

****ABSTRACT NOT FOR CITATION WITHOUT AUTHOR**

PERMISSION. The title, authors, and abstract for this completion report are provided below. For a copy of the full completion report, please contact the author via e-mail at nicholas.sard@oswego.edu. Questions? Contact the GLFC via email at slrp@glfc.org or via telephone at 734-669-3020. **

Simulation-based assessment of a large-scale genetic tagging study to identify adult sea lamprey natal streams

Nicholas Sard², John Robinson³, Scott Miehl⁴, and Kim Scribner^{3,5}

² Biological Sciences Department, SUNY Oswego, Oswego, NY 13126

³ Fisheries and Wildlife Department, Michigan State University, East Lansing, MI 48824

⁴ United States Geological Survey, Hammond Bay Biological Station, Millersburg, MI 49759

⁵ Integrative Biology Department, Michigan State University, East Lansing, MI 48824

June 2022

ABSTRACT:

Identifying the natal origins of adult sea lampreys is a high priority for the Great Lakes Sea Lamprey Control Program. The use of traditional mark recapture techniques to identify natal origins, in part, are not feasible due to poor recovery rates that would require marking and releasing an unrealistic number of larvae into tributaries around a lake. A genomic-scale single nucleotide polymorphism (SNP) panel was recently created for sea lamprey, thereby enabling individuals to be genotyped at 3,446 loci. The increase in loci that could be genotyped in a standardized manner across individuals has provided an unprecedented degree of statistical power to establish individual identity and for use in pedigree reconstruction. Before using the newly developed SNP markers in the field, here, we used simulations to investigate the application of the panel for close-kin markrecapture (CKMR) studies to identify sea lamprey natal origins. First, we compared two pedigree reconstruction programs to determine if a recently created algorithm (Sequoia) would outperform a well-established program (Colony). While Sequoia was faster at reconstructing pedigrees (< 5 minutes), it nearly always incorrectly inferred half-siblings as unrelated. Given the polygamous mating structure of sea lamprey, it will be important to correctly infer half-siblings in CKMR studies. In contrast to Sequoia, Colony correctly inferred the majority of full- and half-siblings across varying levels of breeding adult population sizes ($n_{\text{parents}} = 250, 500, 1000$), offspring sampled ($n_{\text{off}} = 250, 500$), and loci genotyped ($n_{\text{loci}} = 200, 400, 800$). However, the reconstruction could take more than a week to complete using the High-Performance Computing Cluster at Michigan State University. Despite the computational burden, we used Colony in subsequent work because half-siblings were correctly inferred more frequently. Second, we created a forward simulation model in the program Rmetasim to represent the demographic and genetic processes occurring in Lake Erie based on published

papers and technical reports. Effectively the model tracked ~220-414 thousand individuals genotyped at 800 loci in 16 breeding populations of varying sizes (10-10000 adults per locale) across a 17-year time period. We investigated if sea lamprey natal origins could be inferred on a lake-wide scale using sibling relationships inferred by the pedigree reconstructed in Colony using 800 loci. At a Great-Lake-wide scale, full-siblings and unrelated individuals were inferred with high accuracy (median estimate = 0.99); however, half-sibling accuracy was modest (range: 0.36-0.54). Despite errors in pedigree reconstruction, the relative proportion of natal origin assignments among adults collected at traps was positively correlated with the number of age-1 fish present in respective tributaries. While more work needs to be done to improve half-sibling inferences, results suggest that CKMR can provide information useful to management actions.