison simulation studies highlighting better precision for estimates with small true $N_{e}$ (Robinson \& Moyer, 2013). The bias in $N_{s}$ estimates is consistent with known limitations from sample size, since $N_{s}$ is directly based on the number of parent genotypes. Additionally, a small SNP set may lead to more falsely inferred sibling relationships, leading to a downward bias of extrapolated estimates compared to the true $N_{s}$ value. However, when true $N_{b}$ is large, the Chao estimator outperforms the Jackknife, suggesting that the Chao estimator is the better tool to use particularly for large populations. Our simulations are informative for future attempts to sample for $N_{b}$ and $N_{s}$ estimates. If expected $N_{b}$ is large, obtaining a greater number of sampled offspring and generating sufficient numbers of SNPs should be prioritized to ensure minimal bias in estimates.

## Reconstructed Pedigrees and Genetic Estimates

Reconstructed pedigrees were generated for a variety of streams in the Great Lakes region, and they provide a unique look into the diversity of family structure across sea lamprey larval populations in the region. Locations range from a small number of full-sibling families, to groups of mostly unrelated individuals, to large interconnected populations of half-sibling families (Figure 2.4). This variety shows that sea lamprey spawning dynamics are highly variable among lamprey-producing streams across the region, highlighting the importance of evaluating lamprey spawning populations on a per-stream basis.

Accurate separation of cohorts is vitally important for estimation of $N_{b}$ and $N_{s}$. Reconstructed pedigree data, namely Colony clusters, can be used to evaluate individual cohort assignments generated from mixture models and larval length data in semelparous species like sea lamprey. When two individuals were assigned to different cohorts on the basis of length, but connected in the pedigree, we reclassified individuals into cohorts associated with their full- or half-siblings. However, corrections in the opposite direction, where the mixture models indicate a singular cluster when multiple clusters are suspected from patterns in length and/or genomic data, are not possible given this approach (Figure 1.2). This could be the case for the Ocqueoc River, where two clear modes are present in the length-frequency histogram (Figure 2.2) and boxplots of Colony cluster lengths (Figure 2.3) that there are two groups of unrelated individuals with a length cutoff of approximately 37 mm . However, the mixture analysis results indicated one cohort, so the groups were not separated for subsequent analyses. If individuals from multiple cohorts are combined, bias can be introduced into estimates of $N_{b}$ and $N_{s}$. For example, the full population of the Ocqueoc produced an SF $N_{b}$ estimate of 184 and an LD $N_{b}$ estimate of 134.3 but given patterns in the length-frequency histograms and the distributions of lengths of
individuals connected in the reconstructed pedigree for this population, it seems possible that these estimates combine individuals from two cohorts (i.e., 2018 and 2019 year classes). If the two cohorts are separated, the $2019 N_{b}$ estimates (length < 37mm) are 51 (SF) and 33.7 (LD), while the $2018 N_{b}$ estimates are $134(\mathrm{SF})$ and 104.2 (LD). Future research will evaluate other methods for aging lamprey using genomic techniques, or improvements to the mixture analysis that would reduce uncertainty in situations like that seen in the Ocqueoc River.

## $N_{b}$ and $N_{c}$ relationship and Sampling Effects

The relationship between $N_{b}$ and $N_{c}$ estimates is dependent on a large number of sampling and environmental factors, and obtaining a model for that relationship remains difficult. Our models primarily showed the influence of sample size and the number of sampling sites as influences on $N_{b}$ and $N_{s}$ estimates. $N_{b}$ is influenced by many factors that vary across reproductive events, including variation in reproductive success, skewed sex ratios, and the fecundity of spawning individuals. Previous work estimating $N_{c}$ considered variables similar to the environmental data used in the generalized linear models in the study (Mullett et al., 2003), and thus the lack of significance of those environmental variables conflicted with some of our expectations. For example, streams recently treated with TFM should have lower adult recruitment the following year due to weaker larval cue, leading to lower $N_{b}$ and $N_{s}$ estimates, but there was not a significant relationship between our estimates and years since TFM treatment. However, variables like drainage area did have the expected positive relationship with $N_{b}$ using the LD method.

When $N_{c}$ is large and the representative sequenced larvae have significant family structure coupled with small sampled size, there is some concern that estimates may be
representative of a small group of families rather than the full spawning population in a stream (Whiteley et al., 2012). In three of our stream locations, $N_{c}$ was greater than 10,000 , the sampled offspring group was less than 30 , and the sample was collected from a single site in the stream. In these systems, non-random sampling may be leading to downward bias in these estimates. The correlations between $N_{b}, N_{s}$ and $N_{c}$ were conducted with and without these locations and the relationships remained nonsignificant. There is not a universal solution to remove the bias that could exist from family effect, especially since representative family structure is necessary to calculate both $N_{b}$ and $N_{s}$ (Waples \& Anderson, 2017). To combat this potential bias, sampling multiple locations in a stream that are spread across larval habitat in the stream can minimize potential family effect bias and ensure that sampling is representative of the true spawning event. In addition to family effect, there are sources of uncertainty present in the $N_{c}$ estimates that could have further complicated models involving those estimates. Low trap efficiency and variation in trap efficiency across years and index streams, as well as variation in catchability for individual lamprey all contribute to uncertainty in $N_{c}$ (Harper et al., 2018b).

Across systems, $N_{b}$ and $N_{c}$ have not necessarily had a correlative relationship (Bernos \& Fraser, 2016; Whiteley et al., 2015), especially when the population size is large (Waples, 2016). However, some studies have found a relationship when environmental factors and population dynamics could be adjusted for to account for variation in the $N_{e}: N_{c}$ ratio (Ruzzante et al., 2016). In particular, the relationship between sufficient and representative sampling was highlighted (Whiteley et al., 2012), which is consistent with the modeling results found in our study. When the sample size and sampling distance is small, there is the possibility of non-random sampling leading to $N_{b}$ and $N_{s}$ estimates that represent the small number of families in the sample rather than the full stream population, known as the family effect. Additional environmental variables
like the amount of spawning habitat, density of spawning adults and stream flow during spawning could also affect $N_{b}$ and $N_{s}$ estimates and were not included in models for this study (Whiteley et al., 2015). A further potential complication is genetic compensation, which is when variation in reproductive success decreases in small populations, inflating $N_{b}$ estimates compared to $N_{c}$ (Ardren \& Kapuscinski, 2003; Whiteley et al., 2015). The lack of significance in $N_{b}$ and $N_{c}$ estimates in this study, as well as the significant relationship between sampling factors and $N_{b}$ estimates in our models, underscores the need for large and representative sampling when estimating $N_{b}$ from population genomic data.

## Applications in Management

$N_{b}$ and $N_{s}$ can be used to provide additional information about sea lamprey spawning systems as an augmentative annual assessment technique. If an index group of streams are assessed over a larger number of years, families and cohorts can be evaluated as an annual larval and adult assessment technique, which can be utilized to assess larval growth rates and larval dispersal in streams. $N_{b}$ and $N_{s}$ can also be used to evaluate the efficacy and effectiveness of supplemental control techniques like sterile male release and repellant/attractants in push-pull configurations to increase trapping efficiency, and the use of alarm cue as a barrier technique. Additionally, $N_{b}$ provides information on inbreeding, drift, and loss of diversity that is present in the population, all of which can be used to further evaluate control techniques. $N_{s}$ estimates also can provide an annual metric of the minimum number of successfully spawning adults, which could be used as an annual assessment metric for the amount of successful spawning that occurs in streams across the region.

While sea lamprey are one of the most destructive species in the Great Lakes region, the species is under threat in parts of its native range, namely in the Eastern Atlantic. Genetic assessment techniques like $N_{b}$ and $N_{s}$ can be utilized for both control and conservation efforts for sea lamprey and other species. Although connections between $N_{b}$ and $N_{c}$ are complicated by a variety of factors, genetic estimates provide a unique look into the genetic structure of a population that can aid in monitoring efforts to conserve or control a species.

APPENDIX

## APPENDIX

## Table of Acronyms

RAD: Restriction-site Associated DNA

MAF: Minor Allele Frequency
$V_{k}$ : The variance in family size for adults represented in sampling
$N_{e}$ : Effective population size
$N_{b}$ : Effective breeding size
$N_{c}$ : Census size estimate based on mark-recapture trapping methods
$N_{s}$ : The minimum number of spawning adults
$\widehat{N_{s}}$ : The minimum number of spawning adults extrapolated using a 'pedigree reconstruction curve'

RMSE: Root mean squared error
LD: $N_{b}$ estimate using the linkage disequilibrium method
SF: $N_{b}$ estimate using the sibship frequency method
PwoP: $N_{b}$ estimate using the parentage without parents method
SNP: Single nucleotide polymorphism
TFM: 3-triflouromethyl-4-nitrophenol, a selective lampricide
AIC: Akaike information criterion
AICc: Akaike information criterion, corrected for small sample size

LITERATURE CITED

## LITERATURE CITED

Ackerman, M. W., Hand, B. K., Waples, R. K., Luikart, G., Waples, R. S., Steele, C. A., ... Campbell, M. R. (2017). Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. Evolutionary Applications, 10(2), 146160. https://doi.org/10.1111/eva. 12433

Akaike, H. (1974). A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 19(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705

Akaike, H. (1978). On the Likelihood of a Time Series Model, 27(3), 217-235.
Ali, O. A., O’Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., \& Miller, M. R. (2016). Rad capture (Rapture): Flexible and efficient sequence-based genotyping. Genetics, 202(2), 389-400. https://doi.org/10.1534/genetics.115.183665

Applegate, V. C. (1950). Natural history of the sea lamprey, Petromyzon marinus in Michigan. Spec Sci Rep US Fish Wildl Serv, 55, 1-237. Retrieved from http://ci.nii.ac.jp/naid/10010684036/en/

Ardren, W. R., \& Kapuscinski, A. R. (2003). Demographic and genetic estimates of effective population size (Ne) reveals genetic compensation in steelhead trout. Molecular Ecology, 12(1), 35-49. https://doi.org/10.1046/j.1365-294X.2003.01705.x

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE, 3(10), e3376. https://doi.org/10.1371/journal.pone. 0003376

Barber, J., \& Steeves, M. (2019). Sea lamprey control the Great Lakes 2018: Annual report to the Great Lakes Fishery Commission. Detroit, Michigan. Retrieved from http://www.glfc.org/pubs/slcp/annual_reports/ANNUAL_REPORT_2018.pdf

Bernos, T. A., \& Fraser, D. J. (2016). Spatiotemporal relationship between adult census size and genetic population size across a wide population size gradient. Molecular Ecology, 25(18), 4472-4487. https://doi.org/10.1111/mec. 13790

Binder, T. R., \& McDonald, D. G. (2008). The role of temperature in controlling diel activity in upstream migrant sea lampreys (Petromyzon marinus). Canadian Journal of Fisheries and Aquatic Sciences, 65(6), 1113-1121. https://doi.org/10.1139/F08-070

Bolger, A. M., Lohse, M., \& Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.
https://doi.org/10.1093/bioinformatics/btu170

Bravener, G. A., \& McLaughlin, R. L. (2013). A behavioural framework for trapping success and its application to invasive sea lamprey. Canadian Journal of Fisheries and Aquatic Sciences, 70(10), 1438-1446. https://doi.org/10.1139/cjfas-2012-0473

Catchen, J. M., Hohenlohe, P. A., Bassham, S., Amores, A., \& Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology, 22(11), 3124-3140. https://doi.org/10.1111/mec. 12354

Chao, A. (1987). Estimating the Population Size for Capture-Recapture Data with Unequal Catchability. Biometrics, 43(4), 783-791. https://doi.org/10.4081/cp.2017.979

Christie, G. C., Adams, J. V., Steeves, T. B., Slade, J. W., Cuddy, D. W., Fodale, M. F., ... Jones, M. L. (2003). Selecting Great Lakes Streams for Lampricide Treatment Based On Larval Sea Lamprey Surveys. Journal of Great Lakes Research, 29, 152-160. https://doi.org/https://doi.org/10.1016/S0380-1330(03)70484-5

Christie, G. C., \& Goddard, C. I. (2003). Sea Lamprey International Symposium (SLIS II): Advances in the integrated management of Sea Lamprey in the Great Lakes. Journal of Great Lakes Research, 29(SUPPL. 1), 1-14. https://doi.org/10.1016/S0380-1330(03)70474-2

Dawson, H. A., Higgins-weier, C. E., Steeves, T. B., \& Johnson, N. S. (2020). Estimating age and growth of invasive sea lamprey : A review of approaches and investigation of a new method. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2020.06.002

Dawson, H. A., Jones, M. L., Scribner, K. T., \& Gilmore, S. A. (2009). An Assessment of Age Determination Methods for Great Lakes Larval Sea Lampreys. North American Journal of Fisheries Management, 29(4), 914-927. https://doi.org/10.1577/m08-139.1

Dawson, H. A., Quintella, B. R., Almeida, P. R., Treble, A. J., \& Jolley, J. C. (2015). The Ecology of Larval and Metamorphosing Lampreys. In Lampreys: Biology, Conservation and Control (Vol. 1, pp. 75-137). https://doi.org/10.1007/978-94-017-9306-3

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, J. B., \& Ovenden, J. R. (2013). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources, 14(1), 209-214. https://doi.org/doi:10.1111/1755-0998.12157

Duong, T. Y., Scribner, K. T., Forsythe, P. S., Crossman, J. A., \& Baker, E. A. (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(5), 1282-1294. https://doi.org/10.1111/mec. 12167

Grün, B., \& Leisch, F. (2010). BayesMix: an R package for Bayesian mixture modeling. Technique Report, 1-11.

Hardisty, M. W., \& Potter, I. C. (1971). The Biology Of Lampreys. In Academic Press (Vol. 1). New York. https://doi.org/10.1126/science.176.4042.1409

Harper, D. L. M., Horrocks, J., Barber, J., Bravener, G. A., Schwarz, C. J., \& McLaughlin, R. L. (2018). An evaluation of statistical methods for estimating abundances of migrating adult sea lamprey. Journal of Great Lakes Research, 44(6), 1362-1372.
https://doi.org/https://doi.org/10.1016/j.jglr.2018.08.004
Hedrick, P. W., Peterson, R. O., Vucetich, L. M., Adams, J. R., \& Vucetich, J. A. (2014). Genetic rescue in Isle Royale wolves: genetic analysis and the collapse of the population. Conservation Genetics, 15(5), 1111-1121. https://doi.org/10.1007/s10592-014-0604-1

Heltshe, J. F., \& Forrester, N. E. (2009). Estimating Species Richness Using the Jackknife Procedure Published by : International Biometric Society Stable URL: http://www.jstor.org/stable/2530802. Society, 39(1), 1-11.

Hume, J. B., Meckley, T. D., Johnson, N. S., Luhring, T. M., Siefkes, M. J., \& Wagner, C. M. (2015). Application of a putative alarm cue hastens the arrival of invasive sea lamprey (Petromyzon marinus) at a trapping location. Canadian Journal of Fisheries and Aquatic Sciences, 72(12), 1799-1806. https://doi.org/10.2139/ssrn. 2793943

Hurvich, C. M., \& Tsai, C. L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307. https://doi.org/10.1093/biomet/76.2.297

Israel, J. A., \& May, B. (2010). Indirect genetic estimates of breeding population size in the polyploid green sturgeon (Acipenser medirostris). Molecular Ecology, 19(5), 1058-1070. https://doi.org/10.1111/j.1365-294X.2010.04533.x

Jones, A. T., Ovenden, J. R., \& Wang, Y. G. (2016). Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. Heredity, 117(4), 217-223. https://doi.org/10.1038/hdy.2016.19

Jones, O. R., \& Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources, 10(3), 551-555.
https://doi.org/doi:10.1111/j.1755-0998.2009.02787.x
Kaye, C. A., Heinrich, J. W., Hanson, L. H., McDonald, R. B., Slade, J. W., Genovese, J. H., \& Swink, W. D. (2003). Evaluation of Strategies for the Release of Male Sea Lampreys (Petromyzon marinus) in Lake Superior for a Proposed Sterile-Male-Release Program. Journal of Great Lakes Research, 29, 424-434. https://doi.org/https://doi.org/10.1016/S0380-1330(03)70505-X

Kazyak, D. C., Rash, J., Lubinski, B. A., \& King, T. L. (2018). Assessing the impact of stocking northern-origin hatchery brook trout on the genetics of wild populations in North Carolina. Conservation Genetics, 19(1), 207-219. https://doi.org/10.1007/s10592-017-1037-4

Lavis, D. S., Hallett, A., Koon, E. M., \& McAuley, T. C. (2003). History of and advances in barriers as an alternative method to suppress sea lampreys in the Great Lakes. Journal of Great Lakes Research, 29(SUPPL. 1), 362-372. https://doi.org/10.1016/S0380-1330(03)70500-0

Lawrie, A. H. (1970). The Sea Lamprey in the Great Lakes. Transactions of the American Fisheries Society, 99(4), 766-775. https://doi.org/10.1577/15488659(1970)99<766:TSLITG>2.0.CO;2

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 00(00), 1-3. Retrieved from http://arxiv.org/abs/1303.3997

Li, H., \& Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics, 26(5), 589-595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Manion, P. J., \& Smith, B. R. (1978). Biology of larval and metamorphosing sea lampreys, Petromyzon marinus, of the 1960 year class in the Big Garlic River, Michigan, Part II, 1966-1972. Great Lakes Fishery Commission Technical Report, 30, 1-37.

McDonald, D. G., \& Kolar, C. S. (2007). Research to Guide the Use of Lampricides for Controlling Sea Lamprey. Journal of Great Lakes Research, 33, 20-34. https://doi.org/https://doi.org/10.3394/0380-1330(2007)33[20:RTGTUO]2.0.CO;2

McKinney, G. J., Waples, R. K., Seeb, L. W., \& Seeb, J. E. (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(4), 656-669. https://doi.org/10.1111/1755-0998.12613

McLaughlin, R. L., Hallett, A., Pratt, T. C., O’Connor, L. M., \& McDonald, D. G. (2007). Research to Guide Use of Barriers, Traps, and Fishways to Control Sea Lamprey. Journal of Great Lakes Research, 33(2), 7-19. https://doi.org/10.3394/0380-1330(2007)33

Mohammadi, A., Salehi-Rad, M. R., \& Wit, E. C. (2013). Using mixture of Gamma distributions for Bayesian analysis in an $\mathrm{M} / \mathrm{G} / 1$ queue with optional second service. Computational Statistics, 28(2), 683-700. https://doi.org/10.1007/s00180-012-0323-3

Morkert, S. B., Swink, W. D., \& Seelye, J. G. (1998). Evidence for Early Metamorphosis of Sea Lampreys in the Chippewa River, Michigan. North American Journal of Fisheries Management, 18(4), 966-971. https://doi.org/10.1577/1548$\underline{8675(1998) 018<0966: \text { efemos }>2.0 . c o ; 2}$

Mullett, K. M., Heinrich, J. W., Adams, J. V., Young, R. J., Henson, M. P., McDonald, R. B., \& Fodale, M. F. (2003). Estimating Lake-wide Abundance of Spawning-phase Sea Lampreys (Petromyzon marinus) in the Great Lakes: Extrapolating from Sampled Streams Using Regression Models. Journal of Great Lakes Research, 29, 240-252. https://doi.org/10.1016/S0380-1330(03)70492-4

Mullett, K. M., \& Sullivan, P. (2017). Sea Lamprey Control the Great Lakes 2016: Annual Report to the Great Lakes Fishery Commission. Duluth, Minnesota.

Nasserinejad, K., Rosmalen, J. Van, De Kort, W., \& Lesaffre, E. (2017). Comparison of criteria for choosing the number of classes in Bayesian finite mixture models. PLoS ONE, 12(1), 1-23. https://doi.org/10.1371/journal.pone. 0168838

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., ... Szoecs, E. (2019). Package ' vegan .'

Petereit, C., Bekkevold, D., Nickel, S., Dierking, J., Hantke, H., Hahn, A., ... Puebla, O. (2018). Population genetic structure after 125 years of stocking in sea trout (Salmo trutta L.). Conservation Genetics, 19(5), 1123-1136. https://doi.org/10.1007/s10592-018-1083-6

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R foundation for Statistical Computing.

Rawding, D. J., Sharpe, C. S., \& Blankenship, S. M. (2014). Genetic-Based Estimates of Adult Chinook Salmon Spawner Abundance from Carcass Surveys and Juvenile Out-Migrant Traps. Transactions of the American Fisheries Society, 143(1), 55-67. https://doi.org/10.1080/00028487.2013.829122

Renaud, C. (2011). Lampreys of the World: an Annotated and Illustrated Catalogue of Lamprey Species Known To Date. FAO Species Catalogue for Fishery Purposes (Vol. 5).

Robinson, J. D., \& Moyer, G. R. (2013). Linkage disequilibrium and effective population size when generations overlap. Evolutionary Applications, 6(2), 290-302.
https://doi.org/10.1111/j.1752-4571.2012.00289.x
Rousseau, J., \& Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 73(5), 689-710. https://doi.org/10.1111/j.1467-9868.2011.00781.x

Ruzzante, D. E., McCracken, G. R., Parmelee, S., Hill, K., Corrigan, A., MacMillan, J., \& Walde, S. J. (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(1823).
https://doi.org/10.1098/rspb.2015.2601

Sard, N. M., Hunter, R. D., Roseman, E. F., Hayes, D. B., DeBruyne, R. L., \& Scribner, K. T. (n.d.). Extending non-parametric species richness estimators to genetic pedigree rarefaction for breeding adult estimation. In press.

Sard, N. M., Smith, S. R., Homola, J. J., Kanefsky, J., Bravener, G., Adams, J. V., ... Scribner, K. T. (2020). RAPTURE (RAD capture) panel facilitates analyses characterizing sea lamprey reproductive ecology and movement dynamics. Ecology and Evolution, (December 2019), 1-20. https://doi.org/10.1002/ece3.6001

Smith, B. R., \& Tibbles, J. J. (1980). Sea Lamprey (Petromyzon marinus) in Lakes Huron, Michigan, and Superior: History of Invasion and Control, 1936-78. Canadian Journal of Fisheries and Aquatic Sciences, 37(37), 1780-1801. Retrieved from www.nrcresearchpress.com

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., ... Li, W. (2013). Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nature Genetics, 45(4), 415-421. https://doi.org/10.1038/ng. 2568

Smith, J. J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M. C., Parker, H. J., ... Amemiya, C. T. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. Nature Genetics, 50(2), 270-277. https://doi.org/10.1038/s41588-017-0036-1

Steeves, M., \& Barber, J. (2020). Sea Lamprey Control in the Great Lakes 2019.
Strand, A. E. (2002). Metasim 1.0: an individual-based environment for simulating population genetics of complex population dynamics. Molecular Ecology Notes, 2, 373-376. https://doi.org/10.1046/J.1471-8278

Sullivan, P. W., Adair, R., \& Woldt, A. (2016). Sea Lamprey Control in the Great Lakes 2015: Annual Report to the Great Lakes Fishery Commission. Ottowa, Ontario.

Wagner, C. M., Twohey, M. B., \& Fine, J. M. (2009). Conspecific cueing in the sea lamprey: do reproductive migrations consistently follow the most intense larval odour? Animal Behaviour, 78(3), 593-599. https://doi.org/10.1016/j.anbehav.2009.04.027

Waller, D. M., \& Keller, L. F. (2002). Inbreeding effects in wild populations. Trends in Ecology and Evolution, 17(5), 230-241.

Wang, J., \& Santure, A. W. (2009). Parentage and sibship inference from multilocus genotype data under polygamy. Genetics, 181(4), 1579-1594.
https://doi.org/10.1534/genetics.108.100214
Wang, Jinliang. (2009). A new method for estimating effective population sizes from a single sample of multilocus genotypes. Molecular Ecology, 18(10), 2148-2164.
https://doi.org/10.1111/j.1365-294X.2009.04175.x

Wang, Jinliang. (2016). A comparison of single-sample estimators of effective population sizes from genetic marker data. Molecular Ecology, 25(19), 4692-4711. https://doi.org/10.1111/mec. 13725

Waples, R. K., Larson, W. A., \& Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity, 117(4), 233-240. https://doi.org/10.1038/hdy.2016.60

Waples, R. S. (2016). Tiny estimates of the Ne/N ratio in marine fishes: Are they real? Journal of Fish Biology, 89(6), 2479-2504. https://doi.org/10.1111/jfb. 13143

Waples, R. S. (2005). Genetic estimates of contemporary effective population size: To what time periods do the estimates apply? Molecular Ecology, 14(11), 3335-3352. https://doi.org/10.1111/j.1365-294X.2005.02673.x

Waples, R. S. (2016). Making sense of genetic estimates of effective population size. Molecular Ecology, 25(19), 4689-4691. https://doi.org/10.1111/mec. 13814

Waples, R. S., \& Anderson, E. C. (2017). Purging putative siblings from population genetic data sets: A cautionary view. Molecular Ecology, 26(5). https://doi.org/10.1111/mec. 14022

Waples, R. S., \& Antao, T. (2014). Intermittent breeding and constraints on litter size: Consequences for effective population size per generation (ne) and per reproductive cycle (nb). Evolution, 68(6), 1722-1734. https://doi.org/10.1111/evo. 12384

Waples, R. S., Antao, T., \& Luikart, G. (2014). Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics, 197(2), 769-780. https://doi.org/10.1534/genetics.114.164822

Waples, R. S., \& Chi, D. O. (2008). ldne: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources, 8(4), 753-756. https://doi.org/doi:10.1111/j.1755-0998.2007.02061.x

Waples, R. S., \& Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. Evolutionary Applications, 3(3), 244-262. https://doi.org/10.1111/j.17524571.2009.00104.x

Waples, Robin S., Grewe, P. M., Bravington, M. W., Hillary, R., \& Feutry, P. (2018). Robust estimates of a high $\mathrm{Ne} / \mathrm{N}$ ratio in a top marine predator, southern bluefin tuna. Science Advances, 4(July).

Waples, R. S., \& Waples, R. K. (2011). Inbreeding effective population size and parentage analysis without parents. Molecular Ecology Resources, 11(1), 162-171. https://doi.org/10.1111/j.1755-0998.2010.02942.x

Waples, R. S., Waples, R. K., \& Ward, E. J. (2020). Pseudoreplication in genomics-scale datasets. BioRxiv, (November), 1-34. https://doi.org/10.1101/2020.11.12.380410

Whiteley, A. R., Coombs, J. A., Cembrola, M., O’Donnell, M. J., Hudy, M., Nislow, K. H., \& Letcher, B. H. (2015). Effective number of breeders provides a link between interannual variation in stream flow and individual reproductive contribution in a stream salmonid. Molecular Ecology, 24(14), 3585-3602. https://doi.org/10.1111/mec. 13273

Whiteley, A. R., Coombs, J. A., Hudy, M., Robinson, Z., Nislow, K. H., \& Letcher, B. H. (2012). Sampling strategies for estimating brook trout effective population size. Conservation Genetics, 13(3), 625-637. https://doi.org/10.1007/s10592-011-0313-y

## CONCLUSIONS

Overall, the experiments and results detailed above demonstrate the utility of reconstructed pedigrees and $N_{b}$ and $N_{s}$ estimates for evaluating sea lamprey spawning populations in streams. With regards to determining the number of cohorts, a necessary step in calculating $N_{b}$ and $N_{s}$, mixture analysis models using length data alone were insufficient to separate individuals into cohorts, particularly for age 2+ individuals. Reconstructed pedigrees and the presence of family structure can be combined with length data using a decision-making matrix to identify cohorts that are oversplit by the mixture analysis models to more accurately generate cohort assignments for genotyped populations.
$N_{b}$ and $N_{s}$ estimates, along with the analysis of reconstructed pedigrees, are useful in assessing various management actions in the context of invasive species. By using larval genotypes, parental genotypes can be reconstructed to obtain information on adults years after spawning occurs. This allows for assessment of barrier efficacy if larvae are found in subsequent years. In Chapter 1, we determined that despite the presence of larvae above barriers in northern Michigan, spawning of most larvae occurred prior to barrier construction (in the case of the Black Mallard), or were from a group of mostly half-siblings implying a small number of males (in the case of the Ocqueoc) In both cases, results indicated that barriers in two systems did not have a large-scale failure. In Chapter 2, by sequencing a larger number of streams we saw a large variety in the types of family structure as well as $N_{b}$ and $N_{s}$ estimates, showing that spawning populations differ widely from stream to stream. This indicates that the control efforts and methods required to minimize sea lamprey spawning may be different depending on spawning structure.

We also evaluated $N_{b}$ and $N_{s}$ along three parameters relevant to sea lamprey populations: true $N_{b}$, sample size, and SNP set size. We found that genomic estimates generated with large SNP set were robust and accurate even when the true $N_{b}$ of the simulated population was very large. However, sample size and SNP set size becomes an important factor for estimates generated from a reconstructed pedigree at a large true $N_{b}$, particularly for $N_{s}$ estimates. Additionally, our linear models showed that sample size was a significant predictor for $N_{b}$ and $N_{s}$ estimates in our empirical data set, highlighting the importance of sufficient sampling for accurate $N_{b}$ and $N_{s}$ estimation. Additionally, representative sample across potential spawning habitat is vital for obtaining $N_{b}$ and $N_{s}$ estimates that reflect the full stream population. $N_{b}$ and $N_{s}$ estimates are an emerging technique for the assessment of invasive species, and have been established an effective technique for evaluating species of conservation concern. Due to the expanding genomic resources and extensive research efforts, sea lamprey are an emerging model system for long-term management. Providing $N_{b}$ and $N_{s}$ estimates and simulating sea lamprey populations has shown that genomic assessment is a valuable addition to sea lamprey assessment and evaluating control efforts, including new supplemental control techniques.

