

ORIGINAL ARTICLE

Evolution and origin of sympatric shallow-water morphotypes of Lake Trout, *Salvelinus namaycush*, in Canada's Great Bear Lake

LN Harris¹, L Chavarie², R Bajno¹, KL Howland¹, SH Wiley¹, WM Tonn² and EB Taylor³

Range expansion in north-temperate fishes subsequent to the retreat of the Wisconsin glaciers has resulted in the rapid colonization of previously unexploited, heterogeneous habitats and, in many situations, secondary contact among conspecific lineages that were once previously isolated. Such ecological opportunity coupled with reduced competition likely promoted morphological and genetic differentiation within and among post-glacial fish populations. Discrete morphological forms existing in sympatry, for example, have now been described in many species, yet few studies have directly assessed the association between morphological and genetic variation. Morphotypes of Lake Trout, *Salvelinus namaycush*, are found in several large-lake systems including Great Bear Lake (GBL), Northwest Territories, Canada, where several shallow-water forms are known. Here, we assess microsatellite and mitochondrial DNA variation among four morphotypes of Lake Trout from the five distinct arms of GBL, and also from locations outside of this system to evaluate several hypotheses concerning the evolution of morphological variation in this species. Our data indicate that morphotypes of Lake Trout from GBL are genetically differentiated from one another, yet the morphotypes are still genetically more similar to one another compared with populations from outside of this system. Furthermore, our data suggest that Lake Trout colonized GBL following dispersal from a single glacial refugium (the Mississippian) and support an intra-lake model of divergence. Overall, our study provides insights into the origins of morphological and genetic variation in post-glacial populations of fishes and provides benchmarks important for monitoring Lake Trout biodiversity in a region thought to be disproportionately susceptible to impacts from climate change.

Heredity advance online publication, 10 September 2014; doi:10.1038/hdy.2014.74

INTRODUCTION

Intraspecific resource polymorphisms are widespread in numerous taxa and have had an important role in understanding the roles of phenotypic plasticity and natural selection in generating biodiversity, including the process of speciation (Skúlason and Smith, 1995; Smith and Skúlason, 1996). In particular, fishes demonstrate extensive polymorphism in morphology and resource use both in the marine and freshwater realms and from the tropics to the Arctic (Wimberger, 1994; Taylor, 1999). Often, such polymorphisms occur as discrete morphological and behavioural forms that specialize on alternative resource types in distinct habitats (for example, 'pelagic', 'benthic', 'limnetic', 'piscivorous', 'insectivorous' morphotypes or ecotypes). A key feature of the environment strongly associated with resource polymorphisms is the novel ecological opportunity, coupled with low interspecific competition, which is often provided by the emergence of new habitats or when existing habitats are altered or disturbed (Schluter, 1996).

In north-temperate regions, recolonization of newly formed freshwater habitats after the most recent glaciation, beginning about 15 000 years ago (Lindsey and McPhail, 1986; Hewitt, 1996), has apparently provided such novel ecological opportunities, resulting in the

description of a number of sympatric ecotypes in a variety of freshwater fishes (reviewed in Schluter, 1996; Taylor, 1999). In many cases, there are also examples of repeated occurrences of divergent ecotypes within a taxon that appear to have evolved in parallel (Schluter and Nagel, 1995; Pigeon *et al.*, 1997; Østbye *et al.*, 2006), and there is evidence both for evolution of ecotypes in allopatry followed by secondary contact (Bernatchez and Dodson, 1990; Fraser and Bernatchez, 2005) and for sympatric divergence (Taylor and Bentzen, 1993; Præbel *et al.*, 2013). Phenotypic plasticity, the capability of a genotype to exhibit variable phenotypes as influenced by the environment (Whitman and Agrawal, 2009) has also been important in generating morphological and ecological diversity in post-glacial habitats (Robinson and Parsons, 2002).

Where sympatric ecotypes exist, there are several hypotheses that may explain their nature and origin. First, and perhaps most fundamental, is that ecotypes within a locality represent genetically discrete populations and not plastic responses to a heterogenous environment. The most direct test of this hypothesis would involve multi-generation breeding and common garden experiments (Robinson and Parsons, 2002; Lundsgaard-Hansen *et al.*, 2013), but genetic distinctiveness can also be indirectly examined by assessing the

¹Arctic Aquatic Research Division, Fisheries and Oceans Canada, Winnipeg, Manitoba, Canada; ²Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada and ³Department of Zoology, Biodiversity Research Centre and Beaty Biodiversity Museum, University of British Columbia, Vancouver, British Columbia, Canada

Correspondence: LN Harris, Arctic Aquatic Research Division, Fisheries and Oceans Canada, 501 University Crescent, Winnipeg, Canada MB R3T 2N6.

E-mail: Les.N.Harris@dfo-mpo.gc.ca

Received 22 January 2014; revised 24 June 2014; accepted 14 July 2014

level of divergence at neutral genetic markers to determine if ecotypes represent distinct gene pools (for example, Gowell *et al.*, 2012; Stafford *et al.*, 2014). If the latter is the case, then a second fundamental hypothesis concerns the geographic origin of the ecotypes (that is, whether they evolved in allopatry and now exist in sympatry following secondary contact or whether they evolved in sympatry). Testing these hypotheses with molecular markers, however, can often be challenging if processes such as contemporary gene flow after secondary contact is influencing population structure (Lu *et al.*, 2001; Turgeon and Bernatchez, 2001), or if historical gene flow before isolation contributes to a blurring of population structure (Harris and Taylor, 2010; Harris *et al.*, 2013).

Salmonine fishes offer some of the best-known examples of morphological and ecological diversity within sympatric populations (Schluter, 1996; Taylor, 1999). One of the most remarkable examples is the Arctic Char (*Salvelinus alpinus*) in Lake Thingvallavatn, Iceland, where four sympatric morphotypes (from herein ‘morphs’) occur (Skúlason *et al.*, 1989a, b; Kapralova *et al.*, 2011). Similarly, sympatric ecotypes have been documented in the Brook Trout (*S. fontinalis*, for example, Dynes *et al.*, 1999; Fraser and Bernatchez, 2005). The Lake Trout (*S. namaycush*) also exhibits morphological and ecological variation across its range, but the nature and origins of such variability have been less well studied than in its congeners. For example, in the Laurentian Great Lakes (LGLs), up to three depth-segregated morphs differing in morphology and ecology have been reported (Krueger and Ihssen, 1995; Moore and Bronte, 2001). The origin of these morphs is not completely understood, but there appears to be a genetic component to these differences (Page *et al.*, 2004; Goetz *et al.*, 2010). Sympatric morphs in the LGLs have been heavily impacted, even extirpated in some cases, owing to over-harvesting, invasive Sea Lamprey (*Petromyzon marinus*) predation and the long history of Lake Trout hatchery supplementation (Krueger and Ihssen, 1995; Page *et al.*, 2004), making the study of their evolutionary origin problematic.

More recently, however, additional examples of within-lake morphological divergence in Lake Trout have been described throughout the geographical range of this species (Zimmerman *et al.*, 2006, 2007; Northrup *et al.*, 2010) including relatively unperturbed systems from the Canadian Arctic. For example, in Great Bear Lake (GBL; Figure 1), Canada’s largest, virtually pristine system, Blackie *et al.* (2003) showed

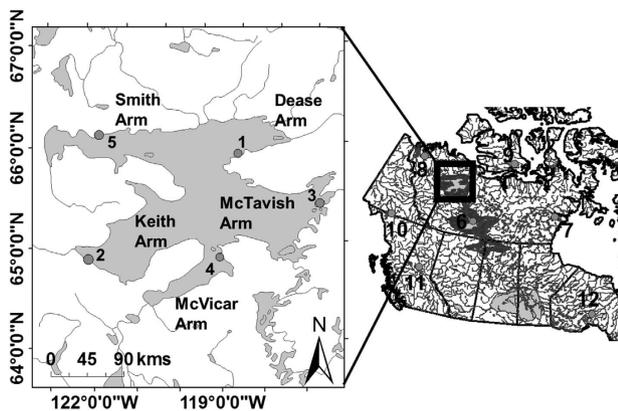


Figure 1 Map of the study area showing sampling locations among arms within GBL including the additional sampling locations throughout Canada. Numbers refer to the locations listed in Table 1. The maximal extent of glacial Lake McConnell that once covered the areas now occupied by GBL, GSL and Lake Athabasca is indicated by the shaded area (modified from Smith, 1994).

that Lake Trout characterized as ‘piscivorous’ and ‘insectivorous’ based on diet that were captured in shallow-water habitats (<20 m) could be differentiated morphologically based largely on differences in upper and lower jaw length, pectoral fin length and caudal peduncle depth. Alfonso (2004) also resolved two groups (called ‘redfin’ and ‘normal’ forms) that were morphologically distinct in body and caudal peduncle depth, as well as a suite of other characters. More recently, Chavarie *et al.* (2013) used a suite of morphological measurements to resolve four shallow-water (<30 m) morphs (referred to as groups 1–4), equalling or exceeding the diversity of the LGLs, but without being related to depth. The levels of ecological and genetic differentiation among these shallow-water forms, and the geography of their origin in GBL, however, remain unknown.

In this study, we evaluated several hypotheses concerning the nature and origin of morphological variation in Lake Trout from GBL. First, we used microsatellite and mitochondrial DNA (mtDNA) to assess levels of genetic divergence among morphs. If the different morphs represent distinct gene pools, rather than solely the result of phenotypic plasticity within a single gene pool, we expected to detect significance divergence between distinct morphs within the same arm, or across arms (Kapralova *et al.*, 2011). Second, we used these data to assess genetic interrelationships among morphs within and among arms of GBL and to test alternative hypotheses on their geographic origins. Under an allopatric model of divergence, we expected that similar morphs found in different arms of GBL would be more closely related to each other than divergent morphs within the same arm and that each such ‘morph-cluster’ would be genetically similar to Lake Trout sampled from outside GBL that originated from distinct glacial refugia (for example, Bernatchez and Dodson, 1991; Lu *et al.*, 2001, Fraser and Bernatchez, 2005; Figure 2a). Alternatively, under a sympatric model of divergence, distinct Lake Trout morphs within GBL should be genetically more similar to each other than to any Lake Trout found outside GBL (the ‘intra-lacustrine’ model; Figure 2b). This scenario would suggest that Lake Trout morphs in GBL diverged from a common ancestor *in situ* soon after colonizing this system post-glacially and subsequently each morph dispersed to the different arms of GBL (Eshenroder, 2008). It is also possible that the multiple occurrences of different morphs within GBL across arms could have originated by multiple bouts of sympatric divergence (the ‘intra-arm’ model). If so, we would anticipate that divergent morphs within the same arm would be each other’s closest relative and that all such arm-specific clusters would form a monophyletic GBL cluster relative to Lake Trout from outside of this system (Hudson *et al.*, 2007; Butlin *et al.*, 2008; Figure 2c). If this model were supported it would provide evidence for parallel divergence of distinct morphs in each arm (see Schluter and Nagel, 1995). Overall, our study strives to better understand the evolution and maintenance of the immense phenotypic and ecological divergence exhibited by fishes occupying post-glacial habitats and to provide insights into whether such variability is a result of genetic discreteness, phenotypic plasticity or combinations of both (Bernatchez and Dodson, 1990; Taylor and McPhail, 2000; Crispo, 2008; Lundsgaard-Hansen *et al.*, 2013).

MATERIALS AND METHODS

Study system, sample collection and previous morph identification

GBL is located in the Northwest Territories of Canada and is the largest lake contained wholly within the country (Figure 1). As the Laurentian and Cordilleran ice sheets receded, numerous glacial lakes were formed throughout much of glaciated North America (Pielou, 1991). One such body of water, glacial Lake McConnell, covered the three basins presently occupied by GBL, Great Slave Lake (GSL) and Lake Athabasca (Craig, 1965; see Figure 1). Glacial

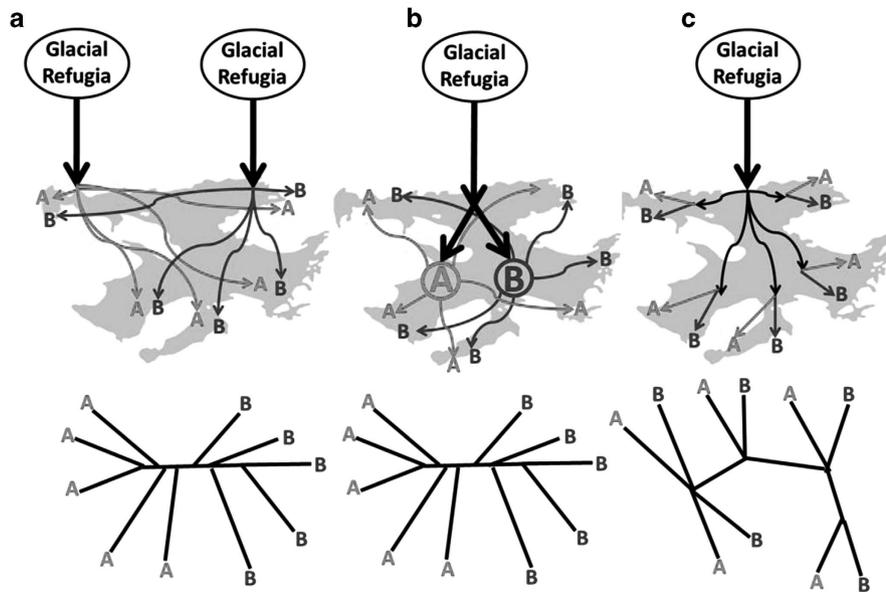


Figure 2 Hypothesized evolutionary scenarios for explaining the origin of morphological variation among shallow-water morphotypes (represented here as 'A' and 'B') of Lake Trout from GBL, Canada. Shown are (a) an allopatric model of divergence, where morphotypes diverged in discrete refugia and then subsequently colonized GBL followed by the colonization of shallow-water habitats within each arm, (b) an 'intra-lacustrine' model of divergence, where the evolution of morphological variation occurred *in situ* soon after colonizing GBL; subsequently, discrete morphs dispersed to occupy the shallow-water habitats within each arm and (c) an 'intra-arm' model of divergence, where the evolution of morphological variation ensued *in situ* within each arm from a common ancestor within GBL (that is, parallel evolution of morphs across arms). Shown below each colonization map are the anticipated phylogenetic relationships.

Lake McConnell formed around 12 000 ybp as the Laurentide ice sheet retreated and existed until around 8500 ybp when isostatic rebound caused it to become separated into GSL, GBL and Lake Athabasca (Lemmen *et al.*, 1994; Smith, 1994). During its existence, glacial Lake McConnell was also impacted by a major flood from glacial Lake Agassiz approximately 9900 ybp that drastically altered the hydrography of this system (Smith, 1994). Presently, GBL is divided into five distinct arms (Dease, Keith, McTavish, McVicar and Smith arms) that are all connected to a central basin. The above description of glacial Lake McConnell raises several important points. First, Lake Trout from what are now separate GBL and GSL would have once existed sympatrically before glacial Lake McConnell became subdivided approximately 8500 ybp. Second, the relatively recent subdivision of glacial Lake McConnell into GBL, GSL and Lake Athabasca implies that Lake Trout from GBL have only been isolated from Lake Trout in other freshwater systems for approximately 8500 years or ~567 generations (based on a 15 year generation time). Finally, as a result of flood water from glacial Lake Agassiz, Lake Trout once isolated in the Mississippian glacial refuge were able to colonize as far north as GBL (see also Rempel and Smith, 1998).

We sampled GBL Lake Trout using paired bottom sets of a 140 mm and multi-mesh (38–140 mm) gill nets set at depths of ≤ 30 m (Chavarie *et al.*, 2013). Owing to the logistical challenges of sampling a large Arctic lake, sampling was localized to one arm per year, rotating each year since 2002. Consequently, the 10-year data set has a full spatial representation of the lake, including temporal coverage in some arms (Table 1).

Four divergent shallow-water (≤ 30 m) sympatric morphs of Lake Trout have been identified, and at least three occur within each arm (Chavarie *et al.*, 2013; Figure 3). Briefly, a digital image of each fish was used to obtain morphological measurements based on 23 landmarks (homologous points), 20 semi-landmarks (used to compare homologous curves), and 12 linear measurement traits. These measurements summarized variation in body shape, head shape, and head and fin lengths of Lake Trout resulting in the resolution of four morphs: groups 1, 2, 3 and 4. Samples were assessed in morphs 1 and 2 from all five arms, whereas morphs 3 and 4 were assessed from only two arms (Dease and McVicar) and one arm (Keith), respectively (Table 1) because of low sample sizes of the latter.

We also obtained tissue samples of Lake Trout from lakes thought to contain lineages that originated from distinct glacial refugia (Wilson and Hebert,

1998): Sandy (Mackenzie River drainage) and Jayko (Victoria Island) lakes (putative Northern Beringian refuge populations); Atlin (Pacific coast drainage) and Nakinlerak (Fraser River drainage) lakes in British Columbia (putative Nahanni or Southern Beringian refuge populations); Peter (Hudson Bay drainage) and Nipigon (Great Lakes drainage) lakes (putative Mississippian refuge populations); and GSL (putative Beringian, Southern Beringian (Nahanni) or Mississippian refuge populations; Table 1, Figure 1). Unfortunately, both genetic and morphological data were only available for Lake Trout samples from GBL (but see Northrup *et al.*, 2010).

Molecular methods

Microsatellite DNA data collection. DNA was extracted using Qiagen Dneasy tissue extraction kits (Qiagen Inc., Valencia, CA, USA). Samples were assayed for variation at 24 microsatellite loci amplified in four multiplexes (Supplementary Appendix 1). An automated sequencer (ABI 3130xl Genetic Analyzer; Applied Biosystems, Foster City, CA, USA) was used for microsatellite analysis using the LIZ 600 size standard. GeneMapper (ver. 4.0, Applied Biosystems) software was used to score all microsatellite data. All scoring data were also visually assessed.

MtDNA data collection. MtDNA data collection methods are described in Harris *et al.* (2013). Briefly, the left domain of the control region (d-loop) was amplified with primers tPro2 (Brunner *et al.*, 2001) and ARCH1 (Alekseyev *et al.*, 2009). The target amplification product was then sequenced with primer tPro2 using the Applied Biosystems Big Dye Terminator v3.1 Terminator Cycle Sequencing kit (Applied Biosystems). Sequencing products were run on an Applied Biosystems 3130xl Genetic Analyzer and aligned to haplotype Snam01 (Genbank accession number: JQ772460) using Seqscape vers. 2.5 (Applied Biosystems).

Statistical analyses

Genetic variation and Hardy–Weinberg and linkage disequilibrium. The program MICRO-CHECKER (ver. 2.2.3; Van Oosterhout *et al.*, 2004) was used to assess the quality of the microsatellite markers by testing for null alleles and large allele dropout. The program FSTAT (ver. 2.9.2.3; Goudet, 2002) was used to compile descriptive statistics (number of alleles (N_A), expected (H_E ,

Table 1 Sampling locations and sample sizes for microsatellite and mtDNA sequencing analyses

Sample (Arm: Morphotype)	Naming code	Map location	Sample size (microsat)	Sample size (mtDNA)	Year(s) collected	Latitude	Longitude
<i>Dease Arm</i>							
Group 1	DES-GRP 1	1	30	11	2005/2010	66°30'	120°22'
Group 2	DES-GRP 2	1	42	16	2005/2010	66°30'	120°22'
Group 3	DES-GRP 3	1	53	5	2005/2010	66°30'	120°22'
<i>Keith Arm</i>							
Group 1	KET-GRP 1	2	59	13	2002/2003	65°10'	123°29'
Group 2	KET-GRP 2	2	44	14	2002/2003	65°10'	123°29'
Group 4	KET-GRP 4	2	20	10	2002/2007/2010	65°10'	123°29'
<i>McTavish Arm</i>							
Group 1	McTAV-GRP 1	3	43	16	2004/2009	66°11'	117°58'
Group 2	McTAV-GRP 2	3	40	16	2004/2009	66°11'	117°58'
<i>McVicar Arm</i>							
Group 1	McVIC-GRP 1	4	45	14	2003/2008	65°24'	120°22'
Group 2	McVIC-GRP 2	4	45	16	2003/2008	65°24'	120°22'
Group 3	McVIC-GRP 3	4	12	10	2003/2008	65°24'	120°22'
<i>Smith Arm</i>							
Group 1	SMI-GRP 1	5	20	16	2006	66°18'	124°18'
Group 2	SMI-GRP 1	5	20	14	2006	66°18'	124°18'
Great Slave Lake	GSL	6	56	17	2010	61°23'	114°26'
Peter Lake	PET	7	50	18	2005	63°06'	92°49'
Sandy Lake	SAN	8	40	21	2004	67°48'	132°14'
Jayco Lake	JAY	9	46	22	2012	69°44'	103°15'
Atlin Lake	ATL	10	54	20	2006	59°28'	133°14'
Nakinlerak Lake	NAK	11	32	13	2005	55°14'	125°14'
Lake Nipigon	NIP	12	60	15	2004	49°52'	88°34'

Abbreviation: mtDNA, mitochondrial DNA.

Map codes refer to those highlighted in Figure 1.

Nei's unbiased gene diversity) and observed (H_O) heterozygosities and the fixation index (F_{IS}) for each locus within each sample. In addition, the program HP-RARE (Kalinowski, 2005) was used to calculate allelic richness at a common sample size (A_R , minimum 22 genes from each sample) and private allelic richness (P_{AR} , 22 genes from each sample). To qualitatively visualize allele frequency differences among all samples and morphs within GBL, bubble plots of allele frequency variation were created for each locus in R (R Core Team, 2013). Differences in A_R , H_E , H_O and F_{ST} when GBL samples were grouped by arm and when samples were grouped by morph were also compared using FSTAT permutation tests. Deviations from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium were assessed using GENEPOP (ver. 4.0; Rousset, 2008). First, we tested for deviations from HWE among all samples and then among only those from GBL when grouped by either arm or morphotype. If large deviations from HWE are observed for one of the grouping scenarios (that is, either by arm or morph), for example, because of a reduction of heterozygosity because of a Wahlund effect, then that grouping hypothesis would be poorly supported. The significance of simultaneous comparisons was initially compared with a nominal alpha of 0.05 and then to an adjusted alpha following the false discovery rate procedure (FDR, Narum, 2006).

Genetic population structure among morphs and arms. Global values of F_{ST} (θ , Weir and Cockerham, 1984) were generated using FSTAT and were calculated among all samples and then among samples in GBL grouped either by arm or by morph. Pairwise estimates of F_{ST} among all samples were compared in ARLEQUIN (ver. 3.1; Excoffier *et al.*, 2005). To further resolve how population structure is best explained in GBL, we also calculated pairwise F_{ST} for samples

when grouped by arm and when grouped by morph. The significance of all pairwise estimates was assessed using 10 000 permutations.

We used several methods to visualize population structure among all samples and then among only those samples from GBL to determine if genetic structure was most evident by morph or by arm. First, a factorial correspondence analysis was performed using GENETIX (ver. 4.02; Belkhir *et al.*, 2004), which graphically represents the genetic distances between individual multi-locus genotypes. Using Cavalli-Sforza and Edwards (1967) chord distance (D_{CE}), bootstrapped neighbour-joining trees were built using PHYLIP (ver. 3.6.9; Felsenstein, 2009). Finally, we used Bayesian clustering implemented in the program STRUCTURE (ver. 2.3.4; Pritchard *et al.*, 2000) to estimate the number of putative populations or clusters (K). For this analysis, we used the admixture model with independent allele frequencies while varying K from 1 to 20. We ran 10 independent runs for each value of K to assess variability of obtained log-likelihood values using a burn-in of 100 000 iterations followed by 100 000 Markov chain Monte Carlo iterations. We performed STRUCTURE analyses for two data sets: (1) one consisting of all samples; and (2) one containing only the GBL samples. Given the limited genetic structure that has previously been resolved among Lake Trout from GBL (Harris *et al.*, 2013), we used the LOCPRIOR option in STRUCTURE for the latter analysis to enhance the likelihood of detecting existent population structure in this system (detailed by Hubisz *et al.*, 2009). To avoid biasing the analysis toward the clustering by morph or arm, LOCPRIOR information was included at the sample level (that is, including 13 distinct entities in our LOCPRIOR information and the 13 localities represent various positions within arms). The program STRUCTURE HARVESTER (ver. 06.6.92; Earl and Vonholdt, 2012) was first used to visualize and compile the results based on both the posterior probability of the data ($\ln P[D]$) and the *post hoc* ΔK statistic of

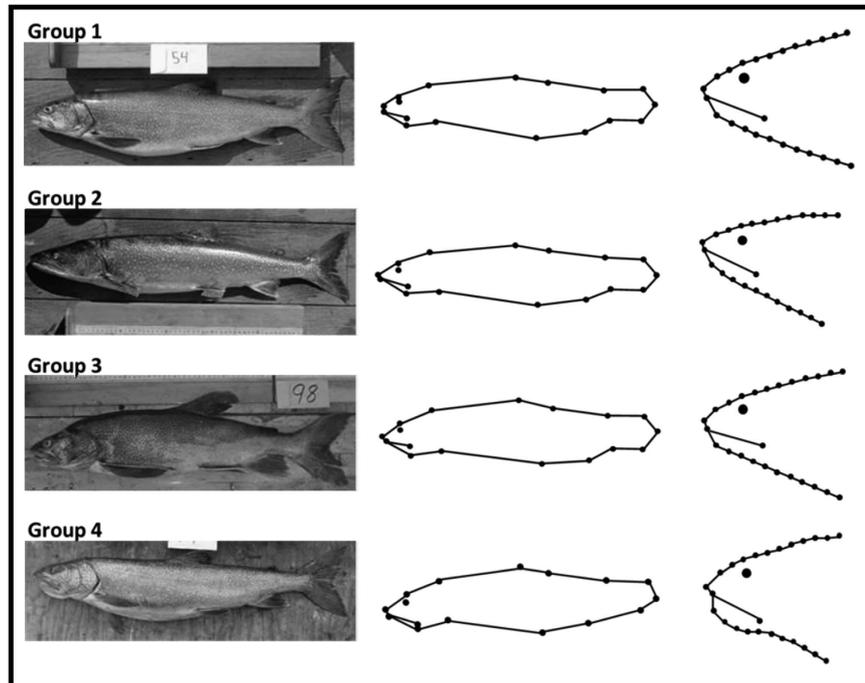


Figure 3 Shallow-water morphotypes of Lake Trout resolved from GBL assessed in this study (from Chavarie *et al.*, 2013). Also shown are the body shape and head shape landmarks used to delineate morphotypes from this system.

Evanno *et al.* (2005). The best alignment of replicate runs was assessed using the program CLUMPP (ver. 1.1; Jakobsson and Rosenberg, 2007) using 1000 permutations and the LargeKGreedy algorithm. The program DISTRICT (ver. 1.1; Rosenberg, 2004) was then used to produce plots of the best alignments for average memberships calculated using CLUMPP. For STRUCTURE analyses, we report the results of both the data in P[D] and the *post hoc* ΔK statistic.

To determine the extent to which genetic variation was partitioned across samples, we conducted a hierarchical analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992) using ARLEQUIN. Using an AMOVA, the percentage of the total genetic variation is partitioned within populations (that is, individuals, Vc), among populations within groups (Vb) and by differences among groups (Va). We first tested a grouping hypothesis based on putative refuge of origin (as described above) to assess large-scale partitioning of genetic variation. We grouped samples by putative refugial origin based on previous assessments for this species (for example, Wilson and Hebert, 1998, see above) and the results of this study. Then, within GBL two grouping hypotheses were tested in which samples were treated by grouping distinct morphs by arm within GBL ('by arm' grouping) and then by grouping similar morphs among arms ('by morph' grouping). If the distinct morphs had arisen by parallel sympatric divergence across each arm of GBL, we expected the greatest amount of variation in allele frequencies to be explained using the 'by arm' grouping scenario.

Finally, to estimate the timing of divergence among the two most common morphotypes in GBL (that is, groups 1 and 2) we used the isolation with migration model (Nielsen and Wakeley, 2001; Hey and Nielsen, 2004) as implemented in the program Ima (Hey and Nielsen, 2007). Ima uses Markov chain Monte Carlo sampling to obtain maximum-likelihood estimates of six parameters, including the timing of divergence (t , as scaled by mutation rate, μ). Initial runs incorporated upper bounds for each parameter ($q1 = q2 = qA = 10$, $m1 = m2 = 10$, $t = 10$) as recommended in the Ima manual. We used heated chains under Metropolis coupling incorporating a geometric heating scheme and parameter estimates were generated under the stepwise mutation model (SMM). After preliminary runs to optimize settings (for example, heating parameters and number of chains), to ensure that the parameter space was fully explored, that the stationary distribution were adequately sampled and that there was adequate mixing (based on low autocorrelation of parameters, high acceptance rates and visualization of plots to ensure no

trends were apparent), five replicate simulations were conducted, each varying the random seed to assess consistency among Ima runs. The final runs included a burn-in of 1×10^6 followed by 1×10^7 Markov chain Monte Carlo steps, with 30 heated chains.

MtDNA genetic analyses. We used MEGA (ver. 5.0; Tamura *et al.*, 2011) to find the best nucleotide substitution model for our mtDNA sequence data as assessed using the Akaike Information Criterion and the Bayesian Information Criterion. ARLEQUIN was used to compute descriptive statistics (that is, haplotype frequency, haplotype diversity (h) and nucleotide diversity (π)), Tajima's D (a test for deviations from neutral expectations; Tajima, 1989) and pairwise F_{ST} between sampling locations (with significance tested using 10000 permutations). In addition, using ARLEQUIN, AMOVA's were performed to determine the extent to which genetic variation (based on mtDNA sequence data) is partitioned according to the grouping hypotheses described above. Evolutionary relationships among haplotypes were visualized by constructing a haplotype network following the statistical parsimony method of Templeton *et al.* (1992) as implemented in the program TCS (ver. 1.20; Clement *et al.*, 2000).

RESULTS

Microsatellite DNA

Genetic variation and HWE and linkage disequilibrium. The locus OMM1128 was monomorphic and the program MICRO-CHECKER consistently identified SnaMSU9 as a locus containing null alleles. Removing these two loci resulted in 22 informative loci that were used in all subsequent analyses across 811 samples. Genetic variation was relatively high with the number of alleles ranging from 4 (Smm21) to 66 (SnaMSU10) alleles per locus and averaging 25.73 alleles per locus across all loci (Supplementary Appendix 2). Bubble plots of allele frequency variation were characterized by differences according to region (that is, putative refugia; Supplementary Appendix 3). Among GBL samples, bubble plots revealed that most samples (morphs) were not characterized by marked differences in allele frequencies or the presence of unique alleles and, qualitatively, no morph-specific

differences were noticeable at any locus (Supplementary Appendix 3). Per locus observed heterozygosity ranged from 0.14 (Sco102) to 0.94 (SnaMSU6), averaging 0.74 across all loci and expected heterozygosity ranged from 0.14 (Sco102) to 0.96 (SnaMSU6 and SnaMSU10), while averaging 0.80 across all loci (Supplementary Appendix 2). Allelic richness ranged from 1.84 (Sco102) to 14.29 (SnaMSU6) and averaged 8.30 across all loci (Supplementary Appendix 2). There were no significant differences ($P > 0.05$) in H_O , H_E , A_R or F_{ST} when comparing GBL samples grouped by arm (Supplementary Appendix 4). Alternatively, when GBL samples were grouped by morph there were significant differences in H_O , H_E and A_R ($P < 0.05$) between morphs, but not in F_{ST} ($P > 0.05$; Supplementary Appendix 4).

When all samples were assessed, HWE was rejected in 61 of a possible 440 population–locus comparisons ($P < 0.05$) but subsequent to adjustments of alpha based on the FDR procedure only 21 deviations were detected ($P < 0.0075$). All deviations were the result of heterozygote deficiencies. When samples from GBL were grouped by arm, 19 of a possible 110 population–locus comparisons (17.2%) were significant after adjustments for multiple comparisons ($P < 0.0095$), whereas when samples were grouped by morph only 12 of a possible 88 comparisons (13.6%) remained significant after FDR adjustments ($P < 0.0099$). Virtually all deviations when samples were grouped by arm or by morph were the result of heterozygote deficiencies. Significant linkage disequilibrium was detected in 188 of 4620 tests ($P < 0.05$), but after using the FDR procedure, it was detected in only 45 comparisons ($P < 0.0055$).

Population structure. Moderate overall differentiation among all samples was resolved (global F_{ST} estimate of 0.071, 95% confidence interval (CI) = 0.056–0.087) but not when comparing only those samples from GBL in which overall differentiation was low (global F_{ST} = 0.008, 95% CI = 0.005–0.011). Pairwise estimates of F_{ST} ranged from virtually no apparent differentiation (among some of the samples from GBL) to 0.435 between lakes from distinct drainages covering putative distinct refugia (NAK and SAN; Supplementary Appendix 5). The majority of these comparisons were significant ($P < 0.05$), even after adjusting for multiple comparisons based using the FDR ($P < 0.0086$). Among GBL samples, however, results differed when samples were grouped by arm or by morph. When samples were grouped by arm, global F_{ST} was 0.002 (95% CI = 0.001–0.003). In contrast, differentiation was almost four times higher when samples were grouped into the four morphs (global F_{ST} = 0.007, 95% CI = 0.004–0.011). When comparing samples grouped by arm, pairwise F_{ST} ranged from 0.0009 (between Keith and Dease arms) to 0.0061 (between McTavish and Smith arms located on opposite sides of GBL; Table 2a). Among morphs, pairwise estimates were higher, ranging from 0.0063 (between GRP 1 and GRP 2 morphs) to 0.0174 (between GRP 2 and GRP 4 morphs; Table 2b). Finally, when GBL samples were combined and compared with those grouped into putative refugia F_{ST} ranged from 0.038 (between GBL and the Mississippian refuge group) to 0.176 (between the Beringian and Southern Beringian (Nahanni) refugial groups; Table 2c). All of these comparisons were significant before ($P < 0.05$) and after adjusting for multiple comparisons ($P < 0.0219$).

The factorial correspondence analysis grouped individuals into clusters suggestive of distinct refugial origins (Figure 4a). Samples from GBL, GSL, PET and NIP (putative Mississippian refuge origin) clearly grouped together, while samples from SAN and JAY (putative Beringian refuge origin) also formed a distinct cluster. Alternatively, the samples from NAK and ATL (putative Southern Beringian refuge origin) were divergent from one another and showed no clear genetic

Table 2 Pairwise F_{ST} (θ) values based on microsatellite (below diagonal) and mtDNA sequence (above diagonal) data between samples from GBL when they are grouped by arm within this system (a), when grouped by morphotype (b) and when all GBL samples were combined and compared with those grouped into putative refugia (that is, Northern Beringian (N-BER), Southern Beringian (S-BER) or Mississippian (MIS) (c))

A	DES	KET	McTAV	McVIC	SMI
DES		−0.0221	0.0884	−0.0191	−0.0247
KET	0.0009		0.0850	−0.0082	−0.0266
TAV	0.0011	0.0030		0.0533	0.0389
VIC	0.0016	0.0018	0.0030		−0.0235
SMI	0.0034	0.0013	0.0061	0.0037	

B	GRP 1	GRP 2	GRP 3	GRP 4
GRP 1		−0.0019	−0.0197	0.0543
GRP 2	0.0063		0.0073	0.1001
GRP 3	0.0038	0.0067		−0.0341
GRP 4	0.0122	0.0174	0.0086	

C	GBL	N-BER	MIS	S-BER
GBL		0.0635	0.0278	0.2139
N-BER	0.1128		0.2038	0.4439
MIS	0.0377	0.1344		0.1010
S-BER	0.1029	0.1762	0.1386	

Arm and morphotype codes are shown in Table 1.

Bold values are significant at $P < 0.05$. Underlined values are still significant following the false discovery rate procedure (FDR) ($P < 0.017$, 0.0219 and 0.0219 for A, B and C respectively).

affinity to any of the other samples. Within GBL, there was some clustering of samples based on morph (Figure 4b) and when only the GRP 1 and GRP 2 morphs were assessed, the genetic distinction between them was even clearer (Supplementary Appendix 6). The neighbour-joining tree also grouped samples by the refuge from which they potentially dispersed and again, GSL, PET and NIP grouped closer in the tree to GBL than any of the other samples (Figure 4c). Some separation of GBL samples by morph was also indicated based on the neighbour-joining tree. Although bootstrap support levels were modest, GRP 1 morphs tended to cluster together and separately from GRP 2 morphs (Figure 4d). The GRP 3 morph was more associated with the GRP 1 morph, whereas the GRP 4 morph was more closely associated with the GRP 2 morph (Figure 4d). The genetic distinction between GRP 1 and GRP 2 morphs was highlighted even further when only those samples were included in the analysis (Supplementary Appendix 6).

Over the entire data set, Bayesian clustering implemented in STRUCTURE suggested the existence of 9 ($\Delta K = 4.67$) or 10 ($\ln P[D] = -70744.18$) genetic clusters (Supplementary Appendix 7). When admixture plots were visualized assuming $K = 9$, distinct clusters that were associated with putative refugial origins were apparent for some samples (Figure 5a). For example, GSL and GBL (putative Mississippian refuge) samples clustered relatively closely together as did the SAN and JAY (putative Beringian refuge) samples. Within GBL, the STRUCTURE analysis suggested the existence of two genetic clusters based on the $\ln P[Data]$ values, whereas results based on ΔK suggest that four clusters as the most likely population structure ($\ln P[Data]$ and ΔK of -43695.97 and 5.75 respectively; Supplementary Appendix 7). Under any of the models of population structure explored (that is, $K = 2$ or 4 as the best models), admixture

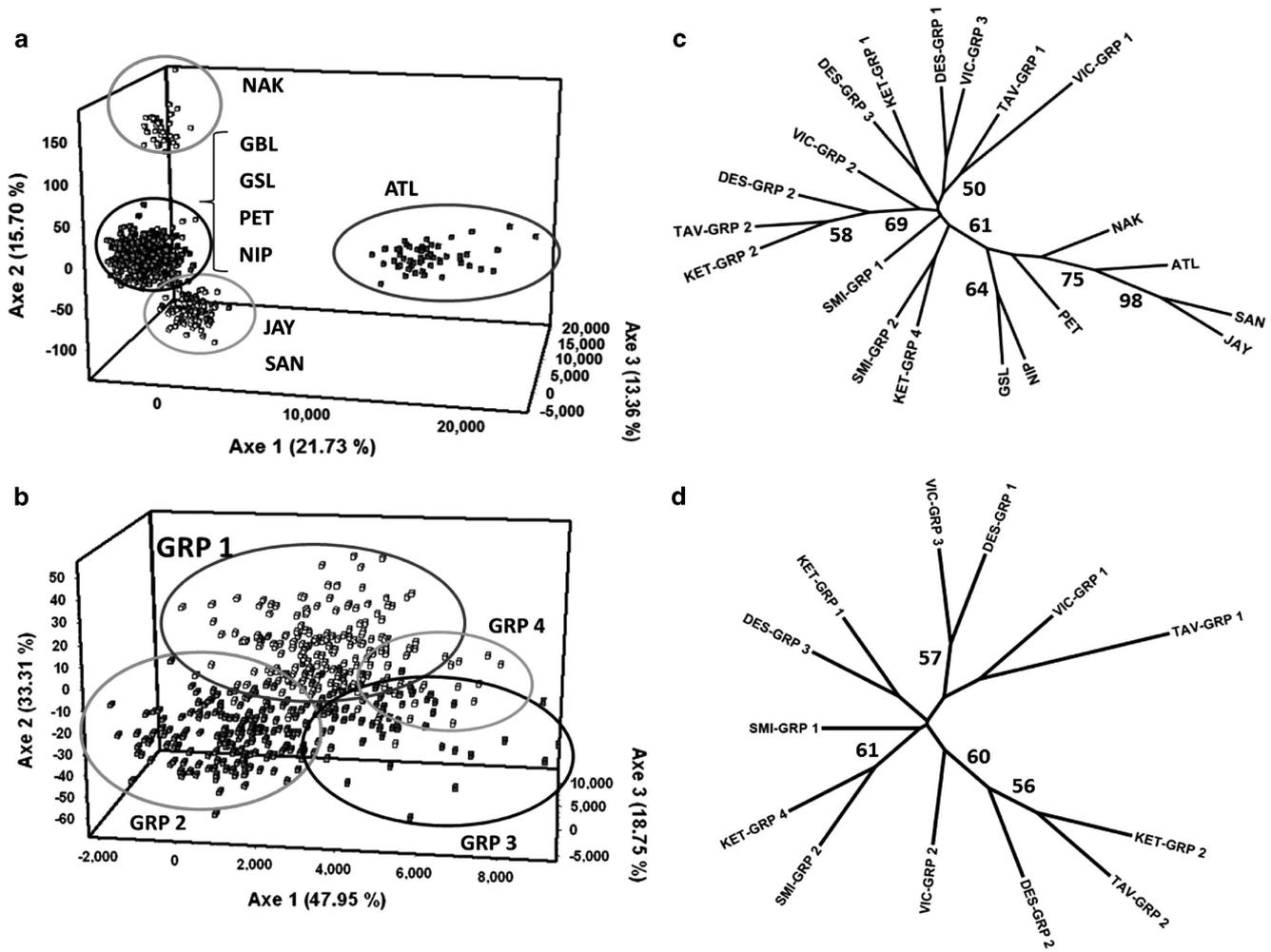


Figure 4 Factorial correspondence analysis (FCA) of microsatellite DNA variation for (a) all samples included in the study and (b) only those from GBL. Also shown is a neighbour-joining (NJ) tree based on Cavalli-Sforza and Edwards (1967) chord distance (DCE) for microsatellite data shown for (c) all samples in the study and (d) only those from GBL. Sample codes are shown in Table 1.

plots indicated strong differentiation in the genetic compositions of GRP 1 and GRP 2 morphs, and that GRP 3 and GRP 4 morphs have greater similarity to the GRP 1 morph (Figure 5b, Supplementary Appendix 6). When only the GRP 1 and GRP 2 morphs were analyzed, both statistics provided support for the existence of two genetic clusters within GBL (ln P[Data] and ΔK of -35263.93 and 6.08 respectively; Supplementary Appendix 7).

AMOVA of microsatellite DNA allele frequencies indicated that 9.5% of the variation was explained by variation among groups when populations were grouped into putative refugial origins, 6.3% of the variation was attributed to variation among populations within putative refugia and 84.2% was explained by individuals within populations (Table 3). Alternatively, the AMOVA revealed that very little of the genetic variation could be explained by the grouping scenarios for samples within GBL; virtually all of the variation was attributed to variation among individuals within each sample ($>99\%$) for both the grouping scenarios ($P < 0.001$) and only 0.54% ($P > 0.05$) and -0.18% ($P > 0.05$) of the variation was attributed to the groupings based on morph and arm within GBL, respectively (Table 3). When samples were grouped into either GRP 1 or GRP 2 morph, results were virtually identical; most of the variation (98.9%) was explained by variation among individuals within populations (Table 3).

Among the five independent runs performed, the posterior probability distribution for divergence time (t) between groups 1 and 2 peaked at 0.195 (0.095–0.255, 90% highest posterior density) and averaged 0.141 (0.077–0.211, 90% highest posterior density; Supplementary Appendix 8) across all five runs. Assuming a microsatellite mutation rate of 1×10^{-4} (Jarne and Lagoda, 1996) resulted in a divergence time of 1950 years (950–2550) using the peak estimate of t and 1410 years (770–2110) using the divergence estimate averaged across all five independent runs. Regardless, both estimates are much less than the predicted age of GBL of 8500 ybp.

MtDNA variation

A 468-base pair region of the mtDNA control region was sequenced for 302 individuals and nine variable sites were found. A total of 12 haplotypes were resolved, 7 of which were new (Snam07-13) and deposited into GenBank under accession numbers KF951407-13. Haplotype Snam01 was the most common haplotype found in 39.2% of all samples, represented in all sampling locations with the exception of ATL and NAK (Supplementary Appendix 9). Snam06, Snam07 and Snam08 were the second, third and fourth most common haplotypes, found in 29.2%, 14.6% and 7.3% of all the samples, respectively. These haplotypes were found in most sampling locations (and morphs within GBL). All other

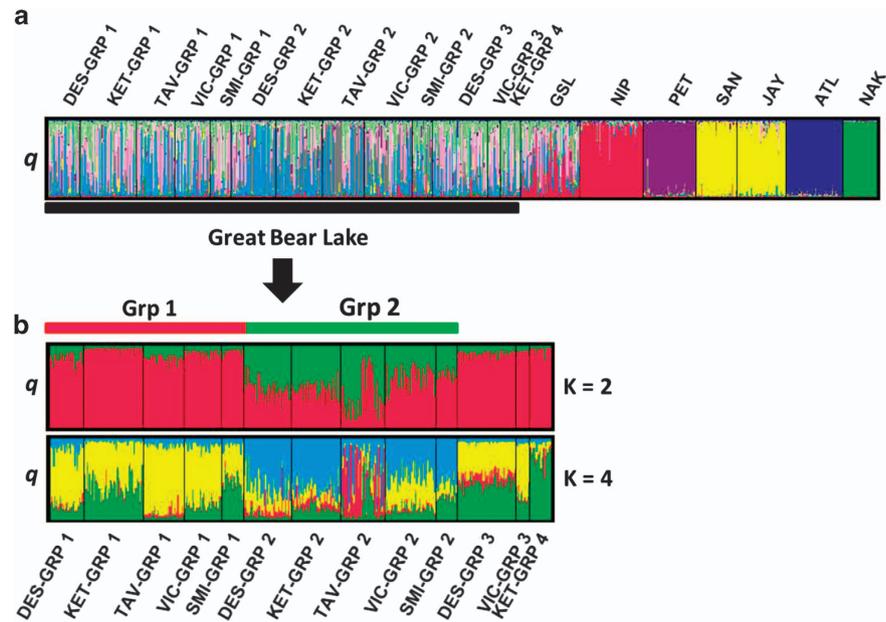


Figure 5 The results of Bayesian clustering analysis implemented in STRUCTURE showing the proportion of the genome (q , admixture coefficient on the y axis) assigned to one of the most likely inferred clusters. Shown are the results when (a) all samples were assessed (shown for $K=9$) and then when (b) only those from GBL were included (shown for $K=2$ and 4 , see Results section). Each column represents a different individual. Sample codes refer to those outlined in Table 1.

Table 3 Results of the hierarchical AMOVA showing the grouping hypotheses tested in this study

Genetic data	Grouping hypothesis	Among groups (V_a)	Among populations within groups (V_b)	Within populations (V_c)
Microsatellite DNA	Among putative refugial groups	9.48**	6.34**	84.17**
	GBL samples grouped by arm	-0.18	0.99**	99.19**
	GBL samples grouped by morphotype	0.54**	0.45**	99.01**
	GBL GRP 1 and GRP 2 samples grouped by morphotype	0.53**	0.48**	98.99**
mtDNA sequence	Among putative refugial groups	8.8*	11.25**	79.95**
	GBL samples grouped by arm	0.25	3.54	96.21*
	GBL samples grouped by morphotype	-0.77	4.29*	96.48*
	GBL GRP 1 and GRP 2 samples grouped by morphotype	-0.48	4.74*	95.75*

Abbreviations: AMOVA, analysis of molecular variance; mtDNA, mitochondrial DNA.

Variation among groups (V_a), variation among populations within groups (V_b) and variation within populations (V_c).

* $P < 0.05$.

** $P < 0.01$.

haplotypes were relatively uncommon and were found in only a handful of samples. The NAK samples were represented exclusively by the Snam06 haplotype and Snam10 was unique to the ATL samples being found in $>50\%$ of these. The Tamura 3-parameter model (Tamura, 1992) was found to be the most appropriate model of nucleotide substitution and was used in all subsequent analyses that required prior substitution information. Estimates of Tajima's D suggest that DNA polymorphism within each sample was consistent with neutral expectations (all $P > 0.05$; Supplementary Appendix 9). Haplotype diversity ranged from 0.00 to 0.90 in the NAK and DES-GRP 3 samples, respectively. Nucleotide diversity ranged from 0.00 to 0.01 in the NAK and DES-GRP 3 samples, respectively.

Pairwise F_{ST} ranged from -0.162 between samples from two distinct arms in GBL (VIC-GRP 3 and DES-GRP 3) to 0.863 between samples from distinct drainages outside GBL (ATL and NAK; Supplementary Appendix 5). Of these comparisons,

57 of 190 were significant ($P < 0.05$) and 40 of these remained significant after adjusting alpha for multiple comparisons ($P < 0.0086$; Supplementary Appendix 5). There were very few comparisons among GBL samples that were significant (4 of 58, $P < 0.0086$) and comparisons between GBL samples and GSL were usually nonsignificant. Finally, both the NAK and ATL sampling locations were highly differentiated from all other samples. When samples were grouped by arm only two comparisons were significant ($P < 0.05$), both of which involved McTavish Arm (Table 2a) and when assessing differentiation among GBL samples grouped by morph, no significant pairwise comparisons of F_{ST} were found ($P > 0.05$; Table 2b). Finally, when GBL samples were grouped and compared with those grouped into putative refugia, F_{ST} ranged from 0.0278 (between GBL and the Mississippian refugia) to 0.444 (between the Beringian and Southern Beringian (Nahanni) refugial groups; Table 2c) and all but the GBL-Mississippian refugial group comparison were significant

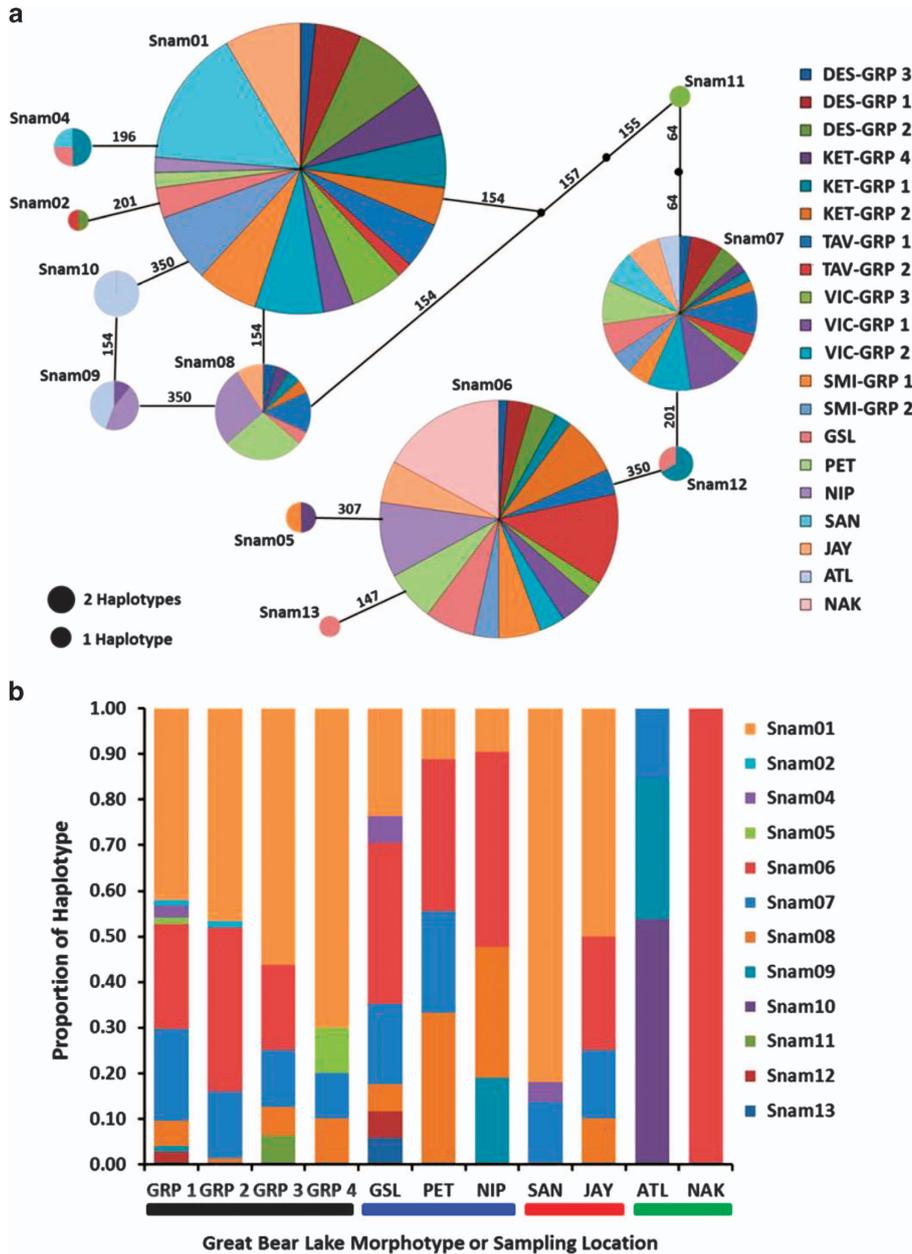


Figure 6 (a) Unrooted haplotype network based on the sequencing of 468-base pairs within the mtDNA control region (d-loop) showing the geographic distribution of haplotypes produced using TCS (Clement *et al.*, 2000). Each circle represents a unique haplotype and lines separating haplotypes represent mutational steps. Small black nodes between haplotypes suggest missing haplotypes. The numbers on the lines between haplotypes indicate the base pair position for the difference. The size of the circle indicates the number of individuals with that haplotype and colours correspond to the sample codes shown in Table 1. (b) The proportion of each haplotype shown per morphotype in GBL (GRP 1, GRP 2, GRP 3 and GRP 4) and within other sampling locations (see Table 1). Coloured bars under sample codes correspond to samples from GBL (black), and putative Mississippian (blue), Northern Beringian (red) and Southern Beringian (Nahanni) (green) glacial refugia origins.

before ($P < 0.05$) and after adjusting for multiple comparisons ($P < 0.0219$).

Similar to the AMOVA on microsatellite data, very little of the variation could be explained by our grouping hypotheses for samples from GBL ($V_a < 1.0\%$ for all hypotheses; Table 3). More of the variation, however, was attributed to populations within groups ($V_b = 3.5\text{--}3.7\%$), but again most of the variation was attributed to variation among individuals within each sample ($> 95\%$, $P < 0.05$ for all scenarios). When grouping populations by putative refugial origin, 8.8% of the variation was

attributed to this grouping hypothesis (V_a , $P < 0.05$), whereas 11.3% and 79.9% was attributed to populations within groups (V_b , $P < 0.01$) and individuals within populations (V_c , $P < 0.01$) respectively.

The haplotype network based on mtDNA sequences highlighted the close interrelationships among haplotypes, which differed from each other by a maximum of seven mutational steps (Figure 6a). Although the frequencies of haplotypes varied across sampling locations, the three most common haplotypes (Snam01, Snam06 and Snam07) were distributed throughout most of the study area.

(Figure 6b). With the exception of ATL (the only location in which Snam10 was present) and NAK (fixed for Sam06), no regions or sampling locations were characterized by unique assemblages of highly divergent haplotypes (Supplementary Appendix 10).

DISCUSSION

Divergence and origin of GBL *S. namaycush* morphs

A remarkable feature of post-glacial, north-temperate freshwater fishes is their extensive phenotypic and ecotypic diversity (Schluter, 1996; Taylor, 1999; Robinson and Parsons, 2002). Lake Trout are well known for polymorphism in the southern latitudes (that is, the LGLs, Moore and Bronte, 2001) but, in general, little is known regarding the evolution and maintenance of morphological and ecological diversity in this species, particularly in the Arctic. Our study has provided evidence that there is an association between morphological and genetically distinct groups of Lake Trout in Canada's GBL and that the latter appear to represent distinct gene pools. Within GBL, inter-arm population structure was low, but when pairwise comparisons were structured by morph, differentiation was much stronger and most analyses appear to genetically group samples by morph better than by arm. Model-based clustering suggested that the greatest degree of genetic divergence in GBL is between GRP 1 and GRP 2 with the other two morphs appearing to be most similar to the GRP 1 morph. Allopatric isolation in, and dispersal from, distinct refugia has been important for generating biodiversity in north-temperate faunas (for example, Bernatchez and Dodson, 1990; April *et al.*, 2012) but our data were consistent with an intra-lacustrine model of morph divergence, as (1) all morphs were genetically most similar to one another rather than to Lake Trout sampled from other areas likely representative of populations that originated distinct glacial refugia and (2) maximum-likelihood estimates of divergence are younger than the formation of GBL suggesting Lake Trout morphological variation evolved *in situ* in this system. Notwithstanding our evidence of intra-lacustrine divergence of GBL Lake Trout, our data suggested that GBL Lake Trout diversity probably ultimately originated by dispersal from a Mississippian refuge.

Few studies have tested for associations between morphology and genetic divergences among morphs in Lake Trout. Goetz *et al.* (2010) assessed growth, morphometry, lipid content and differences in liver transcriptomics among laboratory reared lean and siscowet Lake Trout morphs from Lake Superior and concluded that differences in these traits were genetic in nature. Earlier, Burnham-Curtis and Smith (1994) drew similar conclusions when comparing osteological characters. Alternatively, Northrup *et al.* (2010) found no association between morphologically and genetically distinct groups of Lake Trout from another northern system (Atlin Lake, BC) and Stafford *et al.* (2014) did not detect any genetic differentiation between Lake Trout morphs from Montana, although this latter case involved a non-native population introduced <120 years ago. Thus, in addition to showing that different morphs may represent distinct gene pools, our work represents one of the first studies to use genetic data to evaluate different geographic models for their evolution (but see Krueger and Ihssen, 1995). Specifically, our work supports the idea that post-glacial populations of fishes may diverge into discrete ecomorphological forms in sympatry subsequent to colonization of novel, heterogenous post-glacial habitats. Although empirical examples of divergence occurring sympatrically are still relatively rare (Bolnick and Fitzpatrick, 2007), more studies supporting this scenario are becoming available (for example, Barluenga *et al.*, 2006), including an increasing number involving north temperate fishes (Taylor and

Bentzen, 1993; Gíslason *et al.*, 1999; Lu *et al.*, 2001; Alekseyev *et al.*, 2002; Præbel *et al.*, 2013).

Ecological sympatric speciation, in which barriers to gene flow are driven by divergent selection acting in contrasting directions between ecological niches (Rundle and Nosil, 2005; Nosil, 2012), has been implicated in the evolution of reproductive isolation, genetic differentiation and the evolution of ecological/morphological variation in variety of fish species (Schluter, 1996; Hendry *et al.*, 2007; Langerhans *et al.*, 2007; Præbel *et al.*, 2013). Under an ecological speciation model, upon colonization of GBL or once Lake Trout became isolated in this system, founding populations can be hypothesized to have exploited discrete niches within the same geographical area (Rundle and Nosil, 2005). Divergent selection may have then acted to drive the fixation of advantageous alleles in each morph (Schluter and Conte, 2009) or selection may have resulted in the evolution of so called 'magic traits' to pleiotropically promote reproductive isolation (Servedio *et al.*, 2011). Given that unique alleles were not common in any of the Lake Trout morphs, these populations likely represent incomplete (or initial) stages of reproductive isolation (Nosil *et al.*, 2009). In GBL Lake Trout, natural selection is likely acting on traits important resource/food exploitation and the subsequent adaptations to these new ecological niches may have promoted some degree of assortative mating and reproductive isolation (Hendry *et al.*, 2007). Reinforcement of mate choice from selection against hybrids or immigrant phenotypes that are less fit than the parental morphs (Vamosi and Schluter, 1999) could have then contributed to the maintenance of this divergence. Indeed, in GBL, Blackie *et al.* (2003) identified two morphs that differed in diet (insectivorous vs piscivorous) and suggested that the morphological differences among shallow-water forms were likely the result of divergent foraging strategies. Furthermore, the morphological variation used to delineate the morphs in this study was associated with food acquisition (Chavarie *et al.*, 2013). Thus, the divergence of GBL Lake Trout into discrete morphs is likely, at least partially, related to selection acting on traits important for food exploitation (or habitat use associated with prey distribution). Variation in depth-related habitat use and diet has also been associated with ecotypic/morphological variation in this species in other northern systems (Zimmerman *et al.*, 2007) including GSL (Zimmerman *et al.*, 2006, 2009). GSL has likely shared a similar glacial history to that of GBL, and Lake Trout from these two systems have likely been isolated for approximately 8500 years (Smith, 1994). Unfortunately, no genetic data are available for the previously identified Lake Trout morphs from GSL and we did not have morphological data for our GSL samples that could be compared directly with those from GBL. Given that all morphotypes in GBL were most closely related to one another than to samples from GSL (or any other system) and that GBL appears to exhibit much more diversity (at least four versus two forms, the latter of which were primarily captured in depths >50 m), it is unlikely that the morphological diversity observed in GBL is the result of colonization of divergent morphs from GSL. Furthermore, our estimates for divergence between the morphs for which we had representative samples from all arms, suggests that the Lake Trout morphs assessed in the present evolved *in situ* after post-colonization of GBL.

Interestingly, in the LGLs, sympatric evolutionary hypotheses for divergence have also been suggested for deep-water Lake Trout morphs. For example, Burnham-Curtis (1993) suggested that the siscowet morph diverged from the lean morph in Lake Superior after the Pleistocene (8000 ybp) and that the humper morph was the result of hybridization between the two. Eshenroder (2008) reiterated this idea and proposed that the humper morph diverged post-glacially in

sympatry from an ancestral shallow-water form. Interestingly, a common theme in these studies is that divergence evolved *in situ* as a result of divergent natural selection acting upon traits important for resource exploitation. Regardless of the mechanism responsible for the divergence, our results are consistent with hypotheses proposed for explaining morphological variation among sympatric Lake Trout populations in other large lake systems where divergence appears to have occurred post-glacially in sympatry (Eshenroder, 2008).

Finally, and although widely considered the most important mechanism for the evolution biodiversity (Coyne and Orr, 2004, see also April *et al.*, 2012), our genetic results argue against the fully allopatric model of divergence despite this mechanism being described for the evolution of divergence and morphological variation in numerous North Temperate species (Pigeon *et al.*, 1997; Lu *et al.*, 2001; Fraser and Bernatchez, 2005; April *et al.*, 2012). Under the allopatric hypothesis, we anticipated that similar morphs of Lake Trout across different arms of GBL would be more closely related to each other than divergent morphs within the same arm and that similar morphs would then be associated with Lake Trout from outside of this system that originated in distinct refugia. This, however, was not the case because all morphs from GBL had a genetic affinity to samples that likely originated from the Mississippian refuge (for example, GSL, Peter and Nipigon lakes) suggesting that all morphs of GBL Lake Trout had similar refugial origins. This was also supported by the lack of morph-specific mtDNA haplotypes and by low values of differentiation among morphs from GBL suggesting that they have not spent long periods of time isolated from each other. The microsatellite-based divergence time estimates (<2000 years) also suggest that the two genetic groups of Lake Trout in GBL are also of relatively recent, post-glacial origin. It is, however, impossible to discount the possibility that divergent lineages did, in fact, colonize GBL, but that any historical signature of divergence resulting from isolation in distinct glacial refugia is obscured by recent and/or contemporary gene flow among morphs.

Lake Trout phylogeography

During the height of the Wisconsin glacial, from 85 000 to 10 000 ybp (Lindsey and McPhail, 1986; Pielou, 1991), North American Lake Trout survived in several distinct refugia spanning the continent (Wilson and Hebert, 1998). Lake Trout dispersed from these refugia to occupy its contemporary distribution and although their sample size was small ($N=5$), Wilson and Hebert (1998) hypothesized that Lake Trout colonized GBL from two distinct Beringian refugia (both a southern and northern refuge) to form an area of secondary contact. By contrast, our results suggest that Lake Trout in GBL did not colonize this system from multiple refugia within Beringia, but rather dispersed post-glacially to GBL from one distinct refuge, the Mississippian. This was evidenced by the clustering of all GBL samples with GSL in the STRUCTURE plot and with the GSL, PET and NIP samples in the factorial correspondence analysis.

In addition, the lack of lake-specific haplotypes and the low differentiation among samples from GBL suggests that these populations/morphs were not at one time isolated in distinct glacial refugia (*cf.* Bernatchez and Dodson, 1990). This is consistent with the glacial history of the region suggesting that GBL Lake Trout once occupied glacial Lake McConnell and thus have only been isolated from GSL Lake Trout (and other populations of Lake Trout from glacial Lake McConnell) for ~8500 years. Furthermore, microsatellite and mtDNA differentiation between GBL samples and those from populations that likely originated from the Mississippian refuge was lower than comparisons between GBL and putative Beringian (JAY

and SAN) or southern Beringian (ATL and NAK) samples. On balance, given the clear affinity of GBL samples to those from GSL, PET and NIP, in terms of both microsatellites and mtDNA it seems most plausible that this system was founded by one lineage of fishes that most recently inhabited glacial Lake McConnell, which was colonized by Lake Trout of Mississippian origin (Wilson and Hebert, 1998). This dispersal likely proceeded through the vast system of post-glacial lakes (that is, glacial Lakes Agassiz, Peace, McConnell and others) that permitted dispersal of Mississippian refuge fish to northern and Arctic Canada beginning about 10 000 years ago (Rempel and Smith, 1997; Wilson and Hebert 1998). This inference, combined with our analysis of genetic relationships within GBL and the shallow divergence time estimates for the most common morphs, suggests that the discrete morphs did not arise by allopatric origin in distinct refugia, but rather by intra-lacustrine divergence following colonization by a single glacial lineage of Lake Trout.

CONCLUSIONS

We have shown that morphs of Lake Trout from GBL are genetically differentiated from one another, yet overall they exhibit strong genetic similarities to one another relative to those from outside of GBL. This, combined with divergence time estimates between morphs and the clustering of all GBL Lake Trout samples with putative samples thought to represent a Mississippian glacial lineage, suggests that processes not necessarily involving isolation in allopatry have been important for generating morphological and genetic diversity in this system. Thus, our results provide additional insights into the origin and maintenance of biodiversity, including ecotypic and genetic variation among freshwater fishes, in recently deglaciated, north temperate aquatic habitats, which will be vital for monitoring and assessing potential changes to intraspecific variation if conditions or habitats are altered (for example, Taylor *et al.*, 2006). Furthermore, our data suggest that GBL is an excellent system for further studying the mechanisms involved in divergence that has ensued *in situ* including ongoing contemporary gene flow and the maintenance of reproductive isolation among closely related populations in the initial stages of speciation.

DATA ARCHIVING

Data available from the Dryad Digital Repository: doi:10.5061/dryad.1368p.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the planning and logistic help provided by Déline Renewable Resources Council, Déline Lands and Finance Corporation, the Community of Déline, Polar Continental Shelf Program, DFO in Hay River, and the Federal Department of Environment and Natural Resources in Déline. We especially thank the leading sampling teams over the years consisting of J Chavarie, G Lafferty, Z Martin and Chris Yukon along with the following individuals who helped conduct field sampling in various years or provided Lake Trout samples for this study: J Baptiste, D Baton, M Connor, R Tallman, L Dueck, C Gallagher, R Eshenroder, M Lennie, M Lindsay, G Menacho, I Neyelle, L Neyelle, S Northrup, M Smirle, A Vital, M Yukon and Charity, Cameron and Cyre Yukon. Financial support was provided by Fisheries and Oceans Canada (DFO), Sahtu Renewable Resource Board and Northern Development Canada Northwest Territories Cumulative Impacts Monitoring Program grants. The manuscript was greatly improved by the constructive input from three anonymous reviewers.

- Alekseyev SS, Samusenok VP, Matveev AN, Pichugin MY (2002). Diversification, sympatric speciation, and trophic polymorphism of Arctic charr, *Salvelinus alpinus* complex, in Transbaikalia. *Environ Biol Fish* **64**: 97–114.
- Alekseyev SS, Bajno R, Gordeeva NV, Reist JD, Power M, Kirillov AF *et al.* (2009). Phylogeography and sympatric differentiation of the Arctic charr *Salvelinus alpinus* (L.) complex in Siberia as revealed by mtDNA sequence analysis. *J Fish Biol* **75**: 368–392.
- Alfonso NR (2004). Evidence for two morphotypes of lake charr, *Salvelinus namaycush*, from Great Bear Lake, Northwest Territories, Canada. *Environ Biol Fishes* **71**: 21–32.
- April J, Hanner RH, Dionne-Cote A-M, Bernatchez L (2012). Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Mol Ecol* **22**: 409–422.
- Barluenga M, Stoltzing KN, Salzburger W, Muschick M, Meyer A (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* **439**: 719–723.
- Belkhir K, Borsari P, Chikhi L, Raufaste N, Bonhomme F (2004). Genetix 4.05, logiciel sous windows TM pour la génetique des populations, version 4.05. *Laboratoire Génome, Populations, Interactions, CNRS UMR 5000. Université de Montpellier II: Montpellier, France.*
- Bernatchez L, Dodson JJ (1990). Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial DNA restriction analysis. *Evolution* **44**: 1263–1271.
- Bernatchez L, Dodson JJ (1991). Phylogeographic structure in mitochondrial-DNA of the lake whitefish (*Coregonus-clupeaformis*) and its relation to Pleistocene glaciations. *Evolution* **45**: 1016–1035.
- Blackie CT, Weese DJ, Noakes DLG (2003). Evidence for resource polymorphism in the lake charr (*Salvelinus namaycush*) population of Great Bear Lake, northwest territories, Canada. *Ecoscience* **10**: 509–514.
- Bolnick DI, Fitzpatrick BM (2007). Sympatric speciation: models and empirical evidence. *Annu Rev Ecol Evol S* **38**: 459–487.
- Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L (2001). Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution* **55**: 573–586.
- Burnham-Curtis MK (1993). Intralacustrine speciation of *Salvelinus namaycush* in Lake Superior: an investigation of genetic and morphological variation and evolution of lake trout in the Laurentian Great Lakes. PhD Thesis, University of Michigan, Ann Arbor, MI, USA.
- Burnham-Curtis MK, Smith GR (1994). Osteological evidence of genetic divergence of lake trout (*Salvelinus namaycush*) in Lake Superior. *Copeia* **1994**: 843–850.
- Butlin RK, Galindo J, Grahame JW (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philos Trans R Soc B-Biol Sci* **363**: 2997–3007.
- Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic analysis models and estimation procedures. *Am J Hum Genet* **19**: 233–257.
- Chavarie L, Howland KL, Tonn WM (2013). Sympatric polymorphism in Lake Trout: the coexistence of multiple shallow-water morphotypes in Great Bear Lake. *Trans Am Fish Soc* **142**: 814–823.
- Clement M, Posada D, Crandall K (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1660.
- Coyne JA, Orr HA (2004). *Speciation*. Sinauer Associates: Sunderland, MA, 545 pp.
- Craig BG (1965). Glacial lake McConnell and the surficial geology of parts of Slave River and Redstone River map areas, District of Mackenzie. *Geol Surv Can Bull* **122**: 33.
- Crispo E (2008). Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. *J Evol Biol* **21**: 1460–1469.
- Dynes J, Magnan P, Bernatchez L, Rodriguez MA (1999). Genetic and morphological variation between two forms of lacustrine brook charr. *J Evol Biol* **54**: 955–972.
- Earl DA, Vonholdt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4**: 359–361.
- Eshenroder RL (2008). Differentiation of deep-water lake charr *Salvelinus namaycush* in North American lakes. *Environ Biol Fish* **83**: 77–90.
- Evanno G, Regnault S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform* **1**: 47–50.
- Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes—application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.
- Felsenstein J (2009). PHYLIP (Phylogeny Inference Package) version 3.6.9. *Distributed by the author. Department of Genome Sciences. University of Washington: Seattle.*
- Frankham R, Ballou JD, Briscoe DA (2002). *Introduction to Conservation Genetics*. Cambridge University Press: Cambridge.
- Fraser DJ, Bernatchez L (2005). Allopatric origins of sympatric brook charr populations: colonization history and admixture. *Mol Ecol* **14**: 1497–1509.
- Gislason D, Ferguson M, Skulason S, Snorrason SS (1999). Rapid and coupled phenotypic and genetic divergence in Icelandic Arctic char (*Salvelinus alpinus*). *Can J Fish Aquat Sci* **56**: 2229–2234.
- Goetz F, Rosauer D, Sitar S, Goetz G, Simchick C, Roberts S *et al.* (2010). A genetic basis for the phenotypic differentiation between siscowet and lean Lake Trout (*Salvelinus namaycush*). *Mol Ecol* **19**: 176–196.
- Goudet J (2002). FSTAT: a program to estimate and test gene diversities and fixation indices. Version 2.9.3.2. Available at: <http://www.unil.ch/izea/software/fstat.html>, accessed 2 November 2011
- Gowell CP, Quinn TP, Taylor EB (2012). Coexistence and origin of trophic ecotypes of pygmy whitefish, *Prosopium coulterii*, in a south-western Alaskan lake. *J Evol Biol* **25**: 2432–2448.
- Harris LN, Howland KL, Kowalchuk MW, Bajno R, Lindsay MM, Taylor EB (2013). Microsatellite and mtDNA analysis of Lake Trout, *Salvelinus namaycush*, from Great Bear Lake, Northwest Territories: impacts of historical and contemporary evolutionary forces on Arctic ecosystems. *Ecol Evol* **3**: 145–161.
- Harris LN, Taylor EB (2010). Genetic population structure of broad whitefish, *Coregonus nasus*, from the Mackenzie River, northwest territories: implications for subsistence fishery management. *Can J Fish Aquat Sci* **67**: 905–918.
- Hendry AP, Nosil P, Rieseberg LH (2007). The speed of ecological speciation. *Funct Ecol* **21**: 455–464.
- Hewitt GM (1996). Some genetic cosequences of ice ages, and their role in deivergence and speciation. *Biol J Linnean Soc* **58**: 247–276.
- Hey J, Nielsen R (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**: 747–760.
- Hey J, Nielsen R (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci USA* **104**: 2785–2790.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* **9**: 1322–1332.
- Hudson AG, Vonlanthen P, Müller R, Seehausen O (2007). Review: the geography of speciation and adaptive radiation in coregonines. *Adv Limnol* **60**: 111–146.
- Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.
- Jarne P, Lagoda P (1996). Microsatellites, from molecules to populations and back. *Trends Ecol Evol* **11**: 424–429.
- Kalinowski ST (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.
- Kapralova KH, Morrissey MB, Kristjansson BK, Olafsdottir GA, Snorrason SS, Ferguson MM (2011). Evolution of adaptive diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland. *Heredity* **106**: 472–487.
- Krueger CC, Ihssen PE (1995). Review of genetics of Lake Trout in the great lakes: history, molecular genetics, physiology, strain comparisons, and restoration management. *JGT Lakes Res* **21**: 348–363.
- Langerhans RB, Gifford ME, Joseph EO (2007). Ecological speciation in *Gambusia* fishes. *Evolution* **61**: 2056–2074.
- Lemmen DS, Duk-Rodkin A, Bednarski JM (1994). Late glacial drainage systems along the northwest margin of the Laurentide ice sheet. *Quaternary Sci Rev* **13**: 805–828.
- Lindsey CC, McPhail JD (1986). Zoogeography of fishes of the Yukon and Mackenzie basins. In: Hocutt CH, Wiley EO (eds) *The Zoogeography of North American Freshwater Fishes*. Wiley: New York, pp 639–674.
- Lu G, Basley DJ, Bernatchez L (2001). Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*): relevance for speciation. *Mol Ecol* **10**: 965–985.
- Lundsgaard-Hansen B, Matthews B, Vonlanthen P, Taverna A, Seehausen O (2013). Adaptive plasticity and genetic divergence in feeding efficiency during parallel adaptive radiation of whitefish (*Coregonus* spp.). *J Evol Biol* **26**: 483–498.
- Moore SA, Bronte CR (2001). Delineation of sympatric morphotypes of Lake Trout in Lake Superior. *Trans Am Fish Soc* **130**: 1233–1240.
- Narum SR (2006). Beyond Bonferroni: less conservative analyses for conservation genetics. *Conserv Genet* **7**: 783–787.
- Nielsen R, Wakeley J (2001). Distinguishing migration from isolation. A Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.
- Northrup S, Connor M, Taylor EB (2010). Population structure of Lake Trout (*Salvelinus namaycush*) in a large glacial-fed lake inferred from microsatellite DNA and morphological analysis. *Can J Fish Aquat Sci* **67**: 1171–1186.
- Nosil P (2012). *Ecological Speciation*. Oxford University Press: Oxford.
- Nosil P, Harmon L, Seehausen O (2009). Ecological explanations for (incomplete) speciation. *Trends Ecol Evol* **24**: 145–156.
- Østbye K, Amundsen PA, Bernatchez L, Knudsen A, Knudsen R, Kristoffersen R *et al.* (2006). Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. *Mol Ecol* **15**: 3983–4001.
- Page KS, Scribner KT, Burnham-Curtis M (2004). Genetic diversity of wild and hatchery Lake Trout populations: relevance for management and restoration in the great lakes. *Trans Am Fish Soc* **133**: 674–691.
- Pielou EC (1991). *After the Ice Age. The Return of Life to Glaciated North America*. University of Chicago press: Chicago, IL.
- Pigeon D, Chouinard A, Bernatchez L (1997). Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution* **51**: 196–205.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Præbel K, Knudsen R, Siwertsson A, Karhunen M, Kahilainen KK, Ovaskainen O *et al.* (2013). Ecological speciation in postglacial European whitefish: rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecol Evol* **3**: 4970–4986.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, URL: <http://www.R-project.org/>
- Rempel LL, Smith DG (1998). Postglacial fish dispersal of fish from the Mississippi refuge to the Mackenzie River basin. *Can J Fish Aquat Sci* **55**: 893–899.

- Robinson BW, Parsons KJ (2002). Changing times, spaces, and faces: tests and implications of adaptive morphological plasticity in the fishes of northern postglacial lakes. *Can J Fish Aquat Sci* **59**: 1819–1833.
- Rosenberg NA (2004). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* **4**: 137–138.
- Rousset F (2008). GENEPOP' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* **8**: 103–106.
- Rundle H, Nosil P (2005). Ecological speciation. *Ecol Lett* **8**: 336–352.
- Schluter D (1996). Ecological speciation in postglacial fishes. *Phil Trans R Soc Lond B* **351**: 807–814.
- Schluter D, Conte GL (2009). Genetics and ecological speciation. *Proc Natl Acad Sci USA* **106**: 9955–9962.
- Schluter D, Nagel LM (1995). Parallel speciation by natural selection. *Am Nat* **146**: 292–301.
- Servedio MR, Van Doorn GS, Kopp M, Frame AM, Nosil P (2011). Magic traits in speciation: 'magic' but not rare? *Trends Ecol Evol* **26**: 389–397.
- Skúlason S, Noakes DLG, Snorrason SS (1989a). Ontogeny of trophic morphology in 4 sympatric morphs of Arctic charr *Salvelinus alpinus* in Thingvallavatn, Iceland. *Biol J Linn Soc* **38**: 281–301.
- Skúlason S, Smith TB (1995). Resource polymorphisms in vertebrates. *Trends Ecol Evol* **10**: 366–370.
- Skúlason S, Snorrason SS, Noakes DLG, Ferguson MM, Malmquist HJ (1989b). Segregation in spawning and early life-history among polymorphic Arctic charr, *Salvelinus alpinus*, in Thingvallavatn, Iceland. *J Fish Biol* **35**: 225–232.
- Smith DG (1994). Glacial Lake McConnell: paleogeography, age, duration, and associated river deltas, Mackenzie River Basin, western Canada. *Quaternary Sci Rev* **13**: 829–843.
- Smith TB, Skúlason S (1996). Evolutionary significance of resource polymorphisms in fishes, amphibians, and birds. *Annu Rev Ecol Syst* **27**: 111–133.
- Stafford CP, McPhee MV, Eby LA, Allendorf FW (2014). Introduced Lake Trout exhibit life history and morphological divergence with depth. *Can J Fish Aquat Sci* **71**: 10–20.
- Tajima F (1989). Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Tamura K (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Mol Biol Evol* **9**: 678–687.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.
- Taylor EB (1999). Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation. *Rev Fish Biol Fisher* **9**: 299–324.
- Taylor EB, Bentzen P (1993). Evidence for multiple origins and sympatric divergence of trophic ecotypes of smelt (*Osmerus*) in Northeastern North America. *Evolution* **47**: 813–832.
- Taylor EB, Boughman JW, Groenenboom M, Schluter D, Sniatynski M, Gow JL (2006). Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Mol Ecol* **15**: 343–355.
- Taylor EB, McPhail JD (2000). Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Phil Trans R Soc Lond B* **267**: 2375–2384.
- Templeton AR, Crandall KA, Sing CF (1992). A cladistics analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619–633.
- Turgeon J, Bernatchez L (2001). Mitochondrial DNA phylogeography of lake cisco (*Coregonus artedii*): evidence supporting extensive secondary contacts between two glacial races. *Mol Ecol* **10**: 987–1001.
- Vamosi SM, Schluter D (1999). Sexual selection against hybrids between sympatric stickleback species: evidence from a field experiment. *Evolution* **53**: 874–879.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* **4**: 535–538.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population-structure. *Evolution* **38**: 1358–1370.
- Whitman DW, Agrawal AA (2009). What is phenotypic plasticity and why is it important? In: Whitman DW, Ananthakrishna TN (eds) *Phenotypic Plasticity of Insects: Mechanisms and Consequences*. Science Publishers, Inc.: Enfield, NH, pp 1–63.
- Wilson CC, Hebert PDN (1998). Phylogeography and postglacial dispersal of Lake Trout, *Salvelinus namaycush*, in North America. *Can J Fish Aquat Sci* **55**: 1010–1024.
- Wimberger PH (1994). Trophic plasticity, polymorphisms, and speciation in vertebrates. In: Stouder DJ, Fresh K (eds) *Advances in Foraging Theory and Ecology* Belle Baruch Press: South Carolina.
- Zimmerman MS, Krueger CC, Eshenroder RL (2006). Phenotypic diversity of Lake Trout in Great Slave Lake: differences in morphology, buoyancy, and habitat depth. *Trans Am Fish Soc* **135**: 1056–1067.
- Zimmerman MS, Krueger CC, Eshenroder RL (2007). Morphological and ecological differences between shallow- and deep-water Lake Trout in Lake Mistassini, Quebec. *J Gt Lakes Res* **33**: 156–169.
- Zimmerman MS, Schmidt SN, Krueger CC, Vander Zanden MJ, Eshenroder RL (2009). Onto genetic niche shifts and resource partitioning of Lake Trout morphotypes. *Can J Fish Aquat Sci* **66**: 1007–1018.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)