NRC Research Press

ARTICLE

SNP analyses highlight a unique, imperiled southern walleye (*Sander vitreus*) in the Mobile River Basin

Honggang Zhao, Katherine Silliman, Matthew Lewis, Sarah Johnson, Garret Kratina, Steve J. Rider, Carol A. Stepien, Eric M. Hallerman, Benjamin Beck, Adam Fuller, and Eric Peatman

Abstract: Walleye (*Sander vitreus*) is a popular sportfish threatened by overexploitation, habitat destruction, and loss of genetic integrity due to non-native walleye stocking. Previous studies have identified a genetically distinct lineage of walleye in the Mobile River Basin, but further work is needed to assess population structure and introgression among this southern lineage and northern populations. Here we generated 2782 genome-wide single nucleotide polymorphisms (SNPs) to characterize the genetic uniqueness of southern walleye. We also found strong evidence for a historical declining population trend with reduced genetic diversity and effective population size in a southern walleye population of conservation importance. A 68-SNP panel was developed for rapid identification of genetic integrity and hybrid classification among northern and southern walleye, enabling us to identify an anthropogenic hybrid zone resulting from the previous introduction of northern walleye into the Black Warrior River drainage, Alabama. Our results highlight the need for conservation management of southern walleye in the Mobile River Basin, with our 68-SNP assay already being implemented in ongoing stream survey and captive breeding programs.

Résumé : Le doré jaune (*Sander vitreus*), un poisson prisé des pêcheurs sportifs, est menacé par la surexploitation, la destruction d'habitats et la perte d'intégrité génétique causée par l'ensemencement de dorés non indigènes. Des études antérieures ont cerné la présence d'une lignée de dorés distincte sur le plan génétique dans le bassin du fleuve Mobile, mais d'autres travaux sont nécessaires pour évaluer la structure des populations et l'introgression entre cette lignée méridionale et les populations septentrionales. Nous avons généré 2782 polymorphismes mononucléotidiques (SNPs) pangénomiques dans le but de caractériser l'unicité génétique des dorés du sud. Nous avons aussi relevé des preuves solides d'une tendance démographique passée à la baisse accompagnée de réductions de la diversité génétique et de la taille effective de la population dans une population de dorés méridionaux d'importance pour la conservation. Un panneau de 68 SNP a été mis au point pour la détermination rapide de l'intégrité génétique et la classification d'hybrides au sein des dorés du nord et du sud, nous permettant de cerner une région d'hybrides d'origine anthropique résultant de l'introduction passée de dorés septentrionaux dans le bassin versant de la rivière Black Warrior (Alabama). Nos résultats soulignent la nécessité de la gestion de la conservation des dorés méridionaux dans le bassin du fleuve Mobile, notre épreuve de 68 SNP étant déjà déployée dans des programmes en cours de relevés dans les cours d'eau et d'élevage en captivité. [Traduit par la Rédaction]

Introduction

Freshwater ecosystems are severely threatened by anthropogenic activities such as overexploitation, water pollution, destruction of habitats, and species invasion (Dudgeon et al. 2006). For exploited freshwater fishes of economic or recreational importance, restocking is commonly employed to increase population size and mitigate the risks of genetic collapse (Cochran-Biederman et al. 2015). Effective restocking practices should account for the existing population structure and genetic diversity of the species (Miller and Kapuscinski 2003); however, this is not always the case (Seddon et al. 2007). For example, restocking activities employing nonindigenous hatchery brown trout (*Salmo trutta*) specimens into the wild populations in Iberian Peninsula has resulted in anthropogenic hybridization and a decrease of genetic variability (Madeira et al. 2005). Population genetics is a powerful tool in conservation biology, as it can provide information to reduce risks from loss of genetic diversity, inbreeding depression, and introgression with non-native individuals. Recent developments in genomic approaches, such as genotyping-by-sequencing (GBS), have facilitated the cost-effective detection of genome-wide single nucleotide polymorphisms (SNPs) (Andrews et al. 2016). Owing to improved resolution for describing hybridization–introgression, adaptive genetic variation, and fine-scale demographic structure, SNPs are rapidly becoming the preferred markers for routine genetic identification and hybridization tests in a variety of aquatic species (Lamer et al. 2015; Pritchard et al. 2016; Thongda et al. 2020; Zhao et al. 2019).

Walleye (*Sander vitreus*) is an ecologically important and economically valuable freshwater fish species in family Percidae (Barton 2011). Walleye inhabit a wide range of habitat conditions across North America, from the Mackenzie River in the Arctic to

the US Gulf Coast (Barton 2011). Its east–west distribution is delineated by the eastern continental divide and the Rocky Mountains (Regier et al. 1969). The distribution of walleye in southern latitudes is limited by the cold temperatures required for gonadal maturation (Hokanson 1977). Previous genetic surveys of walleye provided evidence for the existence of five genetically distinct lineages in North America, composed of three northern stocks across the Northwest Lake Plains, Great Lakes watershed, and North Atlantic coastal drainages (Billington and Hebert 1988; Haponski and Stepien 2014; Stepien and Faber 1998; Strange and Stepien 2007; Ward et al. 1989), and two unique lineages identified in the Mobile River Basin (Billington and Strange 1995) and the eastern highland regions (Billington 1996; Stepien et al. 2009). A number of population genetic studies of walleye indicate that northern lineages were shaped by post-Pleistocene recolonization events (Billington and Hebert 1988; Billington et al. 1992; Ward et al. 1989), while populations from unglaciated eastern highlands regions evolved due to isolation in river drainages (Stepien et al. 2009; White et al. 2012). Specifically, the southern walleye in Alabama and Mississippi (Tombigbee River) were identified as a long-isolated historic assemblage of populations that diverged from northern walleye ~1.17 (±0.31) million years ago (Billington et al. 1992; Stepien et al. 2009). Additional genetic investigations using allozymes (Billington and Maceina 1997; Murphy 1990), mtDNA (Billington and Strange 1995; Billington et al. 1992, 1997), and microsatellites (Haponski and Stepien 2014; Stepien et al. 2009) confirmed their unique genetic pattern and suggested that a careful monitoring and management plan should be implemented for these southern populations due to potential introgression with introduced northern walleye (Billington and Strange 1995). Although the genetic divergence between Mobile River Basin walleye and other more northern walleye populations has been initially characterized, knowledge of genetic diversity, population structure, and demographic history in southern populations is minimal. This is due, in large part, to the paucity of modern, verified genetic markers for the southern walleye.

Introgression has already been documented between introduced northern Great Lake stocks and walleye populations in the Ohio River drainage using mitochondrial DNA (Billington et al. 1997; Palmer et al. 2006; White et al. 2012). A previous study conducted by Billington and Maceina (1997) indicated that the integrity of southern walleye mtDNA haplotypes has been well preserved at Hatchet Creek (Coosa River drainage, Alabama). However, this well-preserved southern lineage may be vulnerable to unauthorized introductions of northern walleye by the public (McMahon and Bennett 1996). Ongoing walleye conservation efforts through restocking and captive breeding would benefit considerably from results of a comprehensive survey of existing genetic diversity and introgression among source populations, as well as the development of markers for rapidly characterizing genetic background.

The primary goals of this study were to infer genetic diversity and population structure in southern walleye populations using thousands of GBS-derived SNPs and then develop an SNP assay that could rapidly and accurately identify genetic integrity and hybridization among non-native and southern walleye. To validate the SNP panel for rapid and accurate identification of various hybrid classes in walleye individuals, we focused on walleye from the Black Warrior River system in Alabama, as this system has records of non-native walleye stocking from 1975 to 1985 (Billington and Maceina 1997). Our secondary goal was to utilize the SNP panel for characterizing population structure and introgression in 23 walleye populations. Lastly, we compared historical demographic models using the GBS data to investigate

whether a known pure southern walleye population shows genetic signatures of population decline.

Materials and methods

Sample collection and genotyping-by-sequencing (GBS)

A total of 60 samples representing pure northern (Lake Erie (ERI), $n = 30$), pure southern (Hatchet Creek (HAT), $n = 10$) and hybrid (Blackwater Creek (BLA), $n = 20$) walleye populations (Fig. 1) were collected for GBS library construction and sequencing. We also sampled an additional 545 walleye individuals (fin clips or DNA samples) across 23 northern and southern populations for extensive walleye population genetic structure analysis and hybrid classification (Fig. 1; also refer to online Supplementary material, Table S1[1]). These additional individuals were collected to represent native walleye distribution at lacustrine and river sites, including the Great Lakes watershed (Lakes Erie, Michigan, Superior), Northwest Lake Plains (Mille Lacs Lake at the upper Mississippi River), Mobile River Basin (Coosa and Tombigbee River), and the eastern highlands regions (New River, Rockcastle River, and Big Sandy River). Sites with records of historical restocking (Black Warrior River drainage and sites in the Tennessee River system) were also included in this study for further hybridization and introgression analyses. Fin clips of individuals were collected and stored in 95% ethanol for subsequent DNA extraction.

Following specimen collection, we extracted DNA for downstream GBS library construction and sequencing. Genomic DNA from all samples was extracted from fin clips using the DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's protocol. DNA quality was assessed by running 100 ng of each DNA sample on 1% agarose gels. DNA concentration was determined using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). DNA samples were sent to the University of Minnesota Genomics Center for double-digest GBS library construction and sequencing. Briefly, 100 ng of DNA was digested with 10 units of a combination of *Bam*HI and *Nsi*I enzymes (New England Biolabs; NEB) and incubated at 37 °C for 2 h. Following digestion, samples were then ligated with 200 units of T4 ligase (NEB) and phased adaptors at 22 °C for 2 h to inactivate the T4 ligase. The ligated samples were then amplified for 18 cycles with 2× NEB *Taq* Master Mix along with sample-specific barcodes. Libraries of walleye samples were purified, quantified, pooled, and size-selected for 300- to 744-bp fragments and diluted to 1.7 pmol for sequencing. The pooled libraries were loaded across four lanes of 150-bp single-read sequencing on an Illumina NextSeq 550.

Genome assembly and SNP marker discovery

To perform reference-based SNP calling, we assembled a rough draft genome for walleye. One DNA sample from Blackwater Creek walleye was selected for whole-genome sequencing and sent to the University of Minnesota Genomics Center for library construction and sequencing. During library creation, 100 ng of DNA was fragmented to target a 350-bp insert length using Covaris ultrasonic shearing. The sheared DNA was then end-repaired and subjected to a bead-based size selection. After adaptor and index ligation, the library was amplified using eight cycles of PCR. The amplified library was sequenced across 1.5 lanes of a HiSeq 2500 125-bp paired-end run. A total of 302 million Illumina reads were generated from library sequencing and assembled into a 783 Mb (N50 = 4.13 kb) draft genome using MaSuRCA version 3.2.4 (Zimin et al. 2013). We followed the default parameter settings for genome assembly, except for library insert length (342) and standard deviation (76). These two parameters were estimated using Burrows–Wheeler Aligner version 0.7.17 (Li and Durbin 2009).

1368

Can. J. Fish. Aquat. Sci. Vol. 77, 2020

Fig. 1. Sampling locations of walleye. Populations are labeled as in Table S1[1] and are represented by a pie graph showing the estimated southern (yellow), northern (blue), and eastern highlands (cyan) walleye genomic composition for each location based on STRUCTURE results at *K* = 3. The map only shows walleye native and introduced ranges in the United States. HAT = Hatchet Creek; WHI = White Plains; TOM = Tombigbee River; MUL = Mulberry Fork; BLA = Blackwater Creek; NOR = North River; NAN = Nantahala Lake; FON = Lake Fontana; NOD = Normandy Lake; CHI = Chickamauga Lake; WAT = Watts Bar Lake; DOU = Douglas Lake; CHE = Cherokee Lake; NOS = Norris Lake; PAT = Fort Patrick Henry Lake; MIL = Mille Lacs Lake; FOS = Fosters Falls; ROC = Rockcastle River; LEV = Levisa Fork; HUR = Huron River; ERI = Lake Erie; MUS = Muskegon River; THU = Thunder Bay. Map data sources: Esri (https://esri.com/); walleye native and introduced ranges: USGS (https://nas.er.usgs.gov/). The map was drawn by ArcGIS. [Colour online.]
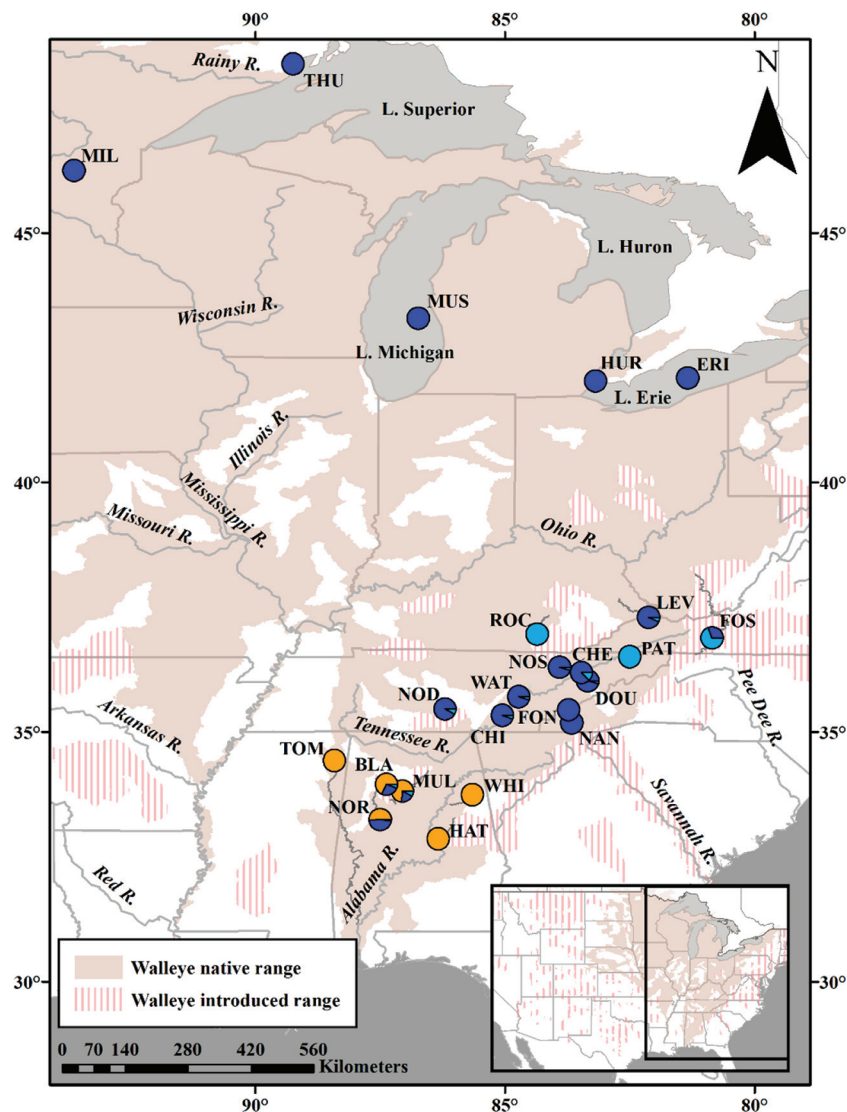
Genome-wide SNPs were called using STACKS version 2.4 (Catchen et al. 2013), with minor changes in parameter settings (described below). GBS reads were cleaned and de-multiplexed using the process_radtags program in STACKS. The process_radtags parameters were set to remove reads with an uncalled base (-c) and low-quality scores (-q). For reference-based SNP calling, we first mapped the de-multiplexed reads to the assembled walleye genome using Burrows–Wheeler Aligner version 0.7.17 with default settings. The mapped reads were then sorted using the "sort" function in SAMTOOLS version 1.6 (Li et al. 2009). The mapped and sorted reads were used to call SNPs with the "gstacks" and "populations" pipelines in STACKS. We used gstacks (--var-alpha 0.05 --gt-alpha 0.05) to assemble the loci using consensus sequence and then call SNPs within each locus. The "population" program was used to merge loci that were produced from the same restriction enzyme cut sites (merge sites) and ex-

port the output as a variant call format (VCF) file (Danecek et al. 2011). We generated two SNP datasets in STACKS: one including all GBS samples (*n* = 60) and the other with Hatchet Creek samples only (*n* = 10). For population genetic analyses, we used the SNP dataset containing all samples, while the demography analysis was performed with SNPs from Hatchet Creek only. For the dataset including all GBS samples, we initially used VCFtools (Danecek et al. 2011) to filter loci with minimum minor allele frequency set to 0.05, minimum minor allele count set to 10, and minimum locus coverage set to 0.1. To ensure that SNPs were informative and reliable for downstream population genetic analyses and marker validation, we used VCFtools and SNPRelate (Zheng et al. 2012) for stringent filtration of SNPs based on the following criteria: (*i*) only SNPs called in 100% of individuals; (*ii*) SNPs with observed heterozygosity larger than 0.6 were removed to avoid paralogous loci in the dataset (Li et al. 2014); (*iii*) SNPs deviating

from Hardy–Weinberg equilibrium (HWE, p value < 0.01) in more than one population were removed; (*iv*) SNP pairs that showed linkage disequilibrium with $r^2 > 0.2$ were pruned (individual SNPs with higher genotype coverage were kept). For the demography dataset, we applied linkage disequilibrium filtering at $r^2 = 0.2$ and kept only SNPs with no missing data.

Diagnostic marker development and validation

To develop SNP assays for rapid and accurate identification of walleye lineages and various hybrid classes, we identified diagnostic SNPs with fixed-allelic differences between representative northern (Lake Erie) and southern (Hatchet Creek) populations (e.g., homozygous "A" in pure southern individuals, homozygous "T" in pure northern individuals, and polymorphic in Blackwater Creek fish). We used GenAlex version 6.5 (Peakall and Smouse 2012) to identify putative neutral loci based on the distribution of SNP F_{ST} values (after exclusion of fixed markers). We ordered all SNPs based on locus-specific F_{ST} (from lowest to highest, across all population pairs) and created three data subsets that fell in different quartiles of the F_{ST} distribution: low global F_{ST} SNPs (below 25th percentile of the F_{ST} distribution corresponding to F_{ST} = 0.001–0.112), intermediate global F_{ST} dataset (between 25th and 75th percentiles of F_{ST} distribution, F_{ST} = 0.112–0.378), and high global F_{ST} (75%–100% percentiles, F_{ST} = 0.378–0.822). An additional outlier scan was performed using the same dataset to identify SNPs showing evidence of divergent or balancing selection. The outlier test was conducted using BAYESCAN version 2.1 with default iteration and burn-in settings, and prior odds set to either 1 or 10 (Foll and Gaggiotti 2008). Here the prior odds of 10 corresponds to a prior belief that the neutral model is 10 times more likely than the model of selection, while 1 represents the equal prior probability for both models. SNPs with a false discovery rate < 10% were considered as putatively under selection.

A MassARRAY System (Agena Bioscience, San Diego, California) was used to validate a subset of diagnostic SNPs identified and genotyped by GBS and to genotype an additional 545 walleye individuals from 23 populations. Using MassARRAY ASSAY DESIGN software and following the protocol described in Zhao et al. (2018), we designed two multiplex assays with 40 SNPs per well (Fig. S1[1]). Amplification and extension reactions were performed using 10 ng of DNA per sample and the iPLEX Gold Reagent Kit (Agena Bioscience) according to the manufacturer's protocol. SNP genotypes were called using the MassARRAY Typer 4 analysis software. This software uses a three-parameter (mass, peak height, and signal-to-noise ratio) model to estimate genotype probabilities. Considering that all types of molecular markers are prone to genotyping errors (Pompanon et al. 2005), we used 59 individuals to test the concordance of SNP genotypes generated from MassARRAY and GBS. A total of 114 individuals (including 58 samples from Black Warrior River drainage) were used as technical replicates (genotyped twice by the MassARRAY system) to test the consistency of genotype calling in the MassARRAY system. Discordant genotypes due to missing data were excluded from this analysis (zero missing data from MassARRAY and GBS comparison; 502 genotypes were not compared in the replicate test due to missing data).

Population genetic analyses

The Bayesian clustering algorithm-based program STRUCTURE version 2.3.4 (Pritchard et al. 2000) was used to characterize population structure for both the GBS and MassARRAY datasets. The admixture model with correlated allele frequencies was applied with a burn-in of 20 000 iterations followed by 200 000 Markov chain Monte Carlo (MCMC) repetitions. We used different numbers of assumed population genetic clusters (K = 1–9 for GBS data, 1–25 for MassARRAY data) to determine the best-supported K values using program KFinder (Wang 2019), repeated 10 times for each K. Three different criterions in KFinder: Pr[$X|K$] (Pritchard et al.

2000), ΔK (Evanno et al. 2005), and parsimony index (Wang 2019) were estimated for both GBS and MassARRAY datasets. Population differentiation was estimated for all pairs of populations in the GBS and MassARRAY datasets using Hudson's estimator of F_{ST} (Hudson et al. 1992) implemented in EIGENSOFT version 7.2.1 (Patterson et al. 2006). Hudson's F_{ST} statistic is not sensitive to uneven population sizes and does not systematically overestimate F_{ST} (Bhatia et al. 2013; Hudson et al. 1992).

Population structure in the MassARRAY dataset was also visualized with a discriminant analysis of principal components (DAPC) implemented in the R package Adegenet (Jombart and Ahmed 2011). The optimal number of principal components was determined by an alpha-score procedure with 20 repeated runs. For only the GBS dataset, population diversity indices for each population were evaluated by computing the observed heterozygosity (H_o), expected heterozygosity (H_e), and inbreeding coefficient (F_{is}) using Arlequin version 3.5 (Excoffier and Lischer 2010). Effective population size (N_e) for each population was estimated using the linkage disequilibrium method implemented in NeEstimator version 2.1 (Do et al. 2014).

Hybridization analyses

One of our goals in this study was to investigate introgression and hybrid status in the Black Warrior River drainage and evaluate the performance of our SNP panel for hybrid classification. For this purpose, we developed a novel framework that combined STRUCTURE, NEWHYBRIDS (Anderson and Thompson 2002), and manual assignment methods for hybrid classification (Fig. S2[1]), with the assumption that no genotyping errors or contamination had occurred in genotyping data (Lamer et al. 2015). Briefly, before the hybrid assignment, we used STRUCTURE to examine the individual introgression status based on membership coefficients (Q values). We assigned individuals with Q values ≥ 0.95 to one of the three major genetic lineages: northern (Great Lakes – upper Mississippi), southern (Mobile River Basin), and eastern highlands; otherwise they were assigned as hybrids depending on lineage proportion (Thongda et al. 2020). Lineage proportion < 0.05 was not considered as an introduced aggressor in cases of hybridization (Lutz-Carrillo et al. 2006; Thongda et al. 2020). Given that we only obtained a small number of SNPs (8 out of 68) that were informative for northern and eastern highlands hybrid classification, we did not monitor hybridization for these individuals. A custom R script (available at https://github.com/hzz0024/walleye) was then used to score three genotype ratios (homozygous AA and BB and heterozygous AB) for each examined fish. Individuals containing a mixture of homozygous loci for each parental group and heterozygous loci were manually assigned to later-generation backcross × backcross hybrids (e.g., F_xS with a majority southern walleye homozygous genotypes or F_xN with a majority northern walleye homozygous genotypes; Table S2[1]), except for those consistent with F_2 proportions estimated from other species (Lamer et al. 2015). The Bayesian framework-based program, NEWHYBRIDS version 1.1 beta, was then used to compute the posterior distribution of individual assignment into 12 different hybrid categories. The hybrid test was conducted using 300 fixed SNPs from the GBS dataset and 68 SNPs in MassARRAY assays, respectively. We trimmed our GBS datasets to 300 unlinked diagnostic SNPs because analyses failed to run with more markers due to an underflow issue (Elliott and Russello 2018). Three independent analyses were conducted with different random subsets of 300 for assignment evaluation. We used HYBRIDLAB version 1.0 (Nielsen et al. 2006) to simulate hybrids from northern and southern baseline populations to evaluate the power of our diagnostic SNPs in hybrid class discrimination. The program generated 15 000 random genotypes for each of the 12 hybrid classes (Table S2[1]): parental northern walleye (ERI), parental southern walleye (HAT), F_1 hybrids, F_2 hybrids ($F_1 \times F_1$), first-generation backcrosses (F_1 hybrids × either parental baseline), second-generation backcrosses (first-

Fig. 2. Demographic models utilized to test declines in the southern walleye population using FASTSIMCOAL. Models from left to right: M1, a consistent model assuming no population size change over time; M2, a continuous decline model; M3, an instantaneous bottleneck model; and M4, a scenario of bottleneck followed by continuous decline. NCUR, the current population size; NANC, the ancestor population size or population size before bottleneck; TBOT, the time of instantaneous population size change. The average estimated effective population sizes are shown at various stages in the best model (M2, in parentheses). Parametric bootstrap estimates are shown in Table S7[1].
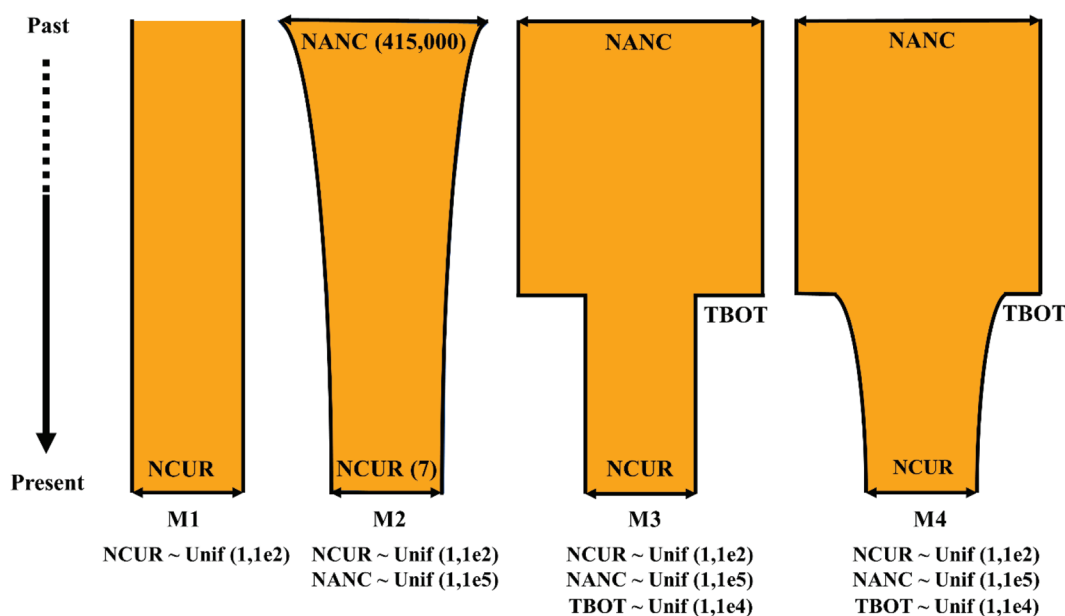
generation backcross × either parental baseline), third-generation backcrosses (second-generation backcross × either parental baseline), and fourth-generation backcrosses (third-generation backcross × either parental baseline). We did not include simulations of hybrids between backcross × backcross in our hybrid tests because of low assignment confidence among these classes. We ran NEWHYBRIDS analyses using an initial 100 000 MCMC burn-in followed by 100 000 MCMC sweeps, with the 12 hybrid categories set as "Jeffreys-like priors". An individual was considered to be robustly classified if its assignment probability was >0.5.

Historical demography analyses of Hatchet Creek

Given the low genetic diversity and effective population size characteristic of southern walleye populations, we conducted model-based demographic analyses on Hatchet Creek samples using the composite likelihood approach implemented in FASTSIMCOAL version 2.6.0.3 (Excoffier et al. 2013). Our goal was to evaluate whether evidence exists for declines in population size and to gain a better understanding of the underlying temporal dynamic in this southern walleye population. The folded site frequency spectrum was estimated from the STACKS dataset with only Hatchet Creek samples and used to compare four demographic models: constant population size, continuous population decline, instantaneous bottleneck, and a scenario of bottleneck followed by continuous decline (Fig. 2). These models were similar to those used in Chattopadhyay et al. (2019), with minor modification to the prior parameters (Fig. 2). To stabilize the estimated values, we conducted 50 independent runs for each model, and each run performed 100 000 simulations and 40 optimization cycles using a conditional maximization algorithm. The maximum-likelihood runs of each model were then compared and the Akaike information criterion (AIC), ΔAIC, and the Akaike's weight were estimated to determine the best-fit demographical model (Excoffier et al. 2013). Because there is no empirical estimate of genome-wide mutation rate for walleye, we applied different values of mutation rate (μ) for demographic model analyses. Three genome-wide mutation rates, human (2.5×10^{-8} per site per generation; Nachman and Crowell 2000), cichlid (3.5×10^{-9} per site per generation;

Malinsky et al. 2018), and Atlantic herring (*Clupea harengus*) (2×10^{-9} per site per generation; Feng et al. 2017), were used to represent "high", "intermediate", and "low" mutation levels, respectively, for model likelihood estimation and comparison. Ultimately, we chose the Atlantic herring mutation rate, as it resulted in more similar estimated and observed likelihood values.

After model determination, we performed an additional 50 FASTSIMCOAL runs with a fixed ancestral N_e to obtain confidence limits for parameter estimates, choosing the parameter estimates from the run with an estimated maximum likelihood closest to the observed likelihood. We estimated the ancestral effective population size of Hatchet Creek walleye based on the equation of $N_e = \theta\pi/2\mu$ (for haploid populations; Chattopadhyay et al. 2019), with genome-wide nucleotide diversity (θ_π) calculated from STACKS and an assumed mutation rate of 2×10^{-9} per site per generation. Using the parameter estimates associated with the best maximum-likelihood run, we performed 100 bootstrap replicates to estimate the confidence limits in parameter estimation.

Results

GBS sequencing and SNP discovery

A total of 69.38 million high-quality reads were generated from Illumina NextSeq sequencing, with a mean of 1.16 million reads for each sequenced sample. During SNP discovery, a total of 16 158 SNPs was identified using STACKS after filtering for minor allele frequency > 0.05, minimum minor allele count > 10, and minimum locus coverage > 0.1 (Fig. S1[1]). Additional stringent filtering steps using VCFtools and SNPRelate packages resulted in a final dataset of 2782 SNPs (Fig. S1[1]). With Hatchet Creek samples only, we obtained a total of 2106 SNPs (total concatenated sequence length = 317 028 bp) from STACKS and used this dataset for demography analysis. To develop SNP assays for rapid and accurate identification of walleye lineages and various hybrid classes, we identified a dataset of 940 diagnostic SNPs showing fixed genetic differences between northern and southern walleye populations. BAYESCAN failed to identify any SNPs showing evidence of balancing or divergent selection, regardless of the settings of

Fig. 3. SNP F_{ST} distribution and population structure results inferred from three GBS data subsets. (*a*) Distribution of locus-specific F_{ST} for GBS SNPs after removal of SNPs fixed between northern and southern walleye. Colour bars below the panel indicate the range of F_{ST} for high (461 SNPs, purple), intermediate (921 SNPs, blue) and low (460 SNPs, orange) SNPs. (*b*) STRUCTURE results using $K = 2$ (left) and parsimony index estimation of most likely K (right). Three datasets (full, diagnostic, and intermediate) were used for STRUCTURE analyses. HAT = Hatchet Creek; BLA = Blackwater Creek; ERI = Lake Erie. [Colour online.]

(a) Frequency distribution of F_{ST} estimates for 1,842 SNPs

(b) STRUCTURE results using three GBS data subsets (K=2)
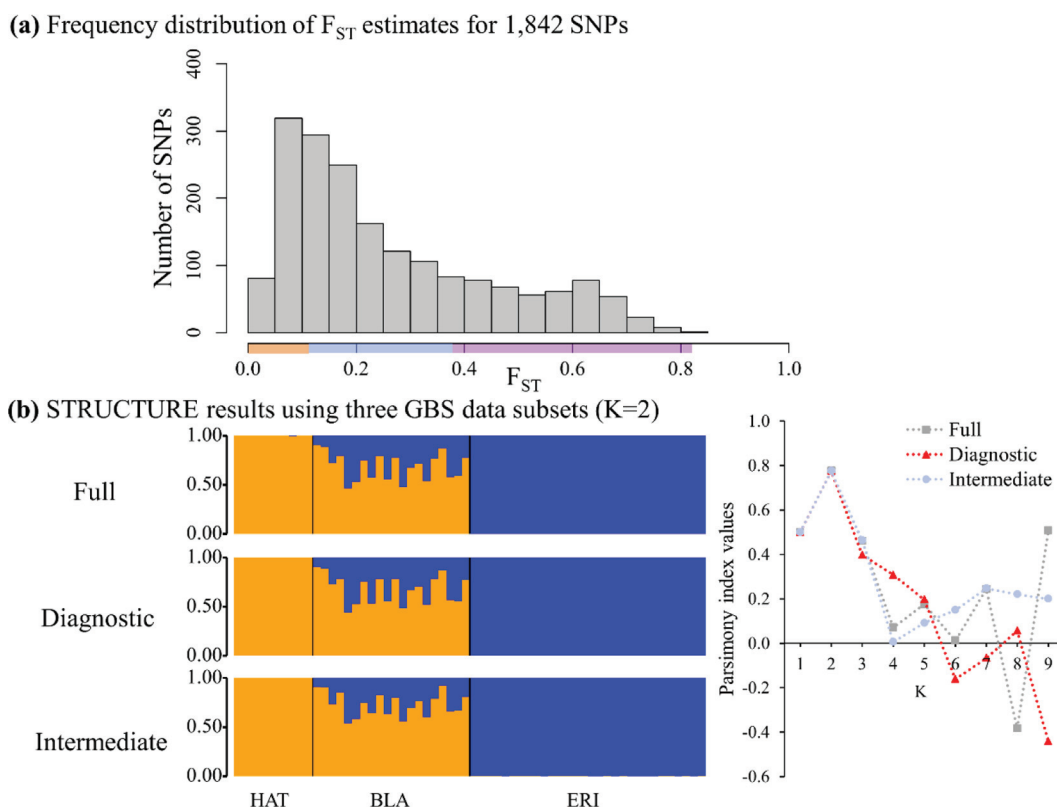
prior odds. The SNP dataset excluding diagnostic loci (1842 SNPs) was categorized into a high-F_{ST} SNP subset (461, $F_{ST} = 0.378$–0.822), intermediate-F_{ST} subset (921 SNPs, $F_{ST} = 0.112$–0.378), and low-F_{ST} subset (460 SNPs, $F_{ST} = 0.001$–0.112; see Fig. 3 for F_{ST} distribution). From here on, we use the terms low, intermediate, and high SNPs to represent these three subsets.

Genetic diversity and population structure using GBS data

We measured the genetic diversity of these three populations through the percentage of polymorphic loci (P_o), observed (H_o) heterozygosity, and expected heterozygosity (H_e) using the full GBS dataset with 2782 SNPs. Blackwater Creek had the largest number of variant SNPs (2652 of 2782, 95.33%), with Hatchet Creek walleye representing the lowest level of marker polymorphism (444 of 2782, 15.96%; Table 1). Similarly, the lowest genetic diversity was found in the Hatchet Creek population, with a mean observed heterozygosity of 0.06, suggesting a limited number of founders and (or) the presence of drift in this population. We found little evidence of inbreeding in the examined populations, as inbreeding coefficients ranged from –0.05 to 0.002 (Table 1).

Pairwise F_{ST} among walleye populations was measured using both the full GBS dataset (2782 SNPs) and SNP subsets based on locus-specific F_{ST}. Using the full GBS dataset, the obtained F_{ST} estimates ranged from 0.238 (between HAT and BLA) to 0.805 (between HAT and ERI; Table 2). Using SNP subsets, the highest level of F_{ST} was observed when we used high SNPs for calculation (ranging from 0.282 between HAT and BLA to 0.841 between HAT and ERI), while pairwise F_{ST} decreased dramatically in the low SNP dataset (0.055–0.145). Pairwise F_{ST} estimated from intermediate SNPs showed moderate population differentiation, with values ranging from 0.128 (between HAT and BLA) to 0.426 (between HAT

and ERI; Table 2). The values generated from intermediate SNPs generally mirror the previous F_{ST} estimates from a large-scale walleye genetic divergence study using microsatellites (global F_{ST} of 0.13 ± 0.00; Haponski and Stepien 2014). Therefore, we utilized the intermediate subset as neutral SNPs for downstream population structure analyses.

We examined the STRUCTURE outputs generated from different SNP datasets (full, diagnostic, and intermediate F_{ST}) based on Pr[X|K], ΔK, and parsimony index; all but the Pr[X|K] estimates had the strongest support when $K = 2$ (Fig. 3; Table S3[1]). In all cases of $K = 2$, walleye individuals from Hatchet Creek and Lake Erie represented pure southern and northern walleye alleles, respectively, while walleye individuals sampled from Blackwater Creek showed consistent hybridization patterns, suggesting the establishment of a hybrid zone along this watershed. Based on Pr[X|K] method, the $K = 3$ results suggested the potential contribution from a third genetic source (likely eastern highlands lineage) to these introgressed and (or) hybrid individuals (Fig. S3[1]).

Development and validation of SNP panels

Following our previously established protocol for SNP assay design (Thongda et al. 2018; Zhao et al. 2018, 2019), we developed two panels of 40 SNP multiplexes for extensive walleye population genotyping. Detailed information on SNP panels, including the SNP ID, position, and primer sequences, are listed in Table S4[1]. Among these SNPs, 12 SNPs were excluded from the final assay because of sampling bias or duplicate sequence issues (Li et al. 2014; Zhao et al. 2018). We observed high concordance of genotype calling between GBS and MassARRAY data, with 99.75% matching genotypes; a similarly high genotype concordance was previously reported for SNP marker development in Florida bass (*Micropterus*

1372

Can. J. Fish. Aquat. Sci. Vol. 77, 2020

Table 1. Number of individuals assayed (N), genetic diversity indices, and estimated effective population size (N_e) across three walleye populations using GBS data.

Population	Abbrev.	N	P (%)	H_o	H_e	F_{is}	N_e (95% CI)
Hatchet Creek	HAT	10	15.96	0.06	0.06	−0.050	10.2 (8.9–11.7)
Blackwater Creek	BLA	20	95.33	0.32	0.32	−0.040	26.1 (25.1–27.1)
Lake Erie	ERI	30	53.34	0.19	0.19	0.002	3103.9 (1068.5–∞)

Note: Diversity indices include the proportion of polymorphic SNPs for each population (P), observed heterozygosity (H_o), expected heterozygosity (H_e), and inbreeding coefficient (F_{is}). The full GBS dataset with 2782 SNPs was used for calculation of diversity indices. The intermediate F_{ST} dataset with 921 SNPs was used for estimation of N_e.

Table 2. Pairwise F_{ST} estimates among walleye populations using GBS SNP data.

	Full		Diagnostic		Intermediate	
	F_{ST}	SD	F_{ST}	SD	F_{ST}	SD
HAT vs BLA	0.238	0.004	0.323	0.003	0.128	0.006
HAT vs ERI	0.805	0.005	1.000	0.000	0.426	0.005
BLA vs ERI	0.490	0.005	0.682	0.003	0.236	0.005

Note: Three datasets, full (2782 SNPs), diagnostic (940 SNPs), and intermediate (or neutral, 921 SNPs), were used for F_{ST} calculation. SD is the standard deviation.

floridanus) using the same MassARRAY system (Zhao et al. 2018). We also examined the consistency of MassARRAY genotype calling among technical replicates using a total of 114 individuals and found 99.88% of genotypes matched across multiple plates.

Hybrid identification and genetic analyses using the 68-SNP assay

Using the 68-SNP assay, a total of 545 additional walleye individuals across 23 populations were genotyped for diagnostic SNP validation, extensive population genetic analysis, and hybrid identification across the landscape. We successfully genotyped these samples with a high genotyping rate (mean 98.98%). Using parsimony index as an indicator, we found an optimal number of genetic clusters of $K = 3$ (Fig. 4b). Based on ΔK method, a strong structure for $K = 2$ was detected (Fig. 4b). We did not consider $K = 4$ (supported by Pr[$X|K$] method) because the STRUCTURE plots showed an even split within representative southern or northern samples, which made it difficult to interpret the results (Fig. S4[1]). Similar to the GBS dataset, the SNP assay result with $K = 2$ differentiated the southern walleye lineages (HAT, WHI, and TOM) from other walleye groups, including Great Lakes (HUR, ERI, MUS, THU), upper Mississippi River (MIL), Tennessee River (NAN, FON, NOD, CHI, WAT, DOU, CHE, NOS, and PAT), and the eastern highlands groups (e.g., FOS and ROC; Fig. 4c). Population structure results using $K = 3$ identified the eastern highlands walleye lineage (Fig. 4c). This result also revealed that several individuals in the Tennessee River system (e.g., NOD, CHI, and DOU) showed signs of admixture between the northern and eastern highlands lineages, probably due to historical restocking in these sites. Moreover, 21 individuals from Black Warrior River drainage are the result of admixture from the three lineages: northern, eastern highlands, and Mobile River Basin.

When the nonparameter-based DAPC method was used for population clustering, four genetic clusters received the strongest support (Fig. 4a), which is represented by bottom right genetic cluster from the Great Lakes (HUR, ERI, MUS, THU), upper Mississippi River (MIL), and Little Tennessee River (NAN and FON); eastern highlands (FOS, ROC) cluster on the upper right; Mobile River Basin cluster (HAT, WHI, TOM) on the left; and hybrid group in the middle (MUL, BAL, NOR).

Population differentiation analyses using Hudson's F_{ST} revealed the same divergent patterns, as the largest level of genetic differentiation was found between the southern (HAT, WHI, and TOM) and upper northern groups (including NAN and FON populations from the Little Tennessee River; Fig. 5). Walleye populations in eastern highlands drainages (e.g., FOS and ROC) were genetically differentiated from all other walleye groups.

Hybrid classification

To evaluate walleye hybridization status in the Black Warrior River drainage, we applied a novel framework that combined STRUCTURE, NEWHYBRIDS, and manual assignment methods for hybrid classification. The hybrid analyses were conducted using 300 fixed SNPs from the GBS dataset and 68 SNPs from MassARRAY assays. Before the analyses, we tested the performance of SNPs in assigning simulated individuals to 12 hybrid classes using NEWHYBRIDS. Using three random subsets of 300 SNPs, the mean accuracy was 100% for F_1, F_2, and first-generation backcrosses (B_x-), 99.3% for second-generation backcrosses (B_x2-), 94.0% for third-generation backcrosses (B_x3-), and 99.4% for fourth-generation (B_x4-) backcrosses (Fig. 6). For simulation using the 68-SNP assay data, a correct assignment was made for 100% of F_1 and F_2 hybrids, 95.0% of first-generation backcrosses (B_x-), 89.0% of second-generation backcrosses (B_x2-), 62.0% of third-generation backcrosses (B_x3-), and 77.0% of fourth-generation backcrosses (B_x4-). The mis-assigned individuals were composed of hybrids from later backcross generations, potentially due to the close genotype probabilities among later generation categories (e.g., B_x3 mis-assigned as B_x2 or B_x4).

Based on 68-SNP assay STRUCTURE results, we identified that 34 out of 58 fish in the Black Warrior River drainage were the hybrids between the northern and southern lineages (Q value ≥ 0.05) and used them for hybridization analyses. We measured the assignment concordance between GBS and MassARRAY results in 12 hybrid walleye individuals. From the comparison, all but one hybrid assignment gave congruent classifications (Table S5[1]). The one mis-assignment was due to rare northern walleye alleles not captured by MassARRAY SNPs. Using the same classification methods, we assigned various hybrid classes to the 34 samples. As shown in Table 3, advanced stages of hybridization were the most commonly observed categories among these hybrid individuals, with early-generation hybrids (B_xS and B_xN) making up only 8.8% of total hybrids sampled (Table 3). Later-generation hybrids (B_x2, B_x3, F_x, F_2) were dominated by the F_2 (51.6%) and F_xS categories (35.5%). We observed some genetically pure walleye in Black Warrior River drainage, with 2 out of 11 fish identified as pure southern walleye in the Mulberry Fork, and one out of three fish as pure northern walleye in the North River.

Historical demographic analyses of Hatchet Creek

We assessed four demographic models to reconstruct the population's demographic history (Fig. 2). The continuous decline model (M2) was the best-supported model, regardless of assigned mutation rates (Fig. 2 and Table S6[1]). Based on this model, we obtained a coalescent diploid N_e estimate of 7 for the current

Fig. 4. Population structure results inferred from DAPC and STRUCTURE using 68-SNP genotyping data. (*a*) Scatterplot output from DAPC for the genetic cluster of walleye individuals; (*b*) estimation of the most likely number of populations (*K*) using Pr[*X*|*K*], Δ*K*, and parsimony index values; (*c*) STRUCTURE results using *K* = 2 and 3 for all genotyped walleye individuals. HAT = Hatchet Creek; WHI = White Plains; TOM = Tombigbee River; MUL = Mulberry Fork; BLA = Blackwater Creek; NOR = North River; NAN = Nantahala Lake; FON = Lake Fontana; NOD = Normandy Lake; CHI = Chickamauga Lake; WAT = Watts Bar Lake; DOU = Douglas Lake; CHE = Cherokee Lake; NOS = Norris Lake; PAT = Fort Patrick Henry Lake; MIL = Mille Lacs Lake; FOS = Fosters Falls; ROC = Rockcastle River; LEV = Levisa Fork; HUR = Huron River; ERI = Lake Erie; MUS = Muskegon River; THU = Thunder Bay. [Colour online.]
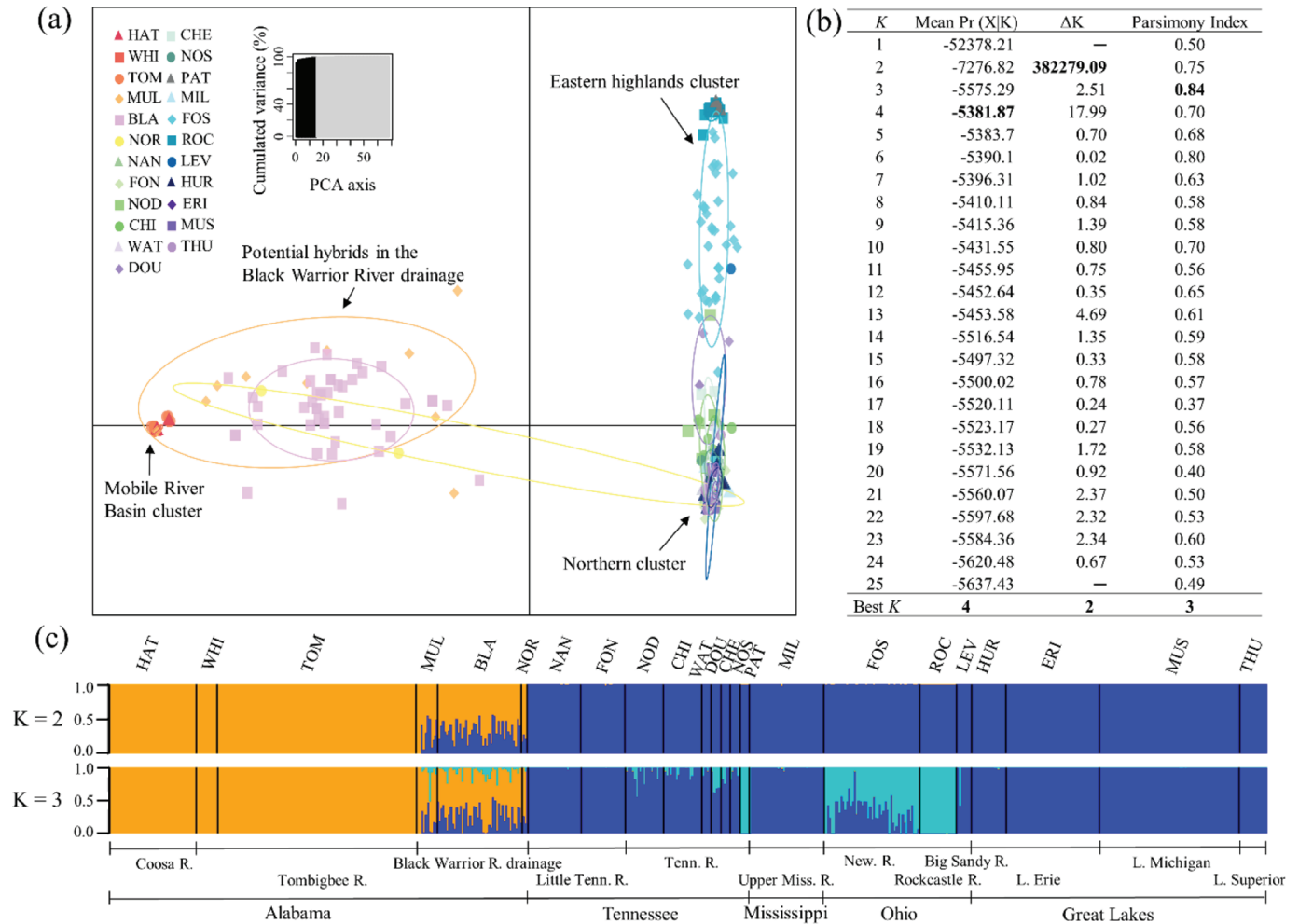
Hatchet Creek walleye population, with a mean population growth rate of 1.33E^{-05} (Table S7[1]). By combining the values for mutation rate (2 × 10^{-9} per site per generation) with nucleotide diversity (θ_π = 0.166%, estimated from STACKS), we estimated a long-term haploid effective population size that approximated 415 000. Given that the long-term harmonic mean of N_e corresponds to $2N_e$ generations (with more weight given to recent generations; Hare et al. 2011), the estimated ancestral haploid N_e can be traced back to ~800 000 generations ago.

Discussion

Walleye is an ecologically and economically valuable freshwater species in North America that is threatened by overexploitation and genetic risks from anthropogenic activities. Using 2782 GBS-derived SNPs, we confirmed the genetic distinctness of Mobile River Basin walleye and identified an anthropogenic hybrid zone that likely resulted from the stocking of northern walleye into the Black Warrior River drainage of Alabama. We also found strong evidence of a historical declining population trend with reduced genetic diversity and effective population size in a pure southern walleye population from Hatchet Creek, Alabama. We have shown that a suite of 68 SNPs can collectively classify

advanced-generation hybrids between northern and southern individuals, which will be useful for conservation and protection of putatively locally adapted stocks in the Mobile River Basin.

Genetic divergence between southern and other walleye groups

A major goal of this study was to characterize the genetic differentiation between northern and southern walleye groups. In the current study, the unique genetic pattern of southern Mobile River Basin walleye was characterized using several approaches. First, we determined that 940 (33.79%) of the 2782 GBS SNPs were fixed between Hatchet Creek (HAT) and Lake Erie (ERI) walleye populations. For comparison, the ratio of diagnostic walleye SNPs identified in our study was remarkably higher than the level of interspecific difference (9.2%, 675 of 7346 loci) observed between Russian (*Acipenser gueldenstaedtii*) and Persian (*Acipenser persicus*) sturgeon (Ogden et al. 2013). These diagnostic loci can be useful for investigating selection and adaptive evolution, heterogeneous genome divergence, and intraspecific introgression or hybridization (Harrison and Larson 2016; Narum et al. 2013).

We also examined population structure and differentiation using both GBS and MassARRAY data and found that Mobile River Basin lineage is strikingly different from other walleye popula-

Fig. 5. Pairwise Hudson's F_{ST} estimate for walleye populations using the 68-SNP assay.
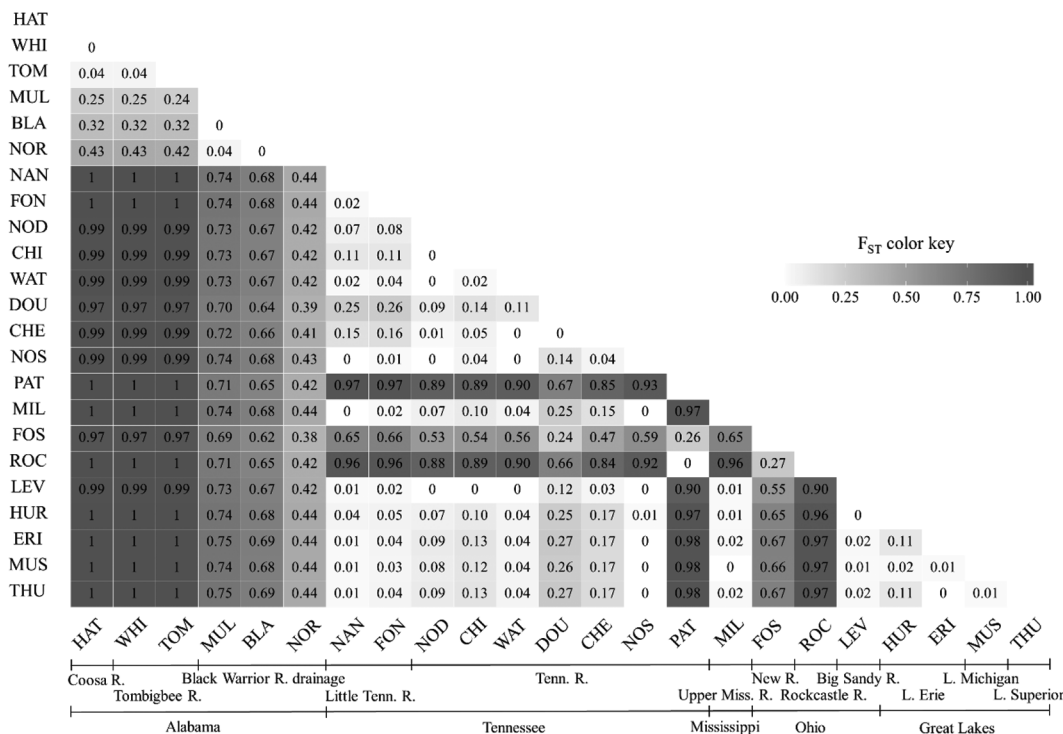

Fig. 6. NEWHYBRIDS simulation analyses based on three GBS datasets with 300 randomly selected diagnostic SNPs and MassARRAY dataset with 68 SNPs. Twelve hybrid categories (listed at the bottom of the figure; defined in Table S2[1]) were set as "Jeffreys-like priors". Values at the bottom of the figure show the accuracy of assignment to each simulated hybrid category. [Colour online.]
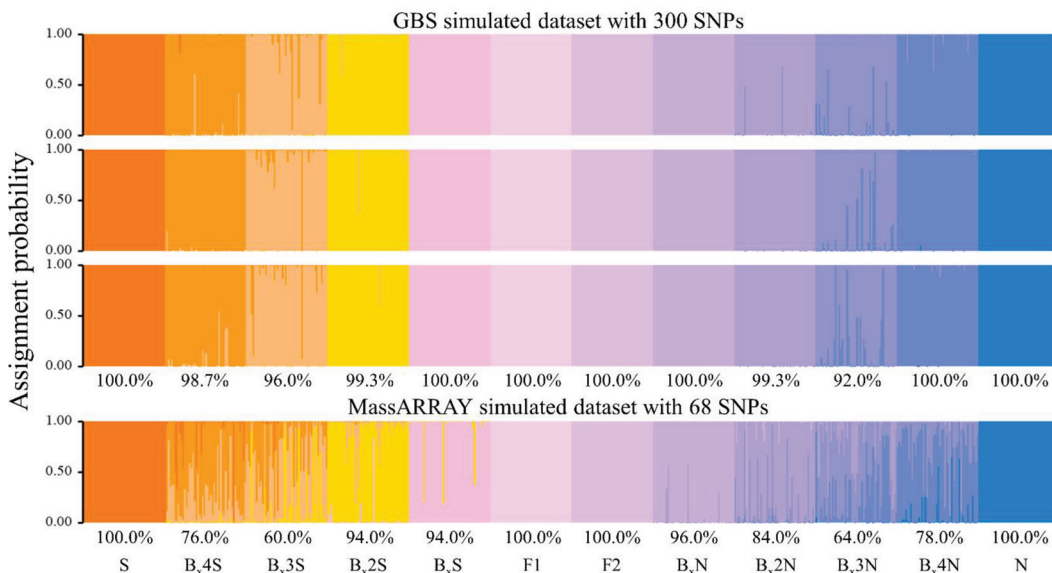
tions. The distinctiveness of this walleye lineage may be due to their long undisturbed history of local adaption and independent evolution in this isolated southern river system ([Petit et al. 2003](); [Stepien et al. 2009]()). Similar genetic structure was previously reported for southerly populations of yellow perch (*Perca flavescens*), a species with a similar native range and life history characteristics to walleye ([Stepien et al. 2015]()). Lastly, our GBS data showed that the walleye population in Hatchet Creek had lower genetic diversity and estimated effective population size compared with the walleye populations sampled from Lake Erie and Blackwater Creek,

reflecting the long-term isolation and potential historical population decline in pure southern groups ([Haponski and Stepien 2014]()).

SNP assay resolution for population structure

Although our 68-SNP assay is able to identify pure and hybrid southern walleye in a rapid and accurate manner, the limited number of representative populations in the GBS data and utilization of fixed markers may restrict the assay application for resolving fine-scale genetic structure among extensive walleye populations. For example, we observed three major genetic clusters (Great Lakes – upper Mississippi, eastern highlands, and Mo-

Table 3. Summary of the hybrid analyses on walleye sampled from Black Warrior River drainage based on manual and NEWHYBRIDS assignment.

	B_x4S	B_x3S	B_x2S	B_xS	F_1	F_2	B_xN	B_x2N	B_x3N	B_x4N	F_xN	F_xS	Total
Mulberry Fork (MUL)													
N	0	0	0	0	0	2	0	0	0	0	0	2	4
%	0.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0	50.0	
Blackwater Creek (BLA)													
N	0	0	4	2	0	14	0	0	0	0	0	9	29
%	0.0	0.0	13.8	6.9	0.0	48.3	0.0	0.0	0.0	0.0	0.0	31.0	
North River (NOR)													
N	0	0	0	1	0	0	0	0	0	0	0	0	1
%	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Note: Twelve hybrid categories (defined in Table S2[1]) were set as "Jeffreys-like priors".

bile River Basin) among examined walleye individuals using the SNP assay (Figs. 4 and 5), which only covers the broad-scale structure patterns previously identified from mtDNA and microsatellites studies (Billington 1996; Billington et al. 1992; Haponski and Stepien 2014; Stepien and Faber 1998; Stepien et al. 2009). Although our SNP assay was not explicitly designed to identify the eastern highlands walleye group, it nevertheless distinguishes walleye spawning in the Ohio River drainage (FOS and ROC) as genetically distinct, reflecting their historical isolation (Palmer et al. 2006; White et al. 2012). However, finer-scale demarcations of walleye populations across the upper northern regions (Northwest Lake Plains, Great Lakes watershed, and North Atlantic coastal) are not recovered. Future work could develop SNP resources for characterizing fine-scale population structure through GBS sequencing of all five major walleye lineages or by using existing GBS data generated from other northern walleye populations (Allen et al. 2018; Chen et al. 2019).

Identification of southern and northern walleye hybrids

Characterizing and detecting the genomic composition of hybrids is critical for studies of hybrid zone dynamics, inheritance of traits, and consequences of stocking and hybridization for evolution, fishery management, and conservation (Fitzpatrick 2012). Empirical data and simulations have demonstrated that 50 or more ancestry-informative markers are needed to accurately identify F_2 hybrids and advanced-generation backcross individuals (Fitzpatrick 2012; Malde et al. 2017). For instance, a bighead (*Hypophthalmichthys nobilis*) and silver (*Hypophthalmichthys molitrix*) carp study using 57 diagnostic SNPs successfully identified advanced-stage hybrids throughout their distribution in the Mississippi River Basin (Lamer et al. 2015). In our study, an SNP assay with 68 diagnostic markers was developed for rapid and accurate identification of genetic purity and classification of various (northern or southern) hybrid classes among walleye individuals. The precision of hybrid classification using our SNP panel was evaluated by three main aspects: reliability of genotypes, the accuracy of simulations, and the repeatability of assignments between GBS and MassARRAY data. Owing to the stringent criteria applied for assay development, we observed high concordance between GBS and MassARRAY genotypes (99.75%) and found only nine discrepancies out of 7250 genotype comparisons among technical replicates (114 samples genotyped twice with MassARRAY), suggesting the high reliability and repeatability of our genotyping data. In both cases, the genotype discrepancy did not impact our assignment results. Secondly, simulation analyses using GBS and MassARRAY datasets showed that our markers have enough discriminatory power to correctly identify up to third-generation hybrids (>89% accuracy). Similar simulation results based on 96 species-specific SNPs were previously reported for two North Atlantic eel species (*Anguilla anguilla* and *Anguilla rostrata*; Pujolar et al. 2014). In addition, we only found one discrepancy when we compared the hybrid assignment results generated from GBS (300 fixed SNPs) and the 68-SNP assay. The assignment discrepancy was potentially due

to the lack of homozygous loci in the least genetically represented hybrid classes in MassARRAY data (i.e., $B_x3S \times B_x3S$ only possesses an average of 1.56% or 0.27 of 68 loci homozygous for northern walleye). In our case, because MassARRAY data failed to detect rare loci homozygous for northern walleye, one individual (among 12 hybrids) identified as F_xS category from GBS data were mis-assigned to B_x2S with 68 MassARRAY markers.

Walleye from the Black Warrior River drainage were mostly later-generation hybrids (31 out of 34 or 91.2%; Table 3), which is unsurprising given the history of non-native walleye stocking in this watershed. Between 1975 and 1985, thousands of northern walleye fingerlings sourced from Seneca Lake (Ohio) and Pymatuning Lake (Pennsylvania) were stocked into Tuscaloosa Reservoir (namely the North River population) and Sipsey River (Billington and Maceina 1997). Previous mitochondrial DNA analyses found evidence of walleye hybridization along these waterways and hypothesized that the introduced fish could potentially pass over Lake Tuscaloosa dam, migrate upstream in the Black Warrior River system, and hybridize with native southern walleye (Billington and Maceina 1997). Given that the generation interval approximates the mean age of breeding individuals for populations with overlapping generations (Hill 1979), it is expected that multiple generations of hybridization have occurred in the Black Warrior River drainage.

Conservation implications in southern walleye

Given the major genetic distinctiveness of walleye in the Mobile River Basin, conserving these populations is critical for resource management and preserving biodiversity. Any efforts at conservation through stocking and genetic rescue for these populations should assess the genomic purity of donor walleye stocks (Chattopadhyay et al. 2019). The SNP panel and MassARRAY system offer a cost-effective and reliable tool for this purpose, with our 68-SNP assay already being implemented in ongoing stream survey and captive breeding programs. Meanwhile, careful monitoring of genetic diversity between donor and recipient walleye stocks should be conducted to stall the erosion of genetic diversity and enhance the long-term survival of southern walleye in the wild. This could be accomplished by designing additional polymorphic SNP panels from the intermediate SNP dataset.

Our analysis of historical demography suggested that the southern walleye population in Hatchet Creek has undergone a continuous decline in population size over ~800 000 generations. We stress that the accuracy of this estimate may be impaired by our modest sample numbers and from the lack of accurate calculations of nucleotide diversity and mutation rate values. This result should also be taken as a qualitative survey of demographic history for Hatchet Creek walleye. The 800 000 generation time (~2.4–3 million years) estimate, derived from a small number of available samples, may be too long, as it is older than some estimates for the southern walleye lineage (~1.17 million years; Billington et al. 1992). We hypothesize that during the Quaternary glaciations (2.58 million years ago; Gibbard et al. 2010), the south-

1376

Can. J. Fish. Aquat. Sci. Vol. 77, 2020

ern walleye lineage contracted into refugia (including the Alabama River system) and expanded out when conditions allowed. Walleye in Alabama may have been isolated, however, since the Appalachian River (the old upper Tennessee, which used to flow southwest into Alabama) changed course near Chattanooga, Tennessee, to take its current course. The Mobile River Basin populations may have contracted up into the cooler headwaters since the end of the Pleistocene (21 000 to 11 000 years ago; Tóth et al. 2019) and evolved independently. This time period would be long enough for favorable combinations of alleles to arise and be retained by strong local selection. Nevertheless, the low effective population size estimated from both demographic (Fig. 2) and genetic diversity (Table 1) analyses suggest a high risk of imminent local extinction, as genetic drift in small populations can have great influence on genetic diversity and population fitness (Franckowiak et al. 2009; Willi et al. 2006). The declining trend in southern walleye population size is also reflected by the low catch rate of wild walleye throughout the Alabama River systems (e.g., only 31 southern walleye were collected in a 2-year survey; Billington et al. 1997) and rare spawning events reported in Luxapallila Creek from the Tombigbee River (Schramm et al. 2004). Several natural and anthropogenic factors may be contributing to the low N_e estimate for the Hatchet Creek walleye population. First, walleye in their natural environment are characterized by low and unequal reproductive success and high mortality rate, all of which negatively affect the effective population size (Ivan et al. 2010). In addition, ecosystem and community changes, including the introduction of predators or competitors, habitat degradation, climate change, and altered hydrologic conditions, may influence rates of growth, survival, and recruitment in the southern walleye populations (Nate et al. 2011). Lastly, illegal fishery or angling exploitation may pose a major risk to walleye populations by negatively impacting recruitment variability, growth rate, and age to maturity (Baccante and Colby 1996; McMahon and Bennett 1996; Spangler et al. 1977).

Our results highlight the need for careful conservation management of southern walleye in the Mobile River Basin. Given the continued threat of habitat loss and climate change, complementary data related to life history features (e.g., generation time, spawning success, survival rate) and population dynamics (e.g., census population size, exploitation rate) in southern walleye need to be collected to facilitate the conservation and management of this unique group. As a final point, it is essential to highlight that while stocking and genetic rescue processes (Whiteley et al. 2015) can help facilitate the conservation and management of southern walleye, they do not address issues of habitat loss and ecological degradation. A well-designed restoration strategy and strict regulation are necessary to help recover the ecosystem and guard against the extinction of southern walleye (Peterson et al. 2003).

Data accessibility

Raw Illumina reads for walleye GBS have been archived at the NCBI SRA under BioProject accession No. PRJNA555395. The raw Illumina reads for walleye whole-genome sequencing have been archived at the NCBI SRA under BioProject accession No. PRJNA555394. The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession SOZC00000000. The version described in this paper is version SOZC01000000.

Author contributions

H.Z. and E.P. conceived and designed the study; S.J.R., C.A.S., and E.M.H. contributed the samples; S.J. processed the samples; G.K. generated the map; H.Z. performed all analyses; and H.Z., K.S., and E.P. wrote the manuscript with contributions and review from other co-authors.

References

Allen, B.E., Bowles, E., Morris, M.R., and Rogers, S.M. 2018. Loss of SNP genetic diversity following population collapse in a recreational walleye (*Sander vitreus*) fishery. Can. J. Fish. Aquat. Sci. **75**(10): 1644–1651. doi:10.1139/cjfas-2017-0164.

Anderson, E.C., and Thompson, E.A. 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics, **160**(3): 1217–1229. PMID:11901135.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., and Hohenlohe, P.A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. **17**(2): 81–92. doi:10.1038/nrg.2015.28. PMID:26729255.

Baccante, D.A., and Colby, P.J. 1996. Harvest, density and reproductive characteristics of North American walleye populations. Ann. Zool. Fenn. **33**: 601–615.

Barton, B.A. 2011. Biology, management, and culture of walleye and sauger. American Fisheries Society, Bethesda, Md.

Bhatia, G., Patterson, N., Sankararaman, S., and Price, A.L. 2013. Estimating and interpreting F_{ST}: the impact of rare variants. Genome Res. **23**(9): 1514–1521. doi:10.1101/gr.154831.113. PMID:23861382.

Billington, N. 1996. Geographical distribution of mitochondrial DNA (mtDNA) variation in walleye, sauger, and yellow perch. Ann. Zool. Fenn. **33**: 699–706.

Billington, N., and Hebert, P.D.N. 1988. Mitochondrial DNA variation in Great Lakes walleye (*Stizostedion vitreum*) populations. Can. J. Fish. Aquat. Sci. **45**(4): 643–654. doi:10.1139/f88-078.

Billington, N., and Maceina, M.J. 1997. Genetic and population characteristics of walleyes in the Mobile drainage of Alabama. Trans. Am. Fish. Soc. **126**(5): 804–814. doi:10.1577/1548-8659(1997)126<0804:GAPCOW>2.3.CO;2.

Billington, N., and Strange, R.M. 1995. Mitochondrial DNA analysis confirms the existence of a genetically divergent walleye population in northeastern Mississippi. Trans. Am. Fish. Soc. **124**(5): 770–776. doi:10.1577/1548-8659(1995)124<0770:NMDACT>2.3.CO;2.

Billington, N., Barrette, R.J., and Hebert, P.D. 1992. Management implications of mitochondrial DNA variation in walleye stocks. N. Am. J. Fish. Manage. **12**(2): 276–284. doi:10.1577/1548-8675(1992)012<0276:MIOMDV>2.3.CO;2.

Billington, N., Strange, R.M., and Maceina, M.J. 1997. Mitochondrial-DNA confirmation of southern walleye in the Mobile Basin, Alabama. *In* Proceedings of the Annual Conference Southeastern Association of Fish and Wildlife Agencies. pp. 12–117.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. **22**(11): 3124–3140. doi:10.1111/mec.12354. PMID:23701397.

Chattopadhyay, B., Garg, K.M., Soo, Y.J., Low, G.W., Frechette, J.L., and Rheindt, F.E. 2019. Conservation genomics in the fight to help the recovery of the critically endangered Siamese crocodile *Crocodylus siamensis*. Mol. Ecol. **28**(5): 936–950. doi:10.1111/mec.15023. PMID:30659682.

Chen, K.Y., Euclide, P.T., Ludsin, S.A., Larson, W., Sovic, M.G., Gibbs, H.L., and Marschall, E. 2019. RAD-seq refines previous estimates of genetic structure in Lake Erie walleye (*Sander vitreus*). Trans. Am. Fish. Soc. **149**: 159–173. doi:10.1002/tafs.10215.

Cochran-Biederman, J.L., Wyman, K.E., French, W.E., and Loppnow, G.L. 2015. Identifying correlates of success and failure of native freshwater fish reintroductions. Conserv. Biol. **29**(1): 175–186. doi:10.1111/cobi.12374. PMID:25115187.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., et al. 2011. The variant call format and VCFtools. Bioinformatics, **27**(15): 2156–2158. doi:10.1093/bioinformatics/btr330. PMID:21653522.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., and Ovenden, J.R. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. Mol. Ecol. Resour. **14**(1): 209–214. doi:10.1111/1755-0998.12157. PMID:23992227.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z., Knowler, D.J., Leveque, C., et al. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol. Rev. Cambr. Philos. Soc. **81**(2): 163–182. doi:10.1017/S1464793105006950. PMID:16336747.

Elliott, L., and Russello, M.A. 2018. SNP panels for differentiating advanced-

generation hybrid classes in recently diverged stocks: a sensitivity analysis to inform monitoring of sockeye salmon re-stocking programs. Fish. Res. 208: 339–345. doi:10.1016/j.fishres.2018.09.001.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14(8): 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x. PMID:15969739.

Excoffier, L., and Lischer, H.E. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10(3): 564–567. doi:10.1111/j.1755-0998.2010.02847.x. PMID:21565059.

Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V.C., and Foll, M. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9(10): e1003905. doi:10.1371/journal.pgen.1003905. PMID:24204310.

Feng, C., Pettersson, M., Lamichhaney, S., Rubin, C.-J., Rafati, N., Casini, M., et al. 2017. Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. Elife, 6: e23907. doi:10.7554/eLife.23907. PMID: 28665273.

Fitzpatrick, B.M. 2012. Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evol. Biol. 12: 131. doi:10.1186/1471-2148-12-131. PMID: 22849298.

Foll, M., and 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. doi:10.1534/genetics.108.092221. PMID: 18780740.

Franckowiak, R., Sloss, B., Bozek, M., and Newman, S. 2009. Temporal effective size estimates of a managed walleye Sander vitreus population and implications for genetic-based management. J. Fish Biol. 74(5): 1086–1103. doi:10.1111/j.1095-8649.2008.02170.x. PMID:20735621.

Gibbard, P.L., Head, M.J., Walker, M.J., and Subcommission on Quaternary Stratigraphy. 2010. Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. J. Quat. Sci. 25(2): 96–102. doi:10.1002/jqs.1338.

Haponski, A.E., and Stepien, C.A. 2014. A population genetic window into the past and future of the walleye Sander vitreus: relation to historic walleye and the extinct "blue pike" S. v. "glaucus". BMC Evol. Biol. 14: 133. doi:10.1186/1471-2148-14-133. PMID:24941945.

Hare, M.P., Nunney, L., Schwartz, M.K., Ruzzante, D.E., Burford, M., Waples, R.S., et al. 2011. Understanding and estimating effective population size for practical application in marine species management. Conserv. Biol. 25(3): 438–449. doi:10.1111/j.1523-1739.2010.01637.x. PMID:21284731.

Harrison, R.G., and Larson, E.L. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Mol. Ecol. 25(11): 2454–2466. doi:10.1111/mec.13582. PMID:26857437.

Hill, W.G. 1979. A note on effective population size with overlapping generations. Genetics, 92(1): 317–322. PMID:17248921.

Hokanson, K.E. 1977. Temperature requirements of some percids and adaptations to the seasonal temperature cycle. J. Fish. Board Can. 34(10): 1524–1550. doi:10.1139/f77-217.

Hudson, R.R., Slatkin, M., and Maddison, W.P. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics, 132(2): 583–589. PMID:1427045.

Ivan, L.N., Rutherford, E.S., Riseng, C., and Tyler, J.A. 2010. Density, production, and survival of walleye (Sander vitreus) eggs in the Muskegon River, Michigan. J. Gt. Lakes Res. 36(2): 328–337. doi:10.1016/j.jglr.2010.02.010.

Jombart, T., and Ahmed, I. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27(21): 3070–3071. doi:10.1093/bioinformatics/btr521. PMID:21926124.

Lamer, J.T., Ruebush, B.C., Arbieva, Z.H., McClelland, M.A., Epifanio, J.M., and Sass, G.G. 2015. Diagnostic SNPs reveal widespread introgressive hybridization between introduced bighead and silver carp in the Mississippi River Basin. Mol. Ecol. 24(15): 3931–3943. doi:10.1111/mec.13285. PMID:26096550.

Li, C., Waldbieser, G., Bosworth, B., Beck, B.H., Thongda, W., and Peatman, E. 2014. SNP discovery in wild and domesticated populations of blue catfish, Ictalurus furcatus, using genotyping-by-sequencing and subsequent SNP validation. Mol. Ecol. Resour. 14(6): 1261–1270. doi:10.1111/1755-0998.12272. PMID: 24797164.

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14): 1754–1760. doi:10.1093/bioinformatics/btp324. PMID:19451168.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16): 2078–2079. doi:10.1093/bioinformatics/btp352. PMID:19505943.

Lutz-Carrillo, D.J., Nice, C.C., Bonner, T.H., Forstner, M.R., and Fries, L.T. 2006. Admixture analysis of Florida largemouth bass and northern largemouth bass using microsatellite loci. Trans. Am. Fish. Soc. 135(3): 779–791. doi:10.1577/T04-221.1.

Madeira, M.J., Gómez-Moliner, B.J., and Machordom, A. 2005. Genetic introgression on freshwater fish populations caused by restocking programmes. Biol. Invasions 7: 117–125. doi:10.1007/s10530-004-9641-x.

Malde, K., Seliussen, B.B., Quintela, M., Dahle, G., Besnier, F., Skaug, H.J., Øien, N., et al. 2017. Whole genome resequencing reveals diagnostic markers for investigating global migration and hybridization between minke whale species. BMC Genomics, 18: 76. doi:10.1186/s12864-016-3416-5. PMID:28086785.

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., and

Durbin, R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. 2(12): 1940–1955. doi:10.1038/s41559-018-0717-x. PMID:30455444.

McMahon, T.E., and Bennett, D.H. 1996. Walleye and northern pike: boost or bane to northwest fisheries? Fisheries, 21(8): 6–13. doi:10.1577/1548-8446(1996)021<0006:WANP>2.0.CO;2.

Miller, L.M., and Kapuscinski, A.R. 2003. Genetic guidelines for hatchery supplement programs. In Population genetics: principles and applications for fisheries scientists. Edited by E.M. Hallerman. American Fisheries Society, Bethesda, Md. pp. 329–355.

Murphy, B.R. 1990. Evidence for a genetically unique walleye population in the upper Tombigbee River system of northeastern Mississippi. In Southeastern Fishes Council Proceedings. pp. 14–16.

Nachman, M.W., and Crowell, S.L. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics, 156(1): 297–304. PMID:10978293.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., and Hohenlohe, P.A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. 22(11): 2841–2847. doi:10.1111/mec.12350. PMID:23711105.

Nate, N.A., Hansen, M.J., Rudstam, L.G., Knight, R.L., and Newman, S.P. 2011. Population and community dynamics of walleye. In Biology, management, and culture of walleye and sauger. American Fisheries Society, Bethesda, Md. pp. 321–374.

Nielsen, E.E., Bach, L.A., and Kotlicki, P. 2006. HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. Mol. Ecol. Notes, 6(4): 971–973. doi:10.1111/j.1471-8286.2006.01433.x.

Ogden, R., Gharbi, K., Mugue, N., Martinsohn, J., Senn, H., Davey, J., et al. 2013. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. Mol. Ecol. 22(11): 3112–3123. doi:10.1111/mec.12234. PMID: 23473098.

Palmer, G.C., Culver, C., Dutton, D., Murphy, B.R., Hallerman, E.M., Billington, N., and Williams, J. 2006. Genetic distinct walleye stocks in Claytor Lake and the upper New River, Virginia. P Southeast Fish Wild Agencies, 60: 125–131.

Patterson, N., Price, A.L., and Reich, D. 2006. Population structure and eigenanalysis. PLoS Genet. 2(12): e190. doi:10.1371/journal.pgen.0020190. PMID: 17194218.

Peakall, R., and Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research — an update. Bioinformatics, 28(19): 2537–2539. doi:10.1093/bioinformatics/bts460. PMID:22820204.

Peterson, G.D., Cumming, G.S., and Carpenter, S.R. 2003. Scenario planning: a tool for conservation in an uncertain world. Conserv. Biol. 17(2): 358–366. doi:10.1046/j.1523-1739.2003.01491.x.

Petit, R.J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., et al. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. Science, 300(5625): 1563–1565. doi:10.1126/science.1083264. PMID:12791991.

Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P. 2005. Genotyping errors: causes, consequences and solutions. Nat. Rev. Genet. 6(11): 847–859. doi:10.1038/nrg1707. PMID:16304600.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, 155(2): 945–959. PMID: 10835412.

Pritchard, V.L., Erkinaro, J., Kent, M.P., Niemelä, E., Orell, P., Lien, S., and Primmer, C.R. 2016. Single nucleotide polymorphisms to discriminate different classes of hybrid between wild Atlantic salmon and aquaculture escapees. Evol. Appl. 9(8): 1017–1031. doi:10.1111/eva.12407. PMID:27606009.

Pujolar, J.M., Jacobsen, M., Als, T.D., Frydenberg, J., Magnussen, E., Jónsson, B., et al. 2014. Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. Heredity, 112(6): 627–637. doi:10.1038/hdy.2013.145. PMID:24424165.

Regier, H.A., Applegate, V.C., and Ryder, R.A. 1969. The ecology and management of the walleye in western Lake Erie. Great Lakes Fishery Commission, Technical Report 15. pp. I–101.

Schramm, H.L., Jr., Hart, J., and Hanson, L.A. 2004. Status and reproduction of Gulf Coast strain walleye in a Tombigbee River tributary. Southeast. Nat. 3: 745–757. doi:10.1656/1528-7092(2004)003[0745:SAROGC]2.0.CO;2.

Seddon, P.J., Armstrong, D.P., and Maloney, R.F. 2007. Developing the science of reintroduction biology. Conserv. Biol. 21(2): 303–312. doi:10.1111/j.1523-1739.2006.00627.x. PMID:17391180.

Spangler, G., Payne, N., Thorpe, J., Byrne, J., Regier, H., and Christie, W. 1977. Responses of percids to exploitation. J. Fish. Board Can. 34(10): 1983–1988. doi:10.1139/f77-265.

Stepien, C.A., and Faber, J.E. 1998. Population genetic structure, phylogeography and spawning philopatry in walleye (Stizostedion vitreum) from mitochondrial DNA control region sequences. Mol. Ecol. 7(12): 1757–1769. doi:10.1046/j.1365-294x.1998.00512.x. PMID:9859203.

Stepien, C.A., Murphy, D.J., Lohner, R.N., Sepulveda-Villet, O.J., and Haponski, A.E. 2009. Signatures of vicariance, postglacial dispersal and spawning philopatry: population genetics of the walleye Sander vitreus. Mol. Ecol. 18(16): 3411–3428. doi:10.1111/j.1365-294X.2009.04291.x. PMID:19659479.

Stepien, C.A., Sepulveda-Villet, O.J., and Haponski, A.E. 2015. Comparative genetic diversity, population structure, and adaptations of walleye and yellow perch across North America. In Biology and culture of percid fishes. Springer. pp. 643–689.

Strange, R.M., and Stepien, C.A. 2007. Genetic divergence and connectivity among river and reef spawning groups of walleye (*Sander vitreus vitreus*) in Lake Erie. Can. J. Fish. Aquat. Sci. **64**(3): 437–448. doi:10.1139/f07-022.

Thongda, W., Zhao, H., Zhang, D., Jescovitch, L.N., Liu, M., Guo, X., et al. 2018. Development of SNP panels as a new tool to assess the genetic diversity, population structure, and parentage analysis of the eastern oyster (*Crassostrea virginica*). Mar. Biotechnol. **20**(3): 385–395. doi:10.1007/s10126-018-9803-y. PMID:29532334.

Thongda, W., Lewis, M., Zhao, H., Bowen, B., Lutz-Carrillo, D.J., Peoples, B.K., and Peatman, E. 2020. Species-diagnostic SNP markers for the black basses (*Micropterus* spp.): a new tool for black bass conservation and management. Conserv. Genet. Resour. **12**: 319–328. doi:10.1007/s12686-019-01109-8.

Tóth, A.B., Lyons, S.K., Barr, W.A., Behrensmeyer, A.K., Blois, J.L., Bobe, R., et al. 2019. Reorganization of surviving mammal communities after the end-Pleistocene megafaunal extinction. Science, **365**(6459): 1305–1308. doi:10.1126/science.aaw1605. PMID:31604240.

Wang, J. 2019. A parsimony estimator of the number of populations from a STRUCTURE-like analysis. Mol. Ecol. Resour. **19**(4): 970–981. doi:10.1111/1755-0998.13000. PMID:30681281.

Ward, R.D., Billington, N., and Hebert, P.D. 1989. Comparison of allozyme and mitochondrial DNA variation in populations of walleye, *Stizostedion vitreum*. Can. J. Fish. Aquat. Sci. **46**(12): 2074–2084. doi:10.1139/f89-257.

White, M.M., Faber, J.E., and Zipfel, K.J. 2012. Genetic identity of walleye in the Cumberland River. Am. Midl. Nat. **167**(2): 373–384. doi:10.1674/0003-0031-167.2.373.

Whiteley, A.R., Fitzpatrick, S.W., Funk, W.C., and Tallmon, D.A. 2015. Genetic rescue to the rescue. Trends Ecol. Evol. **30**(1): 42–49. doi:10.1016/j.tree.2014.10.009. PMID:25435267.

Willi, Y., Van Buskirk, J., and Hoffmann, A.A. 2006. Limits to the adaptive potential of small populations. Annu. Rev. Ecol. Evol. Syst. **37**: 433–458. doi:10.1146/annurev.ecolsys.37.091305.110145.

Zhao, H., Li, C., Hargrove, J.S., Bowen, B.R., Thongda, W., Zhang, D., et al. 2018. SNP marker panels for parentage assignment and traceability in the Florida bass (*Micropterus floridanus*). Aquaculture, **485**: 30–38. doi:10.1016/j.aquaculture.2017.11.014.

Zhao, H., Fuller, A., Thongda, W., Mohammed, H., Abernathy, J., Beck, B., and Peatman, E. 2019. SNP panel development for genetic management of wild and domesticated white bass (*Morone chrysops*). Anim. Genet. **50**(1): 92–96. doi:10.1111/age.12747. PMID:30426519.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, **28**(24): 3326–3328. doi:10.1093/bioinformatics/bts606. PMID:23060615.

Zimin, A.V., Marcais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. 2013. The MaSuRCA genome assembler. Bioinformatics, **29**(21): 2669–2677. doi:10.1093/bioinformatics/btt476. PMID:23990416.