Canadian Science Publishing
NRC Research Press journals | Publishing Services

Canadian Journal of Fisheries and Aquatic Sciences
Journal canadien des sciences halieutiques et aquatiques

Genetic population structure of the round whitefish (Prosopium cylindraceum) in North America: multiple markers reveal glacial refugia and regional subdivision.

Journal:	*Canadian Journal of Fisheries and Aquatic Sciences*
Manuscript ID	cjfas-2016-0528.R2
Manuscript Type:	Article
Date Submitted by the Author:	24-May-2017
Complete List of Authors:	Morgan, Thomas; University of Regina, Biology Graham, Carly; University of Regina, Biology McArthur, Andrew; McMaster University Michael G DeGroote Institute for Infectious Diseases, Department of Biochemistry and Biomedical Sciences Raphenya, Amogelang; McMaster University Michael G DeGroote Institute for Infectious Diseases, Department of Biochemistry and Biomedical Sciences Boreham, Douglas; Northern Ontario School of Medicine, Medical Sciences Manzon, Richard; University of Regina, Biology Wilson, Joanna; McMaster University Lance, Stacey; Savannah River Ecology Laboratory Howland, Kimberly; Department of Fisheries and Oceans Patrick, Paul; Petrudev Inc. Somers, Christopher; University of Regina, Biology
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	Coregoninae, MOLECULAR ECOLOGY < General, PHYLOGEOGRAPHY < General, POPULATION STRUCTURE < General, Prosopium cylindraceum

Genetic population structure of the round whitefish (*Prosopium cylindraceum*) in North America: multiple markers reveal glacial refugia and regional subdivision.

Authors: Thomas D. Morgan (tmorgan89@gmail.com)[1], Carly F. Graham[1] (graham9c@gmail.com), Andrew G. McArthur[2] (mcarthua@mcmaster.ca), Amogelang R. Raphenya[2] (raphenar@mcmaster.ca), Douglas R. Boreham[3] (douglas.boreham@nosm.ca), Richard G. Manzon[1] (Richard.Manzon@uregina.ca), Joanna Y. Wilson[4] (joanna.wilson@mcmaster.ca), Stacey L. Lance[5] (lance@srel.uga.edu), Kimberly L. Howland[6] (Kimberly.Howland@dfo-mpo.gc.ca), Paul H. Patrick[7] (paul.patrick@petrudev.com), Christopher M. Somers[1]** (chris.somers@uregina.ca)

**Corresponding author

[1]Biology Department, University of Regina, 3737 Wascana Pkwy, Regina, SK, Canada, S4S 0A2.

[2]M.G. DeGroote Institute for Infectious Disease Reasearch, Department of Biochemistry and Biomedical Sciences, DeGroote School of Medicine, McMaster University, 1280 Main St W, Hamilton, ON, Canada, L8S 4L8.

[3]Medical Sciences, Northern Ontario School of Medicine, 935 Ramsay Lake Rd, Sudbury, ON, Canada, P3E 2C6.

[4]Department of Biology, McMaster University, 1280 Main St W, Hamilton, ON, Canada, L8S 4L8.

[5] Savannah River Ecology Laboratory, University of Georgia, Aiken, SC, United States of America, 29802.

24 [6]Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada, T6G 2R3.

25 [7] Petrudev Inc., 6345 Bell School Line, Milton, ON L9E0N3.

26

ABSTRACT

Round whitefish (*Prosopium cylindraceum*) have a broad, disjunct range across northern North America and Eurasia, and little is known about their genetic population structure. We performed genetic analyses of round whitefish from 16 sites across its range using nine microsatellites, two mitochondrial DNA (mtDNA) loci, and 4,918 to 8,835 single-nucleotide polymorphism (SNP) loci. Our analyses identified deep phylogenetic division between eastern and western portions of the range, likely indicative of origins from at least two separate Pleistocene glacial refugia. Regionally, microsatellites and SNPs identified congruent patterns in subdivision, and population structure was consistent with expectations based on hydrologic connectivity. Within the Laurentian Great Lakes, Lake Huron and Lake Ontario were identified as key areas of interest. Lake Huron appears to be a contemporary source population for several other Great Lakes, and Lake Ontario contains a genetically discrete group of round whitefish. In all cases, multiple genetic markers yielded similar patterns, but SNPs offered substantially enhanced resolution. We conclude that round whitefish have population subdivision on several scales important for understanding their evolutionary history and conservation planning.

INTRODUCTION

The round whitefish (Salmonidae; Coregoninae; *Prosopium cylindraceum*) is a widespread species that has become a conservation concern in several regions of North America. In the northeastern United States several populations have been extirpated, and the species is listed as Endangered in New York (Steinhart et al. 2007) and critically imperiled in Vermont and New Hampshire (Vermont Department of Fish and Wildlife 2015; Nugent and Carpenter 2015). Round whitefish were once targeted for commercial harvest in the Laurentian Great Lakes (Bailey 1963; Mraz 1964); however, declining catch rates and a limited market have caused a reduction in harvest to an average of less than 4,500 kg per year in Canadian waters (Ontario Commercial Fisheries Association data 1994-2013). Population data for most coregonines is limited in the Great Lakes, but recent evidence suggests local declines of round whitefish (e.g. Ebener 2012). In addition, thermal sensitivity during development has raised concern for this species in areas potentially impacted by industrial once-through cooling processes (Patrick et al. 2013; Graham et al. 2016). Little is known about population trends or conservation status in other areas of the extensive round whitefish range.

Round whitefish populations in North America are likely genetically subdivided on multiple geographic scales relevant to understanding their evolutionary history and management. On the continental scale, the species appears to be divided into two large, disjunct areas, occurring in the west from Alaska to northern Manitoba, and the east from the Laurentian Great Lakes to the Atlantic coast (see Fig. 1A, Scott and Crossman 1973; Global Biodiversity Information Facility). During previous glacial maxima, fishes in northern North America were isolated within at least nine identified glacial refugia (Mandrak and Crossman 1992; Ross 2013; Mee et al. 2015), from which they dispersed after the glaciers receded. The resulting pattern is

reflected in the detection of distinct glacial lineages using genetic markers for other fish species with similar distributions (e.g. Wilson and Hebert 1998; April et al. 2013; Mee et al. 2015). The delineation of separate glacial lineages is of principal importance in determining conservation priorities for species (Dizon et al. 1992; Palsbøll et al. 2007; Funk et al. 2012). The disjunct range of round whitefish in North America has led to the hypothesis that they persisted in at least two glacial refugia: the Beringian in the west, and the Mississippian in the east (McPhail and Lindsey 1970; Scott and Crossman 1973). There are currently no genetic data available to address the hypothesis of separate glacial lineages for round whitefish.

Within the eastern and western portions of their range, round whitefish occupy multiple watersheds with variable hydrologic connectivity, which likely affects gene flow. For example, the upper Great Lakes are contiguous and hydrologically connected, enabling fish movement, but many inland lakes occupied by round whitefish are completely isolated from one another. These hydrological features that affect connectivity alter levels of gene flow and should be reflected in the resulting level of genetic differentiation among fish in different bodies of water (Pringle 2003; Waples and Gaggiotti 2006). Molecular tools have been instrumental in quantifying levels of connectivity between populations, as well as detecting the biogeographic relationships of glacial lineages (e.g. Bernatchez and Wilson 1998; April et al. 2013); however, for round whitefish there has yet to be any analyses of genetic population structure beyond very local, fine-scale applications (see Graham et al. 2016; Wood et al. 2016).

Here we present the first broad-scale study of North American round whitefish population genetics and phylogeography. We applied microsatellite genotyping, mtDNA sequencing, and nextRAD sequencing of single nucleotide polymorphisms (SNP) to address the following major objectives: (1) characterize intraspecific phylogenetic relationships for round

whitefish in North America; (2) relate phylogenetic data for round whitefish to putative glacial refugia during the Wisconsinan glaciation that have been identified for other freshwater fish species; and (3) characterize the genetic population structure of round whitefish at local, regional, and continental scales using traditional and next-generation sequencing (NGS) techniques. In addition, we chose to emphasize analyses of fish from the Laurentian Great Lakes due to the high impact of invasive species and environmental disturbance on the region, as well as its economic importance as one of the largest freshwater fisheries in the world (Kohler and Hubert 1999; FAO 2011).

MATERIALS AND METHODS

Sample Collection and DNA Isolation

Tissue samples (pectoral fin clip, adipose fin clip, or muscle stored in 95% ethanol or lysis buffer) were obtained for round whitefish from various sites across North America and one site in eastern Russia (Table 1; Fig. 1). From the western portion of the round whitefish range, the samples included fish from six different watersheds. The north Alaska site was north of the Brooks Range within the north slope watershed that drains into the Arctic Ocean, and the south Alaska sites were within the Nushagak River Basin. Bennett and Little Salmon Lakes are part of the Yukon River Basin, while Simpson Lake, the Rat River sites, and Great Bear Lake are within watersheds of the Mackenzie River system (the Liard, Peel, and Great Bear Lake watersheds, respectively; Benke and Cushing 2005). In the eastern part of the range, the samples include round whitefish caught in each of the Laurentian Great Lakes, Lake Nipigon, and one site in Labrador, Canada (Fig. 1).

Genomic DNA was extracted from 414 samples of round whitefish tissue following the manufacturer's guidelines (Genomic DNA Isolation Kit, Norgen Biotek Corp., Ontario, Canada). However, we extended lysis to 12-14 hours at 55°C and performed the optional step of treatment with RNase A (Qiagen Inc., Ontario, Canada). DNA concentration was determined using a Qubit 2.0 Fluorometer (Life Technologies Inc., Ontario, Canada). Subsets of the 414-sample collection were selected based on likelihood of capturing representative diversity and assay cost, then analyzed using various molecular techniques as indicated (Table 1).

Mitochondrial DNA Sequencing

Portions of the mitochondrial control region (D-loop) and cytochrome c oxidase subunit I (COI) barcode region were polymerase chain reaction (PCR) -amplified for 124 round whitefish individuals representing 7-12 samples from each site, with the exceptions of Russia (n=3) and the Rat River site in the Yukon (n=4). The Lake Michigan – Door County fish were excluded from this analysis due to the close proximity between this location and the Lake Michigan – Milwaukee site. Loci were amplified as described in Delling et al. (2014; D-loop) and Ward et al. (2005; COI), resulting in amplicons of approximately 400 bp and 655 bp respectively. PCR products were purified using MinElute PCR Purification Kits (Qiagen Inc., Ontario, Canada) and Sanger-sequenced commercially (University of Calgary Core DNA Services; See Table 1 for sample details).

Microsatellite Genotyping

Round whitefish were genotyped at 11 microsatellite loci previously developed for this species (O'Bryhim et al. 2013; Graham et al. 2016; Details in Table S1). We genotyped

individuals from 14 sites that had DNA for 8 or more fish (n=390). Samples were included for five sites in the western range (the Yukon and Alaska), eight sites in the Great Lakes region, and one site in Labrador. Six microsatellite loci (*Prwi6*, *Prwi15*, *Prwi24*, *Prwi25*, *Prwi27*, and *Prwi28*) were amplified and genotyped as described in Graham et al. (2016). Genotypes were determined using GENEMARKER 2.20 software (Softgenetics, State College, PA) with a bin-width of one nucleotide anchored to the median integer value of the raw genotype read. The remaining five loci (*Prwi55*, *Prwi56*, *Prwi60*, *Prwi65*, and *Prwi72*; Graham et al. 2016) were amplified and genotyped as described in O'Bryhim et al. (2013).

Round whitefish genotype data were assessed for scoring errors and null alleles (Micro-Checker v2.2.3; Van Oosterhout et al. 2004), and for conformation to Hardy-Weinberg Equilibrium (HWE) within each of the 14 presumptive populations (GENEPOP v4.3; Rousset 2008). A sequential Bonferroni correction was applied to account for multiple HWE tests. Individuals with complete data for at least seven loci were retained in all subsequent analyses. *Prwi56* and *Prwi72* showed evidence of null alleles; these loci were excluded from subsequent analyses. The nine remaining loci conformed to HWE for all populations and were retained for subsequent analyses; all 390 individuals had data for at least seven of nine loci.

NextRAD Sequencing

Genomic DNA for 190 round whitefish samples from 14 sites was converted into nextRAD genotyping-by-sequencing libraries (SNPsaurus, Oregon, USA; as described by Russello et al. 2015). Sites included in the analysis were those that had reasonable DNA quality (visible fragments larger than 1Kb when extraction run on a gel) for >8 individuals, with the exceptions of Little Salmon Lake in the Yukon and Russia (n=3 in both cases). Briefly,

genomic DNA was first fragmented with Nextera reagent (Illumina, Inc), which fragments the genome using a transposase and also ligates short adapter sequences to the ends of the fragments (Marine et al. 2011). The Nextera reaction was scaled for fragmenting 7 ng of genomic DNA, although 15.75-17.5 ng of genomic DNA were used for input to compensate for degraded DNA in the samples. Fragmented DNA was then amplified, with one of the primers matching the adapter and extending 9 nucleotides into the genomic DNA with the selective sequence 5'-GTGTAGAGC-3'. Thus, only fragments starting with a sequence that hybridized with the selective sequence of the primer will be efficiently amplified. PCR amplification was done with an annealing temperature of 73 °C for 26 cycles. The nextRAD libraries were sequenced on an Illumina HiSeq 2500 with 100bp single-end reads (University of Oregon).

NextRAD data were uploaded to an online Galaxy analysis platform at McMaster University (galaxylab.mcmaster.ca; Afgan et al. 2016). FASTQ files were first processed using Trimmomatic (Bolger et al. 2014); Nextera adaptors were removed, long reads were trimmed to 100bp, and low quality/short reads (<100bp) were removed from subsequent analyses. The remaining reads were analyzed using FastQC (Galaxy version 0.64; Andrews 2010) and individuals containing <100,000 reads were removed. FastQC analysis indicated eight samples with low read numbers, six from Lake Ontario, and one each from Lake Huron and Lake Nipigon. These samples were excluded from subsequent analyses, leaving 182 samples with a mean read number of 1,031,989 ± 452,088 (SD).

Sequences for all individuals used for phylogenetics and population structure (see below) were analyzed using the *de novo* pipeline (no reference genome available) of Stacks 1.41 (Catchen et al. 2013). Putative loci were assembled into 'stacks' using *ustacks* with a minimum stack depth (-m) of 3, and 2 mismatches allowed between stacks (-M). A catalogue of the loci

was then constructed for all individuals using *cstacks* allowing a mismatch (-n) of 3 between sample tags. Sets of stacks were then searched against the catalogue using *sstacks*.

Phylogenetic Analyses

MtDNA

Trace files were compiled and edited in CodonCode Aligner v.6.0.2 (CodonCode Corporation, Centerville, MA, USA), then trimmed for quality and concatenated in MEGA6 (Tamura et al. 2013). Unique haplotypes were identified in the final alignment and concatenated sequences were run through Modeltest using Paup4.0a147 (Swofford 2002) to determine the best model of sequence evolution. A statistical parsimony network was constructed using TCS 1.21 (Clement et al. 2000) for the 13 identified haplotypes (Tables 2 and S2) using 95% connection limits (Fig. S1).

Phylogenetic tree building analysis of mtDNA was first conducted using a maximum-likelihood (ML) approach with the Russian samples designated as the outgroup. ML trees were inferred using the GTRCAT model in the PTHREADS version of RAxML v8.2.8 (Stamatakis 2014), using the default bootstrap procedure with robustness tested using 1000 replicates. A Bayesian search of tree space was also conducted using MrBayes v.3.2.6 (Ronquist et al. 2012) and seeded with the RAxML tree. We used four chains under the HKY+I model until after 7 x 10^6 generations the standard deviation of split frequencies fell below 0.01. Haplotypes for the concatenated sequences were determined and mapped back to their putative population. Sequences for COI were compared to round whitefish sequences already on the International Barcode of Life Database using the BLAST tool to confirm that all samples were the correct species (Ratnasingham and Hebert 2007).

NextRAD

For phylogenetic analyses, a consensus sequence (with IUPAC ambiguity codes) was exported for each population using the Stacks program *populations* in full-sequence Phylip format. To account for unequal sample sizes, consensus sequences were based on the three samples from each site with the most raw reads after Trimmomatic filtering. Loci present in 50% of individuals within a population, in at least 12 of the 14 presumptive populations, and with a minimum read depth of 3x were exported. This maximized the number of captured loci, and allowed for the construction of a supermatrix with a consensus sequence representing each population, the size of which has been shown to positively impact resolution of phylogenetic relationships (Wagner et al. 2013). The matrix of RAD loci was used to determine phylogeographic patterns among locations. Maximum likelihood and Bayesian trees were constructed using the PTHREADS version of RAxML v7.2.7 and MrBayes v3.1.2 on CIPRES (Miller et al. 2010) following the same procedure as for the 2 mitochondrial loci. The samples from Russia were used as an outgroup to root the tree.

Population Structure Analyses

Microsatellites

Continental and regional population structures were analyzed using three types of analyses. First, a distance matrix was created for all sites using an AMOVA (Excoffier et al. 1992) to calculate pairwise F_{ST} between presumed populations (Weir and Cockerham 1984; GENODIVE; Meirmans and Van Tienderen 2004; Table S2). Measures of genetic diversity were

225 determined for each population (GENODIVE), including a measure of allelic richness rarefied to
226 the smallest sample size of eight individuals (HP-Rare; Kalinowski 2005).
227 Population structure was further analyzed using discriminant analysis of principal
228 components (DAPC; adegenet; Jombart 2008, Jombart and Ahmed 2011). Missing genotypes
229 were imputed using a random forest model implemented in *stackr* (Breiman 2001; Ishwaran and
230 Kogalur 2007; Gosselin and Bernatchez 2016) based on genotypes within each population.
231 Ellipses were generated hierarchically for all sites, Alaska and Yukon sites, and Great Lakes
232 sites, and validated using the *optim.a.score* function (PCs retained = 12, 5, and 26, respectively).
233 Group delineations were then determined based on separation along the first and second
234 principal axes.
235 Contemporary migration rates were determined for sites in the Great Lakes region using
236 the program BAYESASS 3.0 (Wilson and Rannala 2003). We ran BAYESASS using 5 x 10^6
237 iterations with a burn-in of 1 x 10^6 and sampling every 2000 iterations; mixing parameters were
238 adjusted to f = 0.35 and a = 0.35 to maintain optimal values between 0.20 and 0.60.
239 The Bayesian clustering program STRUCTURE was run on all samples to determine the
240 most likely number of genetic groups (Pritchard et al. 2000). The analysis was run 10 times with
241 K ranging from 1 to 16 to account for the maximum number of populations (14) and additional
242 unanticipated substructure. Each run consisted of a burn-in of 100,000 followed by 100,000
243 Markov chain Monte Carlo (MCMC) steps. STRUCTURE was then run hierarchically on the
244 western sites (Alaska and the Yukon) with K ranging from 1 to 7, the Great Lakes sites and
245 Labrador with K ranging from 1 to 11, and the Great Lakes sites on their own with K ranging
246 from 1 to 10.

After each STRUCTURE run, the most likely number of genetic groups was determined using the methods of Evanno et al. (2005) in the program STRUCTURE HARVESTER v0.6.94 (Earl and vonHoldt 2012). Results of all 10 runs were combined for the most likely K value using the Greedy algorithm of CLUMPP (Jakobsson and Rosenberg 2007), and membership coefficients for individuals were visualized using the program DISTRUCT (Rosenberg 2004).

NextRAD

Single nucleotide polymorphism (SNP) genotypes (one per locus) for loci present in 50% of individuals, found in at least 3 of 14 populations, with a minor allele frequency of 0.05, and with a read depth of at least 6x were used for population structure analyses. Compared to phylogenetics, these analyses retained the relative frequencies of SNPs within populations and allowed for finer-scale population metrics. Loci were outputted and checked in PLINK1.07 (Purcell et al. 2007) for conformation to Hardy-Weinberg equilibrium ($P<0.05$) in the 12 sites with >8 samples. Loci that did not conform to HWE in two or more populations were used to create an exclusion list for *populations*, and the reduced dataset was re-exported. Patterns in missing data were then analyzed using *stackr* (Gosselin and Bernatchez 2016). Five individuals (one each from Lakes Huron, Superior, and Nipigon, as well as two from south Alaska) were found to be biased based on missingness because of genotyping rates under 5.3% of the total panel, and were removed from subsequent analyses. The overall SNP panel was not biased by missingness, so all SNP loci were retained for subsequent population genetic analyses.

For the nextRAD SNP panel, measures of genetic variation were calculated using *populations*, excluding F_{ST} which was calculated in GENODIVE as described for microsatellites. Hierarchical runs of STRUCTURE were implemented using STRAUTO1.0 (Chhatre et al. 2016)

to determine the number of genetic groups at continental and regional levels until no further substructure was detected. Each analysis was run 10 times with a burn-in of 50,000 followed by 50,000 MCMC steps. K was tested for values ranging from 1 to 15 for all sites, then 1 to 8 for western sites, 1 to 10 for eastern sites, and 1 to 10 for the Great Lakes. The most likely number of genetic groups was determined as described above for microsatellite loci using STRUCTURE HARVESTER. DAPC analyses were implemented hierarchically using *adegenet* as described for the microsatellite dataset. We used the random forest model in *stackr* to impute missing genotypes from within populations and validated the number of retained principal components using *optim.a.score* (9, 2, and 7 for all-sites, Alaska-Yukon sites, and the Great Lakes, respectively).

RESULTS

Phylogenetic Analyses

MtDNA

COI sequences were identified through BLAST as closest to round whitefish for all samples. North American round whitefish returned a >99.5% match to sequences already on BOLD, while round whitefish from Russia returned a top hit of 98.25% sequence similarity to round whitefish. Edited sequences for D-loop and COI yielded 354 and 578bp respectively, for a concatenated sequence of 932 bp (haplotype information in Table S3). The concatenated sequences yielded 11 haplotypes for RWF across North America and two haplotypes for samples from eastern Russia (Fig. 2; Table S3). Five haplotypes were exclusively found in Alaska, Yukon, and Northwest Territories sites (designated *W1, W2, W3, W4* and *W5;* n = 10, 19, 14, 3, and 7, respectively; Table 2). Six haplotypes were found exclusively in the Great Lakes and Labrador sites

(designated *E1-E6*). In the east, the *E3* and *E6* haplotypes represented the majority of samples (n= 30 and 31, respectively of 68 total; Table 2). Tree building analyses were unable to resolve the relationships between most mtDNA haplotypes with high degrees of confidence. However, Russian haplotypes were strongly supported as the outgroup, and within North America there was moderate support for *E1*, *E2*, *E3*, and *E4* sharing a lineage, and strong support for *E5* and *E6* being from a shared lineage more closely associated with *W1, W2*, and *W3* (Fig. 2; Fig. S1). Interestingly, haplotypes *E1-E4* were five mutation steps from the other eastern haplotypes *E5* and *E6* (Fig S1), a greater level of differentiation than observed between eastern and western haplotypes (between two and seven mutational steps).

NextRAD

Stacks analyses yielded a matrix of 4918 loci for phylogenetic analysis. Relationships among major regions resolved with high support (Fig. 3); the eastern sites were all highly differentiated from the western and Russian sites (Maximum-likelihood bootstrap (ML) = 100, Bayesian P = 1.00). There was also high support (Bayesian P = 1.00) for western sites being distinct from the Russian site. The branch lengths between the western sites and the Great Lakes were 0.00112 substitutions/site compared to 0.00041 substitutions/site between the Great Lakes and Labrador (2.7X difference). The Russian group separated from the east by a branch length of 0.00421 substitutions/site, and the west by 0.00367 substitutions/site (3.8X and 3.3X the distance from each other). Within the western region, individual sites resolved with moderate support in the maximum likelihood analysis (ML > 60), and with moderate to high support in the Bayesian analysis (P > 0.80). In the east there was strong support separating Labrador from all the Great Lakes sites (ML = 75, Bayesian P = 1.00). Lake Nipigon and Lake Superior resolved as their

316 own group from the other Great Lakes (ML = 94, Bayesian P = 1.00), and there was moderate to
317 strong support for Lake Ontario being separate from Lake Huron, Lake Michigan, and Georgian
318 Bay (ML >50; Bayesian P = 1.00). Further moderate to high support delineated relationships
319 within Lake Huron, Lake Michigan, and Georgian Bay sites (ML > 60; Bayesian P = 1.00; Fig.
320 3).
321
322 Population Structure
323 *Microsatellites*
324 AMOVA indicated significant differentiation in pair-wise F_{ST} for 26 of the 28 between-
325 site comparisons (Table S2). There was non-significant differentiation between fish from Lake
326 Michigan Door County and those from Lake Michigan Milwaukee, as well as Lake Huron –
327 main basin and northern Georgian Bay. Within the Great Lakes F_{ST} values ranged from 0.020 to
328 0.108, whereas in the Alaska-Yukon region F_{ST} values ranged from 0.049 to 0.283. The nature of
329 the sampling distribution did not permit formal testing of isolation by distance; however, the
330 highest values of F_{ST} observed were for comparisons among the most geographically distant
331 populations. Simpson Lake in Yukon Territory and T-Bone Lake in Labrador had a F_{ST} value of
332 0.416, Simpson Lake and south Georgian Bay 0.321, and south Georgian Bay and Labrador
333 0.336.
334 In the west, southern Alaska and Yukon sites showed similar allelic richness (average
335 ranging from 5.0 to 5.1 alleles per locus after rarefaction; Table S2). Northern Alaska and
336 Simpson Lake had lower values indicating less genetic diversity (3.6 and 3.4 average alleles per
337 locus respectively). Measures of allelic richness were fairly uniform across the Great Lakes
338 region (ranging from 4.4 to 4.8 average alleles per locus), whereas T-Bone Lake in Labrador had

339 the lowest measure at 2.8 average alleles per locus. The total variance retained for the DAPC of
340 all sites, Alaska-Yukon, and the Great Lakes was 43.9%, 35.6%, and 69.3%, respectively (Fig.
341 4). The DAPC of all sites resolved the eastern and western regions well along the first axis, and
342 also separated Labrador from the Great Lakes along the second axis (Fig. 4A). The subsequent
343 DAPCs were for only the western sites (Fig. 4B) and only the Great Lakes sites (Fig. 4C) and
344 once again resolved obvious groups along the two principal axes. For the western sites DAPC
345 ellipses indicated three groups, with north Alaska and Simpson Lake separating distinctly from
346 south Alaska, Bennett Lake, and Little Salmon Lake. Much less separation was apparent between
347 south Alaska, Bennett Lake, and Little Salmon Lake. In the Great Lakes, there was little
348 evidence for clear separation among groups from different sites. However, fish from Lake
349 Ontario and southern Georgian Bay sites showed some evidence of being distinct from the other
350 locations sampled.

351 STRUCTURE analysis of microsatellites returned the same patterns observed in DAPC.
352 Analysis of all samples returned a most likely value of K=2, corresponding to eastern and
353 western regions (membership coefficient, Q, >0.95; Fig. S2A). Subsequent hierarchical runs of
354 STRUCTURE on the Alaska-Yukon region, and Great Lakes and Labrador regions returned
355 most likely values of K=3 in both cases. The three clusters identified in the west corresponded to
356 north Alaska, Simpson Lake in the Yukon (Q of 0.98 and 0.96, respectively), and the three other
357 western sites as one cluster (Q>0.95; Fig. S2B). The three clusters identified in the east
358 delineated T-Bone Lake in Labrador (Q=0.98), and two clusters in the Great Lakes. There was
359 weaker cluster assignment within the Great Lakes; Lake Huron, northern Georgian Bay, and
360 southern Georgian Bay all assigned most closely to cluster 1 (Q=0.81, 0.81, and 0.92,
361 respectively), whereas Lake Nipigon and Lake Ontario assigned to cluster 2 (Q=0.85 and 0.91,

respectively; see Fig. S2C). Lake Michigan and Lake Superior sites did not assign with high confidence to either of the two Great Lakes clusters (Q<0.70). Further hierarchical runs on subgroups in both the east and west returned most likely values of K=1 indicating no further population structure. BAYESASS analysis detected recent migration within the Great Lakes. Connectivity with recent migration was detected for Lake Huron to Lake Superior, Lake Huron to northern Georgian Bay and Lake Michigan, and migration from Lake Superior to Lake Nipigon (Fig. 5).

NextRAD

After filtering SNPs for conformation to HWE and only retaining the first SNP from each locus, an output of 8835 SNP loci was retained for population structure analyses. Measures of genetic variation and F_{ST} values for the nextRAD loci showed similar relationships to those identified using microsatellites (Table S4). The Great Lakes sites were less differentiated from Labrador (F_{ST} ranging from 0.245 to 0.287) than western sites were (F_{ST} ranging from 0.535 to 0.595). Within the Great Lakes F_{ST} values ranged from 0.013 to 0.089, whereas in the Alaska-Yukon region F_{ST} values ranged from 0.006 to 0.301. Samples from the Russian sites showed the highest differentiation from all North American sites (F_{ST} ranging from 0.383 to 0.674); however, pairwise F_{ST} comparisons for Russian samples (as well as those for Little Salmon Lake) were often not significant after Bonferonni correction, likely due to low sample size (n=3).

The DAPC analyses of SNP loci retained only the loci present in all populations for each analysis. The panels were reduced to 343, 801, and 1375 SNP loci for all-sites, Alaska-Yukon sites, and the Great Lakes sites, respectively. The total variance retained for the DAPC of all sites, Alaska-Yukon, and the Great Lakes was 43.1%, 19.5%, and 18.3%, respectively (Fig. 6).

The DAPC of all sites resolved eastern and western regions along the first axis and separated Russian round whitefish from other western sites (Fig. 6A). Labrador round whitefish were separated from Great Lakes individuals along the second axis. The subsequent DAPC on the five Alaska-Yukon sites resolved three groups along the first axis, with Simpson Lake and North Alaska being distinct from the other sites. Along the second axis south Alaska resolved into a separate group from Bennett and Little Salmon Lakes (Fig. 6B). In the Great Lakes DAPC, Lake Ontario round whitefish were separated from the other Great Lakes populations along the first axis, while the other Great Lakes populations separated into three groups along the second axis corresponding to northern Georgian Bay, Lake Huron plus south Georgian Bay and Lake Michigan, and Lake Nipigon with Lake Superior (Fig. 6C).

The STRUCTURE analysis returned a most likely value of K=2, once again corresponding to eastern (Q=1.00) and western sites (Q>0.99), with samples from Russia clustering more closely with western samples (Q=1.00; Fig. S3A). Subsequent hierarchical runs returned K=2 separating Russia (Q=1.00) from all Alaska-Yukon sites (Q=1.00; Fig. S3B), and K=3 within Alaska-Yukon corresponding to north Alaska (Q=1.00), Simpson Lake (Q=1.00), and the other three western sites (Little Salmon Lake, Bennett Lake, and south Alaska; Q>0.92; Fig. S3D). Within the east, STRUCTURE indicated a most likely value of K=2 corresponding to the Great Lakes (Q> 0.97) and Labrador (Q=1.00; Fig. S3C). Analysis within the Great Lakes indicated a most likely value of K=2 with separation of Lake Ontario (Q=0.99) from the other Great Lakes (Q>0.96; Fig. S3E).

DISCUSSION

Phylogeography of Round Whitefish

Eastern and western round whitefish have distinct lineages based on several genetic marker types, supporting the hypothesis that they spread from separate glacial refugia. Separate refugia are likely, based on the distinct mitochondrial haplotypes found in the eastern and western portions of the range, as well as shared haplotypes between the Great Lakes and Labrador (the east-west extent of the eastern range). In addition, analysis of microsatellite loci showed clearly separate clusters with high assignment certainty for eastern and western groups, and highly significant differentiation based on F_{ST} values. Perhaps most convincingly, the over 4900 SNP loci produced by nextRAD sequencing provided much more detection power and resolved the eastern and western groups with higher clarity and confidence than either the mtDNA or microsatellite analyses. Our results are broadly consistent with studies of other postglacial fishes confirmed to originate in multiple glacial refugia, such as lake trout (Wilson and Hebert 1998), pygmy whitefish (Blanchfield et al. 2014), rainbow trout (Tamkee et al. 2010), arctic grayling (Stamford and Taylor 2004), lake sturgeon (McDermid et al. 2011), and lake whitefish (Mee et al. 2015). There is also some evidence of regional meristic differences between eastern and western round whitefish (McPhail and Lindsey 1970; Scott and Crossman 1973), although its relevance is uncertain (Lindsey et al. 1981). We conclude that contemporary round whitefish populations originated from at least two separate glacial refugia similar to other northern fish species.

Our results suggest that the eastern and western round whitefish populations probably stem from isolation in, and postglacial dispersal from, separate Wisconsinan refugia, but further analysis is required to better support this idea. For instance, the incomplete lineage sorting of mtDNA haplotypes we observed may be resolved with sequencing of additional loci. D-loop and COI loci have been sufficiently informative in previous studies of regional relationships for

Nearctic freshwater fish species (e.g., April et al. 2013; Delling et al. 2014; Yamamoto et al. 2014; Overdyk et al. 2015), and occurred as regionally private haplotypes in round whitefish. However, these markers were not as variable in round whitefish as in previous studies of other North American fishes, and our mtDNA analyses alone were unable to resolve distinct regional clades. Sequencing of additional loci, such as ATPase VI (Witt et al. 2011) may help resolve mtDNA relationships, and provide more insight into the apparently deeper divergence of *E5* and *E6* from the other eastern haplotypes. However, NGS techniques provide unprecedented capacity for phylogenetic studies that is far beyond that of traditional techniques (Emerson et al. 2010; Wagner et al. 2013; Bryson et al. 2016). Based on nextRAD NGS data in our study, round whitefish from eastern North America (Lake Nipigon to Labrador) were strongly supported as a distinct clade from western sites. Tree branch lengths within the east were substantially shorter than between the east and west (2.7X difference), despite more than 1700 km separating the Great Lakes and Labrador sites. These results are consistent with expectations for separate glacial lineages in the eastern and western portions of the North American round whitefish range. The wider application of NGS is still in its infancy, and requires continued development (Narum et al. 2013; Schafer et al. 2015; Garner et al. 2016; Schafer et al. 2016); for example, the application of a molecular clock to heterogeneous RAD loci is currently not commonplace. However, given the superior resolution of our nextRAD analyses, we contend that further characterization of round whitefish populations should continue to expand and apply NGS techniques to better reveal informative patterns of genetic diversity.

The western lineage of round whitefish, also strongly suggested by combined genetic marker data, most likely originated from the Beringian refugium (McPhail and Lindsey 1970; Ross 2013). The Beringian lineage of round whitefish likely spread to the Mackenzie River

Basin via proglacial lakes that formed in northern and southern Yukon, similar to dispersal routes used by other fish species that moved to the Peel and Liard River watersheds (Foote et al. 1992; Wilson and Hebert 1998; Stamford and Taylor 2004). Interestingly, the *W5* mtDNA haplotype was only detected in fish from Great Bear Lake. Aquatic systems around Great Bear Lake are a potential mixing zone for lake whitefish from the Nahanni, Beringian, and Mississippian glacial refugia (Foote et al. 1992; Mee et al. 2015), which could explain the *W5* haplotype in round whitefish only being detected at this site. Wider sampling of round whitefish from other western watersheds is necessary to determine whether there are any previously unidentified glacial lineages represented. Previous studies of postglacial fish have identified Nahanni lineages within the lower Liard River system for other fish species (Foote et al. 1992; Wilson and Hebert 1998; Stamford and Taylor 2004); however, round whitefish from this region have not been previously characterized.

Wider genetic characterization of populations east of the Great Lakes is also required to determine the refugial origins within the eastern sub-range. Past analyses of Great Lakes fishes have identified the region as a likely suture zone for glacial lineages from the Mississippian and Atlantic refugia (Mandrak and Crossman 1992; April et al. 2013). However, round whitefish have not been genetically characterized sufficiently to determine biogeographical patterns of diversity within the east. With the noted decline of round whitefish in the northeastern United States (Steinhart 2007; Vermont Department of Fish and Wildlife 2015; Nugent and Carpenter 2015) and the Great Lakes (Ebener 2012), further characterizing genetic diversity, gene flow, and the influence of glacial lineages will be important to preserving diversity of the species in the eastern part of their range. The close relationship detected using nextRAD SNPs, and a common mtDNA haplotype observed among sites (*E6*), could be due to common lineage from one

refugium or secondary contact between multiple eastern refugia, and should be further investigated in order to characterize the patterns of existing genetic diversity. The presence of multiple eastern refugia is hinted at in the divergence observed in eastern haplotypes (e.g. haplotypes *E1-E4* are five mutations from haplotypes *E5* and *E6*; Fig. 2 and Fig. S1). However, wider genetic characterization of round whitefish populations is necessary in order to confirm whether this is due to separate glacial lineages.

Additional key areas of the round whitefish range should be assessed to better characterize post-glacial migration and the range disjunction between the east and west. As the easternmost known populations of the western sub-range, round whitefish from the Churchill and Keewatin River Basins (unsampled in this study) should be genetically characterized to confirm the hypothesized migration routes from the Beringian refugium following glacial retreat. The analyses of western round whitefish in this study represent four of the most western watersheds in North America, and only the most western reaches of the Mackenzie River Basin. The recent discovery of populations of pygmy whitefish in an area that was supposed to represent the gap between disjunct sub-ranges (Blanchfield et al. 2014) extended their known range approximately 320 km further east in northwestern Ontario, and highlights an example where apparent geographic range disjunctions in other similar species have been mischaracterized. Genetic characterization of round whitefish at the extents of the known western range (e.g. further to the east) will inform the glacial lineages of populations in this region.

Elucidating the contributions of glacial Lake Agassiz to round whitefish distribution should be investigated in order to understand the disjunction between eastern and western round whitefish. Lake Agassiz was a glacial lake that variously extended from central Saskatchewan to northern Ontario approximately 12,300 yr BP to 7,500 yr BP (Teller and Clayton 1983); it

500 facilitated the postglacial migration of many fish species across the continent, and secondary
501 contact of western and eastern refugial lineages (Stewart and Lindsey 1983). The apparent
502 absence of round whitefish from the Lake Agassiz region was likely due either to their failure to
503 enter Lake Agassiz, or their extirpation from the region secondarily. The presence of isolated
504 populations of round whitefish in the Severn River system of northern Ontario (Scott and
505 Crossman 1973; not shown in Fig. 1) indicates that round whitefish may have invaded Lake
506 Agassiz; however, they did not persist there despite seemingly suitable habitat (Stewart and
507 Lindsey 1983). Further genetic characterization of round whitefish populations proximate to this
508 region, including those isolated populations, will therefore also inform our understanding of the
509 contribution of Lake Agassiz to current round whitefish phylogeography.
510
511 Regional Population Structure
512 *Alaska and Yukon Region*
513 Round whitefish in the western sites showed greater regional differentiation than in the
514 Great Lakes. Within the Alaska and Yukon sites, we detected population structure and
515 subdivision consistent with the more fragmented hydrologic connectivity of the sampled
516 watersheds using both microsatellites and SNPs (Benke and Cushing 2005). Hydrologic
517 connectivity facilitates contemporary gene flow; proper understanding of this gene flow is
518 integral to informed management (Pringle 2003; Waples and Gaggiotti 2006). In the Alaska-
519 Yukon region, determining the subdivision of round whitefish populations further allows for the
520 determination of genetic stocks across the US-Canada border, informing management between
521 these jurisdictions (Poff et al. 2003; Ban et al. 2013). For the western sites, evidence from both
522 microsatellite and nextRAD SNP analyses indicated a connection between Nugashek River

Basin populations in southern Alaska, and Yukon River Basin populations. A similar level of connectivity was not detected for the site north of the Brooks Range or the site in eastern Yukon, which are both isolated from the Yukon River by mountain ranges. The subdivisions we have identified support the notion that round whitefish populations are delineated on multiple spatial scales in western North America that are not necessarily determined by simple geographic distance, and that these populations can span the American-Canadian border.

We detected strong differentiation of Simpson Lake round whitefish, in eastern Yukon, and those north of the Brooks Range in north Alaska, in both the microsatellite and nextRAD SNP analyses. Simpson Lake is separated from the Yukon River Basin by the Cassiar and Selwyn Mountains. Following glaciation, fish species migrated from the Beringian refugium east to the Mackenzie River Basin and other northern watersheds. Migration was facilitated through the formation of proglacial lakes in northern and southern Yukon that connected the Yukon River Basin and the Mackenzie River Basin (Lindsey et al. 1981; Pielou 1991). These routes of dispersal have been genetically characterized in several other species such as lake whitefish (Foote et al. 1992), lake trout (Wilson and Hebert 1998), and Arctic grayling (Stamford and Taylor 2004). The Simpson Lake population of round whitefish is isolated within the Liard system from others further west in Alaska and Yukon Territory. The Liard system also harbours round whitefish populations near their southern extent in the west, and near the Nahanni refugium (Foote et al. 1992; Wilson and Hebert 1998; Stamford and Taylor 2004). The assessment of Simpson Lake round whitefish relative to other Liard River populations should be prioritized to determine its regional status, degree of isolation, and evolutionary history following glaciation. Isolation may limit the possibility of dispersal among western populations, including potential rescue by migrants in the event of major declines.

546

547 *The Laurentian Great Lakes Region*

548 Our analyses identified levels of subdivision and gene flow consistent with the
549 hydrological connectivity of the Great Lakes. We observed significant genetic differentiation of
550 round whitefish among Lake Michigan, Lake Superior, Lake Nipigon, Lake Ontario, Lake
551 Huron, and Georgian Bay using both microsatellites and SNPs. This subdivision was weak to
552 moderate between the upper Great Lakes (Lakes Superior, Michigan, and Huron), and
553 consistently higher for southern Georgian Bay and Lake Ontario. The analyses indicate
554 significant differences on the level of each lake, consistent with previous studies on species such
555 as lake whitefish (VanDeHey et al. 2009; Bernard et al. 2009; Stott et al. 2010), smallmouth bass
556 (*Micropterus dolomieu*; Stepien et al. 2007), walleye (*Sander vitreus*; Stepien et al. 2009;
557 Haponski and Stepien 2014), and yellow perch (*Perca flavacens*; Kocovsky et al. 2013; Sullivan
558 and Stepien 2014).

559 Our analysis of SNPs using DAPC and phylogenetic approaches supported closer links
560 between Lakes Superior and Nipigon, Lakes Michigan and Huron (including northern Georgian
561 Bay), and isolation of Lake Ontario from the rest of the Great Lakes. DAPC of SNP genotypes,
562 STRUCTURE analysis of SNPs, and tree analysis of nextRAD loci strongly supported that Lake
563 Ontario is isolated relative to the other Great Lakes. In addition, there was moderate support for
564 Lake Ontario being distinct from the other Great Lakes based on DAPC and STRUCTURE
565 analyses of microsatellite genotypes, and higher relative F_{ST} for both microsatellites and SNPs.
566 Lake Ontario is likely disjunct from the other Great Lakes due to Niagara Falls as a barrier to
567 gene flow, which is reinforced by the absence of round whitefish in Lake Erie, which as the
568 shallowest and warmest Great Lake, lacks suitable habitat and tends to support warm-water

569 species (Leach and Nepszy 1976). Phylogenetic analysis of nextRAD SNPs suggests a more
570 recent isolation than separate Wisconsinan glacial refugia because Lake Ontario fish formed a
571 clade within the other Great Lakes, rather than an outgroup, and showed relatively little sequence
572 divergence. Monitoring of round whitefish in each lake is necessary to ensure persistence of the
573 species within the Great Lakes, and to detect any further declines, such as those observed in Lake
574 Huron and Georgian Bay (Ebener 2012). Considering their genetic differentiation and isolation,
575 management plans for the Great Lakes should consider round whitefish populations in Lake
576 Ontario as a distinct genetic stock from those in the upper Great Lakes.
577 Analysis of contemporary migration using microsatellites further highlights the potential
578 importance of Lake Huron and Lake Ontario to the genetic diversity of round whitefish in the
579 Great Lakes. Lake Huron appears to contribute (or previously contributed) to round whitefish
580 populations in Lakes Michigan, Superior, and northern Georgian Bay, as well as indirectly to
581 Lake Nipigon through migrants from Lake Superior. The noted decline of round whitefish in
582 Lake Huron and northern Georgian Bay may therefore have a wider impact on populations
583 within the rest of the Great Lakes (Ebener 2012). Lake Ontario does not appear to exchange
584 migrants with the other Great Lakes, consistent with having disjunct populations. This finding
585 further supports that Lake Ontario should be considered separately as an important and distinct
586 unit for conserving round whitefish genetic diversity in the Great Lakes. The recent declines of
587 round whitefish in Lake Huron and Georgian Bay should be considered with the additional
588 understanding that they may supplement or contribute migrants to the other Great Lakes.
589
590 Russian Round Whitefish

591 Round whitefish from Russia were highly differentiated from those in North America.
592 COI sequences for North American round whitefish were >99.5% similar to others already on
593 BOLD; however, Russian individuals returned a match of only 98.25%. This difference of 1.75%
594 is substantially higher than the average intraspecific difference of 0.73% (SE 0.053) for other
595 North American freshwater fish (April et al. 2011). NextRAD tree analyses also indicated strong
596 differentiation of the Russian site from North American populations with branch lengths >3.3X
597 longer than the differences observed within North American sites. These two analyses support
598 that Russian round whitefish may warrant designation as a separate Evolutionarily Significant
599 Unit (ESU; Ryder 1986; Moritz 1994) or species, especially considering the yet-uncharacterized
600 genetic diversity of populations further west in Russia. Round whitefish are the only species
601 from the genus *Prosopium* found outside of North America, and have likely been isolated from
602 North American populations since at least the Wisconsinan glaciations (McPhail and Lindsey
603 1970). Further investigation into the Eurasian populations of round whitefish will improve
604 understanding of coregonine postglacial migration and contemporary global connectivity.
605

606 **GENERAL CONCLUSIONS**

607 We conclude that round whitefish have important genetic population subdivision across
608 their range at multiple geographic and temporal scales. The major mechanisms driving genetic
609 population structure are most likely: (a) origins in at least two separate glacial refugia
610 representing current eastner and western parts of the round whitefish range; and (b) barriers to
611 gene flow presented by the hydrologic connectivity of currently occupied aquatic systems. Our
612 study provides the first context for understanding round whitefish genetic diversity across North
613 America using microsatellites and mtDNA loci, which can be compared to past studies of

postglacial fishes. In addition, we provide the first examination of the genetic diversity of the round whitefish using a large number of SNPs. Eastern and western glacial lineages were suggested in all of our analyses; however, wider strategic sampling of round whitefish populations across North America, as well as additional genetic characterization (e.g., additional mtDNA loci), is necessary to resolve specific hypotheses about their refugial origins and postglacial migration. We have identified regions where additional focused studies of round whitefish populations will help resolve these relationships. Finally, we conclude that Lake Ontario and Lake Huron are key populations for long-term management of genetic diversity and stock structure in the Great Lakes region. Lake Ontario round whitefish are differentiated from those in the upper Great Lakes, and Lake Huron may be an important feeder lake for others in the system.

DATA AVAILABILITY

Mitochondrial haplotype sequences are available on Genbank under Accession numbers MF278536-MF278561. NextRAD FASTQ sequences, microsatellite genotypes, and associated sample ID files can be accessed on dryad at: doi:10.5061/dryad.2pk34.

ACKNOWLEDGEMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada, Bruce Power, the Canada Research Chairs Program, the Canada Foundation for Innovation, and the University of Regina. Dr. McArthur holds a Cisco Research Chair in Bioinformatics, supported by Cisco Systems Canada, Inc. Computer resources were supplied by the McMaster Service Lab and Repository (MLSR) computing cluster, funded in part by the

637 Canadian Foundation for Innovation Grants to Dr. McArthur. We are grateful to the following
638 people for providing samples: R.L. Eberts, C.L. Fietsch, A. Van Ogtrop, J. Hoyle, S. McNevin,
639 F. Fischer, E. Berglund, D. Montgomery, J. Deeg, A. Nyman, D. Niskanen, T. Maedel, C.
640 Watson, R. Avis, D. Schindelholz, S. Hansen, P. Hirethota, A. López, G.R. McCracken, D.E.
641 Ruzzante, R. Perry, D. Keefe, B. Green, S. Avery, L. Pike, O. Barker, B. Delling, P. Mortensen,
642 C. Gallagher, the University of Alaska Museum, the Yukon Department of the Environment, and
643 the Swedish Natural History Museum. M.E. Cybulski, C.M. Mitz, and J.A. Martino provided
644 technical support. B.A. Lago and B.M. Dave assisted with bioinformatics analysis of nextRAD
645 data. We are grateful to T. Gosselin for advice on analyses of missing data, and the Associate
646 Editor and two anonymous reviewers for helpful comments on an earlier version of this work.
647 This research was partially supported by the U.S. Department of Energy under Award Number
648 DE--FC09--07SR22506 to the University of Georgia Research Foundation.

LITERATURE CITED

Afgan, E., Baker, D. van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., and Goecks, J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res **44**(W1): W3-W10. 10.1093/nar/gkw343.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Babraham Bioinformatics, Cambridge. Available at www.bioinformatics.babraham.ac.uk/ projects/fastqc.

April, J., Mayden, R., Hanner, R., and Bernatchez, L. 2011. Genetic calibration of species diversity among North America 's freshwater fishes. Proc. Natl. Acad. Sci. **108**: 10602–10607. doi:10.1073/pnas.1016437108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1016437108.

April, J., Hanner, R.H., Dion-Côté, A.-M., and Bernatchez, L. 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. Mol. Ecol. **22**(2): 409–22. doi:10.1111/mec.12116.

Bailey, M.M. 1963. Age, growth, and maturity of round whitefish of the Apostle Islands and Isle Royale regions, Lake Superior. Fish. Bull. **63**(1): 63–75.

Ban, N.C., Januchowski-Hartley, S., Alvarez-Romero, J., Mills, M., Pressey, R.L., Linke, S., and de Freitas, D. 2013. Marine and freshwater conservation planning: from representation to persistence. *In* Conservation Planning: Shaping the Future. ESRI Press.

Benke, A.C., and Cushing, C.E. 2005. Rivers of North America. Elsevier, Amsterdam. 1144 pp.

671 Bernard, A.M., Ferguson, M.M., Noakes, D.L.G., Morrison, B.J., and Wilson, C.C. 2009. How
672 different is different? Defining management and conservation units for a problematic
673 exploited species. Can. J. Fish. Aquat. Sci. **66**: 1617–1630. doi:10.1139/F09-106.
674 Bernatchez, L., and Wilson, C.C. 1998. Comparative phylogeography of Nearctic and Palearctic
675 fishes. Mol. Ecol. **7**(4): 431–452. doi:10.1046/j.1365-294x.1998.00319.x.
676 Blanchfield, P.J., Taylor, E.B., and Watkinson, D.A. 2014. Morphological and genetic analyses
677 identify a new record of a glacial relict: pygmy whitefish (*Prosopium coulterii*) from
678 northwestern Ontario. **271**: 267–271.
679 Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina
680 sequence data. Bioinformatics **30**(15): 2114–2120. doi:10.1093/bioinformatics/btu170.
681 Breiman, L. 2001. Random forests. Machine Learning. **45**:5-32. doi:10.1023/A:1010933404324.
682 Brownrigg, R., Minka, T.P., Deckmyn, A. 2016. 'maps': Draw Geographical Maps. R package
683 version 3.1.1.
684 Bryson, R.W., Savary, W.E., Zellmer, A.J., Bury, R.B., and McCormack, J.E. 2016. Genomic
685 data reveal ancient microendemism in forest scorpions across the California Floristic
686 Province. Mol. Ecol. **25**(15): 3731–3751. doi:10.1111/mec.13707.
687 Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. 2013. Stacks: an
688 analysis tool set for population genomics. Mol. Ecol. **22**(11): 3124–40.
689 doi:10.1111/mec.12354.
690 Chhatre, V.E. and Emerson, K.J. 2016. StrAuto - automation and parallelization of
691 STRUCTURE analysis. http://strauto.popgen.org
692 Clement, M., Posada, D., and Crandall, K.A. 2000. TCS: A computer program to estimate gene
693 genealogies. Mol. Ecol. **9**(10): 1657–1659. doi:10.1046/j.1365-294X.2000.01020.x.

694 Delling, B., Palm, S., Palkopoulou, E., and Prestegaard, T. 2014. Genetic signs of multiple
695 colonization events in Baltic ciscoes with radiation into sympatric spring- and autumn-
696 spawners confined to early postglacial arrival. Ecol. Evol.: **4**(22): 4346-4360.
697 doi:10.1002/ece3.1299.
698 Dizon, A.E., Lockyer, C., Perrin, W.F., Demaster, D.P., and Sisson, J. 1992. Rethinking the
699 Stock Concept: A Phylogeographic Approach. Conserv. Biol. **6**(1): 24–36.
700 doi:10.1016/0006-3207(93)90670-V.
701 Earl, D.A., and vonHoldt, B.M. 2012. STRUCTURE HARVESTER: A website and program for
702 visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet.
703 Resour. **4**(2): 359–361. doi:10.1007/s12686-011-9548-7.
704 Ebener, M.P. 2012. Status of whitefish and ciscoes. *In* The state of Lake Huron in 2010. Edited
705 by S.C. Riley. Great Lakes Fish. Comm. Spec. Pub. 13-01, pp. 29-35.
706 Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E., and
707 Holzapfel, C.M. 2010. Resolving postglacial phylogeography using high-throughput
708 sequencing. Proc. Natl. Acad. Sci. U. S. A. **107**(37): 16196–16200.
709 doi:10.1073/pnas.1006538107.
710 Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals
711 using the software STRUCTURE: A simulation study. Mol. Ecol. **14**(8): 2611–2620.
712 doi:10.1111/j.1365-294X.2005.02553.x.
713 Excoffier, L., Smouse, P.E., and Quattro, J.M. 1992. Analysis of molecular variance inferred
714 from metric distances among DNA haplotypes: Application to human mitochondrial DNA
715 restriction data. Genetics **131**(2): 479–491. doi:10.1007/s00424-009-0730-7.
716 Food and Agriculture Organization of the United Nations. 2011. Review of the state of the world

717 fishery resources: inland fisheries. Document prepared by R. Welcomme FAO Fisheries and
718 Aquaculture Circular No. 942, Rev. 2. FIRF/C942, Rev. 2. Accessed at
719 http://www.fao.org/fishery/ecosystems/inland/en#container
720 Foote, C.J., Clayton, J.W., Lindsey, C.C., and Bodaly, R.A. 1992. Evolution of Lake Whitefish
721 (*Coregonus clupeaformis*) in North America during the Pleistocene: Evidence for a Nahanni
722 Glacial Refuge Race in the Northern Cordillera region. Can. J. Fish. Aquat. Sci. **49**: 760–
723 768.
724 Funk, W.C., McKay, J.K., Hohenlohe, P.A., and Allendorf, F.W. 2012. Harnessing genomics for
725 delineating conservation units. Trends Ecol. Evol. **27**(9): 489–496. Elsevier Ltd.
726 doi:10.1016/j.tree.2012.05.012.
727 Garner, B.A., Hand, B.K., Amish, S.J., Bernatchez, L., Foster, J.T., Miller, K.M., Morin, P.A.,
728 Narum, S.R., O'Brien, S.J., Roffler, G., Templin, W.D., Sunnucks, P., Strait, J., Warheit,
729 K.I., Seamons, T.R., Wenburg, J., Olsen, J., and Luikart, G. 2016. Genomics in
730 Conservation: Case Studies and Bridging the Gap between Data and Application. Trends
731 Ecol. Evol. **31**(2): 81–82. doi:10.1016/j.tree.2015.10.009.
732 Global Biodiversity Information Facility Secretariat: GBIF Backbone
733 Taxonomy. doi:10.15468/39omei Accessed via http://www.gbif.org/species/2351534 on
734 2016-07-22.
735 Gosselin, T., and Bernatchez, L. (2016). stackr: GBS/RAD Data Exploration, Manipulation and
736 Visualization using R. R package version 0.3.9. https://github.com/thierrygosselin/stackr.
737 doi : 10.5281/zenodo.53832.
738 Graham, C.F., Eberts, R.L., Morgan, T.D., Boreham, D.R., Lance, S.L., Manzon, R.G., Martino,
739 J.A., Rogers, S.M., Wilson, J.Y., and Somers, C.M. 2016. Fine-Scale Ecological and

Genetic Population Structure of Two Whitefish (Coregoninae) Species in the Vicinity of Industrial Thermal Emissions. PLoS One **11**(1): e0146656. doi:10.1371/journal.pone.0146656.

Haponski, A.E., and Stepien, C.A. 2014. Genetic connectivity and diversity of walleye (Sander vitreus) spawning groups in the Huron-Erie Corridor. J. Great Lakes Res. **40**(2): 89–100. International Association for Great Lakes Research. doi:10.1016/j.jglr.2012.12.006.

Ishwaran, H., Kogalur, U.B., Blackstone, E.H., and Lauer, M.S. 2008. Random survival forests for R, Ann. App. Statist. **2**: 841-860.

Jakobsson, M., and Rosenberg, N.A. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23**(14): 1801–1806. doi:10.1093/bioinformatics/btm233.

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

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.

Kalinowski, S.T. 2005. HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. Mol. Ecol. Notes. **5**: 187-189. doi: 10.1111/j.1471-8286.2004.00845.x

Kocovsky, P.M., Sullivan, T.J., Knight, C.T., and Stepien, C.A. 2013. Genetic and morphometric differences demonstrate fine-scale population substructure of the yellow perch Perca flavescens: Need for redefined management units. J. Fish Biol. **82**(6): 2015–2030. doi:10.1111/jfb.12129.

762 Kohler, C.C., and Hubert, W.A. 1999. The Great Lakes. *In* Inland Fisheries Management in
763 North America.
764 Leach, J.H., and Nepszy, S.J. 1976. The Fish Community in Lake Erie. J. Fish. Res. Board
765 Canada **33**: 622–638.
766 Lindsey, C.C., Patalas, K., Bodaly, R.A., and Archibald, C.P. 1981. Glaciation and the physical,
767 chemical and biological limnology of Yukon lakes. Can. Tech. Rep. Fish. Aquat. Sci. No.
768 966.
769 Mandrak, N.E., and Crossman, E.J. 1992. Postglacial dispersal of freshwater fishes into Ontario.
770 Can. J. Zool. **70**(11): 2247–2259. doi:10.1139/z92-302.
771 Marine, R., Polson, S.W., Ravel, J., Hatfull, G., Russell, D., Sullivan, M., Syed, F., Dumas, M.,
772 and Wommack, K.E. 2011. Evaluation of a transposase protocol for rapid generation of
773 shotgun high-throughput sequencing libraries from nanogram quantities of DNA. Appl.
774 Environ. Microbiol. **77**(22): 8071–8079. doi:10.1128/AEM.05610-11.
775 McDermid, J.L., Wozney, K.M., Kjartanson, S.L., and Wilson, C.C. 2011. Quantifying
776 historical, contemporary, and anthropogenic influences on the genetic structure and diversity
777 of lake sturgeon (*Acipenser fulvescens*) populations in northern Ontario. J. Appl. Ichthyol.
778 **27**(SUPPL. 2): 12–23. doi:10.1111/j.1439-0426.2011.01825.x.
779 McPhail, J.D., and Lindsey, C.C. 1970. Freshwater fishes of northwestern Canada and Alaska.
780 Bull. Fish. Res. Board Can. No. 173.
781 Mee, J.A., Bernatchez, L., Reist, J.D., Rogers, S.M., and Taylor, E.B. 2015. Identifying
782 designatable units for intraspecific conservation prioritization: a hierarchical approach
783 applied to the lake whitefish species complex (*Coregonus spp.*). Evol. Appl. **8**(5): 423–441.
784 doi:10.1111/eva.12247.

785 Meirmans, P.G., and Van Tienderen, P.H. 2004. GENOTYPE and GENODIVE: Two programs
786 for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes **4**(4): 792–794.
787 doi:10.1111/j.1471-8286.2004.00770.x.
788 Miller, M.A., Pfeiffer, W., and Schwartz, T. 2010. Creating the CIPRES Science Gateway for
789 inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments
790 Workshop (GCE), **14**: 1-8.
791 Moritz, C. 1994. Defining "evolutionarily significant units" for conservation. Trends Ecol. Evol.
792 **9**(10): 373–375. doi:10.1016/0169-5347(94)90057-4.
793 Mraz, D. 1964. Age and growth of the round whitefish in Lake Michigan. Trans. Am. Fish.
794 **93**(3): 215–226. doi:10.1577/1548-8659(1964)93.
795 Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., and Hohenlohe, P. A. 2013. Genotyping-
796 by-sequencing in ecological and conservation genomics. Mol. Ecol. **22**(11): 2841–2847.
797 doi:10.1111/mec.12350.
798 Nugent, B., and Carpenter, M. 2015. New Hampshire Wildlife Action Plan Appendix A Fish-146
799 - round whitefish. Available at http://www.wildlife.state.nh.us/wildlife/profiles/wap/fish-
800 roundwhitefish.pdf. Accessed June 2016.
801 O'Bryhim, J., Somers, C., Lance, S.L., Yau, M., Boreham, D.R., Jones, K.L., and Taylor, E.B.
802 2013. Development and characterization of twenty-two novel microsatellite markers for the
803 mountain whitefish, *Prosopium williamsoni* and cross-amplification in the round whitefish,
804 *P. cylindraceum*, using paired-end Illumina shotgun sequencing. Conserv. Genet. Resour.
805 **5**(1): 89–91. doi:10.1007/s12686-012-9740-4.
806 Ontario Commercial Fisheries Association. Commercial catch data 1994-2013. Accessed June
807 2016 http://www.ocfa.ca/fisheries-industry/fisheries-statistics.

808 Overdyk, L.M., Braid, H.E., Crawford, S.S., and Hanner, R.H. 2015. Extending DNA barcoding
809 coverage for Lake Whitefish (Coregonus clupeaformis) across the three major basins of
810 Lake Huron. DNA Barcodes **3**(1): 59–65. doi:10.1515/dna-2015-0007.
811 Palsbøll, P.J., Bérubé, M., and Allendorf, F.W. 2007. Identification of management units using
812 population genetic data. Trends Ecol. Evol. **22**(1): 11–16. doi:10.1016/j.tree.2006.09.003.
813 Patrick, P.H., Chen, E., Parks, J., Powell, J., Poulton, J.S., and Fietsch, C.-L. 2013. Effects of
814 Fixed and Fluctuating Temperature on Hatch of Round Whitefish and Lake Whitefish Eggs.
815 North Am. J. Fish. Manag. **33**(6): 1091–1099. doi:10.1080/02755947.2013.824937.
816 Pielou E.C. 1991. After the ice age the return of life to glaciated North America. University of
817 Chicago Press, Chicago.
818 Poff, N.L.R., Allan, J.D., Palmer, M.A., Hart, D.D., Richter, B.D., Arthington, A.H., Rogers,
819 K.H., Meyer, J.L., and Stanford, J.A. 2003. River flows and water wars: emerging science
820 for environmental decision making. Front. Ecol. Environ. **1**(6): 298–306. doi:10.1890/1540-
821 9295(2003)001[0298:rfawwe]2.0.co;2.
822 Pringle, C. 2003. What is hydrologic connectivity and why is it ecologically important? Hydrol.
823 Process. **17**(13): 2685–2689. doi:10.1002/hyp.5145.
824 Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using
825 multilocus genotype data. Genetics **155**(2): 945–959. doi:10.1111/j.1471-
826 8286.2007.01758.x.
827 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J.,
828 Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. 2007. PLINK: A tool set for whole-
829 genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**(3): 559–
830 575. doi:10.1086/519795.

831 Ratnasingham, S., and Hebert, P.D.N. 2007. BOLD : The Barcode of Life Data System
832 (www.barcodinglife.org). Mol. Ecol. Notes **7**: 355–364. doi:10.1111/j.1471-
833 8286.2006.01678.x.
834 Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B.,
835 Liu, L., Suchard, M.A., and Huelsenbeck, J.P. 2012. Mrbayes 3.2: Efficient bayesian
836 phylogenetic inference and model choice across a large model space. Syst. Biol. **61**(3): 539–
837 542. doi:10.1093/sysbio/sys029.
838 Ross, S.T. 2013. Ecology of North American freshwater fishes. University of California Press,
839 Berkeley, CA, USA.
840 Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for
841 Windows and Linux. Mol. Ecol. Resour. **8**(1): 103–106. doi:10.1111/j.1471-
842 8286.2007.01931.x.
843 Rosenberg, N.A. 2004. Distruct: a Program for the Graphical Display of Population Structure.
844 Mol. Ecol. Notes **4**(1): 137–138. doi:10.1046/j.1471-8286.2003.00566.x.
845 Russello, M.A., Waterhouse, M.D., Etter, P.D., and Johnson, E.A. 2015. From promise to
846 practice: pairing non-invasive sampling with genomics in conservation. PeerJ **3**: e1106.
847 doi:10.7717/peerj.1106.
848 Ryder, O.A. 1986. Species conservation and systematics: the dilemma of subspecies. Trends
849 Ecol. Evol. **1**(1): 9–10. doi:10.1016/0169-5347(86)90059-5.
850 Schafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Brännström, I.,
851 Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fior, S., Hajibabaei, M., Hill, J.A.,
852 Hoezel, A.R., Hoglund, J., Jensen, E.L., Krause, J., Kristensen, T.N., Krützen, M., McKay,
853 J.K., Norman, A.J., Ogden, R., Österling, E.M., Ouborg, N.J., Piccolo, J., Popovíc, D.,

854 Primmer, C.., Reed, F.A., Roumet, M., Salmona, J., Schenekar, T., Schwartz, M.K.,
855 Segelbacher, G., Senn, H., Thaulow, J., Valtonen, M., Veale, A., Vijay, N., Weissensteiner,
856 M., and Wheat, C.W. 2015. Genomics and the challenging translation into conservation
857 practice. Trends Ecol. Evol. **30**: 78–87.
858 Schafer, A.B.A., Wolf, J.B.W., Alves, P.C., Berström, L., Colling, G., Dalén, L., De Meester, L.,
859 Ekblom, R., Fior, S., Hajibabaei, M., Hoezel, A.R., Hoglund, J., Jensen, E.L, Krützen, M.,
860 Norman, A.J., Österling, E.M., Ouborg, N.J., Piccolo, J., Primmer, C.R., Reed, F.A.,
861 Roumet, M., Salmona, J., Schwartz, M.K., Segelbacher, G., Thaulow, J., Valtonen, M.,
862 Vergeer, P., Weissensteiner, M., Wheat, C.W., Vilà, C., and Zieliński, P. 2016. Reply to
863 Garner et al. Trends Ecol. Evol. **31**(2): 83. doi:10.1016/j.tree.2015.10.009.
864 Scott, W.B., and Crossman, E.J. 1973. Freshwater fishes of Canada. Bull. Fish. Res. Board Can.
865 No. 184.
866 Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of
867 large phylogenies. Bioinformatics **30**(9): 1312–1313. doi:10.1093/bioinformatics/btu033.
868 Stamford, M.D., and Taylor, E.B. 2004. Phylogeographical lineages of Arctic grayling
869 (*Thymallus arcticus*) in North America: divergence, origins and affinities with Eurasian
870 Thymallus. Mol. Ecol. **13**(6): 1533–1549. doi:10.1111/j.1365-294X.2004.02174.x.
871 Steinhart, G.B., Mineau, M., Kraft, C.E., and Hall, F. 2007. Status and recovery of round
872 whitefish (Prosopium cylindraceum) in New York , USA. Available at
873 https://www.researchgate.net/publication/228990773_Status_and_recovery_of_round_white
874 fish_Prosopium_cylindraceum_in_New_York_USA.
875 Stepien, C.A., Murphy, D.J., and Strange, R.M. 2007. Broad- to fine-scale population genetic
876 patterning in the smallmouth bass *Micropterus dolomieu* across the Laurentian Great Lakes

877 and beyond: An interplay of behaviour and geography. Mol. Ecol. **16**(8): 1605–1624.
878 doi:10.1111/j.1365-294X.2006.03168.x.
879 Stepien, C.A., Murphy, D.J., Lohner, R.N., Sepulveda-Villet, O.J., and Haponski, A.E. 2009.
880 Signatures of vicariance, postglacial dispersal and spawning philopatry: population genetics
881 of the walleye Sander vitreus. Mol. Ecol. **18**(16): 3411–28. doi:10.1111/j.1365-
882 294X.2009.04291.x.
883 Stewart, K.W., and Lindsey, C.C. 1983. Postglacial dispersal of lower vertebrates in the Lake
884 Agassiz region. *In* Glacial Lake Agassiz. Edited by J.T. Teller and L. Clayton. Geological
885 Association of Canada Special Paper 26, 1983.
886 Stott, W., VanDeHey, J. A., and Sloss, B.L. 2010. Genetic diversity of lake whitefish in lakes
887 Michigan and Huron; sampling, standardization, and research priorities. J. Great Lakes Res.
888 **36**: 59–65. doi:10.1016/j.jglr.2010.01.004.
889 Sullivan, T.J., and Stepien, C.A. 2014. Genetic diversity and divergence of yellow perch
890 spawning populations across the Huron-Erie Corridor, from Lake Huron through western
891 Lake Erie. J. Great Lakes Res. **40**(2): 101–109. Elsevier B.V.
892 doi:10.1016/j.jglr.2012.12.004.
893 Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).
894 Version 4. Sinauer Associates, Sunderland, Massachusetts.
895 Tamkee, P., Parkinson, E., and Taylor, E.B. 2010. The influence of Wisconsonian glaciation and
896 contemporary stream hydrology on microsatellite DNA variation in rainbow trout
897 (*Oncorhynchus mykiss*). Can. J. Fish. Aquat. Sci. **67**: 919–935. doi:doi:10.1139/F10-034.
898 Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular
899 evolutionary genetics analysis version 6.0. Mol. Biol. Evol. **30**(12): 2725–2729.

doi:10.1093/molbev/mst197Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., and Shipley, P. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes **4**(3): 535–538. doi:10.1111/j.1471-8286.2004.00684.x.

Teller, J.T., and Clayton L. 1983. Glacial Lake Agassiz. Edited by J.T. Teller and L. Clayton. Geological Association of Canada Special Paper 26, 1983.

VanDeHey, J. A., Sloss, B.L., Peeters, P.J., and Sutton, T.M. 2009. Genetic structure of lake whitefish (*Coregonus clupeaformis*) in Lake Michigan. Can. J. Fish. Aquat. Sci. **66**(3): 382–393. doi:10.1139/F08-213.

Vermont Department of Fish and Wildlife. 2015. Fish SGCN conservation report – Appendix A3. Accessed May 2016 http://vtfishandwildlife.com.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., and Seehausen, O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. **22**(3): 787–798. doi:10.1111/mec.12023.

Ward, R.D., Hanner, R., and Hebert, P.D.N. 2009. The campaign to DNA barcode all fishes, FISH-BOL. J. Fish Biol. **74**(2): 329–56. doi:10.1111/j.1095-8649.2008.02080.x.

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

Weir, B.S. and Cockerham, C.C. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution **38**(6):1358-1370.

922 Wilson, C.C., and Hebert, P.D.N. 1998. Phylogeography and postglacial dispersal of lake trout
923 (*Salvelinus namaycush*) in North America. Can. J. Fish. Aquat. Sci. **55**(4): 1010–1024.
924 doi:10.1139/f97-286.
925 Wilson, G.A., and Rannala, B. 2003. Bayesian inference of recent migration rates using
926 multilocus genotypes. Genetics **163**(3): 1177–1191.
927 Witt, J.D.S., Zemlak, R.J., and Taylor, E.B. 2011. Phylogeography and the origins of range
928 disjunctions in a north temperate fish, the pygmy whitefish (Prosopium coulterii), inferred
929 from mitochondrial and nuclear DNA sequence analysis. J. Biogeogr. **38**(8): 1557–1569.
930 doi:10.1111/j.1365-2699.2011.02497.x.
931 Wood, J., Hoyle, J.A., Wilson, C.C. 2016. Genetic analysis of round whitefish (*Prosopium*
932 *cylindraceum*) stock structure in Lake Ontario. Ontario Ministry of Natural Resources and
933 Forestry Technical Report TR-11.
934 Yamamoto, S., Maekawa, K., Morita, K., Crane, P.A., and Oleinik, A.G. 2014. Phylogeography
935 of the salmonid fish, Dolly Varden Salvelinus malma: multiple glacial refugia in the North
936 Pacific Rim. Zoolog. Sci. **31**(10): 660–670. doi:10.2108/zs130266.
937

938 **Table 1:** Source information for round whitefish samples from 16 sites across North America and
939 one site in eastern Russia. Information includes site name, GPS coordinates, number of total
940 samples (N), the year collected, the number of individuals included in mtDNA analyses
941 (mtDNA), microsatellite analyses (microsatellites), nextRAD analyses (nextRAD), and
942 additional notes on source including catalog information for museum specimens.

Location	GPS coordinates	N	Year	MtDNA	Microsatellites	NextRAD	Source
North Alaska (N-AK)	N70°00'43", W153°09'11"	8	2014	8	8	8	University of Alaska Museum Ichthyology collection – catalog number 8068
South Alaska (S-AK)	N59°19'16", W156°19'07"	16	2014	8	15	16	University of Alaska Museum Ichthyology collection – catalog number 9114 & 9136
Yukon – Bennett Lake (YK-Ben)	N60°03'23", W156°33'50" N60°04'20", W134°52'26"	20	2014	9	18	8	Environment Yukon – Jul 2014
Yukon – Little Salmon Lake (YK-LSa)	N62°11'10", W134°42'05"	20	2015	3	19	3	Environment Yukon – Jul-Aug 2015
Yukon – Simpson Lake (YK-Sim)	N60°43'28", W129°14'35"	20	2014	8	19	10	Environment Yukon – Jun 2014
Lake Huron (LHU)	N44°23'56", W81°31'51" N44°22'31", W81°33'21" N44°21'22", W81°35'10" N44°20'25", W81°35'32" N44°17'51", W81°36'33" N44°16'54", W81°36'18" N44°15'44" W81°36'52" N44°23'17", W81°31'57"	61	2010-2012	9	60	27	This study
Northern Georgian Bay (NGB)	N45°56'58", W81°30'10"	20	2014	8	19	11	This study
Southern Georgian Bay (SGB)	N44°34'19", W80°04'41" N44°31'00", W80°06'18"	27	2014	7	27	9	This study
Lake Ontario (LON)	N43°51'53", W78°44'10" N43°48'08", W79°03'15"	62	2012, 2014	7	60	32	7 sites within 4 km of this point 6 sites within 1 km of this point This study Dec 2012 and Ontario Ministry of Natural Resources and Forestry – Nov-Dec 2014
Lake Michigan – Milwaukee (LMI-M)	N42°59'11", W87°49'25"	36	2015	8	36	16	Wisconsin Department of Natural Resources – Jun 2015
Lake Michigan - Door County (LMI-D)	N45°06'20", W87°02'46" N45°00'18", W87°07'00" N44°54'08", W87°11'46"	30	2015	0	30	0	Wisconsin Department of Natural Resources – Nov 2015
Lake Nipigon (LNI)	N50°00'24", W88°54'29" N49°53'24", W88°57'57"	15	2015	8	14	15	Ontario Ministry of Natural Resources and Forestry – Sept 2015
Lake Superior (LSU)	N48°23'45", W89°02'02" N48°50'05", W88°06'19" N48°44'36", W86°28'53"	41	2015	12	40	16	Sites within 120 km of each other. Tested for differentiation between sites before being combined; Ontario Ministry of Natural Resources and Forestry – Sept 2015

Labrador – T-Bone Lake (LAB)	N56°09'10", W63°56'21"	25	2010	9	25	16	Dalhousie University – from T-Bone Lake. The system is described in McCracken et al. 2013
Russia – Taniorer River (RUS)	N66°09'17", E175°45'44"	3	2005	3	0	3	Swedish Museum of Natural History Ichthyology collection – catalog numbers NRM52850, NRM57539, NRM57540
Northwest Territories – Rat River (NTRR)	N67°44'57", W136°17'10" N68°17'56", W136°21'20"	5	2013	4	0	0	Department of Fisheries and Oceans
Great Bear Lake (GBL)	N65°08'08", W123°14'40"	10	2012	9	0	0	Department of Fisheries and Oceans

943 **Table 2:** Prevalence of 13 mtDNA haplotypes for n = 124 round whitefish from 15 sites across North America and one site in eastern
944 Russia. Site abbreviations NAK = north Alaska, SAK = south Alaska, YKBen = Bennett Lake, Yukon, YKLSa = Little Salmon Lake,
945 Yukon, YKSim = Simpson Lake, Yukon, LHU = Lake Huron main basin, NGB = north Georgian Bay, SGB = south Georgian Bay,
946 LON = Lake Ontario, LMI-M = Lake Michigan – Milwaukee, LMI-D = Lake Michigan – Door County, LNI = Lake Nipigon, LSU =
947 Lake Superior, LAB = Labrador.

948

		WEST							EAST							
Haplotype	**N**	**RUS**	**NAK**	**SAK**	**YK-Ben**	**YK-LSa**	**YK-Sim**	**NTRR**	**GBL**	**LHU**	**NGB**	**SGB**	**LON**	**LMI**	**LNI**	**LSU**	**LAB**
RWF-E1	2	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.
RWF-E2	2	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.
RWF-E3	30	.	.	.	.	.	.	.	.	3	7	6	.	5	2	7	.
RWF-E4	2	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.
RWF-E5	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
RWF-E6	31	.	.	.	.	.	.	.	.	3	.	.	7	1	6	5	9
RWF-W1	10	.	7	.	1	2	.	.	.	.	.	.	.	.	.	.	.
RWF-W2	19	.	.	7	.	.	8	2	2	.	.	.	.	.	.	.	.
RWF-W3	14	.	1	1	7	5	.	.	.	.	.	.	.	.	.	.	.
RWF-W4	3	.	.	.	1	.	.	2		.	.	.	.	.	.	.	.
RWF-W5	7	.	.	.	.	.	.	.	7	.	.	.	.	.	.	.	.
RWF-R1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RWF-R2	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

949

950

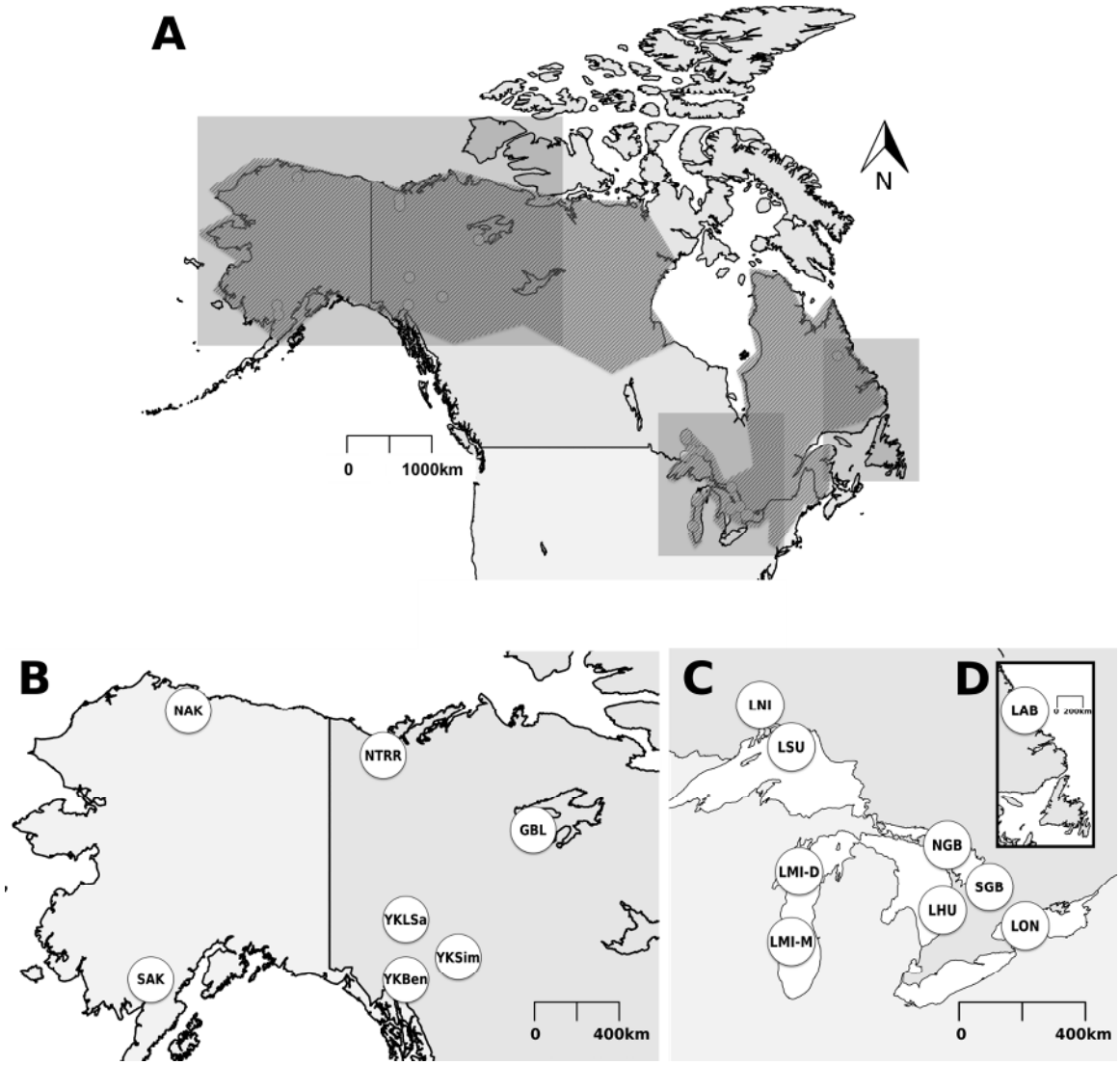
Figure 1: Map of round whitefish range (dark grey; Global Biodiversity Information Facility data) and sampling sites in A: North America, B: Alaska and Yukon regions, C: Great Lakes region, and D: Labrador. Shaded boxes in A indicate areas enlarged in B, C, and D. Site abbreviations NAK = north Alaska, SAK = south Alaska, YKBen = Bennett Lake, Yukon, YKLSa = Little Salmon Lake, Yukon, YKSim = Simpson Lake, Yukon, LHU = Lake Huron main basin, NGB = north Georgian Bay, SGB = south Georgian Bay, LON = Lake Ontario, LMI-M = Lake Michigan – Milwaukee, LMI-D = Lake Michigan – Door County, LNI = Lake Nipigon, LSU = Lake Superior, LAB = Labrador. Maps were generated in R using the 'maps' package (Brownrigg et al. 2016).

Figure 2: Phylogenetic tree for concatenated D-loop-COI sequences (354 and 578 nucleotides) for round whitefish from 15 sites in North America and one site in eastern Russia. *E1-E6* = eastern haplotypes, *W1-W5* =western haplotypes, *R1-R2* = Russian haplotypes. Node supports shown for Maximum-Likelihood bootstrap (>50) and Bayesian values (>0.90). Scale bar showing number of substitutions per base pair.

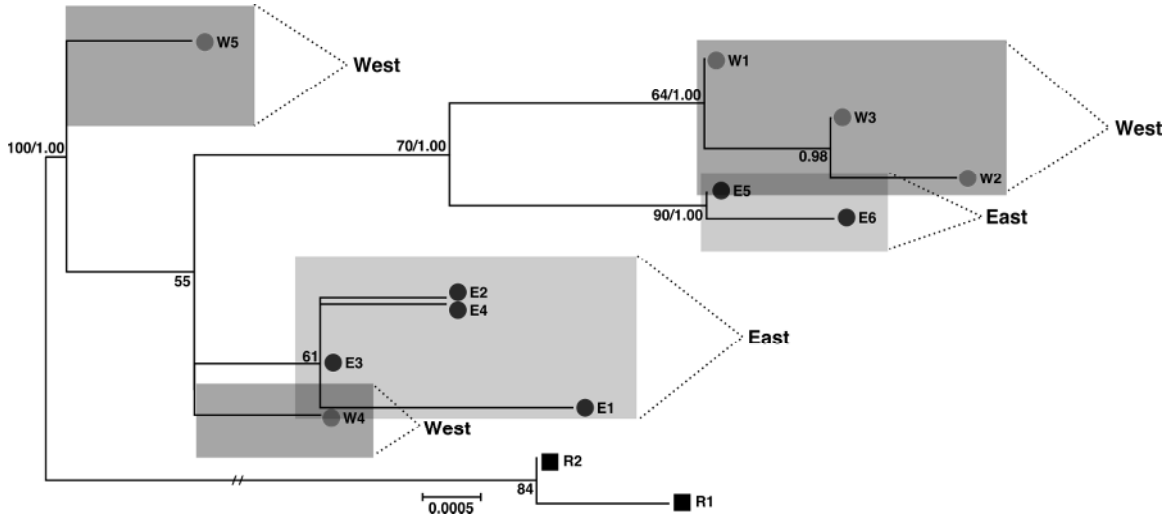
Figure 3: Phylogenetic tree for 4918 loci from 13 sites in North America and one site in eastern Russia. Node supports shown for Maximum-Likelihood bootstrap values (>50) and Bayesian support (>0.90). NAK = north Alaska, SAK = south Alaska, YKBen = Bennett Lake, Yukon, YKLSa = Little Salmon Lake, Yukon, YKSim = Simpson Lake, Yukon, LHU = Lake Huron main basin, NGB = north Georgian Bay, SGB = south Georgian Bay, LON = Lake Ontario, LMI = Lake Michigan - Milwaukee, LNI = Lake

973 Nipigon, LSU = Lake Superior, LAB = Labrador, RUS = Russia. Scale bar showing
974 number of substitutions per base pair.
975
976 **Figure 4**: Discriminant analysis of principal components for nine round whitefish
977 microsatellite loci within A) 13 sites across North America B) Five sites in Alaska and
978 Yukon regions and C) Eight sites in the Laurentian Great Lakes region. NAK = north
979 Alaska, SAK = south Alaska, YKBen = Bennett Lake, Yukon, YKLSa = Little Salmon
980 Lake, Yukon, YKSim = Simpson Lake, Yukon, LHU = Lake Huron main basin, NGB =
981 north Georgian Bay, SGB = south Georgian Bay, LON = Lake Ontario, LMI-M = Lake
982 Michigan Milwaukee, LMI=D = Lake Michigan – Door County, LNI = Lake Nipigon,
983 LSU = Lake Superior, LAB = Labrador.
984
985 **Figure 5**: Contemporary migrations rates (proportion of the population that has migrated
986 per generation in the direction of the arrow with 95% confidence interval) as determined
987 using the program BAYESASS for seven sites in the Laurentian Great Lakes region.
988 Values are shown only for migration rates with 95% confidence not overlapping with
989 zero. LHU = Lake Huron main basin, NGB = north Georgian Bay, SGB = south
990 Georgian Bay, LON = Lake Ontario, LMI = Lake Michigan, LNI = Lake Nipigon, LSU =
991 Lake Superior. Map was generated in R using the 'maps' package (Brownrigg et al.
992 2016).
993
994 **Figure 6**: Discriminant analysis of principal components for 343, 801, and 1375 round
995 whitefish SNP loci within A) 13 sites across North America and one in Russia B) Five

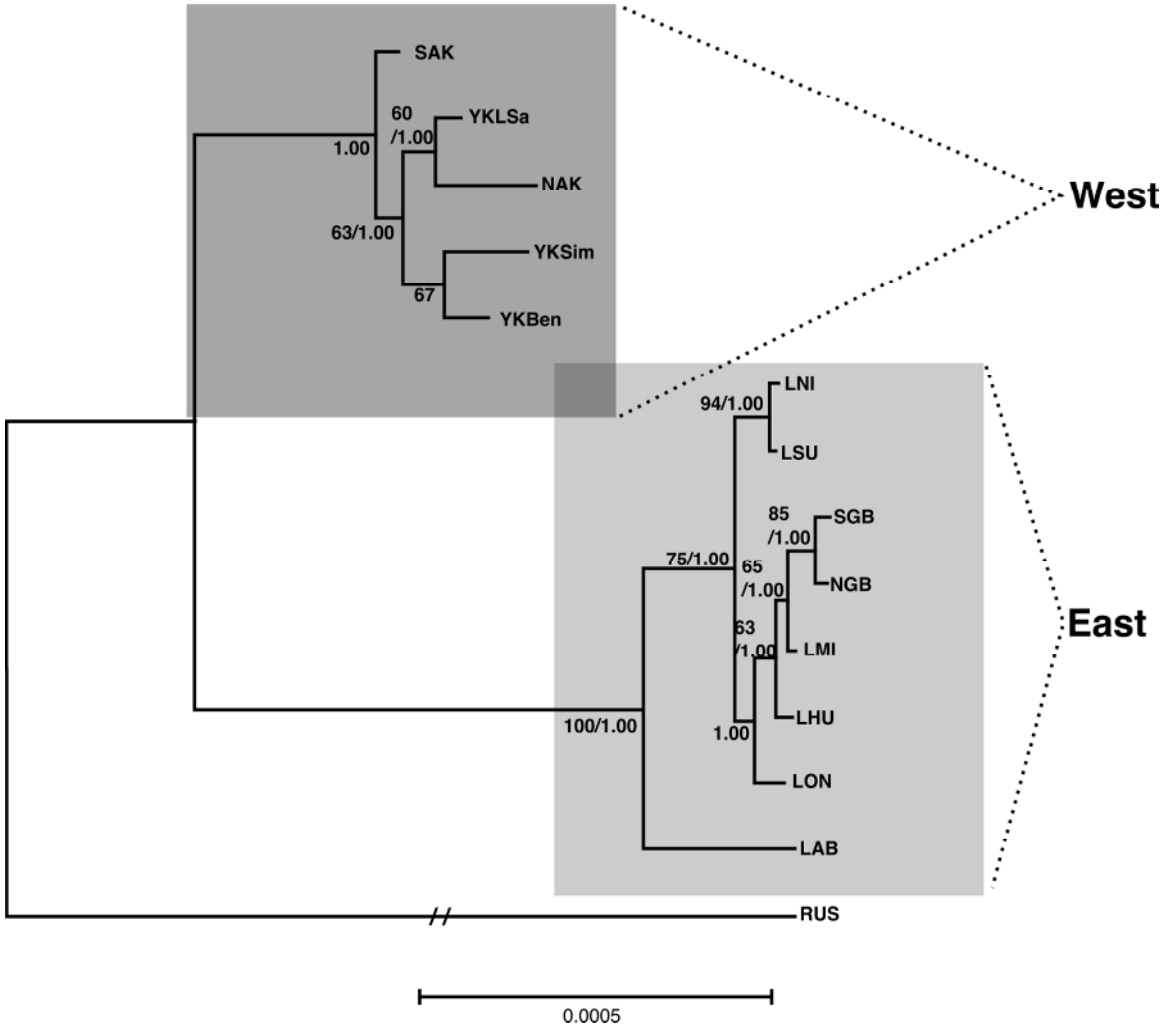
sites in Alaska and Yukon regions and C) Seven sites in the Laurentian Great Lakes region. NAK = north Alaska, SAK = south Alaska, YKBen = Bennett Lake, Yukon, YKLSa = Little Salmon Lake, Yukon, YKSim = Simpson Lake, Yukon, LHU = Lake Huron main basin, NGB = north Georgian Bay, SGB = south Georgian Bay, LON = Lake Ontario, LMI = Lake Michigan, LNI = Lake Nipigon, LSU = Lake Superior, LAB = Labrador, RUS = Russia.


A
N
0
1000km
B
NAK
NTRR
GBL
YKLSa
YKSim
YKBen
SAK
0
400km
C
LNI
LSU
NGB
LMI-D
SGB
LHU
LON
LMI-M
0
400km
D
LAB

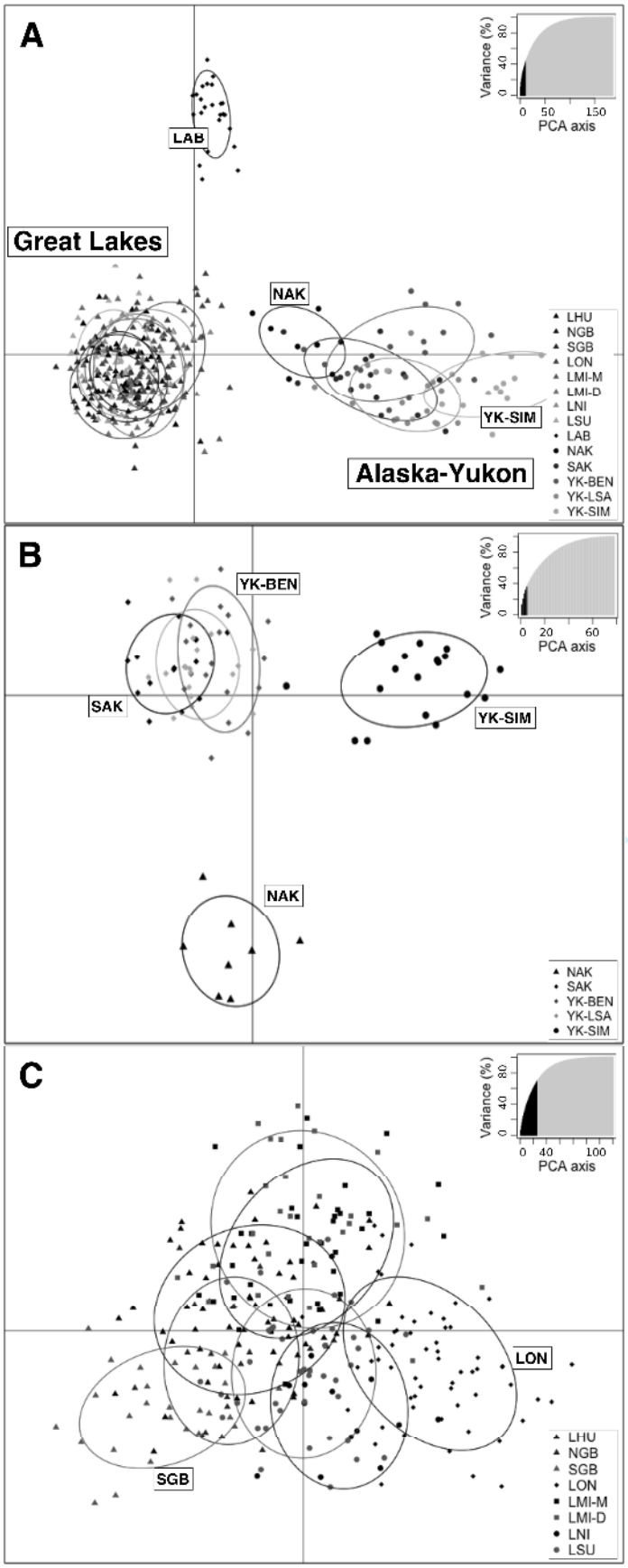
1004



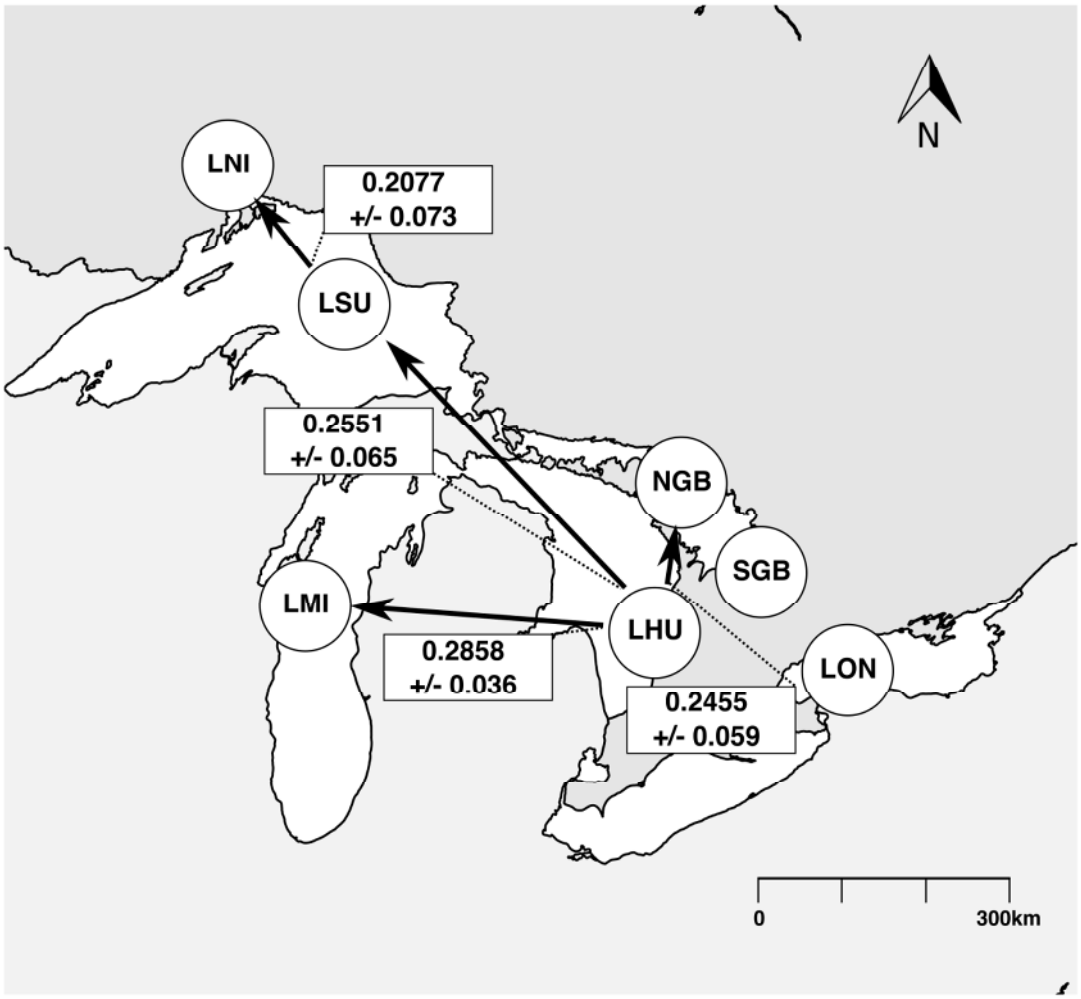
1005
1006


SAK
60
/1.00
YKLSa
1.00
NAK
63/1.00
YKSim
67
YKBen
West
LNI
94/1.00
LSU
85
/1.00
SGB
75/1.00
65
/1.00
NGB
63
/1.00
LMI
East
100/1.00
LHU
1.00
LON
LAB
RUS
0.0005

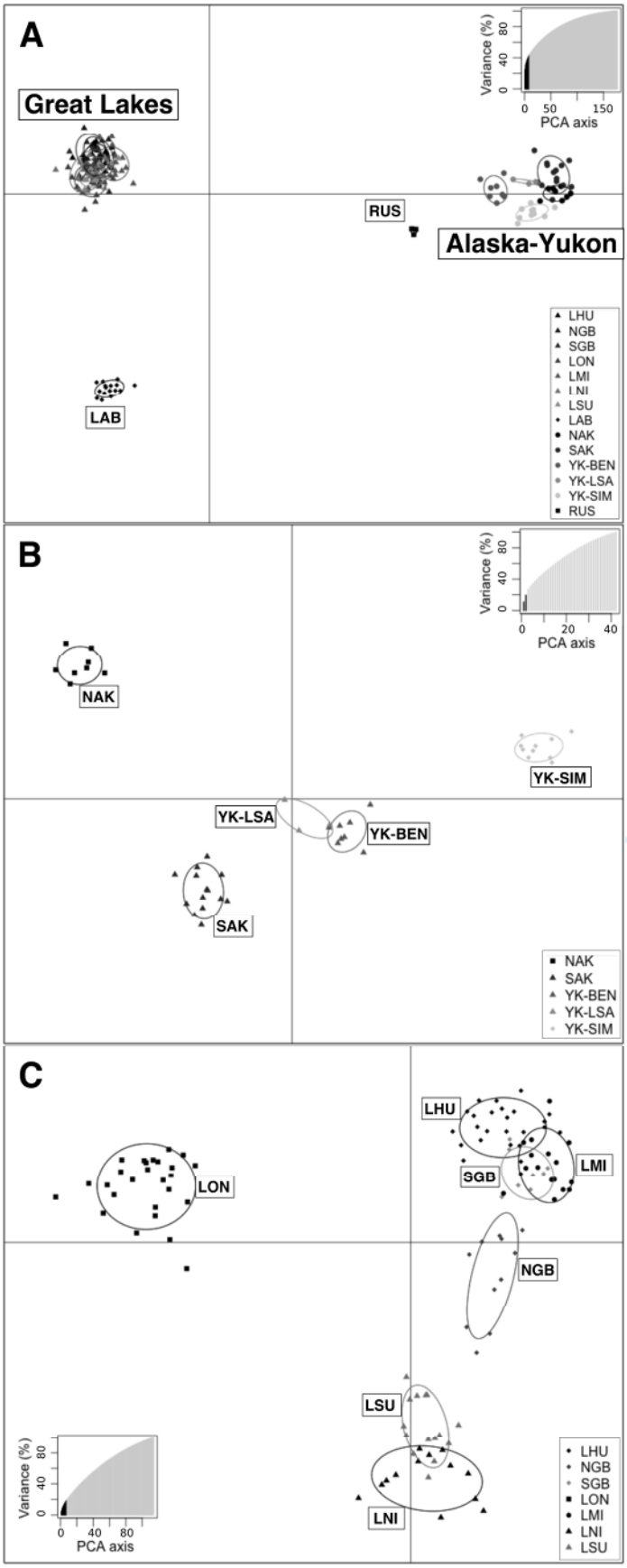
1007
1008


A
Variance (%)
0
40
80
0 50 150
PCA axis
LAB
Great Lakes
NAK
YK-SIM
Alaska-Yukon
LHU
NGB
SGB
LON
LMI-M
LMI-D
LNI
LSU
LAB
NAK
SAK
YK-BEN
YK-LSA
YK-SIM
B
Variance (%)
0
40
80
0 20 60
PCA axis
YK-BEN
SAK
YK-SIM
NAK
NAK
SAK
YK-BEN
YK-LSA
YK-SIM
C
Variance (%)
0
40
80
0 40 100
PCA axis
LON
SGB
LHU
NGB
SGB
LON
LMI-M
LMI-D
LNI
LSU

1009


N
LNI
0.2077
+/- 0.073
LSU
0.2551
+/- 0.065
NGB
SGB
LMI
LHU
0.2858
+/- 0.036
LON
0.2455
+/- 0.059
0
300km

1010
1011


A
Great Lakes
Alaska-Yukon
RUS
LAB
Variance (%)
0
40
80
0
50
150
PCA axis
LHU
NGB
SGB
LON
LMI
LNI
LSU
LAB
NAK
SAK
YK-BEN
YK-LSA
YK-SIM
RUS
B
NAK
YK-SIM
YK-LSA
YK-BEN
SAK
Variance (%)
0
40
80
0
20
40
PCA axis
NAK
SAK
YK-BEN
YK-LSA
YK-SIM
C
LON
LHU
SGB
LMI
NGB
LSU
LNI
Variance (%)
0
40
80
0
40
80
PCA axis
LHU
NGB
SGB
LON
LMI
LNI
LSU

1012