Received: 21 January 2019 | Revised: 17 February 2020 | Accepted: 20 February 2020

DOI: 10.1111/mec.15406

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Is Niagara Falls a barrier to gene flow in riverine fishes? A test using genome-wide SNP data from seven native species

Nathan K. Lujan[1,2] | Jason T. Weir[1] | Brice P. Noonan[2] | Nathan R. Lovejoy[1] | Nicholas E. Mandrak[1]

[1]Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON, Canada

[2]Department of Biological Sciences, University of Mississippi, Oxford, MS, USA

Correspondence
Nathan K. Lujan, Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON, Canada.
Email: nklujan@gmail.com

Present address
Nathan K. Lujan, Department of Ichthyology, American Museum of Natural History, New York, NY, USA

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2014-05226, RGPIN-2016-06221, RGPIN-2016-06538 and 492890; University of Toronto; Canadian Foundation for Innovation; Government of Ontario; Ontario Research Fund – Research Excellence

Abstract

Since the early Holocene, fish population genetics in the Laurentian Great Lakes have been shaped by the dual influences of habitat structure and post-glacial dispersal. Riverscape genetics theory predicts that longitudinal habitat corridors and unidirectional downstream water-flow drive the downstream accumulation of genetic diversity, whereas post-glacial dispersal theory predicts that fish genetic diversity should decrease with increasing distance from glacial refugia. This study examines populations of seven native fish species codistributed above and below the 58 m high Niagara Falls – a hypothesized barrier to gene flow in aquatic species. A better understanding of Niagara Falls' role as a barrier to gene flow and dispersal is needed to identify drivers of Great Lakes genetic diversity and guide strategies to limit exotic species invasions. We used genome-wide SNPs and coalescent models to test whether populations are: (a) genetically distinct, consistent with the Niagara Falls barrier hypothesis; (b) more genetically diverse upstream, consistent with post-glacial expansion theory, or downstream, consistent with the riverscape habitat theory; and (c) have migrated either upstream or downstream past Niagara Falls. We found that genetic diversity is consistently greater below Niagara Falls and the falls are an effective barrier to migration, but two species have probably dispersed upstream past the falls after glacial retreat yet before opening of the Welland Canal. Models restricting migration to after opening of the Welland Canal were generally rejected. These results help explain how river habitat features affect aquatic species' genetic diversity and highlight the need to better understand post-glacial dispersal pathways.

KEYWORDS
fish, fisheries management, genomics/proteomics, invasive species, landscape genetics, population genetics – empirical

1 | INTRODUCTION

Riverine ecosystems are distinguished by having branching dendritic architectures, longitudinal habitat corridors and elevational gradients from headwaters to mouths that produce unidirectional downstream flow and downstream increases in habitat patch size (Paz-Vinas, Loot, Stevens, & Blanchet, 2015; Thomaz, Christie, & Knowles, 2016). Although robust and comprehensive models remain elusive, empirical and simulation studies indicate that variation in each of these riverscape attributes can influence the neutral genetic diversity and spatial population structure of stream-dwelling plants, invertebrates and fishes in a predictable fashion (Paz-Vinas

et al., 2015; Thomaz et al., 2016). For example, many riverine organisms exhibit increased downstream genetic diversity hypothesized to result from combinations of downstream-biased dispersal (drift), upstream-directed colonization (rheotaxis), and increased downstream habitat availability (Paz-Vinas et al., 2015). Due to the longitudinal nature of river channels and the lack of redundant connectivity throughout river networks, barriers – from natural rapids, waterfalls and habitat gaps to anthropogenic alterations – can exert a strong influence on the population structure of lotic organisms (Carrara, Altermatt, Rodriguez-Iturbe, & Rinaldo, 2012; Fuller, Doyle, & Strayer, 2015; Paz-Vinas & Blanchet, 2015). As watersheds across the globe become increasingly fragmented by human-constructed barriers, such as dams, roads, bridges and culverts, a better understanding of how barrier size, network position, and permeability influence the gene flow and genetic diversity of riverine organisms is needed.

North America's two largest drainage basins are the Great Lakes/Saint Lawrence and Mississippi/Missouri watersheds, which border each other and produce similar average annual discharges. Since the retreat of the last (Wisconsin) glaciation exposed the Laurentian Great Lakes beginning in the Middle to Late Pleistocene approximately 14,500 years ago (Calkin & Feenstra, 1985; Dyke, Moore, & Robertson, 2003), the aquatic biodiversity and biogeography of these basins have been intertwined through the dispersal of species away from southern, unglaciated portions of their current or historical ranges (glacial refugia). Of 91 native Great Lakes fish species examined by Bailey and Smith (1981) and Mandrak and Crossman (1992), over 80% were found to have arrived to the basin via dispersal from glacial refugia within southern parts of the modern Mississippi/Missouri watershed. Regionally, this dispersal is thought to have occurred from the west (upstream) across or around Michigan's lower peninsula (Calkin & Feenstra, 1985; Mandrak & Crossman, 1992). Biogeographical patterns suggest that the remaining 20% of native Great Lakes fish species arrived from eastern glacial refugia along the Atlantic Coast of North America via the Saint Lawrence River.

Similar patterns are observed in unionid mussels, with 37 of 39 species native to Lake Erie being shared with the Mississippi/Missouri watershed but not Lake Ontario, and 28 of 30 species native to Lake Ontario being shared with Atlantic Coastal drainages but not Lake Erie (Haag, 2012), highlighted the potential influence of Niagara Falls on aquatic species dispersal. Only two mussel species are widespread throughout these basins, and both are hypothesized to have dispersed upstream from the Atlantic Coast (Haag, 2012), probably via the Welland Canal (Hoffman, Morris, & Zanatta, 2018). Aquatic species in the Great Lakes also exhibit the distinct imprint of opposing historical invasions from western versus eastern glacial refugia in their genetic population structure. Examples include several fish species in which a distinct boundary, or intergrade and hybridization zone between sister lineages occurs in the eastern portion of the lower Great Lakes/St. Lawrence basin, where lineages originating from Atlantic versus Mississippian glacial refugia intermingle (April, Hanner, Dion-Côte, & Bernatchez, 2013; April & Turgeon, 2006).

Landscape-scale drivers of neutral population structure combined with more dynamic processes of population expansion, founder effects, and introgression are predicted to generate complicated patterns of genetic diversity in aquatic species throughout the eastern Great Lakes/St. Lawrence River basin. A better understanding of these processes and their outcomes is needed to guide aquatic species conservation efforts and management priorities throughout the basin. One site where the relative influences of neutral versus dynamic drivers of genetic structure might be clearly distinguished is Niagara Falls (Figure 1), a demonstrated barrier to aquatic species dispersal and presumed barrier to gene flow in the 58 km long Niagara River, which serves as the principal hydrologic corridor between Lake Erie (upstream) and Lake Ontario (downstream). The 58 m high Niagara Falls has interrupted this river for at least 11,000 years, although changes in the height, interconnectivity and drainage patterns of the Great Lakes have affected the falls' size, and constant erosion has caused them to retreat almost 20 km upriver from where they initially formed (Hayakawa & Matsukura, 2009; Tinkler, Pengelly, Parkins, & Asselin, 1994). Moreover, the Niagara River provides an excellent model for testing predictions specific to the unidirectional flow and longitudinal habitat corridor components of riverscape genetics theory, because the river has no significant tributaries and has similar habitat patch sizes above and below Niagara Falls, reducing the confounding effect of these additional drivers of riverscape genetic structure.

In addition to improving the understanding of how populations are structured in the Great Lakes and the Niagara River, a more precise understanding of Niagara Falls' influence on gene flow is needed to understand the role that this barrier might play in restricting the ranges of over a dozen invasive aquatic species that are present in the Great Lakes and have either already had detrimental ecological impacts or are predicted to (Mandrak & Cudmore, 2012). Four fish species (Alewife *Alosa pseudoharengus*, American Eel *Anguilla rostrata*, White Perch *Morone americana*, Sea Lamprey *Petromyzon marinus*) have dispersed upstream from Lake Ontario into Lake Erie within the last 200 years, presumably via the Erie or Welland canals that bypass Niagara Falls (Mandrak & Cudmore, 2012). Three additional species (Blueback Herring *Alosa aestivalis*, Chain Pickerel *Esox niger*, Tench *Tinca tinca*) threaten to invade via a similar pathway in the near future (Mandrak & Cudmore, 2012). Aquatic species that have already dispersed downstream from Lake Erie into Lake Ontario include Bigmouth Buffalo *Ictiobus cyprinellus*, Pink Salmon *Oncorhynchus gorbuscha*, Rainbow Smelt *Osmerus mordax*, and the Mapleleaf Mussel *Quadrula quadrula* (Hoffman et al., 2018; Mandrak & Cudmore, 2012), whereas potential future threats from upstream include Fourspine Stickleback *Apeltes quadracus*, Grass Carp *Ctenopharyngodon idella*, and Eurasian Ruffe *Gymnocephalus cernua* (Mandrak & Cudmore, 2012).

Behavioural research has failed to arrive at a consensus regarding the prevalence and direction of fish movement past Niagara Falls via the Welland Canal. For example, movement of 179 fishes representing 10 species was tracked using acoustic tags and 34 receivers throughout the canal system (Kim & Mandrak,

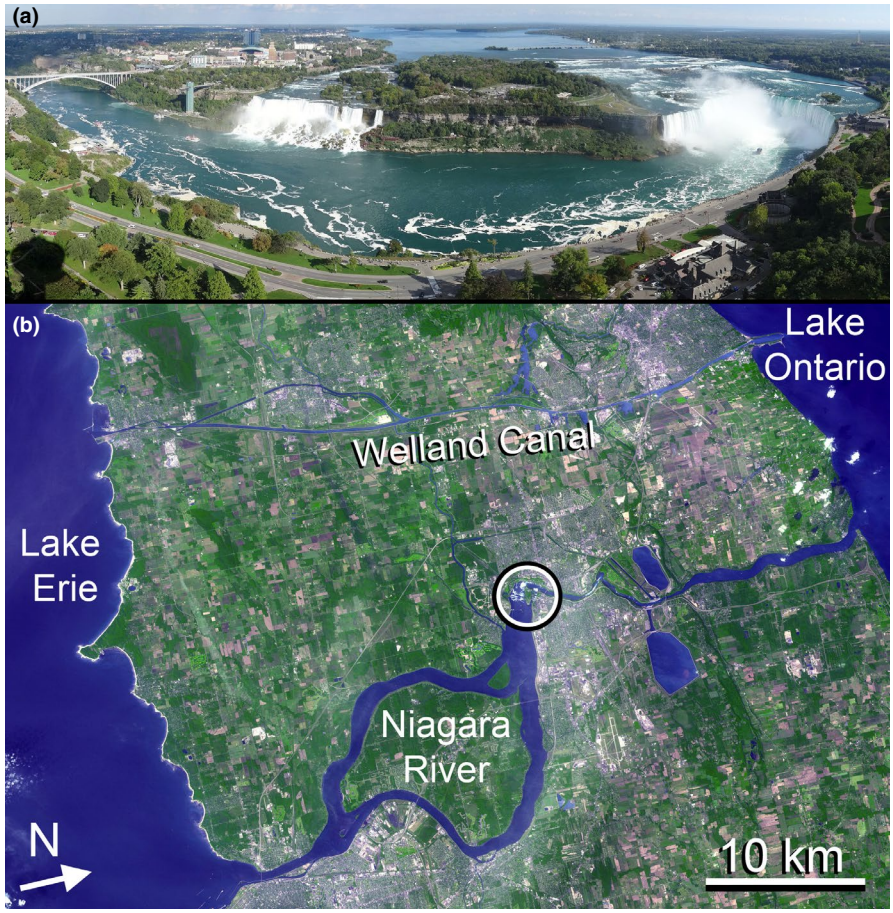

FIGURE 1 Niagara Falls as seen from the west (a), and a satellite image of the Niagara River and Welland Canal, with Niagara Falls circled (b). Photograph by Erwin Meier, satellite image by NASA's Glenn Research Center [Colour figure can be viewed at wileyonlinelibrary.com]

2016). Seven tagged fishes moved from the canal into lakes Erie or Ontario but none moved through the flight locks, a series of three closely spaced, high-lift locks in the middle of the canal system. Kim and Mandrak (2016) concluded that fishes can move through the Welland Canal and out into lakes Ontario or Erie, but the locks, particularly the flight locks, probably limit such dispersal. To better understand the direction, magnitude, and relative influence of Niagara Falls as a barrier, and the Welland Canal as a dispersal corridor, an examination of demographic trends using modern genomic approaches is needed. The need for such research is growing as governments and resource managers evaluate various barrier technologies to impede or prevent future species invasions throughout the Great Lakes.

In addition to extrinsic influences from habitat, historical contingencies, and human interventions, the genetic structure of a species and its ability to navigate barriers is affected by various intrinsic, species-specific traits, such as body size, fecundity, and vagility (Papadopoulou & Knowles, 2016). To address both the theoretical and practical questions related to the role of Niagara Falls as a barrier to gene flow in aquatic species, and to evaluate how intrinsic traits affect dispersal, we selected seven native fish species that are codistributed above and below the falls and vary in their life-history characteristics and habitat associations. For each species, we genotyped individuals from above and below Niagara Falls using a restriction-site associated DNA sequencing (RADseq) approach to obtain sequence data for anonymous loci from throughout the genome. Hundreds to thousands of unlinked single nucleotide polymorphisms (SNPs) were isolated from these loci and used to infer population structure, model gene flow magnitude and directionality, and test the hypotheses that Niagara Falls is a barrier to, and the Welland Canal is a facilitator of, dispersal in aquatic species.

We predicted that if Niagara Falls has throughout its history functioned as a barrier to gene flow in native fishes in a manner consistent with neutral theory of riverscape genetics, then populations below Niagara Falls will be genetically distinct from those above Niagara Falls, will have greater genetic diversity (heterozygosity), and will show evidence of either no migration or exclusively downstream migration past the falls. Alternatively, if the current genetic structure of Niagara River fishes is more strongly influenced by post-glacial population expansion and downstream dispersal from a presumably more genetically diverse upstream Mississippian refugium, then heterozygosity may either be greater upstream (Garner, Pearman, & Angelone, 2004; Hoban et al., 2010) or equilibrated across the falls, and there may be no evidence of population structure or gene flow with respect to Niagara Falls. Finally, if life-history characteristics function as important mediators in the interaction between neutral versus dynamic population-level processes in the Niagara River, we predicted that greater population structure would be apparent in smaller-bodied, less-vagile, less-fecund and more benthic species than in larger, more vagile, fecund, or pelagic species.

TABLE 1 Species examined in this study and their phenotypic traits, sample sizes, and descriptive results. Numbers of individuals for all taxa were equally divided between populations above and below Niagara Falls

	Ambloplites rupestris	*Ameiurus nebulosus*	*Catostomus commersonii*	*Micropterus salmoides*	*Moxostoma macrolepidotum*	*Moxostoma valenciennesi*	*Perca flavescens*
Life history characteristics							
Age at maturity[a]	2	3	4	4	3	5	4
Max age[a]	10	8	15	12	15	12	10
Calculated Generation Time[b]	5	5.5	9.5	8	8.5	8.5	7
Size at maturity (mm)	54	161	253	272	272	300	175
Fecundity[a]	5,500	7,500	20,000	5,000	33,000	40,000	72,800
Habitat	Benthopelagic	Benthic	Benthic	Benthopelagic	Benthic	Benthic	Benthopelagic
Vagility	Low	Low	High	Low	High	High	Low
Parameters and results of reduced representation library data filtering and SNP calling							
N Individuals	14	18	20	14	8	4	18
Min. # Ind./ Locus	6	6	6	6	4	4	6
N Loci for N taxa	12,104	11,155	29,421	12,728	15,513	6,988	13,956
N unlinked Biallelic SNPs	1,954	2,734	13,617	2,030	4,533	810	1,825
Genomic diversity (heterogeneity) within populations and F_{st} between populations							
H_o: above	0.40	0.35	0.29	0.36	0.38	0.45	0.27
H_e: above	0.30	0.27	0.21	0.25	0.25	0.29	0.18
HWE: above	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
H_o: below	0.40	0.34	0.34	0.38	0.41	0.48	0.28
H_e: below	0.30	0.27	0.26	0.30	0.29	0.31	0.22
HWE: below	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
H_O *p*-value: below > above	0.28	1.65	<0.001	0.01	<0.001	0.08	0.09
Hudson's F_{st}	0.104	0.055	0.359	0.213	0.253	0.195	0.339
Hudson's F_{st} 95% CI	0.094–0.113	0.050–0.060	0.354–0.363	0.202–0.224	0.245–0.260	0.166–0.226	0.324–0.352
Parameters for fastSIMCOAL2 model tests							

(Continues)

TABLE 1 (Continued)

	Ambloplites rupestris	*Ameiurus nebulosus*	*Catostomus commersonii*	*Micropterus salmoides*	*Moxostoma macrolepidotum*	*Moxostoma valenciennesi*	*Perca flavescens*
Gene copies per lineage	10	10	10	10	8	N/A	10
N SNPs w no missing data	766	2,160	5,694	815	1,243	N/A	969
N generations since 11 Ka	1,833	2,000	1,158	1,375	1,222	N/A	1,571
N generations since 186 ya	31	34	20	23	21	N/A	27

[a]Data from Portt et al. (1988) and Coker et al. (2001).

[b]Calculated as: (age of maturity + max age)/2 (Randall & Minns, 2000).

2 | MATERIALS AND METHODS

2.1 | Tissue collection

Seven native fish species (Rock Bass *Ambloplites rupestris*, Brown Bullhead *Ameiurus nebulosus*, White Sucker *Catostomus commersonii*, Largemouth Bass *Micropterus salmoides*, Shorthead Redhorse *Moxostoma macrolepidotum*, Greater Redhorse *Moxostoma valenciennesi*, and Yellow Perch *Perca flavescens*) were selected based on their differing life-history characteristics (Table 1). At least four of these species have been detected in the Welland Canal (*Ameiurus nebulosus*, *Catostomus commersonii*, *Moxostoma macrolepidotum*, *Micropterus salmoides*), although none have been observed transiting the canal between lakes (Kim & Mandrak, 2016). Between four and 20 individuals of each species (Table 1, Table S1) divided equally between populations above and below Niagara Falls were collected by the Canada Department of Fisheries and Oceans in 2014 using a boat electrofisher. All fishes were collected within the Niagara River, a 58 km long channel carrying water from Lake Erie to Lake Ontario. Distances between collecting sites above and below Niagara Falls were generally <40 river km for all species. Fishes were visually identified in the field and released alive after having a fin clip removed and preserved in 95% ethanol. Taxonomic identities were confirmed via DNA barcoding (Genbank accession numbers SUB7061508 3748–3361).

All available species and individuals were included in our analyses of genetic diversity and fixation because these analyses tolerate missing data. However, because of computational constraints and intolerance of missing data, we removed one species (*Moxostoma valenciennesi*) and reduced and amalgamated individuals within other species to test alternative demographic models (see below).

2.2 | DNA extraction and quantification

Whole genomic DNA was extracted via salt-ethanol precipitation. Small (3–4 mm^2) fin clips were digested for 3–48 hr (60°C, 600 rpm) in 502 µl cell lysis solution (410 µl extraction buffer, 80 µl 10% SDS, 10 µl 20 mg/ml proteinase K, 2 µl RNase A; stock extraction buffer: 0.5 ml 1 M Tris [ph 8], 1.0 ml 5 M NaCl, 1 ml 0.5M EDTA [ph 8], 47.5 ml ddH_2O). Cellular detritus was removed by centrifuge (5 min, 13,000 rpm), with pellet being discarded and DNA-rich supernatant decanted to tube containing 180 µl 5 M NaCl. Proteins were removed by mixing salt solution via inversion, centrifuging (10 min, 16,000 *g*), then discarding pellet and decanting DNA-rich supernatant to 420 µl chilled isopropanol. DNA was precipitated by mixing via inversion, centrifuging (15 min, 13,000 rpm), then discarding supernatant and retaining DNA pellet. DNA was washed twice with 250 µl chilled 80% EtOH followed each time by centrifugation (15 min, 13,000 rpm), then dried (open air, 2–24 hr), and resuspended in 100 µl 1X TE buffer. Extracted DNA was quantified using a Qubit flourometer (Thermo Fisher Scientific Inc.) and ranged from 115 to >35,000 ng.

2.3 | 3RAD library preparation

Whole genomic DNA was digested, ligated to barcodes and PCR-amplified with Illumina sequencing adaptors using a three-enzyme protocol developed by Glenn et al. (2017). Based on several unpublished trials with a wide range of fish taxa, the enzymes XbaI and EcoR1 were selected to digest the genome, and NheI was used to separate adapter-dimers formed by the phosphorylated adapter, increasing efficiency of adapter ligation to sample DNA (Glenn et al., 2017). One hundred ng of gDNA from each sample was dried to a pellet in a 96-well plate via SpeedVac centrifuge (Thermo Fisher Scientific Inc.). A 15 μl digestion master mix (1.5 μl NEB 10× CutSmart Buffer (New England Biolabs), 10.0 μl dH_2O, 0.5 μl XbaI, 0.5 μl EcoR1, 0.5 μl NheI) was added to each well, followed by 1 ng each of a unique combination of enzyme-cut-site-specific iTru adapters. This plate was sealed using adhesive foil then incubated (37°C, ≥1 hr) to fully digest the genome. Immediately after incubation, a 5 μl ligation master mix (2.0 μl dH_2O, 1.5 μl 10 mM rATP, 0.5 μl 10x ligase buffer, 1.0 μl 0.25× (400 units/μl) DNA ligase) was added to each sample, and adapters were ligated using a thermocycler (22°C for 20 min, 37°C for 10 min, repeat 22°C and 37°C steps once, 80°C for 20 min, hold at 10°C).

Following enzymatic digestion and adapter ligation, libraries underwent bead cleanup. Sixty μl of SpeedBeads (Sigma-Aldrich) were added to each well, the plate was allowed to sit for 10 min then placed on a magnet. After magnetic bead isolation, all fluid was removed and beads were twice washed with 100 μl of 70% ethanol. Following removal of last ethanol wash, libraries were resuspended in 20 μl of IDTE and plate was vortexed to aid bead resuspension. A total of 10 μl of ligation product was aliquoted to a new plate, and a 15 μl PCR master mix was added to each well (5 μl 5× Kapa HiFi Buffer (Sigma-Aldrich), 3.75 μl dH_2O, 0.75 μl Kapa dNTPs (2.5 mM each), 0.50 μl Kapa HiFi DNA polymerase). Two and a half ng each of a unique combination of i5 and i7 barcode primers was added to each well, and libraries with barcodes were amplified via using a thermocycler (95°C for 2 min, 16 cycles of: 98°C for 20 s, 55°C for 15 s, 72°C for 30 s; 72°C for 5 min.; hold at 15°C).

2.4 | Library visualization, pooling and size selection

Each library was visualized by running 2.5 μl on a 1% agarose gel (30 min, 175 v). A total of 10 μl of every library that successfully amplified, as indicated by broad, bright smears on the gel, was combined in pools of 20–25 samples each without regard to species or population identity. The DNA concentration of each pool was quantified using a Qubit flourometer to ensure that it did not exceed binding capacity of a QIAquick PCR purification column (10,000 ng) (Qiagen Inc.). Each pool was then cleaned using a QIAquick PCR purification kit following manufacturer's instructions. A size range of fragments from 340–405 bp was isolated from each library using a Pippin Prep automated size selection instrument (Sage Science, Inc.). By isolating relatively large fragment sizes, we ensured that a manageable number of loci were retained and that the relatively short (~60 bp) sequencing reads obtained from each locus would be separated from all other loci by at least ~280 bp, thus helping to meet the assumption of downstream coalescent-based analyses that loci be unlinked.

2.5 | Sequencing

DNA concentrations of the size-selected 3RAD library pools were quantified using Qubit flourometry and qPCR. Based on an averaged estimate of pool concentration, a standardized volume comprising multiple library pools was loaded on an Illumina NextSeq 550 high-throughput benchtop sequencer (Illumina Inc.) at either the National Center for Natural Products Research in Oxford, MS, or the University of Mississippi Medical Center in Jackson, MS. All pools were sequenced for 75 bp single-end reads using the NextSeq 500/550 High Output 75 cycle kit, which distributes reads across four sequencing flow cell lanes without the option of isolating libraries into individual lanes. Libraries were sequenced one to three times as needed until at least two million demultiplexed reads were obtained for all but a few samples (Table S1). Experiments performed by us on a wide range of fish taxa (unpublished), indicate that for libraries prepared as above, the number of final loci clustered across individuals plateaus at less than two million demultiplexed reads. Data from all sequencing efforts were uploaded to the Illumina BaseSpace Sequence Hub.

2.6 | Demultiplexing and trimming

All sequence data were analysed on the Sequoia supercomputer at the Mississippi Center for Supercomputing Research (MCSR). Data were downloaded using the BaseSpace run downloader and demultiplexed using the Illumina bcl2fastq conversion software programmed to allow for up to two barcode mismatches (--barcode-mismatches 2). Three bp degeneracy was explicitly designed into the barcode sets utilized so that barcodes could have as many as two sequencing errors and still be uniquely assignable to only a single sample source (T. Glenn, personal communication). After demultiplexing, reads from all four sequencing lanes were concatenated and reads for individual samples were combined across runs. All reads were trimmed to a standard 62 base length by removing barcodes and restriction overhangs from the 5′ portion of each read and a compensatory number of bases from the 3′ end using the program FASTQ/A Trimmer within the FASTX-Toolkit (Gordon & Hannon, 2010).

2.7 | Read clustering and SNP calling

The program PYRAD v3 (Eaton, 2014) was used to cluster reads both within and across samples and to identify single nucleotide

polymorphisms (SNPs). To maximize processing efficiency, demultiplexed reads for each sample were run through PYRAD Steps 2–4 (filtering, within-sample clustering, and error rate and heterozygosity estimates) on separate processors allocated 3 GB of memory each. Results of these sample-specific within-group clustering analyses were combined into species-specific groupings for PYRAD steps 5–7 (consensus sequence creation, across-sample clustering, alignment, paralog filtering and output generation), which were run on a single processor allocated 6 GB of memory. The following parameter values were used for all pyRAD analyses: Minimum coverage for a cluster = 10, maximum number of sites with Phred quality score <20 = 5, clustering threshold = 95%, maximum number of individuals with a shared heterozygous site = 75%, maximum number of heterozygous sites in a consensus sequence = 2, maximum number of SNPs per locus = 2 (latter parameters set stringently to exclude potential paralogous loci). For loci with multiple SNPs, a single SNP was drawn at random for downstream analyses.

2.8 | Measuring genetic diversity

Observed (H_o) and expected (H_e) heterozygosities were calculated for each locus and each population above versus. below Niagara Falls using the divBasic function in the R package diveRsity (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013). Overall H_o and H_e were compared within each population and H_o was compared between populations above versus. below Niagara Falls with a two-sample *t* test assuming unequal variance implemented in the Analysis ToolPak in Excel (Microsoft Inc., v16.16.3).

2.9 | Measuring genomic structure and dissimilarity

Intraspecific genomic structure spanning the Niagara River above and below Niagara Falls was visualized by principal coordinates analyses of one SNP per locus from all individuals within a given species. Unphased SNP matrices were analysed using the PCoA function in the program PAST v3 (Hammer, Harper, & Ryan, 2001). Genomic differentiation was measured using Hudson's F_{st} calculated following the equation provided in Bhatia, Patterson, Sankararaman, and Price (2013) using an R script from Barrera-Guzman, Aleixo, Shawkey, and Weir (2018). Bhatia et al. (2013) found that the historically most commonly used fixation indices (i.e., Nei, 1973; Weir & Cockerham, 1984), which take the average of ratios calculated for individual markers, are sensitive to the ratio of sample sizes and can result in large reductions in F_{st} estimates calculated from genomic data. Bhatia et al. (2013) therefore recommended using Hudson's F_{st}, which instead takes the ratio of the average of values calculated for individual markers. Significance of each F_{st} estimate was determined by calculating 95% confidence intervals from a 1,000 replicate bootstrap.

2.10 | Testing demographic models

We used FASTSIMCOAL2 v2.5.2.1 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013) to calculate the likelihood of the observed site frequency spectra under seven alternative demographic models in six of the seven species (excluding *Moxostoma valenciennesi* due to too few individuals), and we compared these models using AIC. FASTSIMCOAL2 requires complete SNP matrices without missing data, which we generated using an R script (Barrera-Guzman et al., 2018) to downsample lineages so that at least eight (*Moxostoma macrolepidotum*) or 10 (*Ambloplites rupestris*, *Ameiurus nebulosus*, *Catostomus commersoni*, *Micropterus salmoides*, *Perca flavescens*) gene copies per lineage were retained at each SNP (Table 1). SNPs that did not possess these numbers of gene copies were discarded. Resulting data sets had between 766 and 5,674 SNPs with no missing data (Table 1). These data sets were used to generate observed, multidimensional site frequency spectra (SFS) for the minor allele at each SNP using Arlequin (Excoffier & Lischer, 2010). Although individuals were modelled as haploid (1n), estimated population sizes and numbers of migrants reported throughout our results have been converted to diploid (2n).

All models had two lineages diverging 11,000 years ago (estimated age of Niagara Falls), with one lineage each representing populations above and below Niagara Falls and numbers of generations varying according to the calculated generation time for each species (Table 1). Models varied only in terms of the timing and direction of migration between populations following divergence. Model 1 assumed no migration after divergence. Models 2–4 assumed that migration occurred continuously since divergence, with the direction of migration varying from unidirectional upstream (Model 2), to unidirectional downstream (Model 3), to bidirectional (Model 4). Finally, Models 5–7 restricted migration to only the 186 years since the first iteration of the Welland Canal was completed in 1829. These models had the same variation in migration directionality (i.e., unidirectional upstream = Model 5, unidirectional downstream = Model 6, bidirectional = Model 7).

Models were tested using an approach similar to Barrera-Guzman et al. (2018) in which generation lengths were estimated a priori. Generation length is defined by the IUCN as 'the average age of parents of the current cohort (i.e., newborn individuals in the population),' which is difficult to measure in contemporary populations, let alone historical. Thus, to evaluate effects of generation length error, all models were run twice, once with a probably underestimated generation length of three years for all species, and once with probably overestimated generation lengths calculated independently for each species from Ontario-specific life-history data in Portt, Minns, and King (1988) and Coker, Portt, and Minns (2001) (Table 1). Following IUCN guidelines, generation lengths were estimated as the mean age at which a species reproduces (IUCN, 2019). Assuming that most fishes reproduce throughout adulthood, generation length was calculated as: (age at maturity + max age)/2, yielding generations lengths for species

TABLE 2 Results of all supported coalescent models comparing aspects of gene flow past Niagara Falls in six native fish species. Evaluated models varied by assuming either no migration since Niagara Falls' formation, unidirectional upstream migration since Niagara Falls' formation, unidirectional downstream migration since Niagara Falls' formation, bidirectional migration since Niagara Falls' formation, unidirectional upstream migration since first completion of the Welland Canal, unidirectional downstream migration since first completion of the Welland Canal, or bidirectional migration since first completion of the Welland Canal. Best supported model results in bold type. Migrants were modelled as haploid individuals in the program fastSIMCOAL2 (Excoffier et al., 2013) using demographic parameters in Table 1. All results reported as diploid. For complete results of all models, including a comparison with models assuming a shorter generation time, see Table S2

			Estimated population sizes						Diploid migrants per generation			
Log(10) likelihood	delta AIC	Akaike weight	Ancestral	95% CI	DS	95% CI	US	95% CI	DS	95% CI	US	95% CI
Ambloplites rupestris												
No migration												
−1,159.2811	**0.000**	**0.348**	**838,015**	**500,041–864,059**	**120,782**	**39,716–301,387**	**67,412**	**46,219–509,719**				
Post-glacial unidirectional upstream migration												
−1,159.2455	1.836	0.139	735,865		197,507		52,613				8.9898	
Ameiurus nebulosus												
Post-glacial unidirectional upstream migration												
−3,288.9506	**0.000**	**0.438**	**425,426**	**432,405–864,082**	**48,084**	**35,807–76,174**	**46,117**	**30,030–64,041**			**1.9003**	**0.0048–3.0433**
Catostomus commersoni												
No migration												
−8,335.5052	**0.000**	**0.256**	**566,030**	**116,009–698,478**	**24,517**	**21,915–32,965**	**58,922**	**35,905–73,052**				
Post-glacial unidirectional upstream migration												
−8,335.3049	1.077	0.149	260,177		25,823		48,803				0.0735	
Post-glacial unidirectional downstream migration												
−8,335.2203	0.688	0.181	206,135		42,980		25,012		0.9539			
Welland Canal unidirectional upstream migration												
−8335.3888	1.464	0.123	277,607		25,160		50,170				0.0240	
Welland Canal unidirectional downstream migration												
−8,335.3737	1.394	0.127	316,608		45,126		25,472		0.1000			
Welland Canal bidirectional migration												
−8,335.0479	1.894	0.099	463,446		49,829		27,721		0.0286		0.0024	
Micropterus salmoides												
No migration												
−1,258.3327	**0.000**	**0.439**	**641,150**	**343,545–840,847**	**61,635**	**26,371–145,947**	**24,422**	**17,852–49,669**				
Post-glacial unidirectional upstream migration												
−1,258.0961	0.910	0.279	832,123		45,189		24,783				1.8624	

(Continues)

TABLE 2 (Continued)

Log(10) likelihood	delta AIC	Akaike weight	Estimated population sizes						Diploid migrants per generation			
			Ancestral	95% CI	DS	95% CI	US	95% CI	DS	95% CI	US	95% CI
Welland Canal unidirectional upstream migration												
−1,258.2786	1.751	0.183	739,915		51,480		23,168				0.1831	
Welland Canal unidirectional downstream migration												
−1,258.0334	0.622	0.322	331,626		430		812		0.0010			
Moxostoma macrolepidotum												
No migration												
−1,609.9001	**0.000**	**0.369**	**155,205**	**39,996–749,179**	**90,191**	**44,534–653,231**	**46,227**	**23,29–142,491**				
Post-glacial unidirectional upstream migration												
−1,609.8581	1.807	0.150	105,186		110,960		45,248				0.0485	
Perca flavescens												
Post-glacial unidirectional upstream migration												
−1,067.4489	1.949	0.197	5,261		46,294		47,958				2.1810	
Post-glacial bidirectional migration												
−1,066.5914	**0.000**	**0.522**	**4,917**	**3,913–6,324**	**38,816**	**26,367–82,479**	**35,982**	**25,650–59,153**	**1.6163**	**1.2038–2.6770**	**1.6937**	**1.2100–3.7386**

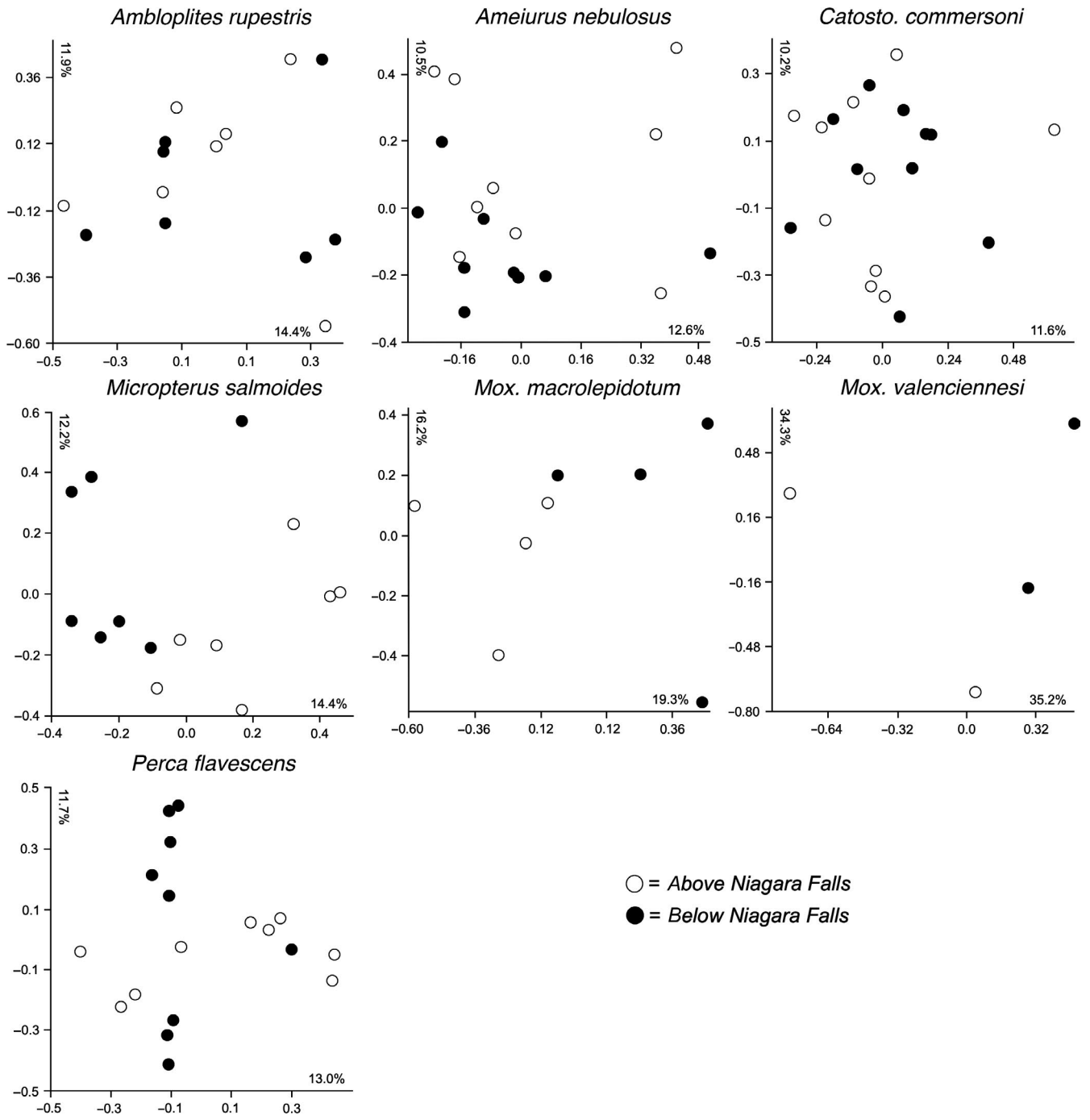

FIGURE 2 Coordinates one (*x*-axis) and two (*y*-axis) of principal coordinates analyses of genetic variation in seven native fish species collected from the Niagara River upstream (open circle) and downstream (closed circle) of Niagara Falls. Percentages of variation explained by each coordinate given inside the axes

in this study ranging from five to 9.5 years (Table 1). Although we present results of both sets of models as Supporting Information (Table S2), the results were generally similar so we focus our discussion on only the second set of models using generation lengths calculated from empirical data.

All models estimated an effective population size for each lineage and for the common ancestor lineage prior to their divergence (three parameters). A total of 120 independent runs were obtained in fastSIMCOAL2, each starting from different randomly drawn starting parameters. Parameters were drawn from log-uniform distributions ranging from 100 to 1,000,000 for effective population size parameters, and 1×10^{-8} to 1×10^{-4} for per capita migration rate parameters. Each independent run used 50 EMC loops and obtained the fit of the observed data using 200,000 simulated SFS spectra. The run and its associated parameter values with the best likelihood were retained for each model. Each run was executed on a separate core of a Lenovo SD530 server at 2.4 Ghz on the Niagara supercomputing cluster (Ponce et al., 2019).

We calculated 95% confidence intervals for parameters of the best fit model for each species using 100 nonparametric bootstrap replicates. For each replicate a bootstrapped SFS was generated and the maximum likelihood parameter values were approximate from the best of 100 independent runs in FASTSIMCOAL2, each starting from a different randomly drawn set of parameters. The 95% confidence intervals were calculated as the 0.025 and 0.975 quantiles of the resulting 100 bootstrap values for each parameter. Only viable model results are reported herein (Table 2), although a comparison of all results from all models is presented in Table S2.

3 | RESULTS

3.1 | Genomic data obtained

We obtained a total of 378,695,994 demultiplexed reads comprising an average of 3,944,750 demultiplexed reads per individual (±2,243,293 *SD*; Table S1). Numbers of loci obtained after within- and between-group clustering varied from 6,988 (*Moxostoma valenciennesi*) to 29,421 (*Catostomus commersonii*), and numbers of "unlinked" biallelic SNPs varied from 810 (*Moxostoma valenciennesi*) to 13,617 (*Catostomus commersonii*; Table 1).

3.2 | Patterns of genetic diversity

In all populations and species, observed heterozygosity (H_o) was significantly greater than heterozygosity expected based on Hardy-Weinberg assumptions (H_e; Table 1). In all highly vagile taxa (*Catostomus commersoni*, *Moxostoma macrolepidotum*, *M. valenciennesi*) and two low vagility taxa (*Micropterus salmoides*, *Perca flavescens*), H_o was significantly ($p \leq .01$; $n = 3$; or close to significantly: $p < .1$; $n = 2$) greater in populations below versus above Niagara Falls. In the two low-vagility taxa *Ambloplites rupestris* and *Ameiurus nebulosus*, there was no significant difference in H_o above versus below Niagara Falls ($p > .2$; Table 1).

3.3 | Patterns of genomic structure

Only three species (*Micropterus salmoides*, *Moxostoma macrolepidotum*, *Moxostoma valenciennesi*) clearly partitioned bivariate genomic ordination space (Figure 2). In all four other taxa, distributions of populations above and below Niagara Falls largely overlapped in the first two coordinates of genomic ordination space; however, the total percent variation explained by these coordinate axes was lower in species that did not partition ordination space (range: 21.8%–26.3%) than in those that did (range: 26.6%–69.5%; Figure 2). Despite the broad overlap in genomic ordination space in four species, bootstrapping of the Hudson's fixation index (F_{st}) found that populations of all taxa were significantly differentiated above versus below Niagara Falls. Hudson's F_{st} values varied over 6.5 fold, from 0.055 (0.050–0.060 95% CI) for *Ameiurus nebulosus* to 0.359 (0.354–0.363 95% CI) for *Catostomus commersonii* (Table 1).

3.4 | Testing models of continuous post-glacial migration past Niagara Falls

The model assuming no migration past Niagara Falls was the best supported model (highest Akaike weight) for *Ambloplites rupestris*, *Catostomus commersonii*, *Micropterus salmoides* and *Moxostoma macrolepidotum* (Table 2), but was rejected (delta AIC > 2) in both remaining taxa (*Ameiurus nebulosus* and *Perca flavescens*) (Table S2). In three of the four taxa for which the no-migration model was best supported (*Ambloplites rupestris*, *Micropterus salmoides* and *Moxostoma macrolepidotum*), upstream-only models were strong alternative contenders (Table 2), with downstream-only and bidirectional migration models being rejected (Table S2). For the fourth taxon (*Catostomus commersoni*), both upstream-only and downstream-only models were strong alternative contenders (Table 2), but bidirectional migration was rejected (Table S2). Upstream-only migration was the only model supported in *Ameiurus nebulosus* (Table 2). For *Perca flavescens*, bidirectional migration received the strongest support, with upstream only migration being a strong alternative contender (Table 2).

The highest estimated number of migrants per generation in a supported model was in the upstream-only migration model for *Ambloplites rupestris* ($n = 17.976$; Table 2). For all other taxa and migration models, estimated numbers of migrants per generation past Niagara Falls did not exceed 3.4667 (Table S2). Excluding *Ambloplites rupestris* and Welland Canal models, estimated numbers of migrants per generation in supported migration models ranged from 0.0485 and 0.0735 (upstream-only migration models for *Catostomus commersoni* and *Moxostoma macrolepidotum*; Table 2) to 1.8624 and 1.9003 (upstream-only migration models for *Micropterus salmoides* and *Ameiurus nebulosus* respectively; Table 2).

3.5 | Testing models of Welland Canal migration past Niagara Falls

All models in which migration was restricted to years since 1829, when the first Welland Canal was completed, were rejected in four of the six taxa examined (*Ambloplites rupestris*, *Ameiurus nebulosus*, *Moxostoma macrolepidotum*, and *Perca flavescens*). In both taxa for which Welland Canal migration models were supported (*Catostomus commersoni* and *Micropterus salmoides*), support for these models was much less than for the no-migration model, and generally less than models in which migration was allowed to occur continuously since glacial retreat. Only in *Micropterus salmoides* did a Welland Canal migration model (unidirectional downstream) receive greater support than the only supported model of continuous post-glacial migration, although the difference was modest (0.322 vs. 0.279

Akaike weight; Table 2) and both models received less support than the no migration model (Table 2).

3.6 | Effects of different generation time estimates

A comparison of model results generated using probably underestimated versus probably overestimated generation length estimates is presented in Table S2. Results of the alternative model sets are generally similar, with the no migration model being most consistently well supported in both sets, and the Welland Canal models being generally rejected in both sets. Only in *Micropterus salmoides* was the no-migration model supported in the second set of model tests, but rejected in the first. In the first set of model tests for *Micropterus salmoides*, a bidirectional model of continuous post-glacial migration was supported most strongly, although the unidirectional upstream migration model was also supported in both sets.

4 | DISCUSSION

Over a distance of <40 river km spanning Niagara Falls (Figure 1), we observed significant genomic differentiation in seven native Great Lakes fish species, evidence for no migration past the falls in four species, yet the likelihood of either unidirectional upstream or bidirectional migration past the falls in two species, yielding new insights into how Niagara Falls has influenced aquatic community population structure and gene flow within the Niagara River. Observed patterns of fish population structure and gene flow contribute to the understanding of how fundamental riverscape features, such as unidirectional water flow or restricted habitat connectivity, can affect the spatial distribution of genetic diversity. Our results also help to explain the role that natural riverine barriers in general, and Niagara Falls specifically, play in creating and maintaining lotic biodiversity, and the role that artificial navigation channels bypassing natural barriers might play in facilitating gene flow. Interspecific variation in population genetics patterns help to explain how morphological, behavioural and life-history traits might affect gene flow in fishes. An immediate, practical benefit of these results is a better understanding of how Niagara Falls, and the neighboring Welland Canal, may affect non-native aquatic species dispersal, which will help guide the development and implementation of strategies to prevent or limit such invasions.

4.1 | Historical versus contemporary drivers of genetic diversity in the Niagara River

The population genetics of many aquatic and terrestrial plant and animal species occupying post-glacial landscapes exhibit well-documented spatial gradients in which concentrations of high genetic diversity in refugial (unglaciated) portions of species' ranges gradually transition to areas of lower genetic diversity in more recently colonized, recently deglaciated portions of species' ranges (Elderkin, Christhian, Vaughn, Metcalfe-Smith, & Berg, 2007; Garner et al., 2004; Hewitt, 2000; Hewitt, Bergner, Woolnough, & Zanatta, 2018; Hoban et al., 2010; Mathias, Hoffman, Wilson, & Zanatta, 2018). Biogeographical patterns among most native aquatic species occupying the Great Lakes indicate that, after glaciers retreated in the Late Pleistocene, the lakes were colonized in a generally upstream-to-downstream or west-to-east direction from the upper Mississippi River basin (Bailey & Smith, 1981; Hewitt et al., 2018; Mandrak & Crossman, 1992; Mathias et al., 2018). Although hydrologic connections among lakes varied through time, the Niagara River is hypothesized to have been the principal dispersal corridor between lakes Erie and Ontario since early in their deglaciated history (Calkin & Feenstra, 1985), thus we predicted that an influence of these historical demographic processes might be detected in fish populations upstream from Niagara Falls having higher genetic diversity.

In contrast to our prediction based on these historical processes, we found that three of the seven examined fish species had significantly greater ($p \leq .01$) genetic diversity downstream of Niagara Falls, two of the seven showed a nonsignificant trend toward greater downstream genetic diversity (p: .08, .09), and populations of two species showed no difference in heterozygosity upstream versus downstream of the falls. Thus, these data suggest that more recent or ongoing contemporary processes have had a greater effect on population genetic structure in the Niagara River than historical processes, a pattern that is also seen in other post-glacial colonizing species and in theoretical models (Vucetich & White, 2003). All prior studies that have observed genetic diversity gradients along post-glacial dispersal corridors (e.g., April & Turgeon, 2006; April et al., 2013; Ginson, Walter, Mandrak, Beneteau, & Heath, 2015; Hoffman et al., 2018) have examined more individuals than this study, and spanned much larger distances, across which genetic differences are predicted to be greater and more easily detected. However, we examined far more loci (thousands vs. <10), which should give our study greater sensitivity to differences across shorter distances.

For fishes in the Niagara River, we hypothesized that greater downstream heterozygosity might be a product of fundamental riverscape habitat features that previous empirical and modelling studies have found to have a similar effect on other lotic plants and animals (Paz-Vinas et al., 2015; Thomaz et al., 2016). Of the various riverine habitat features invoked by these studies (e.g., branching dendritic architectures, limited connectivity due to longitudinal habitat corridors, unidirectional downstream flow, downstream increases in habitat patch size), only unidirectional downstream flow and limited connectivity are particularly applicable to the Niagara River. This river does not branch and it has few tributaries, so habitat patch sizes are similar above and below Niagara Falls, but unidirectional downstream flow and hydrologic connectivity limited to the main channel have probably been defining features of the Niagara River since its formation, with a potential parallel corridor only being opened in 1829 upon first completion of the Welland Canal.

However, effects of habitat features on the genetic diversity of lotic species is contingent upon some degree of genetic connectivity between upstream and downstream populations (Thomaz et al., 2016), which we found no consistent evidence of in the Niagara River. An alternative explanation for the observed pattern is that introgression from extralimital downstream populations, possibly including the Atlantic Coast glacial refugium (April et al., 2013; April & Turgeon, 2006), has elevated downstream genetic diversity with little or no effect on upstream genetic diversity.

4.2 | Contributions of Niagara Falls to the creation and maintenance of riverine biodiversity

Natural in-stream barriers to dispersal and gene flow in lotic organisms can take different forms, from transitions in water physicochemistry (Winemiller, López Fernández, Taphorn, Nico, & Barbarino Duque, 2008) to geomorphic barriers such as rapids and waterfalls (Wofford, Gresswell, & Banks, 2005), with effects that can vary from partial to total limitation of population distributions and gene flow (Winemiller et al., 2008; Wofford et al., 2005). In the Great Lakes, Niagara Falls is known to limit the distributions of many mussel species (Haag, 2012). Similar patterns are documented in fish species near similarly large waterfalls in unglaciated parts of North America (Keck & Near, 2013) and the tropics (Lujan, Agudelo-Zamora, Taphorn, Booth, & López-Fernández, 2013; Lujan, Armbruster, & Lovejoy, 2018). Given the enormous size of Niagara Falls, it is unsurprising that the model assuming no migration past these falls was the best supported in four of the six species examined (Table 2). More surprising is that this model was rejected in two species (*Ameiurus nebulosus*, *Perca flavescens*), and an alternative model of unidirectional upstream dispersal past Niagara Falls was the only model to be supported in all six species (albeit second-ranked in five species; Table 2). Moreover, one hypothesis to explain upstream dispersal past Niagara Falls – that fishes dispersed via the artificial Welland Canal – was rejected in four species and received intermediate to weak support in the remaining two (Table 2).

Significantly higher observed versus. expected heterozygosities in both upstream and downstream populations of all seven examined species (Table 1) is consistent with the hypothesis that both have received admixture from other populations in the relatively recent past, although not necessarily from each other. To better understand the origins of these patterns, more research is needed on historical biogeographical processes at broader spatial and taxonomic scales. For example, using allozymes and the mitochondrial gene COI in a widespread study of the unionid mussel *Amblema plicata* throughout eastern North America, Elderkin et al. (2007) found evidence for a previously unknown historical dispersal corridor between the Ohio River basin and Lake Erie. However, the need for such large-scale research should not diminish the importance of more focused studies such as this. Indeed, data sets spanning large spatial scales can sometimes obscure the effects of migration barriers because of the confounding influence of isolation by distance (Frantz, Cellina, Krier, Schley, & Burke, 2009).

4.3 | Interspecifically variable phenotypic traits that may affect genetic structure

The genetic structure of a species and the spatial scale at which it exchanges DNA can be affected by a wide range of intrinsic traits, such as body size, fecundity, and vagility (Papadopoulou & Knowles, 2016). Species examined in this study were approximately evenly divided between high and low vagility, varied in body size 5.6-fold (from a mature body size of approximately 54 mm in *Ambloplites rupestris* to 300 mm in *Moxostoma valenciennesi*) and varied in fecundity more than 13-fold (from approximately 5,500 eggs in *Ambloplites rupestris* to 72,800 eggs in *Perca flavescens*; Table 1).

There was no obvious or direct correspondence between phenotypic traits and population genetics patterns, but there were some counterintuitive trends. For example, despite *Ameiurus nebulosus* being a relatively low vagility, low fecundity species, it was the only species having the unidirectional upstream migration model be best supported. *Ameiurus nebulosus* and *Ambloplites rupestris*, the latter of which had the unidirectional upstream migration model be ranked second after no migration, were also species with the lowest average F_{st} values (*Ambloplites rupestris*: 0.104, *Ameiurus nebulosus*: 0.055), indicating lower (but still significant) genetic distinctiveness of populations above versus. below Niagara Falls.

4.4 | Insights into the threat that future invasive species will disperse past Niagara Falls

Our results indicate that Niagara Falls has probably functioned as a barrier to both upstream and downstream dispersal of native fishes. Our coalescent models reject hypotheses of post-glacial bidirectional and downstream-only migration in four of six examined species (*Ambloplites rupestris*, *Ameiurus nebulosus*, *Micropterus salmoides*, *Moxostoma macrolepidotum*; Table 2). Moreover, they predict very few (≤1.6163) downstream migrants per generation in the two taxa for which downstream or bidirectional models of continuous post-glacial migration were supported (0.9539 migrants per generation in the unidirectional downstream model for *Catostomus commersoni*, 1.6163 downstream migrants per generation in the bidirectional model for *Perca flavescens*, Table 2). Of these two, the no-migration model was the best supported model for *Catostomus commersoni* but was rejected in *Perca flavescens*.

Our results indicate that the Welland Canal is an insignificant dispersal corridor for native fishes. Welland Canal migration models were rejected in four of the six species examined (*Ambloplites rupestris*, *Ameiurus nebulosus*, *Moxostoma macrolepidotum*, *Perca flavescens*) and were not the best supported models in either of the other two species (*Catostomus commersoni*, *Micropterus salmoides*). In these two species and the Welland Canal migration models that

were supported, estimated numbers of migrants did not exceed 0.1 individuals per generation.

In conclusion, Niagara Falls has probably functioned as a barrier to migration in the Niagara River for much of its history. Demographic models assuming no migration past Niagara Falls were best-supported in four of six species tested. Two remaining species showed strongest support for either unidirectional upstream migration (*Ameiurus nebulosus*), or bidirectional migration (*Perca flavescens*) exclusive of the Welland Canal. Height of Niagara Falls since its formation makes it incomprehensible that any fish could surmount them directly, yet our model tests broadly reject a role for the artificial Welland Canal in facilitating dispersal past the falls. Thus, data from across a broader spatial scale are needed to test for still unknown historical dispersal corridors that may have facilitated upstream migration. Technologies to limit fish movement through the Welland Canal are unlikely to impact the existing population structure of native fishes.

ACKNOWLEDGEMENTS

We thank the Canadian Department of Fisheries and Oceans for supporting this research financially and through the provision of tissue samples, and the Natural Sciences and Engineering Research Council of Canada for funding (NSERC Discovery Grants: RGPIN-2014-05226 to NEM, RGPIN-2016-06221 to NRL, RGPIN-2016-06538 to JTW; NSERC Discovery Accelerator Grant 492890 to JTW). We also thank Arsalan Ahmad for extracting DNA from most of the samples in this study. Computations were performed on the Niagara supercomputer at the SciNet High Performance Computing Consortium, funded by the University of Toronto, the Canadian Foundation for Innovation, the Government of Ontario, and Ontario Research Fund – Research Excellence.

AUTHOR CONTRIBUTIONS

All authors contributed to the design of the study. NKL and BPN collected the data, NKL and JTW analyzed the data, NKL led manuscript preparation with input from all coauthors.

DATA AVAILABILITY STATEMENT

DNA barcode data (cytochrome oxidase I; COI sequence data) for specimens examined in this study can be obtained from Genbank via submission number SUB7061508. Pyrad output (VCF files, unlinked SNP matrices), PCoA input files (.xls), diveRsity input files (Genepop format), and fastSIMCOAL2 model files can be obtained from Dryad via https://doi.org/10.5061/dryad.wdbrv15jn

ORCID

Nathan K. Lujan https://orcid.org/0000-0002-9022-1506

REFERENCES

April, J., Hanner, R. H., Dion-Côte, A. M., & Bernatchez, L. (2013). Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Molecular Ecology*, *22*, 409–422. https://doi.org/10.1111/mec.12116

April, J., & Turgeon, J. (2006). Phylogeography of the banded killifish (*Fundulus diaphanus*): Glacial races and secondary contact. *Journal of Fish Biology*, *69*, 212–228. https://doi.org/10.1111/j.1095-8649.2006.01233.x

Bailey, R. M., & Smith, G. R. (1981). Origin and geography of the fish fauna of the Laurentian Great Lakes basin. *Canadian Journal of Fisheries and Aquatic Sciences*, *38*, 1539–1561. https://doi.org/10.1139/f81-206

Barrera-Guzman, A. O., Aleixo, A., Shawkey, M. D., & Weir, J. T. (2018). Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E218–E225. https://doi.org/10.1073/pnas.1717319115

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F_{st}: The impact of rare variants. *Genome Research*, *23*, 1514–1521. https://doi.org/10.1101/gr.154831.113

Calkin, P. E., & Feenstra, B. H. (1985). Evolution of the Erie-Basin Great Lakes. In P. F. Karrow, & P. E. Calkin (Eds.), *Quaternary evolution of the Great Lakes* (pp. 149–170). St. John's, NL: Geological Association of Canada Special Paper 30.

Carrara, F., Altermatt, F., Rodriguez-Iturbe, I., & Rinaldo, A. (2012). Dendritic connectivity controls biodiversity patterns in experimental metacommunities. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 5761–5766. https://doi.org/10.1073/pnas.1119651109

Coker, G. A., Portt, C. B., & Minns, C. K. (2001). *Morphological and ecological characteristics of Canadian freshwater fishes*. Burlington, ON: Fisheries and Oceans Canada.

Dyke, A. S., Moore, A., & Robertson, L. (2003). *Deglaciation of North America*. Ottawa, ON: Geological Survey of Canada.

Eaton, D. A. R. (2014). PyRAD: Assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*, 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Elderkin, C. L., Christhian, A. D., Vaughn, C. C., Metcalfe-Smith, J. L., & Berg, D. J. (2007). Population genetics of the freshwater mussel, *Amblema plicata* (Say 1817) (Bivalvia: Unionidae): Evidence of high dispersal and post-glacial colonization. *Conservation Genetics*, *8*, 355–372. https://doi.org/10.1007/s10592-006-9175-0

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLOS Genetics*, *9*, e1003905. https://doi.org/10.1371/journal.pgen.1003905

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite v3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? *Journal of Applied Ecology*, *46*, 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Fuller, M. R., Doyle, M. W., & Strayer, D. L. (2015). Causes and consequences of habitat fragmentation in river networks. *Annals of the New York Academy of Sciences*, *1355*, 31–51.

Garner, T. W. J., Pearman, P. B., & Angelone, S. (2004). Genetic diversity across a vertebrate species' range: A test of the central-peripheral hypothesis. *Molecular Ecology*, *13*, 1047–1053. https://doi.org/10.1111/j.1365-294X.2004.02119.x

Ginson, R., Walter, R. P., Mandrak, N. E., Beneteau, C. L., & Heath, D. D. (2015). Hierarchical analysis of genetic structure in the habitat-specialist Eastern Sand Darter (Ammocrypta pellucida). *Ecology and Evolution*, *5*, 695–708.

Glenn, T. C., Bayona-Vásquez, N. J., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., … Faircloth, B. C. (2017). Adapterama III: Quadruple-indexed, triple-enzyme RADseq libraries for about $1USD per sample (3RAD). *bioRxiv*, 205799.

Gordon, A., & Hannon, G. J. (2010). *"Fastx-toolkit"*. FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab. cshl.edu/fastx_toolkit 5.

Haag, W. R. (2012). *North American freshwater mussels: Natural history, ecology, and conservation*. Cambridge, UK: Cambridge University Press.

Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis v3. *Palaeontologia Electronica*, *4*, 1–9.

Hayakawa, Y., & Matsukura, Y. (2009). Factors influencing the recession rate of Niagara Falls since the 19th century. *Geomorphology*, *110*, 212–216. https://doi.org/10.1016/j.geomorph.2009.04.011

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*, 907–913. https://doi.org/10.1038/35016000

Hewitt, T. L., Bergner, J. L., Woolnough, D. A., & Zanatta, D. T. (2018). Phylogeography of the freshwater mussel species *Lasmigona costata*: Testing post-glacial colonization hypotheses. *Hydrobiologia*, *810*, 191–206. https://doi.org/10.1007/s10750-016-2834-3

Hoban, S. M., Borkowski, D. S., Brosi, S. L., McCleary, T. S., Thompson, L. M., McLachlan, J. S., … Romero-severson, J. (2010). Range-wide distribution of genetic diversity in the North American tree *Juglans cinerea*: A product of range shifts, not ecological marginality or recent population decline. *Molecular Ecology*, *19*, 4876–4891. https://doi.org/10.1111/j.1365-294X.2010.04834.x

Hoffman, J. R., Morris, T. J., & Zanatta, D. T. (2018). Genetic evidence for canal-mediated dispersal of Mapleleaf, *Quadrula quadrula* (Bivalvia:Unionidae) on the Niagara Peninsula, Canada. *Freshwater Science*, *37*, 82–95.

IUCN Standards and Protocols Committee (2019). *Guidelines for Using the IUCN Red List Categories and Criteria. Version 14*. Prepared by the Standards and Petitions Committee. Retrieved from http://www.iucnredlist.org/documents/RedListGuidelines.pdf

Keck, B. P., & Near, T. J. (2013). A new species of *Nothonotus* darter (Teleostei: Percidae) from the Caney Fork in Tennessee, USA. *Bulletin of the Peabody Museum of Natural History*, *54*, 3–21.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*, 782–788.

Kim, J., & Mandrak, N. E. (2016). Assessing the potential movement of invasive fishes through the Welland Canal. *Journal of Great Lakes Research*, *42*, 1102–1108. https://doi.org/10.1016/j.jglr.2016.07.009

Lujan, N. K., Agudelo-Zamora, H., Taphorn, D. C., Booth, P. N., & López-Fernández, H. (2013). Description of a new, narrowly endemic South American darter (Characiformes: Crenuchidae) from the central Guiana Shield highlands of Guyana. *Copeia*, *2013*, 454–463. https://doi.org/10.1643/CI-12-079

Lujan, N. K., Armbruster, J. W., & Lovejoy, N. R. (2018). Multilocus phylogeny, diagnosis and generic revision of the Guiana Shield endemic suckermouth armored catfish tribe Lithoxini (Loricaridae: Hypostominae). *Zoological Journal of the Linnean Society*, *184*, 1169–1186.

Mandrak, N. E., & Crossman, E. J. (1992). Postglacial dispersal of freshwater fishes into Ontario. *Canadian Journal of Zoology*, *70*, 2247–2259. https://doi.org/10.1139/z92-302

Mandrak, N. E., & Cudmore, B. (2012). Fish species at risk and non-native fishes in the Great Lakes basin: Past, present, and future. In A. J. Lynch, W. W. Taylor, & N. J. Leonard (Eds.), *Great lakes fisheries policy & management* (pp. 167–202). East Lansing, MI: Michigan State University Press.

Mathias, P. T., Hoffman, J. R., Wilson, C. C., & Zanatta, D. T. (2018). Signature of postglacial colonization on contemporary genetic structure and diversity of *Quadrula quadrula* (Bivalvia: Unionidae). *Hydrobiologia*, *810*, 207–225. https://doi.org/10.1007/s10750-016-3076-0

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*, 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Papadopoulou, A., & Knowles, L. L. (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 8018–8024. https://doi.org/10.1073/pnas.1601069113

Paz-Vinas, I., & Blanchet, S. (2015). Dendritic connectivity shapes spatial patterns of genetic diversity: A simulation-based study. *Journal of Evolutionary Biology*, *28*, 986–994. https://doi.org/10.1111/jeb.12626

Paz-Vinas, I., Loot, G., Stevens, V. M., & Blanchet, S. (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, *24*, 4586–4604. https://doi.org/10.1111/mec.13345

Ponce, M., van Zon, R., Northrup, S., Gruner, D., Chen, J., Ertinaz, F., … Peltier, W. R. (2019). *Deploying a Top-100 Supercomputer for Large Parallel Workloads: The Niagara Supercomputer*. arXiv:1907.13600 [cs.DC] https://doi.org/10.1145/3332186.3332195

Portt, C. B., Minns, C. K., & King, S. W. (1988). *Morphological and ecological characteristics of common fishes in Ontario Lakes*. Burlington, ON: Department of Fisheries and Oceans.

Randall, R. G., & Minns, C. K. (2000). Use of fish production per unit biomass ratios for measuring the productive capacity of fish habitats. *Canadian Journal of Fisheries and Aquatic Sciences*, *57*, 1657–1667.

Thomaz, A. T., Christie, M. R., & Knowles, L. L. (2016). The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution*, *70*, 731–739. https://doi.org/10.1111/evo.12883

Tinkler, K. J., Pengelly, J. W., Parkins, W. G., & Asselin, G. (1994). Postglacial recession of Niagara Falls in Relation to the Great Lakes. *Quaternary Research*, *42*, 20–29. https://doi.org/10.1006/qres.1994.1050

Vucetich, J. A., & White, T. A. (2003). Spatial patterns of demography and genetic processes across the species range: Null hypotheses for landscape conservation genetics. *Conservation Genetics*, *4*, 639–645.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Winemiller, K. O., López Fernández, H., Taphorn, D. C., Nico, L. G., & Barbarino Duque, A. (2008). Fish assemblages of the Casiquiare River, a corridor and zoogeographic filter for dispersal between the Orinoco and Amazon basins. *Journal of Biogeography*, *35*, 1551–1563.

Wofford, J. E. B., Gresswell, R. E., & Banks, M. A. (2005). Influence of barriers to movement on within-watershed genetic variation of coastal cutthroat trout. *Ecological Applications*, *15*, 628–637. https://doi.org/10.1890/04-0095

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Lujan NK, Weir JT, Noonan BP, Lovejoy NR, Mandrak NE. Is Niagara Falls a barrier to gene flow in riverine fishes? A test using genome-wide SNP data from seven native species. *Mol Ecol*. 2020;00:1–15. https://doi.org/10.1111/mec.15406