Lampreys adjust, mammals non-plussed, birds robust: how ecological and environmental features influence genetic diversity

By

Meghan Laurel Mahoney

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
In partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biological Sciences
University of Manitoba
Winnipeg

Abstract

Evolutionary forces are intrinsically tied to environments and ecosystems species occupy. Gene flow is shaped by conduits and impediments in landscape features and conditions, and spatially separate populations often divergently adapt to heterogenous environments. Understanding the nexus of environments and genetic diversity is vital when managing species, especially in the context of rapid global climate change and increasing anthropogenic disturbance. I used two avenues of study to explore this relationship on large spatial scales. I analyzed whole-genome sequencing data for 209 invasive sea lampreys in the Laurentian Great Lakes, and publicly archived, raw microsatellite data from 1,008 bird and mammal population data points across Canada and the United States. For the former, I hypothesized that as an invasive species recently introduced to a novel environmental gradient, the sea lamprey populations would be locally adapted to conditions in the Great Lakes. For the latter, I hypothesized that the human footprint index (HFI), when used as a resistance surface, would best explain genetic distances in bird and mammal populations in North America. For both studies, I used statistical approaches to look for patterns of genetic diversity among and across populations of animals and identify environmental covariates of these patterns. I found evidence of local adaptation in sea lamprey populations, whereby the adaptive divergence of populations significantly correlated with levels of human population density. Bird and mammal populations were also shaped by human influence, with a positive and negative effect, respectively, of the HFI on genetic distance. Though direction and degree of effects on genetic diversity varied across taxonomic groups, these results indicate the overarching influence environmental variables—particularly, human disturbances—have on spatial distribution and genetic diversity across taxonomic groups. Be it an invasive species, or species of conservation concern, understanding the link between environmental gradients and genetic diversity of animal populations is of both biological and managerial interest.

Acknowledgements

I would like to thank my advisors Dr. Colin Garroway and Dr. Margaret Docker for bringing me onto the lamprey project and into their labs. Colin, thank you for your positivity and willingness to hop on Zoom chats on short notice, no matter how trivial the matter. Margaret, thank you for your kindness, and for sharing your infectious love for all things lampreys. I am tremendously grateful for your combined knowledge, patience, and support, especially during a global pandemic. The past few years have certainly been trying, and I could not have hoped for better mentors to learn from. Thank you also to my committee members Dr. Ken Jeffries and Dr. Stephen Petersen for their insight and encouragement. We may not have been able to meet in person, but I sincerely value your feedback and guidance.

Thanks to the members of the Garroway lab, who made every meeting and gathering a celebration. Thank you, Chloe, for assembling synthesis data and lending guidance on the project, as well as being a sci-comm inspiration. Thank you Arfa for your tireless work in getting samples ready to be sent off for sequencing. Thank you, Phil and Jaanus, for the help and training, and thank you, along with Claudio, for your support and friendship and for our weekly Dungeons and Dragons games, which have kept me sane through their chaos. Thanks to Cassandra, for a contagious love of teaching, and general kindness and support. Thank you also to Jessie, whose friendship I count as one of the most important things to come out of my master's degree—thanks for the bus rides, the FaceTimes, and the Coffee Crisp parcels.

A loving thank you to my friends and family, especially my parents, my brother Ryan, and my fiancé Erik. It was their unremitting love and support that kept me going, especially on more challenging days. Thank you also to my cats Maple, my constant companion of three years, and Mochi, our pandemic kitten; much of this thesis was written with one or both on my lap, purring up a storm.

This project was supported by the Great Lakes Fishery Commission, the Natural Sciences and Engineering Research Council of Canada Discovery Grant, and the University of Manitoba Graduate Fellowship. Thank you also to those who collected lamprey samples in the Great Lakes, and to all the tremendous scientists out there whose publicly archived data I used in my analysis.

Finally, I must thank all the feathered, furred, and scaled critters we share this world with—even the lampreys.

Dedication

For my mom, Dee, my biggest supporter who has always known ‘I can’ even when I told myself ‘I can’t’. And for my dad, Duffy, whose effusive love of life and learning has shaped me into the scientist I am today.

Table of Contents

List of Tables

List of Figures

1 | Introduction

My overarching question for this Thesis is: How do landscape elements shape gene flow and local adaptation? We know that animals disperse for many reasons; to take advantage of new resources, find a mate, or avoid competition (Chatterjee 2006). Animals may also disperse to find an environment that best suits a certain phenotype (Edelaar et al. 2008). Biotic, abiotic, and anthropogenic factors shape gene flow, and these three facets combine to make up an environment (Anderson et al. 2010). Often, individuals disperse following environmental gradients, and over time adapt to the new habitats they occupy (Pontarp & Wiens 2017). There are two avenues to study the interactions between environments and genetic structure: neutral and adaptive landscape genetics. I investigated both types of genetic signatures to better understand evolutionary change in wildlife.

Neutral landscape genetics uses neutral loci to infer how landscape factors influence gene flow. Neutral loci, which make up a large part of an organism's DNA, do not affect the fitness or performance of the organism. These alleles are thus considered selectively neutral and are informative as to how species move across a landscape and how populations experience gene flow (Holderegger et al. 2006); when dispersed animals reproduce in new habitats, their movements are detectable in neutral genetic signatures of gene flow.

Contrastingly, adaptive landscape genetics links adaptive loci under selection to environments to explore signatures of local adaptation (Vincent et al. 2013). At loci under selection, a genotype should influence an animal's fitness at certain times and in certain environments. Environments can lead to strong selective pressures, resulting in direct influence on different populations occupying different habitats (Holderegger et al. 2006). When spatially varying environments cause differential survival and reproduction, then signatures of local adaptation become detectable in patterns of adaptive genetic variation (Holderegger et al. 2006).

Populations vary due to the process of local adaptation, as driven by natural selection. Natural selection varies across space, resulting in gene-environment interactions that can increase or decrease fitness. Local adaptation then can occur, despite gene flow with neighboring populations, if the population maintains locally advantageous traits (Savolainen et al. 2013). Thus, environmental features likely play a role in driving evolution, and I explored this relationship in systems of both native and invasive species.

Invasive species are successful in environments they are either preadapted to or can quickly adapt to. Invasives might offer insight into rapid adaptation, as they can be studied within a recent time frame and often exhibit strong responses to selective pressures (Wu et al. 2019). Sea lamprey (*Petromyzon marinus*) have successfully and extensively invaded the Great Lakes ecosystem (Marsden & Siefkes 2019), offering a study system for how invasive species might locally adapt to new environments. This project investigated the adaptive landscape genetics of sea lamprey; by looking for genotype-environment associations within the genome of invasive sea lamprey from across the Laurentian Great Lakes, I was able to assess the extent to which local adaptation has occurred as well as the most influential environmental features within this process. Understanding what environmental factors make lamprey such successful invaders can further help to control that species.

Additionally, understanding the connection between environment and microevolution is vital when managing for and conserving at-risk native species (Rice & Emery 2003). Geography contributes directly to differentiation and local adaptation as organisms adapt to their habitats. In this project, I used landscape genetic tools to study neutral markers; frequencies of non-adaptive loci helped to determine patterns of gene flow among populations (Manthey & Moyle 2015). The second project of this thesis explored how underlying genetic processes contribute to evolution of species. Specifically, by synthesizing microsatellite data of North American mammals, I investigated whether isolation by environment or isolation by distance is a more apt measure of genetic diversity between populations. This, in turn, allowed me to identify the most important corridors to and barriers for gene flow. Gene flow can increase genetic variation, but also dilute adaptive genetic variation and reduce population size (Sexton et. al 2014). Therefore, it is important to study gene flow to predict how vulnerable a population is and the responsiveness of organisms to a changing environment.

Literature Cited

Anderson, C.D., Epperson, B.K., Fortin, M.-J., Holderegger, R., James, P.M.A., Rosenberg, M.S., Scribner, K.T. & Spear, S. (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, 19, 3565-3575.

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

Chatterjee, M. (2006). On the move: What causes animals to disperse? *Resonance,* 11, 55–71.

Edelaar, P., Siepielski, A. M., & Clobert, J. (2008). Matching habitat choice causes directed gene flow: a neglected dimension in evolution and ecology. *Evolution: International Journal of Organic Evolution*, 62(10), 2462-2472.

Holderegger, R., Kamm, U. & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology,* 21, 797–807.

Manthey, J. D., & Moyle, R. G. (2015). Isolation by environment in white-breasted nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: A landscape genomics approach. *Molecular Ecology*, 24(14), 3628–3638.

Pontarp, M., & Wiens, J. J. (2017). The origin of species richness patterns along environmental gradients: uniting explanations based on time, diversification rate and carrying capacity. *Journal of Biogeography*, 44(4), 722-735.

Rice, K. J., & Emery, N. C. (2003). Managing microevolution: Restoration in the face of global change. *Frontiers in Ecology and the Environment*, 1(9), 469–478.

Savolainen, O., Lascoux, M. & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Review Genetics,* 14, 807–820.

Vincent, B., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. (2013). Landscape genomics in Atlantic salmon (*Salmo salar*): Searching for gene-environment interactions driving local adaptation. *Evolution*, 67(12), 3469–3487.

Wu, N., Zhang, S., Li, X., Cao, Y., Liu, X., Wang, Q., & Liu, Q (2019). Fall webworm genomes yield insights into rapid adaptation of invasive species. *Nature Ecology & Evolution*, 3(1), 105-115.

2 | Rapid local adaptation in freshwater populations of invasive sea lamprey (*Petromyzon marinus*)

2.1 | Introduction

2.1.1 | Local Adaptation in Invasive Species

Populations of many species vary spatially and environmentally. Even in species with small range sizes, fine-scale adaptive divergence can occur because of nonrandom, phenotype-dependent dispersal (Edelaar et al. 2008, Garroway et al. 2013). Phenotypic variation, differences in morphology, physiology, and behavior, is in part caused by complex genetic interactions between multiple loci, as well as an organism's sensitivity to environmental conditions (Mackay et al. 2009). Determining genetically distinct populations and the genetic basis for phenotypic variation between them is informative for biological management, especially in the context of invasive species (Lee 2002).

The goal of this study was to assess whether invasive populations of sea lamprey (*Petromyzon marinus*) have become locally adapted across environmental gradients in the Laurentian Great Lakes. North American freshwater species have experienced five times higher rates of extinction than North American terrestrial animals (Ricciardi & Rasmussen 1999), with invasive species significantly contributing to this high rate of decline (Mollot et al. 2017, Wilcove et al. 1998). Understanding how invasive species evolve rapidly in response to novel environmental conditions can aid in understanding how these organisms become established, subsequently spread, and impact their communities (Colautti & Lau 2015).

Environmental-driven genetic variation is known as local adaptation and requires some degree of genetic diversity to occur. To be considered locally adapted, a population should have higher fitness in the environment it lives in when compared to other populations introduced to that site (Savolainen et al. 2013). Local adaptation results in changes in average phenotypes of populations of the same species, and at some loci, allele frequencies may differentiate populations in different environments. When selection pressures from the environment are sufficiently different across the landscape, selected alleles may maintain their frequency despite the occurrence of gene flow (Coop et al. 2010). Observing phenotypic and genetic differentiation along gradients can be indicative of local adaptation (Savolainen et al. 2013).

Biological invasions are exceptional systems in which to study local adaptation (Franks & Munshi-South 2014, Sakai et al. 2001). Invasives, characterized as non-native species

established in an ecosystem where they have an influence, can be a widespread problem for economics, human health, native wildlife, and ecosystems (Mazza et al. 2014, Andersen et al. 2004), where they can alter hydrology, biogeochemistry, and biotic composition (Strayer 2010). Biological invasions are stochastic events involving small populations being introduced into a new environment, where the invasive species can then outcompete native species for resources (Lee 2002). Furthermore, invasive species can cause cascading effects by decreasing native species richness, including in aquatic environments.

It has been demonstrated that invasive species can experience rapid adaptation to new environments in 20 generations or less (Yin et al. 2021, Medley et al. 2019, Colautti & Barrett 2013, Prentis et al. 2008). Local adaptation of invasive species can lead to an increased rate of both survival and reproduction; this increase in fitness can result in the high abundances seen in many populations of noxious invasive species (Colautti & Lau 2015). When climate change, human disturbances, or invasive behaviors cause a species' environment to shift, locally adapted populations can persist (Savolainen et al. 2013). Connecting environmental features to local adaptation can provide vital information on how to control or eradicate invasive species, as well as understand the general features of a successful invasion.

While invasion events reduce genetic diversity and allelic richness via a population bottleneck, the early stages of invasion can lead to an increase in fitness by the removal of deleterious alleles via natural selection (Colautti & Lau 2015). This can lead to rapid evolution because of increased fitness. Also, in the early stages of invasion, directional selection should have a stronger effect on invasive than native species, due to the naivety of invasive populations to novel environmental conditions (Colautti & Lau 2015).

2.1.2 | Sea Lamprey: Life History and Invasion Timeline

While not all non-native species are noxious, invasive sea lampreys fall into this category. In native populations of anadromous sea lamprey, larvae spend 5 to 8 years in freshwater. As juveniles, they live in the ocean for approximately 2 years, adopting a parasitic feeding strategy by latching on to host fishes with serrated oral discs (Applegate 1950). By attaching to host fishes, anadromous sea lampreys often disperse great distances from larval spawning and rearing sites (Moser et al. 2015). Following the parasitic feeding phase, anadromous sea lamprey spawn in freshwater waterways on both sides of the Atlantic Ocean (Potter 1980). Some adults spawn upwards of hundreds of kilometers upstream, and native

populations of anadromous sea lamprey typically access larger rivers over greater distances than the invasive populations of the Great Lakes (Moser et al. 2015).

In the Great Lakes ecosystem, the sea lamprey is an invasive species that was first discovered breeding in a tributary of Lake Ontario in 1835 (Christie et al. 2003). The 1835 record may be dubious, in which case the first credible sighting of sea lamprey in Lake Ontario was in 1888 (Eshenroder 2014, Marsden & Siefkes 2019). It is contested whether sea lamprey entered Lake Ontario during post-Pleistocene glacial retreat (Bryan et al. 2005, Waldman et al. 2006) or invaded more recently following construction of the Erie Canal in 1818 (Smith & Tibbles 1980), and this distinction is important when considering whether these populations adapted over many thousands of years, or more rapidly in the last 200 years. While there is debate about the nativity of sea lamprey to Lake Ontario, the sea lamprey is undoubtedly non-native to the other Great Lakes; before the building of the Welland Canal in 1829, sea lamprey could not pass Niagara Falls (Applegate 1951). Following the construction of the canal, sea lamprey quickly colonized the other Great Lakes by the 1920's and 30's (Docker & Potter 2019).

Sea lamprey have since predated on and decreased populations of native species, a problem exacerbated by the sea lamprey's lack of predators (Sorensen & Bergstedt 2011). While humans also contributed to the decline of fish species such as lake trout (*Salvelinus namaycush*), lake whitefish (*Coregonus clupeaformis*), and burbot (*Lota lota*), lamprey furthered the depletion of these fisheries (GLFC & Christie 1973). With humans targeting the largest of these commercial fishes, smaller fish were left vulnerable to attack and more likely to die from lamprey parasitism (Marsden & Siefkes 2019). By the 1940's, the invasion had shown profound enough effects on fisheries that the Canadian and American governments were spurred into action (Siefkes et al. 2013); the Great Lakes Sea Lamprey Committee was formed in 1946, followed by the Great Lakes Fishery Commission in 1955 (Smith & Tibbles 1980). Still, by the 1960's, the catch of Great Lakes trout had dropped to 2% of what it previously was, leading to a devastation of fisheries, leaving people without jobs, and crippling the region's economy (Smith & Tibbles 1980).

In the 1940's, mechanical and electrical weirs and low barriers were utilized to prevent upstream migrating adults from entering the lakes from tributaries; this method of control is still used today (Marsden & Siefkes 2019, Siefkes et al. 2013). The construction of barriers can reduce access to 15% of available larval habitat (Lavis et al. 2003) and have been shown to

reduce abundance by preventing adult breeding populations from reaching spawning grounds (Avenetti et al. 2006, Marsden & Siefkes 2019). While critical in managing invasive species, barriers have drawbacks as they negatively affect lotic ecosystems and are costly to construct and maintain (McLaughlin et al. 2013). Lampricide treatments aimed at killing lamprey larvae were later developed in the 1950s with the aim of not affecting non-target organisms (Siefkes et al. 2013); now, a combination of lampricide treatments (aimed at killing larvae) and barrier methods (aimed at breeding adults) are used in tandem to control invasive sea lampreys.

Lampricides have since dramatically reduced spawning-phase sea lamprey abundance in all lakes (Siefkes et al. 2013) and are considered highly successful (Siefkes 2017), particularly 3-trifluormethyl-4-nitrophenol (Applegate et al. 1961) and niclosamide (2′,5-dichloro- 4′-nitrosalicylanilide) (McConville et al. 2016). The modern control program largely focuses on suppressing populations rather than total elimination using these methods (Adams et al. 2021). Elimination is thought to be unachievable, as even with barriers and pesticide treatments, some residual larvae survive and contribute to the parasitic juvenile populations that feed in open lakes. Furthermore, surviving larval populations of sea lamprey emit migratory pheromones into the water (Meckley et al. 2014). These pheromone odors disperse through waterways, drawing in adult sea lamprey populations to river mouths (Moser et al. 2015), and then to larval rearing and spawning grounds (Christie et al. 2003, Marsden & Siefkes 2019). Even with a lowered population density, adult lampreys can still follow these traces of pheromones to find mates (Meckley et al. 2014). Despite not being able to wholly eliminate invasive lamprey, these techniques have successfully reduced sea lamprey populations by up to 90% (Marsden & Siefkes 2019). However, to maintain this control, yearly, diligent efforts are required, as well as considerable financial investment from governmental agencies (Adams et al. (in press), Simberloff 2014).

By collecting genetic data from sea lamprey individuals across a variety of landscape features, I aimed to look for patterns of adaptive variation. I tested correlations between genetic variation and environment while estimating confounding factors present in residual levels of background structure using latent factor mixed models. Mixture models, principal component analysis, hidden Markov models, and factor analysis were then combined to estimate population parameters while accounting for presence of confounding variables (Frichot et al. 2013).

The main goal of the study was to examine the patterns of local adaptation between sea lamprey in different populations across the Laurentian Great Lakes using whole genome sequencing. Specifically, I analyzed divergence based on environmental factors to determine whether adaptive genetic differentiation has occurred in these invasive populations. I aimed to test the hypotheses: (a) local adaptation has occurred in invasive sea lamprey in the Great Lakes, and (b) sea lamprey local adaptation has been driven by specific environmental variables. This study is significant in terms of invasive species management, as understanding the rate at which invasive organisms evolve and to which environmental conditions they have adapted are both biologically and managerially important.

2.2 | Materials and Methods

2.2.1 | Study Area

Sea lampreys were sampled from 13 areas across the Laurentian Great Lakes, which are, from west to east: Superior, Michigan, Huron, Erie, and Ontario. The Laurentian Great Lakes are located between 40-50°N, have a combined surface area of 244,000 km^2, and account for 20% of Earth's fresh surface water (Magnuson et al. 1997). Climate is variable across the Laurentian Great Lakes, from boreal areas in the eastern and northwestern regions with cooler winter and summers, to the warmer and more humid seasons in the south-western region (Magnuson et al. 1997). Lake depth is also highly variable, from an average of 19 meters deep in the Lake Erie Basin, to 187 meters deep in Lake Superior (Waples et al. 2008). Approximately 34 million people live in the Great Lakes Basin (Michigan Sea Grant 2018), including in large cities such as Toronto, Chicago, Detroit, Milwaukee, and many smaller metropolitan areas. These densely populated settlements have led to an increase in plastic (Earn et al. 2020, Mason et al. 2020) and organic pollutants (Guo et al. 2018) within the water and sediment of the Great Lakes.

2.2.2 | Sampling

A total of 238 adult individuals were sampled by Great Lakes Sea Lamprey Control Personnel during the spring spawning migration in 2019, 209 of which were used in this analysis: Lake Erie: (Grand River [GO]: 14, Big Creek [BC]: 16); Lake Huron: (Augres River [Augres]: 15, Bridgeland [BL]: 15, Cheboygan [CH]: 16, St. Mary's River [SM]: 17); Lake Michigan: (Grand River [GR]: 15, Manistique [MAN]: 16); Lake Superior: Middle Brule River [BR]: 16, Rock River [RO]: 15, Tahquamenon River [TQ]: 16); and Lake Ontario: (Black River [BLR]: 18, Duffin's Creek [DF]: 20). Coordinates for each sample site were taken at the river

mouth (*Table 1)*. All individuals were frozen, and whole tail fin clips were later collected from the frozen samples and preserved in 95% ethanol for subsequent DNA extraction and whole-genome sequencing and analysis.

2.2.3 | Environmental Measurements

Sea lampreys were captured in stream traps; I used corresponding latitude and longitude values of river mouths when collecting environmental data from several databases. Both the sites, plotted as points, and the environmental datasets were projected in Universal Transverse Mercator (UTM) zone 16 in the WGS84 datum in meters. UTM 16 was selected as it encompasses lakes Superior, Michigan, and Huron, where most of the samples were collected.

Localized environmental data was more readily available at river mouths, as opposed to tributaries. Additionally, the lamprey sampled were all adults, meaning they had completed their juvenile parasitic feeding stage in the open lakes and were sampled as they traveled on the upstream spawning migration. Therefore, while I could determine which lake the adult lamprey spent their parasitic feeding stage in, we couldn't determine in which tributaries these animals spent their larval stages as lampreys do not home to their natal streams. This analysis focused on the environmental conditions experienced by juvenile lampreys, rather than the long-lived larval stages found in smaller streams and tributaries. Many variables, such as latitude, affect all life stages of these organisms in a similar capacity. However, some variables in this analysis affect life stages differently. For example, juvenile lampreys feed on fish in lakes, and larval lampreys live in and feed upon detritus in tributaries. A variable such as primary productivity would have a very different influence on fish-feeding juvenile animals, versus larvae which directly depend on productivity of their local ecosystems. Thus, the use of river mouth coordinates examines conditions of juvenile animals rather than larval animals.

I used latitude values from each of the collection sites as a variable in my model, as it has been found to drive local adaptation in the Pacific lamprey (*Entosphenus tridentatus*) (Hess et al. 2013). Latitude is a surrogate for many variables including temperature, energy availability, and primary productivity (Gillman et al. 2015, Lyons & Willing 2002). For example, lakes at lower latitudes are typically more productive than at higher latitudes and in comparison to oceans (Gross et al. 1988), and I theorized this productivity could be a significant variable driving local adaptation. Latitude as a variable was treated as a proxy for other, untested variables that correlate with latitude and may be driving local adaptation.

To look at whether signatures of local adaptation could be explained by a lake-dependent variable, I assigned each lake a discrete numerical value, and each sample site was assigned this value corresponding with its lake. Like using latitude as a vector, a hierarchical lake level variable is not an explanatory variable in and of itself, but a proxy capturing other environments. I hypothesized that there may be some uncaptured, unique environmental differences in each of the five Great Lakes potentially driving dispersal and local adaptation.

I collected four environmental variables from the Great Lakes Aquatic Habitat Framework (GLAHF), three of which were temperature variables calculated by NOAA's Great Lakes Coastal Forecasting System hydrodynamic models (Schwab & Bedford 1999). Several temperature variables were selected as they explain lamprey distribution at large scales (Potter 1980), influence larval movement (Derosier et al. 2007), larval growth rate and survival (Beamish & Potter 1975, Potter 1980), adult sea lamprey weight (Cline et al. 2014), and feeding efficiency (Morman et al. 1980).

Annual vertical temperature (AVT) was calculated with hydrodynamic models from vertical water temperature using a sigma coordinate system (Wang et al. 2015). I calculated AVT for each of the sample sites by averaging seven years (2006-2012) of mean annual vertical water temperature data in ArcMap 10.3.1 (Esri 2011). While the adults sampled were feeding in the lakes more recently than 2012, data from NOAA's Great Lakes Coastal Forecasting System was only available for the six-year span used in my analysis. Vertical temperature was calculated by taking an average of temperatures from across epilimnion, metalimnion, and hypolimnion lake depths. Summer vertical temperature (SVT) was the same as the AVT variable, except it summarized the average temperature across all lake depths during the summer months (June, July, and August) rather than annually. The third temperature variable collected from GLAHF was annual surface temperature (AST), which was calculated from the mean of eighteen years (1995-2013) of annual surface temperature data of each of the Great Lakes (NOAA Great Lakes CoastWatch). Again, while the lampreys collected in my study inhabited the lakes later than 2013, daily remote sensing estimates were unavailable outside of this eighteen-year range. Annual surface temperature was of interest for two reasons, as in addition to temperature, AST can act as a proxy for primary productivity as higher temperatures and light intensity can act on phytoplankton productivity (Lewandowska et al. 2012). For all three temperature maps, land surfaces were excluded from analysis.

The fourth variable used in analysis was bathymetric readings of lake depths, measured in meters, from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information. Depth has been found to drive differentiation in populations of fish species (King et al. 2006, White et al. 2010); for example, in freshwater populations of sockeye salmon (*Oncorhynchus nerka*), body size was found to be correlated with water depth (Quinn et al. 2001).

I also extracted one environmental variable from the NASA Earth Observations (NEO) Database: human population density (*Figure 5*). This data consists of 1-kilometer grids, with an average human population density corresponding with each square of the grid. The map is based on the Gridded Population of the World dataset, Version 4.11 (GPWv3) (CIESIN 2018), which has data available for the years 2000, 2005, 2010, 2015 and 2020. Because my samples were all collected in 2019, I used the population density estimates for the year 2020. I selected human population density as a variable of interest due to evidence suggesting human disturbance can lead to an increase in invasion success (Jeschke & Strayer 2006). Disturbance undoubtedly influences the spread of many invasive species (Fox & Fox 1986), and because humans are a primary driver of habitat fragmentation and disturbance (Mullu 2016), there is reason to believe human activity in and around the Great Lakes Basin could impact sea lamprey.

All environmental variable data was loaded into the ArcMap 10.3.1 environment, where it was buffered to 10 kilometers surrounding each of the sample sites. The mean value of each environmental layer under this 10-kilometer buffer was then summarized to assign a numerical value to each sample site (*Table 1*). All variables were then assessed to look for correlation using an R^2 coefficient of determination (*Table 2*). Due to the high correlation across most variables, I ran the model once with all variables, and then with each variable individually.

2.2.4 | Genomic Analysis

2.2.4.1 | DNA Extraction and Sequencing

I extracted DNA from the ethanol-preserved whole tail fin clips with DNeasy Blood and Tissue Kit (Qiagen), following the manufacturer's protocol. Samples were eluted in 100 µl of the kit's elution buffer. Sample concentrations and quality were assessed on the NanoDrop spectrophotometer (Thermo Scientific). Quality of the sample was assessed using A260/A280 ratios. Library preparation and high-throughput sequencing was carried out by The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children, Toronto. Libraries were prepared

with the TruSeq PCR-Free Kit (Illumina). Sequencing was conducted using Illumina NovaSeq 6000, using S4 flowcells (one animal per flowcell) and 300 cycles.

2.2.4.2| Filtering and Quality Control

I performed quality control checks on my raw sequencing data using the package FastQC (Andrews 2010, version 0.11.8). I then viewed all the FastQC output together and visualized the quality of the data using the package multiqc (Ewels et al. 2016, version 3.8.0). I removed the adapters introduced in the initial sequencing using the package trimmomatic (Bolger et al. 2014, version 0.36), and FastQC and MultiQC were run again to compare the changes following the removal of adapters. I mapped my reads to the latest sea lamprey genome assembly (kPetMar1.pri, RefSeq identifier GCF_010993605.1) by using the package bowtie2 (Langmead & Salzberg 2012, version 2.3.4.3). This assembly originates from adult sperm collected by Dr. Jeramiah Smith (University of Kentucky) that had been sequenced in the frame of the Vertebrate Genomes Project (VGP). The sea lamprey genome is highly repetitive (Smith et al. 2013) and the VGP assembly consists of 85 chromosomes and 1,195 unplaced sequences. This assembly has a total of 1,434 scaffolds which range in size from approximately 16,000 to 34,000,000 base pairs, scaffold N50 12,997,950.

I removed any duplicate reads using MarkDuplicates from Picard (Picard Toolkit 2018, version 2.20.6), followed by variant calling with the package Platypus (Rimmer et al. 2013, version 0.8.1). The .vcf-formatted output file contained polymorphism data for 1,335 of the scaffolds, and 3,154,585 base pairs of the genome's total ~1 billion base pairs. I have mid-to-low coverage reads of 285 lampreys in total, 209 of which were used in my analysis after filtering out sites with insufficient environmental data. Within these 209 individuals, 1,486,675 polymorphic loci were identified. The data was filtered at 0.05 minimum allele frequency and for biallelic loci only and missing data was removed, using the program PLINK (Purcell et al., 2007, version 1.9b_5.2-x86_64); biallelic loci were used because the tools used in this analysis accepted biallelic format as the default input. PLINK was then used to convert the data to a PED format.

2.2.5 | Statistical Analysis

2.2.5.1 | Outlier Detection Methods

To identify loci under putative selection, I used population differentiation (PD) testing to identify outlier SNPs. PD testing compares a neutral model to genomic regions of high differentiation to detect selection (Luikart et al. 2003, Rellstab et al. 2015). While PD methods

can fail to identify subtle differences in allele frequencies caused by local adaptation, when coupled with genotype-environment association analysis, these approaches can be used to obtain more information on the data (Rellstab et al. 2015). I used two univariate outlier detection methods in this analysis: OUTFLANK (Whitlock & Lotterhos 2015, version 0.2) and PCADAPT (Luu et al. 2016, version 4.3.3) (*Table 3*).

OUTFLANK identifies outliers by finding a trimmed distribution of F_{ST} values from all SNPs in a dataset. Outliers are detected from the tails of this distribution. From the vcf file generated by Platypus, I extracted the genotype data in the form of a character matrix, the positions in base pairs, and chromosome information. Using this information along with the position values, and a list of the population names associated with each individual, I calculated the F_{ST} values for all loci within the dataset using the default option (Whitlock & Lotterhos 2015).

I ran an initial check of plotting heterozygosity by F_{ST} values to determine the presence of neutral loci in the data. I then plotted F_{ST} by the F_{ST} values not corrected for sample size. OUTFLANK requires these F_{ST} values be uncorrected for sample size to fit the F_{ST} distribution to chi-square; correcting for sample size makes the F_{ST} estimate lower than an uncorrected estimate (Whitlock & Lotterhos 2015). Loci deviating from the linear relationship between the F_{ST} and uncorrected F_{ST} values were removed, as they can be misidentified as outliers when they instead come about from small sample sizes.

OUTFLANK then suggests that a set of random and quasi-independent SNPs is used to calculate mean F_{ST} values, as well as degrees of freedom, as non-independent loci can result in an F_{ST} distribution that deviates the expectations of chi-squared (Whitlock & Lotterhos 2015). To thin these SNPs for linkage disequilibrium, I used the program BIGSNPR (Prive et al 2018, version 1.6.1). I converted the data into a code256, the filetype required for input into BIGSNPR. In addition to the code256 data, I also input a list of chromosome names corresponding to each locus. These inputs were used to perform pairwise thinning across the genome (pruning), use summary statistics to incorporate SNP importance (clumping), and remove SNPs within LD regions of the genome (LD thinning) (Prive et al. 2018).

OUTFLANK was run with the thinned data set, estimating neutral FST distribution parameters. I used a threshold value of 0.05 for the q-values and set the minimum heterozygosity for inclusion of a locus' calculation at 0.1. Using the neutral mean FST distribution and degrees

of freedom found for the thinned SNPs, I then applied this analysis to the whole of the dataset to find p- and q-values for all loci. As a result of the LD pruning previously performed, outlier loci within close physical proximity to one another are not likely to be causal (Whitelock & Lotterhos 2015).

I ran PCADAPT as a second avenue for PD testing. PCADAPT is a principal component approach that can be used to detect outliers by looking for a correlation between *K* principal components and genetic variation (Luu et al. 2016). The number of PCs found was based on scree and score plots which limited values of *K* corresponding to relevant levels of background population structure. $K = 2$ was found to be optimal, as beyond $K = 2$ population structure could not be accounted for by greater than 2 principal components (*Figure 3*).

The PCADAPT function, which centers and scales the genotype matrix and then performs a PCA analysis, was run using $K = 2$. Z-scores and p-values were included in the returned object. Manhattan plots displayed -log10 of p-values calculated using PCADAPT. PCADAPT incorporates several options to set a cutoff for outlier detection, and to remain consistent across my analyses, I used the R package QVALUE (Storey et al. 2021, version 2.26.0) to convert p-values from the PCADAPT function into q-values. An alpha value of 0.05 was used as a threshold for expected false discovery rate, and outliers were considered significant if they had a q-value less than alpha. The significant loci from both PD analyses were then compared to one another, and to significant loci identified using genotype-environment association (GEA) testing.

2.2.5.2 | Genotype-Environmental Association Analysis

Genotype-environment association (GEA) testing identifies loci under selection by looking at correlations between genetic data and environmental variables (Forester et al. 2018). Local adaptation results in weak signatures across several loci, which multivariate GEA analyses may be able to detect better than PD methods. When used in tandem, PD and GEA can together reduce the rate of false positives (Rellstab et al. 2015).

LFMM is a program that performs GEA analysis using latent factor mixed modeling (Frichot 2013). Latent factor mixed models are statistical regression models that can test for the correlation between variables of interest (environmental data), and a set of response variables (genotypes or gene expression across individuals). Latent factors, unobserved variables, are

included in the model to correct for confounding effects that could be caused by background population structure (Caye & Francois 2019).

LFMM best balances low error rate with being able to detect weak selection, as opposed to other analyses (Frichot et al. 2013). Using the modeling presented in LFMM, the R package LEA runs large-scale analyses to create ancestry matrices, uses admixture analysis to estimate population structure, scans genomes for signatures of selection, and generates graphs and plots as outputs for the results. LEA also reduces false-positive rates (Frichot & Francois 2015). I used both LEA (version 3.3.2) and LFMM (version 1.0) in analyzing my data for loci under selection.

I ran one initial approach used in LEA, admixture analysis using sparse nonnegative matrix factorization (snmf). The snmf function was run to estimate individual admixture coefficients by assuming *K* ancestral populations. This function provides least-square estimates of ancestry proportions, as opposed to maximum likelihood estimates. It computes an entropy criterion by evaluating the fit of the model to the data. It also uses a cross-validation technique. I projected the snmf results as a cross-entropy plot and in barcharts, looking at projections with *K* equal to 2, 3, 4, and 5 (*Figure 2*). This step was also to check if the data depicted ancestry in the way expected, with the Lake Ontario populations, and those in Lake Erie and the Upper Great Lakes sharing respective common ancestry. As I determined the number of ancestral populations, *K*, was equal to 2 in the cross-entropy plot, I ran the remaining analysis under this assumption.

I fit a latent factor mixed model with $K = 2$ factors using a ridge regression, which is based on minimizing a least-squares problem using an L2 penalty (Caye & Frichot 2019). Using this fitted model, I performed association testing using a calibration of genomic inflation factor (GIF). I used the R package QVALUE (Storey et al. 2021) to convert p-values from the LEA analyses to q-values. An alpha value of 0.05 was used as a threshold for expected false discovery rate, and outliers were considered significant if they had a q-value less than alpha.

I generated Manhattan plots for each of the environmental variables using the R package QQMAN (Turner 2014, version 0.1.4) by plotting the -log10 p-values on the y-axis, and chromosome number on the x-axis (*Figure 4*). SNPs were plotted as points on the graph, displaying a preliminary association between SNPs and environmental variables, with significance shown by peaking -log10 p-values. The outliers identified in the GEA analysis were compared with the outliers of the PD tests to look for a consensus between different methodologies.

2.2.6 | Functional Annotation and Gene Ontology

The identified candidate loci under selection were annotated with Variant Effect Predictor from Ensembl Release 102 (McLaren et al. 2016), with the VGP assembly and annotation of sea lamprey genome as the reference. This allowed us to identify specific sea lamprey genes that the candidate loci could be associated with. Broadly, the genes could be divided in two categories: those that had been given a specific name based on orthology relationship with other vertebrates (examples: 'IGFBP7' and 'CENPE'), and genes for which the exact relationship was less clear (named in a format similar to LOC000012345, sometimes with descriptions such as 'neurexin-1-beta-like' or 'cubilin-like'). In the next step I specifically focused on the first category, the named genes.

The Panther Classification System (Mi et al. 2019) was used to conduct a statistical overrepresentation test (Released 2021.02.24) for gene ontology (GO) terms associated with the gene symbols. PANTHER (Protein Analysis Through Evolutionary Relationships, https://patherdb.org) classifies genes based on evolutionary history data, as well as their molecular and biological functions. As no GO annotation of genes from the VGP sea lamprey genome assembly was available, I ran the list of gene symbols against a reference list of all *Homo sapiens* genes in the Panther database (Ashburner et al. 2000, GO Ontology database 2021). *Homo sapiens* was used as a reference because the known IDs from lamprey genes follow the naming convention used by gene IDs of humans.

I produced PANTHER GO-Slim Biological Process for the resulting annotation dataset, using a Fisher's Exact test type and a correction to calculate FDR (*Table 4*). This step converted uniquely mapped IDs to PANTHER Ontology Terms and determined their biological and molecular functions. PANTHER outputs two types of category IDs: IDs with a prefix of 'PC', Panther's unique ontology terms, and IDs with a prefix of 'GO', GO terms. I removed PANTHER's unique category IDs, as I exclusively used GO terms in my subsequent gene ontology enrichment analysis. Additionally, some IDs could not be mapped to the reference genome or were duplicates; these were not included in the resulting list of GO terms.

From all the GO terms identified, I isolated a list of terms associated with two or more statistical approaches (PCADAPT x OUTFLANK, LFMM x OUTFLANK, PCADAPT x LFMM, or GO terms found across all three approaches). I next input the list of GO terms associated with two or more analyses into REVIGO (Reduce and Visualize Gene Ontology

terms) (Supek et al. 2011, https://revigo.irb.hr). REVIGO uses a computational approach to summarize GO terms, remove redundancy in GO terms with similar functions, and visualize the remaining terms. Any hierarchical GO terms were condensed and simplified. I allowed a medium 'allowed similarity' between GO terms (0.7) and ran the terms against the whole UniProt-to-GO mapping database from the EBI GOA project (*Table 5*). SimRel was the semantic similarity measure selected for the analysis, which uses a common ancestor to inform degree of similarity (Supek et al. 2011). The Revigo analysis produced a list of overrepresented functions, which showed possible biological functions under selection.

2.3 | Results

2.3.1 | SNPs

Our dataset had 6,374,517 SNPs and a mean depth of 15x reads per locus. After filtering out the genetic data for <5% minimum allele frequency (MAF), I was left with 2,178,703 SNPs. By further removing loci with missing data and filtering out non-biallelic sites, I retained 1,486,675 high quality biallelic loci for the final analysis, genotyped in 209 individuals across 13 sampling locations. Analysis of population structure revealed a pattern that was best supported by 2 main lineages across the 13 sampling locations. This genetic structuring correlated very well with geography and historical knowledge of the sea lamprey in the Great Lakes, with one cluster in Lake Ontario and the second in lakes above Niagara Falls, that is, Lake Erie and the upper Great Lakes (Docker et al. 2021).

2.3.2 | Genome Signatures of Selection

Outflank found 153 significant outliers identified as having a q-value less than 0.05. These outlier SNPs from OUTFLANK had a mean F_{ST} value of 0.012, a low differentiation value suggesting these SNPs could be a result of balancing selection, during which polymorphic genes are favored, in turn lowering genetic diversity as measured by F_{ST} (Muirhead 2001).

Of these 153 SNPs, 83 are physically linked to known genes; the criteria for a SNP being physically linked to a gene was location within 5 kb of a gene annotated in the VGP genome assembly, which includes exons, introns, untranslated regions, and the promoter region. A total of 48 of the 153 SNPs were located in intergenic regions greater than 5 kb away from coding sequences, and the remaining 22 SNPs were not annotated. Of the 83 SNPs physically linked to genes, 17 of these genes have an identically named homolog in the human reference genome.

From these 17 identified genes, 47 GO terms were derived for subsequent gene ontology enrichment analysis.

PCADAPT found 3,765 significant outliers identified as having a q-value less than 0.05. 2,408 of these SNPs are physically linked to known genes, using the same criterion listed above. 855 of these SNPs were located in intergenic regions greater than 5 kb away from coding sequences, and the remaining 502 SNPs were not identified as having been annotated. Of the 2,408 SNPs physically linked to genes, 251 of these genes have an identically named homolog in the human reference genome. From these 251 identified genes, 515 GO terms were derived for subsequent gene ontology enrichment analysis.

2.3.3 | Phenotype-genotype Associations

When each environmental variable was run in its own LFMM model, 461 SNPs were identified as being significantly linked to an environmental gradient with a q-value less than 0.05. One of these SNPs was associated with lake depth but was located in an intergenic region farther than 5 kb from a known coding region. The remaining 460 significant outliers were associated with human population density; 328 of these were physically linked to genes, as they were within a known coding region or within 5 kb upstream or downstream of a known coding region. 99 SNPs were intergenic and not affiliated with known coding regions, and 33 SNPs were not annotated. Of the 328 SNPs linked to genes, 36 have an identically named homolog in the human reference genome. 131 GO terms were derived from these 36 genes.

2.3.4 | Genetic Basis of Phenotypic Differentiation

Of the 4,379 unique SNPs identified, none had an overlap across all three methods used. The significant SNPs identified by OUTFLANK and LFMM had no overlap for either environmental variable used in LFMM analysis; however, OUTFLANK and PCADAPT had 2 overlapping significant SNPs. These were located on chromosomes 37 (NC_046105) and 51 (NC-046119). PCADAPT and LFMM (the human population density model) had an overlap of 23 significant SNPs located across 15 chromosomes. The common outliers between these analyses (*Figure 6*), despite lacking overlap between all three methods, show repeatability of finding outliers across methods (Andrew et al. 2018, de Jong 2021).

Though overlap between SNPs identified in the three analyses was limited, there was additional overlap between candidate genes associated with these SNPs, as well as with GO terms derived from candidate genes across all three methods (*Figure 7*). While OUTFLANK and

LFMM did not share any overlapping candidate genes, these two approaches shared 18 GO terms, suggesting selection detected by these methods may be acting on the same biological functions. The OUTFLANK and PCADAPT approaches shared 3 specific candidate genes (CFAP52, EYS, and NCAM1), and 28 GO terms. These overlapping GO terms were in part from the overlapping 3 candidate genes, and genes that were not identical between approaches but shared similar biological functions. Finally, the PCADAPT and LFMM approaches shared 7 overlapping candidate genes (AGK, ARMC2, CDC6, GAB1, MAT2B, MCTP1, and USP47) and 77 overlapping GO terms.

I identified 89 total GO terms that were shared by gene sets identified by at least two statistical approaches (*Table 6*), including 17 GO terms that overlapped between all three approaches. Phosphorylation (16310), protein phosphorylation (6468), protein localization to cilium (61512), intraciliary transport involved in cilium assembly (35735), ciliary transition zone assembly (1905349), and vesicle targeting (97112) were shared across approaches, and 28 additional GO terms were shared by 2 approaches, pointing to possible biological functions under selection (*Figure 8*).

2.4 | Discussion

2.4.1 Evidence of Local Adaptation

Our study indicates that invasive sea lampreys have, after relatively few generations since their widespread invasion of the Great Lakes in the early-to-mid 1900s, undergone adaptive divergence. Sea lampreys have generation times of approximately 7–11 years from larval stage to the end of the adult stage (Hardisty & Potter 1971). Thus, invasive sea lampreys have likely undergone local adaptation in as few as nine to eleven generations, contributing to a growing body of literature suggesting evolution may occur significantly more quickly than on a Darwinian timescale.

Invasive sea lamprey populations do not home to their natal streams, and as a result have previously been thought to exhibit little genetic diversity across their non-native ranges (Waldman et al. 2006). Along with sustaining a more homogenous population, lack of natal homing behavior can lead to more widespread dispersal, which may have helped the sea lamprey to invade successfully (Bryan et al. 2005, Docker et al. 2021). Despite substantial evidence of gene flow in invasive populations of sea lamprey (Bryan et al. 2005) due to lack of natal homing (Waldman et al. 2006), my results display evidence of divergent natural selection. Selective

pressures among these populations have likely been high enough to lead to low levels of local adaptation.

Because local adaptation is shaped by historical and contemporary selective pressures, studies of local adaptation require a comparison between locally adapted populations and their ancestors (Kawecki & Ebert 2004, Williams 1966). In my study, I compared the populations of Lake Ontario, which have been in the Great Lakes system for the longest amount of time either because they are native or because they colonized the lake long ago, to those of Lake Erie and the upper Great Lakes.

Regardless of the origin of Lake Ontario sea lamprey, these animals are ancestral to those that later invaded the rest of the Great Lakes. With the construction of the Welland Canal in 1829, the sea lamprey that bypassed Niagara Falls to invade Lake Erie diverged directly from the Lake Ontario populations (Docker & Potter 2019, Marsden & Siefkes 2019). Such divergence continued as sea lamprey subsequently and quickly colonized the upper Great Lakes. This is evident by ancestry matrices that show a clear distinction between the Lake Ontario populations and the rest of the lakes (*Figure 2*). My findings of two principal components for genetic structuring across the 13 sampled populations (*Figure 3)* was consistent with previous findings (Docker et al. 2021).

It is possible that, because of how quickly sea lamprey propagated in Lake Erie and the upper Great Lakes, and because these populations intermixed with each other, further local adaptation has been slowed by consistent gene flow (Bryan et al. 2005). Local adaptation typically refers to populations that are at connected by gene flow and dispersal, and not populations with total spatial divergence (Kawecki & Ebert 2004), so it is possible local adaptation is still occurring at a less rapid pace following initial invasion. In the absence of natal homing, Pacific lamprey were still found to be limited in dispersal due to geographic distance (Spice et al. 2012); it is likely sea lamprey in the Great Lakes are similarly constrained by isolation-by-distance, resulting in some degree of genetic structuring.

2.4.2 Outlier Detection

OUTFLANK identified significant outliers with a mean F_{ST} value of 0.012, a low differentiation value suggesting these SNPs could be a result of balancing selection. Balancing selection, from the 'balancing hypothesis' of Theodosius Dobzhansky, suggests that selection favors heterozygous traits, and that many genes are polymorphic (Dobzhansky 1982). In addition

to heterozygote advantage, balancing selection can also include frequency-dependent selection in which rarer genes are favored, and selection that varies with temporal or spatial variation (Hendrick 2007). Balancing selection can be identified by observing deviations from the neutral expectation. If divergence is less than one would expect, balancing selection may act across populations; if more than expected, balancing selection may act on spatial divergence (Hendrick 2007). Thus, the low F_{ST} value from my OUTFLANK analysis suggests that multiple alleles of genes are being maintained across the sampled populations.

2.4.3 Genotype-Environment Associations

A novel result of my study was identification of local adaptation correlated with human population density (HPD) above all other environmental variables tested, including localized lake depth and annual vertical temperature. By combining the results from the LFMM analysis with that of PCADAPT, I was able to find 2 statistical outliers that overlapped between the GEA and PD methods. Furthermore, the 7 candidate genes found by both LFMM and PCADAPT, and the 3 candidate genes found by PCADAPT and OUTFLANK, point to possible genes under selection. This finding supports my hypothesis that humans are drivers of selection in invasive sea lamprey populations and aligns with the theory that human activity accelerates evolutionary change in pest species (Palumbi 2001).

The GO terms I identified, under the criteria of being shared by two or more analytical methods (*Table 6*), bolster this evidence of selection due to the high degree of overlap in biological functions. In some instances, different detection methods identified paralogs of the same pathways. For example, LFMM and PCADAPT identified SNPs linked to genes GGT7 and GGT1, respectively; these paralogs contribute to the same biological function, cellular catabolic processes (GO:0044248). The detection tools may lack the precision to identify specific genes within this pathway under selection. However, gene ontology is considered reliable and a standard when working with sequenced genomes of non-model organisms (Škunca et al. 2012). While I was unable to repeatedly pinpoint specific genes under selection, looking at GO terms shared between multiple detection methods allowed us to broadly hypothesize cellular and biological pathways under selection.

The BMP (bone morphogenic protein) signaling pathway (GO:0030509) was identified as a possible biological function under selection by both PCADAPT (TWSG1 gene) and LFMM (GDF10 gene). The BMP signaling pathway can contribute to “patterning of skeletal elements”,

including formation of bone and cartilage (Hoffmann & Gross 2001), and heart development (Rivera-Feliciano & Tabin 2006). Invasive populations of sea lamprey exhibit phenotypic differentiation in body size across the Great Lakes population (Docker et al. 2021); while this could be a plastic response to ecological and/or environmental conditions, it is also possible that identification of the BMP signaling pathway could point to body size because of local adaptation. In a series of studies on Pacific lamprey, Hess et al. (2013) found adaptive markers linked to genes and biological functions associated with body size and skeletal formation, including calcium-dependent cell-cell adhesion, skeletal development, and cartilage development. These finding supports my own and suggest migratory lampreys may become adapted to conditions experienced during migration, such as the length of the spawning run.

Negative regulation of immune system process (GO:0002683) and lymphocyte migration (GO:0072676) were biological functions found by both PCADAPT (LRCH1 gene) and LFMM (GAL gene); these functions regulate immune responses. In terms of human population density as a driver of selection, it is possible that pollution in and around densely populated areas, or use of lampricides, has led to locally adapted populations with differing immune responses. Areas with high human population density also have more aquaculture systems that can lead to considerable numbers and diversity of bacteria in waterways (Watts et al. 2017). In addition to bacterial pathogens in waterways, lampreys are also susceptible to infection from metazoan parasites such as trematodes, cestodes, and acanthocephalans that can exacerbate disease (Shavalier et al. (in press)).

Invasive species have been documented as able to quickly adapt to climatic gradients (Colautti & Barrett 2013, Alford et al. 2009, Prentis et al. 2008), and are exemplary systems in studying rapid evolution (Huey et al. 2005). A number of factors give invasive populations the ability to adapt, including release from predation in new environments, interspecific hybridization with other non-native populations, and selection for greater dispersal capacity (Lee 2002). As locally adapted invasive species can be more difficult to eradicate, it is critical to understand which, if any, environmental features sea lamprey have locally evolved to coincide with. My findings support prior evidence suggesting that co-evolution with humans and human disturbances is correlated with successful invasion events and subsequent spreading of invasive species (Jeschke & Strayer 2006). Experiencing rapid, contemporary evolution may hold an

advantage over competing native species in response to climate change-driven habitat disturbance (Colautti & Lau 2015).

2.4.4 Adaptive Potential Despite Gene Flow

Local adaptation is not limited to species in patchy environments and can be found in randomly selected populations across a continuous range of environmental features (Kawecki & Ebert 2004). In marine species, local adaptation has been historically thought to occur over expansive environmental gradients, due to the high mobility of many marine species that are not recruited into their source populations. Samples from neutral genetic markers of marine animals have revealed low genetic diversity, specifically lower degrees of genetic subpopulation differentiation (Ward et al. 1994). However, among-population genetic variation (Q_{ST}) has been found to be significantly higher than in neutral markers (F_{ST}) for many marine species, indicating a strong likelihood of local adaptation especially with physiological and morphological traits (Conover et al. 2006). This phenomenon could be in part due to natal homing in some marine fishes; homing behavior results in reduced gene flow between populations and higher instances of local adaptation (McDowall 2001). However, this process could also be occurring despite consistent gene flow.

The Atlantic silverside (*Menidia menidia*), for example, is a marine fish with repeated documentation of local adaptation despite extensive dispersal during the species' offshore migration in the winter (Conover & Murawski 1982). Atlantic silverside populations located at higher latitudes have shorter growing seasons than lower-latitude populations, and a number of other traits, such as growth efficiency (Present & Conover 1992) and energy allocation (Billerbeck et al. 2000), also co-vary with latitude. A study by Clarke et al. (2010) found that Atlantic silverside populations had high levels of connectivity despite this adaptive divergence; while the highest proportion of fish from their study were captured at their origin, a significant proportion dispersed distances greater than 200 kilometers.

Another example of this are salmonids, which exhibit local adaptation over small spatial scales despite gene flow from dispersing individuals (Fraser et al. 2011). A study of introduced sockeye salmon (*Oncorhynchus nerka*) by Hendry et al. (2000) found that beach and river populations of salmon were genetically distinct from one another, despite evidence of a subset of individuals migrating from the river to the beach. The study concluded that immigrants from the river had unfavorable body depth among the beach population, and likely had reduced mating

success as a result. This led to reproductive isolation despite geographic connectivity between the two populations, in fewer than 13 generations (Hendry et al. 2000). As illustrated by the sockeye salmon, a possible explanation for local adaptation despite population connectivity is that in some cases, philopatry is associated with higher fitness than is dispersal (Peterson et al. 2014, Bensch et al. 1998, Bélichon et al. 1996). A dispersed individual in a locally adapted population may have low reproductive success, reducing the effect of gene flow and preserving locally adapted traits (Nosil et al. 2005).

Additionally, while gene flow can reduce rates of local adaptation, it also has the potential to spread advantageous genes. As discussed with pesticide resistance, ongoing gene flow can spread new genes and gene combinations throughout all subpopulations, leading to rapid evolution and higher rates of local adaptation (Slatkin 1987).

2.4.5 Future Areas of Study

A further avenue of study could be an investigation of invasive sea lamprey outside of the Laurentian Great Lakes; while I did collect samples from Lakes Seneca and Cayuga, which are the two largest Finger Lakes in upstate New York, localized environmental data was insufficient to include these samples in my analysis. The 11 Finger Lakes of New York are long, thin bodies of water; Seneca and Cayuga each have populations of sea lamprey (Marsden & Siefkes 2019). Both lakes drain into Lake Ontario, making gene flow between the Finger Lakes and Lake Ontario populations possible. It's unclear whether this gene flow is occurring; however, the Seneca and Cayuga populations are slightly genetically distinct from Lake Ontario (Bryan et al. 2005). Comparing freshwater populations of sea lamprey to those in the Atlantic Ocean could similarly have value; adding sample sites outside of the Laurentian Great Lakes could broaden the scope of environmental variability and further illustrate the adaptive capacity of populations of invasive sea lampreys.

I selected variables through the lens of juvenile sea lamprey, which exist in the upper trophic levels of the Great Lakes ecosystem as piscivores, rather than larval sea lamprey which are more reliant on high primary productivity as primary-consumer detritus-feeders (Hayden et al. 2019). Therefore, it is possible primary productivity influences juvenile stages differently than larval stages. Depth is expected to have similar variation in influence. Juvenile lampreys have been captured at a range of depth in lakes and streams ranging from the deepest fished areas to the surface (Johnson & Anderson 1980), with evidence that male and female lampreys have

slight seasonal differences in the depths at which they feed (Hume et al. 2021). Contrarily, larvae burrow in sediment in smaller tributaries. Further analysis of larval lampreys and possible local adaptation to their ecosystems could find that larvae are adapted to an entirely different set of environmental variables. Due to the varying diet, dispersal, and other factors between the life stages of sea lamprey, it is conceivable that environmental pressures will act differently on larval, juvenile, and adult (breeding) animals.

Adaptation can also occur to biotic variables, such as mutualism, commensalism, or predation leading to coevolution (Hoeksema & Forde 2008). A survey of the taxon richness of lampreys' prey base across the Laurentian Great Lakes could determine whether composition of prey fishes drive local adaptation. Phenotypic differences have been documented as variable with prey abundance; stocking of salmonids in the Great Lakes led to larger body sizes of sea lamprey, and lake trout abundance was found to correlate with sea lamprey weight values (Heinrich et al. 1980). It is unclear whether these changes are plastic or adaptive responses but determining whether prey abundance and richness are selective pressures merits future study.

Literature Cited

Adams, J., Birceanu, B., Chadderton, L., Jones, M., Lepak, J., Seilheimer, T., Steeves, T., Sullivan, P, & Wingfield, J. (in press). Trade-offs between suppression and eradication of sea lampreys from the Great Lakes. *Journal of Great Lakes Research*.

Alford, R. A., Brown, G. P., Schwarzkopf, L., Phillips, B. L., & Shine, R. (2009). Comparisons through time and space suggest rapid evolution of dispersal behaviour in an invasive species. *Wildlife Research*, 36(1), 23-28.

Andersen, M. C., Adams, H., Hope, B., & Powell, M. (2004). Risk assessment for invasive species. *Risk Analysis: An International Journal*, 24(4), 787-793.

Andrew, S. C., Jensen, H., Hagen, I. J., Lundregan, S., & Griffith, S. C. (2018). Signatures of genetic adaptation to extremely varied Australian environments in introduced European house sparrows. *Molecular Ecology,* 27(22), 4542-4555.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Applegate, V. C. (1950). Natural history of the sea lamprey, *Petromyzon marinus*, in Michigan (Doctoral dissertation, University of Michigan).

Applegate, V. C. (1951). The sea lamprey in the Great Lakes. *The Scientific Monthly*, 72(5), 275-281.

Applegate, V. C., Howell, J. H., Moffett, J. W., & Smith, M. A. (1961). Use of 3-trifluormethyl-4-nitrophenol as a selective sea lamprey larvicide. *Great Lakes Fishery Commission. Technical Report*, (1), 1-35.

Ashburner et al. (2000). Gene ontology: tool for the unification of biology. *Natural Genetics,* 25(1), 25-29.

Avenetti, L. D., Robinson, A. T., & Cantrell, C. J. (2006). Short-term effectiveness of constructed barriers at protecting Apache trout. *North American Journal of Fisheries Management.*, 26, 213–216.

Beamish, F.W.H. & Potter, I.C. The biology of the anadromous sea lamprey (*Petromyzon mannus)* in New Brunswick. *Journal of Zoology,* 177(1), 57-72.

Bélichon, S., Clobert, J., & Massot, M. (1996). Are there differences in fitness components between philopatric and dispersing individuals? *Acta Oecologia*, 17(6), 503-517.

Bensch, S., Hasselquist, D., Nielsen, B., & Hansson, B. (1998). Higher fitness for philopatric than for immigrant males in a semi-isolated population of great reed warblers. *Evolution*, 52(3), 877-883.

Billerbeck, J. M., Schultz, E., & Conover, D. (2000). Adaptive variation in energy acquisition and allocation among latitudinal populations of the Atlantic silverside. *Oecologia*, 122: 210– 219.

Birceanu, O., McClelland, G., Wang, Y., Brown, J., & Wilkie, M. (2011) The lampricide 3-trifluoromethyl-4-nitrophenol (TFM) uncouples mitochondrial oxidative phosphorylation in both sea lamprey (*Petromyzon marinus)* and TFM-tolerant rainbow trout (*Oncorhynchus mykiss). Comparative Biochemistry and Physiology,* 153(3), 342-349.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

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.

Caprio, M.A., & Tabashnik, B. (1992). Gene flow accelerates local adaptation among finite populations: simulating the evolution of insecticide resistance. *Journal of Economic Entomology*, 85(3): 611-620.

Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution*, 36(4), 852–860.

Center for International Earth Science Information Network - CIESIN - Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

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, 1-14.

Clarke, L., Munch, S., Thorrold, S., & Conover, D. (2010). High connectivity among locally adapted populations of a marine fish (*Menidia menidia)*. *Ecology,* 91(12), 3526-3537.

Cline, T., Kitchell, J., Bennington, V., McKinley, G., Moody, E., & Weidel, B. (2014). Climate impacts on landlocked sea lamprey: Implications for host-parasite interactions and invasive species management. *Ecosphere,* 5(6), 68.

Colautti, R. I., & Barrett, S. C. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, 342(6156), 364-366.

Colautti, R. I., & Lau, J. A. (2015). Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, 24(9), 1999-2017.

Conover, D. O. & S. A. Murawski. (1982). Offshore winter migration of the Atlantic silverside, *Menidia menidia*. *Fishery Bulletin,* 80: 145– 150.

Conover, D., Clarke, L., Munch, S., & Wagner, G. (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implication for conservation. *Journal of Fish Biology,* 69, 21-47.

Coop G., Witonsky D., Di Rienzo A., & Pritchard J.K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, 185(4), 1411-1423.

Darwin, C. (1859) On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. London: J. Murray. [Pdf] Retrieved from the Library of Congress.

de Jong, M. J., Lovatt, F., & Hoelzel, A. R. (2021). Detecting genetic signals of selection in heavily bottlenecked reindeer populations by comparing parallel founder events. *Molecular Ecology*, 30(7), 1642-1658.

Derosier, A., Jones, M., & Scribner, K. (2007) Dispersal of sea lamprey larvae during early life: relevance for recruitment dynamics. *Environmental Biology of Fishes,* 78(3), 271-284.

Dobzhansky, T. (1982). Genetics and the Origin of Species (No. 11). Columbia university press.

Docker, M. F., & Potter, I. C. (2019). Life history evolution in lampreys: alternative migratory and feeding types. In Lampreys: Biology, Conservation and Control (pp. 287-409). Springer, Dordrecht.

Docker, M. F., Bravener, G. A., Garroway, C. J., Hrodey, P. J., Hume, J. B., Johnson, N. S., ... & Zollweg-Horan, E. C. (In press). A review of sea lamprey dispersal and population structure in the Great Lakes and the implications for control. *Journal of Great Lakes Research*.

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.

Earn, A., Bucci, K., & Rochman, C. M. (2020). A systematic review of the literature on plastic pollution in the Laurentian Great Lakes and its effects on freshwater biota. *Journal of Great Lakes Research,* 47(1), 120-133.

Edelaar, P., Siepielski, A. M., & Clobert, J. (2008). Matching habitat choice causes directed gene flow: a neglected dimension in evolution and ecology. *Evolution,* 62(10), 2462-2472.

Eshenroder, R. L. (2014). The role of the Champlain Canal and Erie Canal as putative corridors for colonization of Lake Champlain and Lake Ontario by sea lampreys. *Transactions of the American Fisheries Society*, 143(3), 634-649.

ESRI 2011. ArcGIS Desktop: Release 10.3.1. Redlands, CA: Environmental Systems Research Institute.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016) Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048.

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, 27(9), 2215-2233.

Fox, M.D. & Fox, B.J. 1986. The susceptibility of natural communities to invasion. In Ecology of biological invasions: an Australian perspective, eds. R.H. Groves and J.J. Burdon, 57–66. Canberra: Australian Academy of Sciences.

Franks, S. J., & Munshi-South, J. (2014). Go forth, evolve, and prosper: the genetic basis of adaptive evolution in an invasive species. *Molecular Ecology*, 23(9), 2137-2140.

Fraser, D. J., Weir, L. K., Bernatchez, L., Hansen, M. M. & Taylor, E. B. (2011). Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, 106, 404–420.

Frichot, E. & Francois, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution,* 6(8), 925-929.

Frichot, E., Schoville, S., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution,* 30(7), 1687-1699.

Garroway, C. J., Radersma, R., Sepil, I., Santure, A. W., De Cauwer, I., Slate, J., & Sheldon, B. C. (2013). Fine-scale genetic structure in a wild bird population: The role of limited dispersal and environmentally based selection as causal factors. *Evolution*, 67(12), 3488-3500.

Gillman, L., Wright, S., Cusens, J., McBridge, P., Malhi, Y., & Whittaker, R. (2014). Latitude, productivity, and species richness. *Global Ecology & Biogeography,* 24(1), 107-117.

Great Lakes Fishery Commission, & Christie, W. J. (1973). Review of the Changes in the Fish species composition of Lake Ontario.

Gross, M., Coleman, R., & McDowall, R. (1988) Aquatic productivity and the evolution of diadromous fish migration. *Science,* 239(4845), 1291-1293.

Guo, J., Salamova, A., Venier, M., Dryfhout-Clark, H., Alexandrou, N., Backus, S., & Hites, R. A. (2018). Atmospheric flows of semi-volatile organic pollutants to the Great Lakes estimated by the United States' integrated atmospheric deposition and Canada's Great

Lakes basin monitoring and surveillance networks. *Journal of Great Lakes Research*, 44(4), 670-681.

Hardisty, M., & Potter, I. (1971). The general biology of adult lampreys. In: The Biology of Lampreys (ed. MW Hardisty, IC Potter), 1, 127– 206. Academic Press, New York.

Hayden, B., Ferron, M., Cunjak, R. A., & Samways, K. (2019). Fruit of the forest-larval sea lamprey Petromyzon marinus are fueled by allochthonous resources. *Journal of Fish Biology*, 95(3), 781-792.

Hedrick, P. W. (2007). Balancing selection. *Current Biology*, 17(7), R230-R231.

Heinrich, J. W., Weise, J. G., & Smith, B. R. (1980). Changes in biological characteristics of the sea lamprey (*Petromyzon marinus*) as related to lamprey abundance, prey abundance, and sea lamprey control. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(11), 1861-1871.

Hendry, A. P., Wenburg, J. K., Bentzen, P., Volk, E. C., & Quinn, T. P. (2000). Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. *Science*, 290(5491), 516-518.

Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F., & Narum, S. R. (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, 22(11), 2898-2916.

Hoeksema, J., & Forde, S. (2008) A meta-analysis of factors affecting local adaptation between interacting species. *American Naturalist* 171(3), 275-290.

Hoffmann, A., & Gross, G. (2001). BMP signaling pathways in cartilage and bone formation. *Critical Reviews™ in Eukaryotic Gene Expression*, 11(1-3).

Huey, R., Gilchrist, G., & Hendry, A. (2005). Using invasive species to study evolution. *Species invasions: Insights to ecology, evolution, and biogeography*, 139-164.

Hume, J. B., Bravener, G. A., Flinn, S., & Johnson, N. S. (In press). What can commercial fishery data in the Great Lakes reveal about juvenile sea lamprey (*Petromyzon marinus*) ecology and management? *Journal of Great Lakes Research*.

Jeschke, J. M., & Strayer, D. L. (2006). Determinants of vertebrate invasion success in Europe and North America. *Global Change Biology*, 12(9), 1608-1619.

Johnson, B. G. H., & Anderson, W. C. (1980). Predatory-phase sea lampreys (*Petromyzon marinus*) in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(11), 2007-2020.

Kawecki, T. & Ebert, D. (2004) Conceptual issues in local adaptation. *Ecology Letters,* 7(12), 1225-1241.

King, N. J., Bagley, P. M., & Priede, I. G. (2006). Depth zonation and latitudinal distribution of deep-sea scavenging demersal fishes of the Mid-Atlantic Ridge, 42 to 53 N. *Marine Ecology Progress Series*, 319, 263-274.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357.

Lavis, S., Hallett, A., Koon, E., McAuley, T. (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(1), 362-372.

Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, 17(8), 386-391.

Lennox, R. J., Bravener, G. A., Lin, H. Y., Madenjian, C. P., Muir, A. M., Remucal, C. K., ... & Cooke, S. J. (2020). Potential changes to the biology and challenges to the management of invasive sea lamprey Petromyzon marinus in the Laurentian Great Lakes due to climate change. *Global Change Biology*, 26(3), 1118-1137.

Lewandowska, A. M., Breithaupt, P., Hillebrand, H., Hoppe, H. G., Jürgens, K., & Sommer, U. (2012). Responses of primary productivity to increased temperature and phytoplankton diversity. *Journal of Sea Research*, 72, 87-93.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4(12), 981-994.

Luu, K., Bazin, E., & Blum, M. (2016). Pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67-77.

Lyons, K. & Willig, M. (2002). Species richness, latitude, and scale-sensitivity. *Ecology*, 83(1), 47-58.

Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics,* 10, 565–577.

Magnuson, J. J., Webster, K. E., Assel, R. A., Bowser, C. J., Dillon, P. J., Eaton, J. G., ... & Quinn, F. H. (1997). Potential effects of climate changes on aquatic systems: Laurentian Great Lakes and Precambrian Shield Region. *Hydrological Processes*, 11(8), 825-871.

Marsden, J. E., & Siefkes, M. J. (2019). Control of invasive sea lamprey in the Great Lakes, Lake Champlain, and Finger Lakes of New York. In Lampreys: biology, conservation, and control (pp. 411-479). Springer, Dordrecht.

Mason, S. A., Daily, J., Aleid, G., Ricotta, R., Smith, M., Donnelly, K., ... & Hoffman, M. J. (2020). High levels of pelagic plastic pollution within the surface waters of Lakes Erie and Ontario. *Journal of Great Lakes Research*, 46(2), 277-288.

Mazza, G., Tricarico, E., Genovesi, P., & Gherardi, F. (2014). Biological invaders are threats to human health: an overview. *Ethology Ecology & Evolution*, 26(2-3), 112-129.

McConville, M.B., Hubert, T., & Remucal, C. (2016). Direct photolysis rates and transformation pathways of the lampricides TFM and niclosamide in simulated sunlight. *Environmental Science & Technology*, 50(18): 9998-10006.

McDowall, R. M. (2001). Diadromy, diversity and divergence: implications for speciation processes in fishes. *Fish and Fisheries*, 2(3), 278-285.

McLaughlin, R. L., Smyth, E. R., Castro-Santos, T., Jones, M. L., Koops, M. A., Pratt, T. C., & Vélez-Espino, L. A. (2013). Unintended consequences and trade-offs of fish passage. *Fish and Fisheries*, 14(4), 580-604.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... & Cunningham, F. (2016). The ensemble variant effect predictor. *Genome Biology*, 17(1), 1-14.

Meckley, T. D., Wagner, C. M., & Gurarie, E. (2014). Coastal movements of migrating sea lamprey (*Petromyzon marinus*) in response to a partial pheromone added to river water: implications for management of invasive populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 71(4), 533-544.

Medley, K. A., Westby, K. M., & Jenkins, D. G. (2019). Rapid local adaptation to northern winters in the invasive Asian tiger mosquito *Aedes albopictus*: a moving target. *Journal of Applied Ecology*, 56(11), 2518-2527.

Mi, Huaiyu, et al. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* 47(1): 419-426.

Michigan Sea Grant. (2018). About the Great Lakes. Michigan Sea Grant, Ann Arbor, MI, USA.

Mollot, G., Pantel, J. H., & Romanuk, T. N. (2017). The effects of invasive species on the decline in species richness: a global meta-analysis. *Advances in Ecological Research*, 56, 61-83.

Morman, R. H., Cuddy, D. W., & Rugen, P. C. (1980). Factors influencing the distribution of sea lamprey (*Petromyzon marinus*) in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(11), 1811-1826.

Moser, M. L., Almeida, P. R., Kemp, P. S., & Sorensen, P. W. (2015). Lamprey spawning migration. In Lampreys: biology, conservation, and control (pp. 215-263). Springer, Dordrecht.

Muirhead, C. A. (2001). Consequences of population structure on genes under balancing selection. *Evolution*, 55(8), 1532-1541.

Mullu, D. (2016). A review on the effect of habitat fragmentation on ecosystem. *Journal of Natural Sciences Research,* 6(15), 1-15.

NOAA National Geophysical Data Center. 1999: Bathymetry of Lake Erie and Lake Saint Clair. NOAA National Centers for Environmental Information.

NOAA National Geophysical Data Center. 1999: Bathymetry of Lake Huron. First. NOAA National Centers for Environmental Information.

NOAA National Geophysical Data Center. 1999: Bathymetry of Lake Ontario. First. NOAA National Centers for Environmental Information.

NOAA National Geophysical Data Center. 1999: Bathymetry of Lake Superior (incomplete). NOAA National Centers for Environmental Information.

Nosil, P., Vines, T. H. & Funk, D. J. (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, 59, 705–719.

Palumbi, S. R. (2001). Humans as the world's greatest evolutionary force. *Science*, 293(5536), 1786-1790.

Peterson, D. A., Hilborn, R., & Hauser, L. (2014). Local adaptation limits lifetime reproductive success of dispersers in a wild salmon metapopulation. *Nature Communications*, 5(1), 1-7.

"Picard Toolkit." (2018). Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/; Broad Institute

Potter, I.C. (1980). Ecology of larval and metamorphosing lampreys. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(11), 1641-1657.

Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends in Plant Science*, 13(6), 288–294.

Present, T. M. C., & Conover, D. (1992). Physiological basis of latitudinal growth differences in *Menidia menidia*: Variation in consumption or efficiency? *Functional Ecology*, 6: 23–31.

Privé, F., Aschard, H., Ziyatdinov, A., & Blum, M. G. (2018). Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics*, 34(16), 2781-2787.

Privé, F., Aschard, H., Ziyatdinov, A., & Blum, M.G. (2018). Efficient analysis of large-scale genome-wide data with two R packages: bigstatr and bigsnpr. *Bioinformatics,* 34(16), 2781-2787.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J. & Sham P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.

Quinn, T. P., Wetzel, L., Bishop, S., Overberg, K., & Rogers, D. E. (2001). Influence of breeding habitat on bear predation and age at maturity and sexual dimorphism of sockeye salmon populations. *Canadian Journal of Zoology*, 79, 1782-1793.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL R-project.org.

Rellstab, C., Gugerli, F., Eckert, A., Hancock, A., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology,* 24(17), 4348-4370.

Ricciardi, A., & Rasmussen, J. B. (1999). Extinction rates of North American freshwater fauna. *Conservation Biology*, 13(5), 1220-1222.

Rimmer, A., Phan, H., Mathieson, I., Lunter, G., & McVean, G. (2013). Platypus: A haplotype-based variant caller for next generation sequence data.

Rivera-Feliciano, J., & Tabin, C. J. (2006). Bmp2 instructs cardiac progenitors to form the heart-valve-inducing field. *Developmental Biology*, 295(2), 580-588.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., ... & Weller, S. G. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics*, 32(1), 305-332.

Savolainen, O., Lascoux, M. & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Review Genetics*, 14, 807–820.

Schwab, D. J., & Bedford, K. W. (1999). The Great Lakes forecasting system. *Coastal and Estuarine Studies*, 157-174.

Shavalier, M., Faisal, M., Moser, M., & Loch, T. (in press) Parasites and microbial infections of lamprey (order Petromyzontiformes Berg 1940): A review of existing knowledge and recent studies. *Journal of Great Lakes Research.*

Siefkes, M. (2017). Use of physiological knowledge to control the invasive sea lamprey (*Petromyzon marinus)* in the Laurentian Great Lakes. *Conservation Physiology,* 5(1), cox031.

Siefkes, M., Steeves, T., Sullivan, P., Twohey, M., & Li, W. (2013). Sea lamprey control: past, present, future. *Great Lakes Fisheries Policy and Management,* 651-704.

Simberloff, D. (2014). Biological invasions: What's worth fighting and what can be won? *Ecological Engineering*, 65, 112-121.

Škunca, N., Altenhoff, A., & Dessimoz, C. (2012). Quality of computationally inferred gene ontology annotations. *PLoS Computational Biology,* 8(5), e1002533.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787-792.

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(11), 1780-1801.

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., ... & Morgan, J. R. (2013). Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature Genetics*, 45(4), 415-421.

Sorensen, P. W., & Bergstedt, R. A. (2011). Sea lamprey. In Encyclopedia of Biological Invasions (pp. 619-623). University of California Press.

Spice, E. K., Goodman, D. H., Reid, S. B., & Docker, M. F. (2012). Neither philopatric nor panmictic: microsatellite and mtDNA evidence suggests lack of natal homing but limits to dispersal in Pacific lamprey. *Molecular Ecology*, 21(12), 2916-2930.

Storey, J., Bass, A., Dabney, A., Robinson, D., & Warnes, G. (2021). *Qvalue:* Q-value estimation for false discovery rate control.

Strayer, D. L. (2010). Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biology*, 55, 152-174.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualized long lists of gene ontology terms. *PLoS one,* 6(7), e21800.

Turner, S. (2014). QQMAN: an R package for visualizing GWAS results using Q-Q and Manhattan plots.

Verhoeven, K. J., Macel, M., Wolfe, L. M., & Biere, A. (2011). Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. Proceedings of the Royal Society B: *Biological Sciences*, 278(1702), 2-8.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software*, 36(3), 1-48.

Waldman, J. R., Grunwald, C., & Wirgin, I. (2006). Evaluation of the native status of sea lampreys in Lake Champlain based on mitochondrial DNA sequencing analysis. *Transactions of the American Fisheries Society*, 135(4), 1076-1085.

Wang, L., Riseng, C.M, Mason, L.A, Wehrly, K.E., Rutherford, E.S., McKenna, J.E., Castiglione, C., Johnson, L.B., Infante, D.M., Sowa, S., Robertson, M., Schaeffer, J., Khoury, M., Gaiot, J., Hollenhorst, T., Brooks, C., & Coscarelli, M. (2015). A spatial classification and database for management, research, and policy making: The Great Lakes aquatic habitat framework. *Journal of Great Lakes Research*, 41(2), 584-596.

Waples, J. T., Eadie, B., Klump, J. V., Squires, M., Cotner, J., & McKinley, G. (2008). The Laurentian Great Lakes. In Continental Margins: A synthesis and planning workshop. North American Continental Margins Working Group for the US Carbon Cycle Scientific Steering Group and Interagency Working Group. US Carbon Cycle Program.

Ward, R. D., Woodwark, M., & Skibinski, D. O. F. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology*, 44(2), 213-232.

Watts, J. E., Schreier, H. J., Lanska, L., & Hale, M. S. (2017). The rising tide of antimicrobial resistance in aquaculture: sources, sinks and solutions. *Marine Drugs*, 15(6), 158.

White, T. A., Stamford, J., & Rus Hoelzel, A. (2010). Local selection and population structure in a deep-sea fish, the roundnose grenadier (*Coryphaenoides rupestris*). *Molecular Ecology*, 19(2), 216-226.

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F ST. *The American Naturalist,* 186(S1), S24-S36.

Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A., & Losos, E. (1998). Quantifying threats to imperiled species in the United States. *BioScience*, 48(8), 607-615.

Williams, G. C. (1966). Natural selection, the costs of reproduction, and a refinement of Lack's principle. *The American Naturalist*, 100(916), 687-690.

Yin, X., Martinez, A. S., Perkins, A., Sparks, M. M., Harder, A. M., Willoughby, J. R., ... & Christie, M. R. (2021). Incipient resistance to an effective pesticide result from genetic adaptation and the canalization of gene expression. *Evolutionary applications,* 14(3), 847-859.

Tables & Figures

Table 1. Adult sea lampreys were caught in traps during their spring spawning migration, and nearest river mouth names, abbreviations, and coordinates were used in subsequent analyses. Each of the five Great Lakes were given a corresponding number (Erie = 1, Huron = 2, Michigan = 3, Ontario = 4, Superior = 5). 10 km buffers were calculated around each site's coordinates, and for each site the following variables were averaged and summarized within that buffer distance: Annual vertical temperature (AVT), summer vertical temperature (SVT), and annual surface temperature (AST) were averaged across all lake depths over a span of six years (2006-2012, AVT and SVT) or eighteen years (1995-2013, AST) and measured in degrees Celsius. Lake depth was measured in meters. Human population density was measured in number of humans per square kilometer.

SITE NAME	ABR.	LAT-LON	LAKE	AVT	SVT	DEPTH	HUM POP DENS	AST	# SAMPLES
Grand River, Erie	GO	41.76030816°N, -81.28080382°W	1	11.723925	23.898961	13.037138	290.123008	11.749379	14
Augres River, Huron	Augres	44.028208°N, -83.682204°W	2	11.207741	23.294657	6.357095	15.57745	10.291279	15
Bridgeland, Huron	BL	46.252611°N, -83.560831°W	2	8.973857	15.424706	19.776806	10.233952	8.111231	15
Cheboygan, Huron	CH	45.65651°N, -84.46507°W	2	10.627392	17.692057	12.059733	32.523483	8.418697	16
St. Marys River, Huron	SM	46.50686°N, -84.342001°W	2	8.973857	15.424706	2.084709	304.279348	8.666841	17
Grand River, Michigan	GR	43.058065°N, -86.250187°W	3	10.396035	17.85144	27.807862	171.215774	10.455694	15
Manistique, Michigan	MAN	45.9483086°N, -86.24570579°W	3	8.170958	14.043343	12.236863	19.739718	8.543016	16
Black River, Ontario	BLR	43.9941°N, -76.0638°W	4	11.69487	22.944309	7.961046	41.514474	10.731378	18
Duffin's Creek, Ontario	DF	43.816767°N, -79.034742°W	4	8.129821	8.805038	34.082056	958.830468	9.351117	20
Middle Brule River, Superior	BR	46.6903083°N, -91.827508°W	5	6.619298	8.935644	19.420352	3.395222	7.677738	16
Rock River, Superior	RO	46.46530864°N, -86.91530606°W	5	7.079625	9.822782	40.481473	3.056325	7.869711	15
Tahquamenon, Superior	TQ	46.55500876°N, -85.02940548°W	5	8.088073	12.07401	9.554768	1.310133	8.242073	16
Big Creek, Erie	BC	42.602219°N, -80.45191°W	1	11.118668	19.73941	6.516233	16.12492	10.900946	16

Table 2. All environmental variables were assessed to look for correlation using an R^2 coefficient of determination; because of the correlation between most of the variables, I *ran the model once with all variables, then each variable individually.*

	LAT	**LL**	**SVT**	**DEPTH**	**HPD**	**AVT**	**AST**
LAT	1.0000000	0.53214148	-0.6144377	0.02803401	-0.31147790	-0.7314593	-0.9468592
LL	0.53214148	1.0000000	-0.6843359	0.47185256	-0.03318208	-0.7150600	-0.5606798
SVT	-0.61443767	-0.68433594	1.0000000	-0.60472356	-0.31934836	0.9540128	0.7852941
DEPTH	0.02803401	0.47185256	-0.6047236	1.0000000	0.40919125	-0.5088762	-0.2651476
HPD	-0.31147790	-0.03318208	-0.3193484	0.4091925	1.0000000	-0.1286168	0.1525927
AVT	-0.73145932	-0.71506000	0.9540128	-0.50887616	-0.12861676	1.0000000	0.8443154
AST	-0.94685921	-0.56067976	0.7852941	-0.26514765	0.15259267	0.8443154	1.0000000

Table 3. References, test types (ie. Genotype-Environment Association (GEA) and Population Differentiation (PD)), correction methods and threshold for statistical significance affiliated with each analytical model.

TEST	**REFERENCE**	**TYPE**	**CORRECTION METHOD**	**STATISTICAL SIGNIFICANCE**
LFMM	Frichot et al. 2013	GEA	Latent Factors	q-val < 0.05
PCADAPT	Luu et al. 2017	PD	Number of Principal Components	q-val < 0.05
OUTFLANK	Whitlock and Lotterhos 2015	PD	Trimming	q-val < 0.05

Table 4. Selection summary from PANTHER Classification System detailing analysis type, annotation version and release date, the list analyzed, the list input GO terms were referenced against, the annotation data set, statistical test type, and correction.

Analysis Type	PANTHER Overrepresentation Test (Released 20210224)
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5080993 Released 2021-07-02
Analyzed List:	Client Text Box Input (List of Gene IDs)
Reference List:	Homo sapiens (all genes in database)
Annotation Data Set:	GO biological process complete
Test Type:	Fisher's Exact
Correction:	False Discovery Rate

Table 5. Selection summary from REVIGO (Reduce + Visualize Gene Ontology) detailing list size, removal of obsolete GO terms, which species worked with, and what measure of semantic similarity to use.

Resulting list size:	Medium (0.7)
Remove obsolete GO terms:	Yes (default)
Species worked with:	Whole UniProt database (default)
Semantic similarity measure:	SimRel(default)

Table 6. Significant gene ontology terms identified by at least two analyses. The table consists of the numerical number associated with each term, the name of the function affiliated with the GO term, and columns representing the genes, if any, representing the GO term from each statistical method.

TERM	NAME	OUTFLANK	LFMM	PCA
16310	phosphorylation	FLT1	AGK	AGK, CERKL
7165	signal transduction	IGFBP7	-	SAV1, TCP11L1, TSPEAR
6468	protein phosphorylation	IGFBP7	GDF10	TP53RK, RPTOR
44248	cellular catabolic process	-	GGT7	GGT1
46907	intracellular transport	-	ANTKMT	TMEM167A, FAM199X
6357	regulation of transcription by RNA polymerase II	RFX7	-	WDR77, CBFB, CREB3L1, CTDP1, DMRT3, PHF14, REL
34622	cellular protein-containing complex assembly	TBCE	-	TBCC
6457	protein folding	TBCE	-	TBCC, CCT7
9966	regulation of signal transduction	IGFBP7	GAL	-
45935	positive regulation of nucleobase-containing compound metabolic process	-	ANTKMT	ERCC2
6414	translational elongation	-	EIF2S2	C1QBP, EFTUD2
6749	glutathione metabolic process	-	GGT7	GGT1
398	mRNA splicing, via spliceosome	-	SF3A3	EFTUD2, XAB2
30036	actin cytoskeleton organization	-	MAPKAP1	LDB3
43085	positive regulation of catalytic activity	-	ANTKMT	XRCC4
16579	protein deubiquitination	-	USP47	USP24, USP47
6366	transcription by RNA polymerase II	ISL1, RFX7	IKBKB	-
60341	regulation of cellular localization	-	ANTKMT	XPO5
44782	cilium organization	-	ARMC2	ARMC2, IQCG
723	telomere maintenance	-	ATM	RIF1, RTEL1
6635	fatty acid beta-oxidation	-	PEX2	ECHDC1
2683	negative regulation of immune system process	-	GAL	LRCH1
18105	peptidyl-serine phosphorylation	-	IKBKB	MAST2
46513	ceramide biosynthetic process	-	AGK	SMPD4, AGK
31929	TOR signaling	-	MAPKAP1	RPTOR
30509	BMP signaling pathway	-	GDF10	TWSG1

46928	regulation of neurotransmitter secretion	-	MCTP1	MCTP1
86	G2/M transition of mitotic cell cycle	-	CDC6	CDC6
61512	protein localization to cilium	CFAP52	BBS2	CFAP52
33314	mitotic DNA replication checkpoint signaling	-	CDC6	CDC6
72676	lymphocyte migration	-	GAL	LRCH1
35735	intraciliary transport involved in cilium assembly	CFAP52	BBS2	CFAP52
1905349	ciliary transition zone assembly	CFAP52	BBS2	CFAP52
97712	vesicle targeting, trans-Golgi to periciliary membrane compartment	CFAP52	BBS2	CFAP52

Figure 1. A map of trap sites from 2019, when adult sea lampreys were sampled on the spring spawning run.

Figure 2. Ancestry matrices with K values equal to 2, 3, and 4, respectively. K represents the number of ancestral populations to depict ancestry proportion. Individuals are grouped by sample site, and each bar within a group represents an individual. Sample site abbreviations are explained in Table 1. The proportion of fill color of each bar represents percentage of ancestry from an ancestral population, with respect to K.

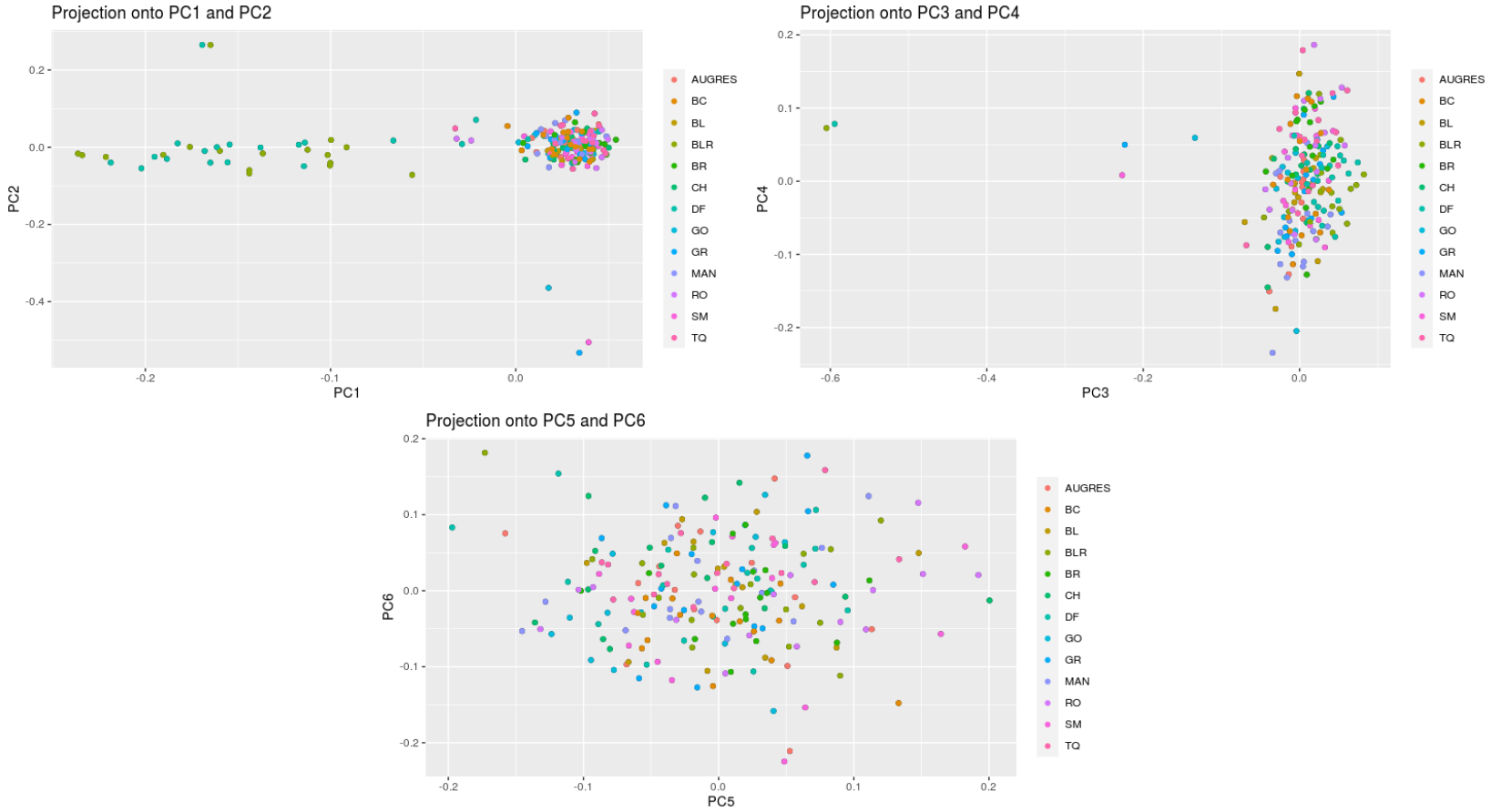

Figure 3. Plots from principal component analysis projecting individuals' shared ancestry on various principal components. The projection onto PC1 and PC2 shows expected clustering, with Ontario populations (BLR and DF abbreviations) distinct from those in Lake Erie and the Upper Great Lakes. With higher principal components, the structuring becomes less clear, indicative of the most likely K value being 2. Sample site abbreviations are explained in Table 1.

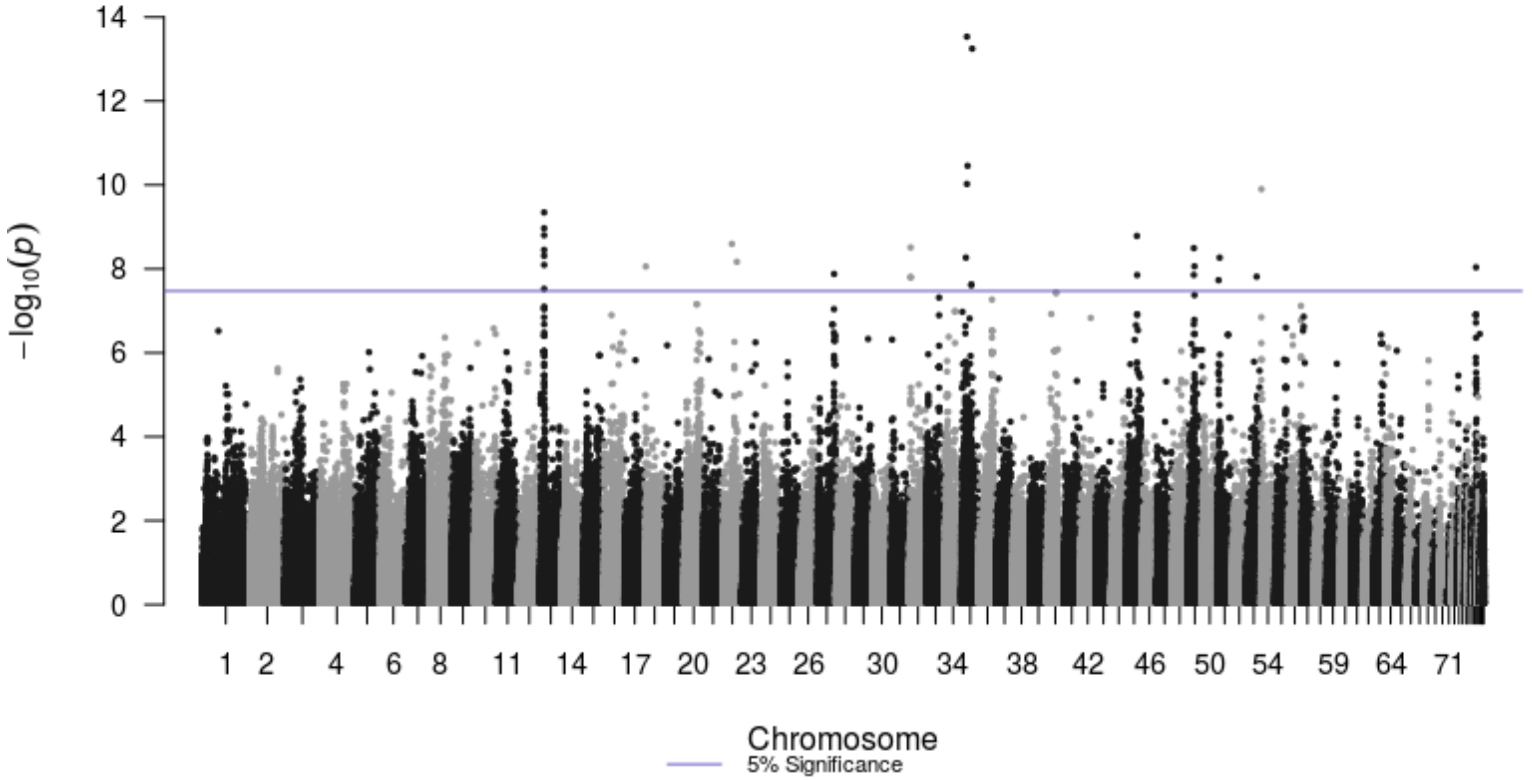

Figure 4. A Manhattan plot of single nucleotide polymorphisms (SNPs) across the lamprey genome, with alternating grey and black bars representing SNPs grouped by chromosome. The y-axis represents the -$\log_{10}$ p-values of SNPs associated with human population density (HPD), and the vertical blue line extending from the y-axis shows a significance value equal to or less than a p-value of 0.05. SNPs plotted above the line of significance suggest strong correlation with HPD.

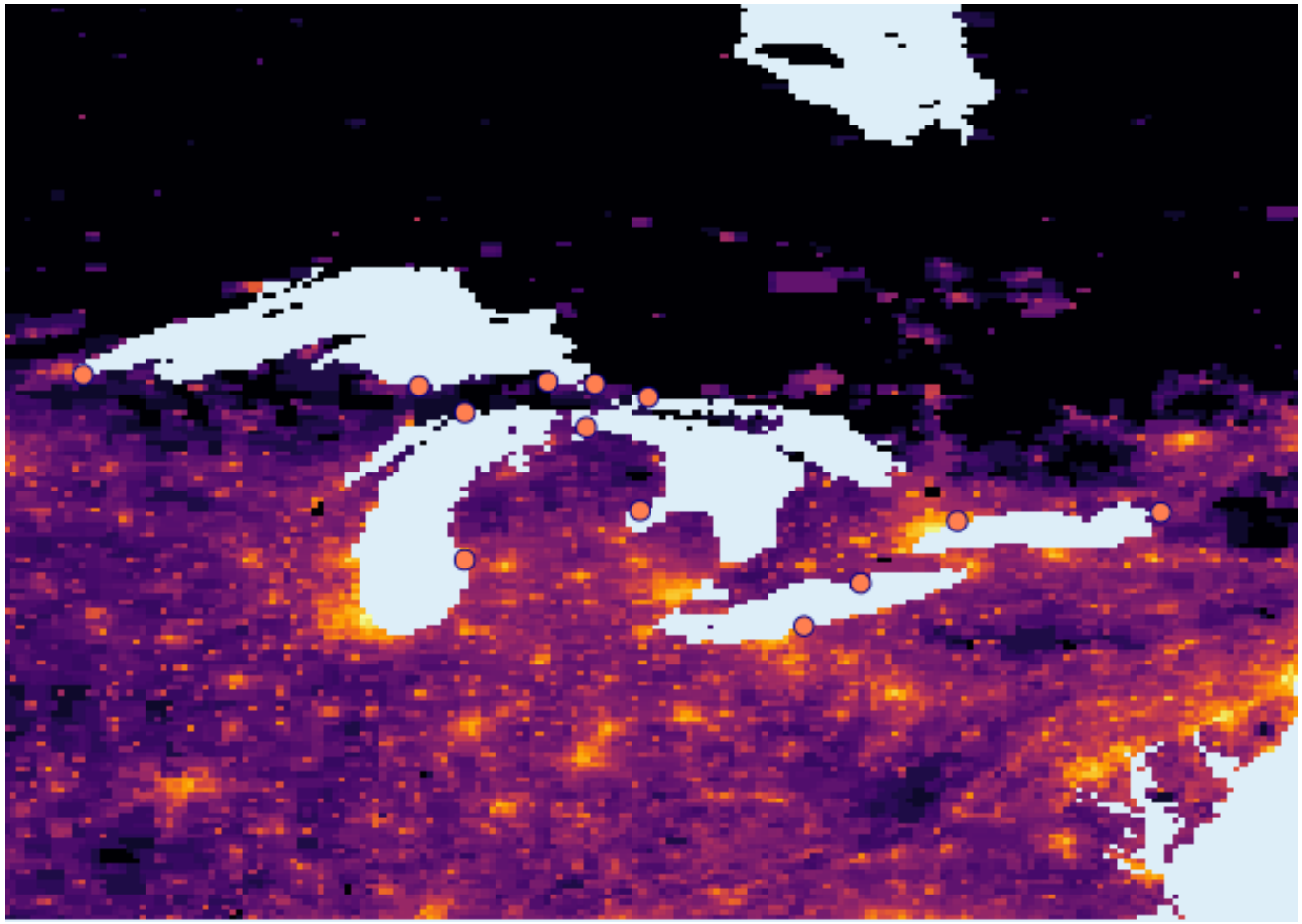

Figure 5. A map of the Great Lakes basin and surrounding areas, with fill color representing human population density. The plots on the map represent river mouth coordinates associated with each sea lamprey trap site, which were used to derive environmental variables for latent factor mixed modelling.

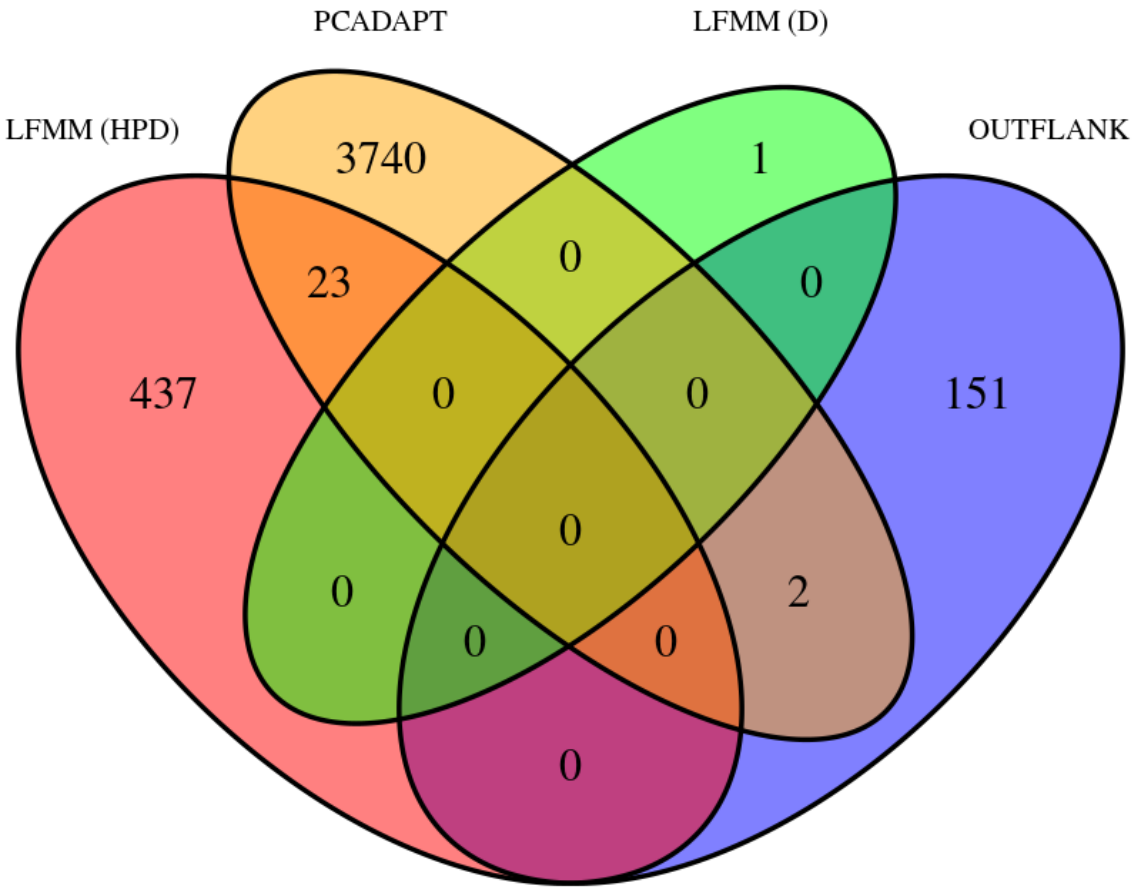

Figure 6. A Venn diagram representing overlap in significant SNPs found by PCADAPT, OUTFLANK, and two different models of LFMM (human population density and localized depth).

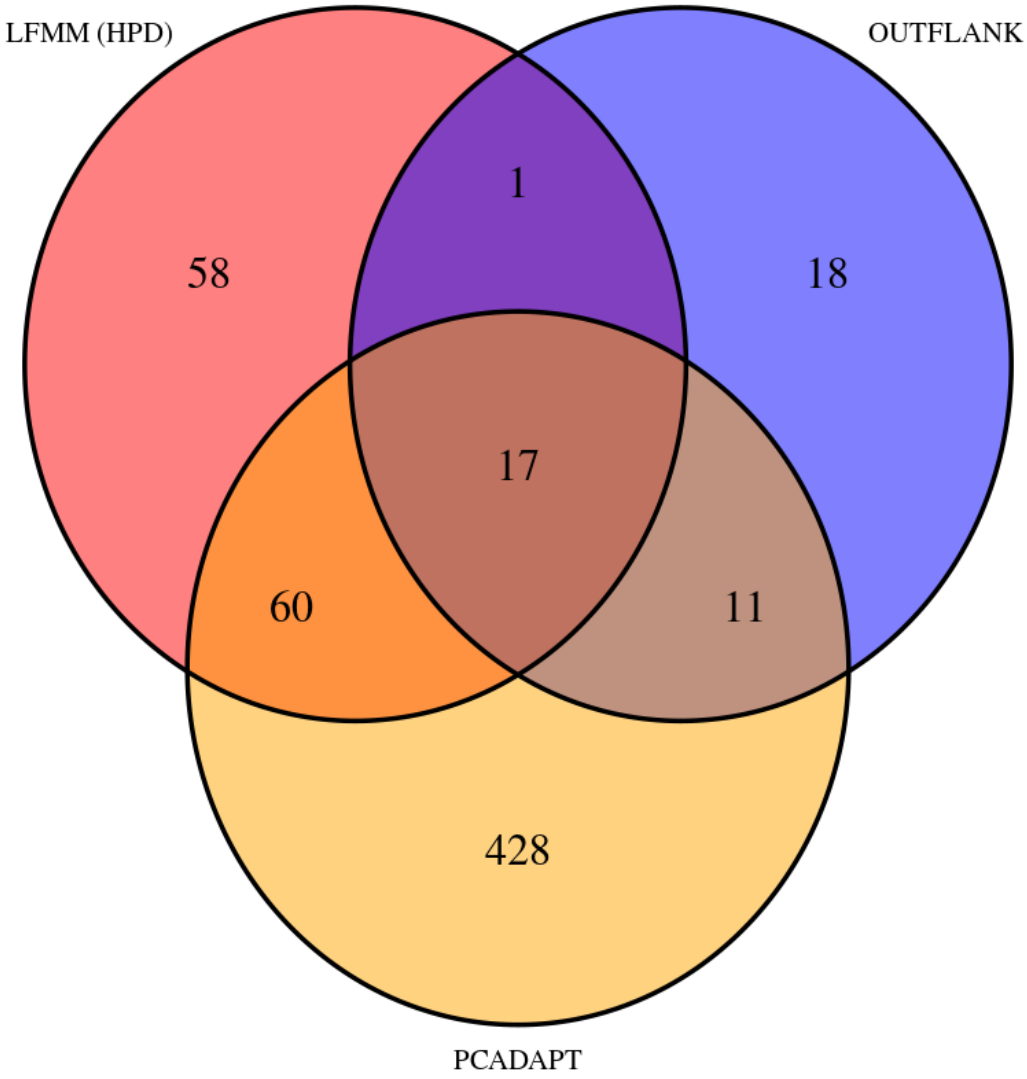

Figure 7. A Venn diagram representing overlap in gene ontology (GO) terms affiliated with significant SNPs found by PCADAPT, OUTFLANK, and LFMM (human population density model).

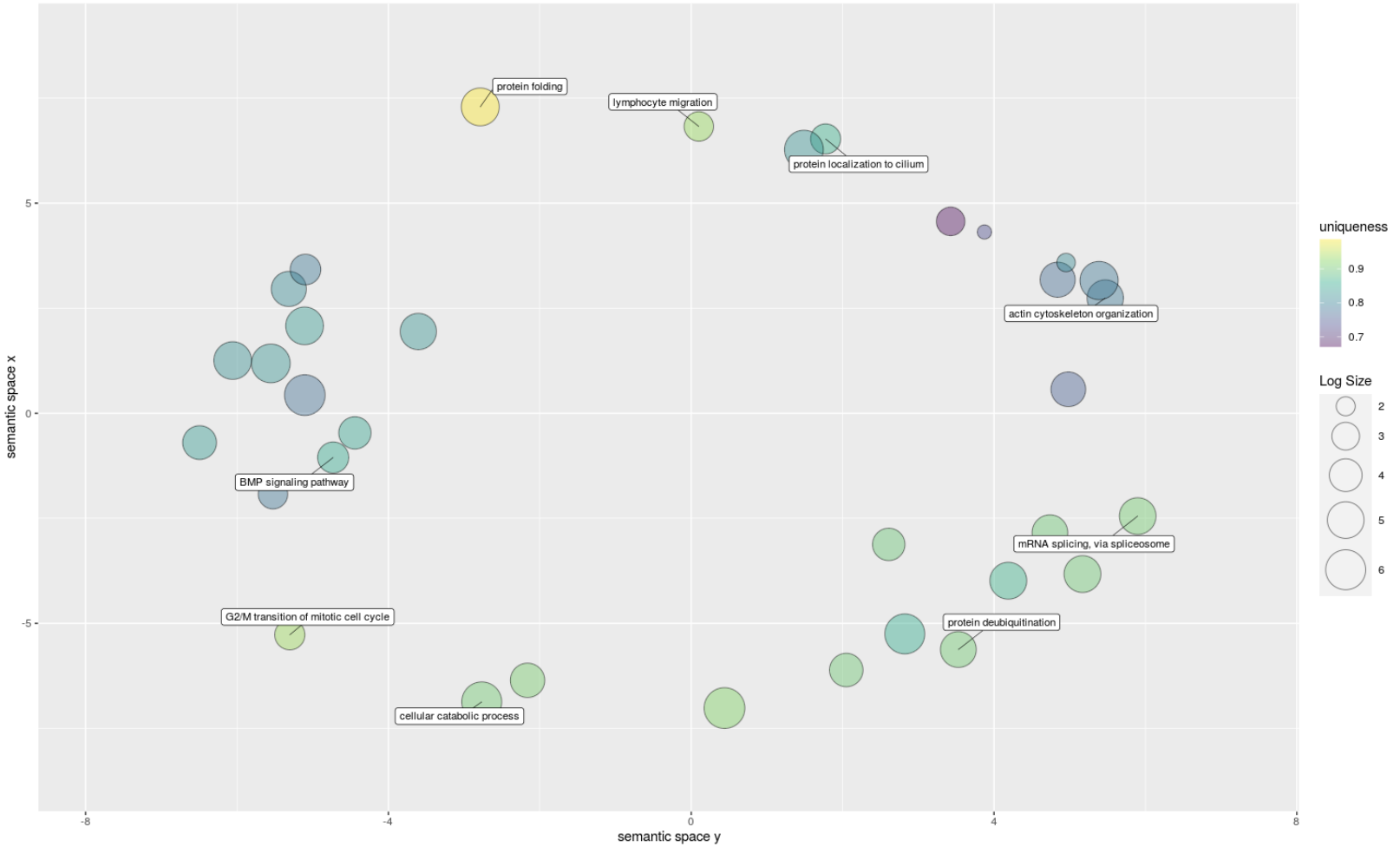

Figure 8. A plot of biological functions associated with significant GO terms. The x- and y- axes do not have intrinsic value, and the size of the plotted circles represents the log size of frequency. SNPs found significant by at least two analytical methods were linked to genes with human homologs, and these genes were affiliated with GO terms. If many genes contributed to one GO term, it was given a higher frequency. The color of the plotted circles represents uniqueness of each GO term, in that its function was unique when compared to the biological functions of other GO terms in the dataset.

Chapter 3 | The human footprint index shapes isolation-by-resistance patterns in North American birds and mammals.

3.1 | Introduction

3.1.1 | Isolation-by-Distance and Isolation-by-Resistance

Gene flow and genetic drift are two of four processes that drive evolution (Wright 1986). Studying these processes is vital when building our understanding of the way animal populations move, change, and adapt over time. This is especially true from a conservation perspective when populations are managed to maintain genetic diversity and facilitate adaptation to local environments. When wildlife populations are isolated from one another and begin to experience the detrimental effects of genetic drift, such as inbreeding depression and loss of adaptive alleles, adding and maintaining corridors for gene flow can introduce new alleles that allow the population to recover (Holt & Gomulkiewicz 1996). In contrast, introducing alleles adapted to other environments can negatively affect local adaptation of a population (Aguillon et al. 2017).

Gene flow between populations can be studied by measuring neutral allele frequencies across populations. As these loci are not adaptive, they can be used to study gene flow and genetic drift while populations exist in habitats with different landscape features (Holderegger et al. 2005). High degrees of similarity in the non-adaptive loci in neighboring populations is thus indicative of the occurrence of gene flow (Frichot et al. 2013). This is a valuable tool when studying how populations experience dispersal and estimating the driving forces behind this process.

In this study, I aimed to determine whether isolation-by-distance (IBD) or isolation-by-resistance (IBR), using human footprint index as a resistance surface, is more effective at explaining broad-scale patterns of genetic diversity in North American mammal and bird species. I hypothesized the human footprint index, a measure of environmental impacts associated with human population density (WCS & CIESIN 2005), to be a more significant predictor of genetic differentiation when compared to a straight-line distance estimate. To investigate the impact of the human footprint index on neutral genetic diversity, in comparison to Euclidean distance between populations, I used publicly archived raw microsatellite datasets for birds and mammals across North America.

To study gene flow occurring between populations and genetic drift occurring at isolated populations, I must first understand how animals are spatially distributed across the landscape.

Dispersion influences biological and genetic processes alike, and individual movements can lead to a shift in population structure over time as new alleles are introduced and shared (Aguillon et al. 2017). Conversely, lack of dispersal can lead to genetic isolation and greater differentiation between populations. Understanding the underlying mechanisms of dispersal and the resulting connectivity between populations is necessary when studying genetic diversity among and between populations (Garner et al. 2005).

Historically, isolation-by-distance has been the null expectation for genetic differentiation between wildlife populations. Introduced by Sewall Wright in 1943, this model predicts that populations geographically closer to one another will be more genetically similar than those further apart. Geographic dispersal of a species is expected to cause an increase in genetic dissimilarity; populations that are farther apart are expected to share less alleles, and populations closer to one another, more genes. IBD uses straight-line Euclidean distance as a proxy for other factors that cause isolation in populations and has been a cornerstone for describing genetic diversity for decades (Jenkins et al. 2010, Wright 1943).

However, more recent studies have looked at how environmental heterogeneity between two populations can predict genetic differentiation. Isolation-by-environment (IBE) and isolation-by-resistance (IBR) are similar concepts to IBD, though more complex. IBE estimates the level of genetic differentiation between two populations, using environmental variation as the distance between populations (Wang & Bradburd 2014). IBR estimates the level of genetic differentiation between two populations using a resistance surface—i.e., surface with features that impede movement or connectivity—as the distance between populations (McRae 2006).

3.1.2 | The Human Footprint

We know that dispersal can occur naturally over time for several reasons; individuals and may disperse to take advantage of limited resources, resources that are spatially and temporally variable, or resources that are required at different stages of their life cycle. Animals also disperse as their environments become less suitable, to find a mate, and to avoid competition or inbreeding (Chatterjee 2006).

Dispersal can be influenced by and occurs because of human development and interference, more so now than ever as humans have altered over 75% of ice-free land around the globe (Ellis & Ramankutty 2008). 55% percent of the global human population now lives in urban areas, and this proportion is expected to continue increasing in the coming decades (United

Nations 2018). This process, known as urbanization, transforms both the environment and the spatial distribution of a population from rural to urban settings (United Nations 2018).

Among other consequences, urbanization has caused habitat fragmentation leading to smaller, more isolated wildlife populations (Sommer et al. 2013). Human development plays a substantial role in increasing dispersal by fragmenting landscapes and hampering movement. These large dispersal distances can lead to inhibited gene flow and an increased rate of genetic drift as populations become disconnected from one another and fragmented across the landscape. This increase in fragmentation between populations of species can, in turn, lead to higher rates of extinction (Crooks et al. 2017).

This hampering of movement and gene flow can be seen across a variety of mammalian taxa on a global scale as human populations have continued to grow and urbanize (Dixo et al. 2009). The effects of humans fragmenting the landscape on genetic diversity can be severe; fragmentation leads to an increase in genetic drift, smaller effective populations sizes, lower genetic diversity, and inability to spread beneficial alleles between populations (Palumbi 2001).

Furthermore, higher degrees of habitat fragmentation have been shown to increase the likelihood of a species being listed as threatened or endangered; terrestrial mammals are sensitive to these kinds of changes, especially those with smaller range sizes or larger body sizes. Specialist species and animals with long generation times (typically large-bodied animals) often cannot adapt to fragmented habitat due to life history constraints, leading to declines in population sizes (Crooks et al. 2017). Additionally, populations in proximity to urbanization have experienced negative effects on within-site mammalian genetic diversity (Schmidt et al. 2020), abundance, and species richness (Sauvajot et al. 1998). It is currently unclear whether loss of genetic diversity is occurring among mammal populations because population sizes are shrinking or because of the loss of connectivity between these sites—or some combination of the two (Bergl et al. 2008).

Conversely, some mammal species—primarily generalists, omnivores, and smaller-bodied carnivores—can adapt in urban areas and utilize food and habitat resources found exclusively in human settlements (McCleery 2010). Urban areas have high habitat heterogeneity at small spatial scales (Savard et al. 2000, McDonnell & Pickett 1990), which can lead to an increase in beta diversity—that is, diversity in species communities between habitat patches (Niemelä 1999).

Birds, too, experience diverse effects when exposed to urbanization. This is heavily dependent on foraging strategies (Marzluff 1997). For example, insectivores often experience a decline in urban areas due to the higher volume of non-native plants, which may support fewer insects (Reichard et al. 2001). However, birds that eat seeds or nectar benefit from bird feeders (Robb et al. 2008) and gardens (Paker et al. 2014) and scavenging species can find food abundant resources in trash (Arnold et al. 2021) and roadkill (Schwartz et al. 2018).

In this study, I was interested in exploring how the human footprint index (HFI) (Venter et al. 2018) acts as a resistance surface to wildlife species' dispersal and subsequent genetic differentiation. Studies across the full scope of animal taxa have shown that there is no uniform response of species to human development. For many species, areas with a high degree of human influence do not provide adequate habitats or ecosystem services (Scanes 2018, Tratalos et al. 2007), and impede gene flow (Fusco et al. 2021, Banks et al. 2013). For others, areas of human development can act as conduits to gene flow (Miles et al. 2019, Crispo et al. 2011) and increase the availability of food and useable habitats (Centeno-Cuadros et al. 2017). Despite the individual nature of positive or negative effects on animal species, I hypothesized the effect of the HFI may be a good metric for predicting genetic diversity among populations.

In surveying the scope of the effect of HFI on genetic distance, I also looked at several potential covariates: average individual body mass, number of populations in a dataset, range of HFI values a population was exposed to, area of minimum convex polygons, and hand-wing index (for birds). By controlling for these variables, I aimed to not only examine the relationship between HFI and genetic diversity, but also determine whether any of these covariates contributed to variation in effect size across species.

The main goal of my study was to create IBD and IBR models with open-source raw microsatellite data and perform a meta-analysis to determine whether Euclidean distance or human development better captures genetic diversity patterns across birds and mammals in North America. I aimed to test the hypotheses: (a) the human footprint index shapes genetic diversity in bird and mammal species and (b) the human footprint index better captures genetic diversity patterns than Euclidean distance between populations.

Teasing apart the effects of geographic distance from environmental distance can help to better understand specific processes that drive microevolution (Sexton et al. 2014). By incorporating anthropogenic features that influence dispersal, I was able to measure whether the

human footprint index had an impact on genetic diversity and gene flow between populations. Specifically, I assessed the extent to which these factors contribute to the IBR hypothesis and compared this to the straight-line distance between the same set of populations.

3.2 | Materials and Methods

3.2.1 | Genetic Datasets

The dataset used in this analysis was synthesized from the open-source repository Data Dryad. It is compiled from genetic studies across North America, consisting of raw microsatellite data from 1,008 population data points and 66 species (Schmidt et al. 2020). For each study, there is information on neutral genetic diversity for each sampled population, and latitude-longitude coordinates associated with each sample site; I transformed the latitude-longitude coordinates into the Albers Equal Area (AEA) projection to remain consistent with the resistance surface. The main benefit of using raw microsatellite data from public data archives was uniformity, in that a direct comparison of neutral genetic diversity could be made across both bird and mammal species despite the original purposes of the datasets.

Microsatellite data were downloaded directly from Dryad; files contained raw genotype data for individuals across multiple sample sites. To read these files into the R programming environment (R Core Team 2017), I used the package GENEPOP (Rousset 2008, version 1.1.7) to convert the structure files into genepop objects, which are R data objects that provide summaries of genotypes for each sample, marker information, and allele frequencies in each population. To then convert the genepop data into the matrix format *genind*, I utilized the R package ADEGENET (version 2.1.4) (Jombart & Ahmed 2011, Jombart 2008, version 2.1.4).

I then calculated pairwise F_{ST} calculations for all studies using Weir & Cockerham's unbiased estimator (1984), a function of the HIERFSTAT package in R (Goudet & Jombart 2020, version 0.5-7). Genetic differentiation is generally overestimated when using small sample sizes to infer F_{ST} (Holsinger & Weir 2009, Kalinowski 2005), largely because most methods cannot have negative F_{ST} values (Willing et al. 2012). The Weir & Cockerham's estimator, however, allows for negative F_{ST} values.

I obtained median adult body mass for each bird and mammal species from the Amniote Life History Database, an aggregation of life history traits from peer-reviewed studies, existing databases, and published books (Myhrvold et al. 2015). Within this database, a median was taken of measurements compiled from several sources. I selected body mass as a variable of interest, as

it is positively correlated with dispersal distance (Tamburello et al. 2015) and negatively correlated with population size (Etienne & Olff 2004). For birds, I downloaded the hand-wing index (HWI) database compiled by Sheard et al. (2020); HWI is an indicator of dispersal capability based on the wing shape of birds. HWI is correlated with both environmental and behavioral elements, most of all seasonality, diet, and habitat type (Sheard et al. 2020). As both body mass and HWI predict dispersal ability and distance, they are vital metrics to consider when estimating the effect of environmental and ecological factors on gene flow and movement.

3.2.2 | Human Footprint Index Dataset

The landscape raster dataset used in this analysis, the Human Footprint Index (HFI) (Venter et al. 2020), was sourced from the Socioeconomic Data and Applications Center (SEDAC), a NASA project. The HFI is a composite surface of eight variables contributing to human influence and development. The surface combines standardized measures of electrical infrastructure, population density, agricultural developments, roads, railways, and waterways, and other human developments, to estimate global human disturbance of natural environments.

I projected the HFI dataset in the Albers equal area conic projection (AEA); this projection is among the most used for the contiguous United States (Snyder 1987). As a conic projection, straight-line meridians are spaced equally and converge at the poles, creating a fan-shape that is downward-oriented in the northern hemisphere. Additionally, as an equal-area projection, shapes and directions experience distortion while preserving the size of regions on the earth. It is most apt for east-to-west expanding land masses (Synder 1997); I selected this projection due to the retention of land mass size, and because most of my data points were collected in the contiguous United States and southern Canada.

For each study, I projected spatial information for each population as points on the resistance surface. I used the RASTER (version 3.4-13) (Hijmans 2021) function buffer() to place a buffer of 200 to 1,000 kilometers around the population points and cropped the HFI surface to the buffer distance surrounding the points. This was to shorten processing time, reducing a continent-sized map to one more appropriate for the dispersal of the species being analyzed. Each study dataset was tested first with a buffer distance of 200 kilometers, and then adjusted to a larger distance to reduce standard errors caused by too small a resistance surface for species with larger home ranges.

Minimum convex polygons (MCP) were calculated for each study by constructing a convex hull of all sample sites therein. MCPs are calculated by drawing a polygon around the outermost points of a set of populations. I used minimum convex polygons as the sample area for each dataset. To obtain this information and I used the chull() command from the package grDevices (R Core Team 2020, version 4.1.0) to draw a convex hull around the outermost points of populations from each study in the meta-analysis. The hull was projected as a polygon using the AEA projection, and the area of the polygon in square kilometers was used in subsequent analysis.

Minimum and maximum HFI values from buffered resistance surfaces were extracted from each RasterLayer object, and these metrics were used to calculate the range of HFI each dataset was exposed to. I used range of HFI as a variable in my meta-analyses, given the variation in HFI each dataset was exposed to. For example, a set of populations farther removed from humans will likely experience different effects than a set of populations of the same species nearer to human settlements. Thus, it was necessary to account for the variation in human footprint index within each sample area.

3.2.3 | Maximum Likelihood Estimation of Isolation-By-Resistance Model

Landscape genetic studies often look to find the effective genetic distance between sample populations, examining whether this is best described as a function of the straight-line distance between populations, or as a function of an ecological matrix (McRae 2006). For the latter, it is frequently not feasible to witness and document the movement of animals across the landscape. As such, functional connectivity of the ecological matrix is used to determine effective distance between populations (Cushman et al. 2006). This functional connectivity is called a conductance surface or matrix, where all environmental features on a spatial layer are given a value regarding whether the feature inhibits or facilitates movement of the study organism (Spear et al. 2010). However, these studies of functional connectivity require parameterization to accurately and meaningfully relate ecological processes to gene flow across a landscape. There are many difficulties when optimizing these resistance surfaces: these surfaces have high dimensionality, and these environmental gradients are not isolated from one another (Peterman 2018).

The R package RADISH aids in these concerns, using gradient-based optimization of resistance surfaces to perform maximum likelihood estimates (MLE) of IBR models (Pope &

Peterman 2020, version 0.0.2). RADISH uses regression of distance matrices or generalized Wishart distributions in its likelihood estimation, in which genetic distance is the observed variable and conductance across the resistance surface is a function of spatial covariates.

The following was repeated for each study: As RADISH requires a resistance surface to be input as a stacked object—as the package can analyze multiple resistance surfaces at once—I used the stack() function on the buffered HFI surface. I then used the conductance_surface() function using the resistance surface and the spatial points associated with each population in the study, which uses a maximum likelihood estimate to fit a conductance model as a function of genetic distance (Pope & Peterman 2020). In the case of HFI, the model estimated whole surface conductances between two points through the urbanization surface. Genetic diversity between these two population datapoints was treated as a function of least-cost distance.

For each study in the meta-analysis, I fit a model (the study's pairwise F_{ST} value ~ HFI) with a maximum likelihood population estimate, using the command radish(). I calculated resistance distances across a grid and used the radish_distance() command to calculate the straight-line distance estimates. A model (pairwise F_{ST} value ~ 1) was fit with a maximum likelihood population estimate, again using the function radish(), and the IBD and IBR models compared using ANOVA tests.

3.2.4 | Meta-Analysis

To look at the overall effect of HFI on genetic diversity, I first fit a linear model using the package STATS (R Core Team 2020, version 3.6.2) to regression coefficients generated by RADISH. Respective coefficient standard errors were used as weights in the linear model.

Meta-analysis was then conducted using the R package METAFOR (Viechtbauer 2010, version 3.0-2). I performed an initial meta-analysis on all data using the function rma(). I fit a fixed effects model using regression coefficients as observed effect sizes, and corresponding standard error values. High variance warnings within the METAFOR package suggested my model would be better fit if less certain values were removed, and I ran the data through a series of filters to remove suspect or unrealistic data.

Population sizes less than $n = 4$ lack sufficient data for pairwise analysis and/or had the potential to lead to biased genetic diversity estimates despite the use of an unbiased F_{ST} estimator (Willing et al 2012). I removed studies with only one to three populations sampled. The resulting model fit was still a poor fit, so I then removed studies with outlying effect sizes and studies with

unrealistically large standard error values. Large standard error values were identified and removed via a univariate approach using boxplot.stats() function from the package STATS. Unrealistic effect sizes were removed from the three highest and three lowest studies; some number of unrealistic effect sizes were to be expected, as the data used in this analysis was not originally collected for the purpose of studying the effect of HFI on genetic diversity. Following the removal of these uncertain values, my model fit without warnings about unusually large variances.

I then looked at moderating variables—mass, range of HFI exposure, minimum convex polygons, and number of populations sampled—and how they might influence the effect HFI has on genetic diversity. I looked for correlations across these four variables and retained only one variable in any correlated pair. The remaining variables were each run as moderators in separate fixed effects models.

3.3 | Results

3.3.1 Mammals

The initial meta-analytic linear model found a negative regression coefficient of -1.35 [$^{+}/_{-}$ 0.23], suggesting a general negative effect of HFI on genetic diversity (*Table 5*). The meta-analytic fixed effects model (*Table 9*) on the unfiltered data showed a significant effect, the HFI impeded gene flow, with an estimate value of 18.78 [$^{+}/_{-}$ 0.0001]. However, the sampling variance ratio was extreme, and the standard error of the model was functionally zero. The AICc value (782,203,682.6) and log Likelihood value (-391,101,840.2) were extremely high and low, respectively, indicating incorrect model fit. The fixed effects model with four or more populations also showed significance at 19.47 [$^{+}/_{-}$ 0.000], but again exhibited an extreme ratio of sampling variance and zero standard error. This model had a larger log Likelihood (-136494.1577) and smaller AICc (272990.4444) with respect to the prior model, however these values were still extreme and thus indicative of incorrect model fit. The final, filtered model, with small population sizes, unrealistic effect sizes, and large standard errors pruned out, had an estimate of -0.1782 [$^{+}/_{-}$ 0.0376]. This model did not have an extreme ratio of sampling variance; it also had log Likelihood (-33.164) and AICc (68.999) values within an acceptable range, indicating the filtered data to be correctly fit. The significant negative effects of HFI on gene flow across species remained no matter the model.

On a scale of -1, perfect negative correlation, to 1, perfect positive correlation, mass, and range of HFI exposure were found to be significantly, negatively correlated (-0.43). The number of populations sampled was also found to be negatively correlated with mass, though to a lesser extent (-0.22). Because each of the fixed effect with moderator models was run independently of one another, none of the variables were excluded despite some degree of correlation (*Table 7*).

The first moderating variable tested, range of HFI exposure, showed a model estimate of 0.92 [$^{+}/_{-}$ 0.67], and a moderator estimate of -0.01 [$^{+}/_{-}$ 0.007]. Mass as a moderating variable showed a model estimate of -0.16 [$^{+}/_{-}$ 0.0450], and a moderator estimate of -0.76 [$^{+}/_{-}$ 0.45]. Number of populations sampled as a moderating variable showed a model estimate of -0.14 [$^{+}/_{-}$ 0.06], and a moderator estimate of -0.001 [$^{+}/_{-}$ 0.001]. However, the p-value of the moderator test for populations sampled was greater than 0.05. The final variable, MCP, showed a model estimate of -0.26 [$^{+}/_{-}$ 0.05] and a moderating variable estimate of 0 [$^{+}/_{-}$ 0]. In sum, no moderators had detectable effects.

3.3.2 Birds

The initial linear model found a negative regression coefficient of 0.19 [$^{+}/_{-}$ 0.53], suggesting a slight positive effect of HFI on genetic diversity (*Table 6*). The meta-analytic fixed effects model on all data showed a significant effect with an estimate value of -0.15 [$^{+}/_{-}$ 0.0023]. The fixed effects model with population size greater than or equal to four also showed significance at 0.19 [$^{+}/_{-}$ 0.06], though with a significantly increased AICc value (128.4038) and reduced log Likelihood (-63.0201) in comparison to the unfiltered model (AICc = 69.2809, log Likelihood = -33.4738). The final model, with small population sizes, unrealistic effect sizes, and large standard errors filtered out, had an estimate of 0.14 [$^{+}/_{-}$ 0.06]. This fit was best out of the three fixed effects models, with the lowest AICc value (63.8892) and highest log Likelihood value (-30.7224). None of the models showed extreme sampling variance ratios, and all had significant estimates.

Mass was significantly, negatively correlated with range of HFI exposure (-0.92), and to a smaller degree, MCP and number of populations sampled (-0.29 for both variables). HWI was positively correlated with mass (0.37), and negatively correlated with number of populations sampled (-0.55), range of HFI exposure (-0.27), and MCP (-0.21). The correlative values across environmental variables are available in Table 8.

Range of HFI exposure showed a model estimate of -1.74 [+/- 1.24], and a moderator estimate of 0.019 [+/- 0.013]. Hand-wing index as a moderating variable showed a model estimate of 0.24 [+/- 0.16], and a moderator estimate of -0.003 [+/- 0.004]. Mass had a model estimate of 0.09 [+/- 0.07] and a moderator estimate of 0.0003 [+/- 0.0002]. All three of these variables had p-values greater than 0.05 and were considered insignificant. Number of populations sampled as a moderating variable showed a model estimate of -0.24 [+/- 0.10], and a moderator estimate of 0.04 [+/- 0.009]. Lastly, MCP had a model estimate of -0.14 [+/- 0.09], and a moderator estimate of -0 [+/- 0]. In summary, number of populations sampled in each dataset was the only variable to have a detectable effect on the model.

3.4 | Discussion

My results indicate a consistent effect of HFI on genetic diversity, though the directional effects of urbanization vary between and among bird and mammal species. Additionally, ANOVAs for both the bird and mammal datasets found the human footprint index to be a better predictor of genetic diversity than straight-line distance for most datasets, suggesting isolation-by-resistance to be a more powerful metric than isolation-by-distance when using data with various sample sizes and species. The body of scientific literature devoted to human impacts on biodiversity is vast, and my findings further contribute to the notion that humans and human development have influenced wildlife on a continent-wide scale.

Urbanization is a leading cause of habitat fragmentation and loss worldwide, causing a marked impact on how species evolve (Otto 2018). Animals can either avoid humans and urbanization or adapt to it (Santini et al. 2019). When animals disperse, fitness can increase in acquiring more readily available resources and avoiding threats at the cost of energy requirements of movement. The converse is true for animals that remain in disturbed areas; in either strategy, behavior, movement, and energy usage are all significantly impacted (Doherty et al. 2021).

Disturbance stemming from urbanization can limit migration and gene flow between animal populations, directly affecting genetic differentiation (Banks et al. 2013). Subdivided populations can experience random genetic drift, which increases genetic differentiation (Keyghobadi 2005). Urbanization can also introduce novel environmental conditions that may be advantageous to some species, particularly birds which peak in species richness and diversity in suburban areas (Blair 1999). One example of this is the urban heat island effect, the well-studied

phenomenon where the air temperature of urban areas is greater than that of surrounding non-urbanized landscapes (Oke 1982). Many bird species overwinter in warmer urban landscapes, leading to a seasonal increase in species richness and occupation of regions previously uninhabitable due to cold temperatures (Greig et al. 2017, Caula et al. 2008). Additionally, animals able to use buildings for nests and generalist species capable of finding food resources in developed areas are favored in urban areas (Marzluff 2001).

3.4.1 Mammals and the Human Footprint Index

My base model indicates a small but significant negative effect of human footprint index on genetic diversity in mammals. This is an expected result, as the negative effects urbanization, habitat fragmentation, and habitat loss have on mammalian diversity and richness are well-documented (Schmit et al. 2020, Lino et al. 2019).

Surprisingly, controlling for moderating variables had no significant bearings on the model results. Mass is known to correlate with dispersal capabilities of some mammals, and larger animals are typically more sensitive to the detrimental effects of urbanization (Lino et al. 2019). The species average masses ranged from the little brown bat (*Myotis lucifugus,* 7.15 grams*)* to the American bison (*Bison bison,* 624.6 kilograms*).* As for the number of sampled populations, the seven smallest datasets had four sites, and the largest had one hundred and twenty-nine. MCP also had a large range of sample values, from an area of 18 km^2 to 12,589,760 km^2. Despite the sizeable sampling variances across these variables, the effect of HFI on each species was remarkably consistent regardless of the mass of the animal, the number of sampled populations in each dataset, and the minimum convex polygon of the sample area.

Additionally, I hypothesized the range of HFI exposure would have a significant effect on the model, given the range of exposure across all datasets. For example, one polar bear (*Ursus maritimus*) dataset had populations sampled exclusively in northern, remote areas of Canada. On a scale of 0 to 100, these bears were exposed to a maximum HFI value of 26.7, indicating relatively little influence of urbanization on these populations in comparison to the rest of the datasets. In contrast, some species such as the snowshoe hare (*Lepus americanus*) and the American black bear (*Ursus americanus*) had maximum HFI values of 100, meaning they occupied areas with the maximum measurement of urbanization in towns or cities. Yet, the p-value associated with controlling for this variable was statistically insignificant. This suggests

that the HFI has a consistent, negative effect across species, despite the level of HFI each species is exposed to.

While the overall trend shown by the model is negative, the effect of humans on mammalian diversity is undoubtedly nuanced. A subset of mammalian species within my dataset experienced a positive effect on genetic diversity, showing a clear distinction between mammals that are adaptive towards urban areas, and mammals that avoid urban areas. Myriad factors are contributing to a species' urban tolerance. For example, a study by Baker et al. (2003) found that small mammals can be limited by predators, such as domestic cats, and habitat fragmentation in the form of disrupted patches of natural vegetation. However, the same study found that small mammal populations can increase in distribution and abundance when gardens were available in place of more naturally occurring habitat patches (Baker et al. 2003), suggesting some animals may be able to adapt to and thrive off vegetation introduced by humans. Cranial capacity—brain size—is known to correlate with behavioral plasticity, and urban populations of some mammals have been shown to have increased cranial capacity in comparison to their rural counterparts (DePasquale et al. 2020). This suggests larger brain size is selected for in urban areas, and animals with greater capacity to change their behaviors are better able to adapt to developed areas.

3.4.2 Birds and the Human Footprint Index

In contrast to my findings on mammalian response to human footprint index, I found a small but significant, positive effect of the HFI on genetic diversity of birds. This result was expected, given the well-studied intermediate hypothesis whereby bird species richness, Shannon diversity, and density are highest in moderately disturbed, suburban areas (Blair 1996). In a study by Møller (2008), urban populations of birds were found to have shorter flight distances when approached by potential predators (in this study, humans) than rural populations; birds with shorter flight distances, in turn, were less susceptible to raptor predation. This tameness, represented by shorter flight initiation distances, shows adaptation of some bird species to urban environments (Møller 2008). Birds have been shown to have little to no response to human-caused disturbance of vegetation in habitats near urban areas (Sauvajot et al. 1998). This is in direct contrast to mammals, which largely have stronger responses to disturbed vegetation.

The direction of effects in bird datasets were highly variable in both body size and geographic metrics. Hand-wing index values ranged from cactus wren (*Campylorhynchus*

brunneicapillus) (13.72) to broad-tailed hummingbird (*Selsphorus platycercus*) (65.18), and mass values ranged from 3.55 grams to 2,387.5 grams. The range of HFI exposure was at minimum 46.94 and at maximum 89.23, and minimum convex polygons of sample areas ranged from 912 km^2 to 13,527,030 km^2 . However, when statistically controlling for these variables, all four were insignificant in their effects on the model. This suggests that bird species are uniformly affected by HFI regardless of HWI, mass, range of exposure to human development, and the minimum convex polygon of the sample area.

The final moderating variable, number of populations sampled, had a small but significant, positive effect on the meta-analytic model. This suggests that the effects of HFI on genetic distance increase with number of populations sampled. A larger number of sampled populations may be necessary to capture the full extent of the effect of HFI on genetic distances.

Much like mammals, the individual effects of HFI on bird species are heavily dependent on taxonomic groups and life history traits. For example, nearness to human infrastructure has been linked to an increase in nest predation and lower productivity of some passerines (Liebezeit et al. 2009). Passerine species (Order Passeriformes) have also been linked to altered sexual selection in cities, in which plumage dichromatism—brightly colored plumage—is negatively associated with urban tolerance (Iglesias-Carrasco et al. 2019). Of my bird datasets used in the final fixed-effects model, three of five passerines displayed a negative effect of HFI on genetic diversity. Two shorebirds (Order Charadriiformes) and one waterfowl (Order Anseriformes) also exhibited a negative effect. Shorebirds, which nest on open beaches and in vegetative patches, are sensitive to human development, in that disturbance can lead to reduced nest attentiveness and increased predation (Frid & Dill 2002). Some Anseriformes, particularly the greater white-fronted goose (*Anser albifrons)* used in my study, are similarly sensitive to human disturbance; conversion of wetland to farmland leads to increased flight distances for geese, increase in energetic costs, and disturbance in migration networks (Xu et al. 2021).

Of the bird species with a positive relationship between genetic diversity and HFI, there was one owl species (Order Strigiformes), one hummingbird (Order Apodiformes), one waterfowl (Order Anseriformes), and two passerines (Order Passeriformes). The variance in individual results clearly displays the species-dependent response to urbanization, and highlights trends in the overall meta-analysis. Overall, birds that do not require a diverse prey base, such as raptors, have particularly plastic responses to urbanization (Isaac et al. 2014, Gehlbach 1996).

Ground-nesting species are more sensitive to human development than species that can nest in buildings and other urbanized structures (Marlzuff 2012), and omnivores, granivores, and habitat generalists are more abundant in urban areas than rural areas (Silva et al. 2016). Urbanization acts as a filter for bird species and homogenizes species within urban habitats (Silva et al. 2016).

3.4.3 Isolation-by-Distance versus Isolation-by-Resistance

My analysis revealed isolation-by-resistance to have greater explanatory power than isolation-by-distance in capturing genetic diversity patterns (*Tables 1 & 2).* For every species in the meta-analysis, the HFI IBR model was found to have at least the same or higher log-likelihood value than respective IBD models. While IBD has been the historical method of explaining the relationship between geography and genetic diversity, my results concur with more recent studies suggesting genetic distances are best explained using the IBR model whereby environmental surfaces act as corridors for or barriers to gene flow.

Isolation-by-distance is not robust across study systems; in a meta-analysis of IBD studies, Jenkins et al. (2010) found IBD to be inversely correlated with sample size, and thus best used with large sample sizes. In my own meta-analysis, many of the datasets used had small sample sizes of as few as four populations, potentially explaining the lesser predictive power of the isolation-by-distance models. Phenotypic traits are also significant in underlying IBD, such as modes of dispersal, endo- or ectothermy, and body size (Jenkins et al. 2010). Given the sheer range of phenotypic traits of the species used in my analysis, it is unsurprising that IBD lacked the predictive power of IBR.

Additionally, IBD suggests limited and random dispersal in contrast to IBR, which suggest that habitat fragmentation and environmental heterogeneity shape genetic differences between populations (Manthey & Moyle 2015). Dispersal is often limited in bird and mammal species, with many animals never dispersing from their natal or breeding site (Greenwood 1980).

3.4.4 Future Areas of Research

There are myriad factors that contribute to the way birds and mammals respond to habitat disturbance. I selected mass as a variable for both birds and mammals, and hand-wing index as a variable for birds, due to the known relationship between these variables and dispersal capability (Sheard et al. 2020, Whitmee & Orme 2013, Sutherland et al. 2000). However, other life history traits are known to influence dispersal. Herbivores have been shown to be especially sensitive to habitat fragmentation (Lino et al. 2019), and diets of animals could influence their response to

the HFI and other disturbances. Habitat use is also a factor in sensitivity to disturbance; arboreal and terrestrial mammals, and forest-dependent species, are more negatively affected by fragmentation (Lino et al. 2019). Human behavior, in addition to disturbance caused by human development, can also have a marked impact on species' response to urbanization (Clucas & Marzluff 2012). For example, in a study comparing behavioral responses of birds in Seattle and Berlin, birds at the American site had exaggerated flight distances suggesting greater wariness of humans; Seattle residents were found to be more repellant of birds, thus driving a behavioral response in urban bird species (Clucas & Marzluff 2012).

Literature Cited

Aguillon, S. M., Fitzpatrick, J. W., Bowman, R., Schoech, S. J., Clark, A. G., Coop, G., & Chen, N. (2017). Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. *PLoS Genetics*, *13*(8), 1–27.

Arnold, Z. J., Wenger, S. J., & Hall, R. J. (2021). Not just trash birds: Quantifying avian diversity at landfills using community science data. *PloS One*, 16(9), e0255391.

Baker, P. J., Ansell, R. J., Dodds, P. A., Webber, C. E., & Harris, S. (2003). Factors affecting the distribution of small mammals in an urban area. *Mammal Review*, 33(1), 95-100.

Balkenhol, N., Pardini, R., Cornelius, C., Fernandes, F., & Sommer, S. (2013) Landscape-level comparison of genetic diversity and differentiation in a small mammal inhabiting different fragmented landscapes of the Brazilian Atlantic Forest. *Conservation Genetics,* 14(2), 355-367.

Banks, S. C., Cary, G. J., Smith, A. L., Davies, I. D., Driscoll, D. A., Gill, A. M., ... & Peakall, R. (2013). How does ecological disturbance influence genetic diversity? *Trends in Ecology & Evolution*, 28(11), 670-679.

Bergl, R. A., Bradley, B. J., Nsubuga, A., & Vigilant, L. (2008). Effects of habitat fragmentation, population size and demographic history on genetic diversity: The cross-river gorilla in a comparative context. *American Journal of Primatology*, 70(8), 848–859.

Blair, R. B. (1999). Birds and butterflies along an urban gradient: surrogate taxa for assessing biodiversity? *Ecological Applications*, 9(1), 164-170.

Blair, R. B. (1996). Land use and avian species diversity along an urban gradient. *Ecological Applications*, 6(2), 506-519.

Brown, J.H. (1984). On the relationship between abundance and distribution of species. *American Naturalist*, 124, 255– 279.

Caula, S., Marty, P., & Martin, J. L. (2008). Seasonal variation in species composition of an urban bird community in Mediterranean France. *Landscape and Urban Planning*, 87(1), 1-9.

Centeno-Cuadros, A., Hulva, P., Romportl, D., Santoro, S., Stříbná, T., Shohami, D., ... & Nathan, R. (2017). Habitat use, but not gene flow, is influenced by human activities in two ecotypes of Egyptian fruit bat (*Rousettus aegyptiacus*). *Molecular Ecology*, 26(22), 6224-6237.

Chatterjee, H. J. (2006). Phylogeny and biogeography of gibbons: a dispersal-vicariance analysis. *International Journal of Primatology*, 27(3), 699-712.

Clucas, B., & Marzluff, J. M. (2012). Attitudes and actions toward birds in urban areas: human cultural differences influence bird behavior. *The Auk*, 129(1), 8-16.

Crispo, E., Moore, J. S., Lee-Yaw, J. A., Gray, S. M., & Haller, B. C. (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals: An examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *BioEssays,* 33(7), 508-518.

Crooks, K. R., Burdett, C. L., Theobald, D. M., King, S. R. B., Di Marco, M., Rondinini, C., & Boitani, L. (2017). Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. *Proceedings of the National Academy of Sciences of the United States of America*, 114(29), 7635–7640.

Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, 168(4), 486-499.

DePasquale, C., Li, X., Harold, M., Mueller, S., McLaren, S., & Mahan, C. (2020). Selection for increased cranial capacity in small mammals during a century of urbanization. *Journal of Mammalogy*, 101(6), 1706-1710.

Devictor, V., Julliard, R., Couvet, D., Lee, A., & Jiguet, F. (2007). Functional homogenization effect of urbanization on bird communities. *Conservation Biology*, 21(3), 741-751.

Dixo, M., Metzger, J. P., Morgante, J. S., & Zamudio, K. R. (2009). Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. *Biological Conservation*, *142*(8), 1560–1569.

Doherty, T. S., Hays, G. C., & Driscoll, D. A. (2021). Human disturbance causes widespread disruption of animal movement. *Nature Ecology & Evolution*, 5(4), 513-519.

Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, 6(8), 439-447.

Etienne, R. S., & Olff, H. (2004). How dispersal limitation shapes species–body size distributions in local communities. *The American Naturalist*, 163(1), 69-83.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699.

Frid, A., & Dill, L. (2002). Human-caused disturbance stimuli as a form of predation risk. *Conservation Ecology*, 6(1).

Fusco, N. A., Pehek, E., & Munshi-South, J. (2021). Urbanization reduces gene flow but not genetic diversity of stream salamander populations in the New York City metropolitan area. *Evolutionary Applications*, 14(1), 99-116.

Garner, A., Rachlow, J. L., & Hicks, J. F. (2005). Patterns of genetic diversity and its loss in mammalian populations. *Conservation Biology*, *19*(4), 1215–1221.

Gehlbach, F. R. (1996). A Model of Raptor Urbanization. Raptors in Human Landscapes: *Adaptation to Built and Cultivated Environments*, 69.

Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, 28(4), 1140-1162.

Greig, E. I., Wood, E. M., & Bonter, D. N. (2017). Winter range expansion of a hummingbird is associated with urbanization and supplementary feeding. *Proceedings of the Royal Society B: Biological Sciences*, 284(1852), 20170256.

Goudet, J. and Jombart, T (2020). hierFSTat: Estimation and Tests of Hierarchical F-Statistics. R package version 0.5-7.

Guralnick, R., Hantak, M. M., Li, D., & McLean, B. S. (2020). Body size trends in response to climate and urbanization in the widespread North American deer mouse, *Peromyscus maniculatus. Scientific reports*, 10(1), 1-13.

Hijmans, R. J. (2021). Geographic Data Analysis and Modeling [R package raster version 3.4-10].

Holderegger, R., Kamm, U., & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, 21(6), 797-807.

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating, and interpreting FST. *Nature Reviews Genetics*, 10(9), 639-650.

Holt, R. D. & Gomulkiewicz, R. (1997). How does immigration influence local adaptation? A reexamination of a familiar paradigm. *American Naturalist,* 149: 563–572.

Iglesias-Carrasco, M., Duchêne, D. A., Head, M. L., Møller, A. P., & Cain, K. (2019). Sex in the city: Sexual selection and urban colonization in passerines. *Biology Letters*, 15(9), 20190257.

Isaac, B., Cooke, R., Ierodiaconou, D., & White, J. (2014). Does urbanization have the potential to create an ecological trap for powerful owls (*Ninox strenua*)? *Biological Conservation*, 176, 1-11.

Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., ... & Tursi, R. (2010). A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, 33(2), 315-320.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070-3071.

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., ... & Purvis, A. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184. *Ecology*, 90(9), 2648-2648.

Kalinowski, S. T. (2005). Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity*, 94(1), 33-36.

Keyghobadi, N., Roland, J. E. N. S., & Strobeck, C. (2005). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology*, 14(7), 1897-1909.

Levin, S. A. (1992). The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, 73(6), 1943-1967.

Liebezeit, J. R., Kendall, S. J., Brown, S., Johnson, C. B., Martin, P., McDonald, T. L., ... & Zack, S. (2009). Influence of human development and predators on nest survival of tundra birds, Arctic Coastal Plain, Alaska. *Ecological Applications*, 19(6), 1628-1644.

Lino, A., Fonseca, C., Rojas, D., Fischer, E., & Pereira, M. J. R. (2019). A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. *Mammalian Biology*, 94(1), 69-76.

Manthey, J. D., & Moyle, R. G. (2015). Isolation by environment in white-breasted nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: A landscape genomics approach. *Molecular Ecology*, 24(14), 3628–3638.

Marzluff, J. M. (2001). Worldwide urbanization and its effects on birds. In Avian ecology and conservation in an urbanizing world (pp. 19-47). Springer, Boston, MA.

McCleery, R. (2010). Urban mammals. *Urban Ecosystem Ecology*, 55, 87-102.

McDonnell, M. J., & Pickett, S. T. (1990). Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology. *Ecology*, 1232-1237.

McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60(8), 1551-1561.

Meillère, A., Brischoux, F., Parenteau, C., & Angelier, F. (2015). Influence of urbanization on body size, condition, and physiology in an urban exploiter: a multi-component approach. *PloS one*, 10(8), e0135685.

Miles, L. S., Rivkin, L. R., Johnson, M. T. J., Munshi-South, J., & Verrelli, B. C. (2019). Gene flow and genetic drift in urban environments. *Molecular Ecology*, 28(18), 4138–4151.

Møller, A. P. (2008). Flight distance of urban birds, predation, and selection for urban life. *Behavioral Ecology and Sociobiology*, 63(1), 63-75.

Myhrvold, N. P., Baldridge, E., Chan, B., Sivam, D., Freeman, D. L., & Ernest, S. M. (2015). An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles: Ecological Archives E096-269. *Ecology*, 96(11), 3109-3109.

Niemelä, J. (1999). Ecology and urban planning. *Biodiversity & Conservation*, 8(1), 119-131.

Nilsen, E. B., Pedersen, S., & Linnell, J. D. (2008). Can minimum convex polygon home ranges be used to draw biologically meaningful conclusions? *Ecological research*, 23(3), 635-639.

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 108(455), 1-24.

Otto, S. P. (2018). Adaptation, speciation, and extinction in the Anthropocene. *Proceedings of the Royal Society B*, 285(1891), 20182047.

Paker, Y., Yom-Tov, Y., Alon-Mozes, T., & Barnea, A. (2014). The effect of plant richness and urban garden structure on bird species richness, diversity, and community structure. *Landscape and Urban Planning*, 122, 186-195.

Palumbi, S. (2001). Humans as the world's greatest evolutionary force. *Science,* 293(5526), 1786-1790.

Peterman, W. E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, 9(6), 1638-1647.

Peterman, W. E., & Pope, N. S. (2021). The use and misuse of regression models in landscape genetic analyses. *Molecular Ecology,* 30(1), 37-47.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL R-project.org.

Reichard, S. H., Chalker-Scott, L., & Buchanan, S. (2001). Interactions among non-native plants and birds. In Avian Ecology and Conservation in an Urbanizing World (pp. 179-223). Springer, Boston, MA.

Robb, G., Mcdonald, R., Chamberlain, D., & Bearhop, S. (2008). Food for thought: supplementary feeding as a driver of ecological change in avian populations. *Frontiers in Ecology and the Environment,* 6(9), 476-484.

Rousset, F., Lopez, J., & Belkhir, K. (2020). Package 'genepop'.

Santini, L., González-Suárez, M., Russo, D., Gonzalez-Voyer, A., von Hardenberg, A., & Ancillotto, L. (2019). One strategy does not fit all: determinants of urban adaptation in mammals. *Ecology Letters*, 22(2), 365-376.

Santini, L., Di Marco, M., Visconti, P., Baisero, D., Boitani, L., & Rondinini, C. (2013). Ecological correlates of dispersal distance in terrestrial mammals. *Hystrix*, 24(2), 181-186.

Sauvajot, R. M., Buechner, M., Kamradt, D. A., & Schonewald, C. M. (1998). Patterns of human disturbance and response by small mammals and birds in chaparral near urban development. *Urban Ecosystems*, 2(4), 279-297.

Savard, J. P. L., Clergeau, P., & Mennechez, G. (2000). Biodiversity concepts and urban ecosystems. *Landscape and Urban Planning*, 48(3-4), 131-142.

Scanes, C. G. (2018). Human activity and habitat loss: destruction, fragmentation, and degradation. In Animals and human society (pp. 451-482). Academic Press.

Schmidt, C., Domaratzki, M., Kinnunen, R. P., Bowman, J., & Garroway, C. J. (2020). Continent-wide effects of urbanization on bird and mammal genetic diversity. *Proceedings. Biological Sciences*, 287(1920), 2019-2497.

Schwartz, A. L., Williams, H. F., Chadwick, E., Thomas, R. J., & Perkins, S. E. (2018). Roadkill scavenging behaviour in an urban environment. *Journal of Urban Ecolog*y, 4(1), juy006.

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, 68(1), 1–15.

Sheard, C., Neate-Clegg, M. H., Alioravainen, N., Jones, S. E., Vincent, C., MacGregor, H. E., ... & Tobias, J. A. (2020). Ecological drivers of global gradients in avian dispersal inferred from wing morphology. *Nature communications*, 11(1), 1-9.

Snyder, J. P. (1987). Map projections--A working manual (Vol. 1395). US Government Printing Office.

Snyder, J. P. (1997). Flattening the earth: two thousand years of map projections. University of Chicago Press.

Spear, S. F., Balkenhol, N., FORTIN, M. J., McRae, B. H., & Scribner, K. I. M. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, 19(17), 3576-3591.

Sutherland, G. D., Harestad, A. S., Price, K., & Lertzman, K. P. (2000). Scaling of natal dispersal distances in terrestrial birds and mammals. *Conservation Ecology*, *4*(1).

Tamburello, N., Côté, I. M., & Dulvy, N. K. (2015). Energy and the scaling of animal space use. *The American Naturalist*, 186(2), 196-211.

Tratalos, J., Fuller, R. A., Warren, P. H., Davies, R. G., & Gaston, K. J. (2007). Urban form, biodiversity potential and ecosystem services. *Landscape and Urban Planning*, 83(4), 308-317.

United Nations, Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2018 Revision, Methodology. Report No. ESA/P/WP.252 (New York, NY, USA, 2018).

Venter, O., E. W. Sanderson, A. Magrach, J. R. Allan, J. Beher, K. R. Jones, H. P. Possingham, W. F. Laurance, P. Wood, B. M. Fekete, M. A. Levy, and J. E. Watson. (2018). Last of the Wild Project, Version 3 (LWP-3): 2009 Human Footprint, 2018 Release. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23(23), 5649–5662.

Whitmee, S., & Orme, C. D. L. (2013). Predicting dispersal distance in mammals: a trait-based approach. *Journal of Animal Ecology*, 82(1), 211-221.

Wildlife Conservation Society - WCS, and Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Willing, E. M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS One,* 7(8), e42649- e42649.

Worton, B. J. (1987). A review of models of home range for animal movement. *Ecological Modelling,* 38(3-4), 277-298.

Wright, S. (1943). Isolation by distance. *Genetics*, 28(2), 114.

Wright, S. (1986). Evolution: selected papers. University of Chicago Press.

Xu, Y., Kieboom, M., Van Lammeren, R. J., Si, Y., & De Boer, W. F. (2021). Indicators of site loss from a migration network: Anthropogenic factors influence waterfowl movement patterns at stopover sites. *Global Ecology and Conservation*, 25, e01435.

Tables & Figures

Table 7. A comparison of log-likelihood values for IBD and IBR (using HFI as the resistance surface) models for each mammalian dataset used in the meta-analysis.

Author	Species	Log-Likelihood IBD	Log-Likelihood HFI
Phuong	*Otospermophilus beecheyi*	7.006	50.875
Serieys	*Lynx rufus*	20.731	23.55
Colson	*Rangifer tarandus*	31.925	31.953
Epps	*Ovis canadensis*	254.61	254.61
Kierepka	*Taxidea taxus*	11.673	50.102
Lee	*Lynx rufus*	15.301	17.993
Rogic	*Peromyscus leucopus*	242	247.63
Marrotte	*Glaucomys volans*	107.9	110.08
Malenfant	*Ursus maritimus*	20.389	21.214
Cronin	*Bison bison*	100.64	100.92
Robinson	*Odocoileus virginianus*	9970.4	9976.4
Ernest	*Puma concolor*	70.774	74.334
Puckett	*Ursus americanus*	1110	1112.9
Serrouya	*Rangifer tarandus*	2115	2116.4
Taylor	*Peromyscus leucopus*	260.18	260.27
Taylor	*Peromyscus maniculatus*	260.18	260.27
Heppenheimer	*Canis latrans*	22244	22246
Cronin	*Ursus arctos*	319.6	319.62
Munshi-South	*Peromyscus leucopus*	302.28	302.51
Marrotte	*Lynx canadensis*	1800.1	1800.3
Bertrand	*Pekania pennanti*	1706.1	1709.3
Prentice	*Lynx canadensis*	29.518	29.74
Puckett	*Ursus americanus*	16.942	17.029
Draper	*Ursus americanus*	40.532	40.57
Latch	*Odocoileus hemionus*	5229.9	5230.1
Burns	*Myotis lucifugus*	553.27	553.27
Chavez	*Tamiasciurus douglasii*	458.33	458.36
Cheng	*Lepus americanus*	1752.4	1752.4
Peacock	*Ursus maritimus*	352.5	352.55
Laidre	*Ursus maritimus*	32.52	32.535
Anderson	*Microdipodops megacephalus*	14.702	14.743
Davy	*Myotis lucifugus*	947.87	950.59
Johnson	*Myotis lucifugus*	1926.6	1926.8
Mager	*Rangifer tarandus*	686.74	692.3
Rico	*Taxidea taxus*	74.359	80.772

Table 8. A comparison of log-likelihood values for IBD and IBR (using HFI as the resistance surface) models for each avian dataset used in the meta-analysis.

Author	Species	Log-Likelihood IBD	Log-Likelihood HFI
Adams	*Poecile atricapillus*	1870.1	1932.9
Adams	*Poecile atricapillus*	104.73	106.79
Barr	*Campylorhynchus brunneicapillus*	200.22	200.23
Durban	*Charadrius melodus*	59.297	60.293
Durban	*Charadrius montanus*	27.837	27.945
Gowen	*Aphelocoma californica*	2406.3	2412.4
Lait	*Poecile hudsonicus*	316.41	316.41
Liu	*Agelaius phoeniceus*	101.85	112.44
Machado	*Tyto alba*	191.62	202.37
Malpica	*Selasphorus platycercus*	61.101	63.217
Miller	*Calidris alpina*	21.164	49.315
Talbot	*Chen canagica*	33.586	34.251
Wilson	*Anser albifrons*	24.357	24.508
Wood	*Rallus obsoletus*	76.506	78.444

Table 9. Variables collected for each mammalian dataset; average adult mass, number of populations sampled in the study, range of HFI values each dataset was exposed to, and total area in square kilometers of minimum convex polygon encompassing all points. Regression coefficients and standard errors represent the effect of HFI on genetic distance without controlling for variables.

Author	**Species**	**Avg Adult Mass (g)**	**# Populations**	**Range of HFI in Sample Area**	**Minimum Convex Polygon (km^2)**	**Regression Coefficient**	**Standard Error**
Phuong	*Otospermophilus beecheyi*	599.6	3	99.99036	7,845.28	-4.539	8.084E-07
Serieys	*Lynx rufus*	9400	4	99.70771	119.92	-3.779	6.563
Colson	*Rangifer tarandus*	101250	5	80.03979	111,241.13	-2.029	13.508
Epps	*Ovis canadensis*	69125	14	99.70771	6,511.74	-1.39	1.254
Kierepka	*Taxidea taxus*	8000	3	100	12,718.29	-1.14	0.000004577
Lee	*Lynx rufus*	9400	4	99.70771	92.35	-1.123	1.09
Rogic	*Peromyscus leucopus*	22.3	11	92.97242	371.24	-1.0922	0.3752
Marrotte	*Glaucomys volans*	65.4	8	93	31,394.91	-1.0099	0.9609
Malenfant	*Ursus maritimus*	371703.81	4	59.41948	1,244,803.00	-0.8029	0.5301
Cronin	*Bison bison*	624577.07	10	99.99036	2,895,006.00	-0.6989	0.6297
Robinson	*Odocoileus virginianus*	65320	64	97	23,855.98	-0.5588	0.1527
Ernest	*Puma concolor*	48000	8	99.99036	182,268.83	-0.3728	0.113
Puckett	*Ursus americanus*	132405	37	100	10,576.90	-0.3369	0.1314
Serrouya	*Rangifer tarandus*	101250	38	96.43398	2,376,943.00	-0.3272	0.1956
Taylor	*Peromyscus leucopus*	22.3	12	93	41,199.39	-0.3203	0.847
Taylor	*Peromyscus maniculatus*	20.5	10	100	42,537.61	-0.3203	0.847
Heppenheimer	*Canis latrans*	11050	129	100	1,680,482.00	-0.3182	0.1699
Cronin	*Ursus arctos*	240500	16	95.60553	5,182,626.00	-0.3177	1.5658
Munshi-South	*Peromyscus leucopus*	22.3	13	95.485046	808.42	-0.2495	0.391
Marrotte	*Lynx canadensis*	9682.82	28	93	296,850.51	-0.2326	0.333
Bertrand	*Pekania pennanti*	3175	34	93	73,438.24	-0.19388	0.06979
Prentice	*Lynx canadensis*	9682.82	5	98.1252	519,231.27	-0.1749	0.2427
Puckett	*Ursus americanus*	132405	4	96.759399	10,845,110.00	-0.1745	0.4021
Draper	*Ursus americanus*	132405	7	95.255563	168,891.00	-0.1181	0.4344
Latch	*Odocoileus hemionus*	57000	71	99.99036	6,518,525.00	-0.05318	0.09715
Burns	*Myotis lucifugus*	7.15	15	98.1252	71,130.31	-0.05012	6.6014
Chavez	*Tamiasciurus douglasii*	229.8	19	94.07375	24,853.03	0.05258	0.25222
Cheng	*Lepus americanus*	1565.61	39	100	12,589,760.00	0.08145	1.03157
Peacock	*Ursus maritimus*	371703.81	14	95.60553	5,050,426.00	0.09044	0.28645
Laidre	*Ursus maritimus*	371703.81	4	26.69551	242,116.61	0.1298	0.8337
Anderson	*Microdipodops megacephalus*	13.405	4	99.67025	82,679.81	0.2031	0.6289
Davy	*Myotis lucifugus*	7.15	21	96.43398	4,196,782.00	0.3966	0.1423
Johnson	*Myotis lucifugus*	7.15	29	98.1252	101,327.40	0.4253	0.7123
Mager	*Rangifer tarandus*	101250	19	80.03979	950,515.58	1.5395	0.3402
Rico	*Taxidea taxus*	8000	8	99.84064	402,258.17	19.47	2.087E-08

Table 10. Variables collected for each avian dataset; hand-wing index, average adult mass, number of populations sampled in the study, range of HFI values each dataset was exposed to, and total area in square kilometers of minimum convex polygon encompassing all points. Regression coefficients and standard errors represent the effect of HFI on genetic distance without controlling for variables.

Author	Species	Hand-Wing Index	Avg Adult Mass (g)	# Populations	Range of HFI in Sample Area	Area Sampled (km^2)	Regression Coefficient	Standard Error
Adams	*Poecile atricapillus*	19.60	10.8	33	86.36091	13,537,030	3.984	0.511
Adams	*Poecile atricapillus*	19.60	10.8	8	83.37277	146,534	-0.1974	0.1113
Barr	*Campylorhynchus brunneicapillus*	13.72	38.9	12	75.36752	15,683	-0.2506	1.4347
Durban	*Charadrius melodus*	54.46	54.41	7	79.466	2,841,817	-0.1667	0.1072
Durban	*Charadrius montanus*	54.15	100.71	4	84.60049	124,910	0.1736	0.4279
Gowen	*Aphelocoma californica*	15.49	82.75	40	83.67086	1,602,368	1.084	0.2711
Lait	*Poecile hudsonicus*	18.43	9.8	13	89.231996	7,034,037	0.011	7.69E-01
Liu	*Agelaius phoeniceus*	23.99	54	7	75.71457	1,774,752	1.2612	0.3932
Machado	*Tyto alba*	41.40	381.6	10	75.96798	6,076,563	0.6592	1.25E-01
Malpica	*Selasphorus platycercus*	65.18	3.55	7	85.29252	394,466	1.6738	0.9294
Miller	*Calidris alpina*	54.34	54.6	3	76.759147	1,226,991	-0.1531	0.000005331
Talbot	*Chen canagica*	49.42	2718	5	46.937446	113,083	1.856	0.9322
Wilson	*Anser albifrons*	52.57	2387.5	4	75.685575	160,183	-0.3082	0.4168
Wood	*Rallus obsoletus*	18.15	294.83	8	67.2432	912	-3.118	1.364

Table 11. Weighted residuals and coefficients for the linear model of mammal datasets. The model estimate and standard error indicate the effect of HFI on genetic distance.

Weighted Residuals					**Coefficients**			
Min	**1Q**	**Median**	**3Q**	**Max**	**Estimate**	**Standard Error**	**T value**	**Pr(>|t|)**
-6.1085	0.3219	0.5520	0.9150	3.4543	-1.3946	0.2308	-6.042	7.6e-07

Table 12. Weighted residuals and coefficients for the linear model of bird datasets. The model estimate and standard error indicate the effect of HFI on genetic distance.

Weighted Residuals					**Coefficients**			
Min	**1Q**	**Median**	**3Q**	**Max**	**Estimate**	**Standard Error**	**T value**	**Pr(>|t|)**
-3.8657	-0.1515	-0.0064	0.6190	2.7107	0.1919	0.5262	0.365	0.721

Table 13. Correlation values between moderating variables from the mammal data using an R^2 coefficient of determination.

	Mass	**HFI Range**	**MCP**	**Populations**
Mass	1.0	-0.43	0.16	-0.22
HFI Range	-0.43	1.0	0.15	0.23
MCP	-0.16	0.15	1.0	0.14
Populations	-0.22	0.23	0.14	1.0

Table 14. Correlation values between moderating variables from the bird data using an R^2 coefficient of determination.

	Mass	**HFI Range**	**MCP**	**Populations**	**Hand-Wing Index**
Mass	1.0	-0.92	-0.29	-0.29	0.37
HFI Range	-0.92	1.0	0.32	0.27	-0.27
MCP	-0.29	0.32	1.0	0.15	-0.21
Populations	-0.29	0.27	0.15	1.0	-0.55
Hand-Wing Index	0.37	-0.27	-0.21	-0.55	1.0

Table 15. Results of the meta-analysis fixed-effects models for the mammal data, and the fixed-effects model while statistically controlling for each potential moderating variable. The first three models were 1) no filtration of data, 2) removal of datasets with small sample sizes, and 3) removal of small sample sizes and suspect effect sizes. Log-likelihood, deviance, and AICc are values assessing model fit; the estimate, standard error, and z- and p-values are indicative of the effect of HFI on genetic distances. The fixed-effects value with moderator estimate, standard error, and z- and p-values depict the effect of potential moderating variables on the overall meta-analytic model.

	Fixed-Effects Model							Fixed-Effects Model with Moderator			
	Log Likelihood	**Deviance**	**AICc**	**Estimate**	**Standard Error**	**Z Value**	**P Value**	**Estimate**	**Standard Error**	**Z Value**	**P Value**
Unfiltered	-391101840.2	782203681.7	782203682.6	18.7772	0.0001	131937.412	<0.001				
Remove ≥ *N=3*	-136494.1577	273006.2607	272990.4444	19.47	0	932918064	<.0001				
Remove ≥ *N=3* and Suspect Effect Sizes	-33.164	68.999	68.4762	-0.1782	0.0376	-4.7384	<.0001				
Range of HFI Exposure	-31.8249	66.3207	68.1113	0.9192	0.6716	1.3686	0.1711	-0.0114	0.007	-1.6366	0.1017
Mass	-32.8762	68.4233	70.2139	-0.1595	0.045	-3.5473	0.0004	-0.7587	0.448	0	0
Populations Sampled	-32.9137	68.4983	70.2889	-0.1418	0.0637	-2.2261	0.026	-0.0009	0.0013	-0.7076	0.4792
Minimum Convex Polygon	-28.4198	59.5104	61.301	-0.2558	0.0453	-5.6511	<.0001	0	0	3.0803	0.0021

Table 16. Results of the meta-analysis fixed-effects models for the bird data, and the fixed-effects model while statistically controlling for each potential moderating variable. The first three models were 1) no filtration of data, 2) removal of datasets with small sample sizes, and 3) removal of small sample sizes and suspect effect sizes. Log-likelihood, deviance, and AICc are values assessing model fit; the estimate, standard error, and z- and p-values are indicative of the effect of HFI on genetic distances. The fixed-effects value with moderator estimate, standard error, and z- and p-values depict the effect of potential moderating variables on the overall meta-analytic model.

	Fixed-Effects Model							Fixed-Effects Model with Moderator			
	Log Likelihood	**Deviance**	**AICc**	**Estimate**	**Standard Error**	**Z Value**	**P Value**	**Estimate**	**Standard Error**	**Z Value**	**P Value**
Unfiltered	-33.4738	64.2361	69.2809	-0.153	0.0023	-66.2595	<.0001				
Remove ≥ *N=3*	-63.0201	123.9015	128.4038	0.189	0.0606	3.116	0.0018				
Remove ≥ *N=3* and Suspect Effect Sizes	-30.7224	62.2599	63.8892	0.1413	0.0611	2.311	0.0208				
Range of HFI Exposure	-29.5689	59.953	64.6379	-1.744	1.2428	1.4033	0.1605	0.0193	0.0127	1.5188	0.1288
Hand-Wing Index	-30.496	61.8072	66.4921	0.2415	0.1609	1.5003	0.1335	-0.0027	0.0039	-0.6728	0.5011
Mass	-29.1388	59.0928	63.7777	0.0911	0.0673	1.3529	0.1761	0.0003	0.0002	1.7796	0.0751
Populations Sampled	-21.8137	44.4425	49.1274	-0.2037	0.1021	-1.9955	0.046	0.0359	0.0085	4.2211	<.0001
Minimum Convex Polygon	-22.2897	45.3945	50.0794	-0.1441	0.0926	-1.5566	0.1196	0	0	4.1067	<.0001

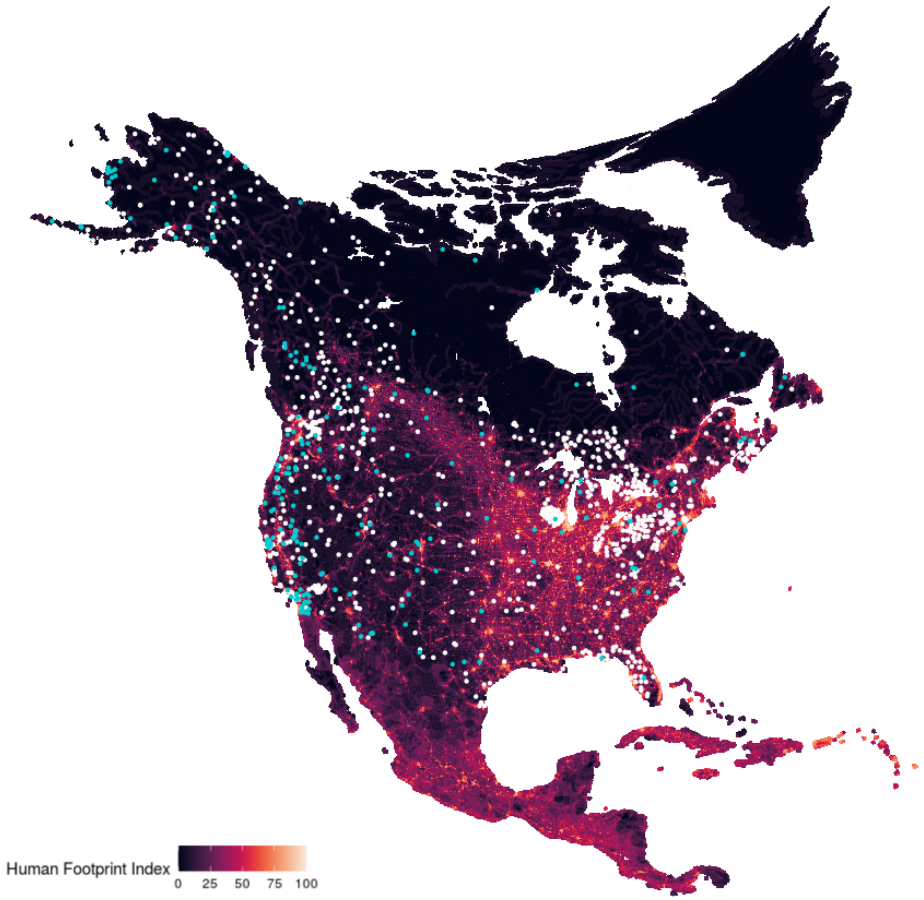

Figure 9. A map of the Human Footprint Index (HFI), on a scale of 0-100, across North America. Sample sites of bird studies used in the meta-analysis are plotted in blue, and mammal sample sites are plotted in white.

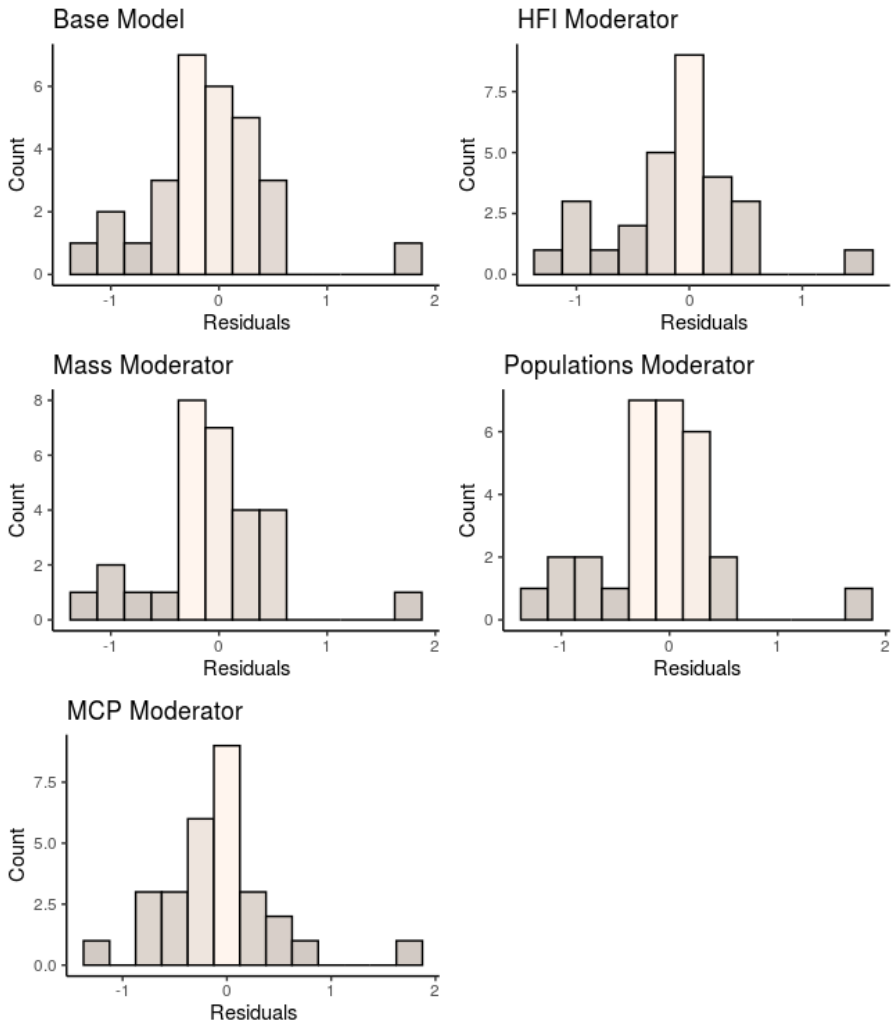

Figure 10. Bar plots of residuals for the meta-analytic fixed-effects mammal model, and the fixed-effects model with each moderator variable.

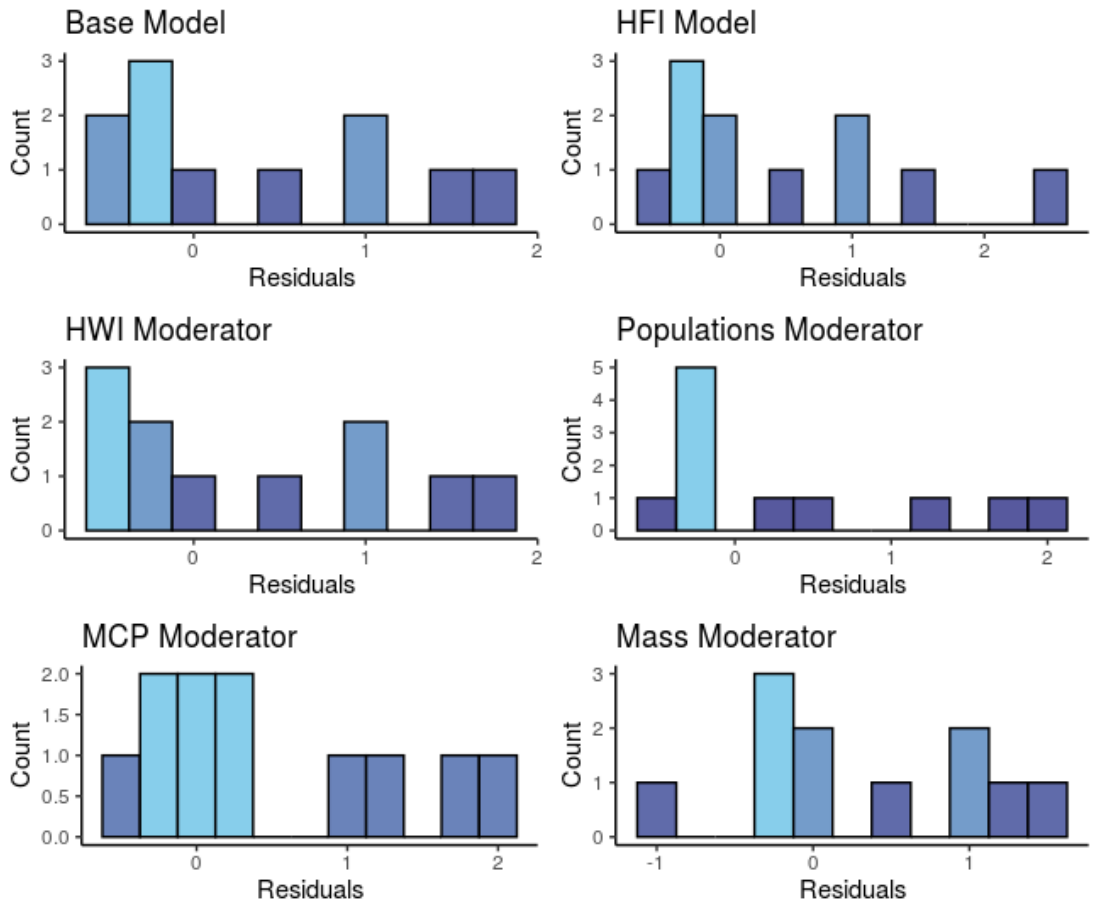

Figure 11. Bar plots of residuals for the meta-analytic fixed-effects bird models, and the fixed-effects model with each moderator variable.

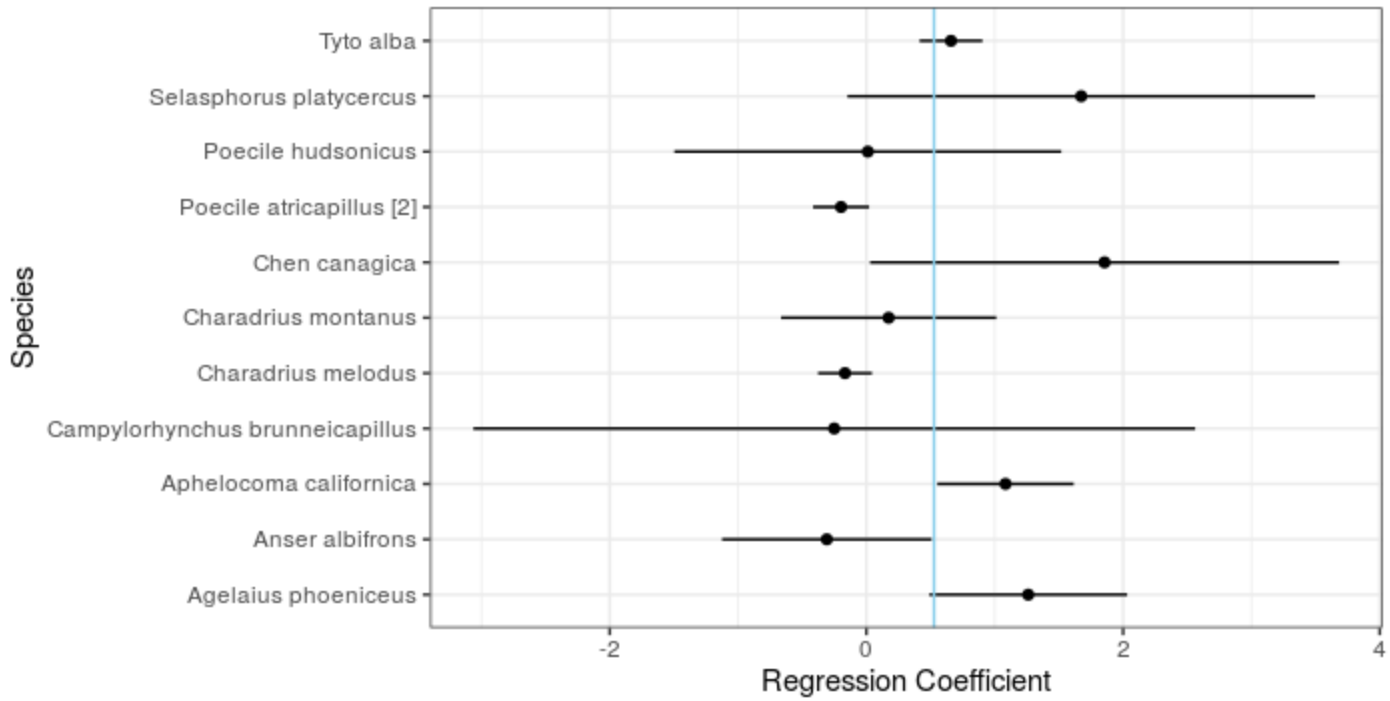

Figure 12. The regression coefficients for each bird species used in the meta-analysis: the points on the plot represent the coefficient value, and the horizontal line on either side of each point represents the standard error associated with the coefficient. The overall trend is depicted as a blue horizontal line.

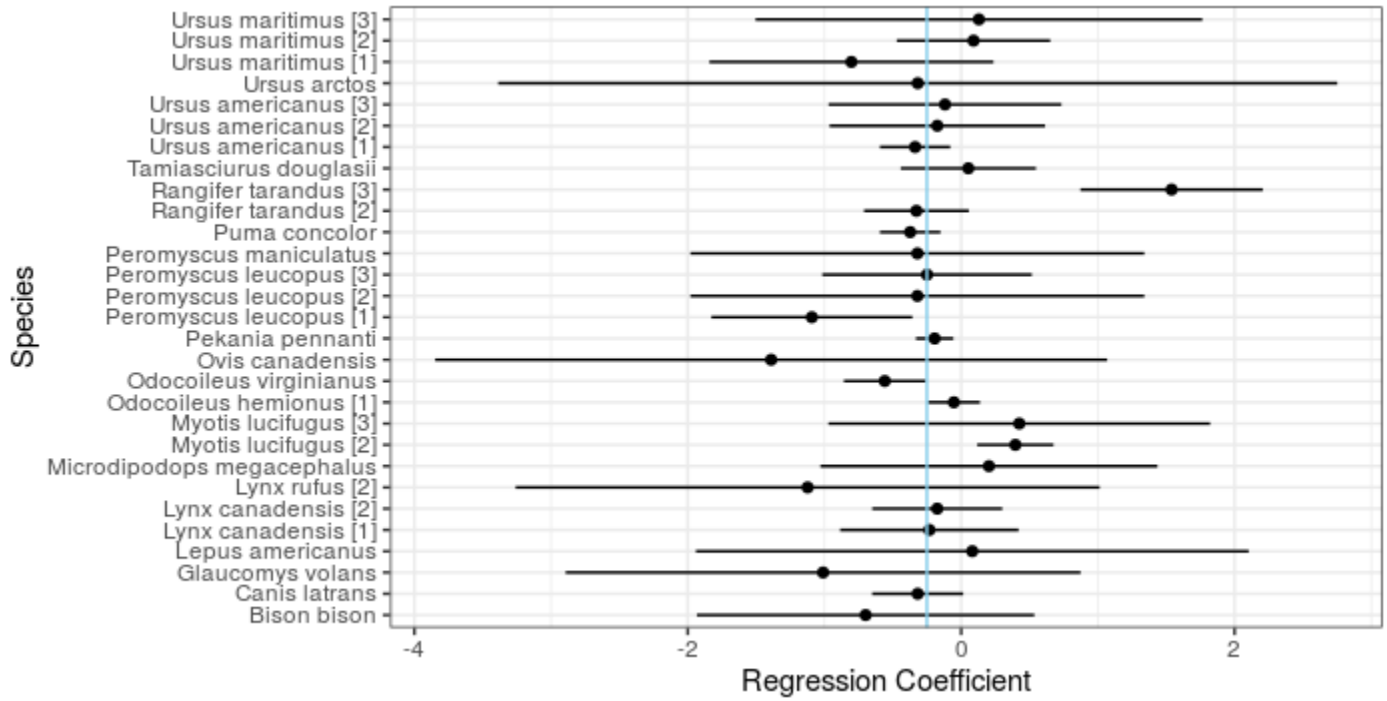

Figure 13. The regression coefficients for each mammal species used in the meta-analysis: the points on the plot represent the coefficient value, and the horizontal line on either side of each point represents the standard error associated with the coefficient. The overall trend is depicted as a blue horizontal line.