**Investigating the genomics underlying fish movement and behaviour**

by

Matthew John Thorstensen

A Thesis submitted to the Faculty of Graduate Studies of the

University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Biological Sciences

University of Manitoba

Winnipeg, Manitoba

Canada

**Abstract**

Movement of organisms through space has large effects on population dynamics, reproduction, and persistence. Movement of genetic material can sway the evolution of populations by influencing local adaptation and population structure. Therefore, information on movement is of fundamental and conservation importance for wild organisms, including freshwater fishes that are especially vulnerable to environmental change. Movement has been connected to environmental, physiological, and genomic factors in fish, although most research has focused on anadromous species. Walleye (*Sander vitreus*) is an economically and culturally important obligate freshwater fish with a native range across North America, in which movement is relatively less well-understood than in anadromous fish. Lake Winnipeg walleye are the largest component of the second-largest freshwater commercial fishery in Canada, contribute to a large recreational fishery, and provide subsistence for First Nations communities around the lake.

To provide a genomic context for movement studies, Lake Winnipeg walleye population structure was established using genomic variants from RNA, where a metapopulation characterized by basin-specific subpopulations was found. Gene flow was found to likely move from south to north in the lake. Gene flow between nearby Lake Manitoba and Lake Winnipeg was found to be slightly stronger going into Lake Winnipeg, but low overall (approximately 0.0027 fish per year) and only started between 550 and 1,219 years ago. Signatures of protein degradation were higher in north basin walleye of Lake Winnipeg, consistent with a possibly longer travel distance needed for those fish to reach spawning areas. Movement within Lake Winnipeg walleye was studied in the context of maximum residency, where fish that stayed in any one ecological zone were given a 1.0 residency value. Fish that spent an even proportion of time among three ecological zones (south basin, narrows, north basin) were given a $0.3\overline{3}$ residency value. No genomic variants were found to strongly affect residency, but 81.6% of variation in maximum residency was explained when using a polygenic approach. These results provide information that can inform future management and demonstrate the evolutionary and biochemical signatures that interact with different perspectives of walleye movement.

## Acknowledgements

This thesis may never have existed without the support of Evelien de Greef. Thank you. I love you, Evelien. Mocha and Opal love you, as well.

Ken Jeffries has been a wonderful advisor. During a global pandemic, Ken went above and beyond in supporting myself and the rest of the Jeffries Lab. I am deeply grateful.

Will Bugg, Jen Jeffrey, Theresa Mackey, Mike Lawrence, Linh Miller, Sonya Michaleski, Tina Schaefer, Carolyn Vandervelde, McKenzie Hauger, Josh Sutherby and the other members of the Jeffries Lab all contributed to an excellent place in which to do a PhD, and all have influenced my scientific thinking. Luke and Lindsey Belding have been lovely neighbors, friends, and Luke introduced me to lake sturgeon biology. I miss you, Luke.

Jay Treberg, Gary Anderson, and Colin Garroway have given me new perspectives on biology in the conversations we shared.

Nick Rice has been a steadfast friend, who has made my time in grad school more fulfilling and joyful. John and Barb Little made my transition to Canada much smoother than it would have otherwise been, and John could have made much more fun of me than he did, for how bad I am at angling for walleye. Bob and Ilsa McLandress opened their cottage to me, and my conversations with Bob about the ecology of Lake Winnipeg continue to help me think about connections in the natural world. My parents, Jay and Nani Thorstensen, have been supportive from the beginning.

## Dedication

For the walleye, and the people who depend on them.

**Table of Contents**

**List of Tables**

**List of Figures**

# Chapter 1. Introduction

## 1.1 Perspectives of Movement

Movement, ‘a change in the spatial location of an individual in time’, can be fundamental to animal survival, reproduction, and persistence (Nathan et al., 2008). Evolutionarily, movement can be considered in terms of gene flow or effective migration rates, where admixture and introgression between populations can change the dynamics of local adaptation and ultimately sway the evolution of populations (Cayuela et al., 2018). Whether gene flow is itself movement, or indicates that organisms moved in the past before reproducing, depends on whether genetic material or individual organisms are considered the units of movement. Methods that observe changes in the spatial location of a whole individual are generally concerned with the individual as a unit, while evolutionary perspectives of movement are concerned with genetic material. Given the importance of movement for individuals and population evolution, information on the factors underlying movement holds both fundamental and conservation significance. Freshwater fish are one group vulnerable to environmental change while providing valuable ecosystem services (Krabbenhoft et al., 2020; Su et al., 2021). However, movement information in obligate freshwater fish is not as extensive as in anadromous species (Hussey et al., 2015; Thompson et al., 2020). Therefore, information on factors underlying freshwater fish movement is both relevant for conservation and can provide important insights for wildlife biology.

Fish movement can be considered at three scales: individual-level movement, population-level movement, and metapopulation-level movement (Bailey & Muths, 2019). At the individual level, behavioural asymmetries (e.g., lateralization of escape responses), habitat use, home ranges, and links between individual movement and morphology or environmental characteristics have been studied (Bailey & Muths, 2019; Clobert et al., 2009; Dadda et al., 2010). Population-level movement has been described in terms of migration, or the back-and-forth movement undertaken to match suitable ecological conditions in variable environments (Cote et al., 2017). Metapopulation-level movement involves multiple populations, in which dispersal (movement of an individual from a birth patch to a breeding patch; Cayuela et al., 2018) is a topic of broad research interest as dispersal has implications for both demographic and

genetic connectivity, along with population dynamics via immigration and emigration (Clobert et al., 2009). Interestingly, dispersal can be considered an individual-level movement with large effects on metapopulations, as dispersing organisms tend to move alone (relative to more synchronized migrations), but the act of dispersal may have large effects on multiple populations (Cote et al., 2017). Such an observation highlights the somewhat arbitrary distinctions among different scales of movement, but nevertheless, distinguishing among these scales provides a useful framework for studying different factors that may contribute to and be affected by fish movement. For example, physiological methods for studying movement will be based at the individual level, such as research that describes the energetic costs of migration with implications for mortality (Cooke et al., 2004). At scales above the individual, evolution and gene flow occur among groups of individuals (Cayuela et al., 2018).

An evolutionary approach to movement is largely concerned with dispersal because connectivity between populations (i.e., metapopulation movement and gene flow) is an important and tractable area of study for population genetics (Cayuela et al., 2018). Here, genotypes are used to estimate migration rates (the proportion of individuals emigrating from one population to another, more accurately called the effective dispersal rate because migration is two-way), which has implications for inbreeding, population differentiation, and ultimately, adaptation and speciation (Cayuela et al., 2018). However, a gap remains in our understanding of the genomics underlying movement within populations, such as predictive models based on genotypes at the individual level. Genomics and population genetics have been effectively used to describe movement between groups of organisms, but direct links between genomics and individual movement tend to be less common (Cayuela et al., 2018; Clobert et al., 2009). Dispersal differences have been linked to individual genetic variation in several cases, such as in rhesus macaques (*Macaca mulatta*), dispersing earlier with a variant of the serotonin metabolism-related gene *slc6a4*, with possible implications for different male behavioural strategies (Trefilov et al., 2000). Migratory versus non-migratory phenotypes have also been linked to chromosomal rearrangements in Atlantic cod (*Gadus morhua*) and genetic correlations with body size in Atlantic salmon (*Salmo salar*) (Debes et al., 2020; Kess et al., 2019). In Pacific salmon (*Oncorhynchus* spp.), association testing (i.e., GWAS) has described a genomic basis for migration timing (Brieuc et al., 2015; Hess et al., 2016; Koch & Narum, 2020; O'Malley et al., 2013; Prince et al., 2017; Thompson et al., 2020). The multiple lines of evidence for a strong

genomic influence on migration timing in salmon are a proof of concept that critical life-history traits may be influenced by genomics and that questions linking individual-level movement to genomics are now tractable.

Physiological and biochemical methods to test predictions associated with movement hypotheses are possible at the individual or population-scale. For example, salmon (*Oncorhynchus* spp. & *Salmo salar*) migration has been studied from physiological perspectives of thermal stress, cardiorespiratory performance, disease states, and stress responses because of the deep importance of salmonids to ecosystems and human societies (Cooke et al., 2012; Stich et al., 2015). Individual-scale connections between 17-β estradiol and migration were found in both direct experimental treatments and in messenger RNA profiles of Atlantic and sockeye salmon, with implications for anthropogenic contaminants changing migratory behaviours (*S. salar* & *O. nerka;* Madsen et al., 2004; Veldhoen et al., 2013). Cortisol has been linked to migration success in Atlantic salmon and sea trout (*S. trutta*), with implications for a tradeoff between stress and migration readiness (Birnie-Gauvin et al., 2019). Cortisol elevation and decline during migration was consistent between landlocked kokanee salmon (*O. nerka kennerlyi*) and anadromous fish, providing evidence that cortisol is part of an endogenous system that affects migration (Carruth et al., 2000). Biochemical and metabolic connections with movement have been observed in several fishes as well, such as in tradeoffs between activity and growth (in yellow perch, *Perca flavescens*; Rennie et al., 2005), increased reliance on protein for energy during extended swimming at elevated speeds (in Nile tilapia, *Oreochromis niloticus*; Alsop et al., 1999), and protein degradation and synthesis each increased with swimming activity (in rainbow trout, *Salmo gairdneri*; Houlihan & Laurent, 1987). Related to protein metabolism, migration has been linked to increased amino acid and carbon flux in sockeye salmon (*O. nerka*) (Mommsen, 2004). While physiology and biochemistry are certainly affected by and affect movement, specific pathways used to address movement likely differ between species, possibly based on phylogenetic constraints and life history strategies.

### 1.2 Methods for Observing Movement

Animal movement can be studied using two general approaches: mark-recapture, where organisms are observed over multiple sampling events, and telemetry, where a transmitter is used

to track an animal through its environment (Bailey & Muths, 2019; Cayuela et al., 2018). In mark-recapture studies with fishes, small tags (e.g., floy tags, Floy Tags, Seattle, Washington, USA) are used to individually identify organisms at a tagging event and in a subsequent recapture. Mark-recapture approaches are useful for estimating census population sizes in addition to dispersal, because an organism's locations of original capture and recapture can show dispersal, and the rates of recapture can be used to model population size (Cayuela et al., 2018; Nichols, 1992). Archival tags and passive integrated transponder tags (PIT-tags) are more advanced methods for studying fish movement, but usually require a subsequent interaction with the fish similar to mark-recapture methods (Thorstad et al., 2013). Pop-up satellite archival tags collect information for a certain period then relay it to receivers via satellites, which are helpful for situations where a receiver array has not been built and do not require retrieval of the organism or tag (Thorstad et al., 2013). Radio telemetry-based studies previously focused on large collars or tags on animals, and global positioning systems (GPS) similarly required relatively large tags (e.g., tigers, *Panthera tigris*; Miller et al., 2010). More recently, radio and GPS tag miniaturization has allowed for their use with smaller animals such as birds, as well (Geen et al., 2019).

Acoustic tags are another approach for tracking movement, and are useful in part because radio waves are blocked by salt water (Hussey et al., 2015). Both radio and acoustic telemetry arrays can be built, which allow for the tracking of multiple tagged individuals without a researcher needing to track each individually (i.e., autonomously) (Cooke et al., 2008; Donaldson et al., 2014). Acoustic telemetry arrays have become commonly used for a broad range of fish species as the tags have become smaller in size over time (Hussey et al., 2015). Satellite-based GPS tags are shrinking in size as well, but the cost-efficiency of acoustic tags for tracking large numbers of individuals carries enormous potential for large-scale studies involving many research organisms (Donaldson et al., 2014; Hussey et al., 2015). The increased variety of fish species that can be tracked using acoustic telemetry arrays has potential for studying interactions among entire fish communities, because multiple species may be tagged and concurrently use the space an array monitors (Enders et al., 2019; Hussey et al., 2015; Rudolfsen et al., 2021).

Predation transmitters are acoustic movement tags implanted in a fish, that react to acidic conditions in a predator's stomach to change their acoustic identifiers (Halfyard et al., 2017). This change in identifier marks a predation event, and tags can subsequently record the predator's movement (Halfyard et al., 2017). These tags were used to assess predation of yellow perch by unknown predators, although sample size may be limited in this approach (e.g., $n$=19 predation events in one study; Weinz et al., 2020). While thus far the technology has only been applied to distinguish movement before and after predation events, it has the potential to describe some aspects of predator feeding behaviour and physiology (see Lidgard et al., 2014 for an example with pinnipeds and predated fish, but inferred from normal acoustic tags).

### 1.3 Methods for Connecting Movement & the Environment

Animal behaviour is strongly influenced by the environment in general, and movement is no exception (Clobert et al., 2009; Hussey et al., 2015; Nathan et al., 2008). Because habitat selection (and therefore, movement) and temperature are intimately linked in fish (Huey, 1991), studies of the consequences of temperature while it moves can provide a wealth of knowledge about fish biology and energy use (Holland et al., 1992). Most fish are ectotherms, and therefore environmental temperature shapes nearly every aspect of fish biology, including biogeographic distributions and individual fitness (Payne et al., 2016). In a laboratory experiment, temperature was found to elicit different movement responses below and above two years of age in Atlantic salmon (*S. salar*), where oxygen consumption in age ≥2 fish plateaued at higher temperatures in conjunction with cool-water aggregation behaviour, compared to fish <2 years of age (Breau et al., 2011). Larval swimming performance in several fish species has been linked to temperature, where hydrodynamic factors are more prominent for larvae exposed to cold water, but physiological factors dominate in warm water exposures (von Herbing, 2002). In walleye (*Sander vitreus*), temperature was shown to play a role for seasonal migrations in a large lake, among other factors (Raby et al., 2018). However, no consistent effect of temperature was identified in wild migrating Atlantic salmon (*S. salar*), although decreasing water flow was associated with increased migratory movement (Trépanier et al., 1996). Salinity is an important environmental factor as well, because ion and water flux in fish are reversed between marine and freshwater environments (McCormick & Saunders, 1987). For example, numerous physiological

changes are associated with anadromous migrations in Pacific and Atlantic salmon (*Oncorhynchus* spp. & *Salmo salar*) as the fish prepare for the salinity transition between the marine and freshwater environments (McCormick & Saunders, 1987). Salinity has been linked to migration barriers in Sacramento splittail (*Pogonichthys macrolepidotus*) in an estuarine environment with fluctuating salinity (Jeffries et al., 2019). In freshwater, salinity has directly been shown to be an important factor in largemouth bass movement (*Micropterus salmoides*; Meador & Kelso, 1989). Given the importance of measuring standard environmental variables to understand movement patterns, including these data in movement studies allows a more comprehensive understanding of habitat use in fishes.

Environmental DNA, where species-specific mitochondrial DNA is sampled directly from water (or now, air; Clare et al., 2021), is another form of environmental sampling that can describe fish movement. For instance, researchers were able to identify spawning behaviour in bigheaded carp (*Hypophthalmichthys* spp.) with downstream water samples (Erickson et al., 2016). Migration using known movement over fish ladders was linked to environmental DNA presence as well, along with signals that revealed aspects of winter habitat use (Yamanaka & Minamoto, 2016). Indirect sampling with DNA from the environment has clear potential for non-invasively describing fish movement, but is limited by sources of error such as degraded DNA or water flow that may lead to false negatives, or contaminated equipment that may lead to false positives (Jerde, 2021).

### 1.4 Methods for Connecting Movement, Gene Expression, & Physiology

Tagging to track an individual through an environment involves an initial interaction where the tag is applied, and this interaction is when biological tissue may be collected. These additional tissue biopsies (here, gill & blood) did not significantly introduce additional mortality in an experiment with sockeye salmon (Cooke et al., 2005). Carbon isotopes measured from non-lethal muscle tissue biopsies were associated with movement after sampling in broadnose sevengill sharks (*Notorynchus cepedianus*), where combined movement and isotope data revealed intrapopulation feeding differences (e.g., inshore vs. offshore individuals) (Abrantes & Barnett, 2011). An enormous body of work exists on teleost movement physiology, especially in anadromous salmon during migrations (Hinch et al., 2005; McCormick et al., 1998). More

recently, combined gill and blood samples that showed that stressed sockeye salmon (*Oncorhynchus nerka*) returned to fresh water earlier (Drenner et al., 2018). Gill tissue was also useful for predicting migration vs residency behaviour in brown trout (*S. trutta*) by analyzing $Na^+/K^+$-ATPase activity (Nielsen et al., 2004). Blood plasma sampled with subsequent radio transmitter and temperature logger tagging revealed how different factors such as temperature, blood glucose, dam discharge, and time of arrival at spawning grounds changed different metrics of reproductive success in female sockeye salmon (Minke-Martin et al., 2018).

Links between gene expression and movement have been studied with quantitative polymerase chain reaction (qPCR), microarrays, and RNA sequencing (Mantione et al., 2014). With these approaches, the abundance of individual genes in groups of fish relative to others are compared (usually migration survival vs mortality), with possible implications for functional response pathways if gene annotations are informative (Connon et al., 2018; Jeffries et al., 2014; Miller et al., 2011). Gene expression was correlated with environmental changes during migration and with successful migrants (*O. nerka*), showing that transcriptional regulation was predictive of survival (Evans et al., 2011). Gene expression has predicted migration success in sockeye salmon (Jeffries et al., 2014; Miller et al., 2011), although a gene expression panel with qPCR did not find links with migration survival among age-1 and age-2 sockeye (Stevenson et al., 2020). Another study showed that viral signatures of gene expression were connected to a higher predation risk in migrating sockeye salmon (Furey et al., 2021), showing that molecular signatures of disease status can be linked to interspecific predator-prey interactions.

Gene expression has often been integrated with other physiological indices. Migration success, infectious agents, gene expression with qPCR, and plasma variables (cortisol, lactate, potassium, sodium, leukocrit, among others) were analyzed in Chinook salmon (*O. tshawytscha*), which provided additional evidence that disease status may be an important predictor of migration success in anadromous fish (Bass et al., 2019). Disease state and stress was interrogated with gene expression (via qPCR) and blood plasma osmolality, sodium, chloride, glucose, lactate, cortisol, testosterone, and 17β-estradiol in migrating sockeye salmon (Drenner et al., 2018). This combination of stress and disease state provides further evidence for the importance of those factors in migration and revealing that stressed fish may enter fresh water earlier, but experience subsequently higher mortality (Drenner et al., 2018). Hatchery steelhead

(*O. mykiss*) migration survival was predicted by gene expression via qPCR as well, with inflammatory response genes associated with mortality (Healy et al., 2018). In each of the gene expression studies described previously, fish were sampled non-lethally for gill tissue. Muscle tissue can be useful as well, such as in linking gene expression, migration success, and several organohalogens in sockeye and chinook salmon (Veldhoen et al., 2010). Here, the sockeye salmon were found to experience greater biological stress than the chinook, in conjunction with endocrine disrupting chemicals (Veldhoen et al., 2010). Brain, gill, muscle, and liver tissue from Coho salmon (*O. kisutch*) revealed how gene expression may change with important life history events, with implications for physiological preparations for migration (Houde et al., 2019). Other work has also connected physiological and transcriptomic patterns, such as staggered increases in protein turnover, protein biosynthesis, immune responses, oxidative phosphorylation, and glycolytic potential as sockeye salmon underwent anadromous migrations (Miller et al., 2009). The gene expression and physiology-based studies discussed here may show a bias towards fish with anadromous migrations, but nevertheless demonstrate the enormous potential for describing how movement may be affected by different factors during fish's lifetimes.

### 1.5 Methods for Connecting Movement & Genetics

Population genetics has been useful for describing the extent to which evolution and movement interact (Cayuela et al., 2018). Here, dispersal is often linked to gene flow and population structure (Cayuela et al., 2018). Effective and non-effective dispersal can be distinguished by a disperser successfully reproducing or not, along with historical versus recent gene flow in the context of effective dispersal (Cayuela et al., 2018). For example, in Arctic charr (*Salvelinus alpinus*), genetic assignment with DNA data revealed estimated dispersal rates, revealing a system with high dispersal but low gene flow (Moore et al., 2013). However, population structure did not distinguish migratory ecotypes in the same species (Moore et al., 2014), but was associated with overwintering habitat choice and genes for muscle and cardiac function (e.g., *mef2*, *pygm*, or in other words, genes that affect muscle energy use or muscle development) (Moore et al., 2017). Population structure described movement in lemon sharks (*Negaprion brevirostris*), where DNA data and acoustic telemetry were used to develop a migration-residency model for predicting aggregations (Kessel et al., 2014). Genetic assignment,

where population structure is used to place individuals in groups, allowed for dispersing brown trout to be identified (Östergren et al., 2012). Quantitative genetics with experimentally-manipulated pedigrees in Atlantic salmon have revealed a 50-70% heritability for migrant probability (Debes et al., 2020). Variance in migratory traits has been linked with genomics using quantitative genetic approaches in arthropods, fish, and birds, although connections with movement were often indirect (Liedvogel et al., 2011). Association testing in migration timing of Pacific salmon (*Oncorhynchus* spp.) has revealed a strong genomic basis for early versus late-run fish linked with the genes *greb1* and *rock1*, an estrogen receptor and rho-associated kinase, respectively (Brieuc et al., 2015; Hess et al., 2016; Koch & Narum, 2020; O'Malley et al., 2013; Prince et al., 2017; Thompson et al., 2020). A chromosomal rearrangement was associated with migration in Atlantic cod (Kess et al., 2019), showing that an association-based approach can also reveal structural genomic variants possibly underlying migration phenotypes. Sea age, or the age at which Atlantic salmon return from marine environments to spawn, was associated with muscle development and metabolic genes, among others, and has implications for connections between life history strategies, movement, and genomics (Johnston et al., 2014).

In summary, four broad types of approaches to studying movement with genetics were described here with different focuses: effective dispersal, population structure, association testing, and quantitative traits. Their perspectives on movement are all ultimately evolutionary, but take different views of how movement can be described by genomics. For instance, investigations focused on effective dispersal are not concerned with the particular genomic regions that may change movement behaviours, but are appropriate for considering demography and population dynamics (Cayuela et al., 2018). Population structure requires assessments of differentiation among groups, and may only be relevant for describing movement if the movement phenotype observed somehow contributes to a reduction in heterozygosity via non-random mating (or correlates with a phenotype that contributes to non-random mating, i.e., Wahlund effect; De Meeûs, 2018; Waples & Allendorf, 2015; Weir & Cockerham, 1984). Association testing and quantitative genetics are related in that they attempt to quantify the extent to which genomics affects variance in movement, and often describe candidate genomic regions that underlie movement phenotypes (e.g., Debes et al., 2020; Kess et al., 2019; Thompson et al., 2020, although Moore et al., 2017 demonstrates a population structure-based approach to movement with candidate regions). Therefore, association tests and quantitative

genetics are useful for a mechanistic view of genomics and movement, similar in approach to gene expression studies that predict migration success based on the differential abundance of specific genes (e.g., Jeffries et al., 2014; Miller et al., 2011).

### 1.6 Walleye

Walleye are a percid fish native to freshwater systems across Canada and the USA, and hold cultural and economic significance to societies across those countries (Baccante & Colby, 1996; Bootsma et al., 2020; Raabe et al., 2020). While a freshwater-obligate species, unlike many of the anadromous fish previously identified as examples of links between movement and genomics, walleye still show broad seasonal movements in the large lakes they inhabit (Munaweera Arachchilage et al., 2021; Raby et al., 2018; Turner et al., 2021). As such, walleye is an excellent species to explore connections between movement and genomics in freshwater systems. There is current debate over the scientific name of walleye being either *Sander vitreus* or *Stizostedion vitreum*. *Sander vitreus* was based on Latvian vernacular after a suspected misunderstanding by an early taxonomist (Bruner, 2021). For consistency throughout this thesis, *Sander vitreus* is used, but debate over the species' scientific name is ongoing.

Walleye in Lake Winnipeg, Manitoba, comprise the largest component of the second-largest freshwater commercial fishery in Canada (Manitoba walleye 4,028 tonnes landed in 2019 worth $162 million CAD; Fisheries and Oceans Canada, 2021), contribute to the largest recreational fishery in the country (26% of total recreational catch nationwide with total $2.5 billion CAD in direct recreational fishing expenditures in 2015; Fisheries and Oceans Canada, 2019), and provide subsistence for First Nations communities around the lake (ECCC & MARD, 2020). Lake Winnipeg is among the largest freshwater lakes in the world based on surface area, and is characterized by a north and a south basin connected by a narrows (Fig. 1.1; Brunskill et al., 1980). No population differentiation in walleye between basins was found in a microsatellite-based study (Backhouse-James & Docker, 2012), although scale morphometry revealed some stock structure characteristic of a metapopulation and interacting subpopulations (Watkinson & Gillis, 2005). As such, Lake Winnipeg walleye may be predicted to move between basins because gene flow has decreased genomic differentiation between basins, despite some morphological differences (Backhouse-James & Docker, 2012; Watkinson & Gillis, 2005). However, between approximately 2.5% and 17.5% movement probability was observed between

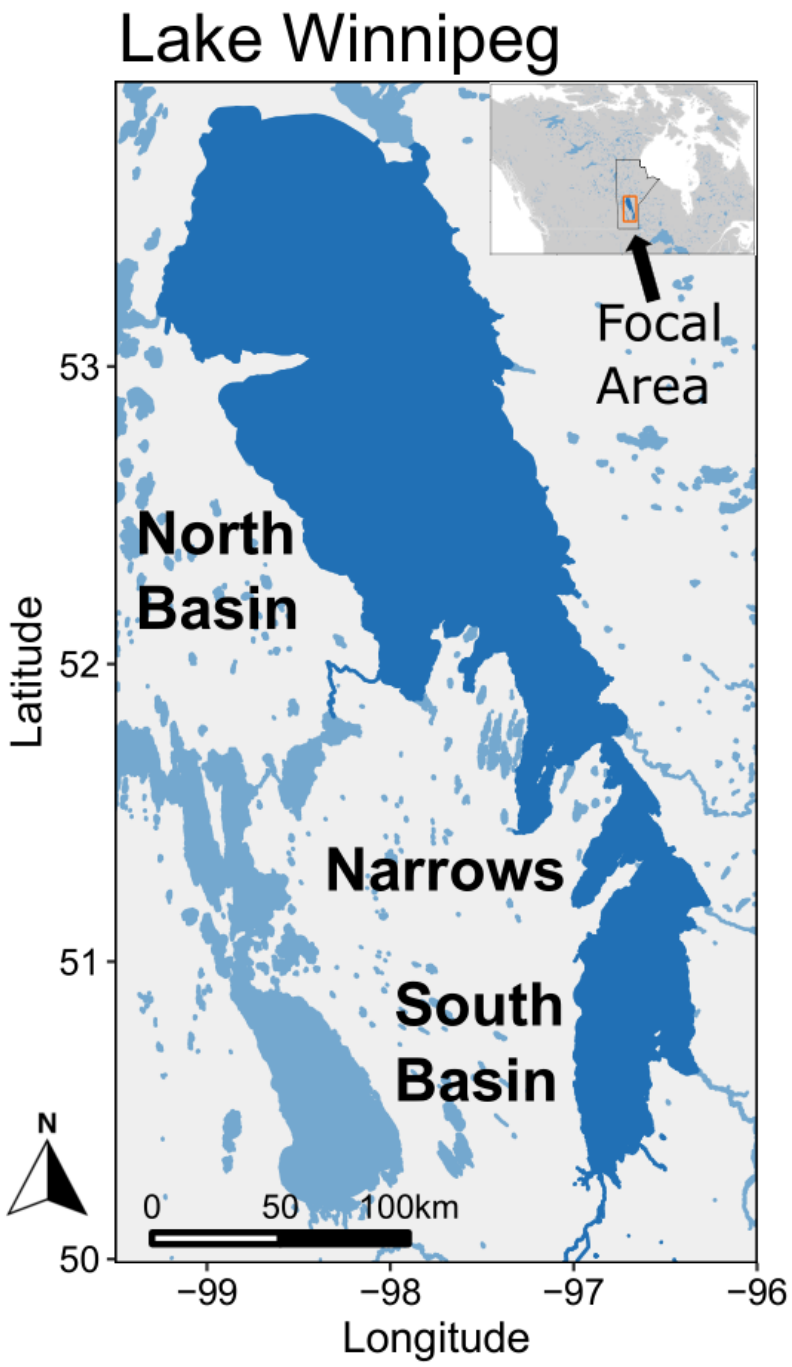


**Fig 1.1** Map of Lake Winnipeg, Manitoba, Canada, with the narrows connecting the North and South basins. Area labels are approximations, see Chapter 5 for precise delineations of ecological zones in Lake Winnipeg. Lake Winnipeg and rivers studied (Red River in the south basin, Dauphin River in the north basin) are coloured in dark blue. Other waterbodies, streams, and rivers are in light blue.

basins in acoustic telemetry array data, indicating low but observable between-basin movement (Figure 2 in Turner et al., 2021). Walleye in the south basin were more likely to move into the north basin than vice versa (Turner et al., 2021), which suggests that habitat use likely differs among groups. Therefore, Lake Winnipeg walleye are a largely undifferentiated population in which movement differences exist among individuals, possibly related to subpopulation structure.

Previous research in movement and genomics has considered migration timing as the movement phenotype (e.g., Thompson et al., 2020), but habitat use, preference, and selection are all critical aspects of organism's life histories (Mayor et al., 2009). Prior research has indicated a potential genetic basis for habitat selection in arthropods and molluscs, although the data preceded high-throughput genomic sequencing (Jaenike & Holt, 1991). In addition, most research connecting movement to genomics and other factors have been in anadromous fishes (especially salmon and trout, genera *Oncorhynchus* & *Salmo*; e.g., Debes et al., 2020; Jeffries et al., 2014; Miller et al., 2011; Minke-Martin et al., 2018; Thompson et al., 2020). Lake Winnipeg walleye are an ideal system in which to investigate this connection between movement and genomics in a freshwater species, where less is known about factors that underlie fish movement.

### 1.7 Thesis Methods & Objectives

Analyses of fish movement from different perspectives connect the studies assembled for this thesis. Chapters 2 and 4 have adopted an evolutionary perspective on movement, applying it to the population level and the metapopulation level, respectively. Each is based on genetic differentiation, although Chapter 2 uses genetic variation in RNA, while Chapter 4 uses genetic variation in DNA. Chapters 3 and 5 have adopted the perspective of movement within the lifetime of individual fish and within a population. Chapter 3 focuses on a panel of nine blood metabolites used to reveal patterns of protein degradation and amino acid breakdown for energy, in conjunction with changing growth rates over time. Chapter 3's connections with movement are indirect, based on the assumption that there are longer distances north basin fish must travel to reach spawning sites. Chapter 5 employs genome-wide association to connect movement and genomics, where genetic variants are associated with movement from acoustic telemetry data. A

polygenic model is used to explore the variance in movement explained by genomic variants of weak effect.

This thesis asks the question, *what are the ecological and evolutionary consequences of seasonal freshwater fish movement, as described with an ecological genomic approach?* The fish studied here are primarily walleye from Lake Winnipeg, along with smaller groups from Lake Manitoba, Manitoba, Canada and Cedar Bluff Reservoir, Kansas, USA.

Four objectives were explicated to answer this question:

**1)** Establish fine-scale patterns of genetic population structure with the implications for movement within Lake Winnipeg and between waterbodies (Chapters 2 & 4).
**Hypothesis:** Lake Winnipeg walleye are a panmictic population.
**Prediction:** Population differentiation between walleye sampled in different basins is low, but significant. Walleye in Lake Winnipeg are therefore one population with subpopulation structure.

**2)** To explore movement using blood samples, spatial signatures of protein degradation and amino acid metabolism were tested (Chapter 3).
**Hypothesis:** North basin walleye must travel further to reach spawning sites, therefore will have traveled further prior to sampling. This movement will be reflected in their metabolites.
**Prediction:** Signatures of amino acid oxidation for energy and protein degradation will be elevated in north basin walleye, relative to south basin walleye.

**3)** To explore movement from an evolutionary perspective by testing evidence of gene flow within Lake Winnipeg and between Lake Manitoba and Lake Winnipeg (Chapters 2 & 4).
**Hypothesis:** Walleye follow water flow from the south basin of Lake Winnipeg to the north, and from Lake Manitoba to the north basin of Lake Winnipeg.
**Prediction:** Signatures of admixture from gene flow in Lake Winnipeg walleye will be more common among individuals sampled in the north basin.

**4)** To explore direct connections between movement in Lake Winnipeg walleye and genomic variants, association testing and polygenic modeling was performed (Chapter 5).

**Hypothesis:** Movement between basins is influenced by numerous factors, extrinsic and intrinsic to a fish. Genomic variants that influence movement do so indirectly and through many mechanisms, indicating a distributed genetic architecture.

**Prediction:** Few or no genomic regions will be strongly associated with between-basin movement. Instead, movement will be moderately influenced by DNA as a polygenic trait, with many variants of weak effect contributing to explaining a portion of variance in the phenotype.

# Chapter 2. Genomic Signals Found Using RNA Sequencing Show Signatures of Selection and Subtle Population Differentiation in Walleye (*Sander vitreus*) in a Large Freshwater Ecosystem.[1]

## 2.1 Abstract

RNA sequencing is an effective approach for studying aquatic species yielding both physiological and genomic data. However, its population genetic applications are not well-characterized. I investigate this possible role for RNA sequencing for population genomics in Lake Winnipeg, Manitoba, Canada, walleye (*Sander vitreus*). Lake Winnipeg walleye represent the largest component of the second-largest freshwater fishery in Canada. In the present study, large female walleye were sampled via nonlethal gill biopsy over two years at three spawning sites representing a latitudinal gradient in the lake. Genetic variation from sequenced mRNA was analyzed for neutral and adaptive markers to investigate population structure and possible adaptive variation. I find low population divergence ($F_{ST}$ = 0.0095), possible northward gene flow, and outlier loci that vary latitudinally in transcripts associated with cell membrane proteins and cytoskeletal function. These results indicate that Lake Winnipeg walleye may be effectively managed as a single demographically connected metapopulation with contributing subpopulations, and suggest genomic differences possibly underlying observed phenotypic differences. Despite its high cost relative to other genotyping methods, RNA sequencing data can yield physiological in addition to genetic information discussed here. I therefore argue that it is useful for addressing diverse molecular questions in the conservation of freshwater species.

[1] A modified version of this chapter has previously been published in open-access format. © Creative Commons 4.0. Matt J Thorstensen conducted the analyses, and Matt J Thorstensen and Ken M Jeffries wrote the first draft of the manuscript. Jennifer D Jeffrey, Matt J Thorstensen, and Douglas A Watkinson collected data. Jason R Treberg, Eva C Enders, and Ken M Jeffries conceptualized the study and acquired funding. All authors were critical in developing the ideas presented and gave final approval for publication. Found in: Thorstensen, M. J., Jeffrey, J. D., Treberg, J. R., Watkinson, D. A., Enders, E. C., & Jeffries, K. M. (2020). Genomic signals found using RNA sequencing show signatures of selection and subtle population differentiation in walleye (*Sander vitreus*) in a large freshwater ecosystem. Ecology and Evolution, 10(14), 7173–7188. https://doi.org/10.1002/ece3.6418

## 2.1 Introduction

Population abundances in aquatic systems are in decline globally, with a 36% decline in the marine Living Planet Index (LPI, http://livingplanetindex.org) and an 81% decline in the freshwater LPI between 1970 and 2012 (WWF, 2016). The LPI is an average rate of change in population size aggregated to the species level, based on population time-series data. The decline in the freshwater LPI is especially alarming for those ecosystems, which cover 2.3% of the earth's global land surface area but are disproportionately high in species richness—for instance, one-third of all described vertebrate species live in freshwater (Reid et al., 2019; WWF, 2018). It is therefore a significant concern that freshwater species are declining in abundance more rapidly than terrestrial or marine species (Reid et al., 2019). This decline underscores an urgent need for research supporting conservation efforts for these diverse freshwater organisms.

To take effective action, conservation practitioners require research on the environmental stressors a population faces, as well as population structure and evolutionary patterns to determine a population's adaptive potential (Connon et al., 2018; Russello et al., 2012; Waples & Gaggiotti, 2006). For instance, RNA sequencing and differential gene expression analysis can be used to conduct population risk assessments by identifying physiological thresholds, thus possibly informing management decisions (Connon et al., 2018). An advantage of using RNA sequencing for conservation research is that it provides information about both genetics and molecular physiology by returning transcript abundances and single nucleotide polymorphisms (SNPs) allowing researchers to gather a diverse array of information within one data set. Moreover, a *de novo* transcriptome can be generated from a dataset of RNA reads using computational tools (e.g. Trinity described in Grabherr et al., 2011), therefore researchers can take a set of samples collected in the field and develop a profile of physiological status and genomic patterns without relying on species-specific probes or primers. These advantages make transcriptomics approaches useful for studying species of conservation concern, especially for species that do not have extensive molecular databases like those available for model species (e.g., zebrafish, *Danio rerio*).

Applications of RNA sequencing to address population genomics questions in non-model species is relatively less characterized compared to DNA sequencing. Several studies that have used RNA sequencing for population genomics describe domestication selection, such as in fish

or plants (Christie et al., 2012; Gros-Balthazard et al., 2019), while others address adaptive variation (Barts et al., 2018; Brown et al., 2018; Passow et al., 2017; Pratlong et al., 2015). Adaptive variation may be useful for conservation by delineating units with functional differences (Funk et al., 2012; Russello et al., 2012). While many genomic methods rely on genes in linkage with outlier SNPs of interest to infer the functional significance of the SNP, the functional significance of SNPs in mRNA is often more readily interpretable because SNPs are more likely to occur in protein-coding sequences than SNPs called from whole-genome data (Verta & Jones, 2019). Within a transcript, the effects of SNPs in open reading frames can be predicted, which allows inferences of how protein function may be modified by genetic variation (Cingolani et al., 2012). Transcriptomics is then valuable for its potential to investigate the dynamic between phenotypic plasticity and evolution (Barts et al., 2018; DeBiasse & Kelly, 2016; Kelly, 2019; Passow et al., 2017; Pratlong et al., 2015).

A large topic of interest in conservation genomics is population structure, or genomic divergence between different groups of individuals, which can support decisions on whether those groups should be managed as a single unit or several units (Funk et al., 2012). Here, population structure in RNA sequencing studies is often considered in the context of adaptation or functional variation (De Wit & Palumbi, 2013; Hoey & Pinsky, 2018), but relatively few transcriptomic studies make population structure a focus of their analyses. Therefore, RNA sequencing may be an effective method for characterizing physiological patterns, population structure, and adaptive variation in species and systems with little prior information available.

Walleye (*Sander vitreus*) in Lake Winnipeg, Manitoba, are the largest component of the second largest freshwater fishery in Canada. Lake Winnipeg is characterized by a north and a south basin connected by a narrow channel (Johnston et al., 2012; Fig 2.1). While previous microsatellite research showed limited population differentiation between groups in each basin except at two out of thirteen sampling locations (Backhouse-James & Docker, 2012), morphological, life history, dietary, and environmental differences among Lake Winnipeg walleye suggest diverging genetic histories (Environment Canada, 2011; Johnston et al., 2012; Moles et al., 2010; Sheppard et al., 2015, 2018; Watkinson & Gillis, 2005). Within Lake Winnipeg, walleye have shown declining biomass and body condition, decreased catches, and commercial harvests above maximum sustainable yields for several years (Manitoba

Government, 2018; Manitoba Sustainable Development, 2018). Observations of dwarf walleye suggest signs of selection against large, economically desirable fish, especially in the south basin of the lake (Johnston et al., 2012; Moles et al., 2010). One possibility is that phenotypic plasticity may underlie walleye responses to environmental differences between basins (Kelly, 2019), and this plasticity may be expressed as changes in walleye morphology, life history, and diet over time and across latitudes. An important first step in investigating possibly plastic patterns is to establish patterns of population structure underlying potential divergence, especially because genomic techniques can have higher resolution than microsatellite markers (Funk et al., 2012). In addition, plastic and evolutionary responses have overlapped in genes showing either changed expression or signatures of selection (Kelly, 2019). Therefore, information on population structure and biological differences in Lake Winnipeg walleye may help distinguish between plastic and evolutionary phenotypic responses, which may support future conservation efforts.

The current study aimed to show how mRNA sequencing can be an effective approach for developing critical pieces of information applicable to fisheries and conservation practitioners. I used RNA sequencing for genetic characterization of Lake Winnipeg walleye sampled from known spawning locations that potentially represent fish from the north and south basins. I also sampled fish collected at the channel that connects the north and south basins as an intermediate site. I hypothesized that walleye populations within Lake Winnipeg show evidence of distinct population differentiation identified using RNA sequencing data, despite the weak signatures from microsatellite data (Backhouse-James & Docker, 2012). I predicted that the walleye population divergence may partially reflect the different environments and natural histories between the north and south basins of Lake Winnipeg.

## 2.3 Methods

### 2.3.1 RNA Extraction and Sequencing

Gill tissue was collected from large (≥ 1.2 kg) predominately female (44 female, 4 unidentified sex) (Table A.1.1) walleye over two years from three sites in the Lake Winnipeg system (Red River, Matheson Island, and Dauphin River, representing sites in the south basin,

channel, and north basin, respectively; Fig 2.1; Table A.1.1; $n$ = 8 per year and site, $n$ = 48 total). With $n$ = 16 individuals per site over both years, sample sizes can predict mean population allele frequency by site, because sample sizes as low as $n$ =12 have been shown to be sufficient with RNA sequencing data (Schunter et al. 2014). These fish were sampled during the spawning season (approximately May through early June in 2017 and 2018). Walleye were collected by electrofishing, held in a live well for no longer than one hour, and anaesthetized using a Portable Electroanesthesia System (PES™, Smith Root, Vancouver, Washington, USA) in accordance with approved animal use protocols of Fisheries and Oceans Canada (FWI-ACC-2017-001, FWI-ACC-2018-001), the University of Manitoba (F2018-019) and the University of Nebraska-Lincoln (Project ID: 1208). Fish were sampled non-lethally for gill tissue, where 2–3 mm of the terminal ends of 3–4 filaments from the left side each fish were collected and placed in RNA*later* (Thermo Fisher Scientific, Waltham, Massachusetts, USA) that was kept at 4 °C for 24 h prior to storage at -80 °C. Total RNA extractions were performed on gill tissue using RNeasy Plus Mini Prep Kits (QIAGEN, Venlo, Netherlands) following manufacturer's protocols, with minor modifications (provided in section A.1.1.1).

The quantity and quality of RNA was assessed with a Nanodrop One Spectrophotometer (Thermo Fisher) and electrophoresis on a 1% agarose gel, respectively. Total RNA was normalized to 80 ng $\mu L^{-1}$ and sent to the McGill University and Génome Québec Innovation Centre sequencing facility (http://gqinnovationcenter.com) for cDNA library preparation and sequencing. A mRNA isolation approach was performed on 250 nanograms of total RNA using the NEBNext Poly(A) Magnetic Isolation Module (New England BioLabs, Ipswich, Massachusetts, USA). From mRNA, stranded cDNA libraries were created using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Each library was individually barcoded with NEBNext dual adaptors (New England Biolabs) prior to sequencing 100 base pair paired end reads. All 48 fish were sequenced on a single lane of a NovaSeq 6000 (Illumina, San Diego, California, USA). 2.17 billion reads total were sequenced, with an average of 45,225,548 reads per sample collected (5,071,090 s.d.) (Table A.1.1).

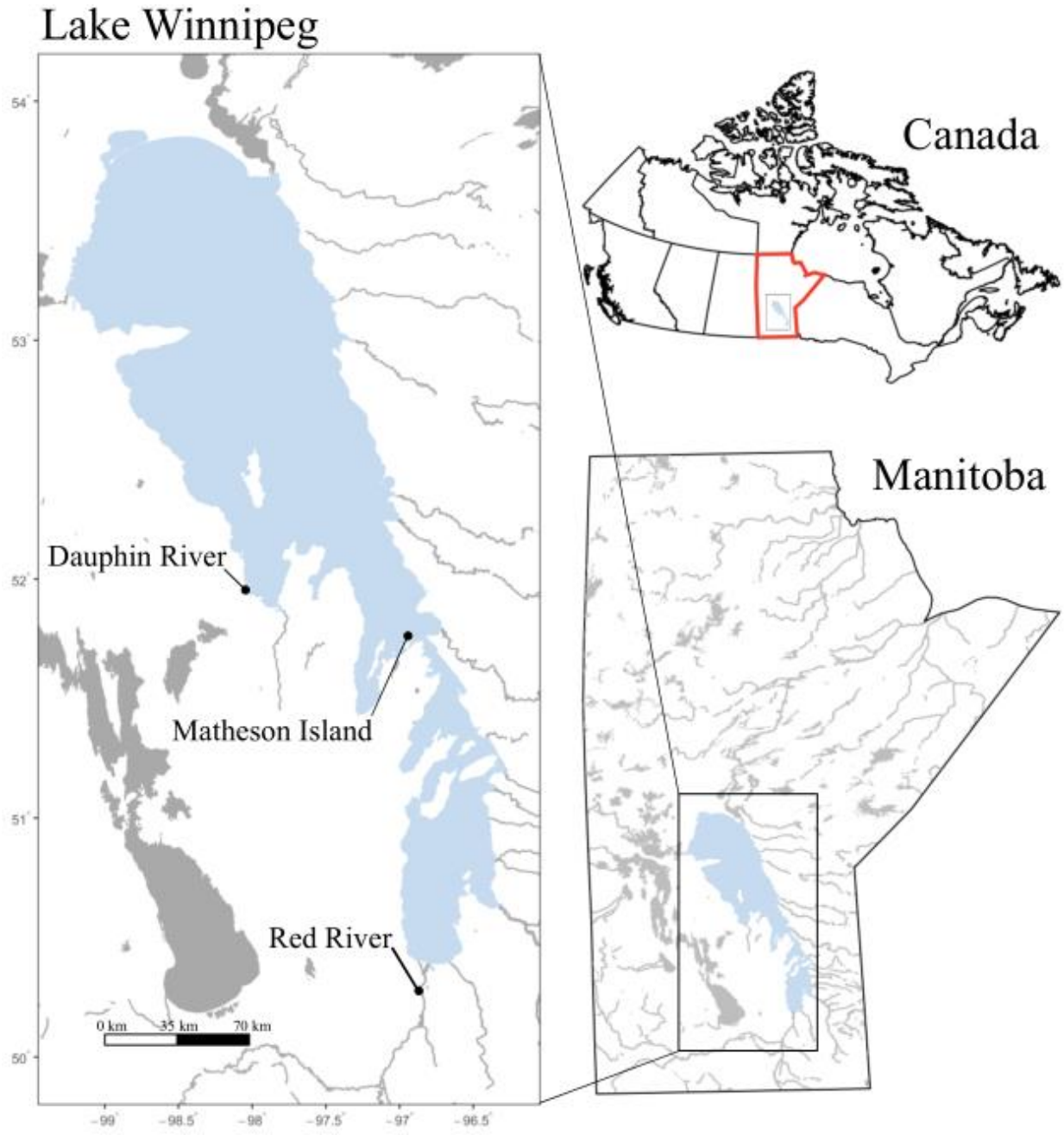


**Fig 2.1** Sampling locations within Lake Winnipeg, Manitoba, Canada. Eight walleye (*Sander vitreus*) per year and per spawning site were collected, for $n$ = 48 fish over 2017 and 2018. The Red River represents the south basin, Matheson Island represents the channel connecting the two lake basins, and the Dauphin River represents the north basin.

### 2.3.2 SNP Calling

Raw read files were uploaded to the Graham and Cedar clusters on the Westgrid section of the Compute Canada partnership (https://www.westgrid.ca/). Read files were checked for quality using FastQC version 0.11.8 (Andrews, 2010) and trimmed using Trimmomatic version 0.36 (Bolger et al., 2014). When using FastQC version 0.11.8 (Andrews, 2010), the program was set to allow two seed mismatches, a palindrome clip threshold of 30 nucleotides, and simple clip threshold of ten nucleotides. A sliding window size of 4 base pairs was used to filter data for a minimum Phred 64 quality of five, with five nucleotides trimmed from both the leading and trailing ends of reads, and a minimum read length of 36. After trimming, FastQC was used again to verify data quality. Scripts used for the analyses in this manuscript are provided at https://github.com/BioMatt/Walleye_RNAseq.

The SuperTranscripts pipeline was used to align reads (Davidson et al. 2017) for SNP calling. First, Salmon version 0.11.3 (Patro et al., 2017) was used to quantify read counts, as compared to a previously assembled reference transcriptome for walleye (Sequence Read Archive Accession SRP150633; Jeffrey et al., 2020). In Salmon, validate mappings, range factorization bins of size 4, sequencing bias, and GC bias options were all used, along with dumping equivalence classes for subsequent steps. Using the count estimates from Salmon, Corset version 1.07 (Davidson & Oshlack, 2014) was used to cluster the data for assembly into SuperTranscripts. A linear representation of the transcriptome was constructed with Lace version 1.00 (https://github.com/Oshlack/Lace) using information from Corset and the original transcriptome, where 263,272 unique gene features were identified using Trinotate (Grabherr et al., 2011) in the original transcriptome were gathered into 148,165 super clusters. Following Lace, STAR version 2.7.0a (Dobin et al., 2013) was used in 2-pass mode to align trimmed reads to the reassembled transcriptome. Here, annotated junctions from Lace were provided along with the new transcriptome, and sjdbOverhang of 99 was chosen following recommended settings of 1 base pair below read length. A minimum of 79.6% reads uniquely mapped to the Lace-clustered transcriptome (mean 81.5% ± 0.5% s.d.) (Table A.1.1).

For calling SNPs, the STAR-aligned reads were processed with Picard version 2.18.9 by adding read groups, splitting cigar ends, and merging bam files (Broad Institute, 2019), then SNPs were called using FreeBayes version 1.2.0 (Garrison & Marth, 2012). Detailed methods for

calling SNPs are provided in section A.1.1.2. This resulted in 2,458,947 SNPs and 586,556 indels, which were used as unfiltered data for subsequent steps. I next filtered the VCF file from FreeBayes in two ways to 1) identify putatively neutral SNPs for population structure analyses and 2) SNPs for which neutrality was not assumed, using for outlier tests and functional analyses. Because purifying selection may be widespread throughout expressed RNA, and I wished to study SNPs as close to neutrality as possible, several filtering steps were taken to create a putatively neutral SNP dataset (Gossmann et al., 2010). VCFtools version 0.1.14 was used to filter for biallelic SNPs of genotype and site quality ≥30, minimum minor allele frequency ≥0.05, no missing data in any individual, and in Hardy-Weinberg Equilibrium with a *p*-value < 0.005 (Danecek et al., 2011). SNPRelate version 1.16.0 was then used to prune SNPs for linkage disequilibrium at a threshold of 0.20, where super clusters were coded as chromosomes for the purposes of linkage disequilibrium pruning (Zheng et al., 2012). These steps resulted in a putatively neutral data set of 52,372 SNPs used for population structure analyses. For a broader subset of SNPs for which neutrality was not assumed, VCFtools was used to filter for genotype and site quality ≥30, minimum minor allele frequency ≥0.05, and a maximum of two out of 48 possible individuals missing data. 222,634 SNPs were retained from these filtering steps, which were then used for outlier tests and functional analyses.

### 2.3.3 Population Structure

To investigate population structure using the 52,372 putatively neutral SNPs, I used a combination of exploratory analyses, either with no prior information or with sampling location provided as priors, and population reassignment and differentiation tests to find genetic clusters despite possible signals of admixture or gene flow. Structure version 2.3.4 (Falush et al., 2003, 2007; Hubisz et al., 2009; Pritchard et al., 2000) was run with no prior location or population information, an initial value of alpha of 1.0, a maximum value of alpha of 10.0, prior mean $F_{ST}$ of 0.01, lambda of 1.0, a burn in period of 10,000 repetitions, and 110,000 Markov Chain Monte Carlo repetitions after burn in. Structure plots were visualized with pophelper version 2.2.7 (Francis, 2017). Ten replicates of K = 2–5 were tested.

For analyses performed in R (R Core Team 2021), the package vcfR was used to format genotype data for use with other programs (Knaus & Grünwald, 2017). Adegenet version 2.1.1

(Jombart et al., 2010) was used in two ways. First, in an exploratory capacity to perform Discriminant Analysis of Principal Components (DAPC), where sampling location was provided for the DAPC as prior population information. Second, population structure was investigated irrespective of sampling location by using cluster identification from successive K-means, as implemented in the find.clusters function in Adegenet. Here, different numbers of clusters were explored in the data (40 principal components were retained for exploratory steps) and evaluated with a Bayesian Information Criterion (BIC), where the most well-supported number of clusters with lowest BIC was 2 (Fig A.1.1). In addition to exploring the two clusters, the population assignments from three clusters were used to explore genetic differentiation in the data because fish were sampled from three sites (Table A.1.1). With Hierfstat version 0.04-22 (Weir & Cockerham, 1984; Yang, 1998), the Weir & Cockerham's pairwise $F_{ST}$ was calculated among the three sampling locations, then between the two reassigned clusters described by Adegenet (Table A.1.1). I generated 95% confidence intervals for these $F_{ST}$ values in Hierfstat using a bootstrap approach over 1,000 iterations.

To visualize population differentiation, I used a PCA as implemented in Adegenet version 2.1.1 (Jombart et al., 2010) and t-SNE as implemented in Rtsne version 0.15 (Van Der Maaten & Hinton, 2012). For the t-SNE, two final dimensions were used, with 100 initial dimensions, 15 perplexity, theta of 0.5, and 5,000 iterations. Perplexity is, approximately, a measure of effective neighbors for a point (Van Der Maaten & Hinton, 2012). These approaches were used with the same settings applied to the putatively neutral SNPs, and visualizations were thus comparable between data sets.

### 2.3.4 Temporal Stability & Kinship

To test for temporal stability in the data, I created subsets of individuals caught in 2017 and 2018. As with the whole dataset, Weir & Cockerham's pairwise $F_{ST}$ was calculated both among sampling locations and between the two reassigned clusters, and generated 95% confidence intervals over 1,000 bootstrapped iterations in hierfstat (Table A.1.1). Modest results that are consistent over time support confidence in a real genetic signal, as opposed to results driven by bias which are more likely to be inconsistent over time (Waples, 1998).

To address the possibility that sample collection, extraction, sequencing, or another process introduced an erroneous year effect into the data, I identified SNPs that differed between fish sampled in 2017 and 2018 with an $F_{ST}$ above 0.01 using hierfstat, then filtered out those SNPs from the data using VCFtools version 0.1.14. The cut-off of 0.01 was chosen by only retaining those SNPs below an inflection point in the distribution of between-year $F_{ST}$ values. Following these steps, 13,640 SNPs (26.04% of 52,372 neutral SNPs total) were identified as having a large effect between years and were thus removed, leaving 38,732 SNPs. Analyses for population structure were then re-run with this smaller set of SNPs. $F_{ST}$ was calculated both between sites and between two reassigned clusters described by Adegenet (Table A.1.1). Data was also visualized by using a PCA as implemented in Adegenet version 2.1.1, and t-SNE as implemented in Rtsne version 0.15.

To test if our estimates of population structure were not driven by family groups (Waples, 1998), I used Colony version 2.0.6.4 to reconstruct pedigrees in our sample of 48 individuals, with consideration of possible full-siblings (Jones & Wang, 2010; Wang, 2004). The putatively neutral SNPs were converted to the Colony format using a script by D. deWaters (https://github.com/dandewaters/VCF-File-Converter). The Colony command-line input file was then generated to run the program with updated allele frequencies, dioecy, inbreeding possible, polygamy allowed, no clones, full sibship scaling, no sibship prior, unknown population allele frequencies, ten runs of medium length, full likelihood inference, and high precision. This Colony input file was generated using a script originally written by M. Ackerman (used in Ackerman et al., 2017), modified and posted with permission for the present study. Independent of Colony's maximum likelihood-based approach, I also used the method of moments as implemented in SNPRelate version 1.18.0 (Zheng et al., 2012) to estimate a kinship coefficient between individuals, also using the putatively neutral SNPs.

### 2.3.5 Outlier SNPs

Using the full list of SNPs filtered for genotype quality 30, minor allele frequency > 0.05 and two missing individuals allowed, but not filtered for Hardy-Weinberg Equilibrium or Linkage Disequilibrium, I tested for outlier SNPs using an unsupervised approach in pcadapt version 4.1.0 (Luu et al., 2017). The unsupervised approach was used because weak population

differentiation and the likely presence of admixed individuals in the data would either lower our sample size by filtering admixed individuals out, or lead to false-positive outlier loci by their inclusion when using a supervised approach with population structure included (Liu et al., 2017). While this may lead to issues of false positives from multiple tests (Foll & Gaggiotti, 2008), I addressed this issue by using a *q*-value of 0.05 and focusing our interpretation on transcripts that contain two or more outlier SNPs. Two PCs were chosen for this analysis by observing the scree plot visualizing K = 1–20 following Cattell's rule, where the point that a smooth decrease in eigenvalues levels off on a scree plot is the last important PC for explaining the data (Cattell, 1966).

By relating the transcript ID of a significant outlier SNP (*q*-value < 0.05) to that transcript's putative function and gene ID from the annotated reference transcriptome, a database of transcripts which diverged by sampling location or year was created for the Lake Winnipeg walleye in the present study. From this database, a list of transcripts relevant to either sampling location or year was used for gene set enrichment analysis using EnrichR (Chen et al., 2013; Kuleshov et al., 2016), thereby summarizing genes by gene ontology (GO) terms. In addition, transcripts were filtered to find those with two or more significant outlier SNPs that diverged by either sampling location or year, and these transcripts were few enough that enrichment analysis was not necessary. By only including genes with multiple outlier SNPs, I sought to reduce the presence of false positive signals in this outlier test.

## 2.4 Results

### 2.4.1 Population Structure

Our data suggested weak but significant population structure between the north and south basins of Lake Winnipeg. The Red River and Matheson Island locations slightly diverged ($F_{ST}$ = 0.0012), while the Dauphin River fish were the most genetically distinct group sampled ($F_{ST}$ = 0.0068 and 0.0043 compared to the Red River and Matheson Island, respectively) (Figs 2.2, 2.3, Table 2.1, and Fig A.1.2). Moreover, Structure and the DAPC returned similar results with respect to which fish were admixed, although membership probabilities differed (Figs 2.2, 2.3).

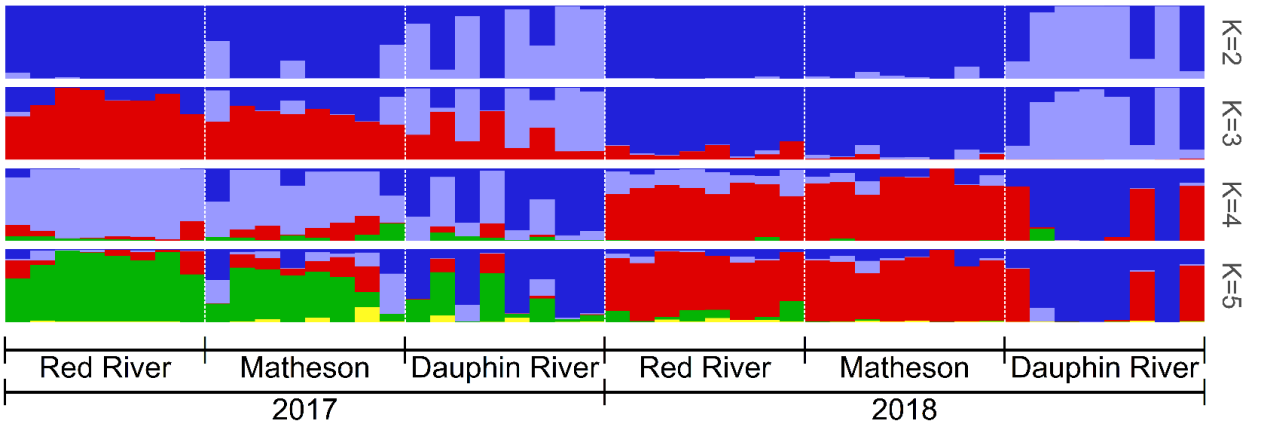


**Fig 2.2** Representative Structure runs from ten replicates testing K = 2–5, organized by collection site (Red River in the south basin, Matheson Island in the channel, and Dauphin River in the north basin) and year collected (2017 and 2018) for all walleye (*Sander vitreus*) used in the present study. Collection site locations are available in Fig 2.1. This analysis was performed with 52,372 Hardy-Weinberg Equilibrium filtered and linkage disequilibrium pruned, putatively neutral single nucleotide polymorphisms.

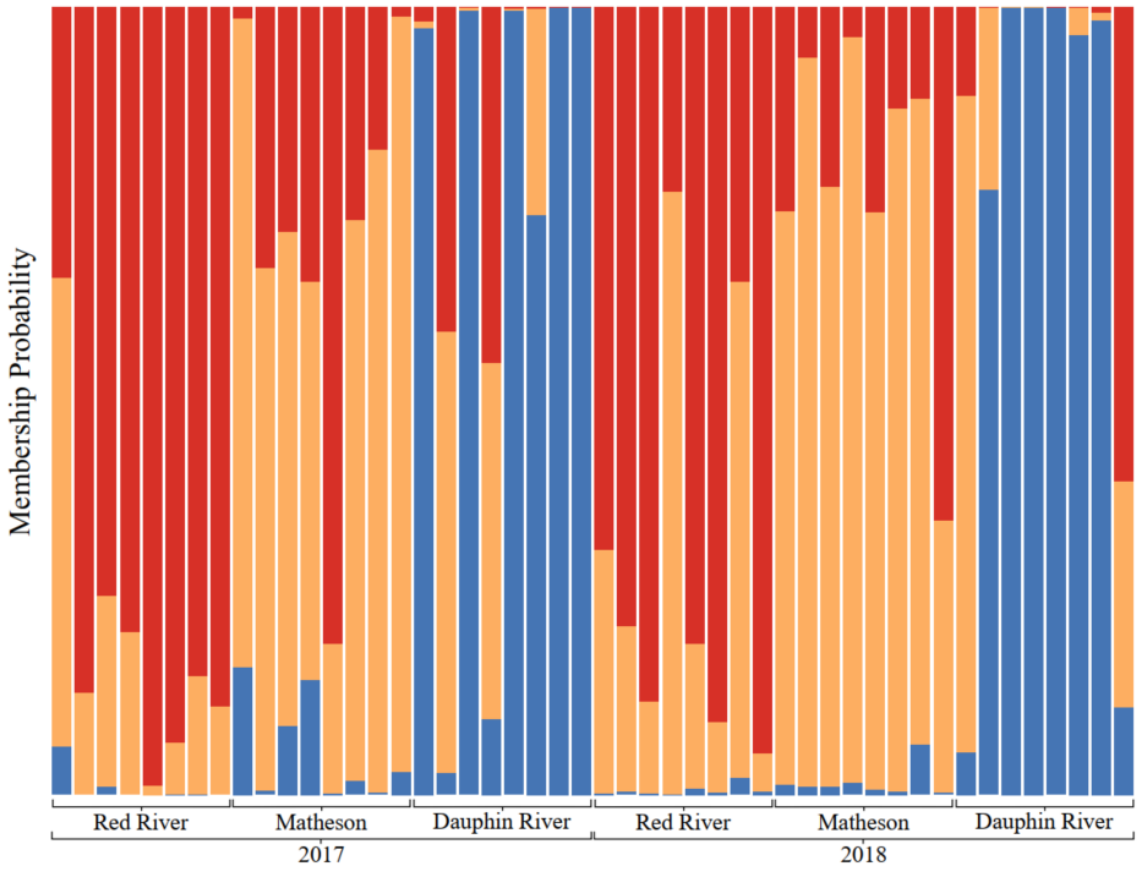


**Fig 2.3** Membership probability plot of discriminant analysis of principal components using prior collection site information (Red River in the south basin, Matheson Island in the channel, and Dauphin River in the north basin) on walleye (*Sander vitreus*) collected over 2017 and 2018, performed using Adegenet. Collection site locations are available in Fig 2.1. This analysis was performed with 52,372 Hardy-Weinberg Equilibrium filtered and linkage disequilibrium pruned, putatively neutral single nucleotide polymorphisms. Red colours represent membership probability in the Red River, pastel represents membership in Matheson, and blue represents membership in the Dauphin River.

**Table 2.1** Weir & Cockerham's pairwise $F_{ST}$ calculated with hierfstat between the Red River in the south basin, Matheson Island in the channel, and Dauphin River in the north basin for all 48 walleye (*Sander vitreus*) sampled in both 2017 and 2018. 95% confidence intervals are provided in parentheses. Collection site locations are available in Fig 2.1. This analysis was performed with 52,372 Hardy-Weinberg Equilibrium filtered and linkage disequilibrium pruned, putatively neutral single nucleotide polymorphisms.

| | **Red River (south basin)** | **Matheson Island (channel)** | **Dauphin River (north basin)** |
|---|---|---|---|
| **Red River** | - | 0.0012 (0.0009–0.0016) | 0.0068 (0.0064–0.0074) |
| **Matheson Island** | | - | 0.0043 (0.0039–0.0048) |
| **Dauphin River** | | | - |

Between K = 2–5, Structure consistently separated the Dauphin River fish from the Matheson Island and Red River fish, while the Red River and Matheson Island fish did not separate from each other by site, but instead separated between years (Fig 2.2).

The PCA and t-SNE used with the putatively neutral SNPs show similar patterns of Matheson and Red River fish separated, but more similar to each other than either with the Dauphin River fish (Fig A.1.2). When comparing the PCA and t-SNE plots between the neutral linkage disequilibrium-pruned SNPs and the broader collection of SNPs used for outlier analyses, genetic differentiation between the Red River and Matheson Island fish disappears when using all of the SNPs with the t-SNE, whereas separation between the two sites persists when only using Hardy-Weinberg Equilibrium filtered and LD-pruned SNPs.

### 2.4.2 Population Assignment

Using two clusters for reassignment from Adegenet, out of 48 fish, 36 clustered in one group (Cluster 1), and twelve in the other (Cluster 2; Table A.1.1). Cluster 1 was characterized by a combined Red River and Matheson Island group of fish with few Dauphin River fish (six were collected from the Dauphin River, 14 from Matheson Island, and 16 from the Red River), while Cluster 2 was characterized by Dauphin River fish and a small number of Matheson Island fish (ten fish from the Dauphin River and two from Matheson Island). Weir and Cockerham's pairwise $F_{ST}$ between these two reassigned clusters was 0.0095 with a 95% confidence interval between 0.0090–0.010.

Using three clusters for reassignments from Adegenet, out of 48 fish, 19 were in one group (Cluster 1), ten fish were in another group (Cluster 2), and 19 fish in a final group (Cluster 3). Clusters 1 and 3 were characterized by a year effect, where every individual in Cluster 1 was captured in 2018 and every individual in Cluster 3 was captured in 2017. Both Clusters 1 and 3 had 16 out of 19 fish coming from the Red River or Matheson Island sites. Meanwhile, all ten fish in Cluster 2 were from the Dauphin River with five fish each collected in 2017 and 2018 (Table A.1.1).

### 2.4.3 Temporal Stability and Kinship

When partitioning individuals by sampling location and year collected, all confidence intervals for between-site pairwise $F_{ST}$ estimates overlapped over both sampling years, indicating consistent patterns of between-site divergence in 2017 and 2018. However, values between the Dauphin River and Matheson Island varied the most, with an estimate of 0.0044 (0.0035–0.0052) in 2017, and 0.0060 (0.0051–0.0070) in 2018 (Table 2.2).

Using the 38,732 SNPs filtered for loci, which showed $F_{ST}$ between years of > 0.01, $F_{ST}$ between the two reassigned clusters found using Adegenet (Table A.1.1) was 0.010 (0.0094–0.011). With these same year effect-filtered SNPs, pairwise $F_{ST}$ between sites did not significantly differ from values found using the neutral SNPs either overall or in a subset by year (Table A.1.2). The PCA and t-SNE on the SNPs filtered for a year effect showed patterns of spatial differentiation consistent with other analyses, with the Dauphin River fish being more separate from the Red River and Matheson Island group of fish (Fig A.1.3).

I found no evidence of kinship using either Colony or the method of moments. Over ten replicate runs in Colony, individuals belonged to separate families with inclusive and exclusive probabilities of 1.0000 each. Using the method of moments implemented in SNPRelate (Zheng et al., 2012), the highest kinship coefficient between two individuals was 0.096 (mean 0.053 ± 0.019 s.d.), where a kinship coefficient of approximately 0.5 would indicate full-siblings.

### 2.4.4 Outlier SNPs

There was site-specific differentiation across Principal Component 1 (PC1) in the pcadapt analysis (Fig 2.4). In total, 1,177 SNPs were outliers at $q < 0.05$, with 386 SNPs contributing to PC1 where fish separated by site, and 791 SNPs contributing to PC2 where fish separated by year (Fig 2.4). For the 386 SNPs associated with PC1 (Fig 2.4), 120 uniquely annotated transcripts were available for enrichment analysis using EnrichR. These transcripts corresponded to GO terms such as purine ribonucleoside triphosphate binding, ATP binding, and adenyl ribonucleotide binding, all significant at Benjamini-Hochberg adjusted *p*-values < 0.05 (Table A.1.3). By filtering for uniquely annotated transcripts with ≥ 2 outlier SNPs associated with PC1, 19 transcripts were identified (Table 2.2) that varied by sampling location. Six of these genes

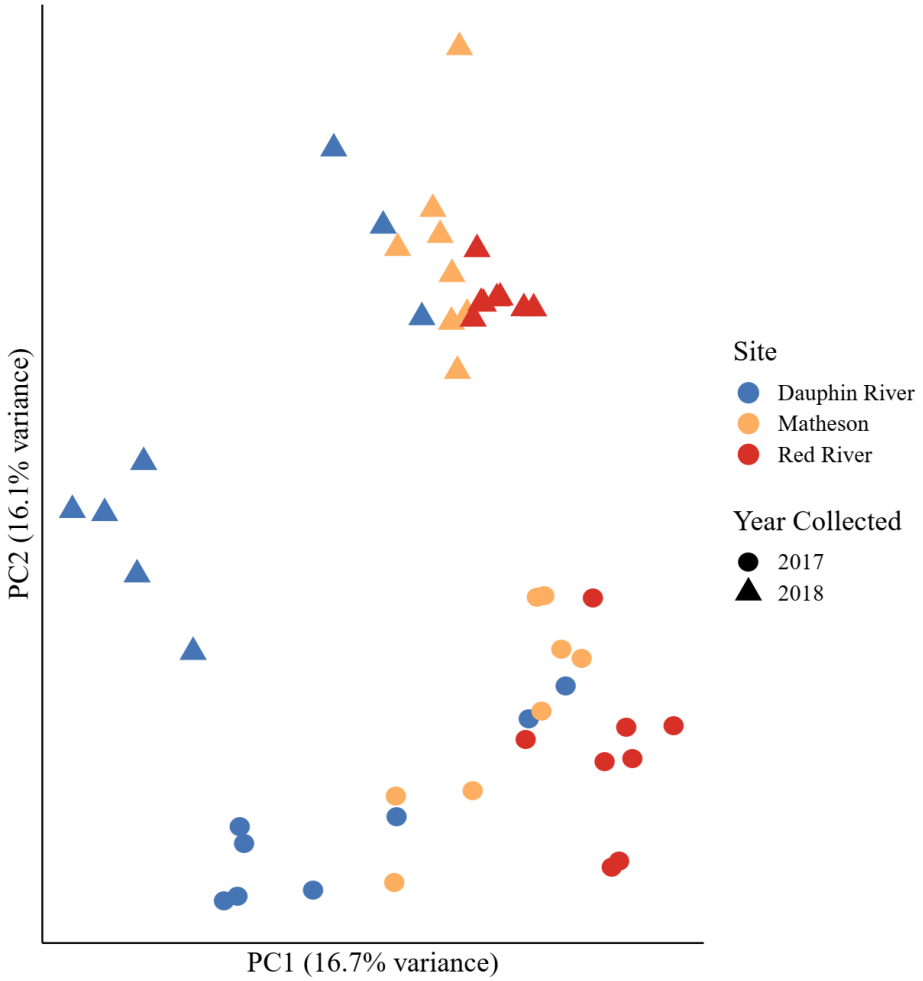


**Fig 2.4** Principal Components Analysis implemented in pcadapt with color showing site collected (red for Red River in the south basin, yellow Matheson Island in the channel, and blue Dauphin River in the north basin), circles showing walleye (*Sander vitreus*) collected in 2017, and triangles showing walleye collected in 2018. Collection site locations are available in Fig 2.1. This analysis was performed using a set of 222,634 single nucleotide polymorphisms (SNPs) filtered for quality, minor allele frequency > 0.05, and a maximum of two out of 48 missing individuals, but not filtered for Hardy-Weinberg Equilibrium or pruned for linkage disequilibrium. These SNPs are, thus, more likely to represent patterns of adaptive variation in the system, and outlier analyses were performed using this set of SNPs. Principal Component 1 (PC1) represents a latitudinal gradient, while Principal Component 2 (PC2) represents a genetic divergence between sampling years. Outlier SNPs that contribute to each of these axes were selected for functional analyses (Tables 2.2, 2.3), with 386 SNPs contributing to PC1 and 791 SNPs contributing to PC2, significant at Benjamini-Hochberg adjusted *p*-values < 0.05.

**Table 2.2** Genes that vary along a latitudinal gradient in Lake Winnipeg walleye (*Sander vitreus*) with ≥ 2 outlier single nucleotide polymorphisms (SNPs) from pcadapt, each significant at a Benjamini-Hochberg adjusted p-value < 0.05 (PC1 in Fig 2.4). SwissProt gene names and corresponding proteins are provided, and general cellular location or function of these genes are described in Summary Function. This analysis was performed using a set of 222,634 single nucleotide polymorphisms (SNPs) not filtered for Hardy-Weinberg Equilibrium or pruned for linkage disequilibrium, unlike the putatively neutral set of SNPs used for population structure analyses.

| SwissProt Gene Name | Protein | Summary Function |
|---|---|---|
| **ABCA12** | ATP-binding cassette sub-family A member 12 | Cell Membrane |
| **ANK3** | Ankyrin-3 | Cell Membrane |
| **atp1a3** | Sodium/potassium-transporting ATPase subunit alpha-3 | Cell Membrane |
| **CLDN10** | Claudin-10 | Cell Membrane |
| **Prf1** | Perforin-1 | Cell Membrane |
| **SLC9A6** | Sodium/hydrogen exchanger 6 | Cell Membrane |
| **CRYBG1** | Beta/gamma crystallin domain-containing protein 1 | Cytoskeleton |
| **IFI44** | Interferon-induced protein 44 | Cytoskeleton |
| **MYH9** | Myosin-9 | Cytoskeleton |
| **TUBB4B** | Tubulin beta-4B chain | Cytoskeleton |
| **DNASE1L1** | Deoxyribonuclease-1-like 1 | DNase |
| **EIF4G1** | Eukaryotic translation initiation factor 4 gamma 1 | Expression regulation |
| **srebf2** | Sterol regulatory element-binding protein 2 | Expression regulation |
| **Znf879** | Zinc finger protein 879 | Expression regulation |
| **CLPX** | ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial | Protease |
| **CXCR3** | C-X-C chemokine receptor type 3 | Signaling |
| **MATK** | Megakaryocyte-associated tyrosine-protein kinase | Signaling |
| **Ralgds** | Ral guanine nucleotide dissociation stimulator | Signaling |
| **Pol** | LINE-1 retrotransposable element ORF2 protein | Transposable Element |

**Table 2.3** Genes that vary between 2017 and 2018 in Lake Winnipeg walleye (*Sander vitreus*) with transcripts containing ≥2 outlier single nucleotide polymorphisms (SNPs) from pcadapt, each SNP significant at a Benjamini-Hochberg adjusted p-value < 0.05 (PC2 in Fig 2.4). SwissProt gene names and corresponding proteins are provided, and general cellular location or function of these genes are described in Summary Function. This analysis was performed using a set of 222,634 single nucleotide polymorphisms (SNPs) not filtered for Hardy-Weinberg Equilibrium or pruned for linkage disequilibrium, unlike the putatively neutral set of SNPs used for population structure analyses.

| SwissProt Gene Name | Protein | Summary Function |
|---|---|---|
| **Gp1bb** | Platelet glycoprotein Ib beta chain | Cell Adhesion |
| **MFAP4** | Microfibril-associated glycoprotein 4 | Cell Adhesion |
| **BTG3** | Protein BTG3 | Cell Division |
| **Ppp6c** | Serine/threonine-protein phosphatase 6 catalytic subunit | Cell Division |
| **Slc12a3** | Solute carrier family 12 member 3 | Cell Membrane |
| **MMD** | Monocyte to macrophage differentiation factor | Expression regulation |
| **Srsf10** | Serine/arginine-rich splicing factor 10 | Expression regulation |
| **Ube2a** | Ubiquitin-conjugating enzyme E2 A | Expression regulation |
| **Znf18** | Zinc finger protein 18 | Expression regulation |
| **St6galnac2** | Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2 | Protein Modification |
| **MYO9A** | Unconventional myosin-IXa | Signaling/Cytoskeleton |
| **Pol** | LINE-1 retrotransposable element ORF2 protein | Transposable Element |
| **RTase** | Probable RNA-directed DNA polymerase from transposon BS | Transposable Element |
| **TC1A** | Transposable element Tc1 transposase | Transposable Element |
| **TN6** | Putative transposase in *Dicentrarchus labrax* (European seabass) | Transposable Element |
| **TY3B-G** | Transposon Ty3-G Gag-Pol polyprotein | Transposable Element |
| **YTX2** | Transposon TX1 uncharacterized protein (Fragment) | Transposable Element |

were associated with ion channels and cell membrane transport, including claudin-10, ankyrin-3, sodium/hydrogen exchanger 6, sodium/potassium-transporting ATPase subunit alpha-3, perforin-1, and ATP-binding cassette sub-family A member 12. Additionally, four genes that varied spatially were associated with the cytoskeleton, such as myosin-9, beta/gamma crystallin domain-containing protein 1, tubulin beta-4B chain, and interferon-induced protein 44.

Using the 791 SNPs associated with Principal Component 2 (PC2), which varied by year (Fig 2.4), 130 uniquely annotated transcripts were available for enrichment analysis; however, no GO terms were significant at an adjusted $p$-value < 0.05. For transcripts with ≥ 2 PC2 outlier SNPs, 17 uniquely annotated genes were identified of which six were either transposons, transposable elements, or fragments of transposons (Table 2.3). Two genes that code for the proteins serine/threonine-protein phosphatase 6 catalytic subunit and protein BTG3, which regulate cell division in the G1 to S phase transition were also identified (Table 2.3).

## 2.5 Discussion

I observed weak population structure characterized by groups collected at the Red River and Matheson Island sampling locations, representing south basin and channel fish, contrasted with a group collected at the Dauphin River, representing north basin fish. As such, the north and south basin walleye in Lake Winnipeg may be separate groups with an $F_{ST}$ of 0.0095, but with gene flow between them primarily at the channel connecting the two basins. Consistent with results in the present study, a study using microsatellites found a similar weak, but significant, differentiation between Lake Winnipeg walleye from some sites in the north and south basins (e.g., $F_{ST}$ = 0.022 between the Grand Rapids in the north and Red River in the south; Backhouse-James & Docker, 2012), suggesting genomic divergence between walleye from the two basins. However, microsatellite data did not resolve overall population structure between basins to a fine-scale as in the present study. That is, microsatellites were limited to describing differentiation between particular sites, while population reassignment with SNPs from RNA data enabled descriptions of structure in the contest of straying and possible metapopulation structure characterized by demographic connectivity. While facilitated gene flow has not been reported for Lake Winnipeg walleye (Manitoba Government, 2018; Manitoba Sustainable Development, 2018), historical stocking programs may have contributed to weak population

structure within Lake Winnipeg by introducing walleye from nearby Lake Manitoba (Backhouse-James & Docker, 2012). This unknown amount of gene flow from other systems, including up to 26.5 million fish annually between 1970 and 1983 (see Chapter 4), may have masked signatures of spatial population differentiation in Lake Winnipeg. While an estimated $F_{ST}$ of 0.0095 between separate groups is low, there is reason to believe that widespread purifying selection in RNA data may decrease $F_{ST}$ estimates in comparison to estimates generated with microsatellites, consistent with observations of negative selection in protein-coding sequences of many plant species (Gossmann et al., 2010).

### 2.5.1 Temporal Differentiation

Differentiation between years was strongest in the south basin, where the Red River and Matheson Island fish separated by year to a greater extent than the Dauphin River walleye. Three hypotheses may explain these patterns of stronger temporal differentiation in the south basin. First, a cohort effect may underlie this pattern, where different year classes were more strongly represented in the lake during a given sampling year. A cohort effect could be the result of greater fishing pressure in the south basin than the north basin, as indicated by smaller allowed net mesh sizes in the south basin (Manitoba Sustainable Development, 2019). Fishing pressure can change population dynamics and age structure in exploited species (Anderson et al., 2008; Murphy & Crabtree, 2001), therefore large fisheries operating since at least 1890 may have affected age structure in Lake Winnipeg fish (Department of Fisheries, 1891). Cohort effects may alternatively be influenced by environmental conditions including predation intensity, water temperature, and time to hatch as observed in Lake Erie, Oneida Lake, and Lake Huron walleye (Busch et al., 1975; Fielder et al., 2007; Forney, 1976). A second hypothesis is that some Lake Winnipeg walleye may engage in unobserved skipped spawning or alternate year spawning, which have been unexpectedly found in several species and may be present in walleye (Carlander et al., 1960; Henderson et al., 1996; Moles et al., 2008; Rideout & Tomkiewicz, 2011). Third, the observed year effect may be an artifact of error introduced during sampling, extraction, sequencing, or bioinformatics. While some error contributing to between-year differentiation is impossible to rule out, the possibility that a particular analysis or filtering method introduced the year effect is reasonably small given that distinct tests showed consistent

year effects. Moreover, the data reveal a consistency in spatial patterns with and without the year effect, demonstrating that at least the spatial population differentiation in Lake Winnipeg walleye is likely real.

While pairwise site $F_{ST}$ values were temporally consistent (i.e., no significant differences between years), the greatest pairwise $F_{ST}$ confidence interval difference between years was between walleye collected at Matheson Island and the Dauphin River, where confidence intervals for $F_{ST}$ estimates overlapped by only 0.0001 between 2017 and 2018. Following Amrhein & Greenland (2019), I interpret here the possibility that the entire range of these confidence intervals reflect meaningful patterns in the data, and that $F_{ST}$ between the Dauphin River and Matheson Island was different between 2017 and 2018. Because Matheson Island represents a narrow channel connecting the north and south basins of Lake Winnipeg, fish which would normally spawn in the Dauphin River may have used the channel more often in 2017, thus, lowering $F_{ST}$ when performing a site-wise comparison. This difference in habitat use may have arisen from an undetermined environmental variable, such as time of ice melt, which in the north basin was ten days later in 2018 than in 2017 (D. Watkinson, unpublished data found using https://zoom.earth/). Nevertheless, confidence intervals for $F_{ST}$ overlapped between years between the Matheson Island and Dauphin River sites, therefore patterns of environment-dependent spawning site fidelity must be corroborated with other data, such as by telemetry.

Notably, gene flow appears to be one way from the southern Red River, northward. Going by capture location, no fish caught in the Red River showed a genetic background consistent with the Dauphin River fish, while with the Adegenet-reassigned clusters, no fish assigned to the mostly Dauphin River group was found in the Red River. On the other hand, fish which showed a genetic background consistent with the Red River group were found in the Dauphin River, both based on capture location and population reassignment.

### 2.5.2 Biological Significance

Several studies report morphological and life history differences between basins in Lake Winnipeg walleye consistent with the two delineated groups found in this study. Furthermore, environmental data show a north-south basin distinction with temperature, turbidity, mean depth,

suspended solid, sulphate, sodium, chloride, and nutrient differences between the two basins (Brunskill et al., 1980; Environment Canada, 2011a). Walleye in the south basin show a bimodal growth pattern, where fisheries-induced selection may have contributed to the observation of dwarf walleye (Johnston et al., 2012; Moles et al., 2010; Sheppard et al., 2018). Harvest-induced genetic changes have been linked to size reductions in other walleye within two generations (Bowles et al., 2019). If walleye were panmictic throughout Lake Winnipeg, I might expect the dwarf morphotype to occur with similar frequency in the north basin. However, out of 616 total walleye caught in 2010 and 2011 (178 in the north basin, 438 in the south basin), only two out of 32 dwarf fish were caught in the north basin (Sheppard et al., 2018). Diet has also been shown to differ between north and south basin walleye, possibly because of prey or turbidity differences between the two basins (i.e., higher turbidity in the south basin) (Brunskill et al., 1980; Sheppard et al., 2015). Between 1979 and 2003, population characteristics such as age and length at 50% maturity were higher, while growth rate was slower in the north basin walleye, suggesting some level of isolation among walleye between basins (Johnston et al., 2012). These population characteristics may no longer be higher in the north basin following the collapse of the rainbow smelt (*Osmerus mordax),* after which walleye body condition has decreased since 2010 (Manitoba Government, 2018). Scale morphometry further suggests differences among spawning aggregations of walleye, especially between the north and south basins (Kritzer & Sale, 2004; Watkinson & Gillis, 2005). Taken together, the results of our study and those of previous studies suggest weak population structure among Lake Winnipeg walleye, with differentiation between walleye in the north and south basins. This pattern of weak population structure, high connectivity, but biologically significant differentiation is common in marine fishes such as the Atlantic cod (*Gadus morhua*) or Atlantic salmon (*Salmo salar*) (Aykanat et al., 2015; Knutsen et al., 2011), and of other walleye such as those observed in Lake Erie (Chen et al., 2019; Stepien et al., 2018).

The results of the present study suggest that the genetic differences between Lake Winnipeg walleye populations may have functional consequences. Out of 19 transcripts that had multiple SNPs that varied by sampling location, eight were related to membrane function, particularly ion channel activity. One of these proteins, Claudin-10 mRNA expression levels have been related ammonia exposure (Connon et al., 2011), rearing density (Sveen et al., 2016), and salinity (Bossus et al., 2015; Kolosov et al., 2013; Marshall et al., 2018) in fishes. Spatial

variation in cell membrane proteins is consistent with environmental differences between basins in chemicals such as sodium, chloride, and phosphorous (Environment Canada, 2011), although the biological impacts of these spatial chemical differences is unknown. Four cytoskeletal proteins were represented in the outlier SNPs that vary by sampling location as well. Cytoskeletal function is connected to cell growth in plants (Hussey et al., 2006; Wasteneys & Galway, 2003), yeast (Li et al., 1995; Pruyne & Bretscher, 2000), mouse cells (Kim et al., 2006; Kim & Coulombe, 2010), and zebrafish (Johnston et al., 2011). Spatial variation in genes related to cell growth may thus be consistent with growth rate differences observed among walleye in Lake Winnipeg, where north basin fish had higher growth rates in 2010 and 2011 (Sheppard et al., 2018). While plastic phenotypes are likely important for walleye responses to environmental differences (Kelly, 2019), the outlier loci found in the present study provide some evidence for adaptive divergence despite weak population differentiation. While the biological significance of the outlier SNPs has not been confirmed, the SNPs themselves or the genes they reside in may be the focus of future research.

### 2.5.3 Limitations

Despite its advantages, there are some limitations to using mRNA sequencing in the context of population genetics. The depth of sequencing required for differential gene expression and differential exon usage leads to greater costs associated with mRNA sequencing studies relative to reduced representation methods such as RAD-seq (Davey & Blaxter, 2010). This often translates to a lower sample size, as is the case in our study. Reduced sample sizes can bias aspects of population genetic analyses, including identifying population structure (Waples & Gaggiotti, 2006) and outlier SNPs (Luu et al., 2017), although sample sizes as low as 12 yield accurate mean population allele frequencies with mRNA sequencing data (Schunter et al., 2014). Second, mutations in mRNA are widely under selection (Chamary & Hurst, 2005), therefore, caution must be exercised when interpreting SNPs from mRNA in genetic tests assuming neutrality. Third, linkage disequilibrium is useful for analyses of selective sweeps and demographic history (Catchen et al., 2017; Garrigan & Hammer, 2006; Hoffmann & Willi, 2008), among other approaches, but mRNA data may not be appropriate for these analyses because the extent of linkage between transcripts and marker density is difficult to characterize

for species without published chromosome-level genome assemblies. Finally, one key element of how transcriptomics was used in the present study is that it measured expressed mRNA in gill tissue. Messenger RNA expression provides useful information for transcript quantification-based analyses, but likely biases SNP discovery toward more highly expressed transcripts in the tissue collected. It is unknown how this expression-specific bias may influence population genomics. Nevertheless, mRNA sequencing has proven useful for recapitulating population structure discovered with traditional genetic methods (Jeffries et al., 2019) and describing previously uncharacterized population structure (Ellison et al., 2011; Yan et al., 2017).

### 2.5.4 Conservation Applications

I used population genetics and outlier detection to characterize weak, but biologically significant population structure, possible one-way gene flow, and genetic variation possibly underlying biological differences among Lake Winnipeg walleye. These results are consistent with observations of behavioural differences leading to fine-scale divergence in the walleye of other systems (Stepien et al., 2009). The low levels of population differentiation and possible gene flow from the south basin northward, indicate that this system may be effectively managed as a demographically connected metapopulation with two contributing populations (Kritzer & Sale, 2004), consistent with conclusions from scale morphology presented in (Watkinson & Gillis (2005) and with observations of subtle stock structure in Lake Erie walleye (Chen et al., 2019; Stepien et al., 2018).

The results from this study provide valuable information for walleye management, especially because the status of Lake Winnipeg walleye is becoming a concern and conservation action may be necessary to sustain the fishery. Signs of a declining fishery include a decrease in biomass and body condition between 2010 and 2015 (Manitoba Government, 2018), possible unnatural selection against larger, economically desirable fish (Allendorf & Hard, 2009; Bowles et al., 2019; Moles et al., 2010), models showing walleye harvests have been above maximum sustainable yields since the early 2000s, and a trend in harvest decline since 2010 (Manitoba Sustainable Development, 2018). The data gathered here, particularly the spatial variation in genes that may drive functional differences among Lake Winnipeg walleye, is useful for generating hypotheses that test and explain organismal responses to environmental stressors,

thereby providing additional information for resource managers. For instance, life-history trait differences can inform conservation in threatened fishes by identifying resilient populations in a system (Hamidan & Britton, 2015). Therefore, possible functional variation identified in this study may underlie heritable genetic differences among Lake Winnipeg walleye that change important traits such as tolerance to environmental conditions and growth rate differences. In addition, with weak population differentiation, the possible functional variation discovered with RNA sequencing may be used in the delineation of conservation units, or population units with adaptive differentiation also considered (Funk et al., 2012). This information may be useful for integrating demographic connectivity and functional differences among walleye into a cohesive management framework.

I have shown how RNA sequencing data can be used for a population genomic scan in a non-model fish, even in a system where little molecular information is available. Filtering for Hardy-Weinberg equilibrium and linkage disequilibrium allows investigators to draw neutral markers from mRNA sequence data, making it useful for classical population genetic approaches. By contrast, the wide selective effects present in species' transcriptomes allow for hypothesis-generating outlier tests that may reveal variation underlying phenotypic differences among populations. Non-lethal sampling makes RNA sequencing useful for species with low population sizes and for follow-up studies, such as the potential to link transcriptomic patterns or genetic data with tagged individuals in the wild using telemetry (Jeffries et al., 2014; Miller et al., 2011; Moore et al., 2017). Despite its high cost relative to other genotyping methods, RNA sequencing data can yield physiological in addition to genetic information discussed here. I therefore argue that it is useful for addressing diverse molecular questions in the conservation of freshwater species.

### 2.5.5 Data Accessibility Statement

Raw sequence reads are available through the National Center for Biotechnology Information Sequence Read Archive (accession #PRJNA596986, https://www.ncbi.nlm.nih.gov/sra/PRJNA596986). Code used for analyses and bioinformatics in this manuscript is available at https://github.com/BioMatt/Walleye_RNAseq.

## 2.7 Acknowledgements

I thank C. Charles, C. Kovachik, D. Leroux, N. Turner, M. Gaudry, S. Glowa, and E. Barker for helping to sample these fish in the field. C. Charles also assisted with a map of Lake Winnipeg. Dr. C. Garroway provided valuable suggestions in data analysis, and Dr. G. Anderson engaged in useful conversations about Lake Winnipeg ecology. I thank the Lake Winnipeg Research Consortium Inc. for supporting the open access publication of this manuscript. Many analyses were enabled by our chance to use computing resources provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). Sequencing was performed at the McGill University and Génome Québec Innovation Centre sequencing facility. This work was supported by a Fisheries and Oceans Canada Ocean and Freshwater Science Contribution Program Partnership Fund grant awarded to JRT, KMJ and Darren Gillis, and Natural Sciences and Engineering Research Council of Canada Discovery Grants awarded to KMJ (#05479) and JRT (#06052). Work by JRT is also supported by the Canada Research Chairs program (#223744) and the Faculty of Science, University of Manitoba (#319254).

# Chapter 3. Morphology and Blood Metabolites Reflect Recent Spatial and Temporal Differences Among Lake Winnipeg Walleye, *Sander vitreus*[2]

## 3.1 Abstract

The invasive rainbow smelt (*Osmerus mordax*) was an abundant food source for Lake Winnipeg walleye (*Sander vitreus*), especially in the north basin of the lake, until the smelt's collapse in approximately 2013. I quantified changing length-at-age (≈ growth rates) and relative mass (≈ body condition) in Lake Winnipeg walleye caught for a gillnet index data set. Here, walleye showed smaller length-at-age, particularly young fish in the north basin, over time. This approach to assessing growth suggests a constraint in the north basin fish, possibly a nutritional limitation between 2017 and 2018, that was not present in the south. I then analyzed a separate group of walleye (≥452 mm in fork length) sampled in 2017 as part of a large-scale tracking study, which had a similar slope in length-mass relationship to large walleye caught in that year for the gillnet index data. A panel of metabolites in whole blood samples associated with amino acid metabolism and protein turnover was compared. These metabolites revealed elevated essential amino acids (i.e., amino acids that cannot be made by the walleye) in fish caught in the Dauphin River, and suggest that protein degradation may be elevated in north basin walleye. Therefore, based on both growth estimates and metabolites associated with protein balance, I suggest there were spatially distinct separations affecting Lake Winnipeg walleye with decreased nutritional status of walleye in the north basin of Lake Winnipeg being of particular concern.

## 3.2 Introduction

Walleye (*Sander vitreus*) are the largest component of the Lake Winnipeg fishery in Manitoba, the second-largest freshwater fishery in Canada (Fisheries and Oceans Canada, 2018).

---

[2] A modified version of this chapter has previously been published. Permission not required for reproduction by the author. © 2020 International Association for Great Lakes Research. Matt J Thorstensen conducted the analyses. Matt J Thorstensen, Lilian M Wiens, Jennifer D Jeffrey, Ken M Jeffries, and Jason R Treberg conceived of the study and wrote the original manuscript. Geoffrey M Klein collected and provided guidance on the gillnet index data. All co-authors participated in revisions of the final manuscript drafts prior to submission. Found in: Thorstensen, M. J., Wiens, L. M., Jeffrey, J. D., Klein, G. M., Jeffries, K. M., & Treberg, J. R. (2020). Morphology and blood metabolites reflect recent spatial and temporal differences among Lake Winnipeg walleye, *Sander vitreus*. Journal of Great Lakes Research, xxxx. https://doi.org/10.1016/j.jglr.2020.06.015

Sustainable management of these walleye is, therefore, of enormous importance to commercial fishing, recreational angling, and the Lake Winnipeg ecosystem. However, annual commercial yield from the Lake Winnipeg walleye fishery has been above maximum sustainable yield since 2002, while commercial harvests have also declined between 2014 and 2018 (Manitoba Sustainable Development, 2018). Relatively recent observations of dwarf walleye, primarily in the south basin, suggest a selective pressure against large individuals—a selective force possibly induced by fisheries (Moles et al., 2010; Sheppard et al., 2018). Taken together, these separate pieces of evidence indicate that the Lake Winnipeg walleye fishery is faced with several issues that may affect its sustainability and suggest that this fishery may need conservation attention.

To address conservation issues the Lake Winnipeg walleye may face, resource managers require many pieces of information to forecast how actions affect the probability of distinct alternative futures (Dudgeon et al., 2006; Gattuso et al., 2015). Simultaneous exposure to multiple stressors, such as eutrophication and invasive species, may lead to cumulative detrimental impacts on a fishery (Schindler, 2001). One key alteration to the Lake Winnipeg ecosystem was the introduction and subsequent crash of non-native rainbow smelt (*Osmerus mordax*). Rainbow smelt were first been observed in Lake Winnipeg in late 1990 (Franzin et al., 1994), and were later found in the stomachs of 82.9% of walleye caught in the north basin, but only 9.3% of walleye caught in the south basin (in 2010 and 2011, see Sheppard et al., 2015). After approximately 2013 however, rainbow smelt have almost disappeared from all of Lake Winnipeg, and their disappearance coincides with walleye body condition (a measure of mass relative to length, or ‘fatness’ of the fish) declines across the lake (Enders et al., 2021; Manitoba Government, 2018). As walleye body condition decreases over time, implicating a reduction in available nutritional resources such as the rainbow smelt, while fishing effort remains constant or rises, the recent declines in Lake Winnipeg walleye abundance may be exacerbated (Manitoba Sustainable Development, 2018). Therefore, understanding how changing food availability may be affecting Lake Winnipeg walleye is a fundamental issue in the system and a useful piece of information for resource management.

Beyond gross morphological measurements like body condition, there are few available tools for evaluating the nutritional status of wild fishes outside of destructive sampling and proximate composition analysis. The current study was undertaken, taking advantage of the

patterns of declining fish condition, to determine if there may be basin-level differences in the nutritional status of walleye in Lake Winnipeg by examining body condition and growth rates over time. Basin-level differences seemed most likely because the lake is characterized by two large basins separated by a narrow channel (Fig 3.1). Moreover, the two basins differ in characteristics such as higher temperature, phosphorous, nitrogen, and prey fish such as emerald shiner (*Notropis atherinoides*) and cisco (*Coregonus artedi*) abundance in the south basin, while sodium and chloride are higher in the north basin (Binding et al., 2018; Environment Canada, 2011; Lumb et al., 2012; Sheppard et al., 2015). Our second goal was to identify potential biomarkers from non-lethal sampling to test for differential nutritional status or energetic demands in walleye from the different regions of the lake. With refinement and validation, such biomarkers could provide managers with tools that may be able to rapidly inform on physiologically relevant thresholds of nutritional constraints that would be useful in risk assessments (Connon et al., 2018).

### 3.2.1 Nutritional Biomarker Strategy

I focused our study on protein metabolism because bulk protein growth is, with certain caveats, analogous to individual fish growth (see Carter & Houlihan (2001) for review). Protein growth is a function of the balance between the rate of synthesis of new proteins from free amino acids and the rate of protein degradation, which releases free amino acids (Fig 3.2). Therefore, if an animal is synthesizing more protein than the rate of protein breakdown, there is net growth. Protein synthesis and degradation are both tightly regulated physiological processes, and as such the balance between both processes are intimately linked to growth and energy balance. I sought a biomarker strategy that could differentiate between amino acid breakdown for oxidation as an energy source (as opposed to re-using amino acids for protein synthesis), and the extent of endogenous protein breakdown in walleye across Lake Winnipeg (Fig 3.1).

Generally, if fish are not taking in sufficient food, stored lipids and carbohydrates are preferentially used to meet energetic demands, with whole-body protein being increasingly mobilized as a fuel after lipid and carbohydrate reserves are depleted (Black & Love, 1986; Collins & Anderson, 1995). Essential, or non-dispensable, amino acids may be useful for insight

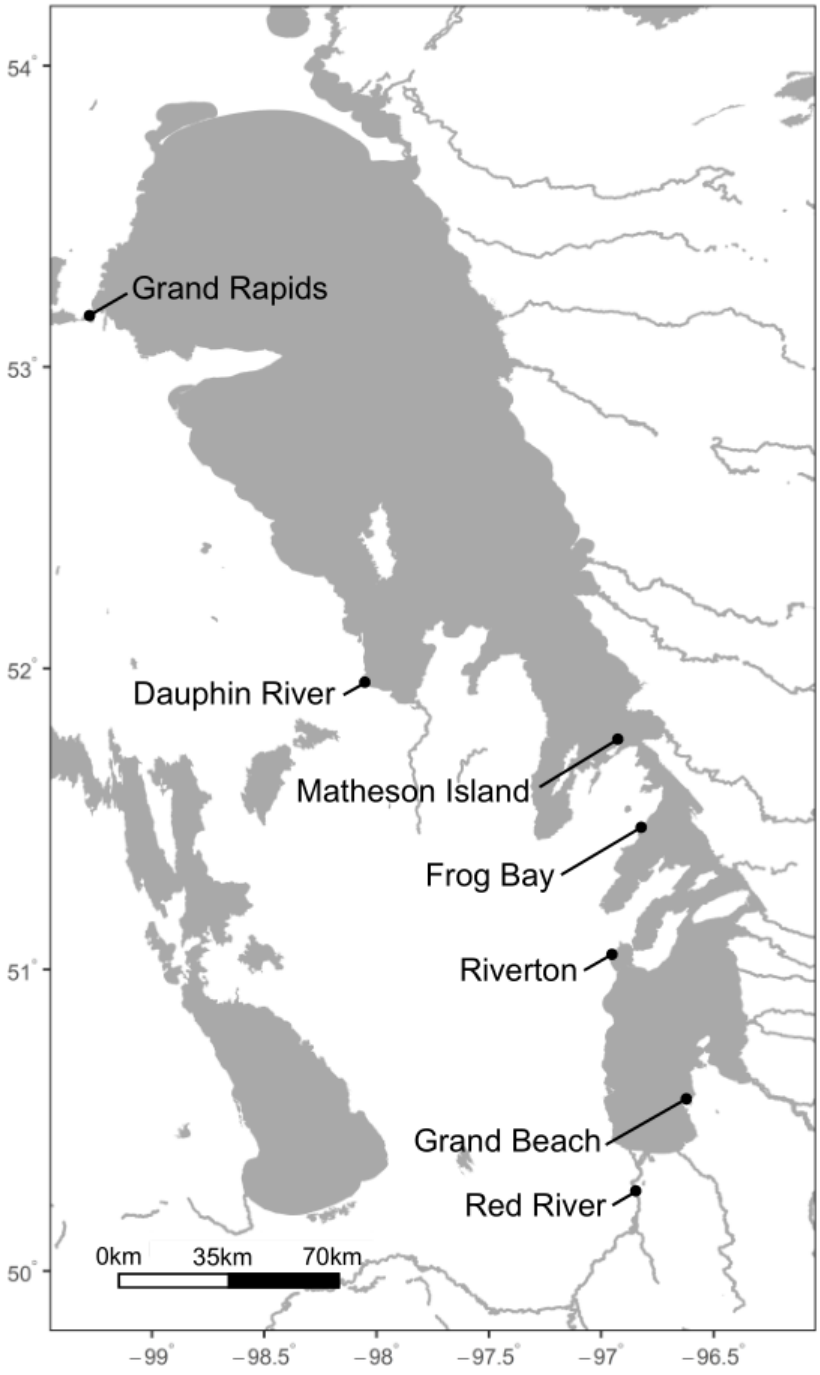


**Fig 3.1** Map of Lake Winnipeg and the sites included in the present study. The Grand Rapids and Dauphin River represent the north basin, Matheson Island and Frog Bay the channel, and Riverton and Grand Beach the south basin, in the gillnet index data collected between 2009 and 2018 by the Government of Manitoba. For the blood metabolite data consisting of large walleye (*Sander vitreus*) caught by electrofishing in 2017, the Dauphin River represents the north basin, Matheson Island the channel, and the Red River the south basin.

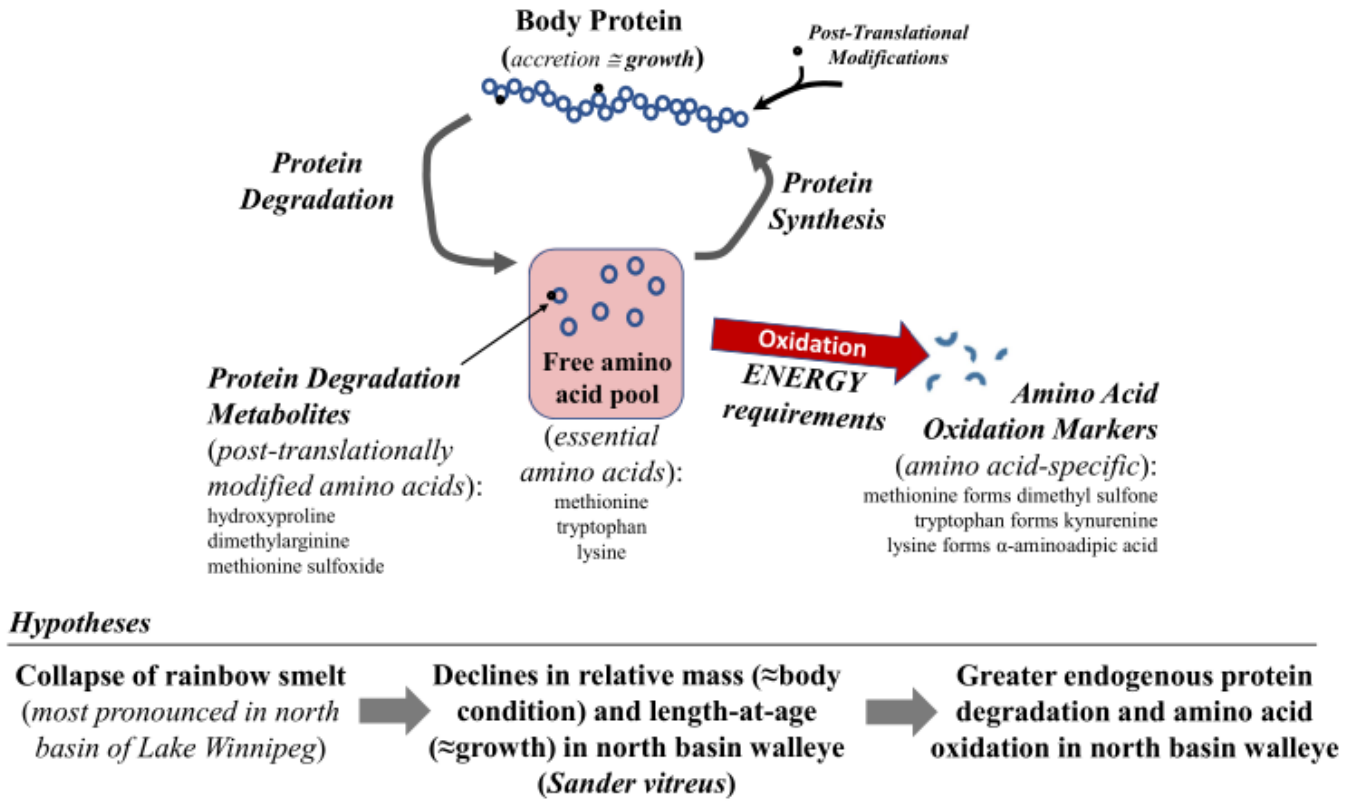


**Fig 3.2** Conceptual diagram linking protein degradation and markers of amino acid breakdown to walleye (*Sander vitreus*) growth, diet, and energy requirements, and diagramming hypotheses tested in the present study. Throughout this manuscript, amino acid breakdown refers to breaking down an amino acid to use it as an energy source. Metabolites specific to the breakdown pathways of specific essential amino acids, indicated below, were selected as markers of these processes. Meanwhile protein degradation refers to proteins that are being separated back into their constituent amino acids. For the current study, I used metabolites in the blood that would come specifically from post-translationally modified amino acids because their release is the result or proteins that have been synthesized and subsequently degraded. I thus hypothesize that the collapse of rainbow smelt, primarily in the north basin, led to declines in relative mass and length-at-age in north basin walleye, which in turn may be reflected as increased signals of endogenous protein degradation and amino acid oxidation by the presence of specific blood metabolites.

into amino acid oxidation in wild fish because the animal must get them through its diet, and cannot rely solely on its own metabolic processes for their supply. If dietary essential amino acid intake is insufficient, the fish cannot grow or may even fail to maintain body mass. For this reason, increased breakdown of essential amino acids would coincide with either excess intake from the diet or a need to catabolize body protein to make up for an energetic deficiency. Amino acid breakdown can be implicated by the presence of metabolites that result from their breakdown, and essential amino acids (specifically methionine, tryptophan, and lysine in the current study) are of particular use in this context because their presence represents only food intake or protein breakdown. Assuming conserved pathways of amino acid breakdown across animals, methionine breakdown results in the formation of dimethyl sulfone (Engelke et al., 2005), tryptophan is linked to kynurenine production (Knox & Mehler, 1950), and lysine to α-aminoadipic acid (Borsook & Deasy, 1948). As such, I argue that the elevation essential amino acids and concomitant increase in levels of their amino acid-specific breakdown products may indicate a physiological need for fish to increase the oxidation of these specific amino acids for energy, over preferential recycling of essential amino acids back into protein.

To assess possible biomarkers of protein degradation, our strategy focused on 'post-translational' protein modifications, which are physical changes to amino acids following their incorporation into a protein molecule. These post-translational modifications are in some cases retained by the amino acid following degradation of the protein back to free amino acids (Fig 3.2). Thus the presence of these modified amino acids in blood may act as potential markers of endogenous protein breakdown. Here, I focus on three modified amino acids: hydroxyproline, a major constituent of collagen (Prockop & Sioerdsma, 1961; Stetten, 1949), dimethylarginines, which are enzymatic modifications of arginine (Vallance & Leiper, 2004), and methionine sulfoxide, which is a non-enzymatic modification of methionine that can occur in 4-8% of all methionine found within proteins (Stadtman et al., 2005). The proteinaceous origin of these post-translationally modified metabolites may thus elucidate the relative rate of endogenous protein degradation in walleye from different areas of Lake Winnipeg (see Fig 3.2).

### 3.2.2 Linking Metabolites and Spatial Patterns

In the present study, I examined both temporal and spatial changes in Lake Winnipeg walleye morphological measurements using gillnet index data collected by Manitoba Government biologists. First, length-at-age was used to assess longer-term trends in growth prior to 2017. Differences in walleye length-at-age may be an effective metric to study the historical collapse of rainbow smelt (approximately 2013), because length-at-age is a summary of prior growth while nutritional deprivation has little effect on length (Caruso et al., 2011; Einen & Thomassen, 1998; Sumpter et al., 1991). On the other hand, body condition or length-mass relationships can vary with a single spawning or feeding event, each of which change a fish's mass. Spatial patterns in length-at-age (≈ growth rate) and length-mass relationships (≈ body condition) are thus assessed concurrently. I then linked length-mass relationships to a survey of walleye in Lake Winnipeg in 2017, where blood was sampled non-lethally for metabolites included in our nutritional biomarker development strategy (see above; Fig 3.2). Spatial differences in the presence of chosen metabolites were then tested with estimated marginal means (Lenth, 2019; Searle et al., 1980) and linear models incorporating length and site collected.

I hypothesized that because the collapse of rainbow smelt was most pronounced in the north basin of Lake Winnipeg (Manitoba Government, 2018), north basin walleye would exhibit the greatest decline in relative mass over time and length-at-age. I also hypothesized that metabolites would implicate higher rates of endogenous protein breakdown (i.e., elevated hydroxyproline, dimethylarginine, and methionine sulfoxide) and amino acid oxidation (i.e., elevated dimethyl sulfone, kynurenine, and α-aminoadipic acid) in the north basin than in the south basin (Fig 3.2). This work thus 1) explored the effect of the collapse of rainbow smelt on Lake Winnipeg walleye morphology and 2) examined the potential for a suite of nine metabolites to reflect walleye nutritional status and be developed into a non-lethal sampling approach for assessing the nutritional state of wild-caught walleye.

## 3.3 Methods

### 3.3.1 Lake Winnipeg Gillnet Index Data

I used the Lake Winnipeg gillnet index data collected by the Government of Manitoba between 2009 and 2018 (data accessed September 2019: https://www.gov.mb.ca/sd/fish_and_wildlife/fish/commercial_fishing/netting_data.html; URL updated May 2021: https://gov.mb.ca/fish-wildlife/fish/commercial_fishing/netting_data.html). This gillnet index has been run annually by the province since 1979 to provide an alternative to commercial fisheries data to better track trends in size and abundance for walleye and sauger (*Sander canadensis*) in Lake Winnipeg. I focused on the data collected between 2009 and 2018 because there was consistent monitoring of the same six sites during that period and these data included age estimates for each fish. To examine walleye relative mass and length-at-age relationships over time, sauger and dwarf walleye (which possibly inhabit a different ecological niche and are responding to selective pressures; see Moles et al., 2010) were filtered out. The remaining data set included year, site, basin collected, gillnet mesh size, fork length, mass, and sex for adult and sub-adult walleye. Fork length and mass were transformed on a $\log_{10}$ scale. Gillnet mesh size was included as an independent variable where possible to account for bias in walleye length and mass caught in gillnets with different mesh sizes. For all linear models, I chose variables based on their biological significance, as opposed to taking a stepwise model selection approach. Statistical modeling was performed in (R Core Team, 2021) using the packages emmeans (Searle et al., 1980), sjstats (Lüdecke, 2019), and tidyverse (Wickham et al., 2019). Scripts used for all statistical analyses in this study are available at: https://github.com/BioMatt/walleye_condition_metabolites.

### 3.3.2 Lake Winnipeg Walleye Length-at-Age Over Time

Changes in growth rate over time were assessed for walleye from seven sites across Lake Winnipeg using length-at-age data from the gillnet index, with a special focus on growth rates before and after the rainbow smelt collapse (in approximately 2013). Length is a useful measure because a fish facing nutritional deprivation decreases in mass, but decreases in length either

negligibly or not at all (Caruso et al., 2011; Einen & Thomassen, 1998; Sumpter et al., 1991). Length-at-age therefore represents more long-term trends in growth than year-specific values for body condition or length-mass relationships. A separate linear model was used for each site sampled for the gillnet index data, with Dauphin River and Grand Rapids representing the north basin, Matheson Island and Frog Bay representing the channel, and Riverton and Grand Beach representing the south basin of Lake Winnipeg (Fig 3.1). These linear models used fork length as the dependent variable, and the independent variables were age interacting with year, sex, and gillnet mesh size. Results were therefore averaged over gillnet mesh sizes to account for sampling bias. Only walleye ages two through six years old as estimated by counting annuli in otoliths were included in the models, because not all sites had sufficient sample sizes available for walleye of older and younger ages ($n \geq 10$ individuals age class$^{-1}$ year$^{-1}$ site$^{-1}$, except the Dauphin River age two fish in 2012 and 2014 were included, $n$=5 and 2, respectively). A total of $n$=2164 individuals were used in the models for Dauphin River, $n$=2745 for Grand Rapids, $n$=844 for Matheson Island, $n$=959 for Frog Bay, $n$=1732 for Riverton, and $n$=1729 for Grand Beach collection sites. Estimated marginal means and estimated marginal trends were used to investigate pairwise mean and slope differences between sites for each model (Lenth, 2019; Searle et al., 1980). Briefly, the analysis of estimated marginal means obtains predictions from a linear model using Tukey's post hoc test and finds meaningful averages to summarize primary factor effects while estimated marginal trends follows the same process but for the interaction between two predictors in a model (see Lenth, 2019).

### 3.3.3 Lake Winnipeg Walleye Relative Mass Over Time

Relative mass over time was examined for walleye across Lake Winnipeg. These relative mass measures were used as metrics analogous to body condition, which is an imperfect approach to our data because walleye mass may change in a single feeding or spawning event, thus changing a fish's body condition and length-mass relationship. However, an examination of relative mass remains useful because it represents a link between inter-annual growth rate trends and metabolite presence, which can vary on much faster timescales. Only walleye ≥375 mm in fork length from the gillnet index data, $n$=5,838 individuals (mean 583.80 walleye ± 220.17 s.d. year$^{-1}$), were included in the model because this value represents the smallest estimate of fork

length for mature individuals among Manitoba lakes (Craig et al., 1995), and is a conservative threshold for modeling the fish that were sampled for metabolites (minimum fork length 452 mm).

To study how relative mass has changed over time, a linear model was used with mass as the dependent variable, and independent variables were fork length and its interactions with site, year, and sex. Eta squared is reported for the independent variables in this model, where analysis of variance applied to the model outputs returns the percentage of variance in the final model accounted for by an independent variable (Levine & Hullett, 2002).

### 3.3.4 Spatial Differences Among Walleye in 2017

In addition to exploring length-mass relationships over time, I modeled spatial relationships among Lake Winnipeg walleye using the gillnet index data set in 2017 to establish possible connections between length-mass relationships and metabolite presence in that year. Here, $n$=286, 127, and 227 individuals remained from the Dauphin River, Matheson Island, and Riverton sites, respectively. These three sites were chosen because they are most similar to the three sites sampled for metabolites (i.e., Dauphin River, Matheson Island, and the Red River). This model used mass as the dependent variable, with the independent variables as the interaction of mass with fork length, site collected, sex, age, and gillnet mesh size.

Another linear model was used to assess the length-mass relationship of walleye caught as part of the blood metabolite study in 2017. The smallest walleye in this data set was 452 mm in fork length. Mass was the dependent variable, with the independent variables fork length interacting with site collected, sex, and the interaction of fork length and sex. Note that gillnet mesh size and age were not included in the model since fish were caught by electrofishing and age was not known for these fish.

To examine how similar the length-mass relationships were of walleye were between those collected for metabolite analysis and those for the gill net index data, I used a linear model with an independent variable describing which study an individual was collected for. Here, mass was the dependent variable, which was related to fork length, basin collected, sex, study (metabolite or gillnet index), the interaction of fork length and basin, the interaction of fork

length and sex, and the interaction between fork length and study. In this model, only gillnet index samples from the Dauphin River, Matheson Island, and Riverton were included. Riverton from the gillnet index data and the Red River from the metabolite data were used to jointly represent the south basin of Lake Winnipeg.

### 3.3.5 Metabolite Data Collection

In 2017, 39 walleye were collected from three sampling locations: the Red River, Matheson Island and Dauphin River representing the south basin, channel, and north basin of Lake Winnipeg, respectively (Fig 3.1). Each sampling location was at a known spawning site during spawning season (May 2$^{nd}$ in the Red River, May 17$^{th}$ and 18$^{th}$ at Matheson Island, and May 29$^{th}$ through 31$^{st}$ at the Dauphin River). Both males and females were collected, with $n$=17 from the Red River (2 males, 15 females), $n$=5 from Matheson island (2 males, 3 females), and $n$=17 from the Dauphin River (5 males, 12 females), and had a minimum mass of 1 kilogram (mean 2.32 kg ± 0.97 standard deviation). Individuals were collected by boat electrofishing, held in a live well for no longer than one hour, and anaesthetized using a Portable Electroanesthesia System (PES™, Smith Root, Vancouver, Washington, USA) in accordance with approved animal use protocols of Fisheries and Oceans Canada (FWI-ACC-2017-001, FWI-ACC-2018-001), the University of Manitoba (F2018-019) and the University of Nebraska-Lincoln (Project ID: 1208).

One milliliter of whole blood was collected for metabolite analysis from anaesthetized walleye by caudal puncture using a heparinized needle and 3 ml syringe, flash frozen in liquid nitrogen immediately after sampling, and stored at -80˚C. Of note, logistics prevented separation of plasma or serum and therefore all metabolite data are for whole blood. This effort was part of a larger study assessing the physiological health, movement, and genetic structure of walleye in Lake Winnipeg. Other tissues collected at the time of blood sampling include a fin clip, gill filaments, the first dorsal spine, scales, and a muscle biopsy. All fish were sampled non-lethally and a VEMCO acoustic tag (VEMCO, Bedford, Nova Scotia, Canada) was surgically implanted prior to release near the collection site.

### 3.3.6 Blood Sample Analyses

Walleye blood samples collected in 2017 (see above) were analyzed using nuclear magnetic resonance (NMR) spectroscopy and a combination of direct injection mass spectrometry with a reverse-phase liquid chromatography with mass spectrometry (DI/LC-MS/MS) assay at the University of Alberta Metabolomics Centre TMIC (Edmonton, AB, Canada) as part of a large scale targeted metabolic study. Here I examine a small subset of analytes to identify potential biomarkers of protein degradation and amino acid breakdown as described in Fig 3.2. Both NMR and DI/LC-MS/MS detected methionine and lysine, so measurements for each metabolite were averaged over the two detection methods. Tryptophan and dimethyl sulfone were only measured by NMR. Kynurenine, hydroxyproline, dimethyl sulfone, and α-aminoadipic acid were detected with DI/LC-MS/MS.

For NMR spectroscopy, deproteinization, involving ultra-filtration, was performed to remove proteins. Filtration, and centrifugation steps were subsequently done to further purify the sample (Psychogios et al., 2011). 250 µL of the blood sample was transferred to a 3 mm SampleJet NMR tube for subsequent spectral analysis following a protocol based on Saude et al., (2006). NMR spectra was collected on a 700 MHz Avance III (Bruker) spectrometer and the spectra acquired at 25°C. NMR spectra were processed and analyzed using the Chenomx NMR Suite Professional software package version 8.1 (Chenomx Inc., Edmonton, AB). DI/LC-MS/MS was done on an API4000 Qtrap® tandem mass spectrometry instrument (Applied Biosystems/MDS Analytical Technologies, Foster City, CA) equipped with an Agilent 1260 series HPLC system (Agilent Technologies, Palo Alto, CA). The samples were delivered to the mass spectrometer by a LC method followed by a direct injection (DI) method. Data analysis was done using Analyst 1.6.2.

### 3.3.7 Modeling Metabolites Differences

Differences in nine metabolites across sampling sites were assessed using separate linear models. Dimethyl sulfone, kynurenine, α-aminoadipic acid, methionine, tryptophan, lysine, hydroxyproline, dimethylarginine, and methionine sulfoxide were used as dependent variables in

their respective linear models. For each model, the independent variables were log fork length and site collected. Estimated marginal means of metabolite presence were calculated to establish pairwise significance between sites. Significance for each fork length and site collected were calculated, along with eta squared (Levine & Hullett, 2002) to report effect size for each independent variable.

## 3.4 Results

### 3.4.1 Lake Winnipeg Walleye Length-at-Age Over Time

Each of the Dauphin River, Grand Rapids, Matheson Island, Frog Bay, Riverton, and Grand Beach sites exhibited a decline in length-at-age for age six walleye between the years 2012 and 2018 (Fig 3.3). However, for the age two and three walleye, while the Dauphin River (north basin) site also showed a shorter length-at-age in 2018 compared to 2012, the south basin sites (Riverton and Grand Beach) showed similar length-at-age in later years (i.e., 2017 and 2018) relative to earlier years (i.e., 2012 and 2013). The Grand Rapids (in the north basin) showed the most consistent decline in length-at-age for all ages over time except for the age two fish sampled in 2017 and 2018 (Fig 3.3).

### 3.4.2 Lake Winnipeg Walleye Relative Mass Over Time

From the years 2009 through 2018, year, sex, and the interaction between fork length and year had a significant relationship with Lake Winnipeg walleye relative mass ($F = 2899$, $p < 2.2 \times 10^{-16}$, adjusted $R^2 = 0.95$, Table A.2.1). While fork length had the greatest effect size as indicated by eta squared, year had a greater effect size than site collected during the time period studied. Confidence intervals determined using estimated marginal means indicated significant differences among the effect of years on mass, with a drop in predicted mass most noticeable between the years 2014 and 2015 (Fig 3.4A).

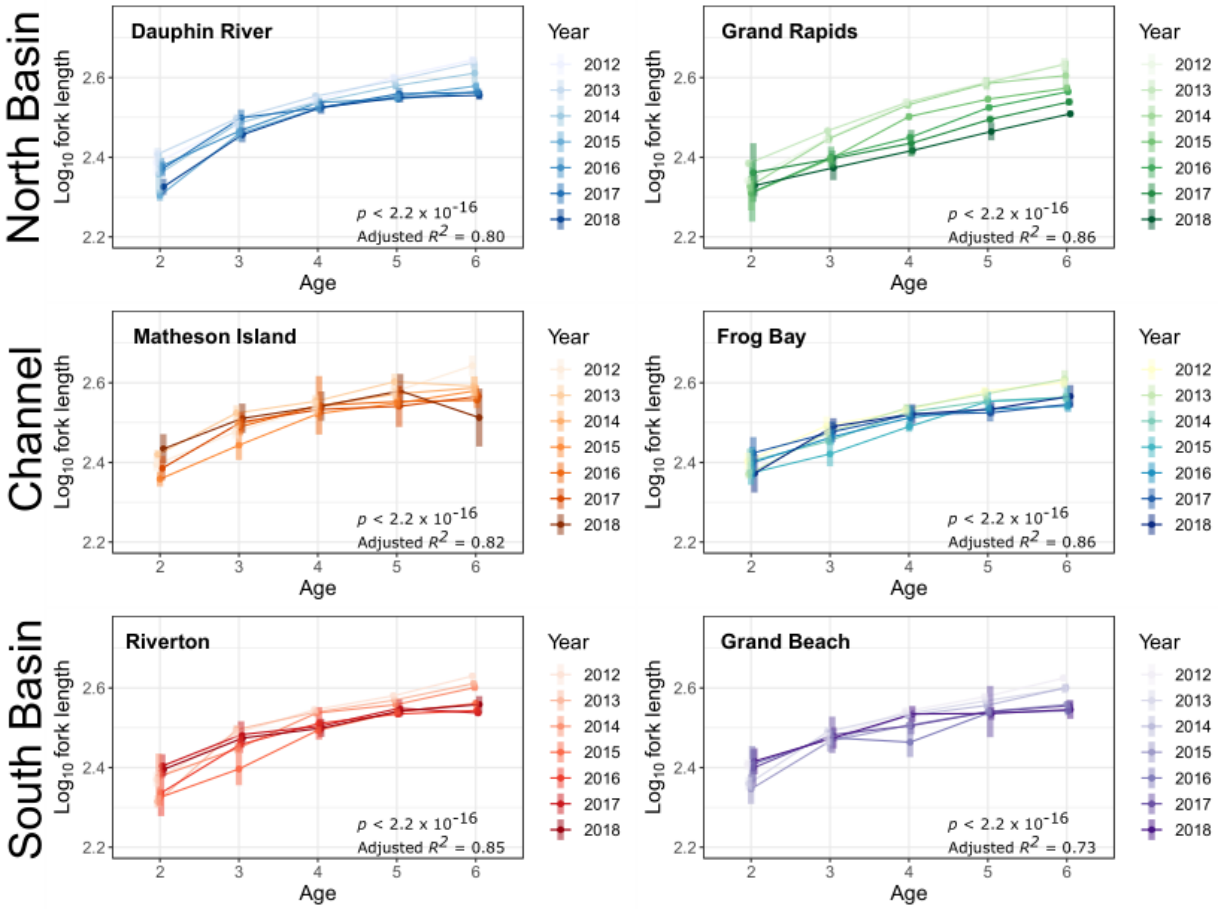


**Fig 3.3** Estimated marginal means of $\log_{10}$ fork length-at-age for walleye (*Sander vitreus*) from 2012 to 2018 across gillnet index collection sites in Lake Winnipeg, showing a decrease in growth rates among walleye of all sites over time that is especially clear in the north basin. The data were collected by the Government of Manitoba. These figures are derived from linear models relating $\log_{10}$ fork length to the interaction of age with year, sex, and gillnet mesh size. A separate model was used for each collection site. 95% confidence intervals are provided as error bars at each age. Overall model significance and adjusted $R^2$ is provided in each panel.

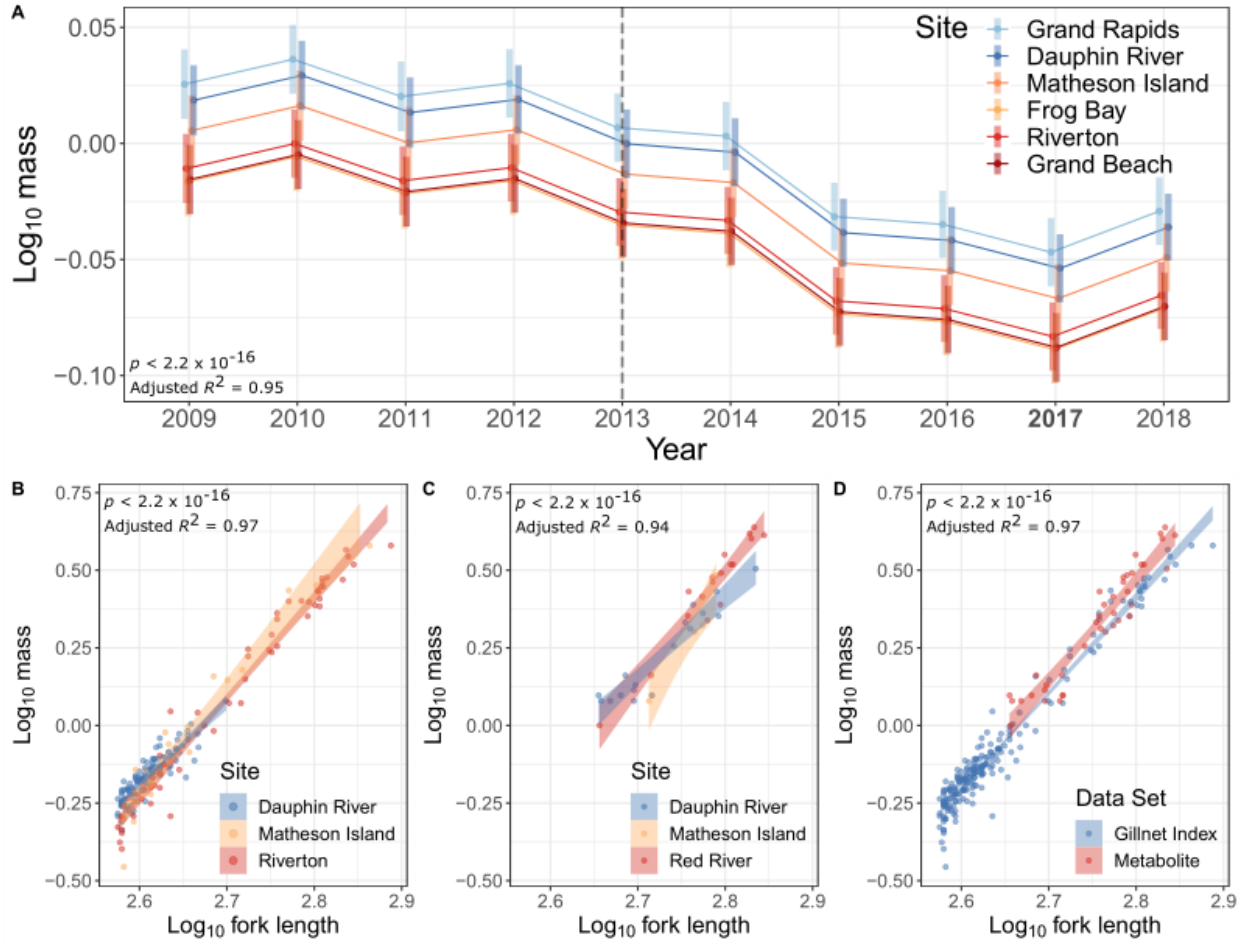


**Fig 3.4** Relative mass over time and length-mass relationships among Lake Winnipeg walleye (*Sander vitreus*) caught in 2017, showing declines in relative mass over time at all sites, different slopes of length-mass relationships for north basin walleye compared to others, and comparability between gillnet index and metabolite datasets. Panel A represents estimated marginal means of $\log_{10}$ mass between 2009 and 2018 across all sites collected for the gill net index data in Lake Winnipeg walleye (*Sander vitreus*). These data are pulled from a linear model relating $\log_{10}$ mass as a dependent variable to $\log_{10}$ fork length, sex, site, and mesh size as independent variables ($F$ = 2686, $p < 2.2 \times 10^{-16}$, adjusted $R^2$ = 0.98). 95% confidence intervals are provided as error bars at each year. A dashed line in the year 2013 represents the approximate time of the rainbow smelt collapse across Lake Winnipeg. For panel B, the linear model related $\log_{10}$ mass to the interaction of $\log_{10}$ fork length and basin collected, sex, mesh size, and age ($F$ = 222, $p < 2.2 \times 10^{-16}$, adjusted $R^2$ = 0.97) using the gillnet index data from the Dauphin River, Matheson Island, and Riverton sites. Panel C represents a linear model on walleye used for fish

from the metabolite measurements, and this linear model related $\log_{10}$ mass in kg to $\log_{10}$ fork length interacting with site collected, sex, and the interaction of $\log_{10}$ fork length and sex ($F$ = 86, $p < 2.2 \times 10^{-16}$, adjusted $R^2$ = 0.94). Panel D shows a linear model comparing length-mass relationships in walleye between those used for metabolites and from the gillnet index data. This model relates $\log_{10}$ mass as the dependent variable, with $\log_{10}$ fork length, basin collected, sex, study (metabolite or gill net index), the interaction of $\log_{10}$ fork length and basin, the interaction of $\log_{10}$ fork length and sex, and the interaction between $\log_{10}$ fork length and study as independent variables ($F$ = 889, $p < 2.2 \times 10^{-16}$, adjusted $R^2$ = 0.97). Standard error is shown with filled colors. Overall model significance and adjusted $R^2$ is provided in each panel. The gillnet index data were collected by the Government of Manitoba and only walleye ≥375 millimeters in fork length are included in these models, while in the metabolite data are large fish (≥452 mm in fork length) caught by electrofishing, from which whole blood metabolites were measured. $\text{Log}_{10}$ mass was calculated in kilograms, and $\log_{10}$ fork length in millimeters.

### 3.4.3 Spatial Differences Among Walleye in 2017

In a linear model using gillnet index data for fish captured in 2017 that were ≥375 mm in fork length and including only the Dauphin River, Matheson Island, and Riverton sites, ($F$ = 222, $p < 2.2 \times 10^{-16}$, adjusted $R^2$ = 0.97), estimated marginal mean length-mass relationships were the same for three sites, but estimated marginal trends were lower in the Dauphin River compared to Matheson Island (Table A.2.2; Fig 3.4B). This suggests a different length-mass relationship for the north basin fish compared to walleye caught in the channel.

Similar to the larger fish sampled as part of the gill net index in 2017, for length-mass relationships among walleye collected for metabolites ($F$ = 86, $p < 2.2 \times 10^{-16}$, adjusted $R^2$ = 0.94), estimated marginal trends were also lower in the Dauphin River compared to Matheson Island (Table A.2.2; Fig 3.4C). However, unlike the linear model used with the gillnet index data, the Red River (south basin) showed a higher estimated marginal mean mass than the Matheson Island (channel) site (Table A.2.2; Fig 3.4C).

When comparing length-mass relationship of larger walleye collected in 2017 as part of the gill net index and metabolite studies, no significant effect of study and of study interacting fork length was found ($F$ = 889, $p < 2.2 \times 10^{-16}$, adjusted $R^2$ = 0.97, Table A.2.3). However, estimated marginal means of mass based on fork length were higher in the walleye from the metabolite data ($p$ = 0.00070) while estimated marginal trends for fork length-mass relationships between studies were not different ($p$ = 0.90). A plot of fork length and mass with these data reveals that the metabolite-measured walleye lie at the upper end of the distribution in length-mass for walleye collected for the gillnet index data, while following a similar slope (Fig 3.4D).

### 3.4.4 Modeling Metabolites Differences

Metabolite presence of the three essential amino acids varied significantly across sampling sites for Lake Winnipeg walleye sampled in 2017. The models predicting the presence of each of the three free essential amino acids were each significant ($F$ = 4.3, 4.7, and 3.7, $p$ = 0.011, 0.0071, and 0.020, adjusted $R^2$ = 0.21, 0.23, and 0.18 for methionine, tryptophan, and lysine respectively, Table A.2.4). In addition, site was a significant independent variable within the overall models for the three essential amino acids while fork length was not, with eta squared

higher for site than for fork length in each case (Table A.2.4). While adjusted $R^2$ were low in an absolute sense, the significance and eta squared for site as an independent variable nevertheless demonstrated predictive power for site and free essential amino acid presence. Estimated marginal means for each of the three essential amino acids revealed significant differences in predicted essential amino acid presence based on site, with higher values in the Dauphin River (north basin) than in the Red River (south basin) (Fig 3.5 methionine, tryptophan, and lysine).

Overall, there was no effect of sampling location or fish size (i.e., fork length) on essential amino acid breakdown biomarkers. Linear models for α-aminoadipic acid, kynurenine, and dimethyl sulfone were not significant ($F$ = 0.32, 2.6, and 3.3, $p$ = 0.81, 0.070, and 0.031, adjusted $R^2$ = 0.057, 0.11, and 0.15, respectively, Table S4), indicating a failure to fit fork length and site to essential amino acid breakdown metabolite presence. The model for dimethyl sulfone was significant ($p$ = 0.031), but neither site nor fork length were significant independent variables ($p$ = 0.36 and 0.088, respectively). Predicted metabolite presence from estimated marginal means showed no significant differences in amino acid breakdown metabolites among sites, as well (Fig 3.5 dimethyl sulfone, kynurenine, and α-aminoadipic acid).

Endogenous protein breakdown metabolite presence varied significantly with collection site in Lake Winnipeg walleye sampled in 2017. Linear models run with hydroxyproline, dimethylarginine, and methionine sulfoxide, potential endogenous protein breakdown biomarkers, were each significant ($F$ = 24.12, 6.2, and 6.3, $p$ = 1.2 x $10^{-8}$, 0.0017, and 0.0016, adjusted $R^2$ = 0.65, 0.29, and 0.29, respectively, Table A.2.4). Moreover, site collected was both a significant independent variable and had a higher effect size than fork length in each model (Table A.2.4). Predicted presence for each candidate metabolite associated with protein breakdown was higher in the northern Dauphin River than the southern Red River (Fig 3.5 hydroxyproline, dimethylarginine, and methionine sulfoxide). Hydroxyproline and dimethylarginine were also higher in fish from the Dauphin River than at Matheson Island in the channel.

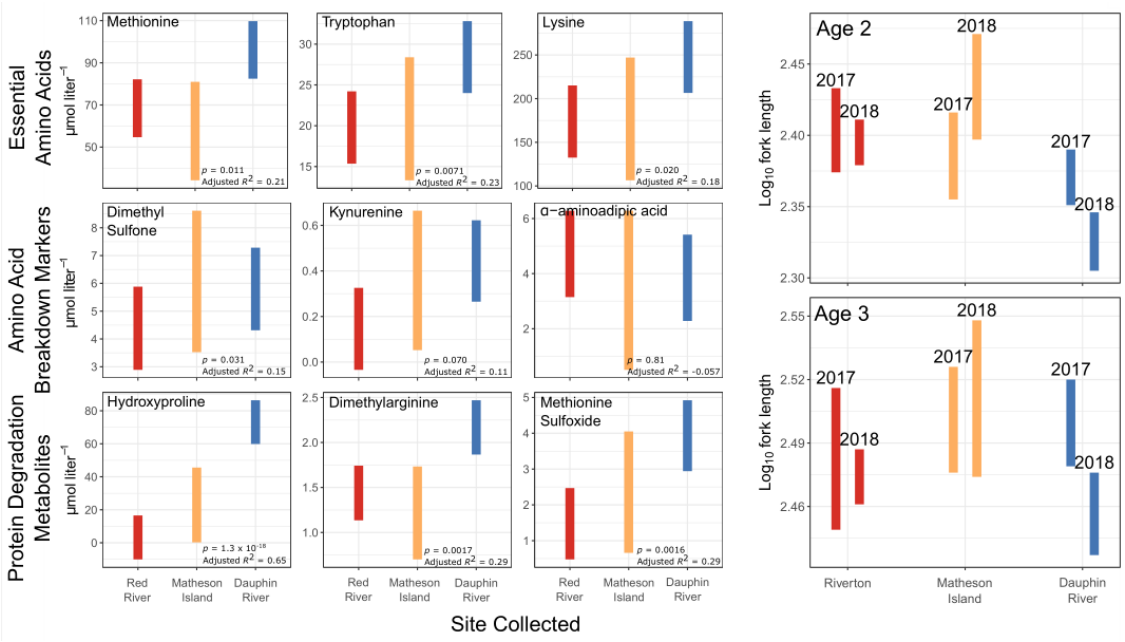


**Fig 3.5** Essential amino acids and protein degradation metabolites are elevated in the Dauphin River (north basin) walleye relative to others, concurrent with decreases in growth rates in the north basin fish, showing that a nutritional limitation may have affected the concentration of these metabolites. Estimated metabolite concentration (µmol $l^{-1}$ in whole blood) for Lake Winnipeg walleye (*Sander vitreus*) by site is predicted by linear models that incorporate metabolite as the dependent variable, with $\log_{10}$ fork length and site collected as the independent variables. The three sites on the horizontal axis for metabolite models are the Red River (in red, south basin), Matheson Island (in yellow, channel), and Dauphin River (in blue, north basin). Here, $n$=39 large fish (≥452 mm in fork length) were caught by electrofishing in 2017. On the vertical axis are estimated marginal means for each metabolite. Overall linear model significance and adjusted $R^2$ is provided in each metabolite panel. Panels Methionine, Tryptophan, and Lysine represent essential amino acids. Panels Dimethyl Sulfone, Kynurenine, and α-aminoadipic acid represent respective amino acid breakdown metabolites that in conjunction with the top three panels may describe amino acid oxidation. Panels Hydroxyproline, Dimethylarginine, and Methionine Sulfoxide represent metabolites associated post-translational modification and therefore may reflect endogenous protein degradation (see Fig. 3.2 for additional details). The panels for Age 2 and Age 3 represent estimated marginal means of length for each age between

the years 2017 and 2018 at three sites from the gillnet index dataset collected by the Government of Manitoba. Ages two and three are emphasized because they represent the growth phase of the walleye lifecycle, so length-at-age would be most sensitive to nutritional deprivation. These estimates are derived from linear models relating $\log_{10}$ fork length to the interaction of age with year, sex, and gillnet mesh size, with a separate model for each site (see Fig. 3.3 for additional details). The three sites on the horizontal axis for length-at-age models are Riverton (in red, south basin), Matheson Island (in yellow, channel), and Dauphin River (in blue, north basin). For all panels, bars represent 95% confidence intervals.

## 3.5 Discussion

### 3.5.1 Morphological Observations

I argue that the walleye used for assessing metabolite levels are likely representative of large walleye captured in the gillnet index data, and that spatial differences in the slopes of length-mass relationships (≈ body condition) and length-at-age (≈ growth rate) are consistent with spatially varying metabolite presence. I used the Lake Winnipeg gillnet index data set to examine length-at-age and relative mass to assess trends in the growth of walleye from 2009 to 2018. The decreases in walleye mass appeared to coincide with the near complete collapse of the rainbow smelt population in 2013 as there was a significant drop in relative mass (analogous to body condition) at all sites in the gillnet index between 2014 and 2015. The pattern of overall decrease in mass remained in fish collected in the index until 2018, the last year of data available at the time of analyses. I attempted to control for as many possible confounding variables in the analyses, however the overall trend showed a decrease in relative mass regardless of sex or gillnet mesh size, a pattern that is consistent with previous analyses (Manitoba Government, 2018). Contrary to our predictions, the general trend of decreasing relative mass was not more severe in the north basin (see Fig 3.5A) where rainbow smelt contributed to a larger portion of the diet of walleye (Sheppard et al., 2015). Given that the decrease in relative mass occurred at all sites sampled, these data suggest ecosystem-wide changes in Lake Winnipeg have contributed to decreases in walleye size over the past decade.

Length-at-age measurements showed a pattern of decreased length in two-year old fish over time, which was most dramatic for fish caught at the Dauphin River site between 2017 and 2018. Reduced food availability or an increase in energetic costs in 2016 and 2017 for young fish at the Dauphin River may have led to this decreased length-at-age. These length-at-age estimates were more stable measurements of changes in growth patterns between 2012 and 2018 because they are less variable than measures of mass that can fluctuate more rapidly (e.g., pre- and post-spawning, post-overwintering). At all sites in the gillnet index, there was also an overall decrease in the length of 6-year-old fish from 2016 to 2018. The pattern of decreased length in fish was most prominent in fish collected at the most northern site sampled, Grand Rapids. I believe that this decrease in fish growth in later years (i.e., 2016–2018) of the six-year-old fish may represent

a delayed effect from an earlier large-scale change in the ecosystem (e.g., collapse of the rainbow smelt population in 2013). The decrease in the length at age in older fish may have further impacts on walleye abundance in Lake Winnipeg as the length of fish and fecundity is highly correlated (Craig et al., 1995; Wolfert, 1969). Therefore, the population-level impacts of the decrease in the size of walleye in Lake Winnipeg may become a bigger issue in the future.

While the present study is focused on the collapse of the rainbow smelt, additional differences between basins may underlie observed spatial patters. Some of these basin-level differences include higher temperatures, precipitation, river discharge, suspended solids, sulphate, phosphorous, and nitrogen in the south relative to the north basin (Environment Canada, 2011). Phosphorous loading and summer surface temperatures may contribute to algal blooms, which were more prevalent in the south basin (Binding et al., 2018; Environment Canada, 2011). Meanwhile, sodium and chloride are twice as high in the north than in the south, likely because of inflow from the Dauphin River (Environment Canada, 2011). Prey fish populations other than the rainbow smelt also differ spatially, with emerald shiner (*Notropis atherinoides*) and cisco (*Coregonus artedi*) more abundant in the south basin (between 2002-2008) as well as in the diets of south basin walleye (in 2010 and 2011) (Lumb et al., 2012; Sheppard et al., 2015). When these differences between basins are considered in conjunction with data showing that rainbow smelt did not historically make up the entirety of walleye diets (Sheppard et al., 2015), it becomes clear that the rainbow smelt collapse is one of many potential factors that have affected walleye growth. Nevertheless, the disparity in growth rate (length-at-age) between basins, the higher abundance of rainbow smelt in the north basin, and its prevalence in north basin walleye diets in 2010 and 2011 (Sheppard et al., 2015) despite high walleye connectivity between basins (Backhouse-James & Docker, 2012; Chapter 2) suggests that the rainbow smelt collapse had a large effect on the Lake Winnipeg walleye fishery.

### 3.5.2 Linking Data Sets

The metabolite and gillnet index datasets were consistent in patterns of possible decreased food availability in the north basin in 2017 (when metabolites were measured), based on analyses of spatially varying growth rates in the 2017 gillnet index data and slopes of length-mass relationships in both datasets. While differences of slope in length-mass relationship

between the Dauphin River and Matheson Island in the metabolite data may be explained by differences in spawning status—many Dauphin River fish, sampled later than others, may have been sampled post-spawn—slopes in length-mass relationships between sites are consistent between the gillnet index data and metabolite data. The difference in estimated marginal mean length-mass relationship may be explained by spawning status as well, where higher 'fatness' may be observed in the fish used for metabolites because of pre-spawn individuals. Spawning status may thus explain differences among length-mass relationships in sampled walleye, providing evidence that the fish used for metabolites represent those used in the gillnet index. Moreover, data set origin (gillnet index or metabolite-measured walleye captured via electrofishing) was not a significant predictor of overall length-mass relationship in a combined linear model, providing additional evidence for no systematic bias in length-mass relationship in one data set or another, even if the walleye used for metabolites were relatively large. As such, morphological observations were mainly consistent between walleye used for gillnet index data and metabolite data.

The time scale in response variables should be considered when relating ecology to length-at-age, length-mass relationships, and metabolite levels across data sets. Length-at-age estimates are summaries of one or more prior years of growth for a cohort of walleye, while mass may change from a single feeding or spawning event, thus changing length-mass relationships. In addition, while I do not have information for the response time of amino acids or their metabolites, other metabolites such as blood glucose and lactate increase with handling on a timescale of minutes, indicating that amino acid metabolite presence may also reflect the past several minutes of a fish's life as opposed to ecological patterns (Chopin et al., 1996; Grutter & Pankhurst, 2000; Lawrence et al., 2018; Meka & McCormick, 2005). However, the consistency in patterns across timescales— stronger signals of protein breakdown, more shallow length-mass relationships, and slower growth in the north basin in 2017—supports the validity of this approach for integrating data from different levels of biological organization. The breadth in timescales analyzed may also be a benefit for integrating information across data sets because each piece of information provides context for other results.

### 3.5.3 Metabolites

I observed from a preliminary analysis of a large-scale targeted metabolomics study (with 163 unique metabolites) that essential amino acids varied in whole blood from walleye caught from different regions of Lake Winnipeg (Wiens, Jeffrey, Treberg unpublished data), which provided the impetus to pursue the nutritional biomarkers strategy that I employed in the current study. I focused on three essential amino acids (methionine, tryptophan and lysine), that were each elevated in the Dauphin River fish relative to the more southern Red River fish (Fig 3.5). Whole blood was necessary because it was not possible for us to separate plasma or serum in the field, and I therefore cannot distinguish between differences at the cellular level (mainly red blood cells), or extracellular component of blood. Despite this limitation, because blood acts as a connection between all organs and tissues, I are nevertheless confident that the more north basin walleye had elevated essential amino acids in circulation.

Another limitation is that measured metabolites may thus have been affected by the approximately one-month range in sampling. Metabolites were sampled at sites in a south-to-north fashion, following both ice melt across Lake Winnipeg, and walleye spawning during the month of May, 2017. It is possible, therefore, that walleye captured at the southern Red River site had not spawned at the time of sampling, while walleye captured at the northern Dauphin River site were post-spawn. Multicollinearity exists between sampling date and sampling site because of the south-to-north sampling; linear models using both variables thus lack statistical power. However, amino acid concentrations in fish blood appear to remain stable after spawning (Gauthey et al., 2015). I therefore have reason to believe that amino acid differences across basins reflect spatial nutritional differences in walleye, as opposed to being artifacts of sampling. An additional limitation is that I could not verify whether sampled individuals had spawned at their collection site, or if they had spawned elsewhere and moved to the collection site.

Because feeding success is one possible source of high essential amino acid presence in blood, I investigated patterns of growth across walleye in Lake Winnipeg and found that age two and three north basin fish in 2017 displayed lower length-at-age, or less growth than other walleye (Fig. 3.3, Fig. 3.5) (Blasco et al., 1991; Costas et al., 2011; Schuhmacher et al., 1995). These results suggest that the northern fish may have had lower feeding success in 2017 compared to the faster growing south basin walleye. Nevertheless, more information beyond

growth rates is needed to provide sufficient context for interpreting plasma amino acid levels. Feeding success directly after spawning, but prior to sampling, may have influenced observations of high essential amino acids as well. However, sampling was conducted at known spawning sites, and walleye of other systems were observed to leave spawning sites for feeding areas within one week of spawning (Rasmussen et al., 2002). In conjunction with observations of length-at-age differences in young fish between 2017 and 2018, the presence of walleye, even post-spawn individuals, at spawning sites suggests they likely had not started feeding to any great extent.

I used a panel of metabolites of amino acid degradation to infer the increased oxidation of amino acids, or in other words, the allocation of amino acids toward energy. The metabolite screening yielded results for metabolites specific to the degradation pathways each of the three essential amino acids I found to be elevated in the blood of north basin fish (Fig. 3.5 dimethyl sulfone, kynurenine, and α-aminoadipic acid). Animals use each specific amino acid for energy metabolism by committing each amino acid to its specific degradation pathway for oxidation. Thus, these degradation metabolites were three independent tests of whether essential amino acids are being directed towards energy metabolism. In all cases, there was no difference in the metabolites of amino acid oxidation across sampling site. Since most amino acid oxidation takes place outside of the blood (Ballantyne, 2001; Jürss & Bastrop, 1995), one caveat is that blood levels of these metabolites may not sufficiently reflect true whole animal amino acid oxidation. For that reason, the possibility that elevated amino acids reflect a greater reliance on amino acids as energy sources in the Dauphin River fish cannot be outright rejected. However, taken together with the lower growth in the north basin walleye in 2017, I can conclude that it is unlikely that the northern walleye have less reliance on protein oxidation than were south basin walleye.

As another source of amino acids in blood, protein turnover was an important subject of investigation, and spatial patterns of protein degradation were consistent with higher rates of protein turnover in the north basin of Lake Winnipeg than the south. To distinguish between amino acids in general and those amino acids that have already been incorporated into proteins, I focused on amino acids that have been modified after they were incorporated into proteins. Three unique protein modifications that arise from three separate routes are focused on: two that come from enzymatic processes (hydroxyproline and dimethylarginine), and one that occurs

spontaneously as exposed methionines are oxidized by reactive molecules such as hydrogen peroxide (methionine sulfoxide). While methionine sulfoxide can be formed from free methionine, most methionine from tissues is bound in proteins, therefore I assume that the bulk of circulating methionine sulfoxide is of proteinaceous origin. Because the means of forming these three post-translational modifications from amino acids represent independent pathways, I had no *a priori* expectation that the levels of all three post-translational modifications would be observed to vary in a consistent fashion across Lake Winnipeg. Nevertheless, all three modifications only present after protein degradation were higher in fish from the north than the south basin (Fig 3.5 hydroxyproline, dimethylarginine, and methionine sulfoxide).

If an increased reliance on protein oxidation for energy is *not* driving the elevated amino acids in north basin fish, then the question remains of why there was an increase in protein turnover in those northern fish. While protein metabolism is controlled by many factors, both protein synthesis and degradation rates in fish are known to increase with the level of swimming activity (Houlihan & Laurent, 1987). If the north basin walleye must spend more time swimming to find sufficient nutritional resources or reach spawning sites, then elevated levels of post-translationally modified amino acids could reflect a greater level of activity for the fish in the north. Relevant to these results, and contrary to what may be intuitive for carnivorous fishes, careful estimations of fuel usage in relation to swimming speed indicate that protein is likely not a major fuel source to supply the increased energy demand of increased swimming in rainbow trout (*Oncorhynchus mykiss*) and Nile Tilapia (*Oreochromis niloticus*) during short-term swimming (Alsop et al., 1999; Alsop & Wood, 1997); although, using the same strategies in Nile Tilapia indicated that prolonged swimming (at ~ 2.7 body lengths/second) over the course of over 48 hours did increase the relative reliance on protein to fuel swimming in unfed fish (Alsop et al., 1999). Because walleye in the wild appear to spend most of their time swimming at speeds below 1.0 body lengths/second (Kelso, 1978), it seems that the extended swimming response in tilapia is not likely applicable to the walleye sampled from the wild in the present study. In addition, a trade-off in growth rate and activity was observed in yellow perch (*Perca flavescens*), where slow growth was directly related to high activity rates in a comparison of wild perch in two lakes (Rennie et al., 2005). I therefore suggest that elevated signals of protein turnover in wild walleye may be due to increased swimming activity, which may be undertaken to reach more distant spawning sites or compensate for decreased food availability.

### 3.5.4 Conclusions

The data presented show declining growth rates and condition in Lake Winnipeg walleye in recent years, especially those in the north basin. These morphological differences are consistent with both the collapse of the rainbow smelt population and blood metabolites, suggesting increased endogenous protein breakdown in the north basin. With validation and refinement, the metabolites identified in the present study thus have potential for further development into molecular markers possibly useful as indicators of nutritional status for the walleye fishery. Molecular indicators of nutritional status would be valuable tools for resource managers for describing physiological thresholds in nutritional status that are predictive of detrimental effects on the walleye fishery (Connon et al., 2018). In other words, a molecular panel describing nutritional status may support the sustainable management of the Lake Winnipeg fishery.

## 3.7 Acknowledgments

I thank E. Enders, D. Watkinson, C. Charles, C. Kovachik, D. Leroux, N. Turner, M. Gaudry, S. Glowa, and E. Barker for their role in sampling the walleye used for metabolites. C. Charles and E. de Greef assisted with a map of Lake Winnipeg, and E. de Greef also provided immense support in the process of writing the manuscript. This work was supported by a Fisheries and Oceans Canada Ocean and Freshwater Science Contribution Program Partnership Fund grant awarded to J.R.T., K.M.J. and Darren Gillis, and Natural Sciences and Engineering Research Council of Canada Discovery Grants awarded to K.M.J. (#05479) and J.R.T. (#06052). Work by J.R.T. is also supported by the Canada Research Chairs program (#223744) and the Faculty of Science, University of Manitoba (#319254).

# Chapter 4. Intrapopulation Signatures of Selection Despite Translocations in Walleye of Three North American Waterbodies[3]

## 4.1 Abstract

Identifying within-population adaptive variation is significant for its relationship with fitness and the ability to consider adaptive variation as a property intrinsic to a population. However, many scans for selection consider adaptive variation in terms of comparisons between populations. Haplotype-based approaches have the potential to describe this within-population adaptive variation. Cross-population extended haplotype homozygosity (XP-EHH) was applied to reduced representation DNA sequencing data in *n*=345 walleye (*Sander vitreus*) of three North American waterbodies: Cedar Bluff Reservoir (Kansas, USA), Lake Manitoba (Manitoba, Canada), and Lake Winnipeg (Manitoba, Canada). Demographic history, gene flow, and population differentiation were analyzed to tease apart a history of possible gene flow (e.g., by stocking or flooding) among populations in this study, with implications for signatures of selection within each waterbody. Outlier tests were used to confirm signatures of selection identified with XP-EHH because of data imputation required for the haplotype-based method, but not outlier tests. Despite patterns of prior gene flow, signatures of selection remained in both Lake Winnipeg and Lake Manitoba, with possible implications for adaptations to cold waters. One region on chromosome 8 was identified to possibly show strong signatures of selection in Lake Winnipeg walleye specifically. Few signatures of selection were identified in the entirely-stocked Cedar Bluff Reservoir walleye. These results show the utility of considering adaptive variation within populations, which has implications for future walleye management in either maintaining local genetic variation or initiating informed translocation programs to maintain population abundance in the face of future environmental change.

[3] This chapter is in preparation for publication, but has not been published. Matt J Thorstensen conducted the analyses and wrote the original manuscript. Matt J Thorstensen, Jennifer D Jeffrey, Ken M Jeffries, Jason R Treberg, Douglas A Watkinson, and Eva C Enders conceived of the study. Yasuhiro Kobayashi collected Cedar Bluff Reservoir samples and contributed to interpreting results. Peter T Euclide and Wesley A Larson developed the sequencing method used and contributed to interpreting results. All authors participated in editing the manuscript.

## 4.2 Introduction

Intrapopulation adaptive variation is rarely considered in conservation genomics. However, within-population adaptive variation allows for adaptive differentiation to be considered as a property intrinsic to a population and not defined by relationships between populations. For species of conservation concern, local adaptation may be relevant because of its relationship with fitness. In several populations, adaptive variation is common, even amidst low neutral variation (Mable, 2019). While putatively neutral variation may be a suitable proxy for adaptive variation in species native to their ranges (Fernandez-Fournier et al., 2021), prior management action such as stocking in freshwater fish may have impacted the adaptive potential of existing variation while possibly even increasing neutral variation (e.g., gene swamping (Lenormand, 2002)). Measuring adaptive variation directly may therefore be useful for conservation or local adaptation-oriented research questions, especially in previously-translocated species. However, the identification of adaptive loci in wild populations can be a challenge, as limitations in sample sizes, prior information, and experimental approaches available diminish the certainty researchers have in characterizing adaptive markers. Moreover, commonly used outlier-based approaches such as principal components analysis (PCA) or $F_{ST}$ outliers (i.e., outliers differentiated between populations) can only identify loci that are different and presumably adaptive between populations (Funk et al., 2012; Luu et al., 2017), as opposed to identifying the population in which selection has acted and in which population a locus is adaptive.

One method with potential for describing adaptive variation in wildlife biology is cross-population extended haplotype homozygosity (XP-EHH), which can be used to identify in which population selection has acted and can identify possible selective sweeps (Sabeti et al., 2007). First, differences in haplotype frequencies between populations are calculated, then possible haplotypes that have undergone recent selection within a population are characterized as unusually long haplotypes that are at or near fixation in one population but not the other (Sabeti et al., 2007). Haplotype length is strongest for scans of recent selection as recombination breaks down haplotype blocks over time (Sabeti et al., 2006), such as their successful application in describing recent parallel adaptations to freshwater habitat in sticklebacks (*Gasterosteus*

*aculeatus*) (Hohenlohe et al., 2012). In addition, the ability for XP-EHH's to identify recent selection on derived alleles is a benefit to selection scans in populations recently diverged, such as fish fauna in North America that dispersed following deglaciation within the past fifteen thousand years (Rempel & Smith, 1998). With native ranges that often span multiple watersheds, North American freshwater fishes may be effective systems to test predictions associated with within-population or local adaptation. The ecology across the range of fishes can guide hypothesis testing and interpretation of adaptive variation identified in multiple waterbodies. By identifying parallel patterns of signatures of selection in multiple waterbodies, the probability that evolutionary processes other than adaptation could explain the pattern is reduced than when viewed as unique pieces of evidence in individual locations (Magalhaes et al., 2021; McGee et al., 2016; Smith, 2016; Soria-Carrasco et al., 2014).

Walleye (*Sander vitreus*, although see Bruner (2021)) is a freshwater fish with a native range from the Northwest Territories (Canada) to Alabama (USA) (Hartman, 2009), and a significant commercial and recreational fishery (Fisheries and Oceans Canada, 2021). Northern populations of walleye (such as those in Manitoba, Canada) are hypothesized to have been recently isolated into smaller lakes following the drainage of glacial Lake Agassiz approximately 7,000 years ago (Rempel & Smith, 1998; Stepien et al., 2009). This event may have resulted in walleye populations that remain closely related at neutral markers, but exhibit local, within-population adaptive divergence. Evaluation of native ancestry and population structure in walleye, however, has been complicated by current and historical stocking programs (Bootsma et al., 2020). Further, walleye cannibalism in captivity has made it difficult to use experimental approaches to determine phenotypes associated with local adaptation (Cuff, 1980). To address these research challenges, methodological advances in demographic history reconstruction and signatures of selection allow for the identification of candidate genomic regions of selection in the context of recent demographic histories (Gutenkunst et al., 2009; Sabeti et al., 2007). The identification of adaptive markers in the context of demographic history may inform programs of assisted gene flow for adaptive introgression in response to changing environments, or provide evidence for maintaining diverse populations overall using the precautionary principle (Aitken & Whitlock, 2013; Funk et al., 2012; Hamilton & Miller, 2016; Persson, 2016). Sustainable management for walleye in changing environments may therefore be usefully informed by analyses and characterization of adaptive markers in the context of demographic history.

Intensive stocking across North America complicates the study of native walleye ancestry, and thus signatures of selection (Bootsma et al., 2020). Genetic homogenization in walleye of Minnesota and Wisconsin (USA) was likely the result of stocking (Bootsma et al., 2020), which may have ramifications for homogenization of stocked walleye populations across North America (Kerr, 2008). For example, walleye were stocked in Lake Winnipeg from Lake Manitoba (Manitoba, Canada) between 1917-2002, with a cumulative 90,015,000 recorded walleye stocked in this way (Fig 4.1; Manitoba Government, 2020). While analyses of population structure may reveal the descendants of stocked individuals in Lake Winnipeg, interpretation of ancestry results is complicated by the shared river connecting these lakes—Lake Manitoba drains into Lake Winnipeg via the Fairford River, Lake St. Martin, and the Dauphin River (Fig 4.2). Moreover, historical flooding in 1882, 1902, 1904, 2011, and 2014 brought waterflow from Lake Manitoba toward Lake Winnipeg (Ahmari et al., 2016), and such flooding could have contributed to fish movement between the waterbodies. As such, several routes of walleye gene flow are possible between these connected lakes. Gene flow may be predicted to be low because of different environmental characteristics creating a migration barrier between Lake Manitoba and Lake Winnipeg (e.g., salinity; Environment Canada, 2011), but because walleye is a mobile species that undergoes broad seasonal migrations, historic gene flow may have been extensive between these waterbodies (Munaweera Arachchilage et al., 2021; Turner et al., 2021).

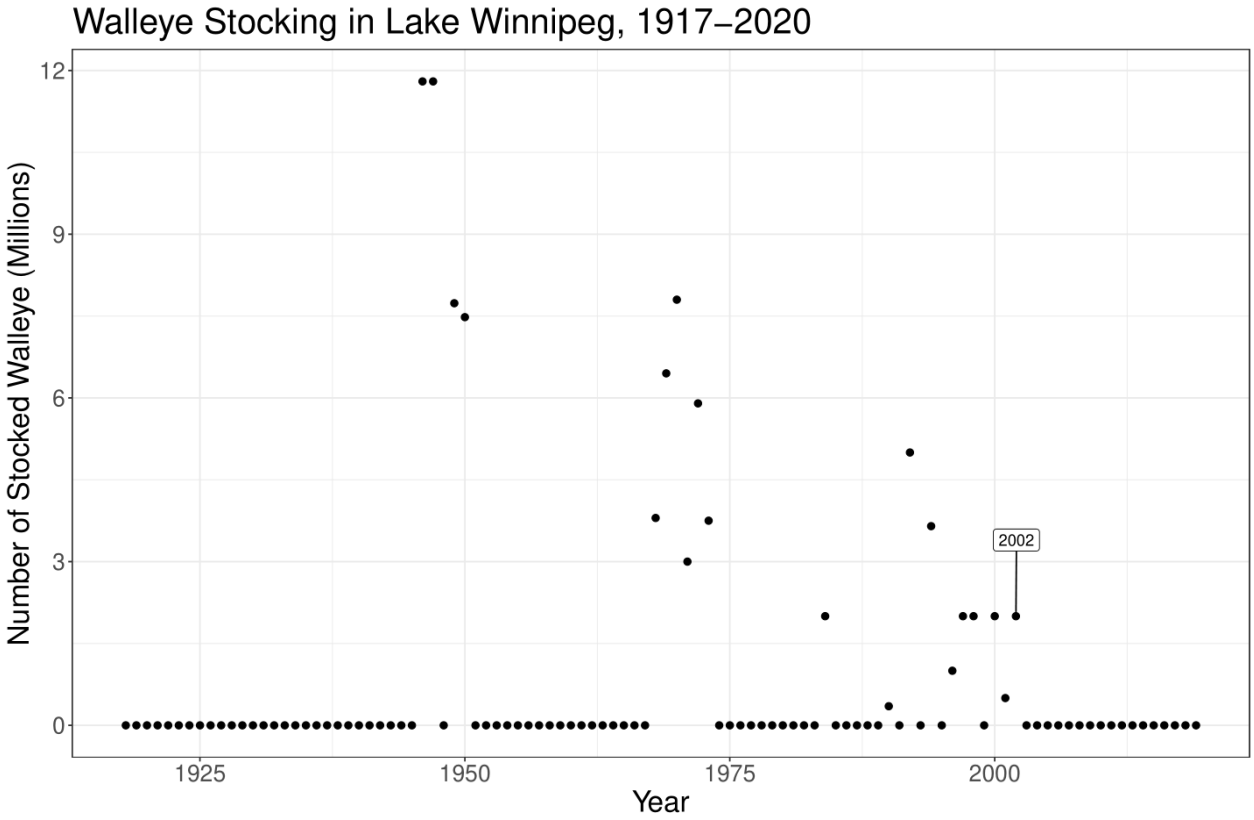


**Fig 4.1** Walleye (*Sander vitreus*) stocking in Lake Winnipeg from Lake Manitoba (Manitoba, Canada) between 1917 through 2020. Stocked fish were fry introduced into the narrows of Lake Winnipeg. Counts of 0 walleye stocked in a year represent years with no walleye stocking conducted. Data accessed from Lake Information for Anglers, Province of Manitoba, November 21, 2020. URL: https://experience.arcgis.com/experience/2557cda82dcc4a348fbb71304cedcf6d/page/page_2/

Another way complex stocking histories likely affect current patterns of genetic variation are in completely introduced populations of walleye. For example, Cedar Bluff Reservoir in Kansas, USA was created in 1951 as a flood control and irrigation resource. Walleye stocking first occurred in 1953 and continued with the establishment of a nearby fish hatchery in 1959 (Simonds, 1998). The origin of Cedar Bluff walleye is unknown and may possibly be from a variety of sources across the United States. Therefore, I cannot assume adaptation to conditions specifically within Cedar Bluff. However, the walleye from Canadian lakes do come from more northern latitudes than any US source population for Cedar Bluff; I reason that signatures of selection unique to both Manitoba lakes compared to Cedar Bluff walleye may thus be informative for potential adaptations to northern conditions. As such, this Cedar Bluff population represents a non-native outgroup (and therefore possibly unadapted group) relative to walleye adaptive variation in populations native to the Canadian lakes.

To study how the complex demographic histories of walleye have shaped modern populations, and possible adaptation to colder, northern lakes, I studied fish from Lake Winnipeg and Lake Manitoba in Manitoba and compared them with walleye from Cedar Bluff Reservoir in Kansas (Fig 4.2). A restriction-associated DNA sequencing (RAD) approach, RAD-Capture (Rapture), was employed to genotype $n$=345 individuals from the three waterbodies, for which population structure and demographic history were explored in conjunction with signatures of selection. I hypothesized that despite recent divergence, opportunities for gene flow, and historical stocking programs, signatures of selection will remain in walleye native to their waterbodies because those signatures of selection may be in regions of the genome contributing to local adaptation, and thus population persistence. Signatures of selection may thus represent candidate genomic regions of adaptive importance to the species.

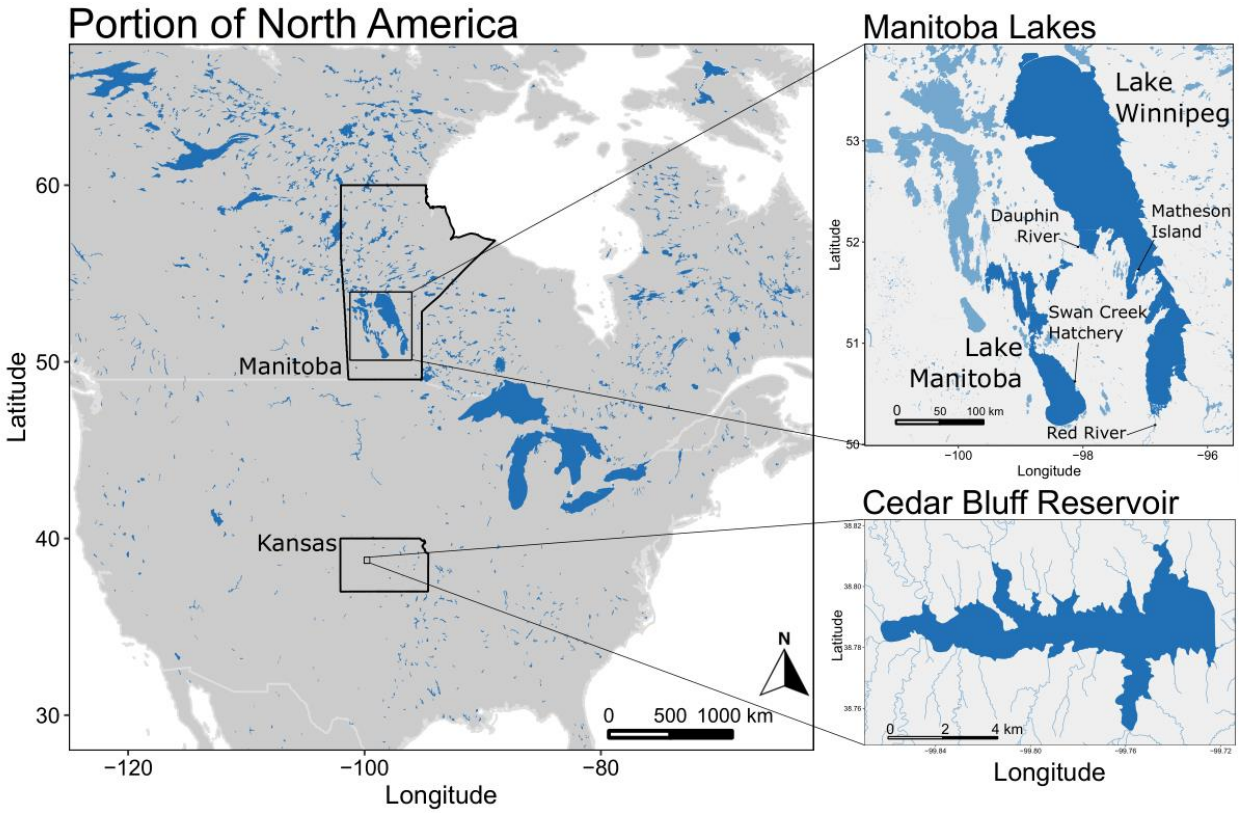


**Fig 4.2** Map of sampled waterbodies. Cedar Bluff Reservoir, Kansas, USA, represents walleye (*Sander vitreus*) in an entirely stocked population of unknown origin. Lake Manitoba and Lake Winnipeg (Manitoba, Canada) walleye represent native populations, although Lake Winnipeg has been stocked by Lake Manitoba walleye throughout the 20$^{th}$ century. In addition, floods in 1882, 1902, 1904, 2011, and 2014 have carried water from Lake Manitoba to Lake Winnipeg, and may have contributed to walleye gene flow.

## 4.3 Methods

### 4.3.1 Sample Collection

Walleye samples from Lake Winnipeg individuals were collected from spawning sites in the Red River, Matheson Island, and Dauphin River, representing the south basin, channel, and north basin, respectively (*n*=129, 51, and 55 respectively; Chapters 2 and 3 describe sampling as well). Lake Winnipeg fish > 1.2 kg were collected by boat electrofishing in Spring 2017 and 2018, and fin clips were sampled non-lethally after. Prior to sampling, a Portable Electroanesthesia System (Smith Root) was used to anesthetize fish following approved animal use protocols of Fisheries and Oceans Canada (FWI-ACC-2017-001, FWI-ACC-2018-001), the University of Manitoba (F2018-019), and the University of Nebraska-Lincoln (Project ID: 1208). Fin clips were kept in 95% ethanol for storage. For Lake Manitoba fish, *n*=50 fry from the 2019-year class were haphazardly sampled at Swan Creek Hatchery and were stored in 95% ethanol. Procedures for sample collection at Cedar Bluff Reservoir were approved by the Fort Hays State University Institutional Animal Care and Use Committee (Protocol # 19-0023). Anal fin clips were taken from 60 sexually mature walleye (n=40 females and n=20 males) during the gamete collection in March 2019. In total, *n*=345 walleye were genotyped for the present study.

### 4.3.2 DNA Extraction, Sequencing, and SNP Calling

DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit (QIAGEN, Venlo, Netherlands) following manufacturer protocols. RAD-seq library preparation was performed with the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, Massachusetts, USA) with the *PstI* restriction enzyme. BestRAD libraries were baited using protocols outlined in Euclide et al., (2021) using a RAD-Capture panel for a different walleye research project in the Laurentian Great Lakes that is still ongoing at the time of publication. The panel includes a single bait for 99,636 SNP loci identified from a preliminary PSTI RAD-sequencing survey of 48 walleye collected in the Great Lakes (*n* = 9 – 10 from each of Lakes Superior, Huron, Michigan, Erie, and Ontario; USA) designed by Arbor Biosciences (Ann Arbor, Michigan, USA). Target loci were selected based on minor allele frequency (>0.01) and

alignment position along a draft walleye genome to retain one SNP approximately every 5000 bp along every contig greater than 0.1 megabases. Bait-captured libraries were then sequenced on a S4 NovaSeq 6000 (Illumina, San Diego, California, USA) at NovoGene (Sacramento, California, USA).

Raw reads were processed with STACKS v2.3 (Catchen et al., 2013; Catchen et al., 2011) using settings in Euclide et al., (2021), with the exception that the yellow perch (*Perca flavescens*) reference genome was used to align reads with bwa v0.7.17 with default settings (PLFA_1.0 at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) BioProject PRJNA514308) (Feron et al., 2020; Li & Durbin, 2009) instead of the loci developed in Euclide et al. (2021). The closely-related yellow perch's reference genome was used because I sought to link SNPs with annotated genes and an annotated walleye reference genome was not available at the time. In addition, the yellow perch reference genome was assembled at the chromosome-level, and the yellow perch and walleye share karyotypes (Bootsma et al., 2020; Danzmann, 1979). Loci were processed with *process_radtags* (-e pstI -i gzfastq -c -q -r --filter_illumina --bestrad -t 140), aligned with bwa, genotyped in *gstacks*, and SNPs were called with *populations* with one random SNP called per locus, a minor allele count ≥2, a ≥60% genotyping rate per site collected, and a ≥60% genotyping rate overall (--min-samples-per-pop 0.6 --min-samples-overall 0.6 --min-mac 2 --write-random-snp). Singletons were excluded from these analyses following best practices for detecting population structure with reduced representation data (Linck & Battey, 2019). The SNPs were filtered for paralogs using the heterozygosity and read ratio deviation method (HDPlot) where SNPs were accepted as non-paralogs with heterozygosity ≤0.60 and read ratio deviation ≤|7| (McKinney et al., 2017). For population structure and demographic reconstruction, the SNP dataset was pruned for linkage with PLINK v1.9 in 50 SNP windows shifted by 10 SNPs per step, filtered at an $r^2$ threshold of 0.1 (--indep-pairwise 50 10 0.1) (Purcell et al., 2007). XP-EHH requires complete data, while phasing vastly increases statistical power in the approach (Gautier et al., 2017). Therefore, he overall SNP dataset was filtered to only 10% missing data. Beagle v5.1 was used to phase and impute unpruned SNPs because of its robust performance on a variety of datasets and observed 95% imputation accuracy with as low as 40% missing data (Browning et al., 2018; Browning & Browning, 2007; Weng et al., 2013; Yang et al., 2014). Because relatedness can introduce bias in association studies, pairwise relatedness was checked using the method of

moments as implemented by PLINK in SNPRelate v1.22 (Zheng et al., 2012). Where appropriate, VCFtools v0.1.16 was used to filter or subset variant call format files, while PGDSpider v2.1.5 was used to convert data between formats (Danecek et al., 2011; Lischer & Excoffier, 2012). The statistical computing environment R v4.0.3, along with the packages Tidyverse v1.3.0 and vcfR v1.12.0, were used throughout these analyses (Knaus & Grünwald, 2017; R Core Team, 2021; Wickham et al., 2019).

To further investigate predicted synteny between the walleye and yellow perch genomes, synteny was analyzed using progressiveMauve and i-ADHoRe following a pipeline published in Doerr & Moret (2018), with details provided in the supplementary materials (yellow perch genome, PLFA_1.0 at NCBI SRA BioProject PRJNA514308; walleye genome, ASM919308v1 at NCBI SRA BioProject PRJNA528354) (Darling et al., 2010; Ghiurcuta & Moret, 2014; Krzywinski et al., 2009; Proost et al., 2012). Weighted and relaxed synteny score distributions were non-normal, thus confidence intervals were calculated with bootstrapping 100 scores over 10,000 iterations, from which 95% confidence intervals were calculated using the percentile method.

### 4.3.3 Population Genetics

Admixture v1.3.0 was run over K values of one through six, with 10-fold cross validation and 1000 bootstrap replicates for parameter standard errors (Alexander et al., 2009). Pophelper v2.3.0 in R was used to visualize and organize Admixture results (Francis, 2017) (Francis et al., 2016). For population assignments, individuals were considered assigned to a cluster when their Q-values were >0.85 for that cluster. Only assigned individuals were used for population differentiation and demographic reconstruction, while all individuals were used for outlier tests.

Hierfstat v0.5-7 (Goudet, 2005) was used to find $\beta_{WT}$ as a measure of population-specific differentiation from the entire pool and Weir & Cockerham's pairwise $F_{ST}$ between populations (Weir & Cockerham, 1984; Weir & Goudet, 2017). 95% confidence intervals were generated for $F_{ST}$ and $\beta_{WT}$ over 1,000 bootstrapped iterations each. In addition, observed heterozygosity ($H_O$), gene diversity ($H_S$), and inbreeding coefficients ($F_{IS}$, 95% confidence intervals generated over 1,000 bootstrapped iterations) for each population were calculated. Tajima's D for each of the

three assigned populations was calculated using δaδi v2.1.1, where singleton SNPs were masked from each site frequency spectrum because a minor allele count filter of ≥2 was previously applied to the data, and projections of 60, 80, and 200 for Cedar Bluff Reservoir, Lake Manitoba, and Lake Winnipeg-assigned individuals, respectively.

NeEstimator v2.1 (Do et al., 2014) was used to study effective population size in 95% confidence intervals for each population using the linkage disequilibrium method. Here, datasets were filtered for no missing data per population using population assignments with vcftools, pulled from the pruned SNP dataset.

### 4.3.4 Demographic History

The diffusion approximation-based program δaδi v2.1.1 was used to explore models of separation, population size change, and migration between the Lake Manitoba and Lake Winnipeg-assigned individuals (Gutenkunst et al., 2009). Because other analyses used a minor allele count of ≥2, singleton SNPs were masked from δaδi analysis. Site frequency spectra generated with δaδi were projected down to 80 and 200 for Lake Manitoba and Lake Winnipeg individuals, respectively, and 210, 220, and 230 grid points were used in extrapolations. Two-population models were chosen from dadi_pipeline v3.1.5 over consecutive rounds of optimizations using default settings (Portik et al., 2017; see also https://github.com/dportik/dadi_pipeline). For each model, 10, 20, 30, and 40 replicates were run over 3, 5, 10, and 15 iterations and 3, 2, 2, and 1 folds, with the highest log-likelihood parameter estimates seeding the parameter starting values for a subsequent run. These values were based on default settings identified to provide an effective trade-off between computational time, accuracy, and precision of runs (https://github.com/dportik/dadi_pipeline). Because of signatures of admixture between fish from Lakes Manitoba and Winnipeg may have possibly been introduced from translocated fish, I chose models based on varying levels of migration on ‘unadmixed’ individuals. The models tested were that of no migration, no migration with one size change step allowed, no migration with two size change steps allowed, symmetrical migration, asymmetrical migration, asymmetrical migration with a population size change, secondary contact with asymmetrical migration, and secondary contact with asymmetrical migration and a population size change (no_mig, no_mig_size, no_contact_size2, sym_mig,

asym_mig, asym_mig_size, sec_contact_asym_mig, and sec_contact_asym_mig_size). In addition, several custom models based on the model with secondary contact with asymmetrical migration and a size change were created because initial tests showed improvements in model fits by including asymmetric migration and size changes independently. The modifications included a second size change step added to the secondary contact model, and rates of asymmetric migration were allowed to 1) exist continuously between populations after secondary contact, 2) vary between time periods of secondary contact and the last size change step, then between the last size change and the present day, and 3) only exist after the last size change step (sec_contact_asym_mig_size2, sec_contact_varying_asym_mig_size2, and sec_contact_late_asym_mig_size2). Akaike Information Criteria (AIC) and negative log-likelihoods were used to compare models.

For the best fit model, parameter confidence intervals were found using 2 megabase genomic chunks over 1,000 bootstrapped iterations, where a Godambe Information Matrix was developed using model settings and optimized parameters to find uncertainty in model parameters (Coffman et al., 2016). Because biological interpretations of optimized parameters depended on $\theta$ in addition to specific model parameters (e.g., $\theta$ and timepoint T, or $\theta$ and population size $N_{1a}$) uncertainties were propagated through the different model parameters (Ku, 1966). A mutation rate $\mu$ of $3.28 \times 10^{-9}$ was used, based on divergence between the blackfin icefish (*Chaenocephalus aceratus*) and dragonfish (*Parachaenichthys charcoti*) (Kim et al., 2019). This mutation rate was chosen based on available mutation rates in the literature, the sequencing method used in Kim et al. (2019) when developing the mutation rate estimate, and the relative evolutionary relatedness between walleye, blackfin icefish, and dragonfish. A walleye generation interval of 4.3 years was used based on results from Franckowiak et al. (2009), which in turn was found using methods for estimating generation times for organisms with overlapping generations in Hill (1979).

### 4.3.5 Linkage Disequilibrium Decay

To test if founder effects led to increased LD in walleye from Cedar Bluff Reservoir and explore interactions between linkage and admixture in the other waterbodies (Reich et al., 2001), LD decay was explored using PopLDdecay v3.40 (Zhang et al., 2019). Pairwise $r^2$ was identified

between SNPs from the unimputed, un-pruned dataset, with separate runs for each of the Lake Winnipeg, Lake Manitoba, and Kansas-assigned individuals, along with unassigned individuals which I suspected to be admixed between Lake Manitoba and Lake Winnipeg walleye. Within each population run, SNPs were only included for LD decay analysis with a maximum of 10% missing data and minor allele frequency ≥ 0.05 because rare alleles lead to increased variance in $r^2$ (Remington et al., 2001). A nonlinear least-squares approach was used to estimate expected $r^2$ between SNPs (see equation 1 in Remington et al., 2001, based on Hill & Weir, 1988). From the nonlinear regression, the distances' half-decay where estimated LD fell to half of its maximum estimated value, and of moderate LD where $r^2 \leq 0.20$, were estimated for each population. For visualization, LD curves were plotted up to 250 kb pairwise SNP distance.

### 4.3.6 Signatures of Selection

Signatures of selection where one genomic region has approached or reached fixation in one population compared to another, were explored with XP-EHH using the R package rehh v3.2.1, with the markers previously described as phased and imputed after filtering for 90% present data (Gautier et al., 2017; Sabeti et al., 2007). Data were unpolarized for analyzing XP-EHH because ancestral and derived alleles were unknown. Individuals with phased and imputed SNPs were split into the three assigned populations using Admixture results (Q>0.85 per individual), and extended haplotype homozygosity (EHH) was analyzed for each population separately (scan_hh). Three pairwise comparisons between each population were used to identify XP-EHH using false discovery rate corrected *p*-values (*q*-values). Significant candidate regions of selection were identified by analyzing 100 kilobase (kb) windows overlapping by 10 kb, in which at least three significant SNPs ($q < 0.05$) showing XP-EHH were found. For visualization, only significant SNPs within candidate regions were highlighted (thus a significant SNP outside a candidate region would be de-emphasized), using the R package ggman v0.99.0 with relative SNP positions to reflect distances between points in the reduced representation data (https://rdrr.io/github/veera-dr/ggman/). Results for XP-EHH were divided into three pairwise tests between each assigned population (Lake Winnipeg, Lake Manitoba, and Cedar Bluff Reservoir), and significant results for each pairwise test were further divided into which population showed elevated XP-EHH scores (e.g., in a pairwise test between Lake Manitoba and

Lake Winnipeg, certain SNPs were significant among haplotypes in Lake Manitoba versus others significant among haplotypes in Lake Winnipeg).

PCAdapt v4.3.3 was used to explore potential signatures of local adaptation with outlier tests (Luu et al., 2017; Privé et al., 2020). Here, a scree plot showed that the optimal choice for principal components (PC) was K=2 based on Cattell's rule, an observation confirmed by a lack of discernible populations structure in K>2 principal components (Cattell, 1966) (Fig A.3.1). Significance for individual SNPs was determined using *q* values implemented in the R package qvalue v2.20.0, where a SNP was accepted as significant at $q < 0.05$ (Storey et al., 2020). The PC associated with a significant SNP was retrieved with *get.pc*, which enabled us to link significant SNPs with different axes of population structure. PCAdapt results were separated into datasets of those significant along PC1 or PC2, corresponding to latitude and longitude, respectively.

Genomic regions and SNPs under selection were identified with XP-EHH, while PCAdapt was used to analyze the PC a SNP was most strongly contributing to and verify that imputed data were not driving significant patterns. Significant SNPs were analyzed for their proximity to genes in the yellow perch genome with Magma v1.08b in windows 15 kb upstream and 5 kb downstream of each gene with data flagged as non-human (--annotate window=15,5 nonhuman) (de Leeuw et al., 2015). For associating candidate regions and SNPs with genes, a conservative 20 kb window around a gene (15 kb upstream, 5 kb downstream) was used in part to limit the possibility that a SNP would be erroneously associated with a gene by limiting the chance that a structural rearrangement between the walleye and yellow perch genomes used may have broken synteny between genomes at those sites (Feron et al., 2020). In addition, synteny was consistently observed between each genome in the present analyses, and the walleye and yellow perch have previously been shown to share karyotypes, providing additional support for using the yellow perch genome to analyze signatures of selection in walleye (Danzmann, 1979). Genes associated with significant SNPs were then compared in the context of their representation in different significant SNP datasets and visualized with the R package UpSetR v1.4 (Conway et al., 2017).

## 4.4 Results

### 4.4.1 DNA Extraction, Sequencing, and SNP Calling

96,955 SNPs were called in STACKS, of which 63,882 were accepted as non-paralogous SNPs after using HDPlot. Pruning non-paralogous SNPs for linkage with PLINK left 46,342 SNPs for use for population structure and demographic reconstruction. The dataset of 63,882 SNPs was filtered for a maximum of 10% missing data and imputed and phased into a dataset of 18,601 SNPs for use with XP-EHH. Synteny between the walleye and yellow perch genomes was high, indicating at least a large degree of concordance between the two genomes (mean weighted synteny score = 0.999, 95% CI [0.997, 1.0]; mean relaxed synteny score = 0.999, 95% CI [0.997, 1.0]) (Fig A.3.2).

### 4.4.2 Population Genetics

Admixture identified K=2 populations based on the lowest cross validation error between K=1 through 6, separating Manitoban and Kansas populations of walleye (Fig A.3.3). However, cross validation error was similar at K=3 and population structure delineated the three different water bodies used in the present study; individuals were thus assigned to populations based on K=3. Using ancestry coefficients Q > 0.85 for population assignments, 189 individuals were assigned to a Lake Winnipeg population, 67 individuals to a Lake Manitoba population, 60 to a Kansas population, and 29 to no population. Notably, 27 out of the 29 unassigned individuals were found in Lake Winnipeg (the 2 remaining in Lake Manitoba), and of those, 18 were caught in the northern Dauphin River, 6 in the central Matheson Channel, and 2 in the southern Red River. In addition, out of the 67 individuals assigned to the Lake Manitoba population, 20 were sampled from the Dauphin River in Lake Winnipeg.

Pairwise $F_{ST}$ showed the greatest differentiation between each Canadian lake and Cedar Bluff Reservoir, and moderate differentiation between Lake Manitoba and Lake Winnipeg (Fig 4.3). Population-specific differentiation from the overall pool, $\beta_{WT}$, was lowest for Cedar Bluff Reservoir (Fig 4.3). Filtering for complete data for NeEstimator v2 from the pruned SNP dataset left 812 SNPs available in with the Lake Winnipeg population, 2,768 SNPs in the Lake Manitoba

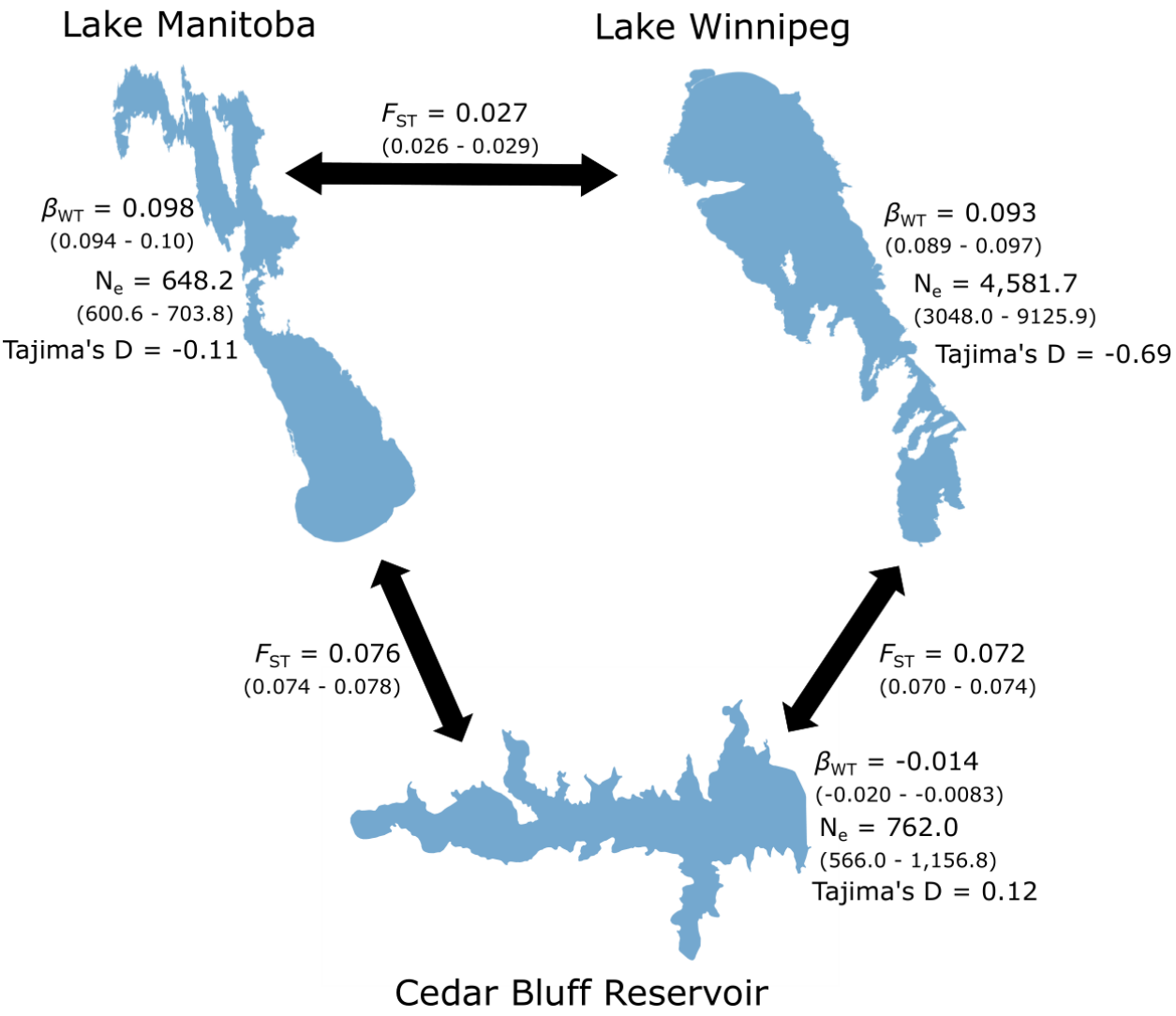


**Fig 4.3** Population structure, effective population size, and Tajima's D for walleye (*Sander vitreus*) in three North American waterbodies. Cedar Bluff Reservoir (Kansas, USA) represents an entirely stocked population of walleye, while Lake Manitoba and Lake Winnipeg (Manitoba, Canada) walleye represent populations with possible gene flow. $F_{ST}$ represents Weir & Cockerham's pairwise $F_{ST}$, while $\beta_{WT}$ represents differentiation from the overall genetic pool, each found using Hierfstat. Effective population size for each population is represented by $N_e$, found using NeEstimator v2. Tajima's D was calculated with δaδi. For $F_{ST}$, $\beta_{WT}$, and $N_e$, 95% confidence intervals are provided in parentheses.

population, and 756 SNPs in the Cedar Bluff Reservoir population. Tajima's D was negative for Lake Winnipeg fish, indicating a possible population expansion, and positive for Cedar Bluff Reservoir fish, indicating a possible population contraction (Fig 4.3). Linkage disequilibrium-based $N_e$ was highest for the Lake Winnipeg population, and approximately 6.5x lower in each the Lake Manitoba and Cedar Bluff Reservoir populations (Fig 4.3). Observed heterozygosity, gene diversity, and inbreeding coefficient ($H_O$, $H_S$, and $F_{IS}$) were each highest for the Cedar Bluff Reservoir population, with $F_{IS}$ approximately two times higher for Cedar Bluff walleye than the walleye in the northern lakes (Fig A.3.4).

### 4.4.3 Demographic History

The most likely demographic model was one testing for a divergence between the Lake Manitoba and Lake Winnipeg populations, a period of no contact, a size change, another period of no contact, a second size change, then asymmetric migration (sec_contact_late_asym_mig_size2 in Table A.3.1; Fig A.3.5). Because singleton SNPs are informative for fitting models with δaδi, confidence intervals for parameter estimates were generally large since the informative singletons were masked from analysis. Nevertheless, δaδi provided additional evidence of a higher current $N_e$ for Lake Winnipeg walleye than Lake Manitoba (Lake Winnipeg $N_e$=25,782.94, 95% CI [20,870.4, 30,695.49], Lake Manitoba $N_e$=1,780.41, 95% CI [-1,027.95, 4,588.78]), and identified possible gene flow starting approximately 884.46 years before present 95% CI [549.82, 1219.10] (Table 4.1).

### 4.4.4 Linkage Disequilibrium Decay

Linkage disequilibrium was highest for the unassigned fish across all expected $r^2$ values fit up to one million base pairs between SNPs, although Cedar Bluff Reservoir showed higher long-range LD than the Lake Winnipeg fish, despite similar LD below approximately 25 kb (Fig 4.4). For Cedar Bluff Reservoir, half-decay was at about 1.9 kb, while moderate LD was at approximately 2.6 kb. Whereas for Lake Winnipeg, half-decay was at approximately 2.4 kb, while moderate LD ($r^2 \leq 0.20$) was at approximately 2.8 kb. For Lake Manitoba, half-decay was at approximately 4.2 kb, while moderate LD was at approximately 5.7 kb. And for unassigned

**Table 4.1** Results from δaδi for the best-fit model tested in demographic history for Lake Manitoba and Lake Winnipeg (Manitoba, Canada) walleye populations. This model represents a population split prior to timepoint one, population size changes at timepoints 1, 2, and 3, and secondary contact after timepoint 3. Migration rates (i.e., effective dispersal) are modeled starting at timepoint 3.

| Model Term | Mean Estimate | Lower Bound | Upper Bound |
|---|---|---|---|
| Lake Manitoba Size 1 | 2106.67 | -16740.87 | 20954.21 |
| Lake Manitoba Size 2 | 13077.74 | 10109.82 | 16045.65 |
| Lake Manitoba Size 3 | 1780.41 | -1027.95 | 4588.78 |
| Lake Winnipeg Size 1 | 5452.95 | -158895.44 | 169801.33 |
| Lake Winnipeg Size 2 | 6855.87 | -29199.42 | 42911.17 |
| Lake Winnipeg Size 3 | 25782.94 | 20870.4 | 30695.49 |
| Migration from Lake Winnipeg to Lake Manitoba starting at Timepoint 3 | 0.00014 | -0.0016 | 0.0019 |
| Migration from Lake Manitoba to Lake Winnipeg starting at Timepoint 3 | 0.00036 | -0.0020 | 0.0027 |
| Timepoint 1 | 4286.65 | -468600.44 | 477173.74 |
| Timepoint 2 | 71989.74 | 46393.36 | 97586.11 |
| Timepoint 3 | 884.46 | 549.82 | 1219.10 |

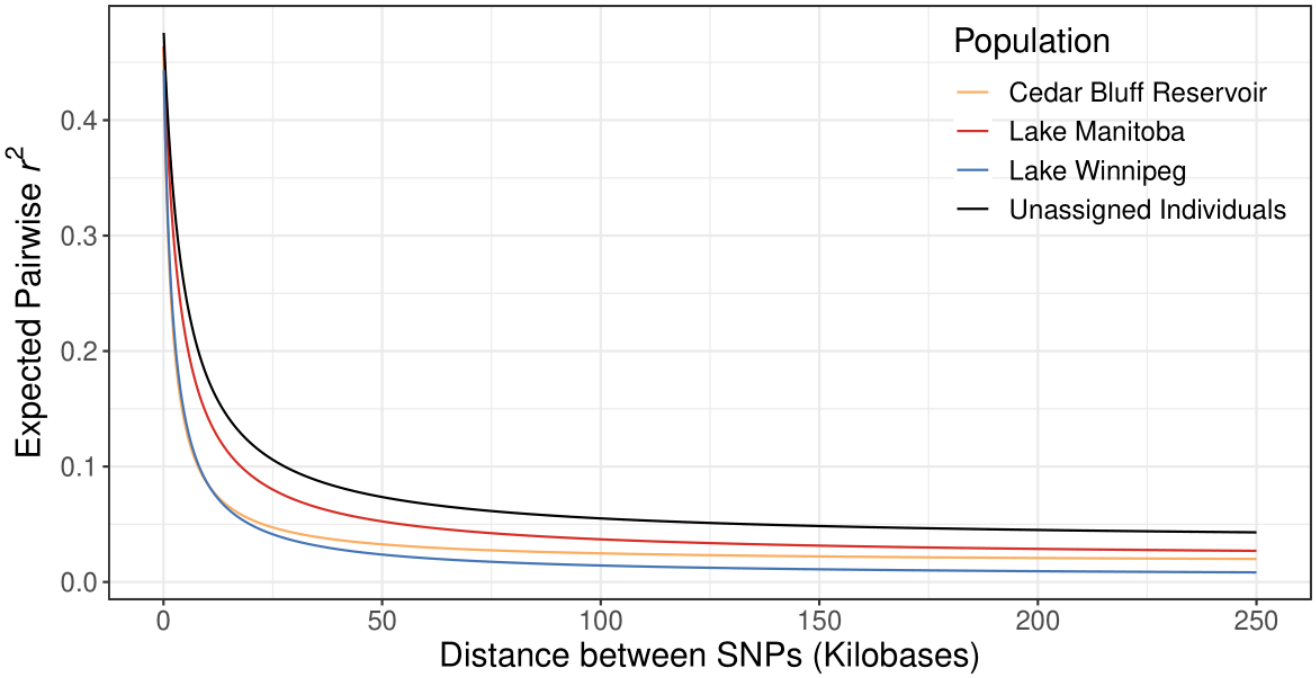


**Fig 4.4** Linkage disequilibrium decay for three populations of walleye (*Sander vitreus*) analyzed in the present study, in addition to unassigned individuals that may have been admixed between Lake Manitoba and Lake Winnipeg individuals. Cedar Bluff Reservoir (Kansas, USA) walleye represent an entirely stocked population of fish of unknown origin, while Lake Manitoba and Lake Winnipeg (Manitoba, Canada) walleye represent native populations with possible gene flow. Unassigned individuals were largely found in the Dauphin River (north basin) of Lake Winnipeg ($n$=18 out of 29 total unassigned individuals). Pairwise $r^2$ was estimated between single nucleotide polymorphisms up to 1,000 kilobase pairs apart with PopLDdecay. A nonlinear regression was used to fit curves for expected pairwise $r^2$. For visualization, linkage disequilibrium curves were truncated to pairwise distances of 250 kilobase distances.

individuals, half-decay was at approximately 5.6 kb, while moderate LD was at about 8.0 kb. Individual pairwise $r^2$, distance, and full nonlinear regressions (not truncated to 250 kb) are visualized for each population in Fig A.3.6.

### 4.4.5 Signatures of Selection

The XP-EHH scores between Lake Winnipeg- and Lake Manitoba-assigned individuals identified 45 SNPs with elevated XP-EHH scores (15 higher in Lake Winnipeg, 30 higher in Lake Manitoba), of which 33 SNPs were in candidate regions under selection (11 with higher XP-EHH scores in Lake Winnipeg, 22 with higher XP-EHH scores in Lake Manitoba) (Fig 4.5). The XP-EHH scores between Lake Winnipeg- and Cedar Bluff Reservoir-assigned individuals identified 26 SNPs with elevated XP-EHH scores (24 higher in Lake Winnipeg, 2 higher in Cedar Bluff Reservoir), of which 18 were in candidate regions under selection (17 with higher XP-EHH scores in Lake Winnipeg, 1 with higher XP-EHH scores in Cedar Bluff Reservoir) (Fig 4.5). Between Lake Manitoba- and Cedar Bluff Reservoir-assigned individuals, 7 SNPs showed elevated XP-EHH scores (7 with higher XP-EHH scores in Lake Manitoba, 10 with higher XP-EHH scores in Cedar Bluff Reservoir), of which 5 were in candidate regions under selection (5 with higher XP-EHH scores in Lake Manitoba, 0 with higher XP-EHH scores in Cedar Bluff Reservoir) (Fig 4.5). There were 10 candidate regions under selection in XP-EHH between Lake Winnipeg and Lake Manitoba, 5 candidate regions between Lake Winnipeg and Cedar Bluff Reservoir, and 0 candidate regions between Lake Manitoba and Cedar Bluff Reservoir (Table A.3.2).

Two principal components (PCs) showed population structure in PCAdapt, with PC1 showing 24.0% variance explained by latitudinal differences between sites and PC2 showing 15.1 variance explained by longitudinal differences (Fig 4.6). 345 SNPs were identified as significant outliers by PCAdapt, of which 226 were significant along PC1 and 119 were significant along PC2 (Fig 4.6).

Genes with SNPs showing elevated XP-EHH scores 15 kb upstream of 5 kb downstream were identified with respect to the three pairwise comparisons performed. For the Lake Winnipeg-Lake Manitoba XP-EHH test, 4 genes were near signatures of selection within Lake

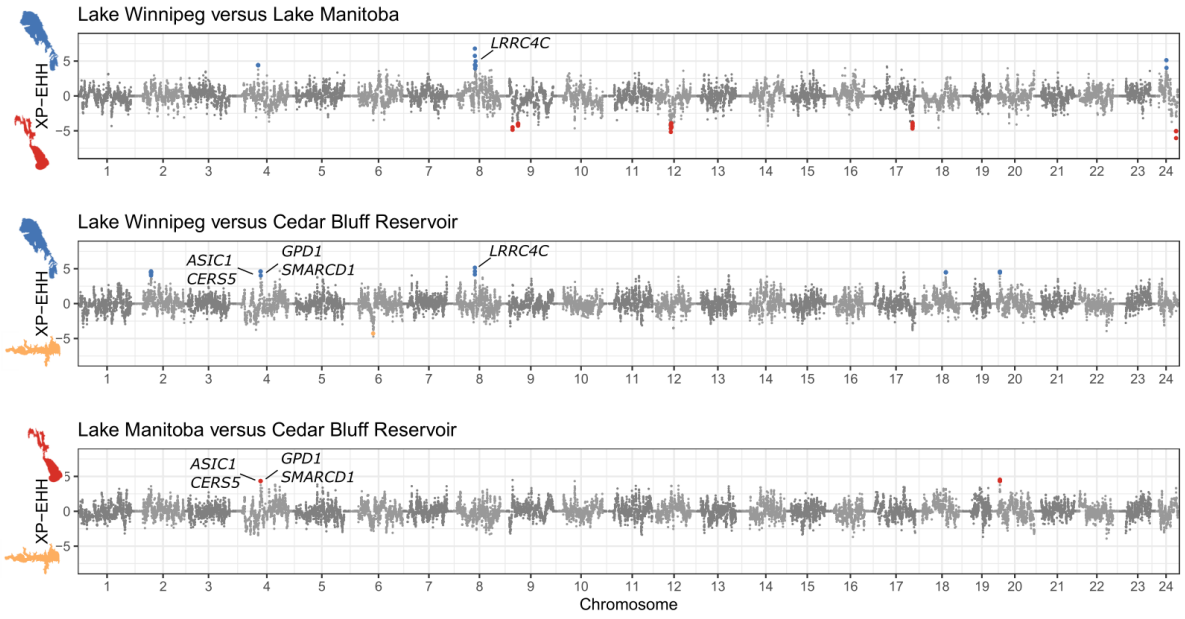


**Fig 4.5** Signatures of selection inferred from cross-population extended haplotype homozygosity (XP-EHH) between each population of walleye (*Sander vitreus*) analyzed in the present study. Cedar Bluff Reservoir (Kansas, USA) represents an entirely stocked population, while Lake Manitoba and Lake Winnipeg (Manitoba, Canada) each represent native populations with possible gene flow. Walleye DNA was aligned to the yellow perch (*Perca flavescens*) reference genome, and chromosome numbers represent yellow perch chromosomes. Unplaced scaffolds are not visualized. Single nucleotide polymorphisms (SNPs) were considered significant with $q<0.05$ and only if the SNPs were in a candidate region defined by 100 kilobase regions of the genome overlapping by 10 kilobases, in which ≥3 SNPs total were significant. Because XP-EHH focuses on haplotypes of unusual length and high frequency in one population compared to another, the population in which haplotypes are long and are at or approaching fixation can be identified. Here, the population in which XP-EHH was significant is identified with colours, where blue represents Lake Winnipeg, red represents Lake Manitoba, and yellow represents Cedar Bluff Reservoir. As a heuristic, genes were associated with significant SNPs in 20 kb windows (15 kb upstream, 5 kb downstream of a gene) using Magma. The gene *lrrc4c* was within the candidate region for selection in Lake Winnipeg on chromosome 8.

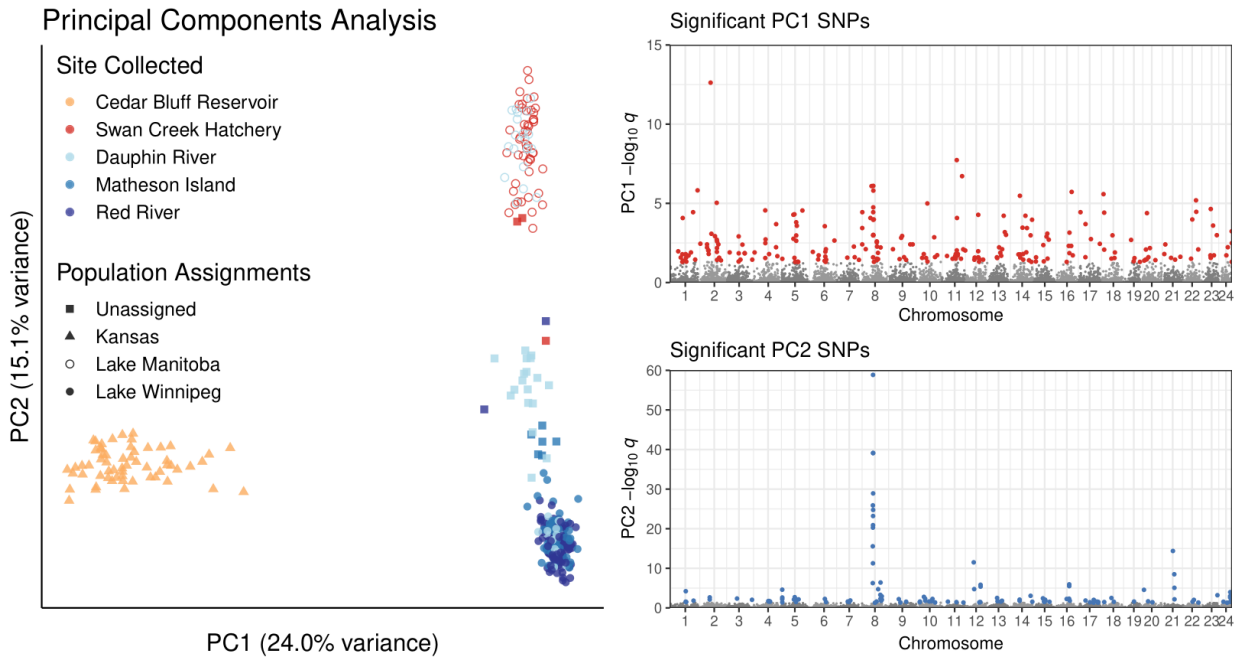


**Fig 4.6** Principal components analysis of walleye (*Sander vitreus*) from three waterbodies across North America. Cedar Bluff Reservoir (Kansas, USA) represent an entirely stocked population of walleye of unknown origin. Lake Winnipeg and Lake Manitoba (Manitoba, Canada) represent native populations of walleye with possible gene flow. Principal components analysis was done using pcadapt, while population assignments were done using Admixture with K=3 groups. Individuals were assigned to a population with Q>0.85 for a population, and individuals with maximum Q≤0.85 were considered unassigned. Significant SNPs were distinguished according to the first two principal components in which they most strongly contributed variation, where principal component 1 (PC1) corresponded to latitudinal differences among populations, while principal component 2 (PC2) was most strongly characterized by variation between Lake Winnipeg and Cedar Bluff Reservoir relative to Lake Manitoba. Only SNPs with -$\log_{10}$ *q*-values <0.05 were highlighted for visualization and retained for downstream analyses. Walleye DNA was aligned to the yellow perch (*Perca flavescens*) reference genome. Unplaced scaffolds were not visualized, and chromosome numbers refer to chromosomes in the yellow perch genome.

Winnipeg while 18 were near signatures of selection in Lake Manitoba. For the Lake Winnipeg-Cedar Bluff Reservoir XP-EHH test, 13 genes showed signatures of selection within Lake Winnipeg, and 1 was identified near signatures of selection in Cedar Bluff Reservoir. For the Lake Manitoba-Cedar Bluff Reservoir XP-EHH test, 4 genes were near signatures of selection in Lake Manitoba, and none were identified near signatures of selection in Cedar Bluff Reservoir. These genes were tested for overlap between genes 15 kb upstream and 5 kb downstream of outlier SNPs identified with PCAdapt, along with overlaps in scanned genes between XP-EHH tests (Fig A.3.7). Four genes, *acid sensing ion channel 1* (*ASIC1*), *ceramide synthase 5* (*CERS5*), *glycerol-3-phosphate dehydrogenase 1* (*GPD1*), and *SWI/SNF-related matrix-associated actin-dependent regulator of chromatin 1 subfamily D member 1* (*SMARCD1*) were associated with XP-EHH in Lake Manitoba and Lake Winnipeg with respect to Cedar Bluff Reservoir, and were associated with outlier SNPs along PC1.

## 4.5 Discussion

In the present study, I used a reduced representation approach to genotype *n*=345 walleye from three different waterbodies in North America at 46,342 genetic markers. Limited historical gene flow was identified between the Canadian lakes despite periodic stocking from Lake Manitoba to Lake Winnipeg throughout the 20$^{th}$ century and flooding between the lakes that potentially facilitated migration. In addition, I found evidence of decreased differentiation associated with stocking in Cedar Bluff Reservoir, and a large effective population size in conjunction with a recent population expansion in Lake Winnipeg walleye. By scanning for selection in both Canadian lakes, I identified signatures of selection possibly associated with adaptations to colder environments. I also found that Lake Winnipeg walleye showed unexpectedly strong signatures of selection on chromosome 8, unique within the connected Lake Manitoba-Lake Winnipeg system.

### 4.5.1 Population Genetics

I observed moderate population differentiation ($F_{ST}$) between two large lakes in Canada (Lake Winnipeg and Lake Manitoba), and comparatively stronger differentiation between the

two lakes in Canada and Cedar Bluff Reservoir, Kansas, USA. Despite these signals of differentiation, admixture analysis and population assignment identified certain individuals sampled from primarily the north basin of Lake Winnipeg as originating from Lake Manitoba. This gene flow into Lake Winnipeg may have potentially been associated with stocking programs from Lake Manitoba to Lake Winnipeg in the 1900s and early 2000s, or from floods in 2011 and 2014 (Ahmari et al., 2016). Historical gene flow was consistently predicted between the two lakes, likely via the Dauphin River and Lake St. Martin. However, the magnitude of this gene flow was small in each model. In the most likely demographic scenario, the magnitude of gene flow amounted to an upper bound of 0.0027 fish traveling from Lake Manitoba to Lake Winnipeg each year (approx. 0.00036 fish from Lake Winnipeg to Lake Manitoba per year), starting between 550 and 1,219 years ago. In addition, outlier loci and population differentiation between Lake Manitoba and Lake Winnipeg individuals supports evidence of limited historical gene flow, as extensive gene flow tends to work in contrast to local selection (Lenormand, 2002). I note that gene flow may actually promote local adaptation in certain contexts, such as by increasing standing variation or by adaptive introgression, and the resulting genetic architecture can maintain local adaptation (Tigano & Friesen, 2016). However, elevated linkage disequilibrium in the unassigned, possibly admixed individuals in Lake Winnipeg is consistent with a recent admixture event as opposed to sustained mixing (Bray et al., 2010; Reynolds & Fitzpatrick, 2013; Wilson & Goldstein, 2000). Therefore, flooding or stocked walleye from Lake Manitoba into Lake Winnipeg may have contributed to the signatures of admixture in Lake Winnipeg walleye. Taken together, I suggest that gene flow between Lakes Manitoba and Winnipeg prior to stocking programs and recent flooding has been low, and that several differentiated genomic regions suggestive of local adaptation remain in each population despite gene flow. This is a conclusion with implications for local adaptation in each water body and for each fishery's population dynamics.

The presence of Lake Manitoba-assigned walleye in the Dauphin River of Lake Winnipeg, along with several individuals likely admixed between the populations, is a potentially significant finding for stock structure and fisheries management in the future because of the possibility that each large lake may be part of a larger metapopulation. This observation is even more striking in the context of Lake Manitoba walleye being represented by individuals sampled at an individual location, Swan Creek Hatchery. The possible introgression of Lake

Manitoba walleye with Lake Winnipeg fish may increase the pool of standing genetic variation in Lake Winnipeg walleye, with possible beneficial effects for walleye responses to a changing lake ecosystem (Dibike et al., 2012). However, it may also swamp patterns of local adaptation by decreasing variation at loci under selection (Hamilton & Miller, 2016; Lenormand, 2002). Therefore, the evidence for local adaptation despite the opportunities for gene flow between these Manitoba lakes suggests that the genetic variation within walleye of each lake is a resource for managing risk from future environmental perturbations. Management action that decreases unique characteristics of walleye in each lake could be irreversible on human timescales, and should be considered with additional precaution (Francis, 1996; Persson, 2016; Schindler et al., 2015). I therefore suggest that if supplementation programs are used to support either fishery in Lake Manitoba or Lake Winnipeg, local individuals should be used for supplementation, such as by repatriation to maintain local variation and maximize portfolio diversity and survival of stocked fish (DuFour et al., 2015; Schindler et al., 2015; Thorstensen et al., 2019).

Effective population size for Lake Winnipeg walleye was several times larger than that for either Cedar Bluff Reservoir or Lake Manitoba, at least in the parental generations for the individuals in the present study (Waples & Do, 2010). In addition, at -0.69, Tajima's D in Lake Winnipeg suggested a recent population expansion. Results from δaδi also suggested a population expansion in Lake Winnipeg walleye in the past several thousand years, although masked singleton SNPs led to wide confidence intervals for most parameter estimates because those SNPs are most informative for the approach. These lines of evidence for a population expansion are interesting in the context of outlier SNPs between the Canadian lakes, identified within a candidate region of selection on chromosome 8 within Lake Winnipeg walleye. A concurrent population expansion with a selective sweep is consistent with evolutionary rescue or adaptive niche expansion as a population colonizes a new habitat (Kim & Gulisija, 2010; Osmond & Coop, 2019). An adaptive niche expansion is plausible given the differentiation between Lake Manitoba and Lake Winnipeg walleye, indicating at least some degree of population isolation that is supported by the geologic history of Lake Winnipeg, that had a reduced northern lakebed and desiccated southern basin between 7,500 and 4,000 radiocarbon years before present (Lewis et al., 2002). In other words, Lake Winnipeg walleye may represent a group isolated following the drainage of glacial Lake Agassiz, which subsequently expanded and adapted as Lake Winnipeg itself grew to its present size.

High $F_{IS}$ at approximately 0.15 indicates an elevated chance of inbreeding depression in the Cedar Bluff walleye. Even though heterozygosity and gene diversity were slightly higher in walleye from the Cedar Bluff Reservoir compared to the Canadian walleye, this may not mitigate this risk of inbreeding depression, as heterozygosity and $F_{IS}$ are weakly correlated, but only $F_{IS}$ may more consistently predict reduced fitness (Chapman et al., 2009; Grueber et al., 2008; Slate et al., 2004). This possible inbreeding is significant when viewed in the context of extended LD in the Cedar Bluff walleye, because inbreeding reduces effective recombination (Hartfield & Bataillon, 2018). The higher linkage disequilibrium was consistent with Cedar Bluff's history as a reservoir and flood control tool created in 1951, initial stocking in 1953, and a stocking program begun in 1959 (Simonds, 1998). Taken together, the combination of elevated LD, elevated $F_{IS}$, and low $\beta_{WT}$ is consistent with a founder event in Cedar Bluff Reservoir where a small group of individuals established a population, with limited admixture from individuals from other waterbodies across the United States, thereby decreasing differentiation overall but failing to mitigate inbreeding in the population descended from stocked founders. As such, if stocking were to continue to supplement the Cedar Bluff Reservoir walleye fishery, identification of the ancestral population that contributed most strongly to the founding group would inform an effective translocation program.

### 4.5.2 Signatures of Selection

The cross-population extended haplotype homozygosity approach (i.e., XP-EHH) was chosen as the basis for identifying signatures of selection because of its power to identify possible selection both within and between populations and control for the effects of demographic history (Gautier et al., 2017; Sabeti et al., 2007). While several methods can identify signatures of selection between populations, possible selection within a population was critical to the present study because I examined parallel patterns of selection in a pairwise way in the waterbodies as evidence of adaptation to local conditions. The success of a parallel approach to analyzing the three waterbodies here was most effective for exploring waterbodies in Manitoba. Walleye in Cedar Bluff Reservoir (Kansas) had lower differentiation from the entire genetic pool, overall inbreeding coefficient approximately two times higher than the other sites, history as an artificially created waterbody with a stocked founding population and continued

stocking (Simonds, 1998). Fewer candidate regions of selection were identified in these Cedar Bluff walleye as well, compared to Lake Manitoba and Lake Winnipeg fish. Nevertheless, genes likely significant for walleye adaptations to northern lakes were identified, along with signatures of selection within Lake Winnipeg.

One candidate region on chromosome 4 contained four genes that may be under selection within both Manitoba lakes. *Ceramide synthase 5* and *glycerol-3-phosphate dehydrogenase 1* showed associations with XP-EHH in both Lake Winnipeg and Lake Manitoba compared to Cedar Bluff Reservoir, and with latitudinal outlier SNPs in the same candidate region for selection on chromosome 4. The gene *GPD1* was upregulated in rainbow trout (*Oncorhynchus mykiss*) gill at cold relative to warmer temperatures (Rebl et al., 2013). *Ceramide synthase 5* has a role in catalyzing the formation of the lipid ceramide, and ceramide species lipids were associated with temperature changes in the brain lipidome of crucian carp (*Carassius carassius*) (Käkelä et al., 2008). Increased ceramide was predicted to decrease food intake by counter-acting ghrelin in rainbow trout, but ceramide interactions with food intake were complicated by other signaling pathways as well (Velasco et al., 2017). These patterns suggest possibly subtle but biologically important differences in thermal sensitivity in northern walleye, relative to southern populations. Cold adaptations are plausible given the low temperatures of Manitoba in the winter, with a mean minimum winter air temperature of -19.9 °C in Lake Winnipeg between 1971 and 2000 (maximum -7.8 °C in the same period; Dibike et al., 2012). These air temperatures are not directly comparable to water temperatures experienced by walleye in Cedar Bluff Reservoir because the walleye in Cedar Bluff are from an entirely stocked population of unknown origin. Nevertheless, walleye from the Manitoba lakes are native to waters further north than any source population for walleye stocking in the United States. *Acid sensing ion channel 1* (*ASIC1*) and *SWI/SNF-related matrix-associated actin-dependent regulator of chromatin 1 subfamily D member 1* were two other genes associated with outlier SNPs latitudinally and XP-EHH within both Canadian lakes compared to Cedar Bluff Reservoir. As a sodium channel activated by drops in extracellular pH, *ASIC1* expression has been used as a marker for water flow sensitivity in rainbow trout and whitefish (*Coregonus maraena*) adipose fin tissue (Koll et al., 2020). The gene *ASIC1* is also among the largely conserved ASIC family genes used to study nociception in zebrafish (*Danio rerio*) (Sneddon, 2019), and has an important role in sodium ion uptake in different freshwater fish (Dymowska et al., 2014). The

gene *SMARCD1* has many roles, among which are chromatin remodeling and neural development, and regulates hematopoietic stem cell development in zebrafish (Huang et al., 2013). While the functional significance of these signatures of selection for these genes is not clear for the walleye in Manitoba, taken together, the presence of these four genes in one candidate region of selection may mean that one or a subset of them contribute to adaptation to the Manitoba lakes. However, the shared signatures of selection within each of these colder lakes supports the conclusion that this region of chromosome four is involved in local adaptation.

One region on chromosome 8 showed outlier SNPs along PC2 of the PCA (outliers between Lake Manitoba and Lake Winnipeg), and XP-EHH provided evidence for a candidate region of selection unique to Lake Winnipeg relative to both other waterbodies approximately between positions 15.3 and 15.9 mb along the chromosome. *Leucine-rich repeat-containing protein 4C* was both within the candidate region of selection and associated with an outlier region between Lake Winnipeg and Lake Manitoba. While the role of *LRRC4C* in walleye is uncertain, the stark results in both haplotype and outlier-based tests around the gene indicate a possibly important signature of selection in Lake Winnipeg. Several other characteristics of Lake Winnipeg walleye are consistent with signatures of selection unique to the waterbody. Observations of recently-evolved dwarf walleye morphotypes possibly due to fishing pressure suggest ongoing selection within Lake Winnipeg (Moles et al., 2010). Additionally, there are patterns of low gonadal investment despite high lipid concentrations in Lake Winnipeg walleye relative to walleye in other lakes (Moles et al., 2008). Therefore, nutrient limitations other than lipids may be significant for these walleye, or gonadal investment may be determined by the recent past as opposed to current nutrient reserves (Moles et al., 2008). While I cannot link genotypes to Lake Winnipeg-specific phenotypes in the present study, the existence of unique Lake Winnipeg morphologies is consistent with present observations of Lake Winnipeg-specific signatures of selection via outlier and haplotype tests, and highlights a valuable resource of genetic diversity present in the Lake Winnipeg walleye. This genetic diversity has likely been useful for the population persisting in Lake Winnipeg, and may be useful for future management, such as by assisted gene flow where walleye may need to be translocated to address changing environmental conditions (Hamilton & Miller, 2016).

### 4.5.3 Conclusion

The utility for within- and between- population comparisons is highlighted by the present study as valuable for informing fisheries management and understanding local adaptation. I suggest that Lake Manitoba and Lake Winnipeg walleye each represent populations adapted to northern lakes, and the candidate genes identified here may be studied in the context of freshwater fish adaptations to cold conditions. These results provide a basis for applying the precautionary principle to stocking among these waterbodies (Francis, 1996; Persson, 2016), as movement of genetic material into these populations may diminish present variation. While adaptive variation is not necessary information for prioritizing populations in non-translocated species (Fernandez-Fournier et al., 2021), results in the present study show how the direct identification of adaptive variation can be useful after a history of stocking and for identifying potential adaptive mechanisms (Connon et al., 2018; Mable, 2019).

## 4.7 Acknowledgements

I thank Dr. Jeff Long for contributing walleye fry from Swan Creek Hatchery and for valuable discussions of Manitoba walleye ecology, diversity, and stocking. I thank Colin Charles, Colin Kovachik, Doug Leroux, Nicole Turner, Mike Gaudry, Sarah Glowa, and Emily Barker, who provided fin clips used for walleye DNA from Lake Winnipeg. I thank Evelien de Greef for helpful guidance on synteny analyses, and Dr. Kristen Gruenthal for sharing her expertise in sequencing library preparation. I thank Dr. Colin Garroway for discussions about $\beta_{WT}$. Many analyses in this manuscript were enabled by our opportunity to use computing resources provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). This work was supported by a Fisheries and Oceans Canada Ocean and Freshwater Science Contribution Program Partnership Fund grant awarded to KMJ, JRT, and Dr. Darren Gillis, and a Natural Sciences and Engineering Research Council of Canada Discovery Grants awarded to KMJ (#05479) and JRT (#06052). Work by JRT is also supported by the Canada Research Chairs program (#223744) and the Faculty of Science, University of Manitoba (#319254).

# Chapter 5. A Polygenic Basis for Habitat Selection in Lake Winnipeg Walleye (*Sander vitreus*)[4]

## 5.1 Abstract

Movement is fundamental to animal survival, reproduction, and population persistence. Movement in terms of habitat selection and the possible genomic architecture underlying tendencies of habitat selection in freshwater fishes may be powerful for understanding why fish select the habitat they do and provide information applicable to conservation. In the present study, *n*=113 walleye (*Sander vitreus*) from Lake Winnipeg, Manitoba, Canada were non-lethally sampled for DNA from fin clips while being surgically implanted with an acoustic transmitter and tracked for ≥180 days (approx. 6 months, or one open-water season when walleye move most). Maximum residency, the greatest proportion of time spent in any of three ecological zones was used as a phenotype for genome-wide association with 21,120 single nucleotide polymorphisms (SNPs). Polygenic scores were investigated, where SNPs associated with movement at a *p*-value threshold of 0.40 were used to develop a polygenic model that explained 81.6% of variation in maximum residency tendency. A functional analysis identified gene ontology terms for underlying genomic mechanisms that may affect movement and have a hypothesis-generating value. Forebrain neuron differentiation (GO:0021879), positive regulation of glutamate receptor signaling pathway (GO:1900451), metanephric nephron tubule development (GO:0072234), and positive regulation of cyclin-dependent protein serine/threonine kinase activity involving G1/S transition of mitotic cell cycle (GO:0031659) were each the highest terms as ranked by a combined score incorporating significance and deviation from expected rank in a Fisher's exact test. An analysis of heterozygosity, gene diversity, and inbreeding coefficients in maximum residency associated SNPs revealed elevated heterozygosity and inbreeding at those genomic variants. These results have implications for a genomic influence on movement between ecological zones in Lake Winnipeg walleye, but the polygenic

[4] This chapter is in preparation for publication, but has not been published. Matt J Thorstensen conducted the analyses and wrote the original manuscript. Matt J Thorstensen and Jennifer D Jeffrey contributed to wild walleye sampling. Lilian M Wiens was first reader for aging scales, and curated age data. Inesh Munaweera analyzed movement data. Jennifer D Jeffrey provided guidance on functional analyses. Ken M Jeffries and Matt J Thorstensen edited the manuscript. Ken M Jeffries and Jason R Treberg acquired funding. All authors contributed to conceiving of and designing the study.

model developed here is likely not generalizable in other systems. Future research may test hypotheses developed by the functional analysis here, and connect genomic variation underlying maximum residency with broader ecological factors.

## 5.2 Introduction

Movement is fundamental to animal survival, reproduction, and ultimately, persistence and evolution (Nathan et al., 2008). Questions of the underlying factors that motivate, enable, and are affected by movement are therefore of both applied and fundamental importance for biologists seeking to conserve threatened populations and understand natural populations through ecosystem dynamics (Nathan et al., 2008). In fishes, movement has been studied at numerous scales, from individual lateral asymmetries in escape responses to broad seasonal migrations (Dadda et al., 2010; Thompson et al., 2020; Turner et al., 2021). A genomic basis of movement could explain the extent to which genomic variation predicts movement phenotypes of species. For instance, genomic regions have been used to characterize migration run timing ecotypes of Pacific salmon (*Oncorhynchus* spp.), where earlier migrants often face disproportionate impacts from fishing pressure (Brieuc et al., 2015; Hess et al., 2016; Koch & Narum, 2020; O'Malley et al., 2013; Prince et al., 2017; Thompson et al., 2020).

While a large genomic effect on migration timing provides hypotheses for future research (85% of variance in freshwater entry timing explained by a small genomic region in Thompson et al., 2020), movement timing is only one of many ecologically significant movement-related phenotypes with a possible genomic basis. Another is habitat selection, or the innate or learned behaviours that an animal makes that influences habitat choice (Hutto, 1985; Krausman, 1999). For ecology and conservation, knowledge of habitat selection has been established as critical for prioritizing habitat for protection (Mayor et al., 2009). This selection behaviour may be inferred from observed habitat occupancy and results in disproportionate use of some habitat resources (Krausman, 1999). Habitat selection has been studied as an important behavioural trait in estuarine and marine fishes, such as in coral reefs (Freitas et al., 2016; Lindberg et al., 2006; Montgomery et al., 2001; Morrissey & Gruber, 1993), but relatively less empirical data is available for habitat selection in obligate freshwater fish (although Ayllón et al., 2010 found brown trout, *Salmo trutta* preferred deeper and colder water at increased sizes in freshwater

rivers with similar ion ionic content). Habitat selection has been linked to population dynamics via emigration in pupfish *Cyprinodon macularius*, supporting the hypothesis that habitat selection has implications for fish populations in general (McMahon & Matter, 2006). Physiology has been connected with habitat selection, as well, such as by linking microhabitat temperature to ectotherm performance at different temperatures (Huey, 1991). Evidence for a genetic tendency for habitat selection was also found in several arthropods and mollusks (Jaenike & Holt, 1991), lending support to the possibility that a genomic tendency for habitat selection may be found in freshwater fish. Therefore, knowledge of habitat selection and the possible genomic architecture underlying tendencies of habitat selection in freshwater fish may be powerful for understanding why fish select the habitat they do and provide information applicable to conservation.

Spatial ecological differences underlying habitat selection often correlate with spatially-varying population structure, common in many fishes (Chistiakov et al., 2006; Jenkins et al., 2010; Reiss et al., 2009). Population structure has been associated with geographic distances between groups since genetic methods were first applied to fishes, such as in studies of isolation by distance (Thompson, 1931; Wright, 1943; see Jenkins et al., 2010 for a review of the topic). With respect to linking genomic variation to habitat selection, any kind of statistical association would need to address genetic variation underlying population structure (if present) because heterogeneity in landscape habitats may be erroneously related to heterogeneity in neutral genetic variation (i.e., isolation by distance; Jenkins et al., 2010). One approach in modeling genetic associations with phenotypes is to include population structure as a covariate (Choi et al., 2020). However, including covariates is not necessarily a sufficient correction, especially in circumstances with relatively few samples, because of multicollinearity between the population structure covariate and underlying habitat heterogeneity (Choi et al., 2020; Jenkins et al., 2010). Given the challenges of linking habitat selection to genomic variation, a first step to addressing the question of what genomic markers may underlie habitat selection is by using habitat selection in a habitat-agnostic manner. Instead of the habitat selected being the phenotype chosen, the tendency for fish to select any single habitat a greater or lesser proportion of time could be the phenotype. This perspective on habitat selection makes genomic associations tractable to models of movement information across a waterbody, even with isolation by distance, if fish from across the spectrum of population structure show variability in habitat selection. To measure how

habitat is selected, residency values may be calculated by the proportion of time spent in different habitats, and the maximum residency value from any of the ecological zones in a dataset can be used as the phenotype representing the magnitude to which a fish shows a tendency to select particular habitats. However, maximum residency is a distinct trait from migration timing described previously, and cannot be assumed to be as discrete as seasonal run differences are in salmon.

Polygenic scores, also called polygenic risk scores or breeding values, may be informative for genomic associations with maximum residency because they allow for identification of associations that do not show classic signatures of large-effect differences more readily identifiable with genome-wide association studies (GWAS) (Pritchard et al., 2010). In classic GWAS, small regions of large effect are identified by peaks in single nucleotide polymorphism (SNP) significance. However, the basic insight of the polygenic perspective is that many loci contribute to many traits and a likelihood for the phenotype of a given trait can be calculated from the aggregate effect of identified loci. This approach is similar in perspective to additive genetic variance associated with phenotypic variance in quantitative genetics (Fernando et al., 1994; Xie & Xu, 1998). Polygenic modeling is not separate from GWAS, either, because allele effect sizes for additive polygenic effects are based on associations established by GWAS (Choi et al., 2020; Healy et al., 2018). Therefore, genomic regions of large effect on maximum residency, if they exist, can still be identified when undertaking a polygenic approach. However, given prior research on the genetic basis of habitat preference and the continuous nature of the phenotype in question (Jaenike & Holt, 1991), habitat selection was hypothesized to be a polygenic trait.

Association testing underlying polygenic modeling also enables analysis of the functional significance of genomic regions underlying maximum residency. The genes *greb1L* and *rock1* have been associated with salmon migration timing in several studies, with possible implications for interactions between fish development and movement (Koch & Narum, 2020; Prince et al., 2017; Thompson et al., 2020). For traits with a possibly polygenic basis however, association testing may not emphasize common variants of small effect (Crouch & Bodmer, 2020). Polygenic models provide thresholds of significance to which a set of SNPs contributes to a polygenic score, and therefore there is a genomic basis for some proportion of variance in a trait.

These sets of SNPs defined by modeled significance thresholds provide an opportunity to apply gene set enrichment analysis for functional regulatory networks possibly underlying maximum residency (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021). Here, enhancer-promoter interactions or linkage with causative SNPs in genes is a key assumption for identifying the functional significance of intergenic SNPs associated with movement (Fulco et al., 2019). While genome scans may be unlikely to find causative SNPs when SNP effect sizes are small (as in polygenic models) (Crouch & Bodmer, 2020), the polygenic perspective of aggregating many loci of low effect is consistent with the system-wide perspective adopted in enrichment analyses (Chen et al., 2013). Therefore, enrichment analysis with polygenic data may not identify causal variants or even individual genes that underlie maximum residency, but may indicate regulatory networks or molecular pathways that are potentially related to freshwater fish movement tendencies.

The ability for polygenic modeling to establish connections between evolutionary, genomic, and phenotypic characteristics of wild fishes may inform how the distribution of heritable traits, demography, population density, and selection operate in positive and negative feedback loops (Bassar et al., 2021). For instance, polygenic modeling allows the identification of a panel of SNPs associated with a trait that may be used to investigate how genetic diversity underlying traits of interest and ecology interact (Hughes et al., 2008). As applied to maximum residency for this study, investigating genetic variation in residency-associated SNPs in conjunction with the variation in residency tendency explained by those SNPs can reveal how genetic variation interacts with habitat selection. Habitat selection is related to habitat use, which can be affected by individual SNPs in the walleye genome that have ecological consequences (Hughes et al., 2008; Mayor et al., 2009; McMahon & Matter, 2006).

In the present study, $n$=113 walleye (*Sander vitreus*) from Lake Winnipeg, Manitoba, Canada were non-lethally sampled for DNA from fin clips while being surgically implanted with an acoustic transmitter and tracked for ≥180 days (Fig 5.1). As an obligate freshwater fish with broad seasonal movement, walleye are an effective species for testing genomic associations with habitat selection (Turner et al., 2021). A panel of 21,120 SNPs were called from DNA and used for associations with walleye residency in three different ecological zones of Lake Winnipeg (defined in Brunskill et al., 1980, based on surface area and depth across Lake Winnipeg).

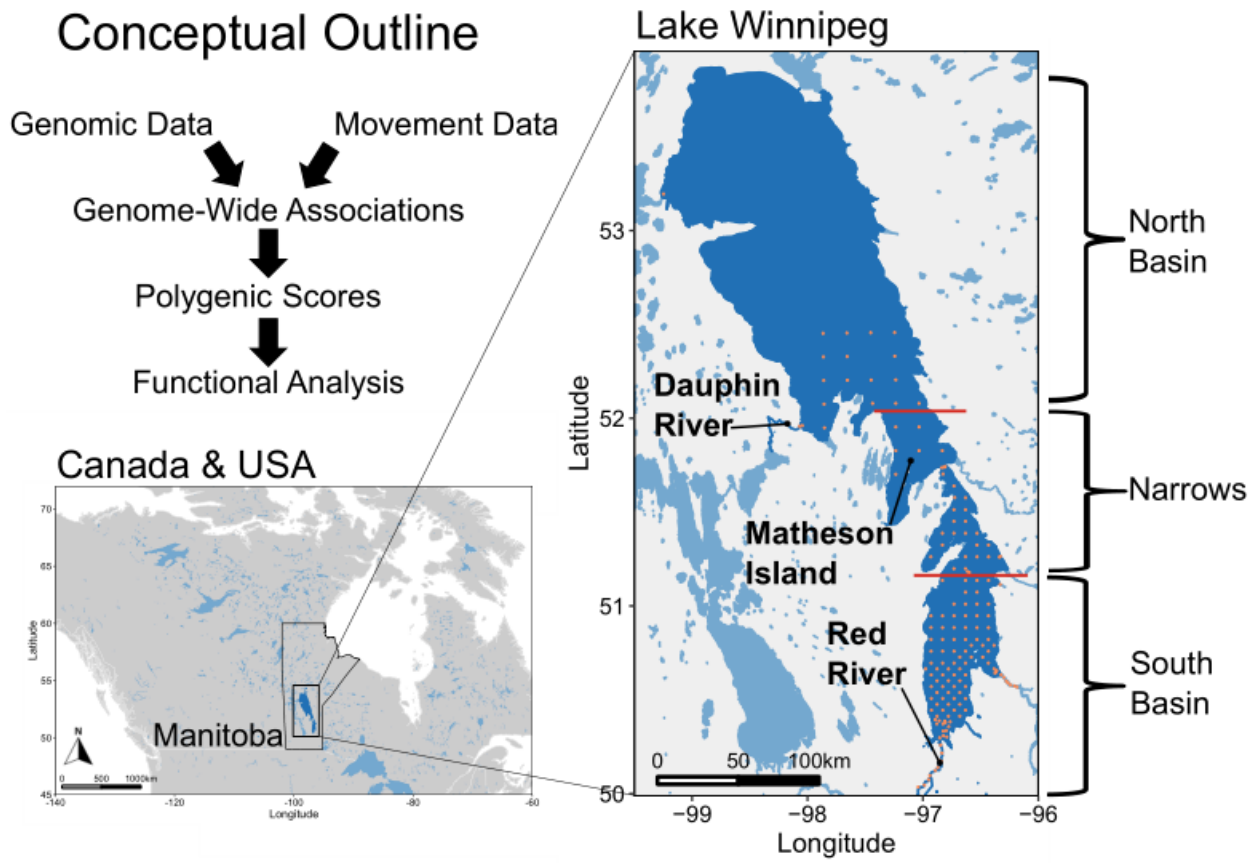


**Fig 5.1** Conceptual outline of the present study, map of portions of Canada and the USA, and map of Lake Winnipeg (Manitoba, Canada). The conceptual outline demonstrates the bioinformatics pipeline used where genomic data and movement data were used to conduct a genome-wide association analysis, from which polygenic scores were developed, and a functional gene set enrichment analysis performed with genomic variants used for the polygenic model. In the map of Lake Winnipeg, orange points are acoustic receiver array receiver locations (VEMCO). Red lines denote separations between ecological zones in Lake Winnipeg used for calculating maximum residency values. Here, the lake is divided into a south basin, narrows, and north basin. Hecla Island, at the northern boundary of the south basin, is not shown.

Residency was calculated as the proportion of time for each fish observed in any of the three ecological zones over the total time observed (i.e., ≥180 days). The maximum of the three residency proportions was used as the maximum residency phenotype for association testing. Using maximum residency proportions is consistent with walleye movement between basins in both historical and contemporary datasets (i.e., 1974-1977 & 2017-2019, 2.5% - 17.5% movement probability between basins) (Turner et al., 2021), because such results indicate variation in walleye movement between basins that has persisted over multiple generations. Polygenic scores were estimated, and a polygenic model was developed, following association testing. To test how genetic variation may interact with maximum residency, heterozygosity ($H_O$), gene diversity ($H_S$), and the inbreeding coefficient ($F_{IS}$) were calculated for SNPs included in the polygenic model, SNPs excluded from the polygenic model, and overall SNPs. Habitat selection was hypothesized to be influenced in part by genomics, with many genomic regions contributing to selection behaviour. Therefore, maximum residency was predicted to be polygenic. Last, functional analyses were conducted with SNPs included in the polygenic model to develop hypotheses for the molecular basis of walleye residency.

## 5.3 Methods

### 5.3.1 Telemetry & Maximum Residency

Movement data for walleye implanted with VEMCO acoustic tags (69 kHz V16, V13, V13TP) in 2017 and 2018 from Lake Winnipeg were collected from a VEMCO acoustic receiver array (VR2W, VR2Tx) concentrated in the south basin of the lake (Fig 5.1) (Vemco, Innovasea, Bedford, NS, Canada) (Munaweera et al., 2021; Turner et al., 2021). Movement paths analyzed in the present study were limited to the year following when a fish was sampled. Fish positions were calculated over time steps of one day. If multiple detections were observed in a day, the mean latitudinal and longitudinal position was taken from those multiple detections to estimate a mean location for the fish in that day. Because walleye move primarily during the open-water season (approx. May-November; Munaweera et al., 2021), fish were included for genomic analyses if at least 180 days (approx. 6 months) of movement data were available by filtering for

fish with detections ≥180 days apart. This filtering was designed to capture the open-water season following sampling, because fish were sampled in May of 2017 and 2018, and ice cover in the south basin began approximately November 17, 2017, and November 16, 2018 (based on satellite observations of Lake Winnipeg from https://zoom.earth/).

A relatively coarse movement phenotype of maximum residency was chosen for this analysis. Here, Lake Winnipeg was divided into three ecologically distinct zones based on Brunskill et al., (1980): a south basin characterized by inflows from the Red and Winnipeg Rivers, a narrows region between the basins, and a north basin characterized by oscillating water inflow from the south basin and flow from the Dauphin River (Fig 5.1). Residency was calculated as the proportion of time a fish spent in each zone out of the total time observed. Because the phenotype of interest was a tendency toward residency overall, rather than a tendency toward residency in any particular zone, the maximum residency value out of the three zones was pulled from the data as the phenotype used for association testing. Therefore, this value ranged from a minimum of $0.3\overline{3}$, or even residency across all three zones, to a maximum of 1.0, which would represent complete residency in any of the three zones.

### 5.3.2 DNA Sampling & Bioinformatics

Walleye sampling for DNA in the present study was described in chapter 4, with an identical bioinformatics pipeline prior to SNP calling, including alignment to the yellow perch genome (*Perca flavescens;* Feron et al., 2020). For calling SNPs, only individuals from Lake Winnipeg for which two movement observations were at least 180 days (i.e., 6 months) apart were included, to limit the analysis to individuals with a full open water season of movement data after being sampled in May. Out of 235 walleye sampled for DNA in Lake Winnipeg, $n$=113 were included for SNP calling for the present study after observing available movement data ($n$=78 from the Red River, 16 from Matheson Island, and 19 from the Dauphin River). A minor allele count >1, 0.60 present data per sampling site, and 0.95 present data overall were required for each SNP. With these settings, 20,146 SNPs were called with 26 SNPs removed because they were on unplaced scaffolds, leaving 21,120 SNPs for phasing and imputation. Beagle v5.1 was used over 20 phasing iterations and assuming an effective population size ($N_e$)

of 4,500 (based on the estimate for Lake Winnipeg walleye $N_e$ in chapter 3) to impute SNPs (Browning et al., 2018; Browning & Browning, 2007).

### 5.3.3 Association Testing

GEMMA v0.98.3 was used to fit linear mixed models for SNP associations with maximum maximum residency (Zhou, 2017; Zhou & Stephens, 2012). Here, covariates included site collected, year sampled, sex, mass, length, principal component 1 (PC1), principal component 2 (PC2), Lake Manitoba ancestry, age, and number of days observed. PC1 and PC2 were intended to represent population structure within Lake Winnipeg (see Chapter 2) and were created with SNPRelate v3.12 using SNPs pruned for linkage disequilibrium at a threshold of 0.2, leaving 10,321 SNPs used in the principal components analysis (Zheng et al., 2012). Lake Manitoba ancestry was included to address the possibility that ancestry may influence movement in a similar manner to population structure, and was estimated using the ancestry coefficients (i.e., Q-values) from ADMIXTURE for fish sampled in Lake Winnipeg in Chapter 3. Age was included to address the possibility that walleye age may influence movement, and was estimated from scales taken at the time of sampling. Scales were measured under a Leica EZ4 (Leica Camera, Wetzlar, Germany) dissecting microscope and pictures were taken using LAS EZ 3.4.0 Sofware (Leica Camera), in duplicate, where at least 2 scales were photographed for comparison. Picture files were analyzed independently by two observers using ImageJ (Schneider et al., 2012), where each annulus was considered a year for the walleye. While scale ages may be unreliable for the walleye age classes used in the present data, length measurements were included as covariates as well to address possibly older walleye included in the data. Last, the number of days in which a walleye was observed in the telemetry array was included to address possible issues with sampling bias, because even though every fish included in the association tests had observations at least six months apart, the number of detections for individual fish varied. A centered relatedness matrix was created with GEMMA using all modeled individuals and default settings. All covariates, the relatedness matrix, 21,120 imputed SNPs for $n$=113 walleye, and maximum residency phenotypes were included for likelihood ratio association tests using univariate linear mixed models in GEMMA. The R packages tidyverse v1.3.0 and ggman v0.99.0 were used to visualize association test results (R Core Team, 2021; Wickham et al.,

2019). A false discovery rate (FDR) adjustment was applied to likelihood ratio test *p*-values to explore multiple test-corrected significance (*q*-value; Storey & Tibshirani, 2003).

### 5.3.4 Polygenic Scores

Polygenic scores were assessed and modeled using a framework described in Choi et al., (2020). Note that polygenic scores are often called polygenic risk scores in the literature on human genetic diseases and breeding values in agricultural research. The term polygenic score is used throughout the present study because the maximum residency phenotype analyzed here is not a disease with an associated risk, and is not being used to develop models used in walleye breeding. Plink v1.90b6.12 was used to clump SNPs 250 kb apart at a LD threshold of 0.1 with *p*-values from the likelihood ratio test for movement associations, to generate a list of unlinked SNPs that represent the lowest *p*-values observed among clumps of SNPs associated with walleye movement (Purcell et al., 2007). Polygenic scores for each individual were then assessed using these clumped SNPs and the beta values for SNP effect sizes over likelihood ratio test *p*-value thresholds of 0.001, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5. Therefore, seven different polygenic scores were estimated for each walleye modeled, one score for each SNP *p*-value threshold for SNP associations with maximum residency. Therefore, the higher the threshold, the more SNPs were included in generating the polygenic score. For visualizing polygenic scores, because Lake Manitoba ancestry may have affected either the polygenic scores themselves or walleye residency in Lake Winnipeg, the Lake Manitoba ancestry coefficients previously used as covariates in GEMMA were plotted in conjunction with polygenic scores and maximum residency.

Polygenic scores were then modeled in R with linear models, where the dependent variable was the maximum residency phenotype used in association testing and the independent variables were polygenic scores and all covariates included in association testing (R Core Team, 2021). Here, linear model significance and $R^2$ was assessed for each *p*-value threshold used for SNP inclusion in the polygenic models. The *p*-value threshold was chosen based on the lowest *p*-value observed among the different linear models tested. Model fit, homoscedasticity, homogeneity of variance, non-normality of residuals, and Cook's distance were assessed with the R package performance v0.7.0 (Lüdecke et al., 2021).

Generalizability of a polygenic model is most rigorously assessed with an independent sample, or more commonly, subdividing data into training and test datasets and performing jackknife partitions among replicate association tests (Choi et al., 2020). However, the limited sample size of the present polygenic model ($n$=113) is far below the sample sizes of thousands or hundreds of thousands of individuals more appropriate for association testing, and no independent walleye genotypes and movement data were available for the present study. As such, generalizability of the polygenic model could not be assessed with an independent sample, and jackknife cross-validation would have unacceptably lowered statistical power for training data using any threshold for training individuals. Instead, predictive $R^2$ was assessed for the polygenic score model with the highest $R^2$. When calculating predictive $R^2$, the predicted residual sum of squares (PRESS; Allen, 1974) is calculated with a leave-one-out cross validation process, then divided over the total sum of squares identified by ANOVA. Predictive $R^2$ is therefore:

$$\text{Predictive } R^2 = 1 - \frac{\text{PRESS}}{\text{Total Sum of Squares}}$$

This predictive $R^2$ value was then interpreted as a coefficient of determination showing the variance in maximum residency explained by polygenic scores, and thus explained by genomic variation (see Kasuya, 2019 for a similar approach to coefficients of determination).

To explore the generalizability of the polygenic model to other walleye in Lake Winnipeg and to test the polygenic scores for individuals not included in developing the model, polygenic scores were calculated for walleye for whom receiver array detections ≥180 days apart were *not* available, but DNA sequencing data represented individuals who disappeared from the movement dataset. These individuals may have disappeared from the data because of death or by evading detection in the telemetry array. To calculate the scores for the set of individuals with unknown movement, a panel of loci and columns within the loci representing SNPs called for the original set of walleye was created (i.e., a whitelist), and genotypes called with the populations function in Stacks v2.3e (Catchen et al., 2013). These SNPs were imputed in a procedure identical to that used for the original data, and polygenic scores identified based on associations developed in GEMMA with the original data.

### 5.3.5 Genetic Variation

Hierfstat v0.5-7 was used to assess observed heterozygosity ($H_O$), gene diversity ($H_S$), and inbreeding coefficients ($F_{IS}$), among the 113 walleye used in the polygenic model. SNPs for these individuals were divided into three subsets: SNPs used to develop the polygenic model (maximum residency-associated SNPs), SNPs *not* used to develop the polygenic model (non-residency-associated SNPs), and all SNPs together (overall SNPs).

### 5.3.6 Functional Analyses

To investigate the genes underlying maximum residency in Lake Winnipeg walleye, the possible functional significance of SNPs incorporated into the polygenic model was investigated. Here, only the clumped SNPs used to develop the polygenic model with the highest $R^2$ with maximum residency were analyzed. Magma v1.08b was used to associate these SNPs with genes in conservative windows 15 kilobase pairs (kb) upstream and 5 kb downstream of a gene; a relatively small window because enhancer-promoter interactions operate at tens to hundreds of kilobase pairs apart (up to over 1000 kb in Fulco et al., 2019). Enrichr v3.0 was used to summarize genes associated with SNPs used for polygenic modeling into gene ontology (GO) terms by querying the list of genes against the biological function 2018, molecular function 2018, and cellular component 2018 databases (Chen et al., 2013; Kuleshov et al., 2016). Revigo was used to identify significant ($q$<0.05) non-redundant GO terms from Supek et al. (2011), weighted by the combined score (incorporating both significance of the GO term and the magnitude of deviation from an expected rank from a Fischer's exact test) for GO terms from EnrichR. In addition, the genes *growth regulation by estrogen in breast cancer 1-like protein* (*greb1L*) and *rho-associated protein kinase 1* (*rock1*) were interrogated for SNPs associated with residency in the previously described 20 kb window (15 kb upstream, 5 kb downstream) around each gene, because the two genes were identified by multiple studies to have associations with salmon migration timing (Koch & Narum, 2020; Prince et al., 2017; Thompson et al., 2020). While migration timing is distinct from maximum residency analyzed in the present study, allocation of time is implicit in our definition of residency as a proportion of time spent in different ecological zones.

## 5.4 Results

### 5.4.1 Telemetry & Maximum Residency

The *n*=113 walleye were observed with detections ≥180 days apart (Fig 5.2A). Among the 113 fish included for movement modeling, maximum residency was 1.0 for 35 individuals, indicating complete residency among any of the three ecological zones used for the present study (Fig 5.2B). No fish sampled at Matheson Island in the intermediate narrows was observed to have 1.0 residency; instead, those full residents were either sampled from the southern Red River (*n*=21) or from the northern Dauphin River (*n*=14) (Fig 5.2B; Fig 5.2C). While fish sampled at Red River and Matheson Island showed intermediate values for maximum residency, fish in the Dauphin River were skewed toward being full residents (Fig 5.2C).

### 5.4.2 Association Testing

No SNPs reached genome-wide significance, with a minimum *p*-value of $8.02 \times 10^{-5}$ and minimum *q*-value of 0.563. No apparent peaks of SNPs with shared significance were visible in a Manhattan plot (Fig 5.3).

### 5.4.3 Polygenic Scores

Out of 20,146 SNPs originally called, 11,604 were included for generating polygenic scores based on clumping for linkage in Plink. Of these 11,604 SNPs, 5,404 were used to develop polygenic scores based on a *p*-value threshold of 0.4 and unadjusted $R^2$ of 0.837, which was the model with the lowest *p*-value for significance in predicting maximum residency (Fig 5.4). Therefore, results from the polygenic model with the 0.4 *p*-value threshold were used for downstream modeling and functional analysis. Predictive $R^2$ for this model was 0.816, indicating 81.6% of variation in maximum residency could be explained by polygenic scores, and that this variation explained was predictive at least within Lake Winnipeg fish (Fig 5.5). Note that this

81.6% variation explained in residency is solely from the polygenic scores included in the linear model; covariates were included for GWAS earlier in analysis but not at this step. Model fit,

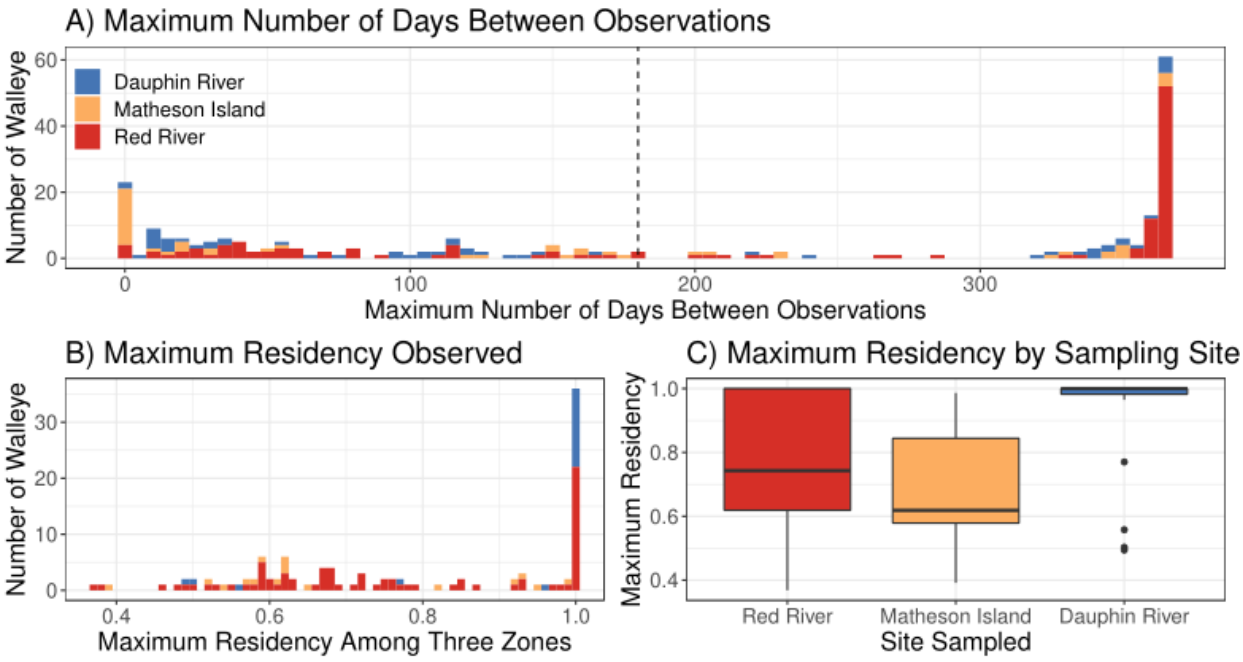


**Fig 5.2** Movement data used for Lake Winnipeg (Manitoba, Canada) walleye (*Sander vitreus*) sampled from the Red River (south basin, *n*=78), Matheson Island (narrows, *n*=16), and the Dauphin River (north basin, *n*=19). Panel A shows a histogram of the maximum number of days between detections available for a fish from an acoustic receiver array. Fish with detections <180 days apart were excluded from the dataset, thus *n*=113 fish had detections ≥180 days apart and were visualized in panels B and C, and used in subsequent analyses. Panel B is a histogram of maximum residency values, where maximum residency is the maximum proportion of time a walleye spent in any of the three ecological zones used in the present study. A fish that stayed in one zone for its entire time detected would have a residency value of 1.0, while a fish that spent time in multiple zones would have a lower value. Panel C shows boxplots of maximum residency values by sampling site.

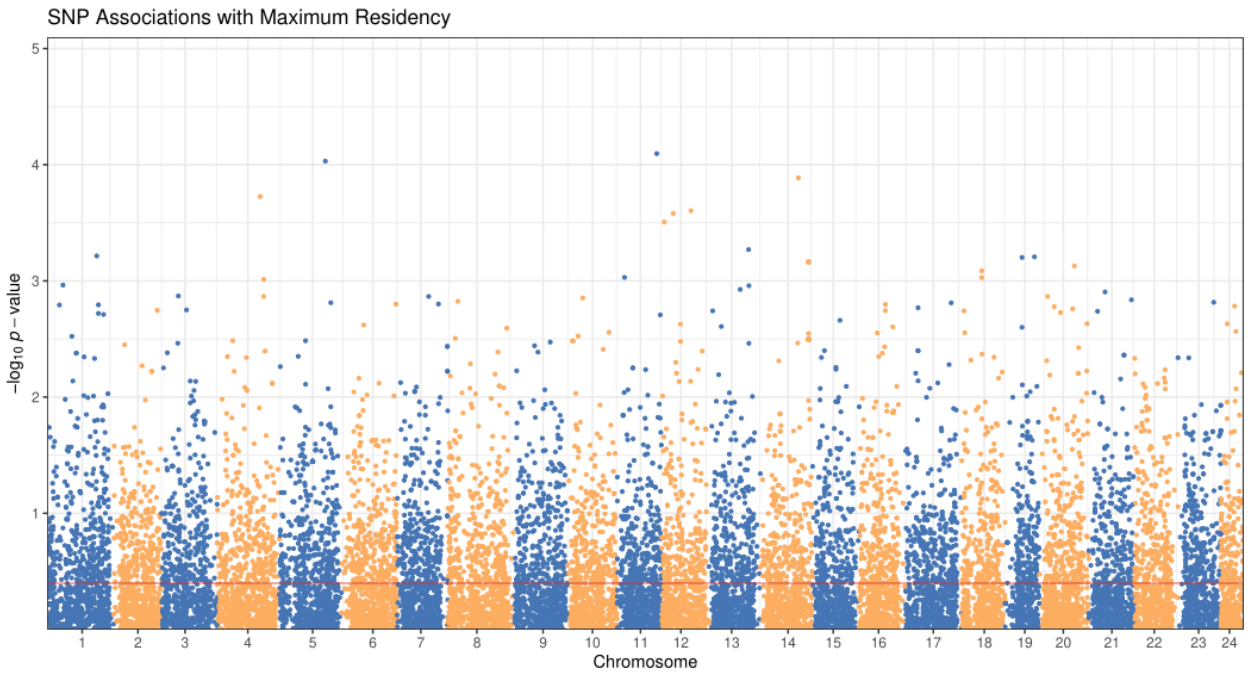


**Fig 5.3** Single nucleotide polymorphism (SNP) associations with maximum residency values for *n*=113 Lake Winnipeg (Manitoba, Canada) walleye (*Sander vitreus*). 20,120 imputed SNPs were associated with maximum residency, where principal component 1, principal component 2 (from a principal components analysis), site collected, year collected, sex, mass, length, ancestry coefficient from Lake Manitoba (from Chapter 4), age, and maximum number of days between detections were used as covariates for association tests. Likelihood ratio test *p*-values were used to evaluate SNP significance, and are plotted on the vertical axis after a -$\log_{10}$ transformation. Chromosomes on the horizontal axis refer to chromosomes in the yellow perch (*Perca flavescens*) reference genome, against which walleye DNA was aligned (Chapter 4 provides details for synteny analysis). The red line represents *p*=0.4, which was the significance threshold identified for SNP inclusion in polygenic modeling. In the present figure, SNPs above the red line were included for developing polygenic scores. Point colours have no biological meaning, and are used to distinguish SNPs from different chromosomes.

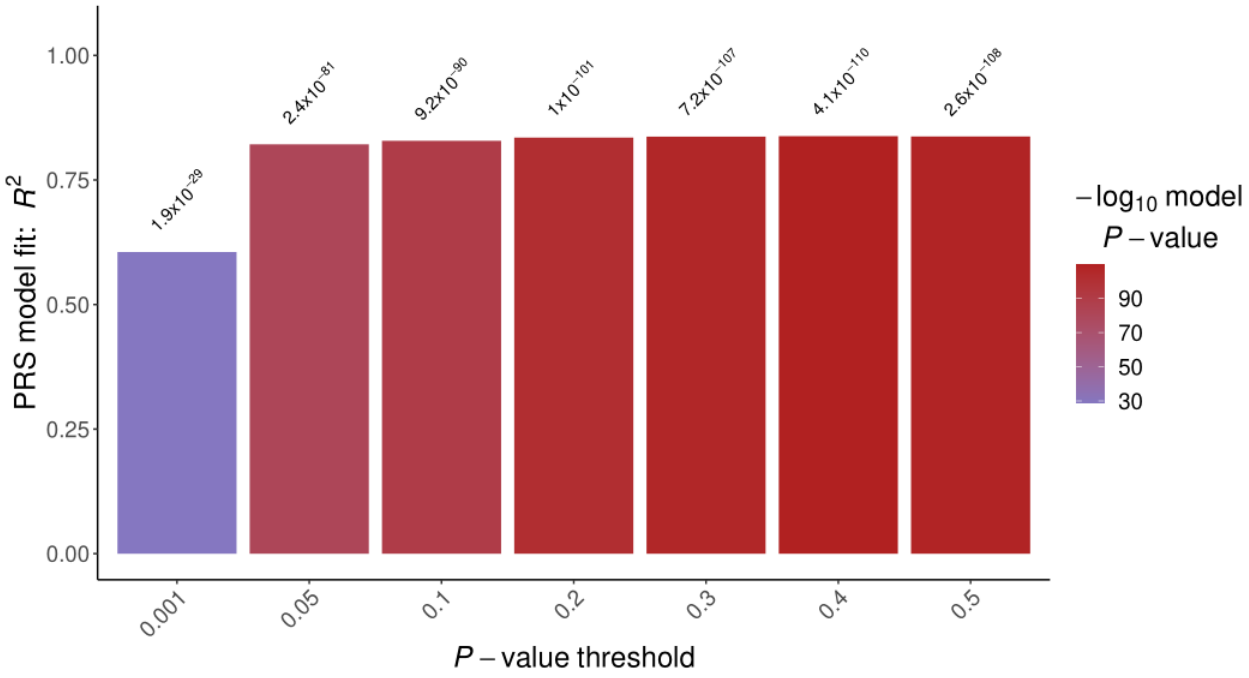


**Fig 5.4** *P*-value thresholds of significance for single nucleotide polymorphism (SNP) inclusion in polygenic models, based on a genome-wide association of SNPs with maximum residency for walleye (*Sander vitreus*) from Lake Winnipeg (Manitoba, Canada). The thresholds of significance represent different thresholds of SNP associations with maximum residency, where SNPs with $p<0.4$ were identified as having the lowest *p*-value in the present analysis ($4.1x10^{-110}$). These thresholds were determined by using linear models with maximum residency as the dependent variable and polygenic scores as the independent variable.

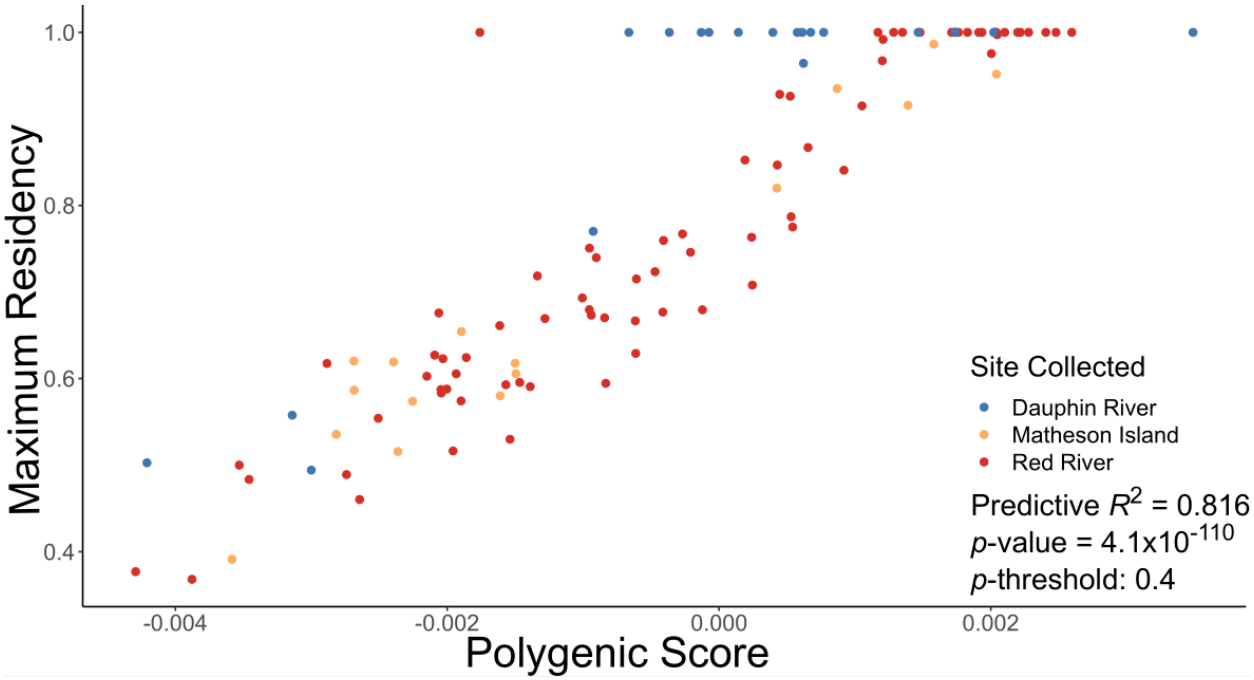


**Fig 5.5** Polygenic scores and maximum residency values for *n*=113 walleye (*Sander vitreus*) sampled from Lake Winnipeg (Manitoba, Canada) from the Red River (south basin), Matheson Island (narrows), and Dauphin River (north basin). Polygenic scores were calculated from additive single nucleotide polymorphism (SNP) effect sizes, based on association testing with maximum residency at a *p*-value threshold of 0.4. A linear model with maximum residency as the dependent variable and polygenic scores the independent variable was developed, from which predictive $R^2$ was calculated using a leave-one-out cross validation process to find the predictive residual sum of squares, divided by the total sum of squares.

homoscedasticity, homogeneity of variance, non-normality of residuals, and Cook's distance did not indicate large biases in the linear model or in outliers in underlying data (Fig A.4.1). Among modeled fish, many outlier fish with higher maximum residency than expected given their polygenic scores were sampled in the Dauphin River (Fig 5.5); when Lake Manitoba ancestry was visualized in conjunction with polygenic scores and residency, outlier fish may largely be driven by ancestry from Lake Manitoba (Fig A.4.2).

Polygenic scores for individuals not included in developing the polygenic model were clustered around the mean for polygenic scores for the original set of individuals (original polygenic score mean $-3.4 \times 10^{-4} \pm 3.3 \times 10^{-4}$ 95% CI; predicted polygenic score mean $-3.3 \times 10^{-4} \pm 3.1 \times 10^{-5}$ CI; Fig 5.6). The tight clustering of predicted individual polygenic scores around the mean for the original dataset may be a failure by the model to identify extreme individuals of either low or high residency values, and therefore a limitation of the model to predict residency in walleye unincluded in developing the polygenic scores. Unfortunately, the residency values for the individuals with observations <180 days apart cannot be considered in the same context as residency values for individuals included in developing the model because less than one open water season of data was available and residency values would likely differ with more data available. As such, the predicted scores and residency values here can *not* be interpreted as a test dataset against which the model and training data may be validated.

### 5.4.4 Genetic Variation

$H_O$ and $H_S$ were slightly higher for the 5,404 maximum residency-associated SNPs used to generate polygenic scores than for the overall set of 20,120 SNPs (Table 5.1). $F_{IS}$ did not significantly differ between each group of SNPs, but tended to be slightly higher for the maximum residency-associated SNPs. $H_O$ and $H_S$ were lowest for the 14,717 non-residency-associated SNPs, with $F_{IS}$ significantly lower for the non-residency associated SNPs relative to the residency-associated SNPs.

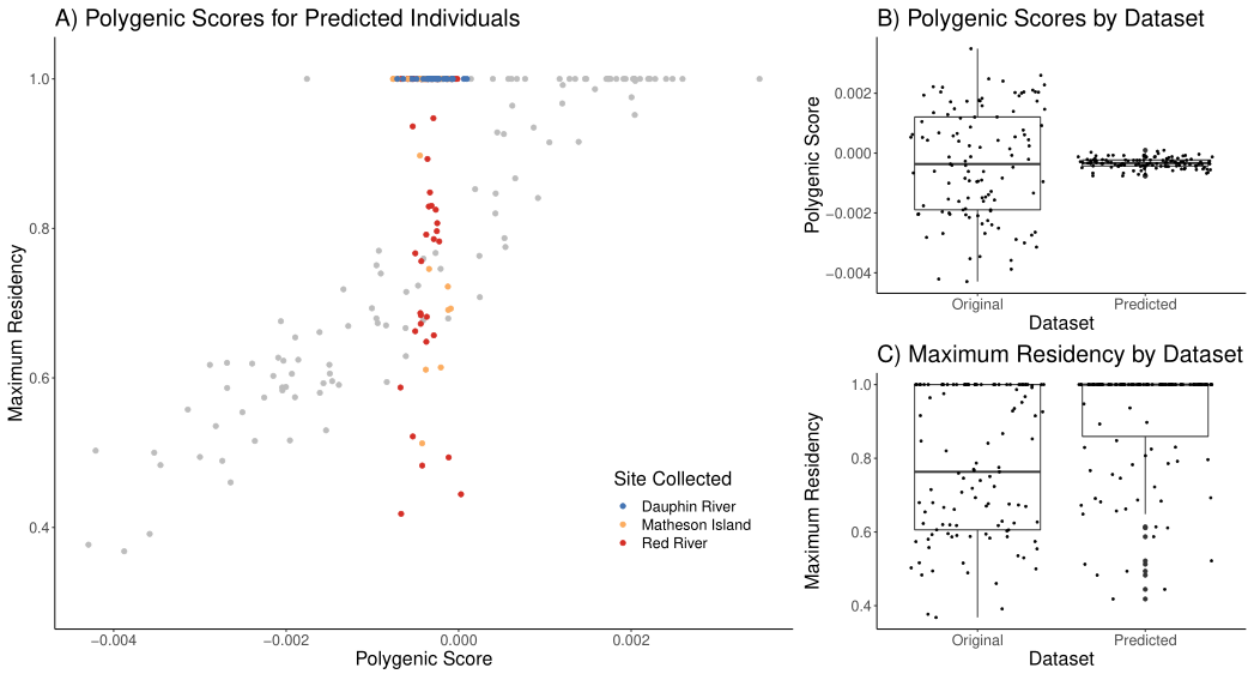


**Fig 5.6** Polygenic scores and maximum residency values for individuals excluded from genomic associations and polygenic modeling because they had observations <180 days apart at the most (i.e., ≤ approx. 6 months of data available). Panel A shows polygenic scores and maximum residency values for walleye (*Sander vitreus*) from Lake Winnipeg (Manitoba, Canada) sampled at the Red River (south basin), Matheson Island (narrows), and Dauphin River (north basin). In gray are polygenic scores and maximum residency values for individuals included in genomic associations and polygenic modeling (e.g., Fig 5.5). Panels B and C also show polygenic scores and maximum residency, respectively, but show boxplots to visualize distributions of each point relative to the individuals included in association testing and polygenic modeling.

**Table 5.1** Observed heterozygosity ($H_O$), gene diversity ($H_S$), and inbreeding coefficients ($F_{IS}$) for single nucleotide polymorphisms (SNPs) overall (21,120 SNPs), SNPs not associated with maximum residency ($p \geq 0.40$; 14,717 SNPs), and associated with maximum residency ($p < 0.40$; 5,404 SNPs). For $F_{IS}$, 95% confidence intervals as estimated over 1,000 bootstrapped iterations in Hierfstat are provided in parentheses. These SNPs represent genomic variants that either have some polygenic influence on variation in maximum residency on Lake Winnipeg (Manitoba, Canada) walleye (*Sander vitreus*).

| | Overall SNPs | Non-Residency-Associated SNPs | Maximum Residency-Associated SNPs |
|---|---|---|---|
| $H_O$ | 0.137 | 0.128 | 0.162 |
| $H_S$ | 0.140 | 0.130 | 0.167 |
| $F_{IS}$ (95% CI) | 0.0187 (0.0149 - 0.0227) | 0.0151 (0.0096 - 0.0204) | 0.0262 (0.0208 - 0.0321) |

### 5.4.5 Functional Analyses

The 5,404 clumped SNPs used to generate polygenic scores were associated with 2,389 annotated genes, from which 52 biological process 2018, 12 molecular function 2018, and 8 cellular component 2018 GO terms were identified with EnrichR and Revigo (Fig 5.7). From the overall GO terms, 31 biological process 2018, 8 molecular function 2018, and 6 cellular component 2018 non-redundant terms were identified with Revigo (Fig 5.7). Forebrain neuron differentiation (GO:0021879), positive regulation of glutamate receptor signaling pathway (GO:1900451), metanephric nephron tubule development (GO:0072234), and positive regulation of cyclin-dependent protein serine/threonine kinase activity involving G1/S transition of mitotic cell cycle (GO:0031659) each were biological process 2018 terms with the highest combined score across all terms. The gene *greb1* had two SNPs associated with maximum residency within a 20 kb window of it while no maximum residency-associated SNPs were identified near *rock1*.

## 5.5 Discussion

Using genomic data, 81.6% of variation in maximum residency tendency among Lake Winnipeg walleye with receiver array detections ≥180 days apart was explained. That no genomic regions of large effect were identified is in contrast to other research that found small genomic regions that influence seasonal run timing in Chinook salmon (*Oncorhynchus tshawytscha*) ecotypes (Thompson et al., 2020). However, spring versus fall-run timing are discrete phenotypes, while maximum residency used in the present study was a continuous phenotype. Additive genetic effects across many loci could be expected for continuous phenotypes under an infinitesimal model first proposed by (Crouch & Bodmer, 2020; Fisher, 1918). A perspective consistent with the hypothesis that habitat preference is polygenic, which is supported by the present results. Moreover, the region of strongest effect found in Thompson et al. (2020) predicted 85% of the variance in freshwater entry timing, a value comparable to the 81.6% identified for residency here. Future research may test whether this 80 – 85% variation in

movement phenotypes explained by genomics is a consistent pattern, or the result of systems with unusually strong genomic influences.

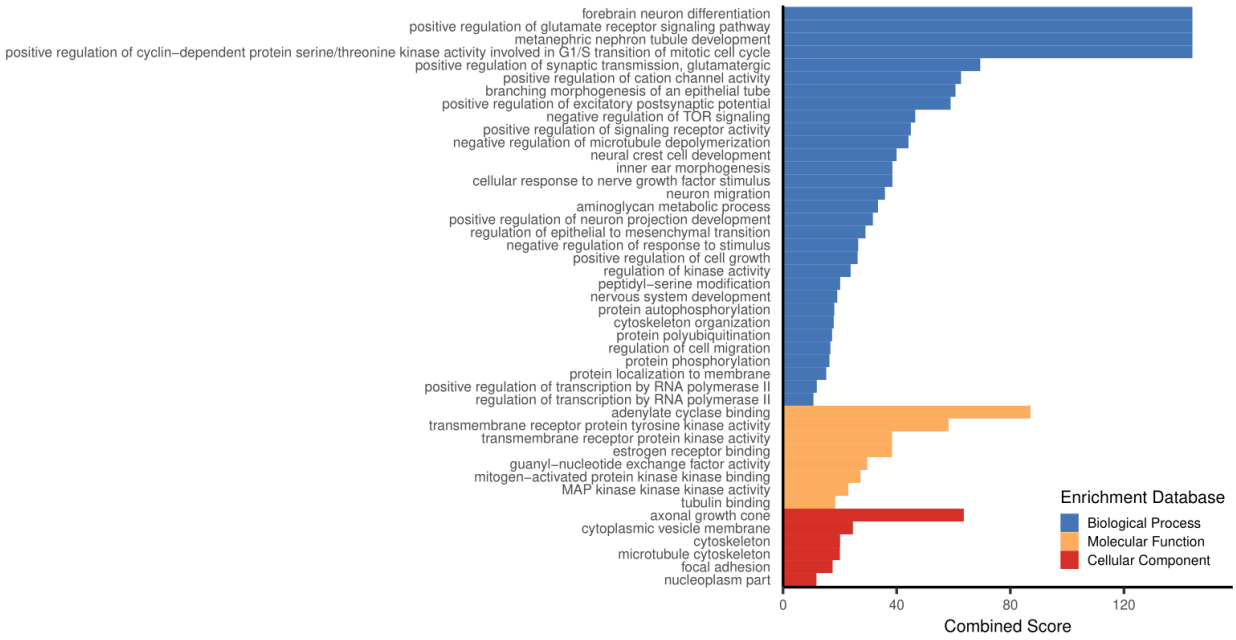


**Fig 5.7** Gene set enrichment analysis for single nucleotide polymorphisms (SNPs) associated with maximum residency at a threshold of $p<0.40$ for Lake Winnipeg (Manitoba, Canada) walleye (*Sander vitreus*). Here, 5,404 SNPs clumped by significance and linkage disequilibrium were identified as possibly associated with maximum residency. These SNPs were associated with genes in windows 15 kilobase pairs upstream and 5 kilobase pairs downstream of a gene. Genes associated with SNPs were analyzed for gene set enrichment via gene ontology (GO) terms in the Biological Process 2018, Molecular Function 2018, and Cellular Component 2018 databases in EnrichR. Non-redundant GO terms with identified with Revigo, using combined scores from EnrichR. Combined scores are used for visualization because they incorporate both significance ($q<0.05$ for all GO terms here) and magnitude of deviation from an expected rank, given Fisher's exact test.

Maximum residency represents a coarse movement phenotype that was tractable for analysis in the present study. While numerous ecologically informative movement metrics were available for analysis, movement phenotypes chosen were limited by statistical power in the genomic data available. For instance, Bayesian state-space modeling and kernel smoothing were used to generate accurate and fine-scale movement paths (Munaweera et al., 2021), but thousands to hundreds of thousands of genomic samples would be needed to describe the genetic architecture underlying subtle phenotypes. Typical wildlife studies do not exist at that level of statistical power, although an analogy can be drawn from human behavioural genetics where the variance explained for depression scores increased five-fold when comparing an under-powered association test with $n \approx 18{,}000$ individuals and a stronger test with $n \approx 163{,}000$ individuals (8% versus 46% variance explained, respectively; Colodro-Conde et al., 2018). The maximum residency phenotype was of ecological interest, as well, because walleye is a mobile species with broad seasonal migrations and Lake Winnipeg walleye consist of a metapopulation divided into two slightly differentiated subpopulations (Chapters 2 & 4; Backhouse-James & Docker, 2012; Turner et al., 2021; Watkinson & Gillis, 2005). As such, the degree to which fish interact with the overall metapopulation or stay within a single ecological zone can have large effects on population emigration-immigration dynamics, along with habitat use and population persistence (McMahon & Matter, 2006).

The implications of the genomic influence on residency for fish ecology and fisheries dynamics are striking, because any genomic influence on residency and habitat selection suggests possible genomic mechanisms to influence walleye behaviour (Healy et al., 2018). For instance, fisheries-induced selection against large fish has been hypothesized to lead to dwarf walleye morphotypes in Lake Winnipeg (Moles et al., 2010). Further, signatures of selection using genomic data have been identified in Lake Winnipeg walleye using both outlier SNPs in RNA data and haplotype frequency differences in DNA data (see Chapters 2 & 4). Selection is therefore likely acting on Lake Winnipeg walleye, and may therefore influence residency. For example, if walleye of high residency in an area of high fishing pressure are caught more often, or if walleye with a low residency tendency are more likely to be caught as they travel between

ecological zones and are exposed to gillnets. Either hypothetical situation would lead to fisheries-induced directional selection on residency in the population, either for or against residency, but both would lead to fisheries-induced changes to walleye behaviour. Genetic drift, another important force in population genetics, may also affect movement. For example, in small populations of walleye, random fluctuations in allele frequency may cause random fluctuations in residency tendencies changing how fish in small populations use habitats in their waterbodies. Another implication of the effect of genomic variation on residency is in the context of gene flow, admixture, and introgression between populations. Walleye from adjacent Lake Manitoba that may have recently admixed with primarily north basin individuals in Lake Winnipeg (see Chapter 4 for details), and visualizing Lake Manitoba ancestry in conjunction with polygenic scores and maximum residency revealed a possible increase in north basin residency from ancestry. While the inability for the polygenic model to predict residency for individuals with increased Lake Manitoba ancestry was likely because of the low sample size for those individuals, maximum residency was consistently higher than expected given their polygenic scores and the simple regression used, especially for individuals sampled from the Dauphin River of Lake Winnipeg. It is therefore possible that gene flow can change residency tendencies and habitat selection in walleye, a potentially significant finding given widespread stocking programs in the species (Bootsma et al., 2020).

An important limitation in the polygenic model presented here was shown by identifying polygenic scores in individuals with receiver array detections <180 days apart. Polygenic scores for those individuals were tightly clustered around the mean for scores from the original set of individuals, indicating a possible inability by the model to identify fish with high or low residency values and therefore, a limitation of the model to be generalizable to other Lake Winnipeg walleye. In other words, the model likely has limited predictive power for other Lake Winnipeg walleye because it is unlikely that polygenic scores for the originally excluded individuals are so tightly clustered if they represent a group undifferentiated from the individuals included in the polygenic model. Therefore, the model is ungeneralizable. This lack of generalizability may be due to the low sample size used to develop the polygenic model ($n$=113). Similar models used in human genetics can be ungeneralizable within populations even when using thousands of samples, and uninformative between populations under any sample size if samples were taken from a single population (Choi et al., 2020; Mostafavi et al., 2020). Further,

polygenic models may be ungeneralizable outside of specific groups studied no matter the sample size used (Mostafavi et al., 2020). Researchers interested in developing predictive models of fish movement may need to partition fish into groups of interest and create polygenic models for each group, as opposed to an overall model as was attempted in the present study. For instance, predictive polygenic models for walleye movement in Lake Winnipeg may be partitioned into north basin and south basin models that also incorporate intra-basin ecological characteristics, but these models would need to be developed with sufficient sample sizes for the genomic associations.

Gene set enrichment revealed several non-redundant GO terms representing collections of genes that may affect walleye maximum residency in Lake Winnipeg. Terms related to nephron tubule development and inner ear morphogenesis were consistent with the renal and inner ear development functions of the gene *greb1*, identified as possibly important for movement both in this study and in research on salmon migration timing (Koch & Narum, 2020; Thompson et al., 2020). Beyond connections with existing data, functional results from polygenic models and SNPs in 20 kb windows around genes are largely speculative. However, the gene set enrichment analysis is used here as a hypothesis-generating tool. Just as the gene *greb1* was identified in several other studies to be associated with migration timing (Koch & Narum, 2020; Thompson et al., 2020), along with associations with maximum residency in the present data, shared patterns of significance across studies may elucidate functional pathways underlying movement phenotypes in the future.

Despite a widespread lack of predictive ability, the utility for polygenic modeling to describe the variation in a phenotype explained by genomic variation holds potential for studies connecting ecology and genetic diversity (Healy et al., 2018; Hughes et al., 2008; Mostafavi et al., 2020). Polygenic modeling involves identifying a set of SNPs that may influence a trait, a process undertaken in the present study to find candidate enrichment terms for a functional basis of walleye movement. However, these sets of SNPs were also used to describe summary statistics of genetic diversity that may provide a direct link between genetic variation, a phenotype, and broader ecological metrics. In the present data, heterozygosity and inbreeding ($H_O$, $H_S$, and $F_{IS}$) were higher among SNPs associated with residency than in SNPs not associated with residency, a seemingly contradictory result given the original definition of

inbreeding partly characterized by a reduction of heterozygosity (Wright, 1950). One study identifying similar concordant patterns of high inbreeding and genetic variation in palm trees of genus *Chamaedorea* attributed them to a low and variable number of mating pairs (Luna et al., 2007). This may suggest for the present study a potential relationship between walleye residency, genetic variation, and breeding behaviour. Unexpectedly high variation in certain genomic regions compared to others is also characteristic of balancing selection, which is plausible if different residency tendencies are maintained over time in walleye (Hiwatashi et al., 2010; Tian et al., 2002). While the present study was limited in its capacity to describe walleye habitat selection in relation to residency tendency and genomic variation, the demonstrated ability to investigate genetic diversity in the context of an ecologically significant trait highlights the potential to link genetic diversity to ecosystem-level variables.

### 5.5.1 Conclusions

A polygenic model predicted 81.6% of the variation in maximum residency tendency for Lake Winnipeg walleye. However the model is likely not generalizable because of low sample size and the few biological groups present in the data used here. Nevertheless, the strong genomic basis for maximum residency holds several implications for walleye biology and management, such as the potential for selection to modify walleye movement in the future, or the potential for admixture to change movement tendencies, and thus habitat use. Functional analyses revealed patterns between the gene *greb1* and the residency-associated SNPs in this study, which was consistent with patterns observed in salmon migration timing research (Koch & Narum, 2020; Thompson et al., 2020). Last, polygenic modeling underscores the importance of genetic variation in fish biology and ecosystems in general, as the present results show how genetic variation in walleye movement is associated with the use of one or more ecological zones in Lake Winnipeg. While other research has discussed the connections between genetic variation and higher levels of biological organization in ecology (Hughes et al., 2008), even non-predictive polygenic models may provide useful information about the ecological consequences of genetic diversity via ecologically-significant traits.

## 5.7 Acknowledgements

Michael E. Yusishen was second reader for aging walleye scales used as covariates for association testing. Darren M. Gillis and Saman Muthukumarana contributed to conceiving of the movement component of the study. I thank Eva C. Enders, Douglas A. Watkinson, Colin Charles, Colin Kovachik, Douglas Leroux, Nicole Turner, Michael Gaudry, Sarah Glowa, and Emily Barker for their role in sampling the walleye. Eva C. Enders and Douglas A. Watkin also contributed to conceiving of the original collaboration. Colin Charles and Evelien de Greef assisted in making a map of Lake Winnipeg. Many analyses in this manuscript were enabled by our opportunity to use computing resources provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). This work was supported by a Fisheries and Oceans Canada Ocean and Freshwater Science Contribution Program Partnership Fund grant awarded to J.R.T., K.M.J. and Darren Gillis, and Natural Sciences and Engineering Research Council of Canada Discovery Grants awarded to K.M.J. (#05479) and J.R.T. (#06052). Work by J.R.T. is also supported by the Canada Research Chairs program (#223744) and the Faculty of Science, University of Manitoba (#319254).

## Chapter 6. Thesis Conclusions

Biochemistry and genomics have been linked to movement in a freshwater fish. Walleye (*Sander vitreus*) in Lake Winnipeg (Manitoba, Canada) were established as a metapopulation with multiple contributing subpopulations, information which provided an evolutionary background against which to understand movement in the fish (Chapters 2 & 4). Despite being a single population, walleye of the north basin in Lake Winnipeg were found to show increased levels of protein degradation relative to the narrows and south basin (Chapter 3). This increased protein degradation is consistent with further distances needed to travel for north basin fish to reach spawning sites, although connections between protein metabolism and movement will need to be verified. Subsequent movement studies identified south basin walleye as more likely to cross between basins than north basin individuals (Chapter 2; Turner et al., 2021). Therefore, increased protein degradation may be indicative of movement directly prior to sampling, as opposed to movement overall.

Gene flow between Lake Manitoba (Manitoba, Canada) and Lake Winnipeg walleye may have started up to 1,219 years ago, following a period of isolation between the populations (Chapter 4). Gene flow was estimated to be low, with approximately 0.0027 walleye traveling from Lake Manitoba to Lake Winnipeg per year. This finding was interesting, given broad seasonal movements observed in the species. An unknown migration barrier therefore exists between Lake Manitoba and Lake Winnipeg walleye, despite a river and small lake system connecting the two larger lakes. Moreover, gene flow was not enough to erase signatures of selection in each lake. Walleye may therefore be locally adapted to Lake Manitoba and Lake Winnipeg, despite a likely recent, in geologic time, divergence and despite gene flow.

Within Lake Winnipeg, 81.6% of variation in a tendency to stay within one ecological zone or move between zones was explained with genomics (Chapter 5). Even if this value is an overestimate, a genomic tendency to remain or move between ecological zones has deep implications for the evolution of fish movement, the capacity for natural selection to change fish movement, and for genomic mechanisms that may influence movement. A hypothesis that selection may affect movement is significant in the context of fisheries-induced selection, where human action may change movement tendencies in populations across generations. Looking back in time, selection and movement may be used to study the evolution of movement and the

evolutionary histories of different movement tendencies. An important limitation of this result was its probable lack of generalizability in the specific genomic variants used to develop the underlying model. Prior research has linked migration timing in anadromous fish with specific genomic variants, with similar variation in the phenotype explained (Thompson et al., 2021). However, my thesis provides novel evidence for a genomic influence on habitat selection in an obligate freshwater fish.

Taken together, these results show the utility of adopting multiple perspectives of fish movement (i.e., gene flow (evolutionary), maximum residency (a whole organism's position in space)) for understanding a system. Future research may connect local adaptation to movement phenotypes, with implications for abiotic and biotic factors shaping the evolution of movement tendencies via selection. Understanding how protein degradation and amino acid oxidation interact has the potential to provide new avenues for studying energy use and fish movement, and the data in wild fish presented here provide a springboard for future research. For instance, laboratory-controlled experiments may help identify the timescales that movement may affect blood metabolite profiles. Such research would contribute to developing biomarkers of fish movement in the wild, for applications in possibly non-lethal assays of prior movement before sampling, and in revealing how different kinds of movement affect fishes' energy use and protein metabolism. Multi-scale movement research has great potential more generally; movement research that integrates different approaches to movement and underlying factors (e.g., physiology, genetics), similar to what has been conducted as part of this thesis, will continue to offer deep insights for conservation and fundamental biology.

# Appendix

## A.1 Supplemental Material for Chapter 2: Genomic Signals Found Using RNA Sequencing Show Signatures of Selection and Subtle Population Differentiation in Walleye (*Sander vitreus*) in a Large Freshwater Ecosystem.

### A.1.1 Supplemental Methods

#### A.1.1.1 RNA Extraction and Sequencing

Specific modifications to the RNeasy Plus Mini Prep Kit (QIAGEN, Venlo, Netherlands) extraction protocol are provided here. Samples were homogenized at 50 rotations per second for 6 min, using one metal bead in a TissueLyser LT (QIAGEN). Lysate was centrifuged for 5 min at 12,775 x g to force tissue to the bottoms of tubes, which I avoided pipetting in further steps. 350 µL of the homogenized lysate was transferred to the gDNA Eliminator spin column and centrifuged for 60 s, the flow-through was added 350 µL of 70% ethanol. 700 µL of the sample were transferred to a RNeasy spin column. All subsequent centrifuge steps were performed for a minimum of 60 s at 12,775 x g. During the Buffer RW1 and Buffer RPE wash steps, tubes were inverted prior to centrifuging to wash the upper portions of the spin columns. After centrifuging, while discarding flow-through from the wash buffers, collection tubes were dabbed dry on paper towels prior to the next step. RNase-free $H_2O$ was heated to 60 °C and 30 µL was used for elution, and incubated on the spin column membrane for at least 1 min prior to the final centrifugation at 12,775 x g.

#### A.1.1.2 RNA SNP Calling

The protocol used for following the SuperTranscripts pipeline (Davidson et al. 2017) is provided here. To call single nucleotide polymorphisms (SNPs) from aligned reads, Picard version 2.18.9 (Broad Institute 2018) was used to add read groups to aligned files, sort reads by coordinate, index read files, then hard-clipped portions of a read that overlapped with regions

outside of a gene. I used FreeBayes version 1.2.0, a haplotype-based program, to call SNPs from the processed file (Garrison & Marth 2012). An index of the transcriptome was created using Samtools version 1.9 (Li et al. 2009), then on a computing cluster (https://www.westgrid.ca/) the maximum number of open file descriptors was set to 2024 (ulimit -n 2024) and the maximum stack size to 102400 (ulimit -s 102400) prior to running FreeBayes in parallel over 48 threads with default options. Scripts used in the analyses for this manuscript are provided at https://github.com/BioMatt/Walleye_RNAseq.

### A.1.2 Supplemental Results

**Table A.1.1** Information for individual Lake Winnipeg walleye (*Sander vitreus*) collected in this study. Sample ID represents sample identifiers used in bioinformatics and analyses. Location represents spawning sites at which fish were collected, with the Red River in the south basin, Matheson Island in the channel connecting two basins, and the Dauphin River representing the north basin. Sex as identified during sub-lethal sampling is provided, along with year collected. The number of reads are total reads for each sample sequenced on an Illumina NovaSeq 6000 (Illumina, San Diego, California, USA), and % uniquely mapped reads is the percent of reads sequenced mapped Star to the Lace-constructed transcriptome, and thus used in subsequent analyses. Adegenet was used to generate 2- and 3-cluster population reassignments.

| Sample ID | Location | Sex | Year Collected | Number of Reads | % Reads Uniquely Mapped | 2 Clusters | 3 Clusters |
|---|---|---|---|---|---|---|---|
| Wall-002 | Red River | F | 2017 | 44,434,267 | 81.71 | 1 | 3 |
| Wall-005 | Red River | F | 2017 | 43,257,271 | 81.29 | 1 | 3 |
| Wall-006 | Red River | F | 2017 | 48,288,672 | 81.03 | 1 | 3 |
| Wall-008 | Red River | F | 2017 | 52,313,120 | 80.69 | 1 | 3 |
| Wall-011 | Red River | F | 2017 | 48,682,209 | 81.33 | 1 | 3 |
| Wall-016 | Red River | F | 2017 | 47,412,164 | 81.31 | 1 | 3 |
| Wall-019 | Red River | F | 2017 | 38,206,953 | 80.60 | 1 | 3 |
| Wall-020 | Red River | F | 2017 | 58,283,154 | 80.91 | 1 | 3 |
| Wall-171 | Matheson Island | F | 2017 | 54,846,547 | 82.40 | 2 | 3 |
| Wall-172 | Matheson Island | F | 2017 | 47,903,022 | 81.80 | 1 | 3 |
| Wall-173 | Matheson Island | F | 2017 | 48,351,276 | 82.07 | 1 | 3 |
| Wall-174 | Matheson Island | F | 2017 | 55,749,392 | 81.47 | 1 | 3 |
| Wall-175 | Matheson Island | F | 2017 | 47,721,630 | 81.78 | 1 | 3 |
| Wall-176 | Matheson Island | F | 2017 | 46,802,684 | 82.15 | 1 | 3 |
| Wall-177 | Matheson Island | F | 2017 | 41,243,681 | 82.00 | 1 | 3 |
| Wall-178 | Matheson Island | F | 2017 | 39,125,672 | 81.39 | 2 | 3 |
| Wall-182 | Dauphin River | F | 2017 | 40,569,210 | 81.88 | 2 | 2 |
| Wall-184 | Dauphin River | F | 2017 | 43,602,697 | 81.08 | 1 | 3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Wall-185 | Dauphin River | F | 2017 | 39,405,325 | 82.04 | 2 | 2 |
| Wall-187 | Dauphin River | F | 2017 | 46,263,666 | 81.34 | 1 | 3 |
| Wall-188 | Dauphin River | F | 2017 | 44,367,660 | 82.12 | 2 | 2 |
| Wall-193 | Dauphin River | F | 2017 | 41,158,198 | 81.47 | 1 | 3 |
| Wall-196 | Dauphin River | F | 2017 | 46,924,282 | 81.53 | 2 | 2 |
| Wall-197 | Dauphin River | F | 2017 | 39,784,120 | 81.75 | 2 | 2 |
| Wall-210 | Red River | F | 2018 | 44,613,544 | 81.76 | 1 | 1 |
| Wall-212 | Red River | F | 2018 | 44,117,274 | 81.99 | 1 | 1 |
| Wall-213 | Red River | F | 2018 | 43,396,269 | 82.50 | 1 | 1 |
| Wall-214 | Red River | F | 2018 | 52,719,989 | 81.76 | 1 | 1 |
| Wall-215 | Red River | F | 2018 | 44,762,039 | 81.71 | 1 | 1 |
| Wall-216 | Red River | F | 2018 | 36,008,614 | 81.31 | 1 | 1 |
| Wall-217 | Red River | F | 2018 | 48,005,006 | 81.55 | 1 | 1 |
| Wall-218 | Red River | F | 2018 | 44,721,946 | 81.36 | 1 | 1 |
| Wall-290 | Matheson Island | F | 2018 | 48,612,467 | 81.44 | 1 | 1 |
| Wall-292 | Matheson Island | F | 2018 | 42,865,563 | 81.46 | 1 | 1 |
| Wall-294 | Matheson Island | F | 2018 | 49,734,421 | 81.36 | 1 | 1 |
| Wall-295 | Matheson Island | F | 2018 | 46,729,515 | 81.87 | 1 | 1 |
| Wall-296 | Matheson Island | UN | 2018 | 39,870,540 | 81.08 | 1 | 1 |
| Wall-297 | Matheson Island | F | 2018 | 36,812,535 | 80.63 | 1 | 1 |
| Wall-301 | Matheson Island | F | 2018 | 50,070,079 | 80.58 | 1 | 1 |
| Wall-302 | Matheson Island | F | 2018 | 44,648,079 | 81.38 | 1 | 1 |
| Wall-343 | Dauphin River | UN | 2018 | 42,217,434 | 81.45 | 1 | 1 |
| Wall-344 | Dauphin River | F | 2018 | 50,429,946 | 80.99 | 2 | 2 |
| Wall-345 | Dauphin River | F | 2018 | 38,506,818 | 82.02 | 2 | 2 |
| Wall-349 | Dauphin River | UN | 2018 | 41,990,932 | 80.54 | 2 | 2 |
| Wall-350 | Dauphin River | UN | 2018 | 40,341,008 | 81.33 | 2 | 2 |
| Wall-353 | Dauphin River | F | 2018 | 44,956,582 | 79.62 | 1 | 1 |
| Wall-354 | Dauphin River | F | 2018 | 37,591,445 | 81.43 | 2 | 2 |
| Wall-355 | Dauphin River | F | 2018 | 52,407,394 | 81.49 | 1 | 1 |

**Table A.1.2** Weir & Cockerham's Pairwise $F_{ST}$ calculated with hierfstat between the Red River in the South Basin, Matheson Island in the channel, and Dauphin River in the North Basin for all 48 walleye (*Sander vitreus*) sampled in both 2017 and 2018, with a data set of 38,732 SNPs that do not show a strong year effect (between-year $F_{ST} \leq 0.01$ for all SNPs used here). 95% confidence intervals are provided in parentheses.

| | **Red River (South Basin)** | **Matheson Island (channel)** | **Dauphin River (North Basin)** |
|---|---|---|---|
| **Red River** | - | 0.0018 (0.0013 - 0.0022) | 0.0072 (0.0067 - 0.0078) |
| **Matheson Island** | | - | 0.0050 (0.0044 - 0.0055) |
| **Dauphin River** | | | - |

**Table A.1.3** Gene ontology terms found using EnrichR, which represent genes that vary along a latitudinal gradient in Lake Winnipeg (PC1 in Fig 2.4). These terms represent 120 uniquely annotated genes with 386 outlier SNPs among them found with pcadapt ($q < 0.05$). Benjamini-Hochberg adjusted p-values and the number of genes represented in each term are provided. This analysis was performed using a set of 222,634 single nucleotide polymorphisms (SNPs) not filtered for Hardy-Weinberg Equilibrium or pruned for linkage disequilibrium, unlike the putatively neutral set of SNPs used for population structure analyses.

| **Gene Ontology Term** | **Adjusted *p*-value** | **Number of Genes** |
|---|---|---|
| purine ribonucleoside triphosphate binding (GO:0035639) | 0.017 | 10 |
| ATP binding (GO:0005524) | 0.017 | 8 |
| adenyl ribonucleotide binding (GO:0032559) | 0.021 | 8 |
| hydrogen-exporting ATPase activity, phosphorylative mechanism (GO:0008553) | 0.084 | 2 |

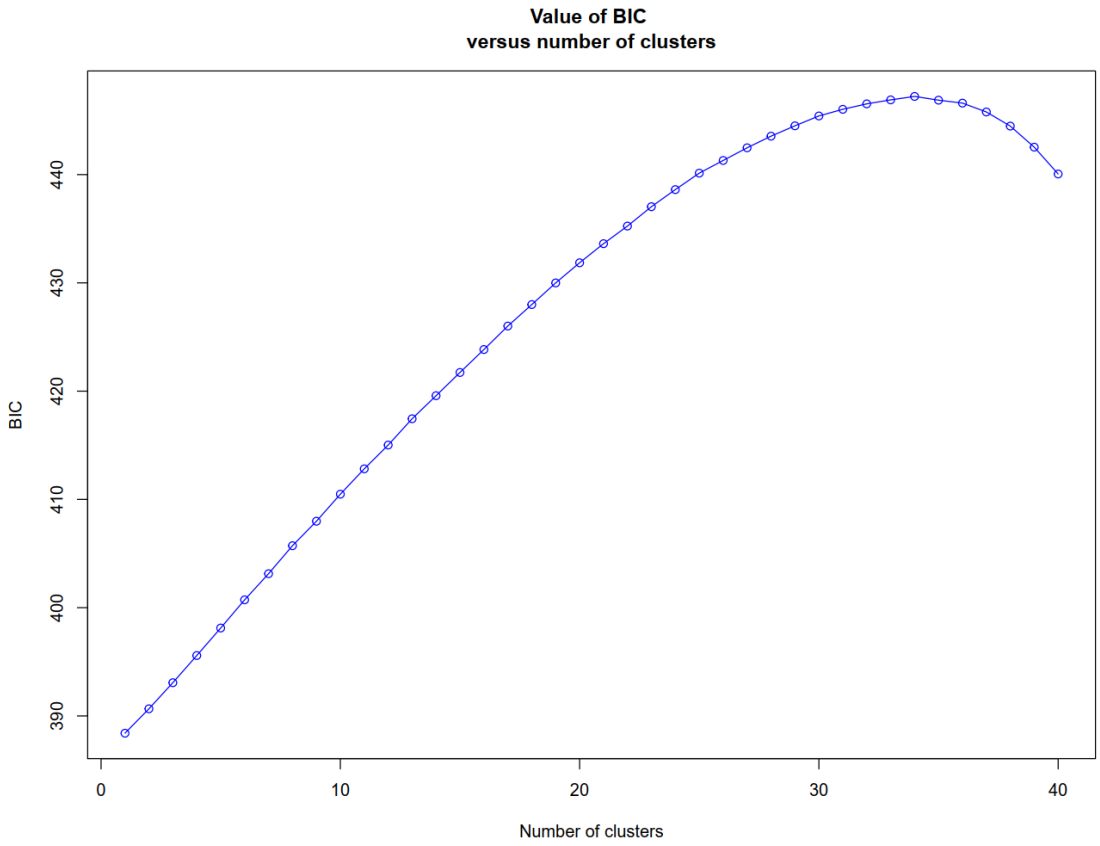


**Fig A.1.1** Bayesian Information Criterion (BIC) values versus number of clusters found when using Adegenet to explore population reassignment in all 48 walleye (*Sander vitreus*) collected in 2017 and 2018, over three sites representing a latitudinal gradient in Lake Winnipeg. These values were found using 52,372 Hardy-Weinberg Equilibrium filtered and linkage disequilibrium pruned, putatively neutral single nucleotide polymorphisms.

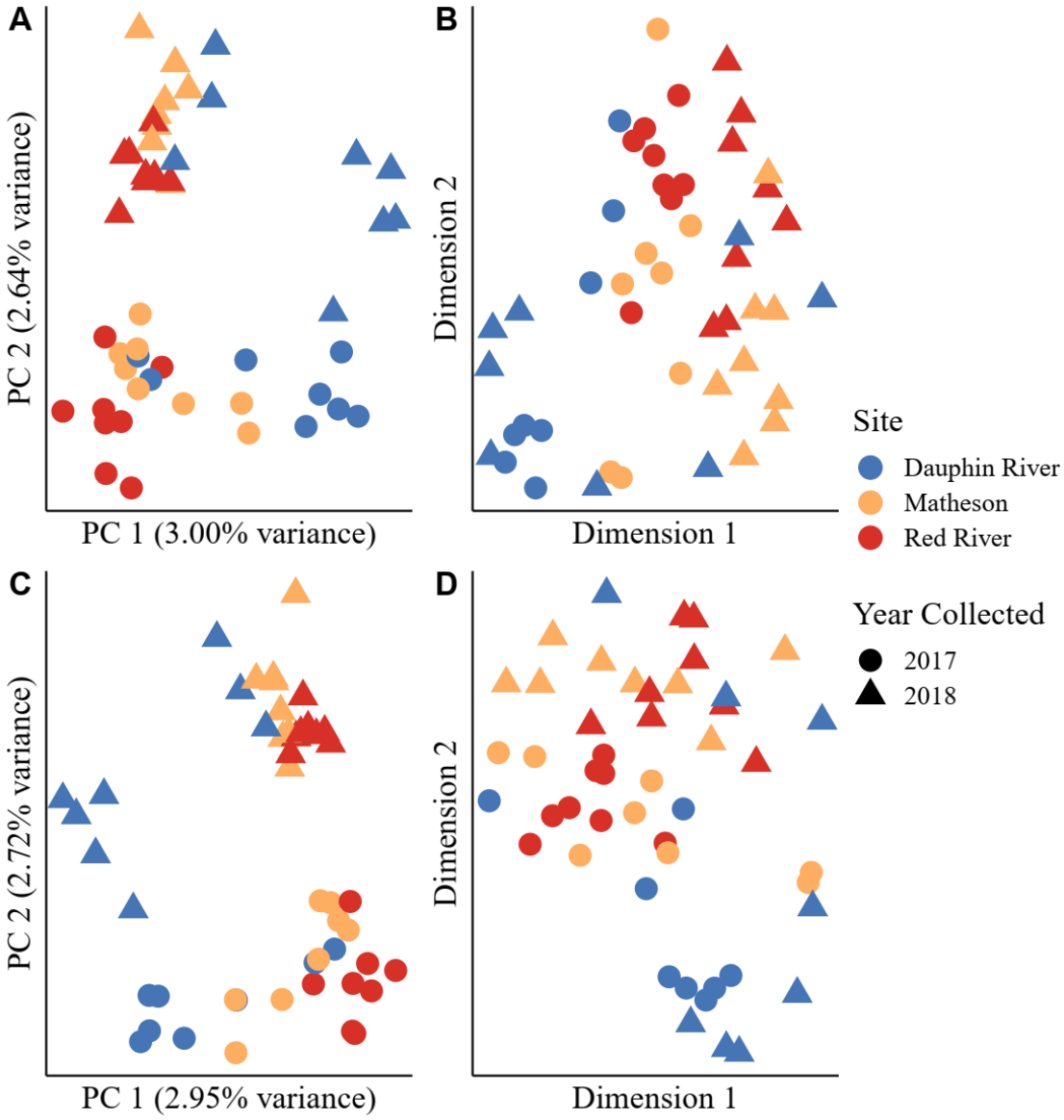


**Fig A.1.2** Principal Components Analysis (PCA) and t-SNE plots using 52,37 putatively neutral and all 222,634 high-quality SNPs performed using Adegenet and Rtsne, respectively. These analyses were performed on all 48 walleye (*Sander vitreus*) sampled from Lake Winnipeg, Manitoba, Canada, used in this study. Color shows site collected (red for Red River in the South Basin, yellow Matheson Island in the channel, and blue Dauphin River in the North Basin), circles showing walleye collected in 2017, and triangles showing walleye collected in 2018. A) Represents the PCA using neutral SNPs, B) represents a t-SNE using neutral SNPs, C) represents a PCA using all high-quality SNPs, and D) represents a t-SNE using all high-quality SNPs.

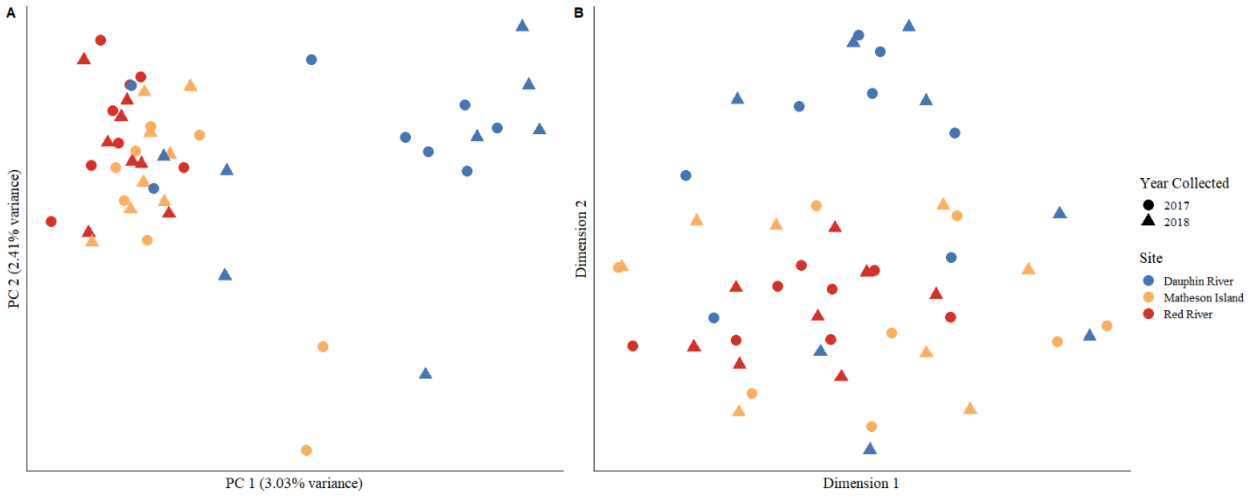


**Fig A.1.3** PCA and t-SNE plots using 38,732 SNPs that do not show a strong year effect (between-year $F_{ST} \leq 0.01$ for all SNPs used here) performed using Adegenet and Rtsne, respectively. These analyses were performed on all 48 walleye (*Sander vitreus*) used in this study. Color shows site collected (red for Red River in the South Basin, yellow Matheson Island in the channel, and blue Dauphin River in the North Basin), circles showing walleye collected in 2017, and triangles showing walleye collected in 2018. A) Represents the PCA, and B) represents the t-SNE.

## A.2 Supplemental Material for Chapter 3: Morphology and Blood Metabolites Reflect Recent Spatial and Temporal Differences Among Lake Winnipeg Walleye, *Sander vitreus*

### A.2.1 Supplemental Results

**Table A.2.1** Results from a linear model describing relative mass over time, relating $\log_{10}$ mass to $\log_{10}$ fork length and its interaction with year, site, and sex, with mesh size controlled for ($F$ = 2899, $p < 2.2 \times 10^{-16}$, $R^2 = 0.95$). The gillnet index data between the years 2009 and 2018 collected by the Government of Manitoba were used for this length-mass model. Only walleye (*Sander vitreus*) ≥375 millimeters in fork length are included.

| | *p*-value | eta squared |
|---|---|---|
| $\text{Log}_{10}$ Fork Length | < 0.05 | 0.25 |
| Year | < 0.05 | 0.0080 |
| Site | 0.11 | 0.0010 |
| Sex | 0.040 | 0.0 |
| Mesh Size | < 0.05 | 0.020 |
| $\text{Log}_{10}$ Fork Length * Year | < 0.05 | 0.0070 |
| $\text{Log}_{10}$ Fork Length * Site | 0.16 | 0.0010 |
| $\text{Log}_{10}$ Fork Length * Sex | 0.053 | 0.0 |

**Table A.2.2** *P*-values of pairwise estimated marginal means and trends between sites for two different linear models. One linear model uses walleye (*Sander vitreus*) ≥375 mm in fork length from the gillnet index and relates $\log_{10}$ mass to $\log_{10}$ fork length and site collected, sex, gillnet mesh size, and age ($F = 222$, $p < 2.2 \times 10^{-16}$, adjusted $R^2 = 0.97$). Data included in this model are from the gillnet index collected by the Government of Manitoba. The other linear model uses data from walleye sampled for metabolites (≥452 mm in fork length), and relates $\log_{10}$ mass to the interaction of $\log_{10}$ fork length and site collected, sex, and the interaction of $\log_{10}$ fork length and sex ($F = 86$, $p < 2.2 \times 10^{-16}$, adjusted $R^2 = 0.94$). *P*-values for means, representing differences in intercepts are above the diagonal and *p*-values for trends, representing differences in slopes are below the diagonal.

| | Gill Net Index | | | Metabolite | | |
|---|---|---|---|---|---|---|
| | Dauphin River | Matheson Island | Riverton | Dauphin River | Matheson Island | Red River |
| Dauphin River | - | 0.60 | 0.082 | - | 0.65 | 0.081 |
| Matheson Island | 0.024 | - | 0.60 | 0.037 | - | 0.040 |
| Riverton/ Red River | 0.16 | 0.41 | - | 0.060 | 0.35 | - |

**Table A.2.3** Results from a linear model relating $\log_{10}$ mass to the $\log_{10}$ fork length, basin collected, study of data origin (gillnet index or metabolite), sex, the interaction of $\log_{10}$ fork length and basin, and the interaction of $\log_{10}$ fork length and study ($F = 889$, $p < 2.2 \times 10^{-16}$, $R^2 = 0.97$). This model represents the length-mass relationship for walleye (*Sander vitreus*) collected for the gillnet index data by the Government of Manitoba and for metabolite information by the authors. These walleye were collected in 2017, and only fish ≥375 millimeters in fork length are included in this model. From the gillnet index data, only walleye from the Dauphin River, Matheson Island, and Riverton sites are included.

| | *p*-value | eta squared |
|---|---|---|
| $\log_{10}$ Fork Length | < 0.05 | 0.74 |
| Basin | 0.0010 | 0.020 |
| Study | 0.98 | 0.0 |
| Sex | 0.65 | 0.0 |
| $\log_{10}$ Fork Length * Sex | 0.62 | 0.0 |
| $\log_{10}$ Fork Length * Basin | < 0.05 | 0.019 |
| $\log_{10}$ Fork Length * Study | 0.90 | 0.0 |

**Table A.2.4** Results from linear models relating metabolite presence to $\log_{10}$ fork length and site collected, with a separate model for each metabolite. Category represents the conceptual framework used to classify the nine metabolites studied (see Figure 2 for details). Overall model *p*-values are provided under their respective metabolites. *P*-values and eta squared are reported for $\log_{10}$ fork length and site collected as independent variables within models. Briefly, methionine, tryptophan, and lysine represent essential amino acids. Dimethyl sulfone, kynurenine, and α-aminoadipic acid represent essential amino acid breakdown. Last, hydroxyproline, dimethylarginine, and methionine sulfoxide represent endogenous protein breakdown. The walleye (*Sander vitreus*) from which metabolites were measured for these linear models were *n*=39 individuals (≥452 mm in fork length) collected by boat electrofishing in 2017 from the Dauphin River, Matheson Island, and Red River representing the north basin, channel, and south basin of Lake Winnipeg, respectively. Metabolites were measured from whole blood.

| Category | Amino acid | Independent variable | *p*-value | eta squared |
|---|---|---|---|---|
| Essential Amino Acids | Methionine | $Log_{10}$ fork length | 0.74 | 0.0020 |
| | $p = 0.011$ | Site collected | 0.0070 | 0.25 |
| | Tryptophan | $Log_{10}$ fork length | 0.36 | 0.019 |
| | $p = 0.0071$ | Site collected | 0.032 | 0.17 |
| | Lysine | $Log_{10}$ fork length | 0.52 | 0.010 |
| | $p = 0.020$ | Site collected | 0.046 | 0.16 |
| Amino Acid Breakdown Markers | Dimethyl Sulfone | $Log_{10}$ fork length | 0.088 | 0.070 |
| | $p = 0.031$ | Site collected | 0.36 | 0.048 |
| | Kynurenine | $Log_{10}$ fork length | 0.80 | 0.002 |
| | $p = 0.070$ | Site collected | 0.092 | 0.13 |
| | α-Aminoadipic Acid | $Log_{10}$ fork length | 0.94 | 0.0 |
| | $p = 0.81$ | Site collected | 0.66 | 0.023 |
| Protein Degradation Metabolites | Hydroxyproline | $Log_{10}$ fork length | 0.69 | 0.002 |
| | $p = 1.3 \times 10^{-18}$ | Site collected | $< 2.2 \times 10^{-16}$ | 0.60 |
| | Dimethylarginine | $Log_{10}$ fork length | 0.97 | 0.0 |
| | $p = 0.0017$ | Site collected | 0.0020 | 0.30 |
| | Methionine Sulfoxide | $Log_{10}$ fork length | 0.47 | 0.011 |
| | $p = 0.0016$ | Site collected | 0.0080 | 0.24 |

## A.3 Supplemental Material for Chapter 4: Intrapopulation Signatures of Selection Despite Translocations in Walleye of Three North American Waterbodies

### A.3.1 Supplemental Methods

#### A.3.1.1 Synteny in Walleye and Yellow Perch Genomes

Because SNP relationships with gene locations were important for functional analyses of signatures of selection and the yellow perch genome was used for these walleye data, I investigated synteny between available walleye and yellow perch genomes using a pipeline published in Doerr & Moret (2018) (yellow perch genome, PLFA_1.0 at NCBI SRA BioProject reference #PRJNA514308; walleye genome, ASM919308v1 at NCBI SRA BioProject reference #PRJNA528354) (Feron et al. 2020). First, progressiveMauve v20150226 build 10 was used to create a backbone to anchor genomic blocks using both genomes (Darling et al. 2010). Then i-ADHoRe v3.0.01 was used to analyze synteny in colinear mode with default settings (Proost et al. 2012). Circos v0.69-8 was then used to visualize synteny as an ideogram, while scripts published in Doerr & Moret (2018) were used to generate both relaxed and weighted synteny scores, along with being useful for data conversion between prior steps (scripts accessed from https://github.com/danydoerr/large_syn_workflow) (for Circos, see Krzywinski et al. 2008; for synteny scores, Ghiurcuta & Moret 2014).

### A.3.2 Supplemental Results

**Table A.3.1** Negative log likelihoods and Akaike information criterion (AIC) values for different models of walleye (*Sander vitreus*) in Lake Manitoba and Lake Winnipeg (Manitoba, Canada) demography tested in δaδi. The most likely model (sec_contact_late_asym_mig_size2) was characterized by a split between walleye of Lake Manitoba and Lake Winnipeg, secondary contact, then gene flow after secondary contact. Population size changes were included prior to secondary contact, then again at secondary contact.

| Model | log-likelihood | AIC |
|---|---|---|
| sec_contact_late_asym_mig_size2 | -8231.28 | 16484.56 |
| sec_contact_varying_asym_mig_size2 | -8973.34 | 17972.68 |
| asym_mig_size | -8987.91 | 17991.82 |
| sym_mig | -9109.83 | 18227.66 |
| sec_contact_asym_mig_size | -9191.2 | 18398.4 |
| sec_contact_asym_mig | -9468.44 | 18948.88 |
| asym_mig | -9652.67 | 19315.34 |
| sec_contact_asym_mig_size2 | -9780.57 | 19583.14 |
| no_mig | -12769.14 | 25544.28 |

**Table A.3.2** Significant candidate regions for cross-population extended haplotype homozygosity (XP-EHH). The comparison represents the two populations used for comparing walleye (*Sander vitreus*) populations. Cedar Bluff Reservoir (Kansas, USA) represents an entirely stocked population, while Lake Manitoba and Lake Winnipeg (Manitoba, Canada) represent native populations with possible gene flow. Candidate genomic regions for XP-EHH were identified by analyzing 100 kilobase (kb) windows overlapping by 10 kb, in which at least three significant SNPs ($q < 0.05$) showing XP-EHH were found. Chromosome represents the chromosome from the yellow perch (*Perca flavescens*) reference genome that walleye DNA was aligned to. Start and end represent beginning and ending base pair positions in the corresponding chromosome, defining the candidate region for selection. Number of markers shows the number of single nucleotide polymorphisms identified in the corresponding genomic region.

| Comparison | Chromosome | Start | End | Number of Markers |
|---|---|---|---|---|
| Lake Winnipeg v Lake Manitoba | 4 | 15420000 | 15590000 | 9 |
| | 8 | 15260000 | 15450000 | 5 |
| | 8 | 15710000 | 15900000 | 6 |
| | 9 | 4430000 | 4610000 | 9 |
| | 9 | 8460000 | 8650000 | 4 |
| | 12 | 14270000 | 14460000 | 8 |
| | 12 | 14610000 | 14900000 | 8 |
| | 17 | 28530000 | 28740000 | 10 |
| | 24 | 8500000 | 8600000 | 3 |
| | 24 | 16030000 | 16220000 | 3 |
| Lake Winnipeg v Cedar Bluff Reservoir | 2 | 13800000 | 14000000 | 11 |
| | 4 | 17220000 | 17410000 | 6 |
| | 8 | 15260000 | 15450000 | 5 |
| | 18 | 19570000 | 19730000 | 6 |
| | 20 | 6280000 | 6470000 | 8 |

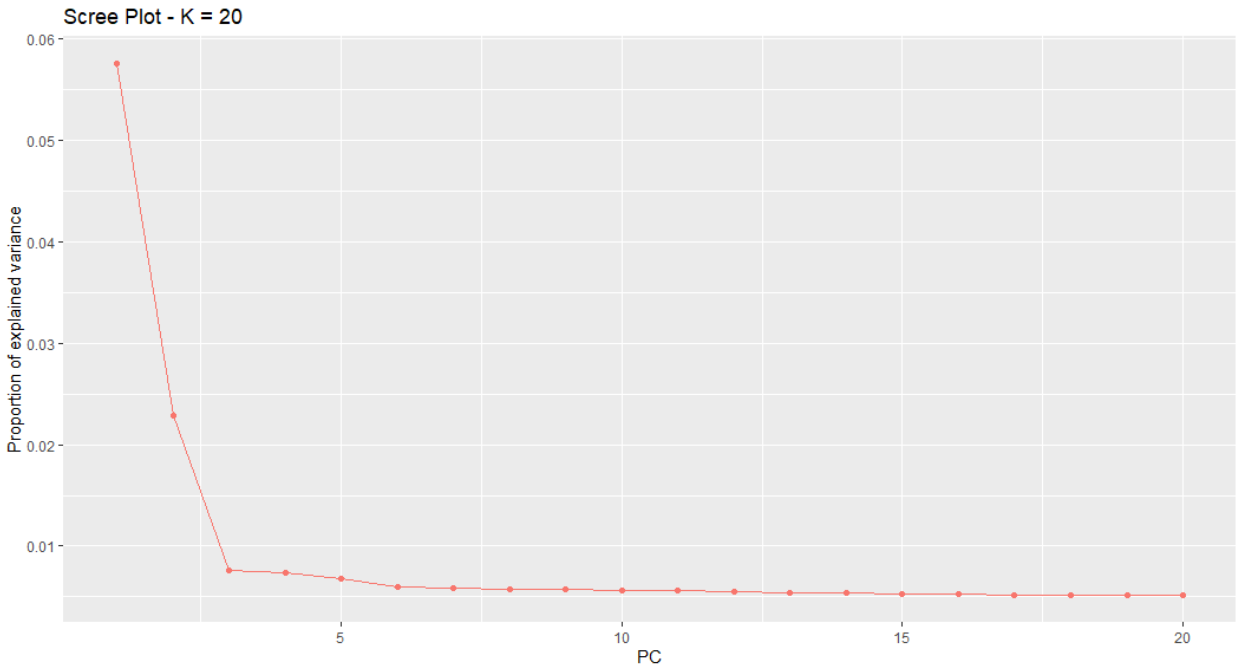


**Fig A.3.1** Scree plot for proportion of explained variance in principal components analysis using pcadapt for walleye (*Sander vitreus*). By Cattell's Rule, K=2 principal components explains a large proportion of variance in the data. These two principal components are used for subsequent analyses of outlier single nucleotide polymorphisms.

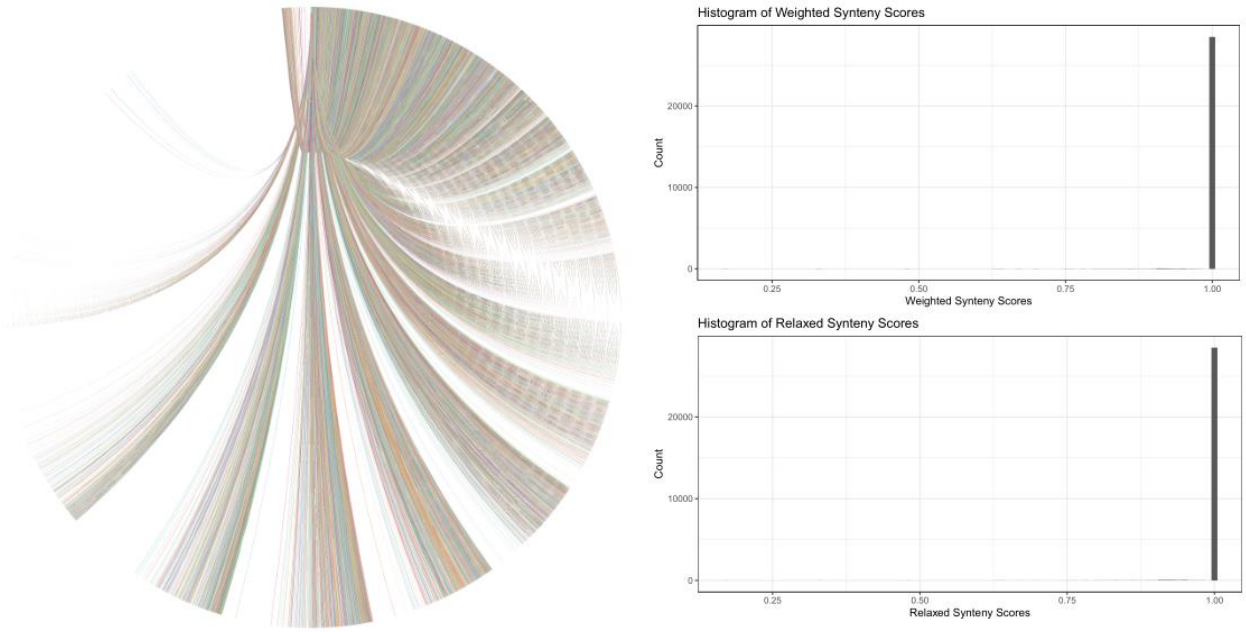


**Fig A.3.2** Synteny analyses between the yellow perch (*Perca flavescens*) and walleye (*Sander vitreus*) genomes. The Circos plot is based on synteny analyses by the programs progressiveMauve and i-ADHoRe. Relaxed and weighted synteny scores were analyzed from those alignments, where higher values near 1 indicate greater complete synteny between the genomes.

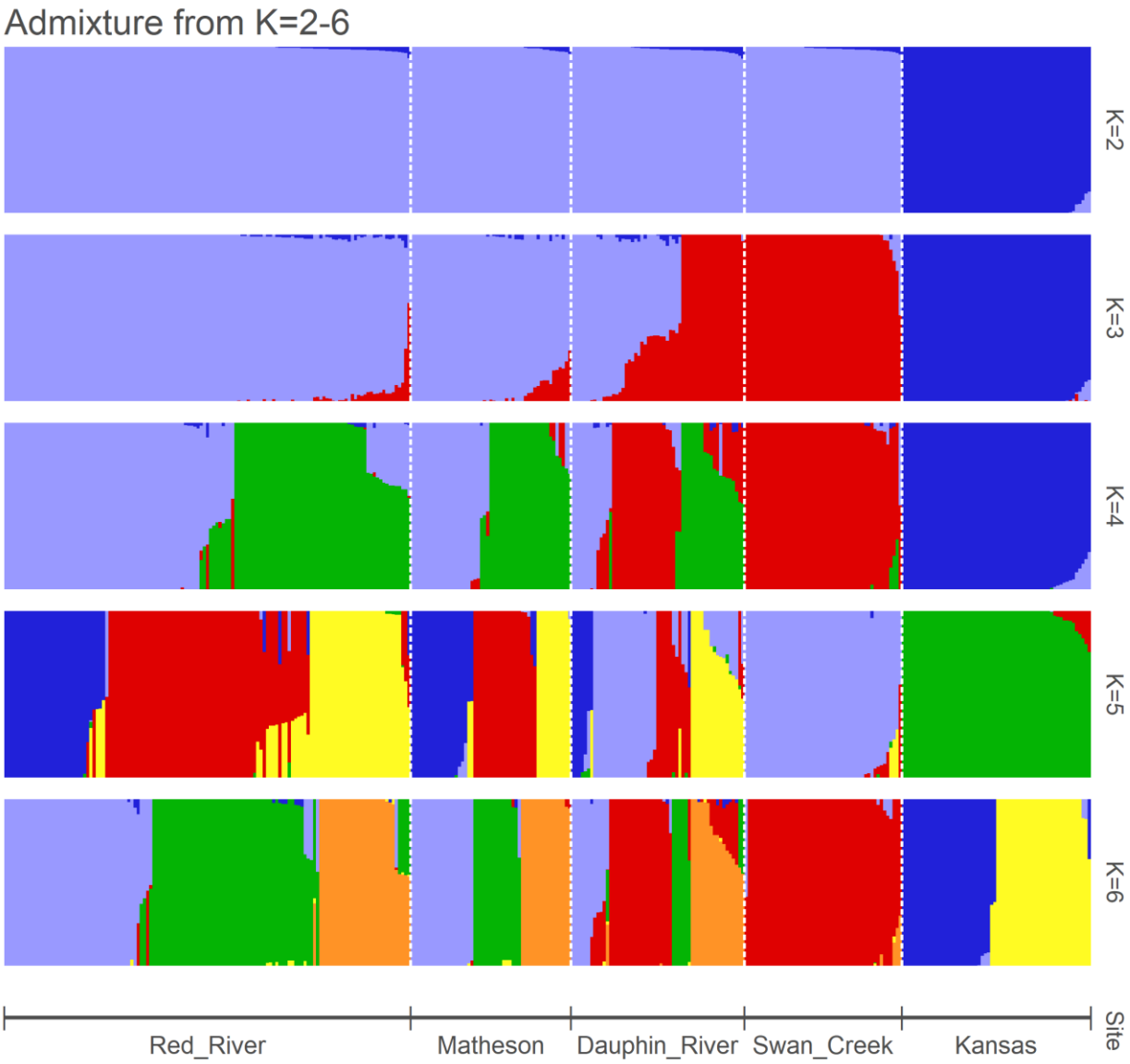


**Fig A.3.3** Admixture results for *n*=345 walleye (*Sander vitreus*) sampled from Cedar Bluff Reservoir (Kansas, USA), Lake Manitoba (Manitoba, Canada), and Lake Winnipeg (Manitoba, Canada). The Red River, Matheson Island, and Dauphin River sites represent south basin, narrows, and north basin sites in Lake Winnipeg. Swan Creek Hatchery represents walleye from Lake Manitoba. Cedar Bluff Reservoir walleye are labeled as originating from Kansas in the present visualization. K=2 through 6 groups were tested for admixture analysis and population assignment.

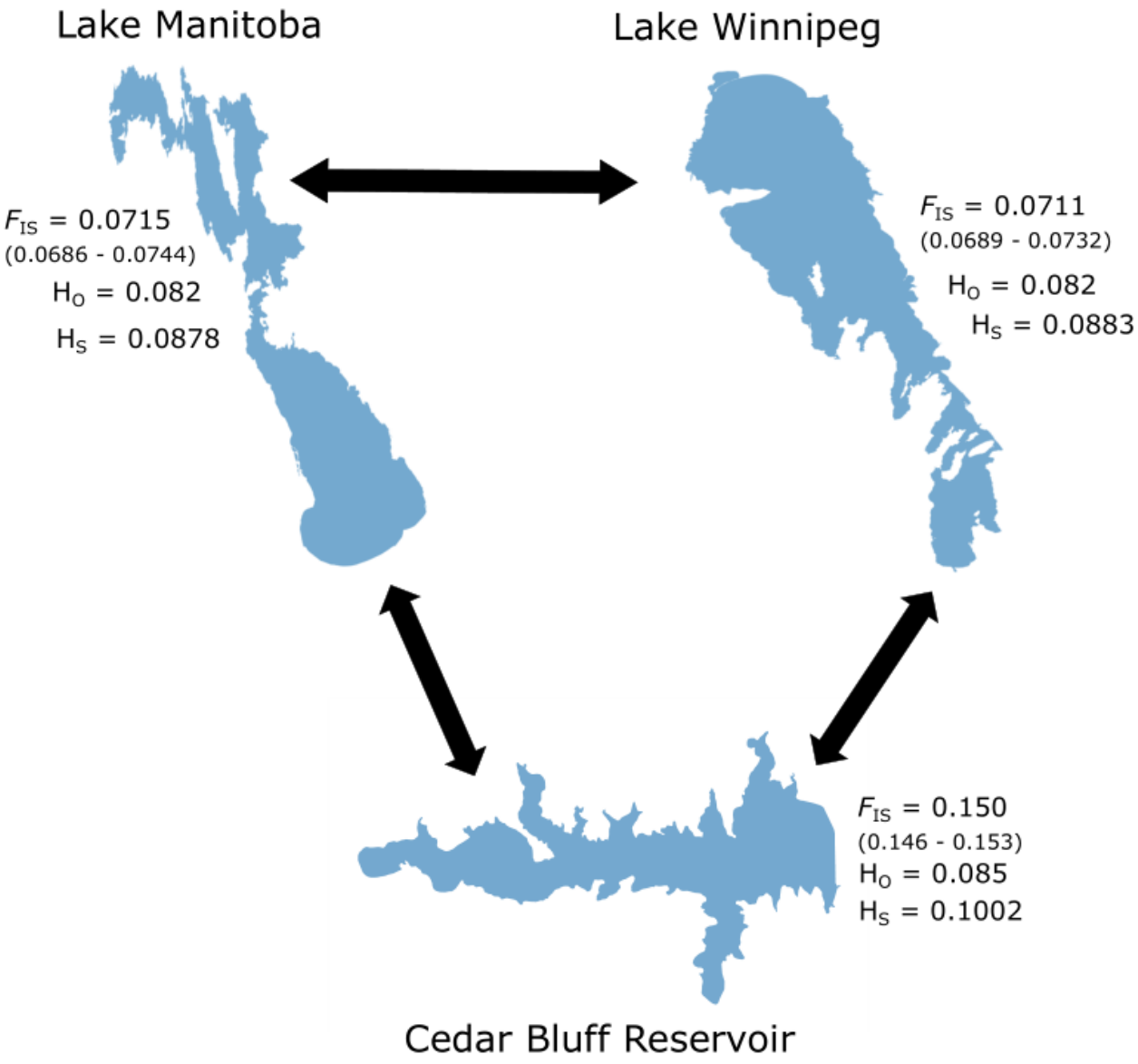


**Fig A.3.4** Inbreeding coefficients ($F_{IS}$), observed heterozygosity ($H_O$), and gene diversity ($H_S$) for three populations of walleye (*Sander vitreus*) from three waterbodies in North America. Cedar Bluff Reservoir (Kansas, USA) represents an entirely introduced walleye population of unknown origin, while Lake Manitoba and Lake Winnipeg (Manitoba, Canada) walleye are native and may experience gene flow between them. For $F_{IS}$, 95% confidence intervals are provided in parentheses.

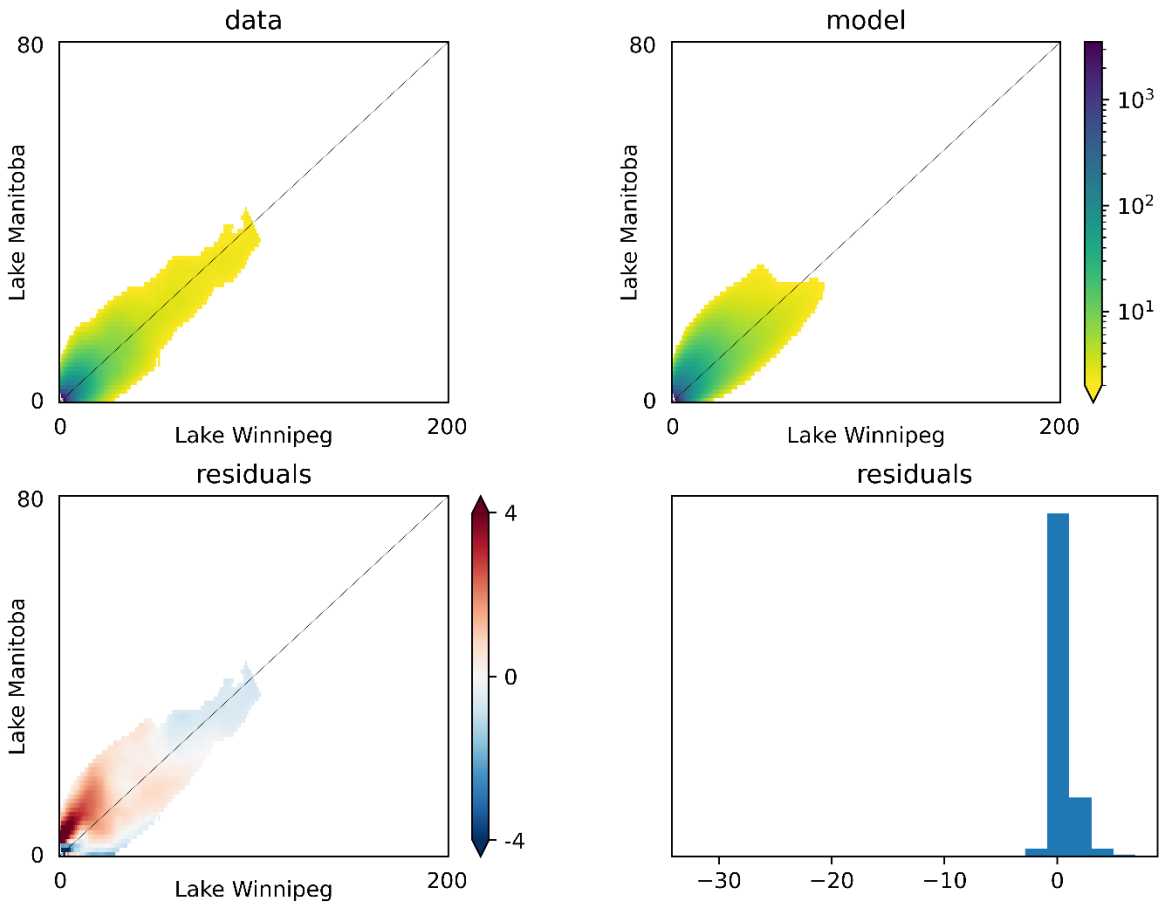


**Fig A.3.5** Visualization of site frequency spectra, residuals, and model fit for the best fit model in δaδi. The most likely model (sec_contact_late_asym_mig_size2) was characterized by a split between walleye (*Sander vitreus*) of Lake Manitoba and Lake Winnipeg (Manitoba, Canada), secondary contact, then gene flow after secondary contact. Population size changes were included at an initial population split, prior to secondary contact, then again at secondary contact.

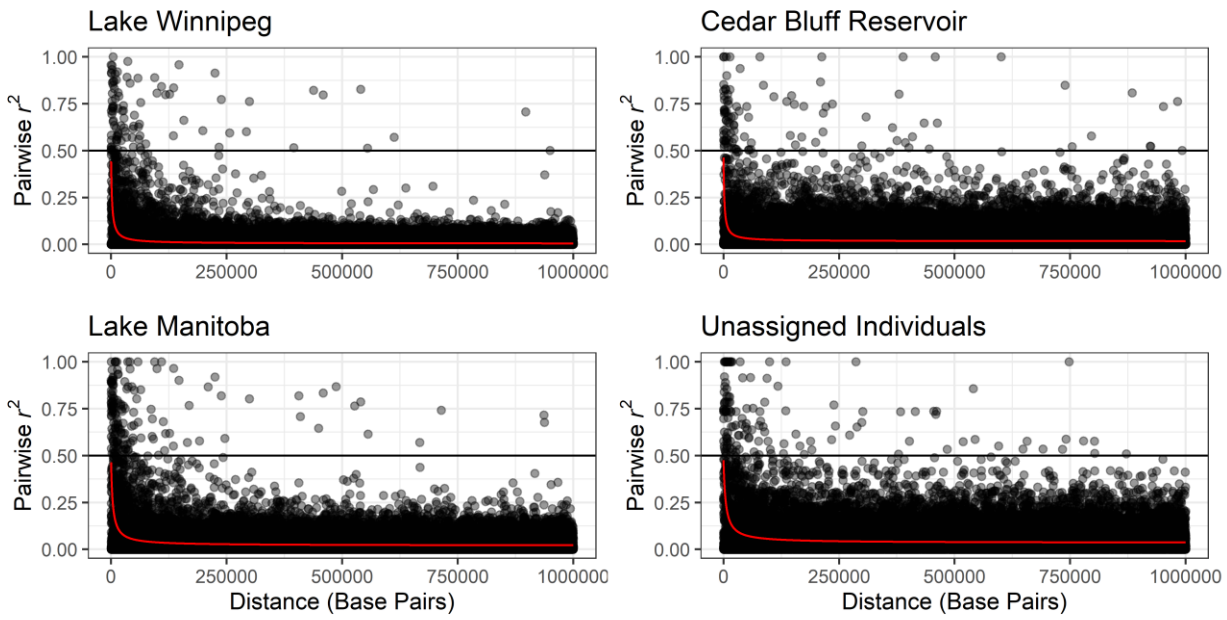


**Fig A.3.6** Linkage disequilibrium decay and pairwise $r^2$ between single nucleotide morphisms up to 1,000,000 base pairs apart in walleye (*Sander vitreus*) of three North American waterbodies. Cedar Bluff Reservoir (Kansas, USA) represents an entirely introduced population of unknown origin, while Lake Winnipeg and Lake Manitoba (Manitoba, Canada) represent native populations with possible gene flow between them. Unassigned individuals are those with ancestry coefficients (Q) ≤0.85 from Admixture for population assignment, all of which were in the Manitoba lakes. Red lines represent expected $r^2$ modeled separately for each group of fish by non-linear regression. In each plot, a line has been drawn at $r^2$=0.50 for emphasis.

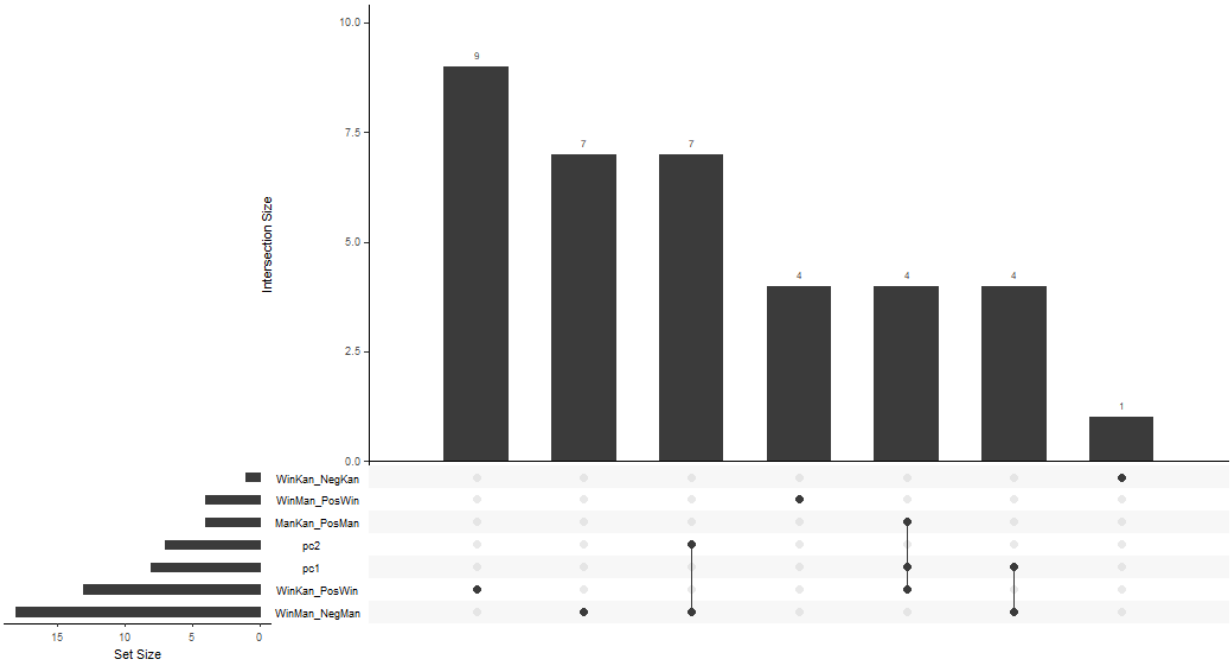


**Fig A.3.7** UpSet plot comparing numbers of genes overlapping between different scans for selection in walleye (*Sander vitreus*) of three North American waterbodies. Cross-population extended haplotype homozygosity (XP-EHH) was used as the basis for identifying candidate genomic regions of selection in the present study. For comparisons associated with this haplotype-based approach, comparisons are described by the name of the set. Lake Winnipeg (Manitoba, Canada) is abbreviated as "Win", Lake Manitoba (Manitoba, Canada) is abbreviated as "Man", and Cedar Bluff Reservoir (Kansas, USA) is abbreviated as "Kan". The location in which selection was inferred by haplotype length is provided following an underscore. The abbreviations "Pos" and "Neg" refer to positive or negative values for XP-EHH, which were arbitrarily assigned but were indicate of the population in which selection may have acted. Therefore, "WinKan_NegKan" refers to a pairwise comparison between Lake Winnipeg and Cedar Bluff Reservoir walleye, in which genes were associated with haplotypes possibly under selection in Cedar Bluff Reservoir relative to Lake Winnipeg fish. The abbreviations "pc1" and "pc2" refer to principal component 1 and principal component 2, respectively, from principal components analysis used to conduct outlier tests for signatures of selection. For these outlier tests, pc1 refers to genes that were associated with latitudinal outliers, while pc2 refers to genes that refers to longitudinal outliers.

## A.4 Supplemental Material for Chapter 5: A Polygenic Basis for Habitat Selection in Lake Winnipeg Walleye (*Sander vitreus*)

### A.4.1 Supplemental Results

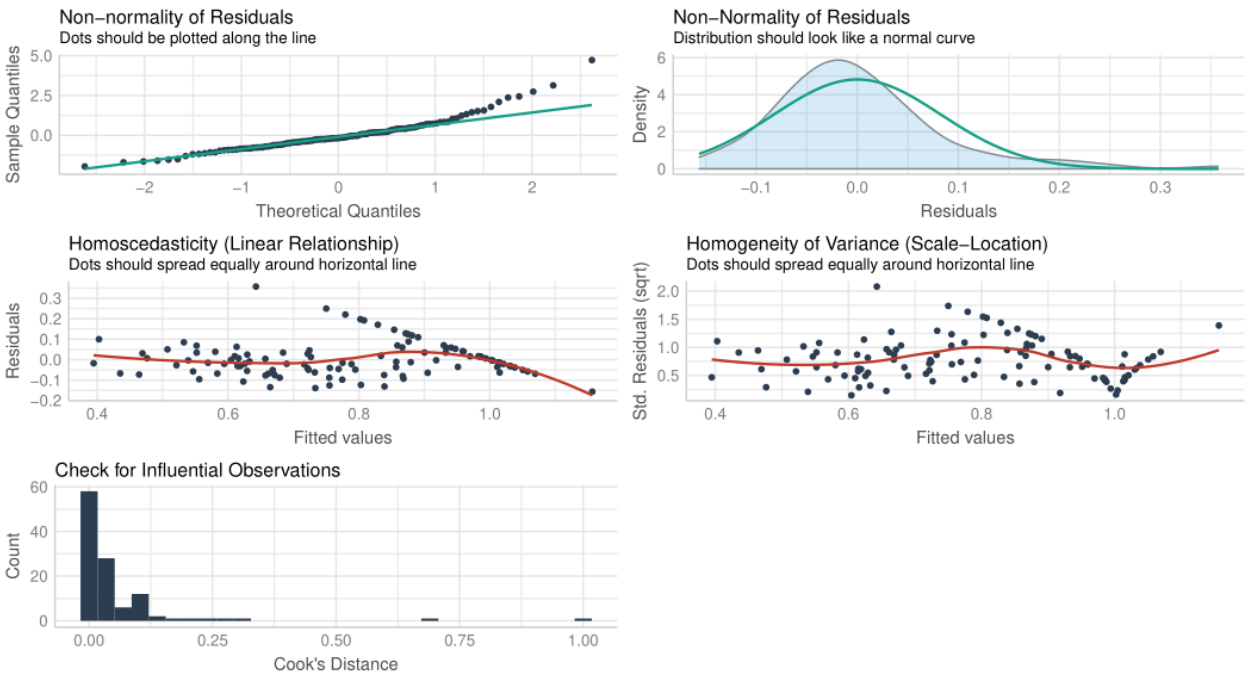


**Fig A.4.1** Non-normality of residuals, homoscedasticity, homogeneity of variance, and Cook's distance for a linear model of predictive $R^2$ in Lake Winnipeg (Manitoba, Canada) walleye (*Sander vitreus*). In this linear model, the dependent variable was maximum residency observed and the independent variable was the polygenic score for each of *n*=113 fish sampled from three sites in the lake.

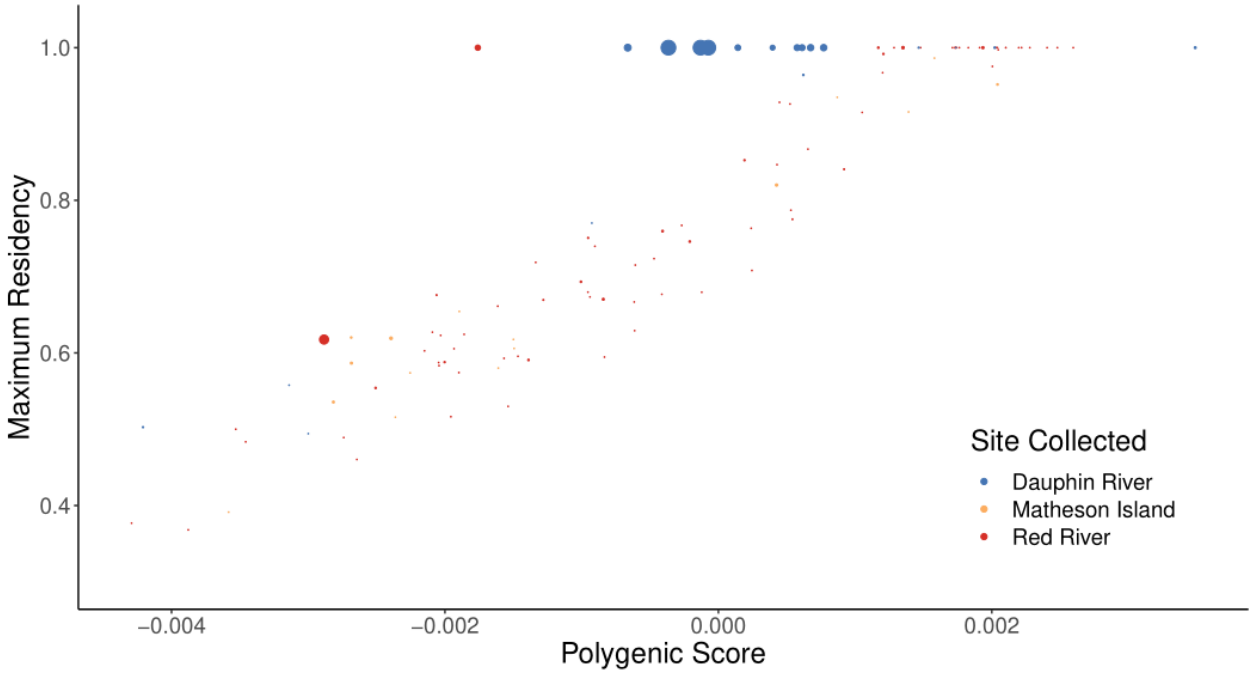


**Fig A.4.2** Polygenic scores and maximum residency values for *n*=113 walleye (*Sander vitreus*) sampled from Lake Winnipeg (Manitoba, Canada) from the Red River (south basin), Matheson Island (narrows), and Dauphin River (north basin). Polygenic scores were calculated from additive single nucleotide polymorphism (SNP) effect sizes, based on association testing with maximum residency at a *p*-value threshold of 0.4. A linear model with maximum residency as the dependent variable and polygenic scores the independent variable was developed, from which predictive $R^2$ was calculated using a leave-one-out cross validation process to find the predictive residual sum of squares, divided by the total sum of squares. This figure is identical to Fig 5.5, except point sizes are determined by ancestry (Q-values) from Lake Manitoba in Lake Winnipeg walleye, determined by Admixture. Ancestry values were taken from Chapter 4, where K=3 groups was chosen for population assignment.

**Literature Cited**

Abrantes, K. G., & Barnett, A. (2011). Intrapopulation variations in diet and habitat use in a marine apex predator, the broadnose sevengill shark *Notorynchus cepedianus*. *Marine Ecology Progress Series*, *442*, 133–148. https://doi.org/10.3354/meps09395

Ackerman, M. W., Hand, B. K., Waples, R. K., Luikart, G., Waples, R. S., Steele, C. A., Garner, B. A., McCane, J., & Campbell, M. R. (2017). Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. *Evolutionary Applications*, *10*(2), 146–160. https://doi.org/10.1111/eva.12433

Ahmari, H., Blais, E. L., & Greshuk, J. (2016). The 2014 flood event in the Assiniboine River Basin: Causes, assessment and damage. *Canadian Water Resources Journal*, *41*(1–2), 85–93. https://doi.org/10.1080/07011784.2015.1070695

Aitken, S. N., & Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Allen, D. M. (1974). The Relationship Between Variable Selection and Data Agumentation and a Method for Prediction. *Technometrics*, *16*(1), 125–127. https://doi.org/10.1080/00401706.1974.10489157

Allendorf, F. W., & Hard, J. J. (2009). Human-induced evolution caused by unnatural selection through harvest of wild animals. *In the Light of Evolution*, *3*, 129–147. https://doi.org/10.17226/12692

Alsop, D. H., Kieffer, J. D., & Wood, C. M. (1999). The effects of temperature and swimming speed on instantaneous fuel use and nitrogenous waste excretion of the Nile tilapia. *Physiological and Biochemical Zoology*, *72*(4), 474–483. https://doi.org/10.1086/316686

Alsop, D. H., & Wood, C. M. (1997). The interactive effects of feeding and exercise on oxygen consumption, swimming performance and protein usage in juvenile rainbow trout (*Oncorhynchus mykiss*). *Journal of Experimental Biology*, *200*(17), 2337–2346.

Amrhein, V., & Greenland, S. (2019). Retire statistical significance Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories. *Nature*, *567*, 305–307.

Anderson, C. N. K., Hsieh, C., Sandin, S. A., Hewitt, R., Hollowed, A., Beddington, J., May, R. M., & Sugihara, G. (2008). Why fishing magnifies fluctuations in fish abundance. *Nature*, *452*(7189), 835–839. https://doi.org/10.1038/nature06851

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

Aykanat, T., Johnston, S. E., Orell, P., Niemelä, E., Erkinaro, J., & Primmer, C. R. (2015). Low but significant genetic differentiation underlies biologically meaningful phenotypic divergence in a large Atlantic salmon population. *Molecular Ecology*, *24*(20), 5158–5174. https://doi.org/10.1111/mec.13383

Ayllón, D., Almodóvar, A., Nicola, G. G., & Elvira, B. (2010). Ontogenetic and spatial variations in brown trout habitat selection. *Ecology of Freshwater Fish*, *19*(3), 420–432. https://doi.org/10.1111/j.1600-0633.2010.00426.x

Baccante, D. A., & Colby, P. J. (1996). Harvest, density and reproductive characteristics of North American walleye populations. *Annales Zoologici Fennici*, *33*(3–4), 601–615.

Backhouse-James, S. M., & Docker, M. F. (2012). Microsatellite and mitochondrial DNA markers show no evidence of population structure in walleye (*Sander vitreus*) in Lake Winnipeg. *Journal of Great Lakes Research*, *38*(SUPPL. 3), 47–57. https://doi.org/10.1016/j.jglr.2011.05.005

Bailey, L. L., & Muths, E. (2019). Integrating amphibian movement studies across scales better informs conservation decisions. *Biological Conservation*, *236*(June), 261–268. https://doi.org/10.1016/j.biocon.2019.05.028

Ballantyne, J. S. (2001). *Amino acid metabolism* (pp. 77–107). https://doi.org/10.1016/S1546-

5098(01)20004-1

Barts, N., Greenway, R., Passow, C. N., Arias-Rodriguez, L., Kelley, J. L., & Tobler, M. (2018). Molecular evolution and expression of oxygen transport genes in livebearing fishes (Poeciliidae) from hydrogen sulfide rich springs. *Genome*, *61*(4), 273–286. https://doi.org/10.1139/gen-2017-0051

Bass, A. L., Hinch, S. G., Teffer, A. K., Patterson, D. A., & Miller, K. M. (2019). Fisheries capture and infectious agents are associated with travel rate and survival of Chinook salmon during spawning migration. *Fisheries Research*, *209*(September 2018), 156–166. https://doi.org/10.1016/j.fishres.2018.09.009

Bassar, R. D., Coulson, T., Travis, J., & Reznick, D. N. (2021). Towards a more precise – and accurate – view of eco-evolution. *Ecology Letters*, *24*(4), 623–625. https://doi.org/10.1111/ele.13712

Binding, C. E., Greenberg, T. A., McCullough, G., Watson, S. B., & Page, E. (2018). An analysis of satellite-derived chlorophyll and algal bloom indices on Lake Winnipeg. *Journal of Great Lakes Research*, *44*(3), 436–446. https://doi.org/10.1016/j.jglr.2018.04.001

Birnie-Gauvin, K., Flávio, H., Kristensen, M. L., Walton-Rabideau, S., Cooke, S. J., Willmore, W. G., Koed, A., & Aarestrup, K. (2019). Cortisol predicts migration timing and success in both Atlantic salmon and sea trout kelts. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-39153-x

Black, D., & Love, R. M. (1986). The sequential mobilisation and restoration of energy reserves in tissues of Atlantic cod during starvation and refeeding. *Journal of Comparative Physiology B*, *156*(4), 469–479. https://doi.org/10.1007/BF00691032

Blasco, J., Fernández, J., & Gutiérrez, J. (1991). The effects of starvation and refeeding on plasma amino acid levels in carp, *Cyprinus carpio L.*, 1758. *Journal of Fish Biology*, *38*(4), 587–598. https://doi.org/10.1111/j.1095-8649.1991.tb03145.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

https://doi.org/10.1093/bioinformatics/btu170

Bootsma, M. L., Miller, L., Sass, G. G., Euclide, P. T., & Larson, W. A. (2020). The ghosts of propagation past: haplotype information clarifies the relative influence of stocking history and phylogeographic processes on contemporary population structure of walleye (*Sander vitreus*). *Evolutionary Applications*, *June 2020*, 1–21. https://doi.org/10.1111/eva.13186

Borsook, H., & Deasy, C. L. (1948). Alpha-Aminoadipic acid; a product of lysine metabolism. *The Journal of Biological Chemistry*, *173*(1), 423. http://www.ncbi.nlm.nih.gov/pubmed/18902413

Bossus, M. C., Madsen, S. S., & Tipsmark, C. K. (2015). Functional dynamics of claudin expression in Japanese medaka (*Oryzias latipes*): Response to environmental salinity. *Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology*, *187*, 74–85. https://doi.org/10.1016/j.cbpa.2015.04.017

Bowles, E., Marin, K., & Fraser, D. J. (2019). Size reductions and genomic changes associated with harvesting within two generations in wild walleye populations. *BioRxiv Preprint*, *September*. https://doi.org/10.1101/787374

Bray, S. M., Mulle, J. G., Dodd, A. F., Pulver, A. E., Wooding, S., & Warren, S. T. (2010). Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16222–16227. https://doi.org/10.1073/pnas.1004381107

Breau, C., Cunjak, R. A., & Peake, S. J. (2011). Behaviour during elevated water temperatures: Can physiology explain movement of juvenile Atlantic salmon to cool water? *Journal of Animal Ecology*, *80*(4), 844–853. https://doi.org/10.1111/j.1365-2656.2011.01828.x

Brieuc, M. S. O., Ono, K., Drinan, D. P., & Naish, K. A. (2015). Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, *24*(11), 2729–2746. https://doi.org/10.1111/mec.13211

Broad Institute. (2019). Picard toolkit. In *Broad Institute, GitHub repository* (p.

http://broadinstitute.github.io/picard/-http://bro). Broad Institute. http://broadinstitute.github.io/picard/

Brown, A. P., Arias-Rodriguez, L., Yee, M.-C., Tobler, M., & Kelley, J. L. (2018). Concordant changes in gene expression and nucleotides underlie independent adaptation to hydrogen-sulfide-rich environments. *Genome Biology and Evolution*, *10*(11), 2867–2881. https://doi.org/10.1093/gbe/evy198

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, *81*(5), 1084–1097. https://doi.org/10.1086/521987

Bruner, J. C. (2021). *Stizostedion* Rafinesque, 1820 (Percidae) is the Valid Generic Name for Walleye, Sauger, and Eurasian Pikeperch. *Fisheries*, *1820*(1816), 1–5. https://doi.org/10.1002/fsh.10582

Brunskill, G. J., Elliott, S. E. M., & Campbell, P. (1980). Watershed data pertinent to the limnology of Lake Winnipeg: Canadian manuscript report of fisheries & aquatic sciences. In *Canadian Manuscript Report of Fisheries and Aquatic Sciences* (Issue 1556). http://hdl.handle.net/1993/23608

Busch, W.-D. N., Scholl, R. L., & Hartman, W. L. (1975). Environmental Factors Affecting the Strength of Walleye (*Stizostedion vitreum vitreum*) Year-Classes in Western Lake Erie, 1960–70 . *Journal of the Fisheries Research Board of Canada*, *32*(10), 1733–1743. https://doi.org/10.1139/f75-207

Carlander, K. D., Whitney, R. R., Speaker, E. B., & Madden, K. (1960). Evaluation of Walleye Fry Stocking in Clear Lake, Iowa, by Alternate-Year Planting. *Transactions of the American Fisheries Society*, *89*(3), 249–254. https://doi.org/https://doi.org/10.1577/1548-8659(1960)89[249:EOWFSI]2.0.CO;2

Carruth, L. L., Dores, R. M., Maldonado, T. A., Norris, D. O., Ruth, T., & Jones, R. E. (2000). Elevation of plasma cortisol during the spawning migration of landlocked kokanee salmon (*Oncorhynchus nerka kennerlyi*). *Comparative Biochemistry and Physiology - C Pharmacology Toxicology and Endocrinology*, *127*(2), 123–131. https://doi.org/10.1016/S0742-8413(00)00140-7

Carter, C. G., & Houlihan, D. F. (2001). *Protein synthesis* (pp. 31–75). https://doi.org/10.1016/S1546-5098(01)20003-X

Caruso, G., Denaro, M. G., Caruso, R., Mancari, F., Genovese, L., & Maricchiolo, G. (2011). Response to short term starvation of growth, haematological, biochemical and non-specific immune parameters in European sea bass (*Dicentrarchus labrax*) and blackspot sea bream (Pagellus bogaraveo). *Marine Environmental Research*, *72*(1–2), 46–52. https://doi.org/10.1016/j.marenvres.2011.04.005

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

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*(3), 171–182. https://doi.org/10.1534/g3.111.000240

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365. https://doi.org/10.1111/1755-0998.12669

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Cayuela, H., Rougemont, Q., Prunier, J. G., Moore, J. S., Clobert, J., Besnard, A., & Bernatchez, L. (2018). Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. *Molecular Ecology*, *27*(20), 3976–4010. https://doi.org/10.1111/mec.14848

Chamary, J. V, & Hurst, L. D. (2005). Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol*, *6*(9), R75. https://doi.org/10.1186/gb-2005-6-9-r75

Chapman, J. R., Nakagawa, S., Coltman, D. W., Slate, J., & Sheldon, B. C. (2009). A quantitative review of heterozygosity-fitness correlations in animal populations. *Molecular Ecology*, *18*(13), 2746–2765. https://doi.org/10.1111/j.1365-294X.2009.04247.x

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128. https://doi.org/10.1186/1471-2105-14-128

Chen, K., Euclide, P. T., Ludsin, S. A., Larson, W., Sovic, M. G., Gibbs, H. L., & Marschall, E. A. (2019). RAD-seq refines previous estimates of genetic structure in Lake Erie walleye (*Sander vitreus*). *Transactions of the American Fisheries Society*, tafs.10215. https://doi.org/10.1002/tafs.10215

Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, *255*(1–4), 1–29. https://doi.org/10.1016/j.aquaculture.2005.11.031

Choi, S. W., Mak, T. S. H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759–2772. https://doi.org/10.1038/s41596-020-0353-1

Chopin, F. S., Arimoto, T., & Inoue, Y. (1996). A comparison of the stress response and mortality of sea bream Pagrus major captured by hook and line and trammel net. *Fisheries Research*, *28*(3), 277–289. https://doi.org/10.1016/0165-7836(96)00512-7

Christie, M. R., Marine, M. L., French, R. A., Waples, R. S., & Blouin, M. S. (2012). Effective size of a wild salmonid population is greatly reduced by hatchery supplementation. *Heredity*, *109*(4), 254–260. https://doi.org/10.1038/hdy.2012.39

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., &

Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Clare, E. L., Economou, C. K., Faulkes, C. G., Gilbert, J. D., Bennett, F., Drinkwater, R., & Littlefair, J. E. (2021). eDNAir: Proof of concept that animal DNA can be collected from air sampling. *PeerJ*, *9*. https://doi.org/10.7717/peerj.11030

Clobert, J., Le Galliard, J. F., Cote, J., Meylan, S., & Massot, M. (2009). Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecology Letters*, *12*(3), 197–209. https://doi.org/10.1111/j.1461-0248.2008.01267.x

Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2016). Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, *33*(2), 591–593. https://doi.org/10.1093/molbev/msv255

Collins, A. L., & Anderson, T. A. (1995). The regulation of endogeneous energy stores during starvation and refeeding in the somatic tissues of the golden perch. *Journal of Fish Biology*, *47*(6), 1004–1015. https://doi.org/10.1111/j.1095-8649.1995.tb06024.x

Colodro-Conde, L., Couvy-Duchesne, B., Zhu, G., Coventry, W. L., Byrne, E. M., Gordon, S., Wright, M. J., Montgomery, G. W., Madden, P. A. F., Ripke, S., Eaves, L. J., Heath, A. C., Wray, N. R., Medland, S. E., & Martin, N. G. (2018). A direct test of the diathesis–stress model for depression. *Molecular Psychiatry*, *23*(7), 1590–1596. https://doi.org/10.1038/mp.2017.130

Connon, R. E., Deanovic, L. A., Fritsch, E. B., D'Abronzo, L. S., & Werner, I. (2011). Sublethal responses to ammonia exposure in the endangered delta smelt; Hypomesus transpacificus (Fam. Osmeridae). *Aquatic Toxicology*, *105*(3–4), 369–377. https://doi.org/10.1016/j.aquatox.2011.07.002

Connon, R. E., Jeffries, K. M., Komoroske, L. M., Todgham, A. E., & Fangue, N. A. (2018). The utility of transcriptomics in fish conservation. *The Journal of Experimental Biology*, *221*(2), jeb148833. https://doi.org/10.1242/jeb.148833

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Cooke, S. J., Crossin, G. T., Patterson, D. A., English, K. K., Hinch, S. G., Young, J. L., Alexander, R. F., Healey, M. C., Van Der Kraak, G., & Farrell, A. P. (2005). Coupling non-invasive physiological assessments with telemetry to understand inter-individual variation in behaviour and survivorship of sockeye salmon: development and validation of a technique. *Journal of Fish Biology*, *67*(5), 1342–1358. https://doi.org/10.1111/j.1095-8649.2005.00830.x

Cooke, Steven J., Hinch, S. G., Donaldson, M. R., Clark, T. D., Eliason, E. J., Crossin, G. T., Raby, G. D., Jeffries, K. M., Lapointe, M., Miller, K., Patterson, D. A., & Farrell, A. P. (2012). Conservation physiology in practice: How physiological knowledge has improved our ability to sustainably manage Pacific salmon during up-river migration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1596), 1757–1769. https://doi.org/10.1098/rstb.2012.0022

Cooke, Steven J., Hinch, S. G., Farrell, A. P., Patterson, D. A., Miller-Saunders, K., Welch, D. W., Donaldson, M. R., Hanson, K. C., Crossin, G. T., Mathes, M. T., Lotto, A. G., Hruska, K. A., Olsson, I. C., Wagner, G. N., Thomson, R., Hourston, R., English, K. K., Larsson, S., Shrimpton, J. M., & Van der Kraak, G. (2008). Developing a Mechanistic Understanding of Fish Migrations by Linking Telemetry with Physiology, Behavior, Genomics and Experimental Biology: An Interdisciplinary Case Study on Adult Fraser River Sockeye Salmon. *Fisheries*, *33*(7), 321–339. https://doi.org/10.1577/1548-8446-33.7.321

Cooke, Steven J., Hinch, S. G., Wikelski, M., Andrews, R. D., Kuchel, L. J., Wolcott, T. G., & Butler, P. J. (2004). Biotelemetry: A mechanistic approach to ecology. *Trends in Ecology and Evolution*, *19*(6), 334–343. https://doi.org/10.1016/j.tree.2004.04.003

Costas, B., Aragão, C., Ruiz-Jarabo, I., Vargas-Chacoff, L., Arjona, F. J., Dinis, M. T., Mancera, J. M., & Conceição, L. E. C. (2011). Feed deprivation in Senegalese sole (*Solea senegalensis* Kaup, 1858) juveniles: Effects on blood plasma metabolites and free amino acid levels. *Fish Physiology and Biochemistry*, *37*(3), 495–504.

https://doi.org/10.1007/s10695-010-9451-2

Cote, J., Bocedi, G., Debeffe, L., Chudzińska, M. E., Weigang, H. C., Dytham, C., Gonzalez, G., Matthysen, E., Travis, J., Baguette, M., & Hewison, A. J. M. (2017). Behavioural synchronization of large-scale animal movements – disperse alone, but migrate together? *Biological Reviews*, *92*(3), 1275–1296. https://doi.org/10.1111/brv.12279

Craig, J. F., Babaluk, J. A., Stevenson, S. G., & Williams, P. C. (1995). Variation in growth and reproduction of walleye (*Stizostedion vitreum*) in three Manitoba lakes. *Canadian Journal of Zoology*, *73*(2), 367–372. https://doi.org/10.1139/z95-040

Crouch, D. J. M., & Bodmer, W. F. (2020). Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proceedings of the National Academy of Sciences*, *117*(32), 18924–18933. https://doi.org/10.1073/pnas.2005634117

Cuff, W. R. (1980). Behavioral aspects of cannibalism in larval walleye, Stizostedion vitreum. *Canadian Journal of Zoology*, *58*(8), 1504–1507. https://doi.org/10.1139/z80-207

Dadda, M., Koolhaas, W. H., & Domenici, P. (2010). Behavioural asymmetry affects escape performance in a teleost fish. *Biology Letters*, *6*(3), 414–417. https://doi.org/10.1098/rsbl.2009.0904

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danzmann, R. G. (1979). The karyology of eight species of fish belonging to the family Percidae. *Canadian Journal of Zoology*, *57*(10), 2055–2060. https://doi.org/10.1139/z79-271

Darling, A. E., Mau, B., & Perna, N. T. (2010). Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0011147

Davey, J. L., & Blaxter, M. W. (2010). RADseq: Next-generation population genetics. *Briefings*

*in Functional Genomics*, *9*(5–6), 416–423. https://doi.org/10.1093/bfgp/elq031

Davidson, N. M., & Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novoassembled transcriptomes. *Genome Biology*, *15*(7), 410. https://doi.org/10.1186/s13059-014-0410-6

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Computational Biology*, *11*(4), 1–19. https://doi.org/10.1371/journal.pcbi.1004219

De Meeûs, T. (2018). Revisiting $F_{IS}$, $F_{ST}$, Wahlund Effects, and Null Alleles. *Journal of Heredity*, *109*(4), 446–456. https://doi.org/10.1093/jhered/esx106

De Wit, P., & Palumbi, S. R. (2013). Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, *22*(11), 2884–2897. https://doi.org/10.1111/mec.12081

Debes, P. V., Piavchenko, N., Erkinaro, J., & Primmer, C. R. (2020). Genetic growth potential, rather than phenotypic size, predicts migration phenotype in Atlantic salmon: Determinants of migration phenotypes. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1931). https://doi.org/10.1098/rspb.2020.0867rspb20200867

DeBiasse, M. B., & Kelly, M. W. (2016). Plastic and evolved responses to global change: What can we learn from comparative transcriptomics? *Journal of Heredity*, *107*(1), 71–81. https://doi.org/10.1093/jhered/esv073

Department of Fisheries. (1891). *Annual Report of the Department of Fisheries for the Year 1890* (Issue 8). https://waves-vagues.dfo-mpo.gc.ca/Library/40758606_1890_pt.1.pdf

Dibike, Y., Prowse, T., Shrestha, R., & Ahmed, R. (2012). Observed trends and future projections of precipitation and air temperature in the Lake Winnipeg watershed. *Journal of Great Lakes Research*, *38*(SUPPL. 3), 72–82. https://doi.org/10.1016/j.jglr.2011.04.005

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–

214. https://doi.org/10.1111/1755-0998.12157

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Doerr, D., & Moret, B. M. E. (2018). Comparative RNA Genomics. In J. C. Setubal, J. Stoye, & P. F. Stadler (Eds.), *Methods in Molecular Biology* (Vol. 1704, pp. 317–329). Humana Press. https://doi.org/10.1007/978-1-4939-7463-4_14

Donaldson, M. R., Hinch, S. G., Suski, C. D., Fisk, A. T., Heupel, M. R., & Cooke, S. J. (2014). Making connections in aquatic ecosystems with acoustic telemetry monitoring. *Frontiers in Ecology and the Environment*, *12*(10), 565–573. https://doi.org/10.1890/130283

Drenner, S. M., Hinch, S. G., Furey, N. B., Clark, T. D., Li, S., Ming, T., Jeffries, K. M., Patterson, D. A., Cooke, S. J., Robichaud, D., Welch, D. W., Farrell, A. P., & Miller, K. M. (2018). Transcriptome patterns and blood physiology associated with homing success of sockeye salmon during their final stage of marine migration. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*(9), 1511–1524. https://doi.org/10.1139/cjfas-2017-0391

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews of the Cambridge Philosophical Society*, *81*(2), 163–182. https://doi.org/10.1017/S1464793105006950

DuFour, M. R., May, C. J., Roseman, E. F., Ludsin, S. A., Vandergoot, C. S., Pritt, J. J., Fraker, M. E., Davis, J. J., Tyson, J. T., Miner, J. G., Marschall, E. A., & Mayer, C. M. (2015). Portfolio theory as a management tool to guide conservation and restoration of multi-stock fish populations. *Ecosphere*, *6*(12). https://doi.org/10.1890/ES15-00237.1

Dymowska, A. K., Schultz, A. G., Blair, S. D., Chamot, D., & Goss, G. G. (2014). Acid-sensing ion channels are involved in epithelial Na+ uptake in the rainbow trout *Oncorhynchus mykiss*. *American Journal of Physiology - Cell Physiology*, *307*(3), 255–265. https://doi.org/10.1152/ajpcell.00398.2013

Environment and Climate Change Canada, Manitoba Agriculture and Resource Development (MARD). (2020). State of Lake Winnipeg (2nd Ed). https://gov.mb.ca/water/pubs/water/lakes-beaches-rivers/state_lake_wpg_report_tech.pdf

Einen, O., & Thomassen, M. S. (1998). Starvation prior to slaughter in Atlantic salmon (*Salmo salar*). *Aquaculture*, *169*(1–2), 37–53. https://doi.org/10.1016/S0044-8486(98)00332-9

Ellison, C. E., Hall, C., Kowbel, D., Welch, J., Brem, R. B., Glass, N. L., & Taylor, J. W. (2011). Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences*, *108*(7), 2831–2836. https://doi.org/10.1073/pnas.1014971108

Enders, E. C., Charles, C., van der Lee, A. S., & Lumb, C. E. (2021). Temporal variations in the pelagic fish community of Lake Winnipeg from 2002 to 2019. *Journal of Great Lakes Research*, *xxxx*. https://doi.org/10.1016/j.jglr.2021.01.004

Enders, E. C., Charles, C., Watkinson, D. A., Kovachik, C., Leroux, D. R., Hansen, H., & Pegg, M. A. (2019). Analysing habitat connectivity and home ranges of Bigmouth Buffalo and Channel Catfish using a large-scale acoustic receiver network. *Sustainability (Switzerland)*, *11*(11). https://doi.org/10.3390/su11113051

Engelke, U. F. H., Tangerman, A., Willemsen, M. A. A. P., Moskau, D., Loss, S., Mudd, S. H., & Wevers, R. A. (2005). Dimethyl sulfone in human cerebrospinal fluid and blood plasma confirmed by one-dimensional 1H and two-dimensional 1H-13C NMR. *NMR in Biomedicine*, *18*(5), 331–336. https://doi.org/10.1002/nbm.966

Environment Canada. (2011). State of Lake Winnipeg: 1999–2007. http://hdl.handle.net/1993/23915

Erickson, R. A., Rees, C. B., Coulter, A. A., Merkes, C. M., McCalla, S. G., Touzinsky, K. F., Walleser, L., Goforth, R. R., & Amberg, J. J. (2016). Detecting the movement and spawning activity of bigheaded carps with environmental DNA. *Molecular Ecology Resources*, *16*(4), 957–965. https://doi.org/10.1111/1755-0998.12533

Euclide, P. T., MacDougall, T., Robinson, J. M., Faust, M. D., Wilson, C. C., Chen, K. Y.,

Marschall, E. A., Larson, W., & Ludsin, S. (2021). Mixed-stock analysis using Rapture genotyping to evaluate stock-specific exploitation of a walleye population despite weak genetic structure. Evolutionary Applications, February, 1–18. https://doi.org/10.1111/eva.13209

Evans, T. G., Hammill, E., Kaukinen, K., Schulze, A. D., Patterson, D. A., English, K. K., Curtis, J. M. R., & Miller, K. M. (2011). Transcriptomics of environmental acclimatization and survival in wild adult Pacific sockeye salmon (*Oncorhynchus nerka*) during spawning migration. *Molecular Ecology*, *20*(21), 4472–4489. https://doi.org/10.1111/j.1365-294X.2011.05276.x

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics*, *164*(4), 1567 LP – 1587. http://www.genetics.org/content/164/4/1567.abstract

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, *7*(4), 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Fernandez-Fournier, P., Lewthwaite, J. M. M., & Mooers, A. (2021). Do We Need to Identify Adaptive Genetic Variation When Prioritizing Populations for Conservation? *Conservation Genetics*, *22*(2), 205–216. https://doi.org/10.1007/s10592-020-01327-w

Fernando, R. L., Stricker, C., & Elston, R. C. (1994). The finite polygenic mixed model: An alternative formulation for the mixed model of inheritance. *Theoretical and Applied Genetics*, *88*(5), 573–580. https://doi.org/10.1007/BF01240920

Feron, R., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., Eché, C., Valière, S., Donnadieu, C., Haffray, P., Bestin, A., Morvezen, R., Acloque, H., Euclide, P. T., Wen, M., Jouano, E., Schartl, M., Postlethwait, J. H., Schraidt, C., … Guiguen, Y. (2020). Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, Perca flavescens. *Molecular Ecology Resources*, *20*(2), 531–543. https://doi.org/10.1111/1755-0998.13133

Fielder, D. G., Schaeffer, J. S., & Thomas, M. V. (2007). Environmental and ecological conditions surrounding the production of large year classes of walleye (*Sander vitreus*) in Saginaw Bay, Lake Huron. *Journal of Great Lakes Research*, *33*(SUPPL. 1), 118–132. https://doi.org/10.3394/0380-1330(2007)33[118:EAECST]2.0.CO;2

Fisher, R. A. (1918). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Fisheries and Oceans Canada. (2018). Freshwater Landings. https://www.dfo-mpo.gc.ca/stats/commercial/land-debarq/freshwater-eaudouce/2017-eng.htm

Fisheries and Oceans Canada. (2019). Survey of recreational fishing in Canada, *2015.* 21.

Fisheries and Oceans Canada. (2021). Freshwater Fisheries - Catches and Landed Values by Species, By Province/Territory, 2019. https://www.dfo-mpo.gc.ca/stats/commercial/land-debarq/freshwater-eaudouce/2019-eng.htm

Foll, M., & Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Forney, J. L. (1976). Year-Class Formation in the Walleye (*Stizostedion vitreum vitreum*) Population of Oneida Lake, New York, 1966–73 . *Journal of the Fisheries Research Board of Canada*, *33*(4), 783–792. https://doi.org/10.1139/f76-096

Francis, J. M. (1996). Nature conservation and the precautionary principle. *Environmental Values*, 257–264.

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Franckowiak, R. P., Sloss, B. L., Bozek, M. A., & Newman, S. P. (2009). Temporal effective size estimates of a managed walleye Sander vitreus population and implications for genetic-based management. *Journal of Fish Biology*, *74*(5), 1086–1103.

https://doi.org/10.1111/j.1095-8649.2008.02170.x

Franzin, W. G., Barton, B. A., Remnant, R. A., Wain, D. B., & Pagel, S. J. (1994). Range Extension, Present and Potential Distribution, and Possible Effects of Rainbow Smelt in Hudson Bay Drainage Waters of Northwestern Ontario, Manitoba, and Minnesota. *North American Journal of Fisheries Management*, *14*(1), 65–76. https://doi.org/10.1577/1548-8675(1994)014<0065:repapd>2.3.co;2

Freitas, C., Olsen, E. M., Knutsen, H., Albretsen, J., & Moland, E. (2016). Temperature-associated habitat selection in a cold-water marine fish. *Journal of Animal Ecology*, *85*(3), 628–637. https://doi.org/10.1111/1365-2656.12458

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., Nguyen, T. H., Kane, M., Perez, E. M., Durand, N. C., Lareau, C. A., Stamenova, E. K., Aiden, E. L., Lander, E. S., & Engreitz, J. M. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics*, *51*(12), 1664–1669. https://doi.org/10.1038/s41588-019-0538-0

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, *27*(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Furey, N. B., Bass, A. L., Miller, K. M., Li, S., Lotto, A. G., Healy, S. J., Drenner, S. M., & Hinch, S. G. (2021). Infected juvenile salmon can experience increased predation during freshwater migration. *Royal Society Open Science*, *8*(3). https://doi.org/10.1098/rsos.201522

Garrigan, D., & Hammer, M. F. (2006). Reconstructing human origins in the genomic era. *Nature Reviews Genetics*, *7*(9), 669–680. https://doi.org/10.1038/nrg1941

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*. 1–9. http://arxiv.org/abs/1207.3907

Gattuso, J. P., Magnan, A., Billé, R., Cheung, W. W. L., Howes, E. L., Joos, F., Allemand, D.,

Bopp, L., Cooley, S. R., Eakin, C. M., Hoegh-Guldberg, O., Kelly, R. P., Pörtner, H. O., Rogers, A. D., Baxter, J. M., Laffoley, D., Osborn, D., Rankovic, A., Rochette, J., … Turley, C. (2015). Contrasting futures for ocean and society from different anthropogenic CO2 emissions scenarios. *Science*, *349*(6243). https://doi.org/10.1126/science.aac4722

Gauthey, Z., Freychet, M., Manicki, A., Herman, A., Lepais, O., Panserat, S., Elosegi, A., Tentelier, C., & Labonne, J. (2015). The concentration of plasma metabolites varies throughout reproduction and affects offspring number in wild brown trout (*Salmo trutta*). *Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology*, *184*, 90–96. https://doi.org/10.1016/j.cbpa.2015.01.025

Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, *17*(1), 78–90. https://doi.org/10.1111/1755-0998.12634

Geen, G. R., Robinson, R. A., & Baillie, S. R. (2019). Effects of tracking devices on individual birds – a review of the evidence. *Journal of Avian Biology*, *50*(2), 1–13. https://doi.org/10.1111/jav.01823

Ghiurcuta, C. G., & Moret, B. M. E. (2014). Evaluating synteny for improved comparative studies. *Bioinformatics*, *30*(12), 9–18. https://doi.org/10.1093/bioinformatics/btu259

Gossmann, T. I., Song, B. H., Windsor, A. J., Mitchell-Olds, T., Dixon, C. J., Kapralov, M. V., Filatov, D. A., & Eyre-Walker, A. (2010). Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular Biology and Evolution*, *27*(8), 1822–1832. https://doi.org/10.1093/molbev/msq079

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Gros-Balthazard, M., Besnard, G., Sarah, G., Holtz, Y., Leclercq, J., Santoni, S., Wegmann, D., Glémin, S., & Khadari, B. (2019). Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. *The Plant Journal*, *100*(1), 143–157. https://doi.org/10.1111/tpj.14435

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2008). Heterozygosity-fitness correlations and their relevance to studies on inbreeding depression in threatened species. *Molecular Ecology*, *17*(18), 3978–3984. https://doi.org/10.1111/j.1365-294X.2008.03910.x

Grutter, A. S., & Pankhurst, N. W. (2000). The effects of capture, handling, confinement and ectoparasite load on plasma levels of cortisol, glucose and lactate in the coral reef fish *Hemigymnus melapterus*. *Journal of Fish Biology*, *57*(2), 391–401. https://doi.org/10.1111/j.1095-8649.2000.tb02179.x

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10). https://doi.org/10.1371/journal.pgen.1000695

Halfyard, E. A., Webber, D., Del Papa, J., Leadley, T., Kessel, S. T., Colborne, S. F., & Fisk, A. T. (2017). Evaluation of an acoustic telemetry transmitter designed to identify predation events. *Methods in Ecology and Evolution*, *8*(9), 1063–1071. https://doi.org/10.1111/2041-210X.12726

Hamidan, N., & Britton, J. R. (2015). Age and growth rates of the critically endangered fish *Garra ghorensis* can inform their conservation management. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *25*(1), 61–70. https://doi.org/10.1002/aqc.2449

Hamilton, J. A., & Miller, J. M. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology*, *30*(1), 33–41. https://doi.org/10.1111/cobi.12574

Hartfield, M., & Bataillon, T. (2018). Selective sweeps under dominance and inbreeding. *BioRxiv*, *10*(March), 1063–1075. https://doi.org/10.1101/318410

Healy, S. J., Hinch, S. G., Bass, A. L., Furey, N. B., Welch, D. W., Rechisky, E. L., Eliason, E.

J., Lotto, A. G., & Miller, K. M. (2018). Transcriptome profiles relate to migration fate in hatchery steelhead (*Oncorhynchus mykiss*) smolts. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*(11), 2053–2068. https://doi.org/10.1139/cjfas-2017-0424

Healy, T. M., Brennan, R. S., Whitehead, A., & Schulte, P. M. (2018). Tolerance traits related to climate change resilience are independent and polygenic. *Global Change Biology*, *24*(11), 5348–5360. https://doi.org/10.1111/gcb.14386

Henderson, B. A., Wong, J. L., & Nepszy, S. J. (1996). Reproduction of walleye in Lake Erie: Allocation of energy. *Canadian Journal of Fisheries and Aquatic Sciences*, *53*(1), 127–133. https://doi.org/10.1139/f95-162

Hess, J. E., Zendt, J. S., Matala, A. R., & Narum, S. R. (2016). Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1830). https://doi.org/10.1098/rspb.2015.3064

Hill, W. G. (1979). A note on effective population size with overlapping generations. *Genetics*, *92*(1), 317–322. http://www.ncbi.nlm.nih.gov/pubmed/17248921

Hill, W. G., & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, *33*(1), 54–78. https://doi.org/10.1016/0040-5809(88)90004-4

Hinch, S. G., Cooke, S. J., Healey, M. C., & Farrell, A. P. (Tony. (2005). Behavioural Physiology of Fish Migrations: salmon as a model approach. *Fish Physiology*, *24*(C), 239–295. https://doi.org/10.1016/S1546-5098(05)24007-4

Hiwatashi, T., Okabe, Y., Tsutsui, T., Hiramatsu, C., Melin, A. D., Oota, H., Schaffner, C. M., Aureli, F., Fedigan, L. M., Innan, H., & Kawamura, S. (2010). An explicit signature of balancing selection for color-vision variation in new world monkeys. *Molecular Biology and Evolution*, *27*(2), 453–464. https://doi.org/10.1093/molbev/msp262

Hoey, J. A., & Pinsky, M. L. (2018). Genomic signatures of environmental selection despite near-panmixia in summer flounder. *Evolutionary Applications*, *11*(9), 1732–1747.

https://doi.org/10.1111/eva.12676

Hoffmann, A. A., & Willi, Y. (2008). Detecting genetic responses to environmental change. *Nature Reviews Genetics*, *9*(6), 421–432. https://doi.org/10.1038/nrg2339

Hohenlohe, P. A., Bassham, S., Currey, M., & Cresko, W. A. (2012). Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 395–409. https://doi.org/10.1098/rstb.2011.0245

Holland, K. N., Brill, R. W., Chang, R. K. C., Sibert, J. R., & Fournier, D. A. (1992). Physiological and behavioural thermoregulation in bigeye tuna (*Thunnus obesus*). *Nature*, *358*(6385), 410–412. https://doi.org/10.1038/358410a0

Houde, A. L. S., Schulze, A. D., Kaukinen, K. H., Strohm, J., Patterson, D. A., Beacham, T. D., Farrell, A. P., Hinch, S. G., & Miller, K. M. (2019). Transcriptional shifts during juvenile Coho salmon (*Oncorhynchus kisutch*) life stage changes in freshwater and early marine environments. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, *29*(October 2018), 32–42. https://doi.org/10.1016/j.cbd.2018.10.002

Houlihan, D. F., & Laurent, P. (1987). Effects of Exercise Training on the Performance, Growth, and Protein Turnover of Rainbow Trout (*Salmo gairdneri*). *Canadian Journal of Fisheries and Aquatic Sciences*, *44*(9), 1614–1621. https://doi.org/10.1139/f87-195

Huang, H. T., Kathrein, K. L., Barton, A., Gitlin, Z., Huang, Y. H., Ward, T. P., Hofmann, O., Dibiase, A., Song, A., Tyekucheva, S., Hide, W., Zhou, Y., & Zon, L. I. (2013). A Network of Epigenetic Regulators Guides Developmental Haematopoiesis in Vivo. *Nature Cell Biology*, *15*(12), 1516–1525. https://doi.org/10.1038/ncb2870

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*(5), 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Huey, R. B. (1991). Physiological Consequences of Habitat Selection. *The American Naturalist*, *137*(June 1991), S91–S115. https://doi.org/10.1086/285141

Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, *11*(6), 609–623. https://doi.org/10.1111/j.1461-0248.2008.01179.x

Hunt von Herbing, I. (2002). Effects of temperature on larval fish swimming performance: The importance of physics to physiology. *Journal of Fish Biology*, *61*(4), 865–876. https://doi.org/10.1006/jfbi.2002.2118

Hussey, N. E., Kessel, S. T., Aarestrup, K., Cooke, S. J., Cowley, P. D., Fisk, A. T., Harcourt, R. G., Holland, K. N., Iverson, S. J., Kocik, J. F., Flemming, J. E. M., & Whoriskey, F. G. (2015). Aquatic animal telemetry: A panoramic window into the underwater world. *Science*, *348*(6240), 1255642. https://doi.org/10.1126/science.1255642

Hussey, P. J., Ketelaar, T., & Deeks, M. J. (2006). Control of the actin cytoskeleton in plant cell growth. *Annual Review of Plant Biology*, *57*, 109–125. https://doi.org/10.1146/annurev.arplant.57.032905.105206

Hutto, R. L. (1985). Habitat selection by nonbreeding, migratory land birds. In *Habitat selection in birds* (pp. 455–476).

Jaenike, J., & Holt, R. D. (1991). Genetic variation for habitat preference: evidence and explanations. In *American Naturalist* (Vol. 137, Issue Suppl.). https://doi.org/10.1086/285140

Jeffrey, J. D., Carlson, H., Wrubleski, D., Enders, E. C., Treberg, J. R., & Jeffries, K. M. (2020). Applying a gene-suite approach to examine the physiological status of wild-caught walleye (*Sander vitreus*). *Conservation Physiology*, *8*(1), 1–12. https://doi.org/10.1093/conphys/coaa099

Jeffries, K. M., Connon, R. E., Verhille, C. E., Dabruzzi, T. F., Britton, M. T., Durbin-Johnson, B. P., & Fangue, N. A. (2019). Divergent transcriptomic signatures in response to salinity exposure in two populations of an estuarine fish. *Evolutionary Applications*, *12*(6), 1212–1226. https://doi.org/10.1111/eva.12799

Jeffries, K. M., Hinch, S. G., Gale, M. K., Clark, T. D., Lotto, A. G., Casselman, M. T., Li, S.,

Rechisky, E. L., Porter, A. D., Welch, D. W., & Miller, K. M. (2014). Immune response genes and pathogen presence predict migration survival in wild salmon smolts. *Molecular Ecology*, *23*(23), 5803–5815. https://doi.org/10.1111/mec.12980

Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., Knight, S., Knox, J., Long, T., Mannino, M., McGuire, M., Riffle, A., Segelsky, S., Shappell, L., Sterner, A., Strickler, T., & Tursi, R. (2010). A meta-analysis of isolation by distance: Relic or reference standard for landscape genetics? *Ecography*, *33*(2), 315–320. https://doi.org/10.1111/j.1600-0587.2010.06285.x

Jerde, C. L. (2021). Can we manage fisheries with the inherent uncertainty from eDNA? *Journal of Fish Biology*, *98*(2), 341–353. https://doi.org/10.1111/jfb.14218

Johnston, I. A., Bower, N. I., & Macqueen, D. J. (2011). Growth and the regulation of myotomal muscle mass in teleost fish. *Journal of Experimental Biology*, *214*(10), 1617–1628. https://doi.org/10.1242/jeb.038620

Johnston, S. E., Orell, P., Pritchard, V. L., Kent, M. P., Lien, S., Niemelä, E., Erkinaro, J., & Primmer, C. R. (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, *23*(14), 3452–3468. https://doi.org/10.1111/mec.12832

Johnston, T. A., Lysack, W., & Leggett, W. C. (2012). Abundance, growth, and life history characteristics of sympatric walleye (*Sander vitreus*) and sauger (*Sander canadensis*) in Lake Winnipeg, Manitoba. *Journal of Great Lakes Research*, *38*(SUPPL. 3), 35–46. https://doi.org/10.1016/j.jglr.2010.06.009

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

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

Jürss, K., & Bastrop, R. (1995). *Chapter 7 Amino acid metabolism in fish* (pp. 159–189). https://doi.org/10.1016/S1873-0140(06)80010-X

Käkelä, R., Mattila, M., Hermansson, M., Haimi, P., Uphoff, A., Paajanen, V., Somerharju, P., & Vornanen, M. (2008). Seasonal acclimatization of brain lipidome in a eurythermal fish (*Carassius carassius*) is mainly determined by temperature. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*, *294*(5), 1716–1728. https://doi.org/10.1152/ajpregu.00883.2007

Kasuya, E. (2019). On the use of r and r squared in correlation and regression. *Ecological Research*, *34*(1), 235–236. https://doi.org/10.1111/1440-1703.1011

Kelly, M. (2019). Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1768), 20180176. https://doi.org/10.1098/rstb.2018.0176

Kelso, J. R. M. (1978). Diel rhythm in activity of walleye, *Stizostedion vitreum vitreum*. *Journal of Fish Biology*, *12*(6), 593–599. https://doi.org/10.1111/j.1095-8649.1978.tb04206.x

Kerr, S. J. (2008). A Survey of Walleye Stocking Activities in North America May. In *Water* (Issue May).

Kess, T., Bentzen, P., Lehnert, S. J., Sylvester, E. V. A., Lien, S., Kent, M. P., Sinclair-Waters, M., Morris, C. J., Regular, P., Fairweather, R., & Bradbury, I. R. (2019). A migration-associated supergene reveals loss of biocomplexity in Atlantic cod. *Science Advances*, *5*(6), eaav2461. https://doi.org/10.1126/sciadv.aav2461

Kessel, S. T., Chapman, D. D., Franks, B. R., Gedamke, T., Gruber, S. H., Newman, J. M., White, E. R., & Perkins, R. G. (2014). Predictable temperature-regulated residency, movement and migration in a large, highly mobile marine predator (*Negaprion brevirostris*). *Marine Ecology Progress Series*, *514*, 175–190. https://doi.org/10.3354/meps10966

Kim, B. M., Amores, A., Kang, S., Ahn, D. H., Kim, J. H., Kim, I. C., Lee, J. H., Lee, S. G., Lee, H., Lee, J., Kim, H. W., Desvignes, T., Batzel, P., Sydes, J., Titus, T., Wilson, C. A.,

Catchen, J. M., Warren, W. C., Schartl, M., … Park, H. (2019). Antarctic blackfin icefish genome reveals adaptations to extreme environments. *Nature Ecology and Evolution*, *3*(3), 469–478. https://doi.org/10.1038/s41559-019-0812-7

Kim, S., & Coulombe, P. A. (2010). Emerging role for the cytoskeleton as an organizer and regulator of translation. *Nature Reviews Molecular Cell Biology*, *11*(1), 75–81. https://doi.org/10.1038/nrm2818

Kim, S., Wong, P., & Coulombe, P. A. (2006). A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. *Nature*, *441*(7091), 362–365. https://doi.org/10.1038/nature04659

Kim, Y., & Gulisija, D. (2010). Signatures of recent directional selection under different models of population expansion during colonization of new selective environments. *Genetics*, *184*(2), 571–585. https://doi.org/10.1534/genetics.109.109447

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

KNOX, W. E., & MEHLER, A. H. (1950). The conversion of tryptophan to kynurenine in liver. I. The coupled tryptophan peroxidase-oxidase system forming formylkynurenine. *The Journal of Biological Chemistry*, *187*(1), 419–430. http://www.ncbi.nlm.nih.gov/pubmed/14794727

Knutsen, H., Olsen, E. M., Jorde, P. E., Espeland, S. H., André, C., & Stenseth, N. C. (2011). Are low but statistically significant levels of genetic differentiation in marine fishes "biologically meaningful"? A case study of coastal Atlantic cod. *Molecular Ecology*, *20*(4), 768–783. https://doi.org/10.1111/j.1365-294X.2010.04979.x

Koch, I. J., & Narum, S. R. (2020). Validation and association of candidate markers for adult migration timing and fitness in Chinook Salmon. *Evolutionary Applications*, *13*(9), 2316–2332. https://doi.org/10.1111/eva.13026

Koll, R., Ribera, J. M., Brunner, R. M., Rebl, A., & Goldammer, T. (2020). Gene profiling in the

adipose fin of salmonid fishes supports its function as a flow sensor. *Genes*, *11*(1). https://doi.org/10.3390/genes11010021

Kolosov, D., Bui, P., Chasiotis, H., & Kelly, S. P. (2013). Claudins in teleost fishes. *Tissue Barriers*, *1*(3), e25391. https://doi.org/10.4161/tisb.25391

Krabbenhoft, T. J., Myers, B. J. E., Wong, J. P., Chu, C., Tingley, R. W., Falke, J. A., Kwak, T. J., Paukert, C. P., & Lynch, A. J. (2020). FiCli, the Fish and Climate Change Database, informs climate adaptation and management for freshwater fishes. *Scientific Data*, *7*(1), 1–6. https://doi.org/10.1038/s41597-020-0465-z

Krausman, P. R. (1999). Some Basic Principles of Habitat Use. *Grazing Behavior of Livestock and Wildlife*, 85–90.

Kritzer, J. P., & Sale, P. F. (2004). Metapopulation ecology in the sea: From Levins' model to marine ecology and fisheries science. *Fish and Fisheries*, *5*(2), 131–140. https://doi.org/10.1111/j.1467-2979.2004.00131.x

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645. https://doi.org/10.1101/gr.092759.109

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. https://doi.org/10.1093/nar/gkw377

Lawrence, M. J., Jain-Schlaepfer, S., Zolderdo, A. J., Algera, D. A., Gilmour, K. M., Gallagher, A. J., & Cooke, S. J. (2018). Are 3 minutes good enough for obtaining baseline physiological samples from teleost fish? *Canadian Journal of Zoology*, *96*(7), 774–786. https://doi.org/10.1139/cjz-2017-0093

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://cran.r-project.org/package=emmeans

Levine, T. R., & Hullett, C. R. (2002). Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research*, *28*(4), 612–625. https://doi.org/10.1093/hcr/28.4.612

Lewis, C. F. M., Forbes, D. L., Todd, B. J., Nielsen, E., Thorleifson, L. H., Henderson, P. J., McMartin, I., Anderson, T. W., Betcher, R. N., Buhay, W. M., Burbidge, S. M., Schröder-Adams, C. J., King, J. W., Moran, K., Gibson, C., Jarrett, C. A., Kling, H. J., Lockhart, W. L., Last, W. M., … Vance, R. E. (2002). Uplift-driven expansion delayed by middle Holocene desiccation in Lake Winnipeg, Manitoba, Canada. *Geology*, *29*(8), 743–746. https://doi.org/10.1130/0091-7613(2001)029<0743:UDEDBM>2.0.CO;2

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, K., Zheng, Y., & Drubin, D. G. (1995). Regulation of cortical actin cytoskeleton assembly during polarized cell growth in budding yeast. *Journal of Cell Biology*, *128*(4), 599–615. https://doi.org/10.1083/jcb.128.4.599

Lidgard, D. C., Bowen, W. D., Jonsen, I. D., & Iverson, S. J. (2014). Predator-borne acoustic transceivers and GPS tracking reveal spatiotemporal patterns of encounters with acoustically tagged fish in the open ocean. *Marine Ecology Progress Series*, *501*, 157–168. https://doi.org/10.3354/meps10670

Liedvogel, M., Åkesson, S., & Bensch, S. (2011). The genetics of migration on the move. *Trends in Ecology and Evolution*, *26*(11), 561–569. https://doi.org/10.1016/j.tree.2011.07.009

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647. https://doi.org/10.1111/1755-0998.12995

Lindberg, W. J., Frazer, T. K., Portier, K. M., Vose, F., Loftin, J., Murie, D. J., Mason, D. M.,

Nagy, B., & Hart, M. K. (2006). Density-dependent habitat selection and performance by a large mobile reef fish. *Ecological Applications*, *16*(2), 731–746. https://doi.org/10.1890/1051-0761(2006)016[0731:DHSAPB]2.0.CO;2

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Liu, L., Ang, K. P., Elliott, J. A. K., Kent, M. P., Lien, S., MacDonald, D., & Boulding, E. G. (2017). A genome scan for selection signatures comparing farmed Atlantic salmon with two wild populations: Testing colocalization among outlier markers, candidate genes, and quantitative trait loci for production traits. *Evolutionary Applications*, *10*(3), 276–296. https://doi.org/10.1111/eva.12450

Lüdecke, D. (2019). *sjstats: Statistical Functions for Regression Models (Version 0.17.7)*. https://doi.org/10.5281/zenodo.1284472

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Lumb, C. E., Franzin, W. G., & Watkinson, D. A. (2012). Temporal and spatial patterns in pelagic trawl fish catches in Lake Winnipeg. *Journal of Great Lakes Research*, *38*, 14–25. https://doi.org/10.1016/j.jglr.2011.07.008

Luna, R., Epperson, B. K., & Oyama, K. (2007). High levels of genetic variability and inbreeding in two Neotropical dioecious palms with contrasting life histories. *Heredity*, *99*(4), 466–476. https://doi.org/10.1038/sj.hdy.6801027

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. https://doi.org/10.1111/1755-0998.12592

Mable, B. K. (2019). Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conservation Genetics*, *20*(1), 89–100.

https://doi.org/10.1007/s10592-018-1129-9

Madsen, S. S., Skovbølling, S., Nielsen, C., & Korsgaard, B. (2004). 17-β Estradiol and 4-nonylphenol delay smolt development and downstream migration in Atlantic salmon, Salmo salar. *Aquatic Toxicology*, *68*(2), 109–120. https://doi.org/10.1016/j.aquatox.2004.03.008

Magalhaes, I. S., Whiting, J. R., D'Agostino, D., Hohenlohe, P. A., Mahmud, M., Bell, M. A., Skúlason, S., & MacColl, A. D. C. (2021). Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nature Ecology and Evolution*, *5*(2), 251–261. https://doi.org/10.1038/s41559-020-01341-8

Manitoba Government. (2018). Fish Populations. *Lake Winnipeg Basin Indicator Series*. https://www.gov.mb.ca/sd/pubs/fish_wildlife/fish/indicators_fishpop.pdf

Manitoba Government. (2020). Lake Information for Anglers. Lake Information for Anglers. https://experience.arcgis.com/experience/2557cda82dcc4a348fbb71304cedcf6d/page/page_2/

Manitoba Sustainable Development. (2018). Manitoba Sustainable Development Annual Report 2017-2018. https://www.gov.mb.ca/sd/pubs/annual-reports/annual_report_2017_18.pdf

Manitoba Sustainable Development. (2019). *Lake Winnipeg Measures to Enhance Sustainability*.

Mantione, K. J., Kream, R. M., Kuzelova, H., Ptacek, R., Raboch, J., Samuel, J. M., & Stefano, G. B. (2014). Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Medical Science Monitor Basic Research*, *20*, 138–142. https://doi.org/10.12659/MSMBR.892101

Marshall, W. S., Breves, J. P., Doohan, E. M., Tipsmark, C. K., Kelly, S. P., Robertson, G. N., & Schulte, P. M. (2018). Claudin-10 isoform expression and cation selectivity change with salinity in salt-secreting epithelia of Fundulus heteroclitus. *Journal of Experimental Biology*, *221*(1). https://doi.org/10.1242/jeb.168906

Mayor, S. J., Schneider, D. C., Schaefer, J. A., & Mahoney, S. P. (2009). Habitat selection at multiple scales. *Ecoscience*, *16*(2), 238–247. https://doi.org/10.2980/16-2-3238

McCormick, S.D, & Saunders, R. L. (1987). Preparatory physiological adaptations for marine life in salmonids: osmoregulation, growth and metabolism. *American Fisheries Society Symposium*, 211–229.

McCormick, Stephen D., Hansen, L. P., Quinn, T. P., & Saunders, R. L. (1998). Movement, migration, and smolting of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(S1), 77–92. https://doi.org/10.1139/cjfas-55-S1-77

McGee, M. D., Neches, R. Y., & Seehausen, O. (2016). Evaluating genomic divergence and parallelism in replicate ecomorphs from young and old cichlid adaptive radiations. *Molecular Ecology*, *25*(1), 260–268. https://doi.org/10.1111/mec.13463

McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*(4), 656–669. https://doi.org/10.1111/1755-0998.12613

McMahon, T. E., & Matter, W. J. (2006). Linking habitat selection, emigration and population dynamics of freshwater fishes: A synthesis of ideas and approaches. *Ecology of Freshwater Fish*, *15*(2), 200–210. https://doi.org/10.1111/j.1600-0633.2006.00130.x

Meador, M. R., & Kelso, W. E. (1989). Behavior and Movements of Largemouth Bass in Response to Salinity. *Transactions of the American Fisheries Society*, *118*(4), 409–415. https://doi.org/10.1577/1548-8659(1989)118<0409:BAMOLB>2.3.CO;2

Meka, J. M., & McCormick, S. D. (2005). Physiological response of wild rainbow trout to angling: impact of angling duration, fish size, body condition, and temperature. *Fisheries Research*, *72*(2–3), 311–322. https://doi.org/10.1016/j.fishres.2004.10.006

Miller, C. S., Hebblewhite, M., Goodrich, J. M., & Miquelle, D. G. (2010). Review of research methodologies for tigers: Telemetry. *Integrative Zoology*, *5*(4), 378–389. https://doi.org/10.1111/j.1749-4877.2010.00216.x

Miller, K. M., Li, S., Kaukinen, K. H., Ginther, N., Hammill, E., Curtis, J. M. R., Patterson, D. A., Sierocinski, T., Donnison, L., Pavlidis, P., Hinch, S. G., Hruska, K. A., Cooke, S. J.,

English, K. K., & Farrell, A. P. (2011). Genomic Signatures Predict Migration and Spawning Failure in Wild Canadian Salmon. *Science*, *331*(6014), 214–217. https://doi.org/10.1126/science.1196901

Miller, Kristina M., Schulze, A. D., Ginther, N., Li, S., Patterson, D. A., Farrell, A. P., & Hinch, S. G. (2009). Salmon spawning migration: Metabolic shifts and environmental triggers. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *4*(2), 75–89. https://doi.org/10.1016/j.cbd.2008.11.002

Miller, Kristina M, Li, S., Kaukinen, K. H., Ginther, N., Hammill, E., Curtis, J. M. R., Patterson, D. A., Sierocinski, T., Donnison, L., Pavlidis, P., Hinch, S. G., Hruska, K. A., Cooke, S. J., English, K. K., & Farrell, A. P. (2011). Genomic signatures predict migration and spawning failure in wild Canadian salmon. *Science, 331*(6014), 214–217. https://doi.org/10.1126/science.1196901

Minke-Martin, V., Hinch, S. G., Braun, D. C., Burnett, N. J., Casselman, M. T., Eliason, E. J., & Middleton, C. T. (2018). Physiological condition and migratory experience affect fitness-related outcomes in adult female sockeye salmon. *Ecology of Freshwater Fish*, *27*(1), 296–309. https://doi.org/10.1111/eff.12347

Moles, M. D., Robinson, B. W., Johnston, T. A., Cunjak, R. A., Jardine, T. D., Casselman, J. M., & Leggett, W. C. (2010). Morphological and trophic differentiation of growth morphotypes of walleye (*Sander vitreus*) from Lake Winnipeg, Canada. *Canadian Journal of Zoology*, *88*(10), 950–960. https://doi.org/10.1139/Z10-062

Moles, Matthew D., Johnston, T. A., Robinson, B. W., Leggett, W. C., & Casselman, J. M. (2008). Is gonadal investment in walleye (*Sander vitreus*) dependent on body lipid reserves? A multipopulation comparative analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(4), 600–614. https://doi.org/10.1139/F07-186

Mommsen, T. P. (2004). Salmon spawning migration and muscle protein metabolism: The August Krogh principle at work. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, *139*(3 SPEC.ISS.), 383–400. https://doi.org/10.1016/j.cbpc.2004.09.018

Montgomery, J. C., Tolimieri, N., & Haine, O. S. (2001). Active habitat selection by pre-settlement reef fishes. *Fish and Fisheries*, *2*(3), 261–277. https://doi.org/10.1046/j.1467-2960.2001.00053.x

Moore, J.-S., Harris, L. N., Le Luyer, J., Sutherland, B. J. G., Rougemont, Q., Tallman, R. F., Fisk, A. T., & Bernatchez, L. (2017). Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic Char. *Molecular Ecology*, *26*(24), 6784–6800. https://doi.org/10.1111/mec.14393

Moore, J., Harris, L. N., Tallman, R. F., & Taylor, E. B. (2013). The interplay between dispersal and gene flow in anadromous Arctic char (*Salvelinus alpinus*): implications for potential for local adaptation. Canadian Journal of Fisheries and Aquatic Sciences, 70(9), 1327–1338. https://doi.org/10.1139/cjfas-2013-0138

Moore, J. S., Loewen, T. N., Harris, L. N., & Tallman, R. F. (2014). Genetic analysis of sympatric migratory ecotypes of arctic charr salvelinus alpinus: Alternative mating tactics or reproductively isolated strategies? *Journal of Fish Biology*, *84*(1), 145–162. https://doi.org/10.1111/jfb.12262

Morrissey, J. F., & Gruber, S. H. (1993). Habitat selection by juvenile lemon sharks, *Negaprion brevirostris*. *Environmental Biology of Fishes*, *38*(4), 311–319. https://doi.org/10.1007/BF00007524

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *ELife*, *9*, 1–52. https://doi.org/10.7554/eLife.48376

Munaweera Arachchilage, I. P., Muthukumarana, S., Gillis, D. M., Watkinson, D. A., Charles, C., & Enders, E. C. (2021). Assessing Movement Patterns using Bayesian State-Space Models on Lake Winnipeg Walleye. *Canadian Journal of Fisheries and Aquatic Sciences*, 1–48. https://doi.org/10.1139/cjfas-2020-0262

Murphy, M. D., & Crabtree, R. E. (2001). Changes in the Age Structure of Nearshore Adult Red Drum off West-Central Florida Related to Recruitment and Fishing Mortality. *North American Journal of Fisheries Management*, *21*(3), 671–678. https://doi.org/10.1577/1548-

8675(2001)021<0671:CITASO>2.0.CO;2

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*(49), 19052–19059. https://doi.org/10.1073/pnas.0800375105

Nichols, J. D. (1992). Capture-Recapture Models. *BioScience*, *42*(2), 94–102. https://doi.org/10.2307/1311650

Nielsen, C., Aarestrup, K., Nørum, U., & Madsen, S. S. (2004). Future migratory behaviour predicted from premigratory levels of gill Na+/K+-ATPase activity in individual wild brown trout (*Salmo trutta*). *Journal of Experimental Biology*, *207*(3), 527–533. https://doi.org/10.1242/jeb.00783

O'Malley, K. G., Jacobson, D. P., Kurth, R., Dill, A. J., & Banks, M. A. (2013). Adaptive genetic markers discriminate migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*) amid continued gene flow. *Evolutionary Applications*, *6*(8), 1184–1194. https://doi.org/10.1111/eva.12095

Osmond, M. M., & Coop, G. (2020). Genetic Signatures of Evolutionary Rescue by a Selective Sweep. Genetics, 215(3), 813–829. https://doi.org/10.1534/genetics.120.303173

Östergren, J., Nilsson, J., & Lundqvist, H. (2012). Linking genetic assignment tests with telemetry enhances understanding of spawning migration and homing in sea trout *Salmo trutta* L. *Hydrobiologia*, *691*(1), 123–134. https://doi.org/10.1007/s10750-012-1063-7

Passow, C. N., Henpita, C., Shaw, J. H., Quackenbush, C. R., Warren, W. C., Schartl, M., Arias-Rodriguez, L., Kelley, J. L., & Tobler, M. (2017). The roles of plasticity and evolutionary change in shaping gene expression variation in natural populations of extremophile fish. *Molecular Ecology*, *26*(22), 6384–6399. https://doi.org/10.1111/mec.14360

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Payne, N. L., Smith, J. A., van der Meulen, D. E., Taylor, M. D., Watanabe, Y. Y., Takahashi, A., Marzullo, T. A., Gray, C. A., Cadiou, G., & Suthers, I. M. (2016). Temperature dependence of fish performance in the wild: Links with species biogeography and physiological thermal tolerance. *Functional Ecology*, *30*(6), 903–912. https://doi.org/10.1111/1365-2435.12618

Persson, E. (2016). What are the core ideas behind the Precautionary Principle? *Science of the Total Environment*, *557–558*, 134–141. https://doi.org/10.1016/j.scitotenv.2016.03.034

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., Rödel, M. O., Blackburn, D. C., & Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245–5263. https://doi.org/10.1111/mec.14266

Pratlong, M., Haguenauer, A., Chabrol, O., Klopp, C., Pontarotti, P., & Aurelle, D. (2015). The red coral (*Corallium rubrum*) transcriptome: a new resource for population genetics and local adaptation studies. *Molecular Ecology Resources*, *15*(5), 1205–1215. https://doi.org/10.1111/1755-0998.12383

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., Hotaling, T. J., Spidle, A. P., & Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances*, *3*(8). https://doi.org/10.1126/sciadv.1603198

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology*, *20*(4), R208–R215. https://doi.org/10.1016/j.cub.2009.11.055

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945 LP – 959. http://www.genetics.org/content/155/2/945.abstract

Privé, F., Luu, K., Vilhjálmsson, B. J., Blum, M. G. B., & Rosenberg, M. (2020). Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. *Molecular Biology and Evolution*, *37*(7), 2153–2154.

https://doi.org/10.1093/molbev/msaa053

Prockop, D. J., & Sioerdsma, A. (1961). Significance of urinary hydroxyproline in man. *Journal of Clinical Investigation*, *40*(5), 843–849. https://doi.org/10.1172/JCI104318

Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van De Peer, Y., & Vandepoele, K. (2012). i-ADHoRe 3.0-fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Research*, *40*(2). https://doi.org/10.1093/nar/gkr955

Pruyne, D., & Bretscher, A. (2000). Polarization of cell growth in yeast. *Journal of Cell Science*, *113*(4), 571 LP – 585. http://jcs.biologists.org/content/113/4/571.abstract

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., … Wishart, D. S. (2011). The Human Serum Metabolome. *PLoS ONE*, *6*(2), e16957. https://doi.org/10.1371/journal.pone.0016957

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 tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.r-project.org/

Raabe, J. K., VanDeHey, J. A., Zentner, D. L., Cross, T. K., & Sass, G. G. (2020). Walleye inland lake habitat: considerations for successful natural recruitment and stocking in North Central North America. *Lake and Reservoir Management*, *36*(4), 335–359. https://doi.org/10.1080/10402381.2019.1697771

Raby, G. D., Vandergoot, C. S., Hayden, T. A., Faust, M. D., Kraus, R. T., Dettmers, J. M., Cooke, S. J., Zhao, Y., Fisk, A. T., & Krueger, C. C. (2018). Does behavioural thermoregulation underlie seasonal movements in lake erie walleye? *Canadian Journal of Fisheries and Aquatic Sciences*, *75*(3), 488–496. https://doi.org/10.1139/cjfas-2017-0145

Rasmussen, P. W., Heisey, D. M., Gilbert, S. J., King, R. M., & Hewett, S. W. (2002). Estimating Postspawning Movement of Walleyes among Interconnected Lakes of Northern Wisconsin. *Transactions of the American Fisheries Society*, *131*(6), 1020–1032. https://doi.org/10.1577/1548-8659(2002)131<1020:EPMOWA>2.0.CO;2

Rebl, A., Verleih, M., Köbis, J. M., Kühn, C., Wimmers, K., Köllner, B., & Goldammer, T. (2013). Transcriptome Profiling of Gill Tissue in Regionally Bred and Globally Farmed Rainbow Trout Strains Reveals Different Strategies for Coping with Thermal Stress. *Marine Biotechnology*, *15*(4), 445–460. https://doi.org/10.1007/s10126-013-9501-8

Reich, D. E., Cargili, M., Boik, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., & Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, *411*(6834), 199–204. https://doi.org/10.1038/35075590

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T. J., Kidd, K. A., MacCormack, T. J., Olden, J. D., Ormerod, S. J., Smol, J. P., Taylor, W. W., Tockner, K., Vermaire, J. C., Dudgeon, D., & Cooke, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, *94*(3), 849–873. https://doi.org/10.1111/brv.12480

Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: Mismatch between biological and fisheries management units. *Fish and Fisheries*, *10*(4), 361–395. https://doi.org/10.1111/j.1467-2979.2008.00324.x

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler IV, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11479–11484. https://doi.org/10.1073/pnas.201394398

Rempel, L. L., & Smith, D. G. (1998). Postglacial fish dispersal from the Mississippi refuge to the Mackenzie River basin. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(4), 893–899. https://doi.org/10.1139/f97-257

Rennie, M. D., Collins, N. C., Shuter, B. J., Rajotte, J. W., & Couture, P. (2005). A comparison of methods for estimating activity costs of wild fish populations: More active fish observed to grow slower. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(4), 767–780. https://doi.org/10.1139/f05-052

Reynolds, R. G., & Fitzpatrick, B. M. (2013). Tests of two methods for identifying founder effects in metapopulations reveal substantial type II error. *Genetica*, *141*(1–3), 119–131. https://doi.org/10.1007/s10709-013-9711-z

Rideout, R. M., & Tomkiewicz, J. (2011). Skipped spawning in fishes: More common than you might think. *Marine and Coastal Fisheries*, *3*(1), 176–189. https://doi.org/10.1080/19425120.2011.556943

Rudolfsen, T. A., Watkinson, D. A., Charles, C., Kovachik, C., & Enders, E. C. (2021). Developing habitat associations for fishes in Lake Winnipeg by linking large scale bathymetric and substrate data with fish telemetry detections. *Journal of Great Lakes Research*, *xxxx*. https://doi.org/10.1016/j.jglr.2021.02.002

Russello, M. A., Kirk, S. L., Frazer, K. K., & Askey, P. J. (2012). Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications*, *5*(1), 39–52. https://doi.org/10.1111/j.1752-4571.2011.00206.x

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D., & Lander, E. S. (2006). Positive natural selection in the human lineage. *Science*, *312*(5780), 1614–1620. https://doi.org/10.1126/science.1124309

Sabeti, Pardis C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., … Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913–918. https://doi.org/10.1038/nature06250

Saude, E. J., Slupsky, C. M., & Sykes, B. D. (2006). Optimization of NMR analysis of biological fluids for quantitative accuracy. *Metabolomics*, *2*(3), 113–123. https://doi.org/10.1007/s11306-006-0023-5

Schindler, D. E., Armstrong, J. B., & Reed, T. E. (2015). The portfolio concept in ecology and evolution. *Frontiers in Ecology and the Environment*, *13*(5), 257–263. https://doi.org/10.1890/140275

Schindler, D. W. (2001). The cumulative effect of climate change and multiple stressors. *Canadian Journal of Fisheries and Aquatic Sciences*, *2958*, 18–29. https://doi.org/10.1139/f00-179

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Schuhmacher, A., Schön, J., Goldberg, M., & Gropp, J. M. (1995). Plasma amino acid levels in rainbow trout (*Oncorhynchus mykiss*). *Journal of Applied Ichthyology*, *11*(3–4), 309–316. https://doi.org/10.1111/j.1439-0426.1995.tb00032.x

Schunter, C., Garza, J. C., Macpherson, E., & Pascual, M. (2014). SNP development from RNA-seq data in a nonmodel fish: how many individuals are needed for accurate allele frequency prediction? *Molecular Ecology Resources*, *14*(1), 157–165. https://doi.org/10.1111/1755-0998.12155

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *The American Statistician*, *34*(4), 216–221. https://doi.org/10.1080/00031305.1980.10483031

Sheppard, K. T., Davoren, G. K., & Hann, B. J. (2015). Diet of walleye and sauger and morphological characteristics of their prey in Lake Winnipeg. *Journal of Great Lakes Research*, *41*(3), 907–915. https://doi.org/10.1016/j.jglr.2015.05.006

Sheppard, K. T., Hann, B. J., & Davoren, G. K. (2018). Growth rate and condition of walleye (*Sander vitreus*), sauger (*Sander canadensis*), and dwarf walleye in a large Canadian lake. *Canadian Journal of Zoology*, *96*(7), 739–747. https://doi.org/10.1139/cjz-2017-0276

Simonds, W. J. (1998). The Cedar Bluff Unit. *Smoky Hill Division. Pick-Sloan Missouri Basin Program.*

Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E., & McEwan, J. C.

(2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: Theoretical expectations and empirical data. *Heredity*, *93*(3), 255–265. https://doi.org/10.1038/sj.hdy.6800485

Smith, R. J. (2016). Explanations for adaptations, just-so stories, and limitations on evidence in evolutionary biology. *Evolutionary Anthropology*, *25*(6), 276–287. https://doi.org/10.1002/evan.21495

Sneddon, L. U. (2019). Evolution of nociception and pain: Evidence from fish models. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1785). https://doi.org/10.1098/rstb.2019.0290

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., Buerkle, C. A., Feder, J. L., Bast, J., Schwander, T., Egan, S. P., Crespi, B. J., & Nosil, P. (2014). Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. *Science*, *344*(6185), 738–742. https://doi.org/10.1126/science.1252136

Stadtman, E. R., Van Remmen, H., Richardson, A., Wehr, N. B., & Levine, R. L. (2005). Methionine oxidation and aging. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1703*(2), 135–140. https://doi.org/10.1016/j.bbapap.2004.08.010

Stepien, C. A., Murphy, D. J., Lohner, R. N., Sepulveda-Villet, O. J., & Haponski, A. E. (2009). Signatures of vicariance, postglacial dispersal and spawning philopatry: Population genetics of the walleye *Sander vitreus*. *Molecular Ecology*, *18*(16), 3411–3428. https://doi.org/10.1111/j.1365-294X.2009.04291.x

Stepien, Carol A., Snyder, M. R., & Knight, C. T. (2018). Genetic Divergence of Nearby Walleye Spawning Groups in Central Lake Erie: Implications for Management. *North American Journal of Fisheries Management*, *38*(4), 783–793. https://doi.org/10.1002/nafm.10176

STETTEN, M. R. (1949). Some aspects of the metabolism of hydroxyproline, studied with the aid of isotopic nitrogen. *The Journal of Biological Chemistry*, *181*(1), 31–37. http://www.ncbi.nlm.nih.gov/pubmed/15390388

Stevenson, C. F., Bass, A. L., Furey, N. B., Miller, K. M., Li, S., Rechisky, E. L., Porter, A. D., Welch, D. W., & Hinch, S. G. (2020). Infectious agents and gene expression differ between sockeye salmon (*Oncorhynchus nerka*) smolt age classes but do not predict migration survival. *Canadian Journal of Fisheries and Aquatic Sciences*, *77*(3), 484–495. https://doi.org/10.1139/cjfas-2019-0113

Stich, D. S., Zydlewski, G. B., Kocik, J. F., & Zydlewski, J. D. (2015). Linking Behavior, Physiology, and Survival of Atlantic Salmon Smolts During Estuary Migration. *Marine and Coastal Fisheries*, *7*(1), 68–86. https://doi.org/10.1080/19425120.2015.1007185

Storey A. J.; Dabney, A.; Robinson, D., J. D. . B. (2020). qvalue: Q-value estimation for false discovery rate control. In *R package version 2.22.0: Vol. null* (Issue null, p. null-null). http://github.com/jdstorey/qvalue

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9440–9445. https://doi.org/10.1073/pnas.1530509100

Su, G., Logez, M., Xu, J., Tao, S., Villéger, S., & Brosse, S. (2021). Human impacts on global freshwater fish biodiversity. *Science*, *371*(6531), 835–838. https://doi.org/10.1126/science.abd3369

Sumpter, J. P., Le Bail, P. Y., Pickering, A. D., Pottinger, T. G., & Carragher, J. F. (1991). The effect of starvation on growth and plasma growth hormone concentrations of rainbow trout, *Oncorhynchus mykiss*. *General and Comparative Endocrinology*, *83*(1), 94–102. https://doi.org/10.1016/0016-6480(91)90109-J

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, *6*(7). https://doi.org/10.1371/journal.pone.0021800

Sveen, L. R., Timmerhaus, G., Torgersen, J. S., Ytteborg, E., Jørgensen, S. M., Handeland, S., Stefansson, S. O., Nilsen, T. O., Calabrese, S., Ebbesson, L., Terjesen, B. F., & Takle, H. (2016). Impact of fish density and specific water flow on skin properties in Atlantic salmon (*Salmo salar* L.) post-smolts. *Aquaculture*, *464*, 629–637. https://doi.org/10.1016/j.aquaculture.2016.08.012

Thompson, D. H. (1931). Variation in fishes as a function of distance. *Transactions of the Illinois State Academy of Science*, *23*, 276–281.

Thompson, N. F., Anderson, E. C., Clemento, A. J., Campbell, M. A., Pearse, D. E., Hearsey, J. W., Kinziger, A. P., & Garza, J. C. (2020). A complex phenotype in salmon controlled by a simple change in migratory timing. *Science*, *370*(6516), 609–613. https://doi.org/10.1126/SCIENCE.ABA9059

Thorstad, E. B., Rikardsen, A. H., Alp, A., & Okland, F. (2013). The Use of Electronic Tags in Fish Research - An Overview of Fish Telemetry Methods. *Turkish Journal of Fisheries and Aquatic Sciences*, *13*(January), 881–896. https://doi.org/10.4194/1303-2712-v13

Thorstensen, M., Bates, P., Lepla, K., & Schreier, A. (2019). To breed or not to breed? Maintaining genetic diversity in white sturgeon supplementation programs. *Conservation Genetics*, *20*(5), 997–1007. https://doi.org/10.1007/s10592-019-01190-4

Tian, D., Araki, H., Stahl, E., Bergelson, J., & Kreitman, M. (2002). Signature of balancing selection in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11525–11530. https://doi.org/10.1073/pnas.172203599

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, *25*(10), 2144–2164. https://doi.org/10.1111/mec.13606

Trefilov, A., Berard, J., Krawczak, M., & Schmidtke, J. (2000). Natal dispersal in rhesus macaques is related to serotonin transporter gene promoter variation. *Behavior Genetics*, *30*(4), 295–301. https://doi.org/10.1023/A:1026597300525

Trépanier, S., Rodríguez, M. A., & Magnan, P. (1996). Spawning migrations in landlocked Atlantic salmon: Time series modelling of river discharge and water temperature effects. *Journal of Fish Biology*, *48*(5), 925–936. https://doi.org/10.1111/j.1095-8649.1996.tb01487.x

Turner, N. A., Charles, C., Watkinson, D. A., Enders, E. C., Klein, G., & Rennie, M. D. (2021). Historical and contemporary movement and survival rates of walleye (*Sander vitreus*) in Lake Winnipeg, Canada. *Journal of Great Lakes Research*, *xxxx*.

https://doi.org/10.1016/j.jglr.2021.01.012

Vallance, P., & Leiper, J. (2004). Cardiovascular Biology of the Asymmetric Dimethylarginine:Dimethylarginine Dimethylaminohydrolase Pathway. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *24*(6), 1023–1030. https://doi.org/10.1161/01.ATV.0000128897.54893.26

Van Der Maaten, L., & Hinton, G. (2012). Visualizing non-metric similarities in multiple maps. *Machine Learning*, *87*(1), 33–55. https://doi.org/10.1007/s10994-011-5273-4

Velasco, C., Moreiras, G., Conde-Sieira, M., Leao, J. M., Míguez, J. M., & Soengas, J. L. (2017). Ceramide counteracts the effects of ghrelin on the metabolic control of food intake in rainbow trout. *Journal of Experimental Biology*, *220*(14), 2563–2576. https://doi.org/10.1242/jeb.159871

Veldhoen, N., Ikonomou, M. G., Dubetz, C., MacPherson, N., Sampson, T., Kelly, B. C., & Helbing, C. C. (2010). Gene expression profiling and environmental contaminant assessment of migrating Pacific salmon in the Fraser River watershed of British Columbia. *Aquatic Toxicology*, *97*(3), 212–225. https://doi.org/10.1016/j.aquatox.2009.09.009

Veldhoen, N., Ikonomou, M. G., Rehaume, V., Dubetz, C., Patterson, D. A., & Helbing, C. C. (2013). Evidence of disruption in estrogen-associated signaling in the liver transcriptome of in-migrating sockeye salmon of British Columbia, Canada. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*, *157*(2), 150–161. https://doi.org/10.1016/j.cbpc.2012.10.007

Verta, J. P., & Jones, F. C. (2019). Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *ELife*, *8*, 1–30. https://doi.org/10.7554/eLife.43785

Wang, J. (2004). Estimating pairwise relatedness from dominant genetic markers. *Molecular Ecology*, *13*(10), 3169–3178. https://doi.org/10.1111/j.1365-294X.2004.02298.x

Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, *89*(5), 438–450. https://doi.org/10.1093/jhered/89.5.438

Waples, Robin S., & Allendorf, F. (2015). Testing for hardy-weinberg proportions: Have we lost the plot? *Journal of Heredity*, *106*(1), 1–19. https://doi.org/10.1093/jhered/esu062

Waples, Robin S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, *3*(3), 244–262. https://doi.org/10.1111/j.1752-4571.2009.00104.x

Waples, Robin S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*(6), 1419–1439. https://doi.org/10.1111/j.1365-294X.2006.02890.x

Wasteneys, G. O., & Galway, M. E. (2003). Remodeling the Cytoskeleton for Growth and Form: An Overview with Some New Views. *Annual Review of Plant Biology*, *54*, 691–722. https://doi.org/10.1146/annurev.arplant.54.031902.134818

Watkinson, D. A., & Gillis, D. M. (2005). Stock discrimination of Lake Winnipeg walleye based on Fourier and wavelet description of scale outline signals. *Fisheries Research*, *72*(2–3), 193–203. https://doi.org/10.1016/j.fishres.2004.11.002

Weinz, A. A., Matley, J. K., Klinard, N. V., Fisk, A. T., & Colborne, S. F. (2020). Identification of predation events in wild fish using novel acoustic transmitters. *Animal Biotelemetry*, *8*(1), 1–14. https://doi.org/10.1186/s40317-020-00215-x

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Weir, B. S., & Goudet, J. (2017). A unified characterization of population structure and relatedness. *Genetics*, *206*(4), 2085–2103. https://doi.org/10.1534/genetics.116.198424

Weng, Z., Zhang, Z., Zhang, Q., Fu, W., He, S., & Ding, X. (2013). Comparison of different imputation methods from low- to high-density panels using Chinese Holstein cattle. *Animal*, *7*(5), 729–735. https://doi.org/10.1017/S1751731112002224

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilson, J. F., & Goldstein, D. B. (2000). Consistent long-range linkage disequilibrium generated by admixture in a bantu-semitic hybrid population. *American Journal of Human Genetics*, *67*(4), 926–935. https://doi.org/10.1086/303083

Wolfert, D. R. (1969). Maturity and Fecundity of Walleyes from the Eastern and Western Basins of Lake Erie. *Journal of the Fisheries Research Board of Canada*, *26*(7), 1877–1888. https://doi.org/10.1139/f69-171

Wright, S. (1943). Isolation by Distance. *Genetics*, *28*(2), 114–138. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209196/

Wright, S. (1950). Genetical Structure of Populations. *Nature*, *166*(4215), 247–249.

WWF. (2016). *Living Planet Report 2016. Risk and resilience in a new era.* (Issue 2). https://www.worldwildlife.org/pages/living-planet-report-2016

WWF. (2018). *Living Planet Report 2018: Aiming higher.* https://www.worldwildlife.org/pages/living-planet-report-2018

Xie, C., & Xu, S. (1998). Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. *Heredity*, *80*(4), 489–498. https://doi.org/10.1046/j.1365-2540.1998.00308.x

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M., & Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. *Current Protocols*, *1*(3), 1–51. https://doi.org/10.1002/cpz1.90

Yamanaka, H., & Minamoto, T. (2016). The use of environmental DNA of fishes as an efficient method of determining habitat connectivity. *Ecological Indicators*, *62*, 147–153.

https://doi.org/10.1016/j.ecolind.2015.11.022

Yan, J., Song, Z., Xu, Q., Kang, L., Zhu, C., Xing, S., Liu, W., Greimler, J., Züst, T., Li, J., & Sang, T. (2017). Population transcriptomic characterization of the genetic and expression variation of a candidate progenitor of *Miscanthus* energy crops. *Molecular Ecology*, *26*(21), 5911–5922. https://doi.org/10.1111/mec.14338

Yang, R.-C. (1998). Estimating Hierarchical F-Statistics. *Evolution*, *52*(4), 950. https://doi.org/10.2307/2411227

Yang, Y., Wang, Q., Chen, Q., Liao, R., Zhang, X., Yang, H., Zheng, Y., Zhang, Z., & Pan, Y. (2014). A new genotype imputation method with tolerance to high missing rate and rare variants. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0101025

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, *35*(10), 1786–1788. https://doi.org/10.1093/bioinformatics/bty875

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zhou, X. (2017). A unified framework for variance component estimation with summary statistics in genome-wide association studies. *Annals of Applied Statistics*, *11*(4), 2027–2051. https://doi.org/10.1214/17-AOAS1052

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310