# MODELING THE MULTIPLE FACETS OF SPECIATION-WITH-GENE-FLOW TOWARDS IMPROVING DIVERGENCE HISTORY INFERENCE OF A RECENT FISH ADAPTIVE RADIATION.

Clément Rougeux*[1], Louis Bernatchez[1], Pierre-Alexandre Gagnaire[2,3]

[1]Département de biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, G1V 0A6, Québec, Canada
[2]Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
[3]Institut des Sciences de l'Évolution de Montpellier - CNRS, UMR 5554, SMEL, 2 rue des Chantiers, 34200 Sète, France
* Corresponding author: clement.rougeux.1@ulaval.ca

August 8, 2016

**Abstract**

The parallel occurrence of replicated species-pairs in similar environmental contrasts may arise through a variety of evolutionary mechanisms. In particular, whether parallelism reflects a common history of divergence or repeated parallel divergence driven by divergent selection needs to be ascertained. Reconstructing historical gene flow is therefore of fundamental interest to understand how demography and selection jointly shape the genomic landscape of species divergence. Here, we extend the current modeling framework to explore the multiple facets of speciation-with-gene-flow using demo-genetic divergence models that capture both temporal and chromosomal variation in migration rate and effective population size. We implement this approach to investigate the divergence history of a young adaptive radiation involving five sympatric Lake Whitefish limnetic (dwarf) and benthic (normal) species pairs (*Coregonus clupeaformis*) characterized by variable degrees of ecological divergence and reproductive isolation. Genome-wide SNPs obtained by RAD-sequencing were used to document the extent of genetic differentiation and of allele sharing among species pairs. Using a composite likelihood approach, we then compared the predictions of 26 divergence models to the unfolded joint allele frequency spectrum of each species-pair. We found strong evidence indicating that a recent (circa 3000-4000 generations) asymmetrical secondary contact between expanding post-glacial populations has accompanied independent Whitefish diversification. Our results suggest that heterogeneous genomic differentiation patterns have emerged through the combined effects of linked selection generating variable rates of lineage sorting across the genome during the allopatric phase, and heterogeneous introgression eroding divergence at different rates across the genome upon secondary contact. This study provides a new retrospective insight into the historical demographic and selective processes that shaped a continuum of divergence associated with ecological speciation. This study also illustrates the efficiency of the newly developed models towards improving the understanding of historical events in the process of adaptive evolution and speciation in any other species complex.

## Introduction

Historical changes in the geographical distribution of species have been an important driver of diversification in many taxa [1]. In particular, the pronounced climatic variations that occurred during the late Pleistocene caused major shifts in the distribution ranges of many species. These shifts are responsible for the divergence of ancestral lineages that survived in different glacial refugia, and then possibly came into secondary contact during interglacial periods [2-4]. The signature of post-glacial recolonization is still apparent in well-known terrestrial and aquatic suture zones, where multiple contacts between expanding post-glacial lineages tend to overlap and form hybrid zones hotspots for many species [5-10].

In some cases, secondary contacts have resulted in the sympatric enclosure of previously allopatric, partially reproductively isolated lineages, for instance within post-glacial lakes. This sympatric coexistence should have facilitated gene flow compared to parapatric populations, eventually leading to complete genetic homogenization of the original glacial lineages. This is not the case, however, for several north temperate freshwater fishes in which sympatric glacial lineages have phenotypically further diverged in the form of reproductively isolated species pairs [11-16]. These cases of ecological speciation have been hypothesized to reflect adaptive responses to minimize competitive interactions and outbreeding depression through ecological niche segregation and hybridization avoidance among previously allopatric lineages [17].

The evolutionary processes responsible for the phenotypic diversification of these incipient sympatric species remains contentious, especially with regards to the relative contributions of genetic differences that evolved in allopatry compared to more recent genetic changes occurring in sympatry [18]. To gain a more thorough understanding of the evolutionary processes leading to divergence, it is crucial to simultaneously take into account the historical demographic events that accompanied divergence and the subsequent genetic exchanges that occurred in sympatry. Genome-wide polymorphism data now provide the opportunity to infer complex demographic histories [19-21] and investigate the evolutionary processes leading to the formation of nascent sympatric species.

Many aspects of a populations' evolutionary history are influenced by demography [22], such as the rate of lineage sorting and gene exchange [23]. Several approaches have been developed to infer the history of population divergence from genetic data obtained from contemporary populations. These approaches usually rely on demographic models capturing the effects of population size, splitting time and migration between two populations exchanging genes [24-26], and may include more complex histories such as divergence and admixture among multiple populations [27,28]. These methods have been mainly used to infer the demographic history of speciation in the presence of gene flow [27], with several studies including further refinements to account for temporal variation in the intensity of gene flow during divergence (*e.g.*, [29]).

An important facet of the speciation process is that a non-negligible proportion of the genome may be affected by selection [30]. Studies investigating the genomic landscape of species divergence revealed that genome-wide differentiation patterns are shaped by multiple factors leaving heterogeneous footprints across the genome [31,32]. Localized selective effects may generate chromosomal variation in the rate of lineage sorting [33] and

effective migration [34,35], which are usually not taken into account by inference methods assuming homogeneous demographic parameters among loci. Approaches allowing for variation in effective migration rate ($m_e$) have been developed to capture the barrier effect of speciation genes [23,36,37]. Less effort has been devoted, however, towards accounting for the effects of background selection and selective sweeps [38-40] on linked neutral diversity, despite their influence on patterns of genome-wide diversity in relation to local recombination rate [41]. To our knowledge, only one study has explicitly addressed these effects by allowing for local reductions in effective population size [23]. Therefore, the consequences of not accounting for linked selection when making inferences about demographic divergence remain unclear [42].

North American Lake Whitefish (*Coregonus clupeaformis*) represents a valuable model to study the role of past allopatric isolation on recent, sympatric ecological divergence. The St. John River drainage (southeastern Québec, northeastern Maine), where benthic (normal) and limnetic (dwarf) Whitefish sympatric species pairs occur, corresponds to a suture zone where two glacial lineages (Atlantic/Mississippian and Acadian) have been hypothesized on the basis of mitochondrial DNA phylogeography to have come into secondary contact during the last glacial retreat [11]. Thus, in some lakes, phenotypic divergence between sympatric dwarf (limnetic) and normal (benthic) populations is still partly associated with the mitochondrial DNA lineages characterizing the different glacial origins of the sympatric populations [43]. Dwarf whitefish are most often associated with the Acadian mitochondrial lineage and are only found in sympatry with the normal species. Moreover, fish from the Acadian lineage have a normal phenotype outside the contact zone, which supports the hypothesis that the dwarf species has been derived postglacially from an Acadian genetic background within the contact zone and always in sympatry with normal Whitefish [11,17,44]. The evolution of further divergence in sympatry suggests that character displacement [45] may have been facilitated by the contact between genetically differentiated lineages. However, since a single mtDNA lineage was observed in two of the lakes harboring sympatric dwarf and normal whitefish, the alternative scenario of single colonization and sympatric speciation could not be ruled out in these cases. Moreover, the different dwarf-normal species pairs found in the contact zone are arrayed along a continuum of phenotypic differentiation, which closely mirrors the potential for niche segregation and exclusive interactions within lakes [46-50]. This continuum is also evident at the genomic level, with increased baseline genetic differentiation and larger genomic islands of differentiation being found from the least to the most phenotypically and ecologically differentiated species pair [51,52]. Finally, quantitative trait loci (QTL) underlying adaptive phenotypic divergence map preferentially to genomic islands of differentiation [51-53], suggesting that selection acting on these traits contribute to the barrier to gene flow. Despite such detailed knowledge on this system, previous studies did not allow clarifying how the genomic landscape of dwarf-normal divergence has been influenced by the relative effects of directional selection on these QTLs and post-glacial differential introgression. Consequently, it is fundamental to elucidate the demographic history of Whitefish populations to disentangle the evolutionary mechanisms involved in their diversification.

The main goal of this study was to use a genome-wide single nucleotide polymorphism (SNP) dataset to infer the demographic history associated with the origin of five independently evolved sympatric dwarf and normal lake whitefish species pairs. Using RAD-seq SNP data to document the Joint Allele Frequency Spectrum (JAFS) in each species pair, we specifically test for the role of temporal and chromosomal variations in the rate of gene flow for each species pair separately, controlling for both effective population size and migration. The history of gene flow was then investigated among lakes to decipher the origin of divergence parallelism at the genome level.

## Results

*Comparisons among divergence models*

For each of the five sympatric whitefish species pairs (Fig 1), 26 alternative divergence models (Fig 2) were fitted to polymorphism data and compared to each other. The unfolded joint allele frequency spectrum (JAFS) of each species pair was constructed using orientated SNPs for which we could distinguish the ancestral and the derived variant. The five JAFS obtained highlighted the continuum of divergence existing among lakes (Fig 3A). Namely, the density of shared polymorphisms located along the diagonal decreased from the least divergent (Témiscouata, East and Webster) to the most divergent (Indian and Cliff) species pairs, while the variance of SNP density around the diagonal increased accordingly. In addition, non-shared polymorphisms (*i.e.*, private SNPs) located on the outer frame of the spectrum were mostly found in Indian and Cliff lakes, which were also the only lakes to display differentially fixed SNPs between dwarf and normal whitefish (*i.e.*, $F_{st}$=1).

Comparisons among demographic divergence models accounting for temporal and chromosomal variation in gene flow (separating $N$ and $m_e$ effects) depicted informative trends. First, the comparison of model scores among lakes showed the importance of considering temporal changes in effective population size, since the models including population growth (-G) generally provided better fits to the data (ANOVA, $F_{GT}$ = 6.557, $p$ = 0.015*; $F_{GI}$ = 7.018, $p$ = 0.012* and $F_{GC}$ = 11.31, $p$ = 0.002** for Témiscouata, Indian and Cliff, respectively but not significantly for East Lake; $F_{GE}$ = 0.931, $p$ = 0.342), except for Webster Lake (Fig 4A). Similarly, accounting for heterogeneous migration rates across the genome (-2m models) improved the average model scores for each lake (ANOVA, $F_{mT}$ = 3.525, $p$ = 0.07., marginally significant ($p$<0.1); $F_{mE}$ = 0.021, $p$ = 0.885; $F_{mW}$ = 0,017, $p$ = 0.897; $F_{mI}$ = 4.466, $p$ = 0.043* and $F_{mC}$ = 6.245, $p$ = 0.018*), although not significantly so in East and Webster lakes (Fig 4B). Moreover, models integrating heterogeneous effective population size at the genomic level (-2N models) provided significant improvements for the three most divergent species pairs (ANOVA, $F_{NW}$ = 4.737, $p$ = 0.037*; $F_{NI}$ = 8.937, $p$ = 0.005** and $F_{NC}$ = 8.008, $p$ = 0.008**) (Fig. 4C).

Using a criterion of $\Delta AIC_i \leq 10$ for model selection, we retained two best models for Témiscouata (SC2mG and SIG), Indian (SC2N2mG and SC2NG) and Cliff (SC2N2mG and SC2mG) lakes, with the second fittest model in each lake showing a $\Delta AIC$ value of 1.54 (SIG), 5.78 (SC2NG) and 1.77 (SC2mG), respectively. Four best models were retained for East Lake SC2N2mG, IM2mG ($\Delta AIC$=2.77), AM2N2m ($\Delta AIC$=7.71), AMG ($\Delta AIC$=8.86). Finally, eight models were retained for Webster Lake (AM2m, IMG, IM, IM2NG, SIG, SI2NG, IM2mG and SC2NG) (Table 1, Table S1). This number of retained models illustrates the difficulty to reliably identify a probable scenario for Webster Lake, which displayed a particularly reduced level of polymorphism compared to the four other lakes. The analysis of Akaike weights ($w_{AIC}$) provided a more precise picture of the relative support for the different models in each lake (Fig 4D, Table 1, Table S1). For Témiscouata, the best model was a secondary contact with heterogeneous migration contemporary to population size change (SC2mG), which had a probability of 0.68, while the second retained model (SIG) received a probability of 0.32. The best model for Webster Lake was a scenario of ancient migration with heterogeneous migration (AM2m),

which showed a much higher probability (0.74) compared to the second best model (0.13 for IMG). Finally, East Lake and the two most differentiated species pairs (Indian and Cliff) received the highest support for the same best model, which was a secondary contact with heterogeneous migration and effective population size contemporary to population size change (SC2N2mG; $w_{AIC} = 0.77$ in East, $w_{AIC} = 0.94$ in Indian and $w_{AIC} = 0.53$ in Cliff).

*Inferences of model parameters*

The inferred proportion of correctly orientated markers in the unfolded JAFS (parameter *O*) ranged from 95.4% to 99%, suggesting that the vast majority of ancestral allelic states were correctly inferred using the European whitefish as an outgroup, and thus supporting the validity of our marker datasets for making inferences (Table 1, Table S1).

Considering only the best fit model for each lake, some general patterns emerged from the comparisons of inferred model parameters among lakes. First, asymmetric effective population sizes between dwarf and normal populations after splitting from the ancestral population was inferred in all lakes except East Lake (ANOVA, $F_{NeT} = 6.33$, $p = 0.016$*; $F_{NeE} = 0.186$, $p = 0.67$; $F_{NeW} = 297.7$, $p = 2.10^{-12}$***; $F_{NeI} = 16.55$, $p = 1.13\text{x}10^{-4}$*** and $F_{NeC} = 32.28$, $p = 1.78\text{x}10^{-7}$***, Table 1, Table S1). In all lakes except Webster, $N_e$ was larger for dwarf compared to normal whitefish. Taking into account population growth in the four lakes for which this varied ($N_e$ was constant in Webster) revealed variable patterns of population size changes. A recent demographic expansion was found in all populations except for the normal whitefish of Témiscouata. A more pronounced demographic expansion was inferred for dwarf compared to normal whitefish in Témiscouata, East and Indian lakes, while the opposite was found for Cliff Lake. Contemporary effective population size, given by the product of theta*$N_i$*$b_i$, was larger in dwarf than in normal populations in the four lakes where a secondary contact was inferred (*i.e.*, Témiscouata, East, Indian and Cliff; ANOVA, $F_{NebT} = 2.95$, $p = 0.095$., marginally significant $p<0.1$; $F_{NebE} = 4.4$, $p = 0.047$*; $F_{NebI} = 4.38$, $p = 0.04$* and $F_{NebC} = 5.42$, $p = 0.02$*).

Asymmetric migration rates were also found in all five lakes for the two categories of loci assumed in heterogeneous migration models (-2m) (Table 1). The best models for the three most divergent species pairs (Webster, Indian and Cliff lakes) involved similar asymmetric migration directions for both categories of loci, with higher rates from dwarf to normal whitefish ($m_{e21} > m_{e12}$ and $m_{e'21} > m_{e'12}$) in Webster and Indian lakes and the opposite in Cliff Lake. In these three lakes, the proportion of loci exhibiting reduced effective migration rates generally followed the divergence continuum (Webster: 0.05; Indian: 0.41; Cliff: 0.52). In contrast, we found that in Témiscouata and East lakes, the best models revealed opposite migration rates between the two categories of loci considered.

We then obtained a measure of gene flow estimate by calculating the product of the contemporary effective population size and effective migration rate weighted by the relative fraction of the genome falling in each category (*i.e.*, average-gene flow = $N*b*(P*m_e+(1-P)*\ m_{e'})$ in each direction). This provides an estimation of the contemporary number of migrants exchanged per generation from one population to the other. We found more pronounced gene flow from dwarf to normal populations in all lakes. Also, differences in gene flow intensities were significant in all lakes but not in Cliff Lake where gene flow was most restricted overall (ANOVA, $F_{NmT} = 13.75$, $p = 9\text{x}10^{-4}$*** ;

$F_{NmE} = 24.34$, $p = 4x10^{-4}$*** ; $F_{NmW} = 46.55$, $p = 1.99x10^{-9}$***; $F_{NmI} = 5{,}414$, $p = 0.023$* and $F_{NmC} = 0.328$, $p = 0.57$ ) (Fig 5).

The best models for East Lake and the two most divergent pairs (*i.e.*, Indian and Cliff) included heterogeneous effective population size at the genomic level (Fig 1C, Table 1). The fraction of the genome with a reduced $N_e$ (*i.e.*, Q) was estimated to about 16% in East, and 40% in Indian and Cliff, and the degree of reduction in $N_e$ (*i.e.*, the Hill-Robertson factor, *hrf*) was about 11%, 20% and 6% for East, Indian and Cliff lakes, respectively.

Time parameters, namely the duration of allopatric isolation ($T_S$) and gene flow ($T_{SC}$ and $T_{AM}$), were converted into absolute time estimates (years). Summing over periods of strict isolation and gene flow (summing the higher parameter values of $T_S$ and $T_{SC}/T_{AM}$, from East Lake) revealed a recent divergence history, which was younger than 55,000 yrs in all five lakes (*e.g.*, *Ts* was approximatively 30,000; 36,000; 36,000; 29,000; 32,000 yrs for Témiscouata, East, Webster, Indian and Cliff lakes respectively) (Table 1). The inferred time of secondary contact in Témiscouata, East, Indian and Cliff lakes coincided roughly with the last glacier retreat following the Wisconsinian glaciation (7,200; 19,600; 8,500 and 9,200 ybp, respectively, Table S1) [54].

*Comparisons of genetic variation among lakes*

Of the 42,582 SNPs that were genotyped in total (all lakes combined), only 2.5% corresponded to shared polymorphic loci across all lakes (*i.e.*, 'shared across lakes' category; Fig 2). Reciprocally, about 25% of the SNPs were private to Témiscouata or East lakes, whereas Webster, Indian and Cliff lakes each contained ~10% of private SNPs. Consequently, loci segregating in at least two lakes represented 73-88% of the dataset. Over the five lakes, 34% and 38% of all the loci were private to dwarf and normal populations, respectively, while the remaining 28% SNPs were shared between species. Within lakes, the highest proportions of SNPs private to normal whitefish were found in Témiscouata (51%) and Webster (69%), where ~37% and 26% of SNPs were private to the dwarf whitefish. The three other lakes displayed the opposite pattern with smaller proportions of SNPs that were private to normal (East ~31%, Indian ~33%, Cliff ~32%) and larger proportions of SNPs private to dwarf (~52%, ~50% and ~61%, respectively). Shared variation within lakes represented only 6% to 16% of the SNPs.

Partitioning genetic variation within and among lakes using a dAPC revealed distinct signals on the first four axes (Fig 6A). On the first axis (LD1, explaining 39.5% of the variance), populations clustered by lakes according to their geographical distribution, roughly separating the three most southern lakes (Webster, Indian and Cliff, negative coordinates) from the other two lakes (Témiscouata and East, positive coordinates). The second axis (LD2, explaining 22.5% of the variance) mostly separated dwarf and normal species from Cliff. The third axis (LD3, explaining 16% of the variance) separated species-pairs and clustered dwarf and normal whitefish in two distinct groups. Positive coordinates were occupied by normal whitefish of Mississippian/Atlantic origin, except for East Lake where normal whitefish is fixed for the Acadian mitochondrial haplotype [43]. The two most extremes populations values on that axis corresponded to normal species, the least introgressed relict of the Mississippian/Atlantic lineage [53], and to dwarf species associated with the Acadian lineage. Finally, the fourth axis (LD4, explaining 11% of the

variance) separated the dwarf and normal species from each lake, underlying some level of parallelism in the divergence between species among lakes (Fig 6B).

The genetic relationships among populations analyzed with *TreeMix* revealed two levels of signal (Fig 7). The first level was directly linked with genetic distance between species, since the population tree rooted with the normal population from Cliff, the most divergent population that best reflects the ancestral state of the Mississippian/Atlantic lineage [55], separated normal populations from Cliff (CN) and Indian (IN) and the dwarf populations from Cliff (CD) and Indian (ID), which clustered together and separately from all other populations. The second level of signal (geographic signal) grouped population pairs by lake in the remaining part of the tree, most likely reflecting the effect of gene flow and admixture following secondary contact.

Inferred migration links were represented by arrows, the color of which indicates their relative weights (Fig 7). Migrations links between sympatric species-pairs from Cliff and Indian lakes suggested contemporary gene flow between dwarf and normal populations within each of these two lakes. Other migration links between allopatric populations of the same species illustrated the genetic proximities of species from distinct lakes. For instance, the dwarf population from Webster (WD) was related to the dwarf population from Indian lake (ID), and the same link was found between the normal populations of these same lakes (WN and IN). Finally, the ancestral population of East Lake was related with the dwarf population from Indian, whereas the normal population from East Lake (EN) was linked with the normal population from Indian Lake, thus supporting a common genetic background between the normal populations of East and Indian lakes.

## Discussion

*Revisiting the dynamics of secondary contact in whitefish species pairs*

The resolved patterns of genomic differentiation in the five normal and dwarf lake whitefish species pairs provide new insights into the demographic history of a well-studied model of ecological speciation [17]. The approach implemented here relied on inferring the divergence history of each species pair separately, using the JAFS as a summary statistics of genome-wide differentiation patterns. In order to maximize the amount of available information, each JAFS was oriented using the closely related European whitefish as an outgroup species, thus providing increased power to detect demographic processes that generate asymmetric distributions of derived variants about the diagonal of the JAFS.

The secondary contact scenario was the most likely divergence history inferred for four of the five species pairs. Moreover, this scenario was detected both in the least divergent pairs (Témiscouata and East) and the most divergent ones (Indian and Cliff lakes). Therefore, our ability to detect the secondary contact was not entailed by the small degree of genetic divergence between the least differentiated species pairs. This indicates that the erosion of past allopatric isolation by gene flow, which typically generates an excess of shared intermediate frequency alleles, can be retrieved from the JAFS.

The secondary contact scenario is concordant with previous phylogeograhic studies in North-East America, inferred by classical mtDNA, but also clarified several uncertainties about the evolutionary history of the whitefish adaptive radiation. The geographic area where sympatric whitefish species pairs corresponds to a well-known suture zone where glacial lineages came into contact in several freshwater species as they were recolonizing from different refugia after the Laurentide ice sheet retreat [10,54]. In whitefish, this zone corresponds to a phylogeograhic transition between Acadian and Atlantic/Mississippian mitochondrial lineages [11,43]. Interestingly, the Allegash River basin (including the studied lakes), which represents the core of this contact zone, is the only area where sympatric populations of lake whitefish are observed, since no dwarf population, either in allopatry or sympatry, has been reported outside this region [11]. Therefore, phenotypic and ecological divergence, and in particular the occurrence of the dwarf species, is tightly linked with the secondary contact zone.

The frequency of Acadian and Atlantic/Mississippian mitochondrial lineages within lakes was shown to be partly associated with dwarf and normal whitefish level of phenotypic divergence, with variable amounts of mitochondrial admixture being found among lakes [11,43]. At one extreme, the least phenotypically divergent pair from East Lake is characterized by the fixation of the Acadian mitochondrial haplogroup in both dwarf and normal whitefish, which has previously been taken as a support for sympatric divergence in this particular lake [43]. Although our inferences based on the JAFS could not definitely rule out the IM model (IM2mG, $w_{AIC} = 0.19$), we obtained a much stronger evidence in favor of the secondary-contact scenario (SC2N2mG, $w_{AIC} = 0.77$). A possible explanation for the loss of the Atlantic/Mississippian haplogroup in East Lake involves the higher contemporary demographic expansion inducing an asymmetrical $N_e$ that was detected for the dwarf population, which likely contributed to the fixation of the Acadian lineage. Indeed, the preferential gene introgression between sympatric populations with

asymmetrical $N_e$ experiencing gene flow occurs from the larger to the smaller population [56]. Added to the observed direction of the effective gene flow observed for nuclear DNA (nDNA) (Fig 5), the introgression rate for mitochondrial DNA is expected to be higher and faster, again favored by asymmetrical $N_e$ between populations [56]. This interpretation is also supported by the closely similar scenario inferred in the neighboring lake of Témiscouata, which harbours the second least divergent pair. Témiscouata is also dominated by the Acadian haplogroup, but a small proportion of its normal population is still associated with the Atlantic/Mississippian lineage. Since we also inferred an expansion of the dwarf population (but not in the normal whitefish) following secondary contact in this lake, it is likely that this demographic imbalance explains the predominance of Acadian mitochondrial haplotypes in the northern part of the contact zone. At the other extreme, in Cliff Lake, where species divergence is the most pronounced, shows differential fixation of Acadian and Atlantic haplotypes in dwarf and normal populations, respectively [11]. Thus, there is a perfect association in this lake between glacial lineage origin and phenotypic divergence, which was also attributed to a secondary contact in our demographic inferences. Similarly to East Lake, Indian Lake harbored populations fixed for the Acadian haplogroup [55]. However, nDNA markers (*i.e.*, microsatellites) previously indicated admixture between the two sympatric populations [55], suggesting a faster and higher introgression rate for mtDNA than nDNA, during the secondary contact period. Despite higher genetic and phenotypic differentiation due to lower effective gene flow (Fig 5), our analysis of the JAFS also confirmed that two distinct glacial lineages have come into contact in this lake.

Only Webster Lake received unclear results regarding its inferred demographic divergence history. In this lake, both the normal and the dwarf populations display mitochondrial admixture [43]. Consistent with bi-directional gene flow after secondary contact, the Acadian haplogroup is more frequent in the dwarf population, whereas the Atlantic/Mississippian is the most common lineage in the normal population. However, the JAFS was best explained by the AM2m model ($w_{AIC} = 0.74$). Contrary to Cliff and Indian lakes which are both located close to head watersheds, Webster Lake is located downstream of Chamberlain Lake, and its recent history was affected by human activities since the early 1800's. A dam was constructed on the outlet Chamberlain Lake to raise the water level, and a channel was dug between Chamberlain and Webster lakes (Jeremiah Wood, pers. comm.). Then, a dam built on Chamberlain Lake, for water control, connecting the Allagash drainage and the Penobscot drainage, two watersheds that were previously isolated. Therefore, we cannot exclude that contemporary whitefish populations from Webster Lake have been impacted by recent contacts between distinct populations from different lakes. This recent history may have confounded the detection of the secondary contact in this lake, which nevertheless ranks among the best-retained models (SC2NG, $\Delta$AIC < 10).

*A shared history of divergence before independent evolution within lakes*

A global analysis including all five pairs simultaneously was necessary to understand the extent to which replicate whitefish species pairs share a common history of divergence.

The secondary contact scenario implies that the different population pairs are derived from the same two glacial lineages, the evidence of which is partly supported by mitochondrial data [11,43]. However, whether whitefish species pairs share a common history before secondary contact has never been assessed using nuclear markers.

Grouping populations based on their overall genetic similarities with *Treemix* produced two different types of grouping in the population tree. Populations from the three least divergent species pairs were grouped by lake (*i.e.*, TN grouped with TD, EN with ED, and WN with WD), while populations from the two most divergent species pairs were grouped by ecotypes (*i.e.*, IN with CN, and ID with CD). This complex picture likely reflects the relative importance of gene flow between species within lakes and genetic drift among lakes, and is in itself insufficient to distinguish contemporary admixture from shared ancestry during lakes colonization. Inferring migration events among populations enabled us to detect current gene flow between the divergent species of Indian and Cliff lakes. However, the other inferred links connecting populations of the same species but from different lakes rather indicated shared genetic variation due to common ancestry. Namely, inferred links between Webster and Indian indicated the sharing of ancestral variation between WN and IN (and therefore with CN), as well as between WD and ID (and therefore with CD). This supports the view that the different populations of each species in these three lakes, which are not connected today by gene flow, were genetically similar before being isolated in their respective lakes. An additional link inferred between EN and IN (itself connected with WN and CN) confirmed that normal whitefish from East Lake share ancestral variation with other normal populations from the southern part of the contact zone. This provides further evidence that the secondary contact inferred in East Lake has occurred between the same two glacial lineages as for the other lakes despite the lack of Atlantic/Mississippian mitochondrial lineage in this lake. Finally, the ancestral population from East Lake was linked to the dwarf population from Indian (and therefore to WD and CD), indicating that both populations from East Lake share much of the ancestral variation originating from the Acadian lineage. This is also consistent with the genetic swamping hypothesis proposed for explaining the lack of mitochondrial polymorphism in this lake.

The analysis of overall diversity patterns performed with the dAPC (Fig 6) was an alternative way to disentangle remaining signals of genetic differentiation between glacial lineages (axis 3) from genetic differentiation among lakes (axes 1). On the third axis, the projection of dwarf and normal populations from Cliff Lake indicated the positions of the two least introgressed populations of our dataset. Therefore, they could be used to define an Acadian (negative coordinates) and an Atlantic/Mississippian (positive coordinates) reference for comparisons with other lakes. Consistent with the preferential direction in gene flow from dwarf to normal, but the increasing gene flow from normal to dwarf in least differentiated lakes, which was inferred among lakes (Fig 5), dwarf populations were generally shifted towards the Atlantic/Mississippian reference position. Interestingly, both populations from East and Indian lakes occupied intermediate positions, which is concordant with a higher proportion of Acadian ancestry in these lakes as suggested by mitochondrial data [43,55].

In summary, the most parsimonious overall scenario supported by our analyses corresponds to a secondary contact in all lakes, with variable contributions of Acadian and Atlantic/Mississippian lineages due to demographic contingencies. The secondary contact was accompanied by spatial population expansions that were detected for all lakes, excluding Webster. These expansions were observed for both dwarf and normal populations (except for the TN population), consistent with the idea that the two glacial lineages were undergoing spatial expansions after the last glacial retreat, which provoked the secondary contact. Overall, population expansion was more pronounced for dwarf relative to normal populations, still reflected today by the generally higher contemporary abundance of dwarf whitefish in all lakes (L. Bernatchez, unpubl. data), which also impacted the main direction of gene flow. Asymmetric gene flow was consistently inferred between dwarf and normal populations in *∂a∂i*, *Treemix* and dAPC analyses. In addition, the demographic parameters inferred in *∂a∂i* included the effective population size of each population ($N$), their growth parameter ($b$) and effective migration rate ($m_e$), which could be used to estimate the contemporary effective number of migrants per generation (*i.e.*, $Nbm_e$). Supporting previous results, each species pair experienced asymmetrical effective gene flow which was more pronounced from dwarf to normal whitefish than the reverse. As a consequence, an important amount of shared ancestral polymorphism between dwarf and normal populations should correspond flowing variation (*i.e.*, genetic variation coming from gene exchanges) between lineages due to introgression, in addition to incomplete lineage sorting (see Fig 2 for '*shared polymorphism across lakes*' proportions).

*An extended framework for inferring speciation-with-gene-flow*

The concept of speciation-with-gene-flow embraces a large diversity of divergence scenarios with regards to the timing of gene flow, which in turn pertains to different modes of speciation that have long been recognized in the speciation literature [1,30]. Diverging populations can experience temporal variations in effective size ($N_e$) and migration rate ($m$), which both influence the temporal dynamics of gene flow. Consequently, demographic inferences methods that account for these temporal variations have the potential to provide a better understanding of the historical demographic events that shaped the unfolding of speciation.

For the lake whitefish as for other species with a pan-Arctic distribution, the history of divergence has been strongly impacted by quaternary climatic oscillations [2]. Glaciations have drastically restricted species distribution areas provoking geographic isolation among bottlenecked populations [57-59], while inter-glacial periods have allowed secondary contacts between populations expanding from their glacial refugia [4]. Here, accounting for temporal variation in migration rate and effective population size allowed us to determine that the secondary contact between whitefish glacial lineages has occurred contemporarily with population expansions. This later point is of prime importance for understanding the evolution of reproductive isolation, since bottlenecked populations undergoing demographic expansions are likely to fix deleterious alleles [60-62], which could translate into genetic incompatibilities upon secondary contact.

Another important aspect of divergence-with-gene-flow relates to the extent to which the previously described demographic effects interact with selection. The speciation

genomics literature is increasingly integrating the influence of selective processes in historical divergence models (*e.g.*, [23,36,37]), and more generally, in the analytical approaches to relate genomic divergence patterns to the underling evolutionary processes [33]. These selective effects can be separated in two categories. First, genetic barriers caused by local adaptation and reproductive isolation loci can resist introgression, hence reducing gene flow at linked loci [34,35]. This generates local reductions in the effective migration rate of the genomic regions involved in speciation. The second category groups the effect of positive [39] and background selection [63], which cause local reductions in genetic diversity at both selected sites and linked neutral sites. This later selective effect rather corresponds to a reduction in the effective population size of the genomic regions influenced by selection, irrespective to the role that they play in the speciation process. Since gene flow depends both on effective population size and migration rate, both types of selective effects are likely to impact genomic divergence patterns during speciation. Here, we captured these effects separately using divergence-with-gene-flow models that take into account in a simple way the effects of genetic barriers and linked selection.

Accounting for variation in effective migration rate across the genome generally improved the fits to empirical data whatever the model considered (Fig 4B), and the best models for all lakes also included heterogeneous migration rates (-2m). This strongly suggests that the rate of introgression between whitefish glacial lineages has been highly variable across their genome since the beginning of secondary contact, as already observed in other species[37,64,65]. Besides, integrating heterogeneous effective population size in the models (-2N) also generally improved model scores for the two most divergent species pairs (Fig 4C). Moreover, the best models included variation in $N_e$ for three of the five species pairs (East, Indian and Cliff). Therefore, our results also support the view that that linked selection has influenced the patterns of genomic divergence in whitefish. As proposed in earlier studies, this mechanism may be particularly efficient in low-recombining chromosomal regions [33]. In some of our models (SC and AM), we have integrated genome-wide variation in effective population size and effective migration rate. The rationale behind this is that only models that both contain a period of isolation and gene flow enable to dissociate the influence of both sources of chromosomal variation, since only linked selection is at play during periods of geographic isolation. However, it is currently unclear how much the signal contained in empirical polymorphism data can retain distinct signatures for these two selective effects. This will need to be addressed using simulations.

In sum, our approach illustrates the need to take into account both temporal and chromosomal variations in effective population sizes ($N_e$) and migration rates ($m$) when inferring the history of speciation. The 26 divergence models considered here enabled us to evaluate a large diversity of scenarios, and to dissociate selective and demographic effects to improve the inference of the divergence history.

*Understanding the divergence continuum in whitefish*

Lake whitefish species pairs offer a rare opportunity to understand the influence of selection and historical demography on a continuum of phenotypic divergence associated with speciation. Previous works have provided mounting evidence for the role of selection

in shaping genetic and phenotypic divergence across this continuum [17,47,48,52,66,67]. However, the role played by historical demography has never been fully resolved since previous studies largely depended on mitochondrial DNA alone.

Our study brings new evidence supporting previous findings based on mitochondrial DNA that the onset of this young adaptive radiation matched the last glacial period [11,44]. Using a mutation rate of $10^{-8}$ mutations/site/generation and a generation time of 3.5 years, the average divergence time estimated among lakes using our genome-wide SNP dataset was 41,600 years (s.d. 8,100). This corresponds to the late Wisconsin glacial episode during which the Laurentide ice sheet covered the studied region in eastern North America. The average time of secondary contact was dated to 11,200 years (s.d. 5,700), which also corresponds to the glacial retreat period when waters of lake Ontario drained down to the Atlantic [54]. Therefore, the inferred secondary contact between glacial lineages closely matches the chronology of the climatic events in eastern North America. Because of this shared history, contemporary genetic variation in whitefish result from a combination of standing ancestral variation and gene flow between glacial lineages after secondary contact.

Our results suggest that demographic differences among lakes have contributed to shaping the divergence continuum observed among the five lakes. Inferred demographic parameters such as ancestral ($N_1$ and $N_2$) and contemporary ($b_1N_1$ and $b_2N_2$) effective population sizes showed similar asymmetry among lakes (except for Webster) with a generally higher effective population size for the dwarf populations (mainly of Acadian origin). In parallel, the predominant direction of gene flow was from dwarf to normal populations, and the effective number of migrants per generation decreased in the most divergent species pairs. Consequently, even if no correlation could be established between the divergence continuum and effective population size per se, introgression rates tended to be higher in the least divergent species pairs, resulting into a weaker genetic differentiation. Yeaman et al. [68] recently showed that the formation of genomic islands by erosion of divergence following secondary contact depends on the amount of linkage disequilibrium (LD) among selected loci and the intensity of effective migration. Here, we showed that effective migration rate was generally higher in the least differentiated lakes (*i.e.*, Témiscouata, East and Webster), while at the same time, increased LD among islands has been documented in the most divergent lakes [52]. Therefore, the divergence continuum likely implies both the antagonistic effects of divergent selection maintaining LD and introgression eroding past divergence.

Our study also provides new insights on the role of linked selection in shaping patterns of genomic divergence observed among the whitefish species pairs. Namely, we inferred that some genomic regions have experienced a reduction in $N_e$, as predicted under the effect of selection at linked sites [33]. The increasing proportion of genomic regions affected by Hill-Robertson effects along the divergence continuum indicated that the divergence continuum among lakes was also influenced by linked selection. In the light of those observations, we propose that the continuum of genetic divergence in whitefish species pairs is the evolutionary result of a complex interplay between (*i*) genetic divergence between glacial lineages through lineage sorting and mutation accumulation, (*ii*) reduced introgression in genomic regions involved in reproductive isolation due to the accumulation of incompatibilities, (*iii*) divergent selection on phenotypes maintaining LD,

and (*iv*) the independent contingency of demographic events among lakes. The heterogeneous landscape of species divergence in the whitefish system was thus likely built by a combination of selective and demographic factors. Our inferences allowed us to disentangle part of this complex interplay, although many aspects remain to be clarified. In particular, whether selection at linked sites also plays a role in facilitating the accumulation of incompatibilities during allopatry will need to be scrutinized into more details. To conclude, this study provides a clear illustration of the potential benefits to apply the improved models developed here towards disentangling the relative role of selective and demographic processes towards elucidating the complexity of species divergence in any other taxonomic group.

## Materials and Methods

*Sampling and genotyping*

We used RAD-sequencing data from Gagnaire et al. [52] to generate a new genome-wide polymorphism dataset. Previous studies based on these data [52,67] only focused on a subset of 3438 RAD markers that were included in the Whitefish linkage map [53]. Here, we used the total amount of sequence data ($1.7\times10^9$ reads of 101 bp) to document genome-wide variation in five sympatric species pairs occurring in five lakes from the Saint John River basin (Fig 1). For each pair, 20 normal and 20 dwarf individuals were used for RAD-sequencing, but five individuals that received poor sequencing coverage were removed from the dataset. Consequently, the following analyses were performed with 195 individuals, each having an average number of $8.7\text{x}10^6$ of reads.

We also sequenced six European whitefish (*Coregonus lavaretus*), the sister species of the North American lake whitefish, to be used as an outgroup for identifying ancestral and derived alleles at each polymorphic site within lake whitefish. European whitefish were sampled in Skrukkebukta Lake (Norway, 69°34'11.6"N-30°02'31.9"E), which also harbors postglacial sympatric whitefish species pairs [69,70]. RAD libraries were prepared for three individuals from each species, using the same procedure as for American lake whitefish [52].

The *C. lavaretus* raw sequence dataset was filtered using the same criteria as for *C. clupeaformis* [52]. After sequence demultiplexing, the reads were trimmed to a length of 80 bp to avoid sequencing errors due to decreasing data quality near the end of reads. We then used the *Stacks* pipeline (v1.24) for *de novo* RAD tags assembly and individual genotyping [71]. We used a custom *Perl* script to determine an optimal set of assembly parameters for *Ustacks*. A minimal coverage depth of 5x per allele (m=5) and a maximal number of six mismatches between two haplotypes at a same locus within individuals (M=6) were set. We then allowed a maximal number of six mismatches between individuals in *Cstacks* (n=6) to merge homologous RAD tags from different samples. Finally, we used the program *Population* to export a VCF file containing the genotypes of all individuals.

Several filtering steps were then performed with *VCFtools* v0.1.13 [72] to remove miscalled and low-quality SNPs, as well as false variation induced by the merging of paralogous loci. We first removed SNPs with more than 10% missing genotypes in each on the 10 *C. clupeaformis* populations. A lower exclusion threshold of 50% was used for *C. lavaretus* to retain a maximum of orthologous loci with the outgroup. We then filtered for Hardy-Weinberg disequilibrium within each of the 10 *C. clupeaformis* populations, using a p-value exclusion threshold of 0.01. Finally, we merged the filtered datasets of dwarf and normal populations within each lake together with the European whitefish outgroup and kept only loci that passed the previous filters in all three samples. This resulted in five lake-outgroup datasets containing 14812, 22788, 5482, 26149, and 14452 SNPs for Témiscouata, East, Webster, Indian and Cliff lakes, respectively. Variation in SNP number for Webster Lake was associated to a lower coverage that reduced the amount of detectable markers and also an overall reduced level of polymorphism in that lake. Finally, we determined the most parsimonious ancestral allelic state for loci that were monomorphic in the outgroup but polymorphic in *C. clupeaformis*. The resulting oriented SNP datasets

contained 11985, 11315, 5080, 13905 and 9686 SNPs for Témiscouata, East, Webster, Indian and Cliff lakes respectively, that were used to build the unfolded Joint Allele Frequency Spectrum (JAFS) of each lake.

*Inferring the demographic history of divergence with gene flow*

The demographic histories of the five species pairs were inferred using a custom version of the software *∂a∂i* v1.7 [19]. In each lake, the JAFS was projected down to 13 individuals (*i.e.*, 26 chromosomes) per population to avoid remaining missing genotypes and optimize the resolution of the JAFS. We considered 26 models (see Fig 2) that were built on four basic models representing alternative modes of divergence: Strict Isolation (SI), Isolation-with-Migration (IM), Ancient Migration (AM), and Secondary Contact (SC). Each model consists of an ancestral population of size $N_{ref}$ that splits into two populations of effective size $N_1$ and $N_2$ during $T_S$ (SI, IM), $T_{AM}$+$T_S$ (AM), or $T_S$+$T_{SC}$ (SC) generations, possibly exchanging migrants during $T_S$ (IM), $T_{AM}$ (AM), or $T_{SC}$ (SC) generations at rate $m_{e12}$ from population 2 (*i.e.*, normal populations) into population 1 (*i.e.*, dwarf populations), and $m_{e21}$ in the opposite direction. These models were extended to integrate temporal variation in effective population size (-G) by enabling exponential growth using current-to-ancient population size ratio parameters $b_1$ (for dwarf populations) and $b_2$ (for normal populations) to account for expansions or bottlenecks. Variation in effective population size across the genome due to Hill-Robertson effects [38] - *i.e.* local reduction in $N_e$ at linked neutral sites due to the effect of background [63] and positive selection [39] - was modeled by considering two categories of loci (-2N) occurring in proportions $Q$ and 1-$Q$. In order to quantify a mean effect of selection at linked sites, we defined a Hill-Robertson scaling factor (*hrf*), relating the effective population size of loci influenced by selection ($N'_1$=$hrf$×$N_1$ and $N'_2$=$hrf$×$N_2$) to that of neutral loci ($N_1$ and $N_2$). Then, models of divergence with gene flow were extended to account for heterogeneous migration across the genome by considering two categories of loci (-2m). In addition to a first category of loci evolving neutrally (*i.e.*, with migration rates $m_{e12}$ and $m_{e21}$) and occurring in proportion $P$, we considered a second category of loci that occur in proportion 1-$P$, experiencing different effective migration rates $m_{e'12}$ and $m_{e'21}$ due to their linkage with nearby selected genes [37]. Because migration and drift influence gene flow during the whole divergence time in the IM model, the effects of heterogeneous migration and population effective size were evaluated separately (IM2N *vs*. IM2m). However, these effects could be estimated jointly in AM and SC models using the period without gene flow to decouple the effects of migration and drift (-2N2m in addition to -2N and -2m). All models with heterogeneous gene flow were also implemented to allow for population growth (-2NG, -2mG and -2N2mG). Finally, in order to take into account potential errors in the identification of ancestral allelic states, predicted JAFS were constructed using a mixture of correctly oriented and mis-orientated SNPs occurring in proportions $O$ and 1-$O$, respectively.

The 26 models were fitted independently for each lake using successively a hot and a cold simulated annealing procedure followed by 'BFGS' optimization [37]. We ran 25 independent optimizations for each model in order to check for convergence and retained the best one to perform comparisons among models based on Akaike information criterion (AIC). A conservative threshold was applied to retain models with $\Delta AIC_i = AIC_i -$

$AIC_{min} \leq 10$, since the level of empirical support for a given model with a $\Delta AIC_i > 10$ is essentially none [73]. For each lake, the difference in AIC between the worse and the best model $\Delta_{max} = AIC_{max} - AIC_{min}$ was used to obtain a scaled score for each model using:

$$model\ score = \frac{(\Delta_{max} - \Delta AIC_i)}{\Delta_{max}} \tag{1}$$

such that for each lake the worst model takes a score of 0 and the best model takes a score of 1. Akaike weights ($w_{AIC}$) were also computed following equation (2) to estimate the probability that each given model is actually the best one, where $R$ corresponds to the number of models ($R = 26$).

$$w_{AIC} = \frac{e^{\frac{(-\Delta AIC_i)}{2}}}{\sum_{i=1}^{R} e^{\frac{(-\Delta AIC_i)}{2}}} \tag{2}$$

Finally, we converted estimated demographic parameters into biologically meaningful values. We used the optimal multiplicative scaling factor *theta* ($\theta$, obtained as reference from demographic inferences) between model and data to estimate the ancestral effective population size ($N_{ref}$) before split for each lake:

$$N_{ref} = \frac{\theta}{4L\mu} \tag{3}$$

with $\mu$ being the mutation rate (fixed at $10^{-8}$ mutations/site/generation) and $L$ the effective length of the genome explored by our RAD-Seq experiment and estimated as:

$$L = \frac{zy80}{x} \tag{4}$$

where $x$ is the number of SNPs that were originally detected from $y$ RAD tags of 80 bp present in the initial dataset, and $z$ is the number of SNP retained for *∂a∂i* analyses in the lake considered. Estimated times in units of $2N_{ref}$ generations were converted in years assuming a generation time of 3.5 years (*i.e.*, the average between 3 years for dwarf and 4 years for normal whitefish), and estimated migration rates were divided by $2N_{ref}$ to get the proportion of immigrants received by each population every generation.

*Patterns of shared ancestry and admixture among lakes*

We combined the genomic datasets of the five lakes to search for signatures of shared ancestry among replicate species pairs. The five lake-specific datasets used for demographic inferences were merged together and we applied a threshold to focus on

polymorphic loci that were retained after filtering in all five lakes (42558 SNPs in total). For each lake, we determined the fraction of private polymorphisms, the fraction of SNPs shared with at least one other lake or shared across all five lakes. We then measured the proportion of SNPs that were shared between species and private to each species within each lake, as well as among lakes after pooling all dwarf populations together and all normal populations.

To visualize the overall genetic structure and relationships among lakes and species, we performed a discriminant analysis of principal components (dAPC) in *Adegenet* v2.0.0 [74]. We first imputed missing genotypes within each population using a Random Forest regression approach, which provides a more accurate imputation than the replacement of missing genotypes by population mean allele frequency [75]. Imputation was performed using ten iterations with 150 trees using the *randomForestSRC* v1.6.1 package in *Stackr* v.0.1.5 [76], and imputed subdatasets, composed by all individual loci, were subsequently merged to perform the dAPC on a dataset containing 56 812 SNPs.

Finally, we used *TreeMix* v1.12 [77] to infer historical relationships among populations. This method uses the covariance structure of allele frequencies between populations and a Gaussian approximation for genetic drift to build a maximum likelihood graph relating sampled populations to their common ancestor, taking migration events into account to improve the fit to the inferred tree. Migration events between populations are modeled in *TreeMix* by discrete mixture events. Such events may either reflect gene exchange between populations within lakes and/or genetic correlations between geographically isolated populations of the same species, due to the retention of shared ancestral polymorphism within whitefish populations among lakes. In order to avoid interpreting spurious migration signals (*e.g.*, between lakes that are currently not connected), we focused on the main events of gene flow that received the highest weights, which likely correspond to the largest admixture proportions. We thus allowed six migration events to be inferred among branches of the whitefish population tree. The fixed number of migration events was determined to represent only events with important migration weight, for statistical robustness. For this analysis, we used a 20% missing genotype rate per population without imputing missing genotypes to avoid potential biases in the covariance matrix.

**Acknowledgements**
We would like to thank Nicolas Bierne, Anne-Laure Ferchaud, Martin Laporte and Anne-Marie Dion-Côté for insightful discussions, as well as Anne C. Dalziel for commenting an earlier version of this manuscript. We are grateful to Guillaume Côté, Melissa L. Evans and the staff of Maine Department of Inland Fisheries & Wildlife (William Adam, David J. Basley and Jeremiah Wood) for all of their help sampling whitefish and sharing information about watersheds histories. This research was supported by a discovery research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to L.B. L.B also holds the Canadian Research Chair in genomics and conservation of aquatic resources, which funded the research infrastructure for this project. This paper is a contribution to Québec-Océan.

| LAKE | MODEL | MLE | AIC | ΔAIC | *w*AIC | *θ* | Nref | $N_1$ | $N_2$ | *b1* | *b2* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Témiscouata | SC2mG | 1656.47 | 3336.93 | - | 0.68 | 532.15 | 7611.39 | 110.8 | 42.1 | 20.3 | 0.3 |
| | | | | | | | | [32.7; 189.0] | [0.0; 84.6] | [0.0; 68.6] | [0.0; 32.4] |
| Témiscouata | SIG | 1663.23 | 3338.47 | 1.54 | 0.32 | 570.56 | 4350.13 | 33.4 | 49.2 | 95.4 | 0.2 |
| | | | | | | | | [21.7; 45.0] | [47.1; 51.4] | [81.0; 109.8] | [0.0; 1.5] |
| East | SC2N2mG | 909.87 | 1847.74 | - | 0.77 | 212.86 | 1622.96 | 0.4 | 0.4 | 37.0 | 2.2 |
| | | | | | | | | [0.0; 9.5] | [0.0; 6.8] | [0.0; 104.7] | [0.0; 41.1] |
| East | IM2mG | 914.25 | 1850.50 | 2.77 | 0.19 | 982.85 | 7493.63 | 1.0 | 0.1 | 1.8 | 84.9 |
| | | | | | | | | [0.0; 2.2] | [0.0; 1.2] | [0.0; 3.7] | [82.1; 87.7] |
| East | AM2N2m | 916.30 | 1856.60 | 7.71 | 0.02 | 406.95 | 3102.73 | 8.6 | 4.3 | - | - |
| | | | | | | | | [0.0; 12.2] | [0.0; 5.1] | - | - |
| East | AMG | 918.72 | 1855.45 | 8.86 | 0.01 | 775.93 | 5915.95 | 0.6 | 0.2 | 25.8 | 96.3 |
| | | | | | | | | [0.0; 40.0] | [0.0; 11.2] | [7.0; 44.6] | [53.6; 139.0] |
| Webster | AM2m | 510.09 | 1040.18 | - | 0.74 | 125.21 | 3837.23 | 32.2 | 116.7 | - | - |
| | | | | | | | | [32.2; 32.2] | [116.7; 116.8] | - | - |
| Webster | IMG | 513.78 | 1043.55 | 3.38 | 0.14 | 121.13 | 3712.25 | 106.7 | 20.4 | 0.2 | 9.9 |
| | | | | | | | | [105.6; 107.8] | [11.5; 29.3] | [0.0; 1.4] | [0.0; 24.6] |
| Webster | IM | 517.00 | 1046.01 | 5.83 | 0.04 | 119.97 | 3676.57 | 35.4 | 111.3 | - | - |
| | | | | | | | | [35.2; 35.6] | [110.8; 111.9] | - | - |
| Webster | IM2NG | 513.24 | 1046.47 | 6.30 | 0.03 | 127.36 | 3903.28 | 50.9 | 12.8 | 0.5 | 27.1 |
| | | | | | | | | [47.5; 54.4] | [8.4; 17.1] | [0.0; 4.5] | [18.4; 35.7] |
| Webster | SIG | 518.14 | 1048.27 | 8.09 | 0.01 | 119.45 | 3660.85 | 30.3 | 29.5 | 2.0 | 39.9 |
| | | | | | | | | [30.2; 30.4] | [28.3; 30.7] | [1.6; 2.3] | [37.3; 42.5] |
| Webster | SI2NG | 516.20 | 1048.41 | 8.23 | 0.01 | 129.62 | 3972.43 | 53.7 | 19.2 | 0.9 | 64.7 |
| | | | | | | | | [53.2; 54.1] | [18.8; 19.7] | [0.5; 1.2] | [64.0; 65.4] |
| Webster | IM2mG | 513.76 | 1049.52 | 9.35 | 0.01 | 127.30 | 3901.23 | 105.8 | 16.8 | 0.2 | 29.4 |
| | | | | | | | | [104.2; 107.4] | [10.9; 22.7] | [0.0; 2.1] | [17.8; 41.0] |
| Webster | SC2NG | 513.99 | 1049.98 | 9.81 | 0.01 | 114.12 | 3497.43 | 74.4 | 25.2 | 0.4 | 11.7 |
| | | | | | | | | [73.5; 75.4] | [24.5; 25.9] | [0.0; 1.7] | [10.8; 12.7] |
| Indian | SC2N2mG | 1089.03 | 2206.07 | - | 0.94 | 158.61 | 933.35 | 0.6 | 0.3 | 64.0 | 6.3 |
| | | | | | | | | [0.6; 0.6] | [0.0; 2.7] | [62.1; 65.9] | [4.1; 8.6] |
| Indian | SC2NG | 1094.92 | 2211.85 | 5.78 | 0.05 | 1201.86 | 7072.67 | 0.2 | 0.2 | 43.2 | 4.4 |
| | | | | | | | | [0.0; 28.7] | [0.0; 70.8] | [42.4; 44.0] | [0.0; 77.7] |
| Cliff | SC2N2mG | 1006.59 | 2039.18 | - | 0.53 | 123.06 | 1482.41 | 4.1 | 0.9 | 24.7 | 19.6 |
| | | | | | | | | [1.9; 6.4] | [0.4; 1.4] | [20.9; 28.5] | [16.9; 22.3] |
| Cliff | SC2mG | 1007.72 | 2039.43 | 0.25 | 0.46 | 232.67 | 2802.70 | 7.3 | 1.0 | 3.6 | 10.7 |
| | | | | | | | | [0.0; 53.0] | [0.0; 12.1] | [0.0; 132.8] | [0.0; 86.9] |

| LAKE | MODEL | *hrf* | $m_{e12}$ | $m_{e21}$ | $m_{e'12}$ | $m_{e'21}$ | *Tsplit* | *Tpost-split* | *P* | *Q* | *O* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Témiscouata | SC2mG | - | 0.011 | 0.000 | 0.000 | 0.042 | 0.56 | 0.14 | 0.93 | - | 0.99 |
| | | - | [0.000; 21.107] | [0.000; 21.680] | [0.000; 3.610] | [0.000; 3.858] | [0.00; 3.39] | [0.00; 0.91] | [0.22; 0.95] | - | [0.99; 0.99] |
| Témiscouata | SIG | - | - | - | - | - | 0.97 | - | - | - | 0.99 |
| | | - | - | - | - | - | [0.91; 1.02] | - | - | - | [0.84; 0.99] |
| East | SC2N2mG | 0.1 | 18.900 | 37.344 | 1.979 | 1.249 | 3.16 | 1.73 | 0.60 | 0.16 | 0.98 |
| | | [0.0; 0.6] | [7.861; 29.940] | [15.639; 59.050] | [0.000; 7.340] | [0.000; 5.661] | [0.00; 7.50] | [0.93; 2.52] | [0.26; 0.94] | [0.01; 0.37] | [0.96; 0.98] |
| East | IM2mG | - | 14.780 | 0.007 | 0.046 | 5.821 | 0.28 | - | 0.77 | - | 0.97 |
| | | - | [14.570; 14.991] | [0.000; 4.195] | [0.000; 0.546] | [5.493; 6.148] | [0.00; 0.65] | - | [0.59; 0.95] | - | [0.94; 0.99] |
| East | AM2N2m | 0.0 | 26.409 | 2.112 | 0.001 | 4.454 | 0.19 | 0.00 | 0.06 | 0.50 | 0.96 |
| | | [0.0; 0.3] | [14.941; 37.877] | [0.000; 12.290] | [0.000; 3.853] | [0.000; 12.517] | [0.06; 0.33] | [0.00; 0.13] | [0.87; 0.86] | [0.26; 0.74] | [0.96; 0.96] |
| East | AMG | - | 11.344 | 0.037 | - | - | 0.21 | 0.04 | - | - | 0.96 |
| | | - | [0.260; 22.428] | [0.000; 1.202] | - | - | [0.08; 0.34] | [0.00; 0.15] | - | - | [0.95; 0.97] |
| Webster | AM2m | - | 1.337 | 18.091 | 0.010 | 0.214 | 1.33 | 0.00 | 0.05 | - | 0.99 |
| | | - | [1.330; 1.345] | [0.000; 18.091] | [0.000; 2.971] | [0.152; 0.276] | [1.32; 1.34] | [0.00; 2.44] | [0.01; 0.09] | - | [0.98; 0.99] |
| Webster | IMG | - | 0.000 | 0.259 | - | - | 1.36 | - | - | - | 0.99 |
| | | - | [0.000; 2.415] | [0.000; 1.000] | - | - | [1.23; 1.50] | - | - | - | [0.99; 0.99] |
| Webster | IM | - | 0.001 | 0.168 | - | - | 1.40 | - | - | - | 0.99 |
| | | - | [0.000; 2.089] | [0.000; 1.019] | - | - | [1.26; 1.54] | - | - | - | [0.98; 0.99] |
| Webster | IM2NG | 0.3 | 0.019 | 0.140 | - | - | 1.26 | - | - | 0.01 | 0.99 |
| | | [0.1; 0.5] | [0.000; 1.720] | [0.000; 1.389] | - | - | [1.07; 1.45] | - | - | [0.00; 0.03] | [0.98; 0.99] |
| Webster | SIG | - | - | - | - | - | 1.48 | - | - | - | 0.99 |
| | | - | - | - | - | - | [1.30; 1.66] | - | - | - | [0.984; 0.995] |
| Webster | SI2NG | 0.5 | - | - | - | - | 1.30 | - | - | 0.04 | 0.99 |
| | | [0.1; 1.0] | - | - | - | - | [1.08; 1.52] | - | - | [0.02; 0.06] | [0.98; 0.99] |
| Webster | IM2mG | - | 0.170 | 0.776 | 0.066 | 0.011 | 1.28 | - | 0.21 | - | 0.99 |
| | | - | [0.000; 2.748] | [0.569; 0.983] | [0.000; 0.833] | [0.000; 0.995] | [1.20; 1.37] | - | [0.01; 0.95] | - | [0.99; 0.99] |
| Webster | SC2NG | 0.6 | 0.000 | 0.169 | - | - | 0.20 | 1.30 | - | 0.04 | 0.99 |
| | | [0.5; 0.7] | [0.000; 2.939] | [0.000; 1.572] | - | - | [0.00; 0.96] | [1.15; 1.46] | - | [0.01; 0.50] | [0.98; 0.99] |
| Indian | SC2N2mG | 0.2 | 0.809 | 1.906 | 0.155 | 1.014 | 4.50 | 1.31 | 0.60 | 0.41 | 0.97 |
| | | [0.0; 0.4] | [0.000; 1.708] | [1.053; 2.760] | [0.000; 1.098] | [0.994; 1.033] | [3.65; 5.35] | [0.96; 1.65] | [0.19; 0.95] | [0.39; 0.42] | [0.94; 1.00] |
| Indian | SC2NG | 0.2 | 1.676 | 5.703 | - | - | 0.13 | 0.34 | - | 0.50 | 0.95 |
| | | [0.0; 0.4] | [0.000; 42.686] | [0.000; 48.432] | - | - | [0.11; 0.15] | [0.00; 8.00] | - | [0.48; 0.50] | [0.29; 0.99] |
| Cliff | SC2N2mG | 0.2 | 0.088 | 0.024 | 0.074 | 0.008 | 3.12 | 0.89 | 0.48 | 0.40 | 0.99 |
| | | [0.0; 1.0] | [0.000; 1.159] | [0.000; 1.894] | [0.000; 0.494] | [0.000; 7.252] | [2.79; 3.45] | [0.84; 0.94] | [0.36; 0.61] | [0.00; 2.17] | [0.98; 0.99] |
| Cliff | SC2mG | - | 14.629 | 34.813 | 0.085 | 0.006 | 0.65 | 0.49 | 0.05 | - | 0.96 |
| | | - | [0.000; 35.145] | [14.934; 54.692] | [0.000; 3.306] | [0.000; 2.323] | [0.11; 1.20] | [0.00; 0.99] | [0.00; 0.26] | - | [0.95; 0.97] |

**Table 1. Results of model fitting.**

For all lakes, statistics and model parameter values are provided for the fittest models under the fixed threshold of $\Delta AIC_i < 10$ are included. This table contains by ranking order the maximum likelihood (MLE) obtained from model with the smallest Akaike information criterion (AIC), the $\Delta AIC$ value for the corresponding model and the weighted AIC ($w_{AIC}$). Then, the inferred demographic parameters values are scaled by Theta ($\theta$): the ancestral effective population size before population split (Nref); the effective population size after split for dwarf ($N_1$) and normal ($N_2$) populations; the growth coefficient for dwarf ($b_1$) and normal ($b_2$) populations. The $b$ parameter is defined as the ratio of contemporary to past effective population size, with past population size corresponding to splitting time for models SI, IM and AM, and of time of secondary contact for models SC. Exponential populations growths are associated with $b_i>1$ and population effective size decrease with $b_i<1$; the Hill-Robertson factor (*hrf*) corresponds to the degree to which the effective population size is locally reduced due to the effect of background selection and selective sweep effects; migrations rates from normal population to dwarf population ($m_{e12}$) and reciprocally ($m_{e21}$); the effective migration rates for genomic regions under selection ($m_{e'12}$) and ($m_{e'21}$); Time parameters for split of the ancestral population ($Ts$) and migration stop (*i.e.*, $T_{AM}$ for AM models) and start after secondary contact (*i.e.*, $T_{SC}$ for SC models). Finally, the table also contains proportion parameters such as the proportion ($Q$) of the genome with a reduced effective population size due to selection at linked sites; the proportion ($P$) of the genome evolving neutrally and the proportion ($O$) of accurate orientation of our SNPs.

## Figures

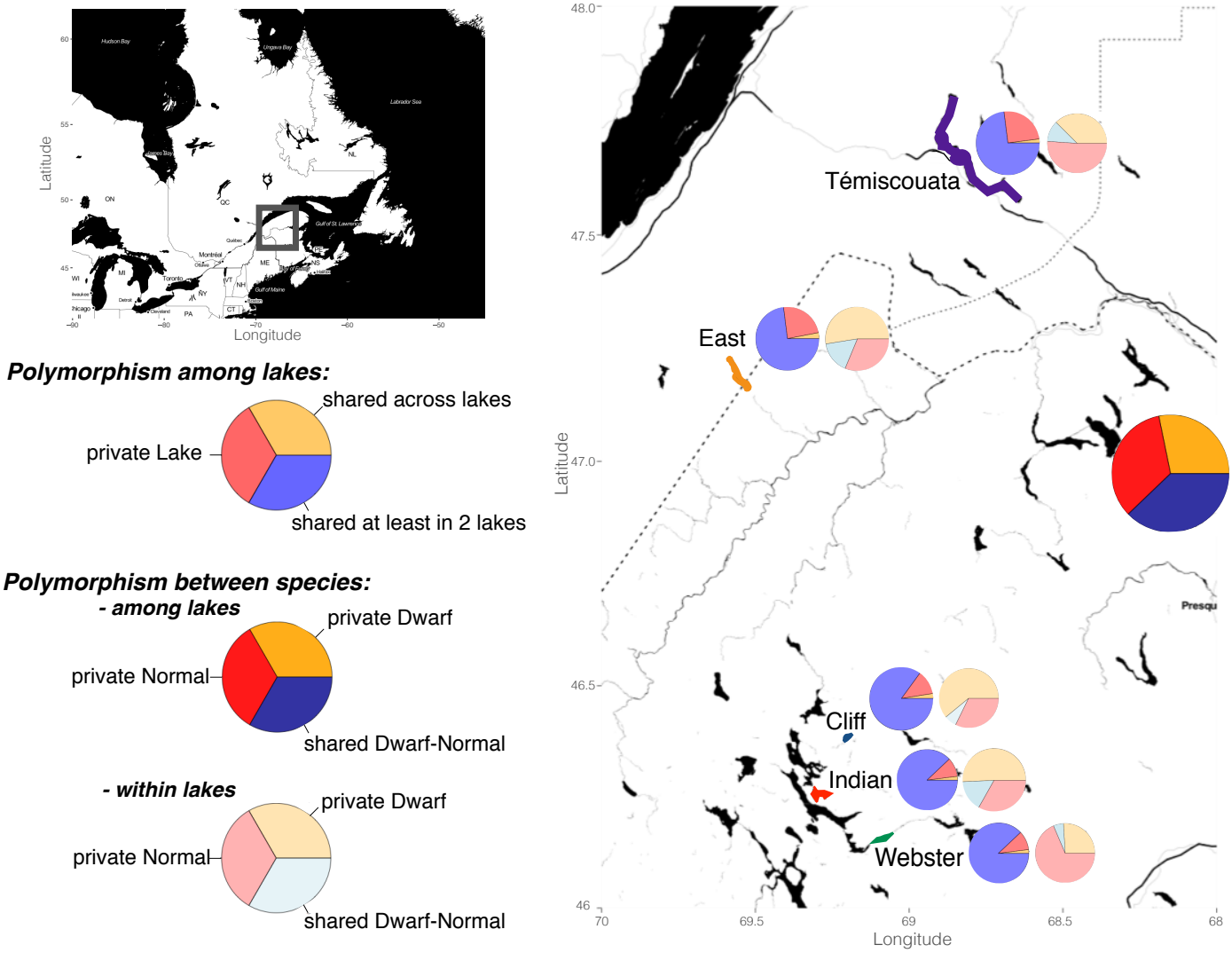


**Figure 1: Geographic locations of lakes where sympatric lake whitefish species pairs were sampled, and overview of the extent of shared polymorphism.** For these five lakes of the southern Québec and Northern Maine (USA) composing the system, each pie chart corresponds to the shared and private polymorphism among lakes and between species ecotypes. The first pie chart aimed to document the polymorphism among lakes in the Lake Whitefish system (‘*Polymorphism among lakes*’), composed by three categories of alleles: ‘private to a lake’, alleles ‘shared in at least two lakes’ and ‘alleles shared across all lakes’. The other pie chart aimed to describe the ‘*Polymorphism between species*’ at two levels; among lakes and within lakes where alleles were categorized as private to a species or shared between species.

| Homo | G | Hetero | | | Hetero.G | | |
|---|---|---|---|---|---|---|---|
| SI | SIG | SI2N | - | - | SI2NG | - | - |
| IM | IMG | IM2N | IM2m | - | IM2NG | IM2mG | - |
| AM | AMG | AM2N | AM2m | AM2N2m | AM2NG | AM2mG | AM2N2mG |
| SC | SCG | SC2N | SC2m | SC2N2m | SC2NG | SC2mG | SC2N2mG |

G: population size change

2N: Heterogeneous effective population size across the genome (linked selection)

2m: Heterogeneous effective migration rate across the genome (differential introgression)

**Figure 2: The 26 models implemented in the study.** All the models implemented in this study were based on the four classical models of divergence ('Simple Isolation'-SI, 'Isolation with Migration'-IM, 'Ancient Migration'-AM and 'Secondary Contact'-SC), accounting for temporal variation in gene flow. These models correspond to the homogeneous models (HOMO) category in our analysis. The second category of models (HOMO.G) includes temporal variation in effective populations size by integrating the -G parameters, allowing independent expansion or reduction of the diverging populations. The third category of models correspond the heterogeneous models (HETERO), which integrate parameters allowing chromosomal variations in migration rate (-2m) and effective population size (-2N) to account for genetic barriers and selection at linked sites. The last category includes temporal variation in effective population size in models of heterogeneous gene flow (HETERO.G).

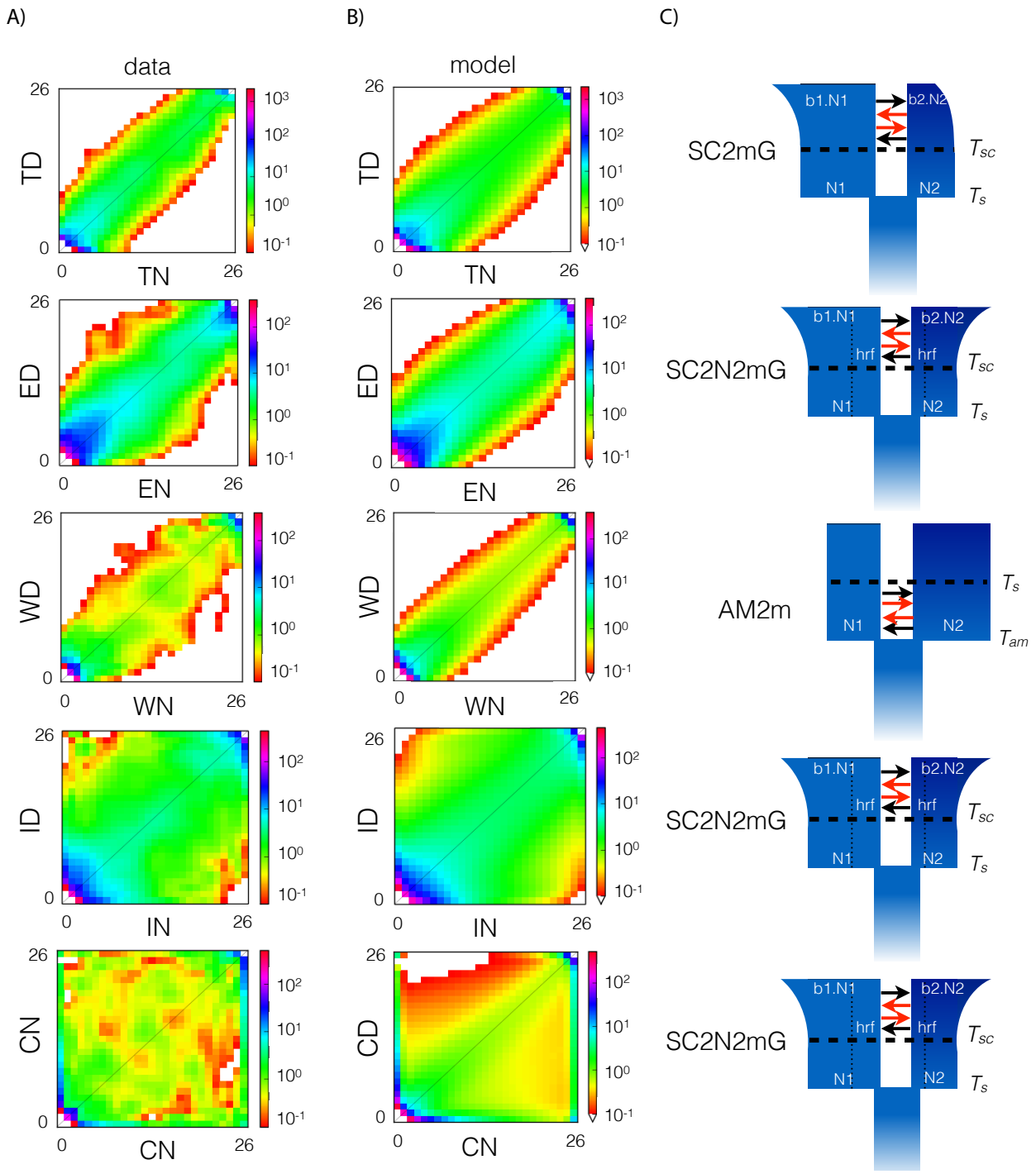


**Figure 3: Historical demography of the lake whitefish species-pairs.** (**A**) Observed joint allele frequency spectrum (JAFS) for normal (-N; x-axis) and dwarf (-D; y-axis) populations for each lake (T=Témiscouata, E=East, W=Webster, I=Indian and C=Cliff), obtained by projection of empirical data. For each JAFS, the color scale corresponds to the probability of occurrence of the derived allele from 13 individuals from each population. (**B**) Predicted JAFS of the fittest model per lake. (**C**) Representation of the fittest model corresponding for each lake.

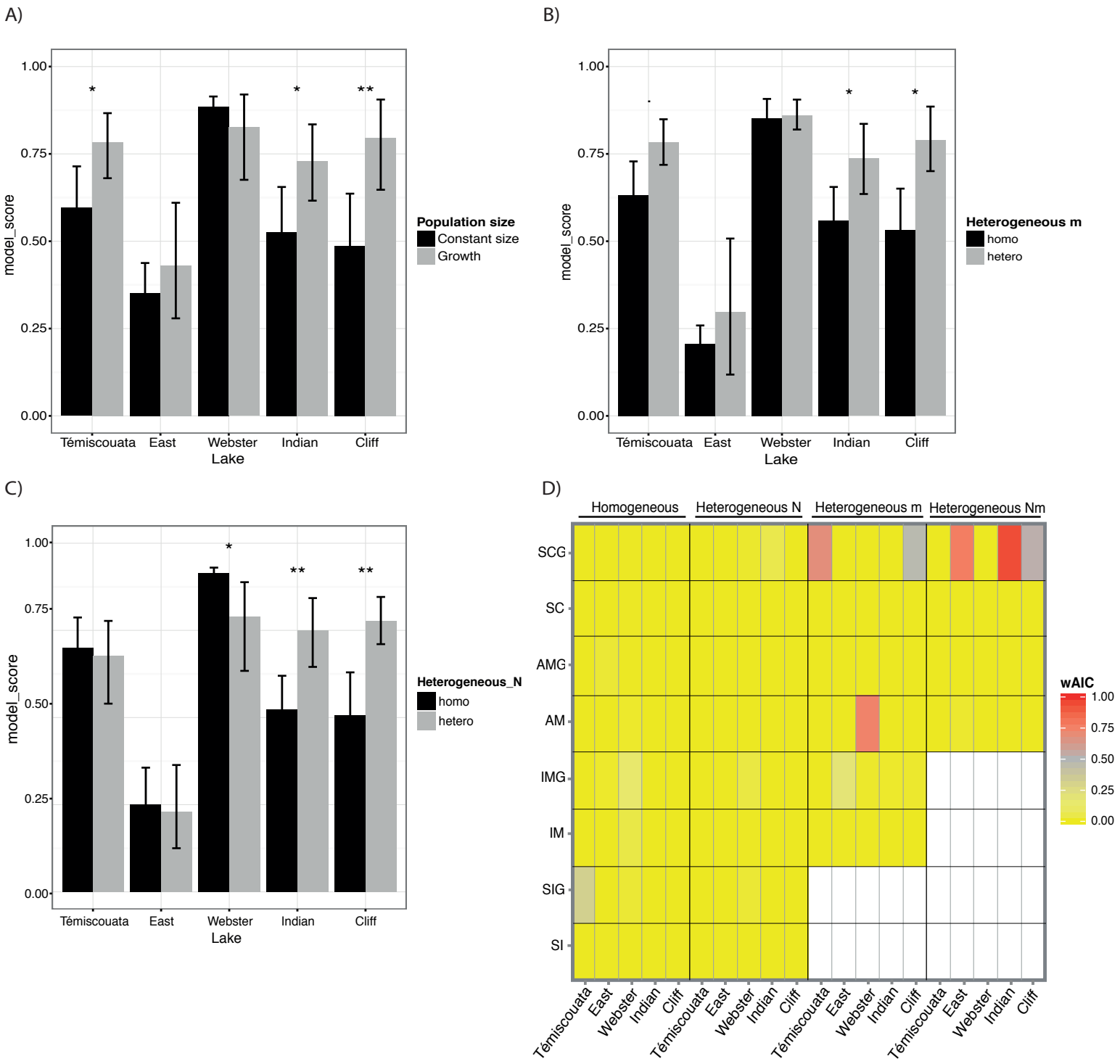


**Figure 4: Parameters effects on inferred models and comparisons among models.** Barplots showing the effect of taking into account particular demographic or selective aspects in the models, assessed using model scores, with (**A**) the effect of including temporal variation in population effective size (-G), (**B**) heterogeneous migration rates among loci (-2m) and (**C**) heterogeneous effective population sizes among loci (-2N). (**D**) Heat-map of the weighted AIC ($w_{AIC}$) showing the relative weights of the 26 models for each lake. The color scale corresponds to the $w_{AIC}$ values ranging from 0 to 1. Warmer colors indicate the best models for each lake.

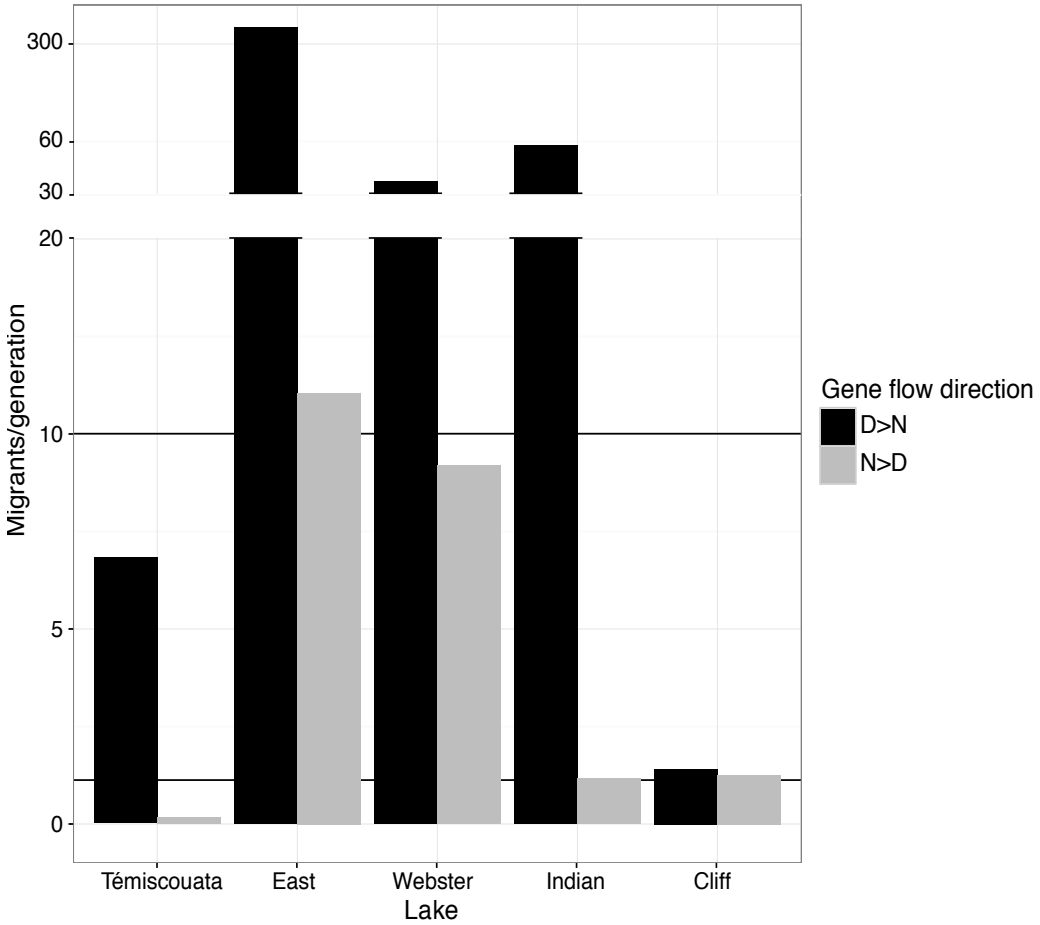


**Figure 5: Asymmetrical effective gene flow between species among lakes.**
Bar plot of the number of migrants per generation for both directions from dwarf to normal (black) and reciprocally (gray), obtained from estimated parameters from the fittest model for each lake.

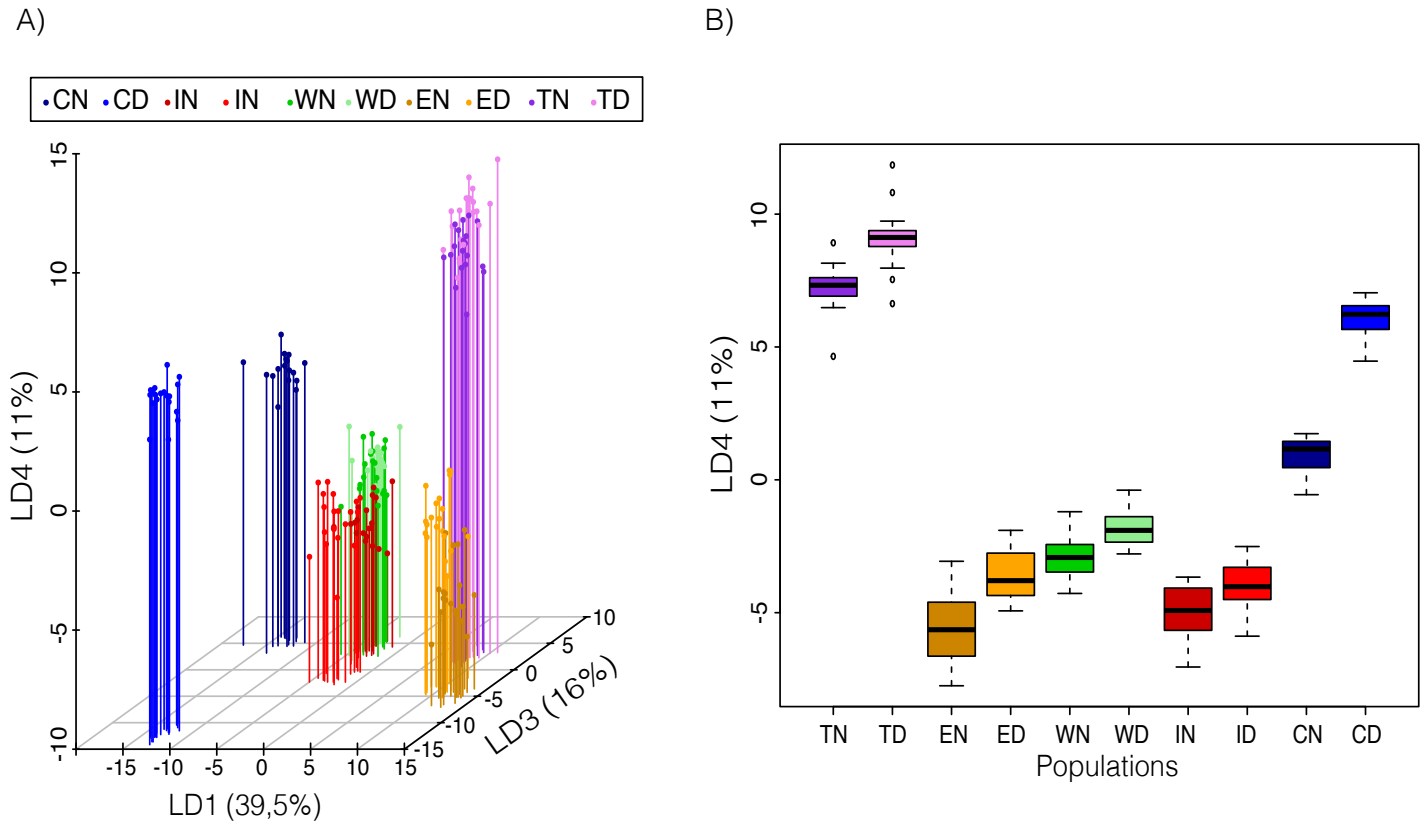


**Figure 6: Genetic structure and relationship among lakes and species.** (**A**) Discriminant analysis of principal components (dAPC) of the different lakes (Cliff, Indian, Webster, East, Témiscouata) for each species (D or N), describing the relationship between populations on three dimensions. First axis (LD1-39.5%), clustering of populations by lakes, corresponds to geographical distribution of the lakes. The LD3 (16%) separates species-pairs and clusters dwarf and normal populations in two distinct groups associated to glacial lineages. Positive coordinates were occupied by normal populations of Atlantic origin and negative coordinates by dwarf populations of Acadian origin. The third axis, LD4 (11%) separated species within each lake. (**B**) Boxplot of the fourth dAPC axis (LD4), highlighting the divergence parallelism between species among lakes.

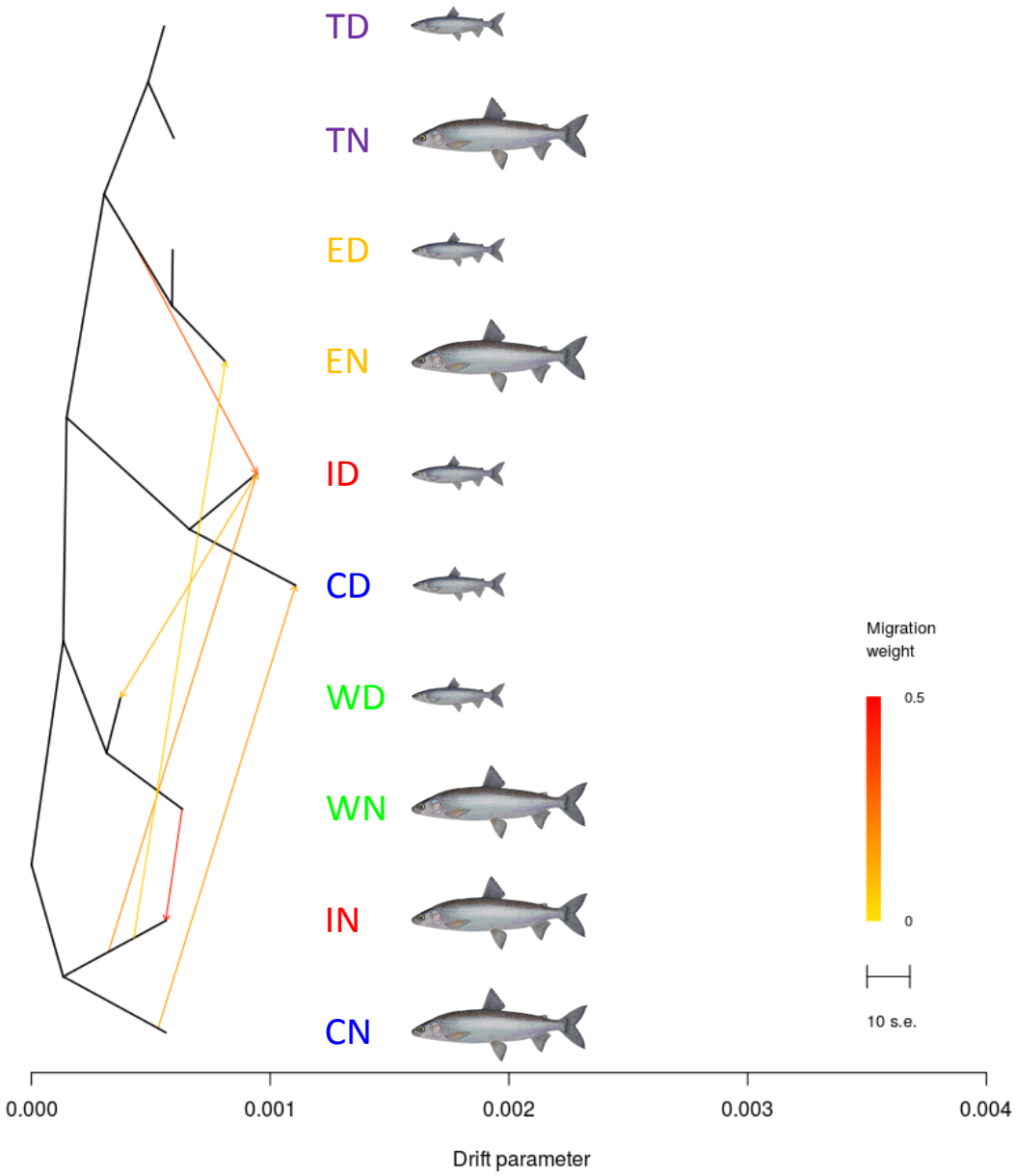


**Figure 7**: **Genetic relationships and inferred migration links among populations.** Color-scale indicates the inferred weight migration events. Inferring migration events among populations enabled us to detect current gene flow between the divergent species-pairs of Indian (IN-ID) and Cliff (CN-CD) lakes. The other inferred links connecting populations of the same species but from different lakes (*e.g.*, WN-IN; WD-ID; IN-EN; ID-ED) rather indicated shared genetic variation due to common ancestry. The drift parameter axis is used as relative temporal measure where the scale bar indicates 10 times the average standard error (s.e.) of the relatedness among populations based on the variance-covariance matrix of allele frequencies.

# Table S1

| LAKE | MODEL | MLE | AIC | ΔAIC | *w*AIC | *θ* | Nref | $N_1$ | $N_2$ | *b1* | *b2* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Témiscouata | SC2mG | -1656.47 | 3336.93 | - | 0.68 | 532.15 | 7611.39 | 2952603.4 | 1120233.2 | 20.3 | 0.3 |
| | | | | | | | | [870315.0; 5034891.7] | [266.4; 2253298.1] | [0.0; 68.6] | [0.0; 32.4] |
| Témiscouata | SIG | -1663.23 | 3338.47 | 1.54 | 0.32 | 570.56 | 4350.13 | 507815.2 | 749808.8 | 95.4 | 0.2 |
| | | | | | | | | [330990.9; 684639.4] | [717244.8; 782372.8] | [81.0; 109.8] | [0.0; 1.5] |
| East | SC2N2mG | -909.87 | 1847.74 | - | 0.77 | 212.86 | 1622.96 | 2325.3 | 2453.9 | 37.0 | 2.2 |
| | | | | | | | | [56.8; 54009.3] | [56.8; 38360.6] | [0.0; 104.7] | [0.0; 41.1] |
| East | IM2mG | -914.25 | 1850.50 | 2.77 | 0.19 | 982.85 | 7493.63 | 27310.7 | 2249.4 | 1.8 | 84.9 |
| | | | | | | | | [262.3; 57150.4] | [262.3; 30703.2] | [0.0; 3.7] | [82.1; 87.7] |
| East | AM2N2m | -916.30 | 1856.60 | 7.71 | 0.02 | 406.95 | 3102.73 | 93159.0 | 47124.4 | - | - |
| | | | | | | | | [108.6; 131971.6] | [108.6; 55127.0] | - | - |
| East | AMG | -918.72 | 1855.45 | 8.86 | 0.01 | 775.93 | 5915.95 | 12292.9 | 4330.7 | 25.8 | 96.3 |
| | | | | | | | | [207.1; 827979.7] | [207.1; 232806.3] | [7.0; 44.6] | [53.6; 139.0] |
| Webster | AM2m | -510.09 | 1040.18 | - | 0.74 | 125.21 | 3837.23 | 432166.7 | 1567970.7 | - | - |
| | | | | | | | | [432166.7; 432166.7] | [1567633.8; 1568307.6] | - | - |
| Webster | IMG | -513.78 | 1043.55 | 3.38 | 0.14 | 121.13 | 3712.25 | 1386634.4 | 264957.2 | 0.2 | 9.9 |
| | | | | | | | | [1372001.6; 1401267.2] | [148792.5; 381122.0] | [0.0; 1.4] | [0.0; 24.6] |
| Webster | IM | -517.00 | 1046.01 | 5.83 | 0.04 | 119.97 | 3676.57 | 455767.2 | 1432599.8 | - | - |
| | | | | | | | | [452848.5; 458685.9] | [1425538.1; 1439661.6] | - | - |
| Webster | IM2NG | -513.24 | 1046.47 | 6.30 | 0.03 | 127.36 | 3903.28 | 695909.3 | 174486.3 | 0.5 | 27.1 |
| | | | | | | | | [648867.1; 742951.5] | [115410.6; 233562.0] | [0.0; 4.5] | [18.4; 35.7] |
| Webster | SIG | -518.14 | 1048.27 | 8.09 | 0.01 | 119.45 | 3660.85 | 387673.3 | 378214.1 | 2.0 | 39.9 |
| | | | | | | | | [386316.3; 389030.4] | [362835.0; 393593.2] | [1.6; 2.3] | [37.3; 42.5] |
| Webster | SI2NG | -516.20 | 1048.41 | 8.23 | 0.01 | 129.62 | 3972.43 | 746073.5 | 267102.4 | 0.9 | 64.7 |
| | | | | | | | | [740020.1; 752126.8] | [260992.5; 273212.2] | [0.5; 1.2] | [64.0; 65.4] |
| Webster | IM2mG | -513.76 | 1049.52 | 9.35 | 0.01 | 127.30 | 3901.23 | 1444451.6 | 229073.5 | 0.2 | 29.4 |
| | | | | | | | | [1422965.2; 1465938.0] | [148184.6; 309962.4] | [0.0; 2.1] | [17.8; 41.0] |
| Webster | SC2NG | -513.99 | 1049.98 | 9.81 | 0.01 | 114.12 | 3497.43 | 911342.4 | 308342.2 | 0.4 | 11.7 |
| | | | | | | | | [899950.7; 922734.1] | [300249.2; 316435.2] | [0.0; 1.7] | [10.8; 12.7] |
| Indian | SC2N2mG | -1089.03 | 2206.07 | - | 0.94 | 158.61 | 933.35 | 2048.1 | 957.3 | 64.0 | 6.3 |
| | | | | | | | | [1984.1; 2112.1] | [32.7; 8971.6] | [62.1; 65.9] | [4.1; 8.6] |
| Indian | SC2NG | -1094.92 | 2211.85 | 5.78 | 0.05 | 1201.86 | 7072.67 | 5306.7 | 4851.4 | 43.2 | 4.4 |
| | | | | | | | | [247.5; 709290.9] | [247.5; 1753775.2] | [42.4; 44.0] | [0.0; 77.7] |
| Cliff | SC2N2mG | -1006.59 | 2039.18 | - | 0.53 | 123.06 | 1482.41 | 21376.3 | 4560.6 | 24.7 | 19.6 |
| | | | | | | | | [9640.8; 33111.8] | [1982.4; 7138.8] | [20.9; 28.5] | [16.9; 22.3] |
| Cliff | SC2mG | -1007.72 | 2039.43 | 0.25 | 0.46 | 232.67 | 2802.70 | 71589.0 | 9480.8 | 3.6 | 10.7 |
| | | | | | | | | [98.1; 519641.2] | [98.1; 118734.2] | [0.0; 132.8] | [0.0; 86.9] |

| LAKE | MODEL | *hrf* | $m_{e12}$ | $m_{e21}$ | $m_{e'12}$ | $m_{e'21}$ | *Tsplit* | *Tpost-split* | *P* | *Q* | *O* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Témiscouata | SC2mG | - | 0.000 | 0.000 | 0.000 | 0.000 | 29996.49 | 7246.04 | 0.93 | - | 0.99 |
| | | - | [0.000; 0.005] | [0.000; 0.005] | [0.000; 0.001] | [0.000; 0.001] | [0.00; 180578.93] | [0.00; 48478.87] | [0.22; 0.95] | - | [0.99; 0.99] |
| Témiscouata | SIG | - | - | - | - | - | 29506.96 | - | - | - | 0.99 |
| | | - | - | - | - | - | [48741.01; 31157.00] | - | - | - | [0.84; 0.99] |
| East | SC2N2mG | 0.11 | 0.020 | 0.040 | 0.002 | 0.001 | 35907.02 | 19601.58 | 0.60 | 0.16 | 0.98 |
| | | [0.01; 0.56] | [0.008; 0.032] | [0.017; 0.064] | [0.000; 0.008] | [0.000; 0.006] | [0.00; 85224.95] | [10542.92; 28660.24] | [0.26; 0.94] | [0.01; 0.37] | [0.96; 0.98] |
| East | IM2mG | - | 0.003 | 0.000 | 0.000 | 0.001 | 14597.50 | - | 0.77 | - | 0.97 |
| | | - | [0.000; 0.004] | [0.000; 0.001] | [0.000; 0.000] | [0.001; 0.001] | [0.00; 33933.42] | - | [0.59; 0.95] | - | [0.94; 0.99] |
| East | AM2N2m | 0.04 | 0.015 | 0.001 | 0.000 | 0.003 | 4204.02 | 0.00 | 0.06 | 0.50 | 0.96 |
| | | [0.01; 0.34] | [0.008; 0.021] | [0.000; 0.007] | [0.000; 0.002] | [0.000; 0.007] | [3263.59; 7077.67] | [0.00; 2730.24] | [0.87; 0.86] | [0.26; 0.74] | [0.96; 0.96] |
| East | AMG | - | 0.003 | 0.000 | - | - | 8725.63 | 1536.75 | - | - | 0.96 |
| | | - | [0.000; 0.007] | [0.000; 0.000] | - | - | [4276.72; 14127.18] | [0.00; 6368.43] | - | - | [0.95; 0.97] |
| Webster | AM2m | - | 0.001 | 0.008 | 0.000 | 0.000 | 35653.78 | 0.00 | 0.05 | - | 0.99 |
| | | - | [0.001; 0.001] | [0.000; 0.008] | [0.000; 0.001] | [0.000; 0.000] | [35401.86; 35905.70] | [0.00; 65494.31] | [0.01; 0.09] | - | [0.98; 0.99] |
| Webster | IMG | - | 0.000 | 0.000 | - | - | 35415.16 | - | - | - | 0.99 |
| | | - | [0.000; 0.001] | [0.000; 0.000] | - | - | [31876.76; 38953.56] | - | - | - | [0.99; 0.99] |
| Webster | IM | - | 0.000 | 0.000 | - | - | 35990.39 | - | - | - | 0.99 |
| | | - | [0.000; 0.001] | [0.000; 0.000] | - | - | [32383.78; 39596.99] | - | - | - | [0.98; 0.99] |
| Webster | IM2NG | 0.28 | 0.000 | 0.000 | - | - | 34499.01 | - | - | 0.01 | 0.99 |
| | | [0.08; 0.47] | [0.000; 0.001] | [0.000; 0.001] | - | - | [29327.19; 39670.83] | - | - | [0.00; 0.03] | [0.98; 0.99] |
| Webster | SIG | - | - | - | - | - | 37893.82 | - | - | - | 0.99 |
| | | - | - | - | - | - | [33197.67; 42589.97] | - | - | - | [0.98; 0.99] |
| Webster | SI2NG | 0.53 | - | - | - | - | 36252.68 | - | - | 0.04 | 0.99 |
| | | [0.09; 0.96] | - | - | - | - | [30137.19; 42368.17] | - | - | [0.02; 0.06] | [0.99; 0.99] |
| Webster | IM2mG | - | 0.000 | 0.000 | 0.000 | 0.000 | 35073.11 | - | 0.21 | - | 0.99 |
| | | - | [0.000; 0.001] | [0.000; 0.000] | [0.000; 0.000] | [0.000; 0.000] | [32731.26; 37414.95] | - | [0.01; 0.95] | - | [0.99; 0.99] |
| Webster | SC2NG | 0.59 | 0.000 | 0.000 | - | - | 4805.63 | 31892.69 | - | 0.04 | 0.99 |
| | | [0.47; 0.72] | [0.000; 0.001] | [0.000; 0.001] | - | - | [0.00; 23478.82] | [28130.88; 35654.50] | - | [0.01; 0.50] | [0.98; 0.99] |
| Indian | SC2N2mG | 0.20 | 0.002 | 0.004 | 0.000 | 0.002 | 29401.99 | 8541.27 | 0.60 | 0.41 | 0.97 |
| | | [0.01; 0.43] | [0.000; 0.003] | [0.002; 0.005] | [0.000; 0.002] | [0.002; 0.002] | [23833.27; 34970.72] | [6279.02; 10803.53] | [0.19; 0.95] | [0.39; 0.42] | [0.94; 1.00] |
| Indian | SC2NG | 0.19 | 0.000 | 0.001 | - | - | 6230.14 | 16705.74 | - | 0.50 | 0.95 |
| | | [0.01; 0.38] | [0.000; 0.011] | [0.000; 0.012] | - | - | [5259.77; 7200.51] | [0.00; 396069.76] | - | [0.48; 0.50] | [0.29; 0.99] |
| Cliff | SC2N2mG | 0.17 | 0.000 | 0.000 | 0.000 | 0.000 | 32375.74 | 9245.76 | 0.48 | 0.40 | 0.99 |
| | | [0.00; 1.00] | [0.000; 0.001] | [0.000; 0.002] | [0.000; 0.001] | [0.000; 0.009] | [28981.58; 35769.90] | [8722.07; 9769.46] | [0.36; 0.61] | [0.00; 2.17] | [0.98; 0.99] |
| Cliff | SC2mG | - | 0.009 | 0.022 | 0.000 | 0.000 | 12818.87 | 9679.91 | 0.05 | - | 0.96 |
| | | - | [0.000; 0.022] | [0.009; 0.034] | [0.000; 0.002] | [0.000; 0.001] | [2190.82; 23446.92] | [0.00; 19505.76] | [0.00; 0.26] | - | [0.95; 0.97] |

**Table S1. Translated results of model fitting, demographic and evolutive process parameters.**

For all lakes, statistics and demographic parameters details of the fittest models under the fixed threshold of $\Delta AIC_i < 10$. The table contains in order the maximum likelihood (MLE) obtained from model with the smallest Akaike information criterion (AIC), the $\Delta AIC$ value for the corresponding model and the weighted AIC ($w_{AIC}$). Then, the inferred demographic parameters values scaled by Theta ($\theta$): the ancestral effective population size before population split (Nref); the effective population size after split for dwarf ($N_1$) and normal ($N_2$) populations; the growth coefficient for dwarf ($b_1$) and normal ($b_2$) populations. The $b$ parameter defined as a ratio of contemporary effective population size of the ancestral populations at time of split (models: SI, IM and AM) and at time of secondary contact (models SC). Population exponential growth is associated with $b_i>1$ and reduction in population effective size with $b_i<1$; the Hill-Robertson factor (hrf) which corresponds to the degree to which the effective population size is locally reduced due to the effect of background selection or selective sweep effects; migrations rates from normal population to dwarf population ($m_{e12}$) and reciprocally ($m_{e21}$); the effective migration rates for genomic regions under selection ($m_{e'12}$) and ($m_{e'21}$); temporal estimates (in years) regarding the split of the ancestral population (Tsplit) and the time when the migration stopped (*i.e.*, $T_{AM}$ for the AM based models) and started after secondary contact (*i.e.*, $T_{SC}$). The table also include parameters related to genomic characteristics: the proportion (Q) of the genome with a reduced effective size due to selection at linked sites; the proportion (P) of the genome evolving neutrally and the proportion (O) of accurate orientation of our SNPs.

**References**


1. Coyne JA, Orr HA. Speciation. Sunderland MA Sinauer Associates Inc; 2004.

2. Bernatchez L, Wilson CC. Comparative phylogeography of Nearctic and Palearctic fishes. Molecular Ecology. 1998;7: 431–452.

3. Avise JC. Phylogeography: the history and formation of species. Cambridge MA Harvard University Press; 2000.

4. Hewitt GM. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. Molecular Ecology. 2001;10: 537–549. doi:10.1046/j.1365-294x.2001.01202.x

5. Hewitt GM. Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society. 1996;58: 247–276. doi:10.1111/j.1095-8312.1996.tb01434.x

6. Hewitt G. The genetic legacy of the Quaternary ice ages. Nature. 2000;405: 907–913. doi:10.1038/35016000

7. Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2004;359: 183–95–discussion 195. doi:10.1098/rstb.2003.1388

8. Swenson NG, Howard DJ. Clustering of Contact Zones, Hybrid Zones, and Phylogeographic Breaks in North America. Am Nat. 2005;166: 581–591. doi:10.1086/491688

9. Bierne N, Welch J, Loire E, Bonhomme F, David P. The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology. Blackwell Publishing Ltd; 2011;20: 2044–2072. doi:10.1111/j.1365-294X.2011.05080.x

10. April J, Hanner RH, Dion-Côté A-M, Bernatchez L. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. Molecular Ecology. 2013;22: 409–422. doi:10.1111/mec.12116

11. Bernatchez L, Dodson JJ. Allopatric origin of sympatric populations of lake whitefish (Coregonus clupeaformis) as revealed by mitochondrial-DNA restriction analysis. Evolution. 1990;24: 890–908. doi:10.1111/mec.13031

12. McPhail JD. Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. Canadian Journal of Zoology. 1992;70: 361–369. doi:10.1139/z92-054

13. Taylor EB, Bentzen P. Evidence for Multiple Origins and Sympatric Divergence of Trophic Ecotypes of Smelt (Osmerus) in Northeastern North America. Evolution. 1993;47: 813. doi:10.2307/2410186

14. Schluter D. Ecological Speciation in Postglacial Fishes. Philosophical Transactions of the Royal Society of London B: Biological Sciences. The Royal Society; 1996;351: 807–814. doi:10.1098/rstb.1996.0075

15. Wood CC, Foote CJ. Evidence for Sympatric Genetic Divergence of Anadromous and Nonanadromous Morphs of Sockeye Salmon (Oncorhynchus nerka). Evolution. 1996;50: 1265. doi:10.2307/2410667

16. Taylor EB. Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation. Reviews in Fish Biology and Fisheries. 1999;9: 299–324.

17. Bernatchez L, Renaut S, Whiteley AR, Derome N, Jeukens J, Landry L, et al. On the origin of species: insights from the ecological genomics of lake whitefish. Philos Trans R Soc Lond, B, Biol Sci. 2010;365: 1783–1800. doi:10.1098/rstb.2009.0274

18. Welch JJ, Jiggins CD. Standing and flowing: the complex origins of adaptive variation. Molecular Ecology. 2014;23: 3935–3937. doi:10.1111/mec.12859

19. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. McVean G, editor. PLoS Genet. 2009;5: e1000695. doi:10.1371/journal.pgen.1000695

20. Excoffier L, Dupanloup I, Huerta-SÃ nchez E, Sousa VC, Foll M. Robust Demographic Inference from Genomic and SNP Data. Akey JM, editor. PLoS Genet. 2013;9: e1003905–17. doi:10.1371/journal.pgen.1003905

21. Butlin RK, Saura M, Charrier G, Jackson B, André C, Caballero A, et al. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. Evolution. 2014;68: 935–949. doi:10.1111/evo.12329

22. Nadachowska-Brzyska K, Burri R, Olason PI, Kawakami T, Smeds LA, Ellegren H. Demographic Divergence History of Pied Flycatcher and Collared Flycatcher Inferred from Whole-Genome Re-sequencing Data. Payseur BA, editor. PLoS Genet. 2013;9: e1003942. doi:10.1371/journal.pgen.1003942.s007

23. Sousa V, Hey J. Understanding the origin of species with genome-scale data: modelling gene flow. Nature Publishing Group. 2013;14: 404–414. doi:10.1038/nrg3446

24. Hey J, Nielsen R. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics. Genetics Society of America; 2004;167: 747–760. doi:10.1534/genetics.103.024182

25. Hey J, Nielsen R. Integration within the Felsenstein Equation for Improved Markov Chain Monte Carlo Methods in Population Genetics. Proc Natl Acad Sci USA. National Academy of Sciences; 2007;104: 2785–2790. doi:10.2307/25426552?ref=search-gateway:7de81229ed7a0e39440b24fe4990ef55

26. Becquet C, Przeworski M. A new approach to estimate parameters of speciation models with application to apes. Genome Research. Cold Spring Harbor Lab; 2007;17: 1505–1519. doi:10.1101/gr.6409707

27. Pinho C, Hey J. Divergence with gene flow: models and data. Annual review of ecology. 2010;41: 215–230. doi:10.1146/annurev-ecolsys-102209-144644

28. Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, et al. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics. Oxford University Press; 2014;30: 1187–1189. doi:10.1093/bioinformatics/btt763

29. Duvaux L, Belkhir K, Boulesteix M, Boursot P. Isolation and gene flow: inferring the speciation history of European house mice. Molecular Ecology. Blackwell Publishing Ltd; 2011;20: 5248–5264. doi:10.1111/j.1365-294X.2011.05343.x

30. Nosil P. Speciation with gene flow could be common. Molecular Ecology. 2008;17: 2103–2106. doi:10.1111/j.1365-294X.2008.03715.x

31. Feder JL, Egan SP, Nosil P. The genomics of speciation-with-gene-flow. Trends in Genetics. 2012;28: 342–350. doi:10.1016/j.tig.2012.03.009

32. Harrison RG, Larson EL. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Molecular Ecology. 2016;: n/a–n/a. doi:10.1111/mec.13582

33. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology. 2014;23: 3133–3157. doi:10.1111/mec.12796

34. Barton NH, Bengtsson BO. The barrier to genetic exchange between hybridising populations. Heredity. 1986;57: 357–376. doi:10.1038/hdy.1986.135

35. Feder JL, Nosil P. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution. Society for the Study of Evolution; 2010;64: 1729–1747. doi:10.2307/40664014?ref=search-gateway:d9aab692ab6140e4c7d2470dfd85ad92

36. Roux C, Tsagkogeorga G, Bierne N, Galtier N. Crossing the Species Barrier: Genomic Hotspots of Introgression between Two Highly Divergent Ciona intestinalis Species. Molecular Biology and Evolution. 2013;30: 1574–1587. doi:10.1093/molbev/mst066

37. Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST, et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nature Communications. Nature Publishing Group; 2014;5: 1–10. doi:10.1038/ncomms6770

38. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8: 269–294. doi:10.1017/S0016672300010156

39. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. Cambridge University Press; 1974;23: 23–35. doi:10.1017/S0016672300014634

40. Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res. 1997;70: 155–174.

41. Corbett-Detig RB, Hartl DL, Sackton TB. Natural Selection Constrains Neutral Diversity across A Wide Range of Species. Barton NH, editor. PLoS Biol. Public Library of Science; 2015;13: e1002112. doi:10.1371/journal.pbio.1002112

42. Ewing GB, Jensen JD. The consequences of not accounting for background selection in demographic inference. Molecular Ecology. 2016;25: 135–141. doi:10.1111/mec.13390

43. Pigeon D, Chouinard A, Bernatchez L. Multiple Modes of Speciation Involved in the Parallel Evolution of Sympatric Morphotypes of Lake Whitefish (Coregonus clupeaformis, Salmonidae). Evolution. 1997;51: 196. doi:10.2307/2410973

44. Bernatchez L, Dodson JJ. Phylogeographic structure in mitochondrial DNA of the lake whitefish (Coregonus clupeaformis) and its relation to Pleistocene glaciations. Evolution. 1991;45: 1016–1035.

45. Bernatchez L. Ecological theory of adaptive radiation: an empirical assessment from coregonine fishes (Salmoniformes). Pages 175-207 *in* A.P. Hendry and S.C. Stearns, eds. Evolution illuminated: salmon and their relatives. Oxford Univ. Press, Oxford ; 2004

46. Lu G, Bernatchez L. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (Coregonus clupeaformis): support for the ecological speciation hypothesis. Evolution. 1999;53: 1491–1505.

47. Rogers SM, Gagnon V, Bernatchez L. Genetically based phenotype-environment association for swimming behavior in lake whitefish ecotypes (Coregonus clupeaformis Mitchill). Evolution. 2002;56: 2322–2329.

48. Landry L, VINCENT WF, Bernatchez L. Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. J Evol Biol. 2007;20: 971–984. doi:10.1111/j.1420-9101.2007.01304.x

49. Rogers SM, Bernatchez L. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonus sp. Salmonidae) species pairs. Molecular Biology and Evolution. 2007;24: 1423–1438. doi:10.1093/molbev/msm066

50. Landry L, Bernatchez L. Role of epibenthic resource opportunities in the parallel evolution of lake whitefish species pairs (Coregonus sp.). J Evol Biol. 2010;23: 2602–2613.

doi:10.1111/j.1420-9101.2010.02121.x

51. Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM, et al. Genome-wide patterns of divergence during speciation: the lake whitefish case study. Philos Trans R Soc Lond, B, Biol Sci. 2012;367: 354–363. doi:10.1098/rstb.2011.0197

52. Gagnaire P-A, Pavey SA, Normandeau E, Bernatchez L. The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution. 2013;67: 2483–2497. doi:10.1111/evo.12075

53. Gagnaire P-A, Normandeau E, Pavey SA, Bernatchez L. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (Coregonus clupeaformis). Molecular Ecology. 2013;22: 3036–3048. doi:10.1111/mec.12127

54. Curry RA. Late glacial impacts on dispersal and colonization of Atlantic Canada and Maine by freshwater fishes. Quaternary Research. 2007;67: 225–233. doi:10.1016/j.yqres.2006.11.002

55. Lu G, Basley DJ, Bernatchez L. Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (Coregonus clupeaformis); relevance for …. Molecular Ecology. 2001;10: 965–985. doi:10.1046/j.1365-294X.2001.01252.x

56. Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P. Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (Microtus agrestis). J Evol Biol. 2011;25: 400–408.

57. Bernatchez L, Dodson JJ, Boivin S. Population bottlenecks: influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. Journal of Fish Biology. 1989;35: 233–244. doi:10.1111/j.1095-8649.1989.tb03066.x

58. Ambrose SH. Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. Journal of Human Evolution. 1998;34: 623–651. doi:10.1006/jhev.1998.0219

59. Aoki K, Kato M, Murakami N. Glacial bottleneck and postglacial recolonization of a seed parasitic weevil, Curculio hilgendorfi, inferred from mitochondrial DNA variation. Molecular Ecology. 2008;17: 3276–3289. doi:10.1111/j.1365-294X.2008.03830.x

60. Luikart G, Allendorf FW, Cornuet JM, Sherwin WB. Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity. Oxford University Press; 1998;89: 238–247. doi:10.1093/jhered/89.3.238

61. Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L. On the accumulation of deleterious mutations during range expansions. Molecular Ecology. 2013;22: 5972–5982. doi:10.1111/mec.12524

62. Lohmueller KE. The impact of population demography and selection on the genetic architecture of complex traits. Williams SM, editor. PLoS Genet. 2014;10: e1004379. doi:10.1371/journal.pgen.1004379

63. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134: 1289–1303.

64. Le Moan A, Gagnaire PA, Bonhomme F. Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. Molecular Ecology. 2016;25: 3187–3202. doi:10.1111/mec.13627

65. Rougemont Q, Gagnaire P-A, Perrier C, Genthon C, Besnard A-L, Launey S, et al. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Molecular Ecology. 2016. doi:10.1111/mec.13664

66. Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM, et al. Genome-wide patterns of divergence during speciation: the lake whitefish case study. Philos Trans R Soc Lond, B, Biol Sci. 2012;367: 354–363. doi:10.1098/rstb.2011.0197

67. Laporte M, Rogers SM, Dion-Côté A-M, Normandeau E, Gagnaire P-A, Dalziel AC, et al. RAD-QTL Mapping Reveals Both Genome-Level Parallelism and Different Genetic Architecture Underlying the Evolution of Body Shape in Lake Whitefish (Coregonus clupeaformis) Species Pairs. G3 (Bethesda). 2015;5: 1481–1491. doi:10.1534/g3.115.019067

68. Yeaman S, Aeschbacher S, Bürger R. The evolution of genomic islands by increased establishment probability of linked alleles. Abbott RJ, Barton NH, Good JM, editors. Molecular Ecology. 2016;25: 2542–2558. doi:10.1111/mec.13611

69. Amundsen P-A, Bøhn T, Våga GH. Gill raker morphology and feeding ecology of two sympatric morphs of European whitefish (Coregonus lavaretus). Annales Zoologici Fennici. 2004;: 291–300. doi:10.2307/23736213

70. Østbye K, Amundsen P-A, Bernatchez L, Klemetsen A, Knudsen R, Kristoffersen R, et al. Parallel evolution of ecomorphological traits in the European whitefish Coregonus lavaretus (L.) species complex during postglacial times. Molecular Ecology. 2006;15: 3983–4001. doi:10.1111/j.1365-294X.2006.03062.x

71. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Molecular Ecology. 2013;22: 3124–3140. doi:10.1111/mec.12354

72. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330

73. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical

information-theoretic approach. 2002.

74. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11: 94. doi:10.1186/1471-2156-11-94

75. Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, et al. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. The Plant Genome Journal. 2012;5: 103–11. doi:10.3835/plantgenome2012.06.0006

76. Thierry Gosselin, Louis Bernatchez, (2016). stackr: GBS/RAD Data Exploration, Manipulation and Visualization using R. R package version 0.2.1. https://github.com/thierrygosselin/stackr. doi : 10.5281/zenodo.19647

77. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012;8: e1002967. doi:10.1371/journal.pgen.1002967