A chromosomal inversion may facilitate adaptation despite periodic gene flow in a freshwater fish

Matt J. Thorstensen[1], Peter T. Euclide[2], Jennifer D. Jeffrey[1,3], Yue Shi[2,4], Jason R. Treberg[1], Douglas A. Watkinson[5], Eva C. Enders[5], Wesley A. Larson[2,6], Yasuhiro Kobayashi[7,8], and Ken M. Jeffries[1]

[1] Department of Biological Sciences, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

[2] U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, Stevens Point, WI 54481-3897, USA

[3] Department of Biology, University of Winnipeg, Richardson College, 2RC047, 515 Portage Avenue, Winnipeg, Manitoba, R3B 2E9, Canada

[4] College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, 17101 Point Lena Loop Road, Juneau, AK 99801, USA

[5] Freshwater Institute, Fisheries and Oceans Canada, Winnipeg, MB R3T 2N6, Canada

[6] National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, 17109 Point Lena Loop Road, Juneau, AK 99801, USA

[7] Department of Biological Sciences, Fort Hays State University, 600 Park Street, Hays, KS 67601, USA

[8] Department of Biology, The College of St. Scholastica, 1200 Kenwood Avenue, Duluth, MN 55811, USA

Abstract

Genomic architecture, such as chromosomal inversions, may play an important role in facilitating adaptation despite opportunities for gene flow. One system where chromosomal inversions may be important for eco-evolutionary dynamics are in freshwater fish, which often live in heterogenous environments characterized by varying levels of connectivity and varying opportunities for gene flow. In the present study, reduced representation sequencing was used to study possible adaptation in *n*=345 walleye (*Sander vitreus*) from three North American waterbodies: Cedar Bluff Reservoir (Kansas, USA), Lake Manitoba (Manitoba, Canada), and Lake Winnipeg (Manitoba, Canada). Haplotype and outlier-based tests revealed a putative chromosomal inversion that contained three expressed genes and was nearly fixed for alternate genotypes in each Canadian lake. These patterns exist despite several opportunities for gene flow between these proximate Canadian lakes, suggesting that the inversion may be important for facilitating adaptative divergence between the two lakes despite gene flow. Our study illuminates the importance of genomic architecture for facilitating local adaptation in freshwater fish and provides additional evidence that inversions may facilitate local adaptation in many organisms that inhabit connected but heterogenous environments.

Keywords

selection, haplotype, fishery, genomic architecture, adaptive variation, walleye, *Sander vitreus*

Introduction

Genomic architecture such as structural variants may help to facilitate local adaptation, and are therefore key mechanisms of eco-evolutionary processes (Dorant et al., 2020; Shi et al., 2021; Tigano & Friesen, 2016; Wellenreuther & Bernatchez, 2018). One type of structural variant, chromosomal inversions, are taxonomically widespread reversed regions of DNA that have been linked to both adaptation and speciation (see Wellenreuther & Bernatchez, 2018 for a review). Inversions can effectively isolate large chromosomal regions containing hundreds of genes from recombination, facilitating local adaptation even in the face of high gene flow (Wellenreuther & Bernatchez, 2018). Most inversions found in previous studies have been relatively large because sequencing methods can more easily detect large inversions (Wellenreuther & Bernatchez, 2018). However, inversions of any size may be important, as intermediate-to-large inversions have been linked to local adaptation, but small inversions were preferentially fixed in directly beneficial mutation models and contributed the most to genome evolution in neutral and underdominant scenarios according to a simulation study (Connallon & Olito, 2021). Among fishes, chromosomal inversions have largely been characterized in well studies species such as salmonids (Arostegui et al., 2019; Leitwein et al., 2017; G. McKinney et al., 2020; Pearse et al., 2014) and gadids (Berg et al., 2017; Berg et al., 2016; Kirubakaran et al., 2016; Puncher et al., 2019; Sinclair-Waters et al., 2018; Sodeland et al., 2016), but little information exists on the importance of inversions in most taxa especially those inhabiting freshwater (Shi et al., 2021).

Freshwater fish often live in heterogenous environments characterized by varying levels of connectivity and therefore, varying opportunities for gene flow (Griffiths, 2015; Mushet et al., 2019). Gene flow tends to work in contrast to local adaptation (Lenormand, 2002) but

nevertheless, evidence for local adaptation has been observed among different fish that use freshwater habitats (see Fraser et al., 2011 for a review of salmonids; Shi et al., 2021). While genomic architecture has been proposed to facilitate local adaptation (Tigano & Friesen, 2016), few studies test for this possibility in freshwater fish (Shi et al., 2021). Freshwater habitats contribute a disproportionate richness in species diversity to global species diversity, given that only 0.01% of the planet is fresh water (Balian et al., 2008). Fish are a significant fraction of overall freshwater species diversity (approx. 10% of overall freshwater species and 70% of vertebrate freshwater species are fish; Balian et al., 2008). Therefore, an understanding of the mechanisms of local adaptation in freshwater fish can contribute substantially to our overall understanding of local adaptation in heterogenous freshwater environments.

Walleye (*Sander vitreus*, but see Bruner, 2021) is a freshwater fish with a native range from the Northwest Territories (Canada) to Alabama (USA) (Hartman, 2009) that support important commercial and recreational fisheries (Fisheries and Oceans Canada, 2019, 2021). The economic importance of walleye underscores the need for research supporting the sustainable management of walleye fisheries, while its wide native distribution provides an opportunity to study the genomic basis of adaptation to a variety of freshwater environments. In addition, northern populations of walleye (such as those in Manitoba, Canada) were suspected 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 show evidence of divergence at adaptive markers, such as in chromosomal inversions.

To study how heterogenous environments with varying connectivity may have shaped modern walleye populations, and the potential for genomic architecture to facilitate adaptation

despite opportunities for gene flow, we studied fish from Lake Manitoba and Lake Winnipeg in Manitoba (Canada) and compared these populations with an outgroup from Cedar Bluff Reservoir in Kansas (USA) (Fig 1). Walleye in Lake Manitoba and Lake Winnipeg live in largely separated waterbodies with a shared river—Lake Manitoba drains into Lake Winnipeg via the Fairford River, Lake St. Martin, and the Dauphin River (Fig 1). As a mobile species that undergoes broad seasonal migrations, historic gene flow may have been extensive between these waterbodies (Backhouse-James & Docker, 2012; Munaweera Arachchilage et al., 2021; Thorstensen et al., 2020; Turner et al., 2021). In addition, 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. Last, walleye fry were opportunistically stocked in Lake Winnipeg from Lake Manitoba (Manitoba, Canada) between 1917-2002 (Manitoba Government, 2020). While intensive stocking complicated analyses of walleye ancestry in walleye of Minnesota and Wisconsin (USA) (Bootsma et al., 2020), stocking of walleye fry, which likely face large mortality, had little to no effect in South Dakota and Missouri (USA) (Fielder, 1992; Koppelman et al., 1992). Therefore, the genetic impacts of fry stocking in this system are likely minimal.

Here, we employed Rapture, a reduced representation approach, to genotype *n*=345 individuals from the three waterbodies described above, and population structure and demographic history were explored to provide context for signatures of selection. Simulations with EASYPOP were used to estimate historical migration rates between Lake Manitoba and Lake Winnipeg (Balloux, 2001). Candidate regions of adaptive variation were identified with XP-EHH, and outlier tests were used to confirm haplotype-based results. In addition, gene

expression in one candidate chromosomal inversion was assessed for Lake Winnipeg walleye to show the potential for functional significance of genomic architecture.

Methods

Sample Collection

Walleye anal fin clips from Lake Winnipeg individuals were collected by boat electrofishing in Spring 2017 and 2018 from spawning sites in the Red River and possible spawning sites near Matheson Island and in the Dauphin River, representing the south basin, channel, and north basin, respectively (*n*=129, 51, and 55 respectively). 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 stored in 95% ethanol. For Lake Manitoba fish, *n*=50 fry from the 2019-year class were sampled at Swan Creek Hatchery and stored in 95% ethanol. From Cedar Bluff Reservoir, anal fin clips were taken from *n*=60 adult female walleye. 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. Fish were captured using seven 25 mm-mesh trap nets and four 76 mm mesh gill nets (91.44 X 1.83 m panel per net). Fin clips were collected after gametes were harvested. In total, *n*=345 walleye were genotyped for the present study. See Figure 1 for information on sampling locations.

DNA Extraction, Sequencing, and Single Nucleotide Polymorphism 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), and aligned to the yellow perch genome with bwa v0.7.17 (Feron et al., 2020; Li & Durbin, 2009). Single nucleotide polymorphisms (SNPs) were called following best practices with RAD data (Linck & Battey, 2019), and filtered for paralogy with HDPlot (McKinney et al., 2017). Data for population structure and demographic history were pruned for linkage with PLINK v1.9 (Purcell et al., 2007). 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). Because XP-EHH requires complete data, imputation a set of SNPs with 10% missing data was performed with Beagle v5.1 (Browning et al., 2018; Browning & Browning, 2007; Weng et al., 2013; Yang et al., 2014). VCFtools v0.1.16, PGDSpider v2.1.5, 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 (Danecek et al., 2011; Knaus & Grünwald, 2017; Lischer & Excoffier, 2012; R Core Team, 2021; Wickham et al., 2019). Full details of SNP calling and processing are provided in the supplemental materials.

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).

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). 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.

Hierfstat v0.5-7 was used to find β_{WT} as a measure of population-specific differentiation from the entire pool and Weir & Cockerham's pairwise F_{ST} between populations (Goudet, 2005; Weir & Cockerham, 1984; Weir & Goudet, 2017). Ninety five percent confidence intervals were generated for F_{ST} and β_{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.

NeEstimator v2.1 was used to estimate effective population size in 95% confidence intervals for each population using the linkage disequilibrium method and only comparing SNPs on different chromosomes (Do et al., 2014). Here, datasets were filtered for 10% missing data per population using population assignments with vcftools, pulled from the pruned SNP dataset.

EASYPOP v2.0.1 (Balloux, 2001) was used to test different scenarios of historic gene flow that may have led to contemporary F_{ST} between Lake Winnipeg and Lake Manitoba. Migration rates varied between 0, 0.0001, 0.001, 0.002, 0.003, 0.004, 0.005, 0.01, 0.02, and 0.03 individuals per generation, with migration rates held constant within each simulation run. Population sizes of 4,500 individuals in Lake Winnipeg and 350 individuals in Lake Manitoba were assumed constant with equal males and females, based on estimates of N_e (see results). 1,000 simulated loci with two allelic states were used with free recombination allowed, using a KAM mutation model with mutation rate μ of 3.28×10^{-9} based on divergence between the blackfin icefish (*Chaenocephalus aceratus*) and dragonfish (*Parachaenichthys charcoti*) (Kim et al., 2019). Maximal variability (i.e., randomly assigned alleles) were chosen for the initial population. Each simulation was run over 930 generations, based on the 4.3 year generation time estimated in Franckowiak et al. (2009) and an approximate lower bound of 4,000 radiocarbon years since Lake Winnipeg had a reduced northern lakebed (Lewis et al., 2002). This reduced

northern lakebed would have prevented walleye migration between populations via the Dauphin River and Lake St. Martin in the north basin (Fig 1). 100 replicate runs of EASYPOP were used for each migration rate tested, with 800 simulation runs total. For each run, Weir & Cockerham's pairwise F_{ST} was measured with Hierfsat. These F_{ST} values were compared to the F_{ST} observed between Lake Winnipeg and Lake Manitoba walleye populations using linkage disequilibrium - pruned SNPs.

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 (Cedar Bluff Reservoir, Lake Manitoba, and Lake Winnipeg), 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).

The program 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 S1). 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.

Hierfstat v0.5-7 was used to provide supporting evidence for selection at individual SNPs using F_{ST}' with *basic.stats* on the same set of SNPs used with pcadapt. Here, F_{ST}' was measured in pairwise comparisons between individuals between each of three populations defined by admixture. In addition, absolute allele frequency differences were analyzed between populations. VCFtools was used to calculate allele frequencies for each population. Minor alleles in Lake Winnipeg-assigned individuals were used for comparisons across populations.

Putative Inversion Analysis

To assess evidence that a region of interest that we identified (chromosome 8, position 15,260,000 - 15,900,000; see results for more detail) was a putative inversion, we conducted the following analyses as suggested by Huang et al. (2020) and Shi et al. (2021). First, to test whether the region showed elevated linkage disequilibrium or LD (r^2), we calculated r^2 using PLINK v1.9 for SNPs on chromosome 8 (--r2 --ld-window-r2 0.05 --ld-window 999999999 --ld-window-kb 30000) and generated a LD heatmap (Chang et al., 2015; Purcell et al., 2007). Second, we conducted regional PCA analysis using SNPs within the region to identify whether individuals were grouped into three clusters along PC1, which is characteristic of chromosomal inversions. The discreteness of the clustering was calculated as the proportion of the between-cluster sum of squares over the total using the R function *kmeans* in *adegenet* (Jombart, 2008). Third, we compared heterozygosity (the proportion of heterozygotes) among the three identified PCA clusters using Wilcoxon tests ($\alpha = 0.05$) to further confirm that individuals in the middle PCA cluster is heterozygous for the putative inversion. Lastly, we calculated genotype frequencies of the putative inversion in all populations and assumed that the more derived inversion arrangement would have lower heterozygosity given its relatively recent origin compared to the ancestral state (Knief et al., 2016; Laayouni et al., 2003; Twyford & Friedman, 2015). Individual assignments for the three identified PCA clusters for the putative inversion were then visualized with respect to population assignments and site collected, to assess the frequency of the inversion across the different waterbodies studied.

To provide evidence of potential functional significance for the identified putative inversion on chromosome 8, gene expression data was collected for genes within the inversion. RNA-seq data was used from n=48 walleye in three sites in Lake Winnipeg, using previously

published reads (NCBI SRA database accession #PRJNA596986) (Thorstensen et al., 2020). Three genes in the yellow perch genome were within the putative inversion on chromosome 8: *PDHX*, *EHF*, and *LRRC4C*. Therefore, expression of these genes was assessed with counts per million (CPM) values from the walleye transcriptome-aligned data (Jeffrey et al., 2020; Patro et al., 2017; Soneson et al., 2015). Differential gene expression was not tested for because the putative inversion is nearly fixed in the Lake Winnipeg-assigned walleye; differential expression would thus be expected between populations and not within populations. Details for RNA-seq are provided in supplementary materials.

Results

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 a large degree of concordance between the two genomes (mean weighted synteny score = 0.999, 95% CI [0.997, 1.0]; mean relaxed synteny sore = 0.999, 95% CI [0.997, 1.0]) (Fig S2).

Population Genetics

Admixture identified K=2 populations based on the lowest cross validation error between K=1 through 6, separating Canadian and Kansas populations of walleye (Fig S3). 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, 60 individuals were assigned to the Kansas population, 67 individuals to the Lake Manitoba population, 189 individuals to a Lake Winnipeg 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 2). Population-specific differentiation from the overall pool, β_{WT}, was lowest for Cedar Bluff Reservoir (Fig S4). Filtering for 90% present data for NeEstimator v2 from the pruned SNP dataset left 11,784 SNPs available in with the Cedar Bluff Reservoir population, 16,436 SNPs in the Lake Manitoba population, and 13,213 SNPs in the Lake Winnipeg population. Linkage disequilibrium-based N_e was highest for the Lake Winnipeg population, and approximately 20-35x lower in each the Lake Manitoba and Cedar Bluff Reservoir populations (Lake Winnipeg N_e=12,084, 95% CI [9,986, 15,282]; Lake Manitoba N_e=336, 95% CI [331, 342]; Cedar Bluff Reservoir N_e=587, 95% CI [568, 608]) (Fig 2). 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 2, Fig S4).

Simulations with EASYPOP showed observed F_{ST} between Lake Manitoba and Lake Winnipeg (between 0.026 and 0.029, Fig 2) most consistent with continuous migration of 0.001 individuals per generation starting 930 generations prior to the present (Fig S5).

Signatures of Selection

The XP-EHH scores between Lake Winnipeg- and Lake Manitoba-assigned individuals identified 45 SNPs with elevated XP-EHH scores (15 SNPs 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 3). 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 3). 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 3). 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

S1). Prominent among candidate regions was a section of chromosome 8 between approximately 15.26 Mb and 15.90 Mb in Lake Winnipeg with high differentiation compared to both Lake Manitoba and Cedar Bluff Reservoir.

Two principal components (PCs) showed population structure in pcadapt, with PC1 showing 5.76% variance explained by latitudinal differences between sites and PC2 showing 2.28% variance explained by longitudinal differences (Fig 4). 345 SNPs were identified as significant outliers by pcadapt, of which 226 were significant along PC1 and 119 were significant along PC2 (Fig 4). Along PC2, between approximately 15.35 Mb and 15.91 Mb, chromosome 8 showed a region of strong differentiation between Lake Manitoba and Lake Winnipeg.

The region between 15.26 and 15.90 Mb along chromosome 8 was notable for its candidate regions identified in XP-EHH that were unique to Lake Winnipeg (Table S1), outlier SNPs in pcadapt (Fig 4), F'_{ST} outlier SNPs (Fig S6), and SNPs with absolute allele frequency differences (Fig 3, Fig S7) between Lake Manitoba and Lake Winnipeg. Many of the outlier SNPs in the chromosome 8 peak for F'_{ST} and absolute allele frequency differences overlapped with PC2 outlier SNPs from pcadapt at $q < 0.05$ (Fig 3, Fig 4, Fig S6, Fig S7).

Putative Inversion Analysis

We found strong evidence of the genomic region 15.26 - 15.90 Mb on chromosome 8 as a putative chromosomal inversion (Fig 5). This region demonstrated elevated LD with a mean r^2 of 0.30 and values up to 0.96 (Fig 5B). Regional PCA showed that individuals were clustered into three distinct groups along PC1 with a discreteness of 0.9626 (Fig 5C) and the middle PCA

cluster displaying significantly higher heterozygosity than the other two clusters (Fig 5D, Fig S8). There were also significant differences in heterozygosity between the two homokaryotypes (Fig 5D) and the arrangement with lower heterozygosity (cluster 2) was assumed to be the derived inverted type. This putative inversion (cluster 2) was nearly fixed in the Red River and Matheson Island sites of Lake Winnipeg (toward the channel and south basin; Fig 1), was at intermediate frequency in the Dauphin River (north basin) of Lake Winnipeg, and was nearly fixed or heterozygous for the opposite genotype near Swan Creek Hatchery in Lake Manitoba and in Cedar Bluff (Fig 5E). Frequency of the putative inversion visualized in relation to individual assignments and site collected revealed the inversion is nearly fixed in Lake Winnipeg-assigned walleye for one genotype, nearly fixed for the other genotype among Lake Manitoba-assigned walleye, and among unassigned, likely admixed fish in Lake Winnipeg, nearly fixed for the Lake Winnipeg genotype (Fig S9).

Three genes were identified within the putative inversion: *pyruvate dehydrogenase protein X component* (*PDHX*), *ETS homologous factor* (*EHF*), and *leucine-rich repeat-containing protein 4C* (*LRRC4C*). Each of these genes had transcripts expressed in Lake Winnipeg walleye, with counts per million between 0.4 and 20 (Fig S10).

Discussion

We used a reduced representation approach to genotype *n*=345 walleye from three different waterbodies in North America at 46,342 genetic markers. Walleye of the Canadian lakes diverged from each other despite three opportunities for gene flow: a connecting river, stocking of fry throughout the 20th century, and flooding in 2011 and 2014 (and possibly

periodically over evolutionary timescales). Historic gene flow likely occurred in sporadic pulses, as we observed evidence of limited but present contemporary gene flow possibly facilitated by flooding in 2011 and 2014. These hypothesized pulses of gene flow were consistent with population differentiation (F_{ST}) between the Canadian lakes that was lower in magnitude than each Canadian lake compared to a waterbody in Kansas, USA. Lake Winnipeg walleye showed unexpectedly strong signatures of selection on chromosome 8, with concurrent evidence for a chromosomal inversion between 15.26 and 15.90 Mb. An inversion may have played a role in adaptive divergence between Lake Winnipeg and Lake Manitoba walleye (Wellenreuther & Bernatchez, 2018; Shi et al., 2021), and three expressed genes within it indicate at least the potential for functional significance of the putative inversion.

Population Genetics

We 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. This relatively stronger differentiation between the Canadian lakes used here and the waterbody in Kansas was similar in magnitude to differentiation observed between walleye in Minnesota and Wisconsin (USA), differences which were attributed to Pleistocene lineages (Bootsma et al., 2020). Despite moderate signals of differentiation between the Canadian lakes, admixture analysis and population assignment identified 11% 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). However,

individuals possibly admixed between Lake Manitoba and Lake Winnipeg were likely the direct descendants (i.e., F_1 crosses) from individuals originating from each lake, based on admixture analyses (Bray et al., 2010; Reynolds & Fitzpatrick, 2013; Wilson & Goldstein, 2000). Therefore, we believe flooding in 2011 and 2014 is the more likely route for movement than stocking, because stocking ended in 2002, approximately 3.5 generations before sampling in 2017 and 2018, given an estimated 4.3 year generation time (Franckowiak et al., 2009). Simulations confirmed low gene flow between each lake, assuming continuous gene flow post-glaciation. Taken together, we suggest that gene flow between Lakes Manitoba and Winnipeg prior to stocking programs and recent flooding has been sporadic, and that several differentiated genomic regions suggestive of local adaptation remain in each population despite recent gene flow.

Walleye population differentiation between Lake Winnipeg and Lake Manitoba was consistent with certain other walleye populations within watersheds (e.g., Grand River, Ontario, Canada vs other Lake Erie stocks in Euclide et al., 2020; Sarah Lake vs Lake Koronis, Minnesota, USA & in Bootsma et al., 2020). Retreating glacial refugia dispersed walleye into different watersheds approximately 7,000-10,000 years ago (Bootsma et al., 2020; Stepien et al., 2009). Therefore, the population structure identified in the present study represents differentiation following geologically recent changes to the overall landscape. This differentiation is consistent with high spawning site fidelity previously observed in walleye (Horrall, 1981; Jennings et al., 1996; Stepien et al., 2009; Stepien & Faber, 1998), despite the species' otherwise broad movements in their waterbodies (Munaweera Arachchilage et al., 2021; Raby et al., 2018; Turner et al., 2021).

Given widespread stocking and movements in walleye, it may be surprising that the species shows as much population differentiation as has been observed between waterbodies in multiple watersheds (Bootsma et al., 2020; Euclide et al., 2021; Munaweera Arachchilage et al., 2021; Raby et al., 2018; Turner et al., 2021). While stocking has led to extensive homogenization in some walleye, the lack of gene flow from stocking in the present data and among other waterbodies implicates potential mechanisms by which gene flow is prevented (Bootsma et al., 2020). One possible mechanism is environmental. Here, environmental factors impede gene flow because local adaptation may select against migrants (Hendry, 2004; Sexton et al., 2014). In a comparison between environmental mechanisms of isolation and isolation by distance, isolation by environment was more common in all taxa, including vertebrates (Sexton et al., 2014). As such, environmental differences between source and stocked waterbodies may have precluded gene flow in different walleye systems. Genomic architecture is another possible mechanism that can maintain differentiation despite stocking. Here, physical linkage or chromosomal rearrangements have been hypothesized to play an important role in local adaptation in the face of gene flow (Tigano & Friesen, 2016). In other words, walleye in waterbodies with historical pulses of gene flow or recent stocking may maintain differentiation because of genomic architecture facilitating local adaptation (Tigano & Friesen, 2016). Moreover, the hypotheses that environment or genomic architecture may have maintained differentiation despite opportunities for gene flow are not mutually exclusive. Because genomic architecture has played a role in local adaptation in several freshwater fishes (Shi et al., 2021), and environment is a key variable in local adaptation (Coop et al., 2010; Forester et al., 2018), the two mechanisms may often work in tandem.

Putative Chromosomal Inversion

A 640 Kb region on chromosome 8 showed evidence as an outlier region between Lake Manitoba and Lake Winnipeg by PCA, absolute allele frequency differences, and F'_{ST}. For example, the top 10 outlier SNPs (by -$\log_{10}$ q-value) out of 119 significant outliers ($q < 0.05$) along principal component 2 in a pcadapt analysis are all from this small outlier region (Luu et al., 2017; Privé et al., 2020). Similarly, the top 10 SNPs by absolute allele frequency difference between Lake Manitoba and Lake Winnipeg are also from this small outlier region. As a haplotype-based test, XP-EHH provided evidence that this region was selected for not in Lake Manitoba, but in Lake Winnipeg instead (Sabeti et al., 2007). In a different study on walleye in Wisconsin and Minnesota (USA) (Bootsma et al., 2020), this chromosome 8 region was also a consistent outlier with F_{ST} and BAYESCAN analyses (Foll & Gaggiotti, 2008). However, SNP data were too sparse to interrogate the genomic architecture of this outlier region. In the present study, a dense RAPTURE panel of SNPs provided sufficient resolution to investigate the possibility that this outlier region was a chromosomal inversion.

Linkage disequilibrium, regional PCA, discreteness of clustering, and heterozygosity provided evidence for a putative chromosomal inversion at higher frequency in Lake Winnipeg relative to both other waterbodies. This putative inversion is nearly fixed in Lake Winnipeg-assigned walleye, and at intermediate frequencies in walleye assigned to the other populations in this study. Among Lake Manitoba-assigned fish caught in Lake Winnipeg, that we hypothesize were brought by recent floods into Lake Winnipeg, the putative inversion is nearly fixed for the alternate genotype. If the putative inversion had a neutral effect on walleye biology, an intermediate frequency would be expected among admixed fish given the parental populations almost fixed for opposite genotypes. Instead, among possibly Lake Winnipeg-Lake Manitoba

admixed fish found in Lake Winnipeg (the majority of which were found in the northern Dauphin River), the putative inversion is nearly fixed for the Lake Winnipeg genotype. Therefore, this putative inversion may play a role in walleye adaptation to Lake Winnipeg.

Three genes were within the putative inversion: *pyruvate dehydrogenase protein X component* (*PDHX*), *ETS homologous factor* (*EHF*), and *leucine-rich repeat-containing protein 4C* (*LRRC4C*). All three were expressed in Lake Winnipeg walleye, indicating at least some gene transcription activity related to the putative inversion. However, the specific functional significance of these genes is uncertain. 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; Sheppard et al., 2018). Additionally, there are patterns of low gonadal investment despite high lipid concentrations in Lake Winnipeg walleye relative to walleye in other lakes, indicating a possibly unusual phenotype in walleye of Lake Winnipeg relative to others (Moles et al., 2008). Walleye habitats are different between Lake Manitoba and Lake Winnipeg, as well. Lake Winnipeg is one of the largest lakes in the world based on surface area, but is more shallow than many other freshwater lakes (Brunskill et al., 1980; ECCC & MARD, 2020). However, Lake Manitoba is shallower still, with a 4.9 m mean depth (compared with a mean depth for Lake Winnipeg of 13.3 m and 9 m in the north and south basins, respectively) (ECCC & MARD, 2020; Rawson, 1952). Drainage from Lake Manitoba to Lake Winnipeg's north basin increases salinity there, but not in Lake Winnipeg's south basin (Brunskill et al., 1980; ECCC & MARD, 2020; ECCC, 2011). Possible local adaptation in Lake Winnipeg may be an evolutionary response to environmental differences in the waterbody.

Conclusion

The identification of a putative inversion contributes to a broad literature on chromosomal inversions in diverse taxa (see Wellenreuther & Bernatchez, 2018 for a review), although information on inversions is relatively scarce for freshwater fish (Shi et al., 2021). In addition, many observed inversions have been >1 Mb in length, possibly because reduced representation sequencing methods may be biased for detecting large inversions (Wellenreuther & Bernatchez, 2018). The present data show a small putative inversion in an obligate freshwater fish discovered *via* reduced representation sequencing (i.e., Rapture). That this putative inversion may have played a role in recent divergence between fish of two connected lakes, and that its frequency among likely admixed fish is consistent with an adaptive role in Lake Winnipeg walleye, indicates a potential importance for chromosomal inversions in freshwater fish more generally. Inversions have been previously related to environmental adaptation, such as in freshwater emerald shiner with high gene flow (Wellenreuther & Bernatchez, 2018; Shi et al., 2021). Heterogeneous habitats with opportunities for gene flow are common for freshwater fishes (Griffiths, 2015; Mushet et al., 2019), and chromosomal inversions, along with other genomic architectures, may facilitate local adaptation in many organisms that live in such connected environments.

Acknowledgements

We thank Dr. Jeff Long for contributing walleye fry from Swan Creek Hatchery and for valuable discussions of Manitoba walleye ecology, diversity, and stocking. We 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. We thank Evelien de Greef for guidance on synteny analyses and the map figure, and Dr. Kristen Gruenthal for sharing her expertise in sequencing library preparation. We thank David Splausbury and his staff from the Kansas Department of Wildlife, Parks, and Tourism for their assistance in the collection of samples at Cedar Bluff Reservoir. We thank Dr. Colin Garroway for discussions about β_{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).

References

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

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

Arostegui, M. C., Quinn, T. P., Seeb, L. W., Seeb, J. E., & McKinney, G. J. (2019). Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative freshwater ecotypes in rainbow trout. *Molecular Ecology*, *28*(6), 1412–1427. https://doi.org/10.1111/mec.15037

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

Balian, E. V., Segers, H., Lévèque, C., & Martens, K. (2008). The Freshwater Animal Diversity Assessment: an overview of the results. *Hydrobiologia*, *595*(1), 627–637. https://doi.org/10.1007/s10750-007-9246-3

Balloux, F. (2001). EASYPOP (version 1.7): A computer program for population genetics simulations. *Journal of Heredity*, *92*(3), 301–302. https://doi.org/10.1093/jhered/92.3.301

Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., Jentoft, S., & Jakobsen, K. S. (2017). Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity*, *119*(6), 418–428. https://doi.org/10.1038/hdy.2017.54

Berg, Paul R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., Jakobsen, K. S., & Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports*, *6*(1), 23246. https://doi.org/10.1038/srep23246

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

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

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

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

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

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 1–16. https://doi.org/10.1186/s13742-015-0047-8

Connallon, T., & Olito, C. (2021). Natural selection and the distribution of chromosomal inversion lengths. *Molecular Ecology*, *July*, 1–15. https://doi.org/10.1111/mec.16091

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*(4), 1411–1423. https://doi.org/10.1534/genetics.110.114819

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

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

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

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

Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., Normandeau, E., Rochette, R., & Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype-temperature association in a marine species. *Molecular Ecology*, *July*, 1–18. https://doi.org/10.1101/2020.01.28.923490

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

ECCC & MARD. (2020). *State of Lake Winnipeg, 2nd Edition*.

Euclide, P. T., MacDougall, T., Robinson, J. M., Faust, M. D., Wilson, C. C., Chen, K., 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*, *14*(5), 1403–1420. https://doi.org/10.1111/eva.13209

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. (1992). Evaluation of Stocking Walleye Fry and Fingerlings and Factors Affecting Their Success in Lower Lake Oahe, South Dakota. *North American Journal of Fisheries Management*, *12*(2), 336–345. https://doi.org/10.1577/1548-8675(1992)012<0336:eoswfa>2.3.co;2

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

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*(9), 2215–2233. https://doi.org/10.1111/mec.14584

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

Fraser, D. J., Weir, L. K., Bernatchez, L., Hansen, M. M., & Taylor, E. B. (2011). Extent and scale of local adaptation in salmonid fishes: Review and meta-analysis. *Heredity*, *106*(3), 404–420. https://doi.org/10.1038/hdy.2010.167

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

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

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

Griffiths, D. (2015). Connectivity and vagility determine spatial richness gradients and diversification of freshwater fish in North America and Europe. *Biological Journal of the Linnean Society*, *116*(4), 773–786. https://doi.org/10.1111/bij.12638

Hartman, G. F. (2009). A Biological Synopsis of Walleye (Sander vitreus). *Canadian Manuscript Report of Fisheries and Aquatic Sciences*, *2888*, v–48. http://www.dfo-mpo.gc.ca/Library/337847.pdf

Hendry, A. P. (2004). Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evolutionary Ecology Research*, *6*(8), 1219–1236.

Horrall, R. M. (1981). Behavioral Stock-Isolating Mechanisms in Great Lakes Fishes with Special Reference to Homing and Site Imprinting. *Canadian Journal of Fisheries and Aquatic Sciences*, *38*(12), 1481–1496. https://doi.org/10.1139/f81-201

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, *29*(14), 2535–2549. https://doi.org/10.1111/mec.15428

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

Jennings, M. J., Claussen, J. E., & Philipp, D. P. (1996). Evidence for Heritable Preferences for Spawning Habitat between Two Walleye Populations. *Transactions of the American Fisheries Society*, *125*(6), 978–982. https://doi.org/10.1577/1548-8659(1996)125<0978:EFHPFS>2.3.CO;2

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

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 & Evolution*, *3*(3), 469–478. https://doi.org/10.1038/s41559-019-0812-7

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., De Rosa, M. C., Righino, B., Johansen, T., Otterå, H., Sonesson, A., Lien, S., & Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, *25*(10), 2130–2143. https://doi.org/10.1111/mec.13592

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

Knief, U., Hemmrich-Stanisak, G., Wittig, M., Franke, A., Griffith, S. C., Kempenaers, B., & Forstmeier, W. (2016). Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biology*, *17*(1), 1–22. https://doi.org/10.1186/s13059-016-1056-3

Koppelman, J. B., Sullivan, K. P., & Jeffries, P. J. (1992). Survival of Three Sizes of Genetically Marked Walleyes Stocked into Two Missouri Impoundments. *North American Journal of Fisheries Management*, *12*(2), 291–298. https://doi.org/10.1577/1548-8675(1992)012<0291:SOTSOG>2.3.CO;2

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

Laayouni, H., Hasson, E., Santos, M., & Fontdevila, A. (2003). The evolutionary history of Drosophila buzzatii. XXXV. Inversion polymorphism and nucleotide variability in different regions of the second chromosome. *Molecular Biology and Evolution*, *20*(6), 931–944. https://doi.org/10.1093/molbev/msg099

Leitwein, M., Garza, J. C., & Pearse, D. E. (2017). Ancestry and adaptive evolution of anadromous, resident, and adfluvial rainbow trout (Oncorhynchus mykiss) in the San Francisco bay area: application of adaptive genomic variation to conservation in a highly impacted landscape. *Evolutionary Applications*, *10*(1), 56–67. https://doi.org/10.1111/eva.12416

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

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

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

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

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

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

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

McKinney, G., McPhee, M. V., Pascal, C., Seeb, J. E., & Seeb, L. W. (2020). Network analysis of linkage disequilibrium reveals genome architecture in chum salmon. *G3: Genes, Genomes, Genetics*, *10*(4), 1553–1561. https://doi.org/10.1534/g3.119.400972

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

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

Mushet, D. M., Alexander, L. C., Bennett, M., Schofield, K., Christensen, J. R., Ali, G., Pollard, A., Fritz, K., & Lang, M. W. (2019). Differing Modes of Biotic Connectivity within Freshwater Ecosystem Mosaics. *Journal of the American Water Resources Association*, *55*(2), 307–317. https://doi.org/10.1111/1752-1688.12683

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

Pearse, D. E., Miller, M. R., Abadía-Cardoso, A., & Garza, J. C. (2014). Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1783). https://doi.org/10.1098/rspb.2014.0012

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

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

Puncher, G. N., Rowe, S., Rose, G. A., Leblanc, N. M., Parent, G. J., Wang, Y., & Pavey, S. A. (2019). Chromosomal inversions in the Atlantic cod genome: Implications for management of Canada's Northern cod stock. *Fisheries Research*, *216*(March), 29–40. https://doi.org/10.1016/j.fishres.2019.03.020

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/

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

Rawson, D. S. (1952). Mean Depth and the Fish Production of Large Lakes. *Ecology*, *33*(4), 513–521. https://doi.org/10.2307/1931525

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

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

Sabeti, P. 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

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, *68*(1), 1–15. https://doi.org/10.1111/evo.12258

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

Shi, Y., Bouska, K. L., McKinney, G. J., Dokai, W., Bartels, A., McPhee, M. V, & Larson, W.

A. (2021). Gene flow influences the genomic architecture of local adaptation in six riverine fish species. *BioRxiv*. https://doi.org/10.1101/2021.05.18.444736

Sinclair-Waters, M., Bradbury, I. R., Morris, C. J., Lien, S., Kent, M. P., & Bentzen, P. (2018). Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic Cod in the northwest Atlantic. *Molecular Ecology*, *27*(2), 339–351. https://doi.org/10.1111/mec.14442

Sodeland, M., Jorde, P. E., Lien, S., Jentoft, S., Berg, P. R., Grove, H., Kent, M. P., Arnyasi, M., Olsen, E. M., & Knutsen, H. (2016). "Islands of Divergence" in the atlantic cod genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution*, *8*(4), 1012–1022. https://doi.org/10.1093/gbe/evw057

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*(2), 1521. https://doi.org/10.12688/f1000research.7563.1

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., & Faber, J. E. (1998). Population genetic structure, phylogeography and spawning philopatry in walleye (Stizostedion vitreum) from mitochondrial DNA control region sequences. *Molecular Ecology*, *7*(12), 1757–1769. https://doi.org/10.1046/j.1365-294x.1998.00512.x

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

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

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

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

Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower Mimulus guttatus is maintained by a chromosomal inversion. *Evolution*, *69*(6), 1476–1486. https://doi.org/10.1111/evo.12663

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, Bruce 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

Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends in Ecology and Evolution*, *33*(6), 427–440. https://doi.org/10.1016/j.tree.2018.04.002

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

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

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

Data Accessibility

All scripts used for this manuscript are available at: https://github.com/BioMatt/walleyeDNA

Author Contributions

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. Jason R Treberg contributed to interpreting candidate genes. Eva C Enders and Douglas A Watkinson contributed to information about Lake Winnipeg ecology. 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. Yue Shi analyzed the putative inversion. All authors participated in editing the manuscript.

Tables and Figures

Figure 1. Map of waterbodies included in the present study. 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 20th 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.

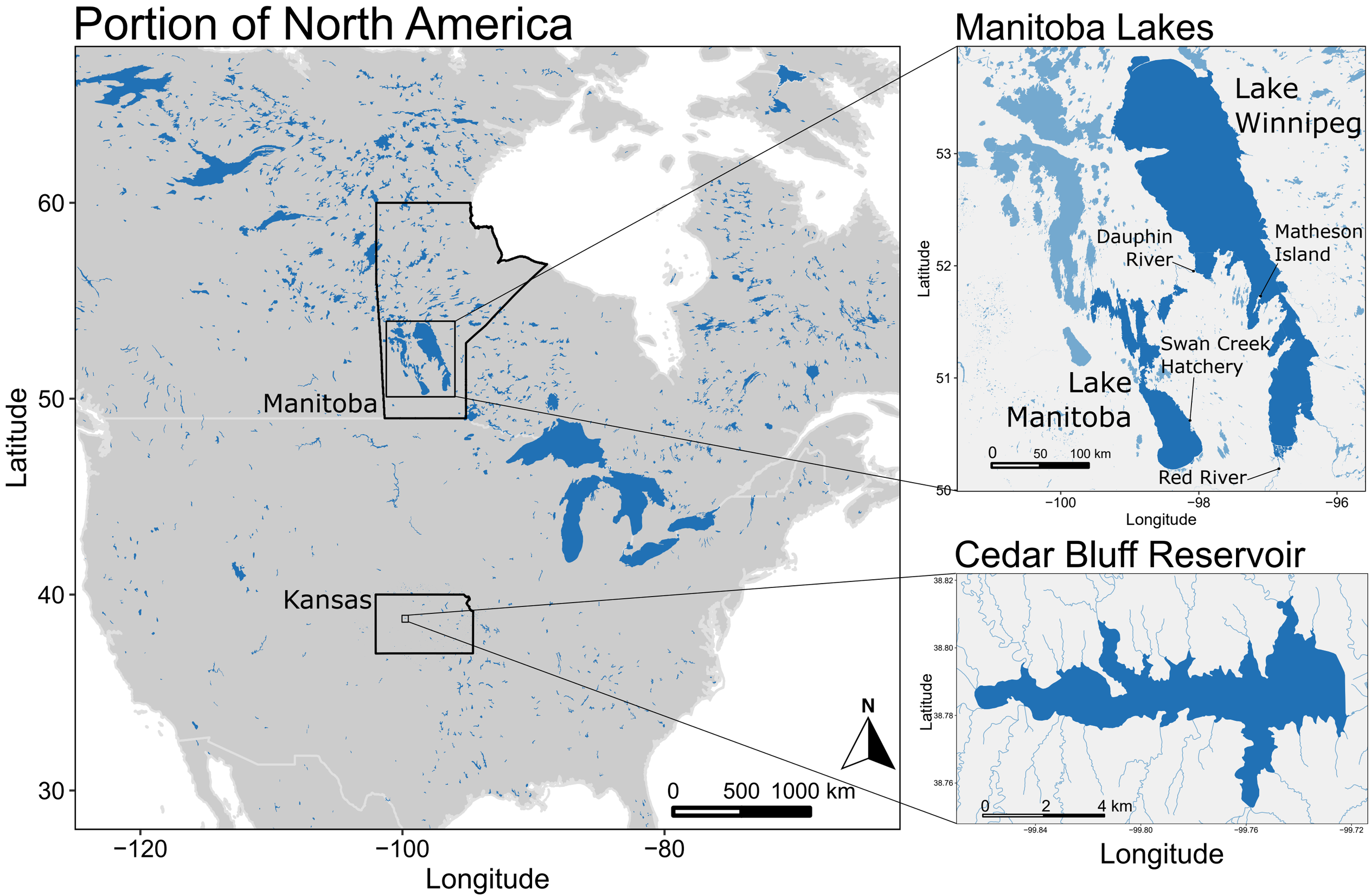
Figure 2. Population differentiation, effective population size, and observed heterozygosity for walleye (*Sander vitreus*) in three North American waterbodies. Cedar Bluff Reservoir (Kansas, USA) represents an entirely stocked population of walleye, while Lake Manitoba and Lake Winnipeg (Manitoba, Canada) walleye represent populations with possible gene flow. F_{ST} represents Weir & Cockerham's pairwise F_{ST}, while H_O 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. For F_{ST} and N_e, 95% confidence intervals are provided in parentheses.

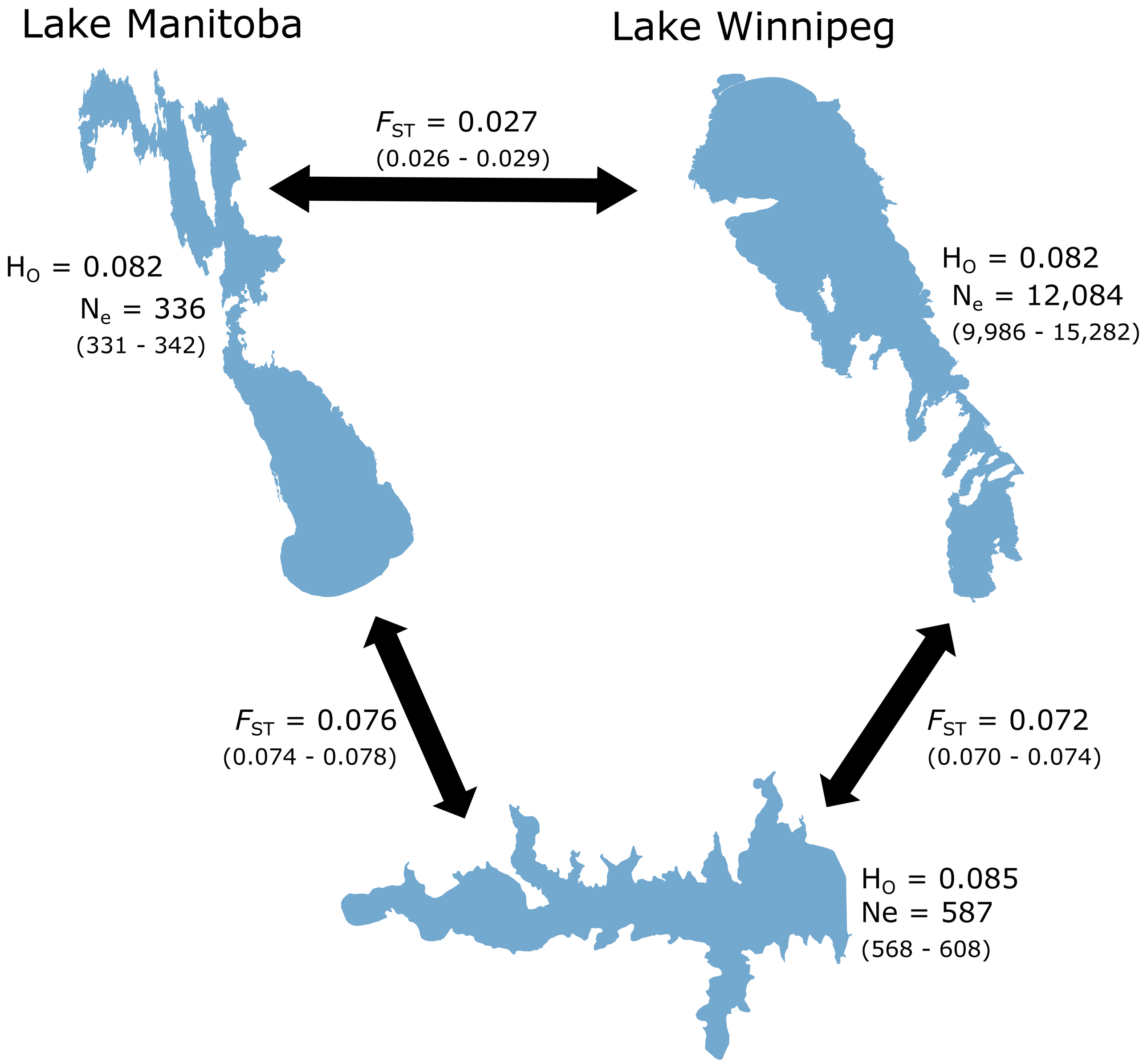
Figure 3. Signatures of selection inferred from cross-population extended haplotype homozygosity (XP-EHH) between each population of walleye (*Sander vitreus*) analyzed in the present study. Cedar Bluff Reservoir (Kansas, USA) represents an entirely stocked population, while Lake Manitoba and Lake Winnipeg (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. Absolute allele frequency differences were calculated between Lake Manitoba and Lake Winnipeg-assigned walleye. PC1 and PC2 refer to significant SNPs from principal component 1 or 2 from pcadapt (see Fig. 4). The genes *PDHX*, *EHF*, and *LRRC4C* were within the candidate region for selection in Lake Winnipeg on chromosome 8.

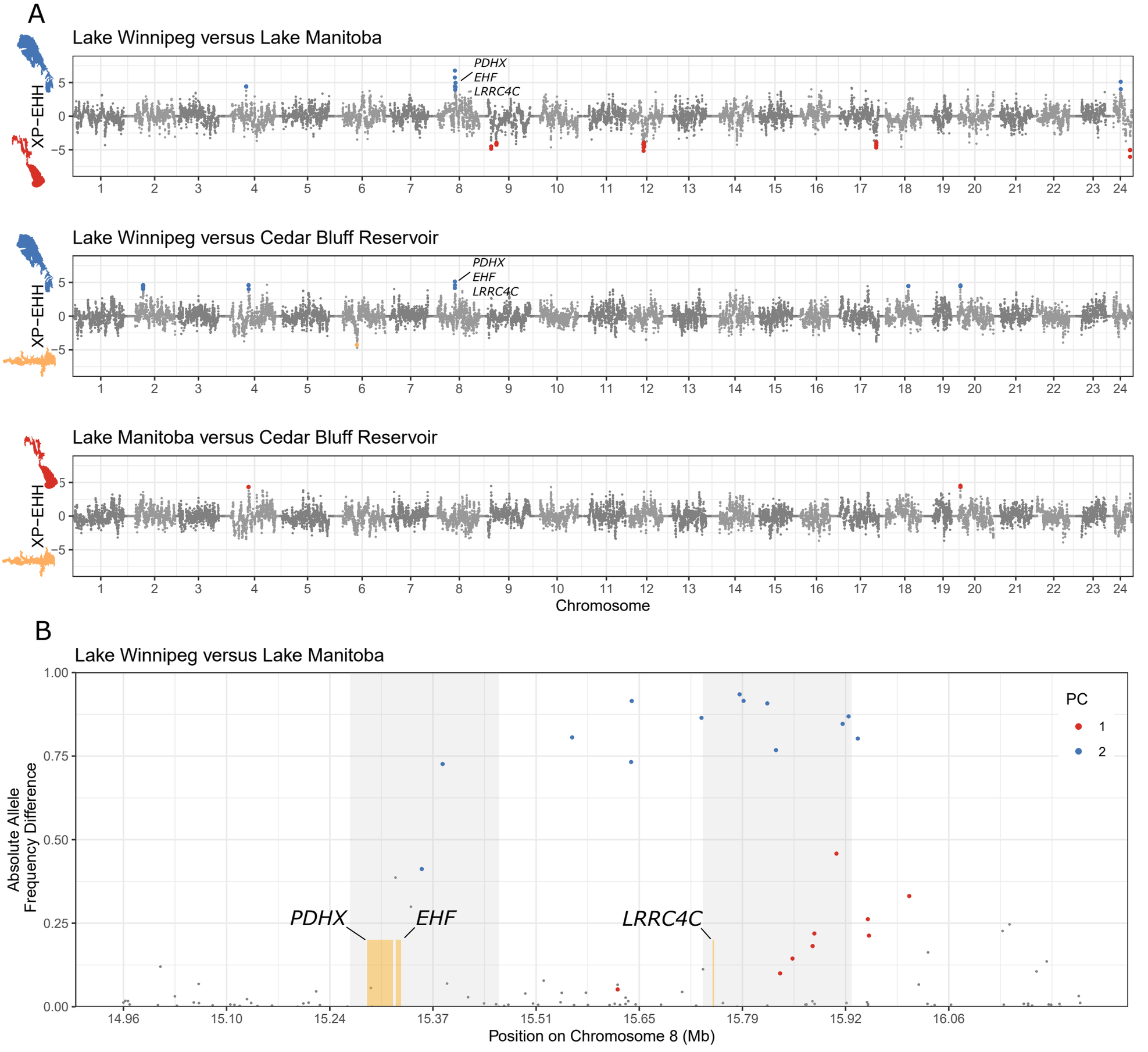
Figure 4. Principal components analysis of walleye (*Sander vitreus*) from three waterbodies across North America. Cedar Bluff Reservoir (Kansas, USA) represent an entirely stocked population of walleye of unknown origin. Lake Winnipeg and Lake Manitoba (Manitoba,

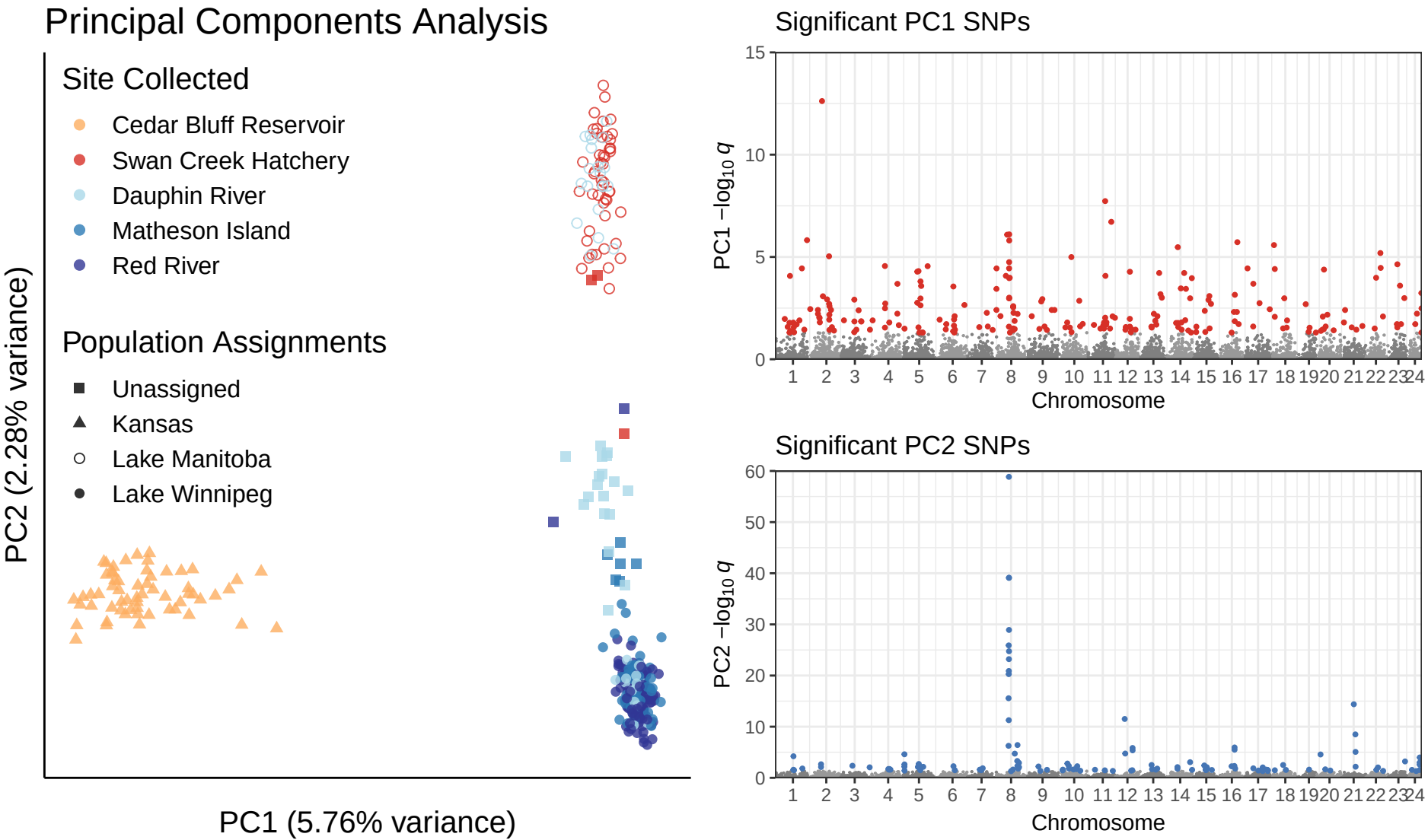
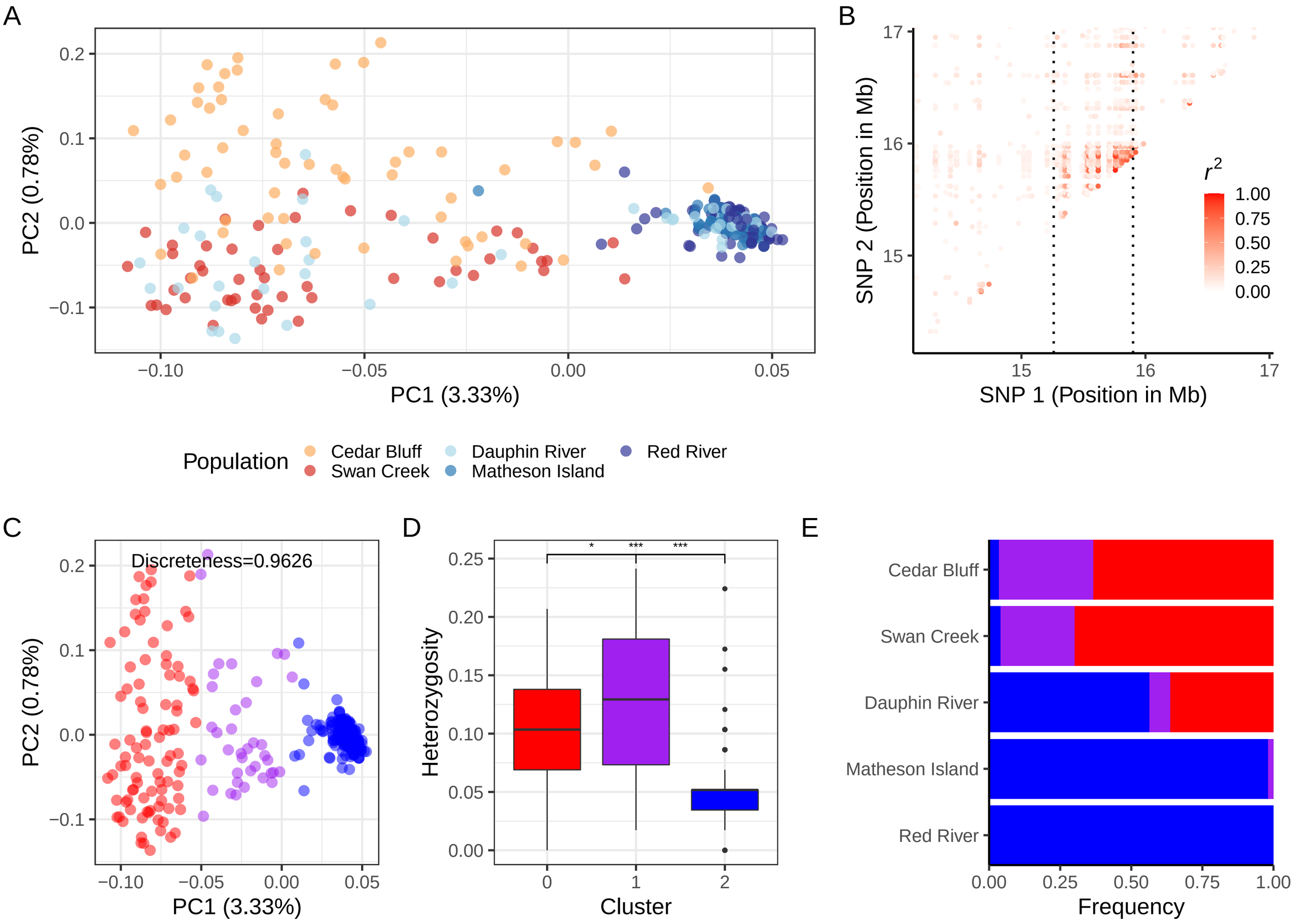
Canada) represent native populations of walleye with possible gene flow. Principal components analysis was done using pcadapt, while population assignments were done using Admixture with K=3 groups. Individuals were assigned to a population with Q>0.85 for a population, and individuals with maximum Q≤0.85 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.

Figure 5. Characterization of the putative inversion on chromosome 8 (15.26 - 15.90 Mb). A) PCA based on SNPs within the putative inversion, coloured by site. B) Linkage disequilibrium heatmap for SNPs within the putative inversion, with boundary of the putative inversion emphasized by dashed lines. C) Clusters identified using k-means clustering represent two homozygous groups, red and blue, and a heterozygous purple group. D) Observed individual heterozygosity within each cluster identified. E) Genotype frequency distribution by collection site and cluster. Bars represent the proportion of individuals from each k-means cluster.


Portion of North America
Manitoba
Kansas
60
50
40
30
Latitude
−120
−100
−80
Longitude
N
0
500
1000 km
Manitoba Lakes
Lake Winnipeg
Dauphin River
Matheson Island
Swan Creek Hatchery
Lake Manitoba
Red River
0
50
100 km
53
52
51
50
Latitude
−100
−98
−96
Longitude
Cedar Bluff Reservoir
38.82
38.80
38.78
38.76
Latitude
0
2
4 km
−99.84
−99.80
−99.76
−99.72
Longitude


Lake Manitoba
Lake Winnipeg
F_{ST} = 0.027
(0.026 - 0.029)
H_O = 0.082
N_e = 336
(331 - 342)
H_O = 0.082
N_e = 12,084
(9,986 - 15,282)
F_{ST} = 0.076
(0.074 - 0.078)
F_{ST} = 0.072
(0.070 - 0.074)
H_O = 0.085
Ne = 587
(568 - 608)
Cedar Bluff Reservoir


A
Lake Winnipeg versus Lake Manitoba
Lake Winnipeg versus Cedar Bluff Reservoir
Lake Manitoba versus Cedar Bluff Reservoir
XP-EHH
PDHX
EHF
LRRC4C
Chromosome
B
Lake Winnipeg versus Lake Manitoba
Absolute Allele Frequency Difference
Position on Chromosome 8 (Mb)
PC
1
2
PDHX
EHF
LRRC4C


Principal Components Analysis
Site Collected
Cedar Bluff Reservoir
Swan Creek Hatchery
Dauphin River
Matheson Island
Red River
Population Assignments
Unassigned
Kansas
Lake Manitoba
Lake Winnipeg
PC2 (2.28% variance)
PC1 (5.76% variance)
Significant PC1 SNPs
PC1 −log10 q
Chromosome
Significant PC2 SNPs
PC2 −log10 q
Chromosome

Chromosome 8 (Position 15260000 − 15900000)