MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 409–422 doi: 10.1111/mec.12116

Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes

JULIEN APRIL,* ROBERT H. HANNER,† ANNE-MARIE DION-CÔTÉ* and LOUIS BERNATCHEZ*

Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Biologie, Université Laval, Québec, QC, Canada G1V 0A6, †Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada N1G 2W1

Abstract

Allopatric speciation may be the principal mechanism generating new species. Yet, it remains difficult to judge the generality of this process because few studies have provided evidence that geographic isolation has triggered the development of reproductive isolation over multiple species of a regional fauna. Here, we first combine results from new empirical data sets (7 taxa) and published literature (9 taxa) to show that the eastern Great Lakes drainage represents a multispecies suture zone for glacial lineages of freshwater fishes with variable levels of genetic divergence. Second, we performed amplified fragment length polymorphism analyses among four pairs of lineages. Results indicate that lineages with relatively deep levels of mtDNA 5′ COI (barcode) sequence divergence (>2%) developed strong reproductive barriers, while lineages with lower levels of divergence show weaker reproductive isolation when found in sympatry. This suggests that a threshold of 2% sequence divergence at mtDNA could be used as a first step to flag cryptic species in North American freshwater fishes. By describing different levels of divergence and reproductive isolation in different co-occurring fishes, we offer strong evidence that allopatric speciation has contributed significantly to the diversification of north-eastern American freshwater fishes and confirm that Pleistocene glacial cycles can be viewed as a 'speciation pump' that played a predominant role in generating biodiversity.

Keywords: amplified fragment length polymorphism, comparative phylogeography, DNA barcode, hybrid, mtDNA, suture zone

Received 6 May 2012; revision received 23 August 2012; accepted 29 August 2012

Introduction

Understanding the underlying diversification mechanisms that generated biodiversity represents a fundamental goal of both evolutionary and conservation biology. The formation of new species through geographic isolation, or allopatric speciation, has been widely considered the most common mechanism generating new species (Coyne & Orr 2004). Support for this affirmation is largely based on observational inference whereby species borders and genetic discontinuities within species often correspond to geographic barriers (Jordan 1908; Coyne & Orr 2004). Yet, despite the fact that demonstrating the presence of reproductive isolation in secondary contact zones represents the ultimate 'acid test' of allopatric speciation (Coyne & Orr 2004), such empirical studies are surprisingly scarce. Speciation studies have even more rarely detailed and compared patterns of allopatric divergence followed by genetic introgression in secondary contact zones over multiple species (Edmands 2002). Based on the literature, it therefore remains difficult to empirically support and thus generalize the role of allopatric speciation in the diversification process of a specific fauna. Strong empirical evidence of the general importance of allopatric speciation would consist of demonstrating that many co-occurring taxa are at different stages of allopatric divergence and speciation (Coyne & Orr 2004). The underlying rationale is that 'if all phases of the process

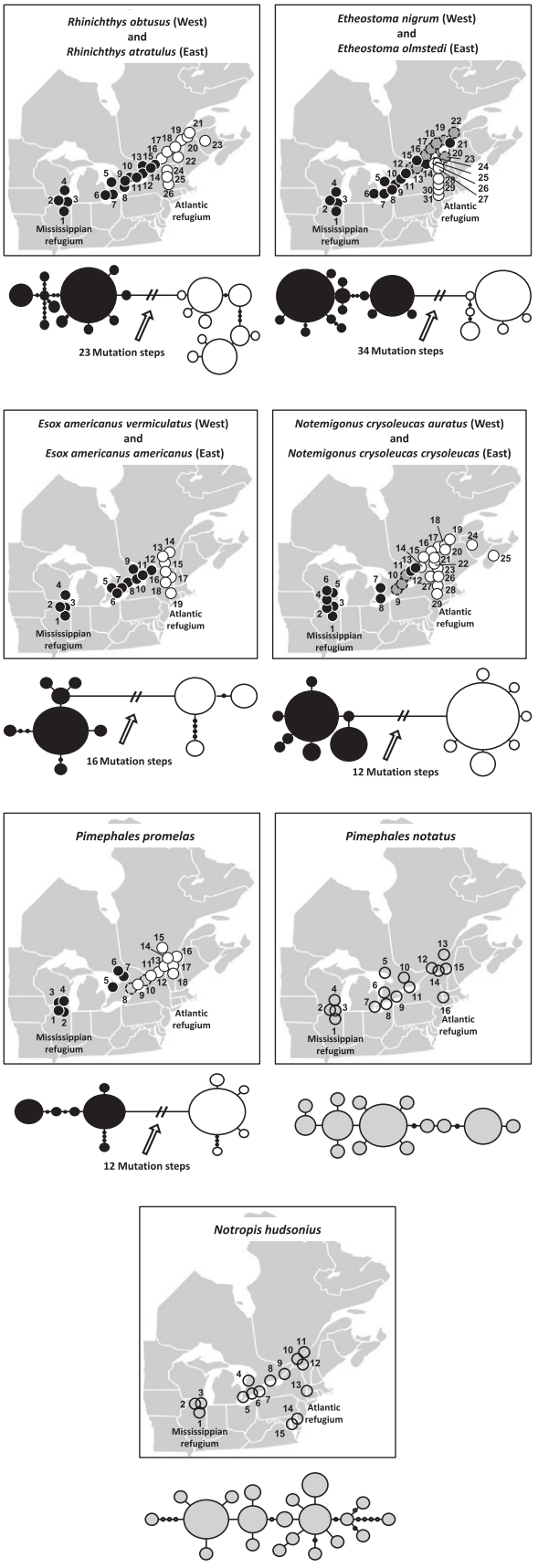
Correspondence: Julien April, Fax: 1 418 656 7176;
E-mail: julien.april.1@ulaval.ca

can be seen in nature, then it is reasonable to conclude that the entire process can occur in a single taxon' (Jiggins & Mallet 2000; Coyne & Orr 2004).

The clustering of many secondary contact zones in nature has been described as a 'suture zone' (Remington 1968) and represents a particularly suitable natural laboratory for the study of speciation. Each of the individual secondary contact zones represents a test of reproductive isolation facilitating the study of permeability of gene exchange between the previously allopatric genetic lineages (Barton & Hewitt 1985; Harrison 1993). The outcome of such secondary contact may be highly variable. On the one hand, some hybrid zones may extend over large geographic areas and be characterized by a high frequency of hybrids compared with the frequency of individuals from both parental groups. They typically represent cases where there is weak or no selection against hybrids or against gene transfer (Barton & Hewitt 1985; Jiggins & Mallet 2000). On the other hand, some hybrid zones are geographically very narrow and characterized by an excess of genetically pure individuals compared with hybrids. This usually suggests cases where reproductive isolation is pronounced and speciation almost completed (Barton & Hewitt 1985; Jiggins & Mallet 2000). However, suture zones have rarely been identified (Swenson & Howard 2004) and even less often analysed in a speciation context (but see Moritz *et al.* 2009; Dasmahapatra *et al.* 2010).

The repeated glacial cycles of the Pleistocene have long been viewed as having the potential to drive a 'speciation pump' (*sensu* Haffer 1969) by promoting the formation of new species through geographic isolation (Haffer 1969; Avise *et al.* 1998; Bernatchez & Wilson 1998; Johnson & Cicero 2004). This theory is the subject of debate however, and several studies have concluded that most species are older than the Pleistocene (Klicka & Zink 1997; Knapp & Mallet 2003). Therefore, it remains unclear whether geographic isolation can trigger the onset of reproductive isolation in a matter of just a few hundreds of thousands of generations. It also remains unclear whether many different species of the same fauna commonly respond to climatic variation and diversification process in a similar fashion (Solstis *et al.* 2006).

Pleistocene glacial advances and retreats may have promoted the formation of 'suture zones' among North American freshwater fishes. According to zoogeographic studies based on geographic distribution and morphological data, more than 30 North American temperate fish taxa may comprise distinct evolutionary lineages that presumably diverged in the same two important glacial refugia (Bailey & Smith 1981; Hocutt & Wiley 1986; Mandrak & Crossman 1992). The Mississippian refugium (Fig. 1) is considered as the most important refugium in North America for freshwater fishes and was located in



the Mississippian River basin, south of the ice sheet near the contemporary states of Missouri and Illinois. The Atlantic coast refugium (Fig. 1) was located on the Atlantic coastal plain, east of the Appalachians Mountains, south of Long Island. For nine of these thirty taxa, phylogeographic and population genetic analyses have confirmed the existence of Atlantic and Mississippian glacial lineages. Most of these taxa with glacial lineage pairs meet between the locations of the two refugia, in the Laurentian Great Lakes watershed.

Probing pairs of Atlantic and Mississippian glacial lineages could illuminate key steps in the process of allopatric speciation among North American freshwater fishes. Levels of morphological divergence between populations (presumed to belong to either Mississippian or Atlantic glacial lineages) vary greatly among taxa (Bailey & Smith 1981; Hocutt & Wiley 1986; Mandrak & Crossman 1992). While most glacial lineages are morphologically undifferentiated and do not have particular taxonomic status, eight pairs of potential glacial lineages are recognized as distinct subspecies or species based on morphological differences. To date, phylogeographic studies have only been conducted on a single fish species (*Fundulus diaphanus diaphanus*/*F. d. menona*) putatively comprising morphologically differentiated glacial lineages associated with the Atlantic and Mississippian refugia. Specifically, April & Turgeon (2006) confirmed that these subspecies represent glacial lineages of Mississippian and Atlantic coast origin that presumably originated during the Pleistocene. Notably, the levels of mtDNA sequences divergence between glacial lineages documented in published phylogeographic studies are variable (ranging from 0.4% to 3.2%), strengthening the idea that those pairs of evolutionary lineages are not at the same step in the allopatric speciation process.

In this study, we compare the genetic structure of codistributed freshwater fishes to investigate the role of glacial cycles in allopatric speciation and the diversification process. The main objective was to test the general hypothesis that geographic isolation of glacial lineages during the Pleistocene has triggered the onset of reproductive isolation across several freshwater fish taxa. First, we documented phylogeographic patterns in seven taxa using mitochondrial DNA and tested whether each of these is comprised of Atlantic and Mississippian glacial lineages. These were selected as possibly representing different levels of divergence. More precisely, the choice was based on the taxonomic status of fish (to include putative glacial lineages recognized as species, subspecies and without taxonomic rank) as well as on the distribution range (occurring from the Mississippi watershed to the Atlantic coast watershed). Thus, we analyse one species, *Pimephales promelas*, which may comprise two morphologically similar glacial lineages from the Mississippian and Atlantic refugia (Hocutt & Wiley 1986; Mandrak & Crossman 1992). We also include two other taxa that may be comprised of Mississippian and Atlantic glacial lineages that are recognized as distinct subspecies because they show significant, yet subtle, morphological variation (*Notemigonus crysoleucas auratus*/*N. c. crysoleucas* and *Esox americanus americanus*/*E. a. vermiculatus*; Hubbs & Lagler 2004). We further analysed two taxa that may comprise glacial lineages that are recognized as distinct species based on relatively pronounced morphological differentiation (*Rhinichthys atratulus*/*R. obtusus* and *Etheostoma nigrum*/*E. olmstedi*) (Chapleau & Pageau 1985; Hubbs & Lagler 2004; Fraser *et al*. 2005). We also included two species, *Notropis hudsonius* and *Pimephaes notatus* that are believed to have survived the Pleistocene glaciations in the Mississippi refugium only because they exhibit limited distribution near the Atlantic coast (Hocutt & Wiley 1986). Yet, they are currently geographically codistributed with the five other taxa with which they coexist from the Hudson River and eastern Saint Lawrence River to the Mississippi watershed. Therefore, *N. hudsonius* and *P. notatus* could also potentially have survived in both the Mississippian and Atlantic refugia, but distinct lineages within each taxon remained morphologically similar.

In the second part of the study, we analysed nuclear DNA variation using amplified fragment length polymorphism (AFLP) markers in four of these taxa for which the occurrence of Mississippian and Atlantic glacial lineages was confirmed by the mtDNA phylogeographic survey (*R. atratulus*/*R. obtusus*, *E. nigrum*/*E. olmstedi*, *N. c. crysoleucas*/*N. c. auratus* and *P. promelas*). We thus tested the hypothesis that glacial lineages identified using mitochondrial DNA are also divergent at the nuclear DNA level. We then tested whether distinct glacial lineages within each taxon remain

Fig. 1 Map of samples' localities (above) and mtDNA haplotype network (below) for each taxon. For taxa characterized by two distinct mtDNA clades, black and white circles were used to distinguish them. In the maps, sample sites where haplotypes from both mtDNA clades were found are filled in grey (see Table S1 for details). For taxa characterized by a single clade, sample sites are represented by empty circles in the maps. In the haplotype network, circles represent documented haplotypes and their size is proportional to the number of samples with the given haplotype. Lines connecting haplotypes correspond to a single mutational step. Small dots represent unobserved but inferred haplotypes.

reproductively isolated following secondary contact and documented for each their pattern of introgressive hybridization in the putative suture zone located in the upper St. Lawrence River drainage.

Material and methods

Biological material

The sampling design follows a transect ranging from the Mississippi R. watershed (in the vicinity of the former Mississippian refugium) to the Atlantic coast watershed (near the former Atlantic coast refugium), passing through the Great Lakes and the St. Lawrence River watershed (Fig. 1). For seven taxa (or pairs of sister taxa), we collected between 54 and 278 samples from 15 to 31 different locations (Table 1 and S1). Fish were caught using beach seines or backpack electrofishing gear and preserved in 95% ethanol.

All specimens (n = 1378) were first analysed using the mitochondrial 5′ cytochrome *c* oxidase subunit I (COI) gene region (Table 1 and S1). To standardize the mitochondrial phylogeography analyses across taxa, we first report results based on a maximum of ten individuals by sampling sites (randomly chosen as the first one sequenced; Fig. 1, Table S1) and results from the other samples are presented in the hybrid zone analyses (Table 1). For the hybrid zone analysis, which also included the analysis of AFLP markers, we used the results from mtDNA to choose six to twelve sampling sites near and within the putative secondary contact zones, as well as from the parental populations in the zones of allopatry. We primarily selected populations with many individuals (Table 1). Thus, *E. a. americanus*/*E. a. vermiculatus* was not analysed by AFLP because the number of samples per sampled location was too small to accurately document patterns of hybridization (Bonin *et al.* 2007).

Mitochondrial DNA analyses

We extracted DNA, then amplified and sequenced a 652-bp fragment of COI spanning the region typically used for animal DNA barcoding (Hebert *et al.* 2003), following the protocol described by April *et al.* (2011). DNA sequences were aligned using Clustal X (Thompson *et al.* 1997). The relationship between haplotypes was established by constructing a mutational network using Arlequin (Excoffier *et al.* 2005). Networks resolved for each fish were then compared with maximum-likelihood (ML) trees using Mega 5 (Tamura *et al.* 2011). The ML trees were produced using mutational models selected by BAIC (Nei & Kumar 2000), and one thousand pseudoreplications (bootstraps) were performed. We calculated the mean number of base differences per site between the genetic clusters (p distance) inferred by the mutational networks using Mega 5 (Tamura *et al.* 2011). Standard error estimates were obtained by a bootstrap procedure (1000 replicates).

We then used the software msBayes (Hickerson *et al.* 2007) to test for simultaneous divergence across codistributed pairs of taxa (Hickerson *et al.* 2007). More specifically, this software uses coalescence simulations to characterize the congruence in divergence/colonization times across the codistributed taxon pairs (Hickerson *et al.* 2007). The hyper-posteriors were obtained from 1000 accepted draws from 500 000 simulated draws from the hyper-priori. The acceptance/rejection was performed using a local regression. To compare the support of competing models (numbers of diversification waves = 1, 2, 3, 4 or 5), we used Bayes factor (Kass & Raftery 1995). To verify whether divergence events most likely occurred during the Pleistocene, we estimated time since divergence using the following formula: $T_{div} = \tau \times (\theta_{AVE} \div \mu)$, where τ is the divergence time expressed in global coalescent units, θ_{AVE} is a constant determined by the mean of the subprior for θ, and μ is the per generation mutation rate. We first used the classic mitochondrial DNA mutation rate value of 2% per million years (Brown *et al.* 1979) that has been validated in North American freshwater fishes (Centrarchidae: Near *et al.* 2003; Percidae: Near & Benard 2004), marine fishes (Bermingham *et al.* 1997; Bowen *et al.* 2001), as well as in other vertebrates (Pacheco *et al.* 2011). We also report divergence times computed using a wide range of possible mitochondrial DNA mutation rates (1–4% per million years) to account for the uncertainty in our estimation.

Amplified fragment length polymorphism analysis

The AFLP markers were obtained following the procedure of Vos *et al.* (1995). We used chemicals provided in the AFLP-based Genotyping Kits and Reagents (Applied Biosystems). We first digested genomic DNA obtained by salt extraction (Aljanabi and Martinez 1997) using the restriction enzyme *Eco*RI and *Mse*I. We then used four different primer pairs for each taxon (Table S2) in a selective PCR. Amplified fragments were run in an ABI 3010 xl sequencer (Applied Biosystems) and visualized using Genemapper (Applied Biosystems). Peak scoring was performed by a single individual (Anne-Marie Dion-Côté, M.Sc. Biology) and followed the objective procedure suggested by Whitlock *et al.* (2008) using the software AFLPSCORE (Whitlock *et al.* 2008). Ten per cent of the samples, randomly selected from nearly all populations, were run twice from the DNA extraction step to estimate the error rate.

Table 1 Basic data and assignment results based on amplified fragment length polymorphisms (AFLPs) for the secondary contact analyses. Assignments are based on the program NewHybrids and He refers to gene diversity. Locations of site are provided on Fig. 1 and Table S1

Sites	n	mtDNA sequences (proportion from clade W)	AFLP								
			Polymorphic loci (%)	He	SE (He)	Pure W	BC-W	F1	F2	BC-E	Pure E
Rhinichthys obtusus (west) and *Rhinichthys atratulus* (east)											
7	27	27/27	50.0	0.153	0.014	27					
10	33	33/33	53.1	0.185	0.015	29	4				
14	30	30/30	40.7	0.125	0.012	30					
15	42	42/42	60.5	0.201	0.015	42					
16	49	0/49	43.2	0.162	0.014						49
18	30	0/30	36.4	0.141	0.014						30
20	30	0/30	29.6	0.118	0.013					2	28
Total	241		162 loci								
Etheostoma nigrum (west) and *Etheostoma olmstedi* (east)											
6	25	25/25	44.8	0.161	0.014	24			1		
8	30	30/30	50.6	0.175	0.015	29	1				
10	31	31/31	39.0	0.146	0.014	30			1		
12	19	16/19	50.6	0.158	0.013	16	3				
13	29	29/29	47.7	0.161	0.013	29					
14	31	26/31	43.6	0.158	0.014						31
23	24	2/24	54.7	0.180	0.013						24
25	29	0/29	48.8	0.165	0.013						29
26	30	1/30	48.8	0.171	0.014						30
31	30	0/30	47.1	0.188	0.015						30
Total	278		172 loci								
N. c. auratus and N. c. crysoleucas											
3	7	7/7	58.5	0.204	0.014	7					
4	12	12/12	53.4	0.201	0.013	12					
5	11	11/11	51.8	0.200	0.013	11					
6	26	26/26	45.1	0.159	0.013	26					
9	29	20/28	53.4	0.177	0.013	28					
10	14	11/14	47.7	0.189	0.014	14					
11	32	28/32	52.3	0.167	0.013	32					
15	37	9/35	54.9	0.206	0.014	3			25	8	1
20	28	0/28	50.3	0.179	0.014				2	1	25
21	21	0/21	53.4	0.186	0.014						21
23	9	0/9	58.0	0.209	0.014				1	1	7
Total	226		193 loci								
Pimephales promelas											
3	35	34/34	52.6	0.172	0.009	35					
4	17	17/17	52.9	0.167	0.009	17					
8	34	5/34	57.4	0.187	0.009				14	20	
10	8	2/8	53.1	0.185	0.010				8		
12	37	0/37	40.9	0.146	0.009				1		36
16	34	0/34	46.6	0.144	0.009				1	13	20
18	35	0/35	23.1	0.103	0.009						35
Total	200		350 loci								

We first used the software AFLP-surv version 1.0 (Vekemans *et al.* 2002) to quantify, for each population, the proportion of polymorphic loci at the 5% level and gene diversity (He) assuming Hardy–Weinberg equilibrium. This software was also used to calculate F_{ST} values following the method of Lynch & Milligan (1994). The STRUCTURE software (Falush *et al.* 2007) was then used to infer the number of genetically different groups within a taxon or within pairs of taxa following the method of Evanno *et al.* (2005). Without any a priori information about localities, we thus performed independent runs using a number of inferred populations (K) varying from one and up to the total number of sampling sites. Those simulations were run for 20 000

steps after 10 000 initial burn-in steps and with 10 iterations. Then, a longer run was performed using the most likely *K* value (100 000 steps and 50 000 burn-in steps). This allowed assigning individuals to genetically different groups and inferring their individual admixture proportion (*Q* value). We then used the program NewHybrids (Anderson & Thompson 2002) to classify all analysed individuals in the six following genotype categories: pure individual from group 1 (Pure-W), pure individual from group 2 (Pure-E), F1, F2, backcross with group 1 background (BC-W) and backcross with group 2 background (BC-E). This was performed using Jeffrey type priors and by running 50 000 steps after 5000 steps of burn-in. To evaluate the power of our sets of AFLP loci to classify individuals in different hybrid categories, we used AFLP data from parental individuals to simulate and then assign 1000 individuals from each of the 5 genotype categories according to Duchesne & Bernatchez (2002). We then used the frequencies of hybrids and pure individuals to evaluate whether hybrid zones corresponded to unimodal or bimodal hybrid zones (Jiggins & Mallet 2000).

Cline analyses

Geographic cline analyses were performed using the software Cfit (Gay *et al.* 2008). This software allows modelling the geographic variation of allele frequency by a three-stepped cline, consisting of a central sigmoidal curve with two exponential tails (Szymura & Barton 1986; Gay *et al.* 2008). By performing a likelihood search, it further allows estimating the centre as well as the slope of clines at individual loci. Within a particular hybrid zone, different slopes or centres at individual loci would indicate that they are differentially affected by selection and or genomic incompatibilities (Barton & Hewitt 1985; Payseur 2010). We thus compared four different models where (i) the slope and the centre are unconstrained, (ii) the centre is constrained, (iii) the slope is constrained, and (iv) both the slope and the centre are constrained. Model choices were then made by comparing AICc values (Akaike 1974). These analyses were conducted using the 5% of loci with the most pronounced differentiation in allele frequencies between putative pure glacial lineages (including mtDNA haplotypes) to standardize the analyses among taxa (e.g. Gagnaire *et al.* 2011).

Using the cline width values obtained from Cfit (Gay *et al.* 2008), we estimated the strength of selection against hybrids. This estimation is based on the tension model of Barton & Hewitt (1985), whereby hybrid zones are shaped by the balance between the effect of migration and selection. The model is defined by: $s^* = 8(\sigma/w)^2$, where s^* is the difference in mean fitness between populations at the centre and the edge, w is the cline width, and σ^2 is the standard deviation of the distance between parents and offspring. In the absence of direct estimate of the dispersal parameters σ for the analysed species, we used the same value of 1 km/gen$^{1/2}$ for all taxa. This assumption is supported by mark recapture studies conducted on four small and codistributed North American freshwater fish species that found only little dispersal differences between species (Skalski & Gilliam 2000). Similar dispersal values have also been obtained in other studies of small freshwater fishes (Mundahl & Ingersoll 1983; Chenuil *et al.* 2000).

Results

Mitochondrial DNA

For each of the seven taxa analysed, we obtained between 625 and 652 bp of COI sequences, which revealed from 10 to 21 different haplotypes per taxa (Fig. 1). Mean nucleotide diversity varied from 0.001 to 0.020. Two divergent clades, separated by 12–34 mutation steps, were clearly defined by the mutational network in five of the seven taxa. Bootstrap values from the ML tree also highly supported (>99%) the genetic groups identified with the mutational network (trees not shown). The mean divergence between clades varied from 1.2% to 6.1% (*R. obtusus*/*R. atratulus* = 3.8% (SE = 0.7), *E. nigrum*/*E. olmstedi* = 6.1% (SE = 0.9), *E. a. vermiculatus*/*E. a. americanus* = 2.8% (SE = 0.6), *N. c. auratus*/*N. c. crysoleucas* = 1.2% (SE = 0.4), *P. promelas* = 2.1% (SE = 0.5)). In the two other taxa, *P. notatus* and *N. hudsonius,* no distinct clade was apparent and no haplotype diverged from others by more than five mutation steps (Fig. 1). The two most divergent haplotypes (4 and 5 mutation steps), recovered in *N. hudsonius*, were each observed in a single individual.

The geographic distribution of clades was structured on an east/west axis (Fig. 1). For *R. obtusus*/*R. atratulus* and *E. a. vermiculatus*/*E. a. americanus*, there was an abrupt transition from the western clade to the eastern clade without overlap in distribution or co-occurrence of both clades within any sampling site. The transition zones between those clades were located in the St. Lawrence River watershed near the limit of Quebec and Ontario. For *E. nigrum*/*E. olmstedi*, *N. c. auratus*/*N. c. crysoleucas* and *P. promelas*, the eastern and western clades distribution overlapped in the St. Lawrence River and the eastern Great Lakes region. For those taxa, there was at least one sample site, in which haplotypes from the two clades were recovered. For *P. notatus* and *N. hudsonius*, there was no apparent geographic pattern in the distribution of haplotypes.

The five taxa with pairs of clades matching the expected distribution of Atlantic and Mississippian glacial lineages did not diverge simultaneously. This is first indicated by the highly variable level of mtDNA sequence divergence (from 1.2% to 6.1%). This is also corroborated with the approximate Bayesian computation that did not support the hypothesis of a single common event of divergence (Bayes factor = 0.43), while there was substantial support for the hypothesis involving from two to four waves of divergence (Bayes factor = 4.62). The best support was obtained for the model involving three distinct events of diversification (Bayes factor = 3.35). Thus, a second analysis where the number of diversification events was constrained to three suggested that a single taxon diverged in the older diversification wave (*E. nigrum*/*E. olmstedi*), two in the second one (*R. obtusus*/*R. atratulus* and *E. a. vermiculatus*/*E. a. americanus*) and two in the most recent event (*N. c. auratus*/*N. c. crysoleuca*s and *P. promelas*). Based on the tau values of the constrained analyses ($\tau_1 = 0.4863$, $\tau_2 = 0.2712$ and $\tau_3 = 0.0880$) and using a mutation rate estimate of 2% per million years (Brown *et al.* 1979; Near *et al.* 2003; Near & Benard 2004), those waves of diversification would have most likely occurred around 470 000, 262 000 and 85 000 years before present, respectively. Even using mutation rate values ranging from 1% to 4% per million years, we found that all three divergence events most likely occurred between 43 000 and 940 000 years ago, that is during the Pleistocene.

Amplified fragment length polymorphism analysis

A total of 162–350 polymorphic peaks (frequency between 1% and 99%) were unambiguously scored for 200–278 individuals of each of the four taxa for which the analysis was performed (Table 1). The per-locus error rate by species was low, with a mean of 3.4%. Globally, the proportion of polymorphic loci within sampling sites ranged from 23% to 61%, and the gene diversity within sites (assuming HWE) varied from 0.103 to 0.209 (Table 1).

Assignment tests using simulated hybrids first showed that assignment success rate of individuals as pure-W, pure-E or hybrid was almost perfect ($\geq$ 99.9%). Assignment success was more variable, although still generally high (mean = 89.2%), when classifying simulated individuals into different hybrid categories (Table S3). The lower assignment success systematically involved the F2 category. As we could not specifically test for the occurrence of post-F2 hybrids other than the two backcross categories, it is probably that some individuals classified as F2 may in fact correspond to post-F2 hybrids with nearly proportional amount of loci from both genetic groups.

Analyses of population structure conducted using STRUCTURE showed that all four taxa analysed were characterized by two geographically structured genetic groups, in full agreement with the mtDNA data. Following Evanno *et al.* (2005), we found that the likelihood of the data rose from $K = 1$ to $K = 2$ and then reached a plateau that later declined when using higher values of K. F_{ST} values between genetic groups varied from 0.434 in *R. obtusus*/*R. atratulus* to 0.169 in *N. c. auratus*/*N. c. crysoleucas* and were intermediate in *E. nigrum*/*E. olmstedi* (0.282) and *P. promelas* (0.301). Results obtained with STRUCTURE were also highly similar to those obtained with NewHybrids (Fig. 2, Table 1). For all four taxa, most individuals were assigned to one or the other genetic cluster with strong confidence. The majority of individuals from the eastern most sampling sites were assigned to one genetic group, while all, or nearly all, individuals from the western most sampling sites were assigned to the second genetic group. Depending on taxa, however, the number of individuals with mixed ancestry varied greatly (Fig. 2, Table 1).

The frequency of pure and hybrid individuals in secondary contact for the *R. obtusus*/*R. atratulus* and *E. nigrum*/*E. olmstedi* taxa pairs corresponded to bimodal hybrid zones (Fig. 2). We found a very restricted number of hybrids, while pure individuals predominated in all sites. For *N. c. auratus*/*N. c. crysoleucas* and *P. promelas*, we found at least one site where hybrids predominated (Fig. 2). For these taxa, combining all sampling sites where hybrids were recovered and that are not separated by any obvious geographic barrier, we found both hybrid and pure individuals. This suggests that hybrid zones for these taxa are trimodal and represent an intermediate stage between unimodal and bimodal hybrid zones (Jiggins & Mallet 2000; Gay *et al.* 2008).

Cline analyses

The geographic cline of *R. obtusus*/*R. atratulus*, when constrained to reflect all loci, was 46 km wide, and its centre was located in the St. Lawrence River about 200 km east of Lake Ontario. For *E. nigrum*/*E. olmstedi*, the cline was 112 km wide and its centre was at approximately the same location as that for *R. obtusus*/*R. atratulus*. For *N. c. auratus*/*N. c. crysoleucas*, the cline was wider (372 km) and its centre was located more to the west, approximately 70 km east of Lake Ontario. The cline observed in *P. promelas* was also wide (235 km), and its centre was located close to that of *N. c. auratus*/*N. c. crysoleucas*, about 50 km west of the St. Lawrence River outlet. The selection coefficient s* for each of the four hybrid zones, as calculated with

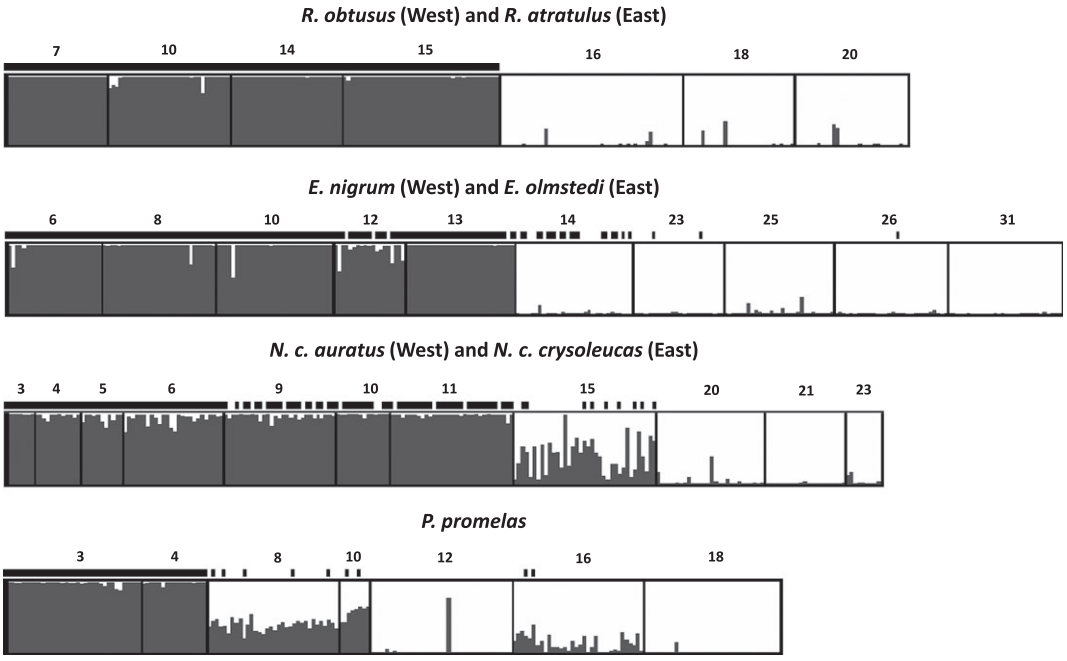


Fig. 2 Bar plot showing the admixture proportion of each specimen based of amplified fragment length polymorphism data and calculated using the program STRUCTURE. Sampling site numbers are given above bar plots. Sites analysed using mtDNA only are not include here, but rather in Fig. 1 and Table S1. Each individual corresponds to a vertical bar, and the proportion of the genome from each of the genetic groups is represented by its corresponding colour (black = west and white = east). Individuals possessing mtDNA haplotypes from the western genetic group are marked with a black line above bar blots.

the tension zone model (Barton & Hewitt 1985) using the constrained slope obtained from Cfit and modelling the overall genetic pattern, varied by a factor of 65, the strongest being between *R. obtusus/R. atratulus* and the weakest between *N. c. auratus/N. c. crysoleucas*. In decreasing order, the s* value was 3.78E−03 for *R. obtusus/R. atratulus*, 6.38E−04 for *E. nigrum/E. olmstedi*, 1.45E−04 for *P. promelas* and 5.78E−05 for *N. c. auratus/N. c. crysoleucas*.

For *R. obtusus/R. atratulus*, AICc values suggested that the best geographic cline model implied a constrained cline centre but varying slope among loci (Table S4). The slopes varied from −0.0101 (for AFLP locus K5.RH.X23) to −0.8846 (mtDNA), and the centre was located in the St. Lawrence River about 200 km east of Lake Ontario. For *E. nigrum/E. olmstedi*, *N. c. auratus/N. c. crysoleucas* and *P. promelas*, the best models implied different centre and slope for each of the allelic cline (Table S4). More precisely, the allelic cline slopes varied from −0.0060 (X61.ET.K4) to −0.0243 (mtDNA) in *E. nigrum/E. olmstedi* and the centres varied from 67 km west (X1.ET.K6) to 89 km east (mtDNA) of the constrained centre. For *N. c. auratus/N. c. crysoleucas*, the allelic cline slopes varied from −0.0017 (X98.NOCR.K2) to −0.0118 (X38.NOCR.K6) and the centres location varied from 524 km west (X87.NOCR.K10) to 289 km east (X38.NOCR.K6) of the constrained centre. For *P. promelas*, the slopes varied from −0.0167 (X23.PIPR.K6) to −0.0034 (X113.PIPR.K1) and the centres location varied from 399 km west (X62.PIPR.K10) to 271 km east (X23.PIPR.K6) of the constrained centre.

Discussion

Allopatric origin and distribution of glacial lineages

We identified pairs of distinct genetic groups showing a pronounced east–west difference in geographic distribution with limited overlap in five of the seven taxa investigated (Fig. 1). The observed levels of genetic divergence between these groups and the use of different mtDNA molecular clocks further indicated that isolation between the east and west lineages occurred during the Pleistocene in each case. Thus, these groups represent distinct glacial lineages that survived the Pleistocene glaciations in both the Mississippian and the Atlantic refugia and remained largely geographically isolated. However, while they were presumably all isolated in the two refugia during the Wisconsinan (~14 000–110 000 years before present), isolation for many species is older and was estimated to date back to the Illinoian (~130 000–200 000 years before present) and Kansan glaciations (~300 000–455 000 years before present).

The postglacial colonization of the Great Lakes basin probably started around 14 000 years before present when the Laurentide ice sheet began to melt (Dyke & Prest 1987). As suggested by other zoogeographic studies and geological data, fish coming from the Mississippian refugium may have entered the region of the actual Great Lakes by the Chicago outlet (region of the actual Lake Michigan) and the outlet of the glacial Lake Maumee (region of the actual Lake Erie) (Bailey & Smith 1981; Hocutt & Wiley 1986; Dyke & Prest 1987; Mandrak & Crossman 1992). Fish from the Atlantic coast refugium most likely used the Hudson River and glacial Lakes Albany and Vermont (Hocutt & Wiley 1986; Dyke & Prest 1987; Mandrak & Crossman 1992) to reach the actual Lake Champlain and St. Lawrence River.

For *P. notatus* and *N. hudsonius*, mtDNA data support the previous hypothesis based on distribution data (Bailey and Smith 1981, Mandrak & Crossman 1992) by indicating a single glacial origin in the study area. Indeed, we found a single mtDNA cluster, and the geographic distribution of the recovered haplotype group did not follow any apparent pattern that could suggest a period of allopatry. Thus, as suggested by the limited distribution of this species along the Atlantic Coast (Bailey and Smith 1981, Mandrak & Crossman 1992), those species probably used the Mississippian refugium and later dispersed eastward.

The cumulative evidence from this and previous studies clearly indicates that Lake Ontario and the St. Lawrence River represent an important suture zone for north-eastern temperate freshwater fishes. Indeed, we observed that distinct genetic lineages (defined in all five studied taxa) that apparently survived in both the Mississippian and Atlantic refugia form a secondary contact zones located in this region. In addition to the species analysed in this study, Lake Ontario and the upper St. Lawrence River region also correspond to the secondary contact zones of glacial lineages in nine other species (Hocutt & Wiley 1986; Mandrak & Crossman 1992; Todd and Hatcher 1993; Stepien *et al*. 2007; see Table 2 for species analysed using mtDNA). Together, these species represent eight different families, including Esocidae, Cyprinidae, Ictaluridae, Percidae, Fundulidae, Centrarchidae, Salmonidae and Catostomidae. In fact, from a total of 14 species possessing Atlantic and Mississippian lineages that have been the subject of phylogeographic or population structure study (using allozymes, microsatellites or mtDNA), all except the lake whitefish (13/14) possess a secondary contact zones in this region. It would therefore be relevant to check whether Lake Ontario and the St. Lawrence River represent a secondary contact zones for many other species of the 30 or so that are thought to comprise both Atlantic and Mississippian glacial lineages (Hocutt & Wiley 1986; Mandrak and Crossman 1992). As a whole, such generally high congruence in the geographic distribution of glacial lineages among North American codistributed fish species has only been described in the classical case of the south-eastern USA (Bermingham & Avise 1986).

We observed a tendency for taxa with deeper genetic divergence to show increased levels of morphological differentiation. Thus, the level of genetic divergence observed in *R. obtusus/R. atratulus* (3.8%) and *E. nigrum/E. olmstedi* (6.1%), the two glacial lineages that are considered as distinct sister species based on morphological differences, is higher than the level observed in all other fish that have used the Mississippian and Atlantic refugia and that have been analysed so far (Table 2). Genetic divergence was also relatively high in species comprising glacial lineages that are recognized as subspecies (mean = 2.0%), but low in species where no subspecific designation has been given between populations based on morphology and that are associated with either the Mississippian or the Atlantic glacial lineages (mean = 1.1%). Overall, we observed a significant tendency for genetic divergence to increase with taxonomic status (Spearman's rank correlation where species, subspecies and lineages without taxonomic rank are, respectively, given a value of 2, 1 and 0: P-value = 0.0033). Interestingly, it also seems that lineages from species with small adult body sizes tend to have significantly deeper genetic divergence than larger ones (Pearson correlation using maximum total length data from literature (Scott and Crossman 1998): P-value = 0.0283).

The glacial lineages that were first depicted by mtDNA analysis were also differentiated at the nuclear DNA level (AFLP). Indeed, the large-scale population genetic structure recovered with about 200 randomly distributed nuclear loci was highly congruent to that recovered with the mtDNA sequences. Although the number of replicates is admittedly too low for statistical analyses, there also seems to be a tendency for nucDNA divergence to covary with mtDNA divergence (Table 2). Therefore, while some doubt has been cast on the reliability of mtDNA sequences to recover 'real' evolutionary lineages, our results show that in many freshwater fishes at least, the analysis of mitochondrial DNA still represents a reliable first step towards describing the overall geographic distribution of young independently evolving lineages.

Reproductive isolation between glacial lineages

Based on genetic data collected in this study, two lines of evidence support the hypothesis that some of the glacial lineages have remained reproductively isolated

Table 2 Divergence between Mississippian and Atlantic glacial lineages from different freshwater fish taxa. Data from two other codistributed species (*P. notatus* and *N. hudsonius*) are also included. References are given when genetic data were obtained from the literature

Taxa	Morphological differences	Taxonomic designation	Divergence DNAmt (%)	Divergence DNAnuc (Nei's)	Selection against hybrids (s*)	Maximum total lenght
E. nigrum/*E. olmstedi*	Yes	Species	6.1	0.0929	6.38E−04	11
R. atratulus/*R. obtusus*	Yes	Species	3.8	0.1774	3.78E−03	10
E. americanus americanus/ *E. a. vermiculatus*	Yes	Subspecies	2.8			38
F. diaphaus diaphanus/ *F. d. menona* (April & Turgeon 2006)	Yes	Subspecies	2			13
N. crysoleucas crysoleucas/ *N. c. auratus*	Yes	Subspecies	1.2	0.0477	5.78E−05	30
A. nebulosus (Billington & Hebert 1988)	No	NA	3.2			50
P. promelas	No	NA	2.1	0.0856	1.45E−04	10
C. commersoni (Lafontaine & Dodson 1997)	No	NA	0.7			64
C. artedi (Turgeon & Bernatchez 2001)	No	NA	0.5			57
S. vitreum (Murdoch & Hebert 1997)	No	NA	0.5			80
C. clupeaformis (Bernatchez & Dodson 1991)	No	NA	0.4			91
S. namaycush (Wilson & Hebert 1998)	No	NA	0.4			126
P. notatus	No	NA	—	—	—	11
N. hudsonius	No	NA	—	—	—	15

in their zone of secondary contact. First, the distribution of genotypic classes suggests that the secondary contact zones of both *R. obtusus*/*R. atratulus* and *E. nigrum*/*E. olmstedi* correspond to bimodal hybrid zones, reflecting important reproductive barriers (Jiggins & Mallet 2000). For *N. c. auratus*/*N. c. crysoleucas* and *P. promelas*, the hybrid zones appear to correspond to an intermediate stage between unimodality and bimodality. This may indicate cases where some, albeit weaker, intrinsic reproductive isolation barriers have developed (Jiggins & Mallet 2000; Gay *et al.* 2008).

The second line of evidence suggesting that introgressive hybridization is highly restricted in some taxa comes from cline width analyses. According to the tension zone model (Barton & Hewitt 1985), cline width tends to be narrow when there is selection against hybrids and larger when hybridization occurs with limited constraints. Everything else being equal, selection against genetic admixture between *R. obtusus* and *R. atratulus* would be 65.4 times higher than that between *N. c. auratus* and *N. c. crysoleucas*. In contrast, selection against hybridization between *E. nigrum* and *E. olmstedi* and between *P. promelas* distinct clades would, respectively, be 11.0 and 2.5 times higher than that observed for *N. c. auratus*/*N. c. crysoleucas*.

Allelic cline analyses at individual loci show that hybrid zones correspond to semi-permeable barriers to gene exchange and could point to cases of cytonuclear incompatibility (e.g. Lane 2010). Indeed, we found in all taxa that the selection coefficients vary significantly among loci, suggesting that some parts of the genome may be more permeable to genetic exchange than others, perhaps due to genomic incompatibilities (Barton & Hewitt 1985). Among the four taxa analysed with AFLP data, the most important interlocus variation was observed in *R. obtusus*/*R. atratulus*, where the selection against hybrids acting on mtDNA was apparently 1656 times more pronounced than the mean selection strength acting on nucDNA loci. In *E. nigrum*/*E. olmstedi*, mtDNA was also the locus showing the highest selection coefficient against hybrids, being seven times higher than the mean selection acting on the other loci. Even if it was not the locus with the steepest allelic cline slope in *N. c. auratus*/*N. c. crysoleucas* and *P. promelas*, selection against hybrids acting on mtDNA was still seven and two times higher than the mean selection

over nucDNA markers. Admittedly, because mtDNA is haploid and only transmitted by females (Avise 2000), gene flow at equilibrium is expected to be fourfold higher for nucDNA compared with mtDNA. This could partially explain the pattern observed in some taxa. However, in *R. obtusus*/*R. atratulus* at least, the high selection strength acting on mtDNA strongly suggests that the mitochondrial genome cannot freely cross the barrier from one lineage to the other and could therefore be directly involved in the reproductive isolation between them. Indeed, mtDNA from *R. obtusus*/*R. atratulus* was observed in association with the alternate nuclear DNA background in only six backcross individuals. Even then, mtDNA was always associated with a genetic background mostly composed of its corresponding genetic group of origin (Fig. 2 and Table 1). Given that many nucDNA and mtDNA genes must co-evolve and interact to form enzymes involved in the oxidative phosphorylation pathway (Gershoni *et al.* 2009), and because mtDNA evolves rapidly (Brown *et al.* 1979; Gershoni *et al.* 2009), it is plausible that the genetic incompatibilities observed in *R. obtusus*/*R. atratulus* arose because of the interaction of foreign mtDNA and nucDNA genes. Cases of cytonuclear incompatibilities leading to hybrid breakdown have previously been reported in fish (Bolnick *et al.* 2008; Gagnaire *et al.* 2012) as well as in various other taxonomic groups, including insects (Innocenti *et al.* 2011), copepods (Ellison & Burton 2008) and yeast (Lee *et al.* 2008). Our results along with others (reviewed in Gershoni *et al.* 2009) thus point towards an important role for mtDNA in the development of genetic incompatibilities and speciation. In this respect, it is perhaps noteworthy that the mtDNA may represent a particularly suitable 'DNA barcode' to identify animal species because of its potential to be directly involved in the development of genetic incompatibilities (Lane 2010).

Our comparative study supports the idea that a threshold of 2% sequence divergence at mtDNA (e.g. DNA barcodes) can be used in North American freshwater fishes as a first step to flag potential cryptic species that may need deeper taxonomic investigation (Hebert *et al.* 2003; April *et al.* 2011). This idea was first based on the observation that most sister species pairs diverge by 2% or more (Avise *et al.* 1998; Hebert *et al.* 2003) and that such a level of divergence is probably indicative of approximately a million years of unique evolutionary legacy (Brown *et al.* 1979; Avise 2000). Here, by directly testing for reproductive isolation in a secondary contact zones, we show that taxa possessing DNA barcodes that diverge by over 2% from any other named species are probably engaged in the speciation process and may have already developed clear signs of reproductive isolation (Table 2).

Allopatric speciation and biodiversity conservation

By revealing that many co-occurring taxa are at different stages of allopatric divergence and speciation caused by the same vicariance processes, this study supports the hypothesis that the glacial cycles of the Pleistocene acted as an allopatric speciation pump in north-eastern American freshwater fishes. Our comparative phylogeographic analyses, as well as the zoogeography literature, show that at least 30 north temperate freshwater fish taxa may have been isolated in the Mississippian and Atlantic refugia during the last glacial cycle (Wisconsinan), and it appears most likely that isolation dates back to older glacial cycles of the Pleistocene for several of them. Some of the taxa characterized by glacial lineages with deep genetic divergences (at both the mtDNA and nucDNA levels) have probably been isolated for a long period of time and have also developed morphological differences. Following their isolation and subsequent secondary contact in the Great Lakes/upper St. Lawrence River watershed, at least two pairs of lineages with moderate levels of mtDNA and nucDNA divergence and little or no morphological differences showed signs of reproductive isolation. For two other species with relatively high levels of genetic differentiation and moderate morphological differences, we found signs of strong reproductive isolation preventing the mixing of glacial lineages in the secondary contact zones, most likely reflecting the fact that they represent a relatively complete stage of speciation.

The formation of evolutionary lineages and species following isolation in glacial refugia could thus be common. In fishes for instance, reproductive isolation between glacial lineages has been described in two other species also occurring in north-eastern North America but having survived in the Atlantic and Acadian refugia (Lecomte & Dodson 2004; Bernatchez *et al.* 2010). Allopatric speciation caused by glacial cycles may also frequently occur in many other aquatic taxa in North America. For crustaceans (Dooh *et al.* 2006), aquatic insects (Heilveil & Berlocher 2006) and bivalves (Elderkin *et al.* 2008) at least, phylogeographic studies conducted in North America have shown that they may have used the same glacial refugia and consequently display phylogeographic structures similar to those documented in fishes.

Our results also bring relevant information for conservation strategies, given the importance of preserving not only current biodiversity, but also the underlying evolutionary processes (Davis *et al.* 2008). First, our results highlight several evolutionary lineages that are morphologically cryptic and/or not completely reproductively isolated but may still merit designation

in conservation strategy. Second, we show that the Mississippian and Atlantic coast watersheds represent a geographic template that produced and is still producing new species. Finally, a future research avenue would be to use this comparative system to investigate the role of habitat and ecological differences and divergent natural selection in the development of reproductive isolation.

Acknowledgements

We wish to thank Caroline Durand, Geneviève Ouellet-Cauchon and Guillaume Côté for their valuable help during fieldwork and Scott Pavey for his comments on the manuscript. We also thank Nicholas Mandrak from University of Toronto, as well as Huguette Massé from Ministère des Ressources naturelles et de la Faune Québec, for their valuable help in collecting samples. We thank Heather Braid for assistance in generating DNA sequences. We are also grateful to Julie Turgeon, Julian Dodson, Nicolas Derome and Pierre-Alexandre Gagnaire for constructive discussions and three anonymous referees for their constructive comments. We are grateful to Paul Hebert for his support and interest throughout this study. This research was supported through funding to the Canadian Barcode of Life Network from the Natural Sciences and Engineering Research Council of Canada and other sponsors (listed at http://www.BOLNET.ca), as well as the Canadian research chair in genomics and conservation of aquatic resources. This study represents a contribution to the research program of Québec-Océan.

Conflict of interest

Authors declare no conflict of interest.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Aljanabi SM, Martinez I (1997) Universal and rapid salt extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692–4693.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.

April J, Turgeon J (2006) Phylogeography of the banded killifish (Fundulus diaphanus): glacial races and secondary contact. *Journal of Fish Biology*, **69**(Suppl B), 212–228.

April J, Mayden RL, Hanner RH, Bernatchez L (2011) Genetic calibration of species diversity among North America's freshwater fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10602–10607.

Avise JC (2000) *Phylogeography, the History and Formation of Species*. Harvard University Press, Cambridge, MA, pp. 447.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of Royal Society B*, **265**, 1707–1712.

Bailey RM, Smith GR (1981) Origin and geography of the fish fauna of the Laurentian Great Lakes basin. *Canadian Journal of Fish and Aquatic Science*, **38**, 1539–1561.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Bermingham E, Avise JC (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics*, **113**, 939–965.

Bermingham E, McCafferty SS, Martin AP (1997) Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: *Molecular Systematics of Fishes* (eds Kocher TD & Stepien CA), pp. 113–128. Academic Press, New York.

Bernatchez L, Dodson JJ (1991) Phylogeographic structure in mitochondrial DNA of the Lake Whitefish *Coregonus clupeaformis*) in North America and its relationships to Pleistocene glaciations. *Evolution*, **45**, 1016–1035.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of nearctic and palearctic freshwater fishes. *Molecular Ecology*, **7**, 431–452.

Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of whitefish. *Philosophical Transactions of the Royal Society of London*, **365**, 1783–1800.

Billington N, Hebert PDN (1988) Mitochondrial DNA variation in Great Lakes Walleye (Stizostedion vitreum) populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **45**, 643–654.

Bolnick DI, Turelli M, Lopez-Fernandez H *et al.* (2008) Accelerated mitochondrial evolution and ''Darwin's corollary'': asymmetric viability of reciprocal F1 hybrids in Centrarchid fishes. *Genetics*, **178**, 1037–1048.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR (2001) Phylogeography of the trumpetfish (Aulostomus spp.): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 1967–1971.

Chapleau F, Pageau G (1985) Morphological differentiation of *Etheostoma olmstedi* and *E. Nigrum* (Pisces: Percidae) in Canada. *Copeia*, **1985**, 855–865.

Chenuil A, Crespin L, Pouyaud L, Berribi P (2000) Movements of adult fish in a hybrid zone revealed by microsatellite genetic analysis and capture-recapture data. *Freshwater Biology*, **43**, 121–131.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA, USA.

Dasmahapatra KK, Lamas G, Simpson F, Mallet J (2010) The anatomy of a 'suture zone' in Amazonian butterflies: a coalescent-based test for vicariant geographic divergence and speciation. *Molecular Ecology*, **19**, 4283–4301.

Davis EB, Koo MS, Conroy C, Patton JL, Moritz C (2008) The California Hotspots Project: identifying regions of rapid diversification of mammals. *Molecular Ecology*, **17**, 120–138.

Dooh RT, Adamowicz SJ, Hebert PDN (2006) Comparative phylogeography of two North American 'glacial relict' crustaceans. *Molecular Ecology*, **15**, 4459–4475.

Duchesne P, Bernatchez L (2002) AFLPOP: a computer program for simulated and real population allocation based on AFLP data. *Molecular Ecology Notes*, **3**, 380–383.

Dyke AS, Prest VK (1987) Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Geographie Physique et Quaternaire*, **41**, 237–263.

Edmands S (2002) Does parental divergence predict reproductive compatibility? *Trends in Ecology and Evolution*, **17**, 520–527.

Elderkin CL, Christian AD, Metcalfe-Smith JL, Berg DJ (2008) Population genetics and phylogeography of freshwater mussels in North America, *Elliptio dilatata* and *Actinonaias ligamentina* (Bivalvia: Unionidae). *Molecular Ecology*, **17**, 2149–2163.

Ellison CK, Burton RS (2008) Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution*, **62**, 631–638.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Fraser BA, Mandrak NE, McLaughlin RL (2005) Lack of morphological differentiation in eastern (*Rhinichthys atratulus*) and western (*Rhinichthys obtusus*) blacknose dace in Canada. *Canadian Journal of Zoology*, **83**, 1502–1509.

Gagnaire PA, Minegishi Y, Zenboudji S, Valade P, Aoyama J, Berrebi P (2011) Within-population structure highlighted by differential introgression across semipermeable barriers to gene flow in *Anguilla marmorata*. *Evolution*, **65**, 3413–3427.

Gagnaire PA, Normandeau E, Bernatchez L (2012) Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American eels. *Molecular Biology and Evolution*, **29**, 2909–2919.

Gay L, Crochet PA, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, **62**, 2789–2806.

Gershoni M, Templeton AR, Mishmar D (2009) Mitochondrial bioenergetics as a major motive force of speciation. *BioEssays*, **31**, 642–650.

Haffer J (1969) Speciation in Amazonian forest birds. *Science*, **165**, 131–137.

Harrison RG (1993) *Hybrid Zones and the Evolutionary Process*, Oxford University Press, New York, USA.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of Royal Society B*, **270**, 313–321.

Heilveil JS, Berlocher SH (2006) Phylogeography and postglacial range expansion in Nigronia serricornisSay (Megaloptera: Corydalidae). *Molecular Ecology*, **15**, 1627–1641.

Hickerson MJ, Stahl E, Takebayashi N (2007) msBayes: pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *BMC Bioinformatics*, **26**, 268.

Hocutt CH, Wiley EO (1986) *The Zoogeography of the North American Freshwater Fishes*, pp. 880. Wiley, New York, USA.

Hubbs CL, Lagler KF (2004) *Fishes of the Great Lakes region*, pp. 213. University of Michigan Press, Ann Arbor, Michigan, USA.

Innocenti P, Morrow EH, Dowling DK (2011) Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science*, **332**, 845–848.

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, **15**, 250–255.

Johnson NK, Cicero C (2004) New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American Birds. *Evolution*, **58**, 1122–1130.

Jordan DS (1908) The law of germinate species. *American Naturalist*, **42**, 73–80.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of American Statistical Association*, **90**, 773–795.

Klicka JZ, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. *Science*, **277**, 1666–1669.

Knapp S, Mallet J (2003) Refuting refugia? *Science*, **300**, 71–72.

Lafontaine P, Dodson JJ (1997) Intraspecific genetic structure of white sucker (*Catostomus commersoni*) in northeastern North America as revealed by mitochondrial DNA polymorphism. *Canadian Journal of Fisheries and Aquatic Science*, **54**, 555–565.

Lane N (2010) Biodiversity: on the origin of bar codes. *Nature*, **462**, 272–274.

Lecomte F, Dodson JJ (2004) Role of early life-history constraints and resource polymorphism in the ecological segregation of sympatric populations of an estuarine fish. *Evolutionary Ecology Research*, **6**, 631–658.

Lee HY, Chou JY, Cheong L *et al.* (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell*, **135**, 1065–1073.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Mandrak NE, Crossman EJ (1992) Postglacial dispersal of freshwater fishes into Ontario. *Canadian Journal of Zoology*, **70**, 2247–2259.

Moritz C, Hoskin CJ, MacKenzie JB *et al.* (2009) Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proceedings of Royal Society B*, **276**, 1235–1244.

Mundahl ND, Ingersoll CG (1983) Early autumn movements and densities of Johnny (*Etheostoma nigrum*) and fantail (*E. flabellare*) darters in a southwestern Ohio stream. *Ohio Journal of Science*, **83**, 103–108.

Murdoch MH, Hebert PDN (1997) Mitochondrial DNA evidence of distinct glacial refugia for brown bullhead (Ameiurus nebulosus) in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 1450–1460.

Near TJ, Benard MF (2004) Rapid allopatric speciation in logperch darters. *Evolution*, **58**, 2798–2808.

Near TJ, Kassler TW, Koppelman JB, Dillman CB, Phillip DP (2003) Speciation in North American black basses, Micropterus (Actinopterygii: Centrarchidae). *Evolution*, **57**, 1610–1621.

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Pacheco MA, Battistuzzi FU, Lentino M, Aguilar R, Kumar S, Escalante AA (2011) Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Molecular Biology and Evolution*, **28**, 1927–1942.

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.

Remington CL (1968) Suture-zones of hybrid interaction between recently joined biotas. In: *Evolutionary Biology* (eds Dobzhansky T, Hecht MK & Steere WC), pp. 321–428. Plenum Press, New York, USA.

Scott WB, Crossman EJ (1998) *Freshwater Fishes of Canada*. Galt House Publications, Oakville, Canada.

Skalski GT, Gilliam JF (2000) Modeling diffusive spread in a heterogeneous population: a movement study with stream fish. *Ecology*, **81**, 1685–1700.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Stepien CA, Murphy D, Strange RM (2007) Broad- to fine-scale genetic patterning in smallmouth bass *Micropterus dolomieu* across the Laurentian Great Lakes and beyond: an interplay of behavior and geography. *Molecular Ecology*, **16**, 1605–1624.

Swenson NS, Howard DJ (2004) Do suture zones exist? *Evolution*, **58**, 2391–2397.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution*, **40**, 1141–1159.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Todd TN, Hatcher CO (1993) Genetic variability and glacial origins of yellow perch (*Perca flavescens*) in North America. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 1828–1834.

Turgeon J, Bernatchez L (2001) Mitochondrial DNA phylogeography of lake Cisco (*Coregonus artedi*): evidence supporting extensive secondary contacts between two glacial races. *Molecular Ecology*, **10**, 987–1001.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Vos P, Hagers R, Bleeker M *et al.* (1995) AFLP: new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Research*, **8**, 725–735.

Wilson CC, Hebert PDN (1998) Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1010–1024.

J.A. and L.B. designed the research. J.A., R.H.H. and L.B. performed the research. J.A. and A.-M.D.-C. performed the laboratory work. J.A. analysed the data and wrote the article. L.B. and R.H.H. revised and commented the article.

Data accessibility

DNA sequences: GenBank accessions (JX516792–JX517194, JX516102–JX516786)

Individual AFLP and GenBank accession numbers, and sampling location data: Dryad accession (doi:10.5061/dryad.km246).

Sampling sites (and sizes) used for phylogeographic analyses uploaded as online supplemental material (Table S1).

Primer combinations used in PCR of AFLP uploaded as online supplemental material (Table S2).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Information on sampling sites (and sizes) used for phylogeographic analyses. For mtDNA, numbers in brackets represent the number of haplotypes from the western clade and cases where haplotypes from two clades have been recovered are in gray. Sampling sites analysed using AFLPs are marked with an asterisk (*).

Table S2 Primer combinations used in the selective PCR of the amplified fragment length polymorphism analyses. *Mse*I primer: 5′-GATGAGTCCTGAGTAANNN-3′. *Eco*RI primer: 5′-GACTGCGTACCAATTCNNN-3′.

Table S3 Assignment success of simulated hybrids into different hybrid categories.

Table S4 Likelihood and AIC values for geographic cline models. Parameters were obtained from the program Cfit (Gay *et al.* 2008). The best models, based on AICc scores, are highlighted in bold.