doi:10.1111/evo.12075

THE GENETIC ARCHITECTURE OF REPRODUCTIVE ISOLATION DURING SPECIATION-WITH-GENE-FLOW IN LAKE WHITEFISH SPECIES PAIRS ASSESSED BY RAD SEQUENCING

Pierre-Alexandre Gagnaire,[1,2,3] Scott A. Pavey,[1] Eric Normandeau,[1] and Louis Bernatchez[1]

[1] *Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Biologie, Université Laval, Pavillon Charles-Eugène-Marchand, Québec, Canada G1V 0A6*

[2] *Institut des Sciences de l'Evolution – Montpellier (ISEM), Université Montpellier II, Place Eugène Bataillon 34095, Montpellier Cedex, France*

[3] *E-mail: pierre-alexandre.gagnaire.1@ulaval.ca*

Received October 21, 2012
Accepted January 31, 2013
Data Archived: Dryad doi:10.5061/dryad.f7125

During speciation-with-gene-flow, effective migration varies across the genome as a function of several factors, including proximity of selected loci, recombination rate, strength of selection, and number of selected loci. Genome scans may provide better empirical understanding of the genome-wide patterns of genetic differentiation, especially if the variance due to the previously mentioned factors is partitioned. In North American lake whitefish (*Coregonus clupeaformis*), glacial lineages that diverged in allopatry about 60,000 years ago and came into contact 12,000 years ago have independently evolved in several lakes into two sympatric species pairs (a *normal* benthic and a *dwarf* limnetic). Variable degrees of reproductive isolation between species pairs across lakes offer a continuum of genetic and phenotypic divergence associated with adaptation to distinct ecological niches. To disentangle the complex array of genetically based barriers that locally reduce the effective migration rate between whitefish species pairs, we compared genome-wide patterns of divergence across five lakes distributed along this divergence continuum. Using restriction site associated DNA (RAD) sequencing, we combined genetic mapping and population genetics approaches to identify genomic regions resistant to introgression and derive empirical measures of the barrier strength as a function of recombination distance. We found that the size of the genomic islands of differentiation was influenced by the joint effects of linkage disequilibrium maintained by selection on many loci, the strength of ecological niche divergence, as well as demographic characteristics unique to each lake. Partial parallelism in divergent genomic regions likely reflected the combined effects of polygenic adaptation from standing variation and independent changes in the genetic architecture of postzygotic isolation. This study illustrates how integrating genetic mapping and population genomics of multiple sympatric species pairs provide a window on the speciation-with-gene-flow mechanism.

KEY WORDS: **Barrier strength, demography, effective migration rate, RAD-Seq, speciation-with-gene-flow, whitefish.**

© 2013 The Author(s). *Evolution* © 2013 The Society for the Study of Evolution.

Elucidating the nature of the evolutionary processes leading to progressive accumulation of reproductive isolation barriers is central to our understanding of speciation, especially when populations are connected by gene flow (Endler 1977; Rice and Hostert 1993). The so-called "speciation-with-gene-flow" mechanism encompasses a variety of speciation modes ranging from primary divergence in sympatry to the evolution of additional reproductive barriers following secondary contact between taxa that have partially diverged in allopatry (Smadja and Butlin 2011). The central issue related to speciation-with-gene-flow is to understand how selection generates and maintains associations between alleles against the disruptive effects of recombination (Felsenstein 1981), and how these associations contribute to reduce the effective migration rate on a local or global genome scale (Barton and Bengtsson 1986). Theoretical studies have shown that the evolution of linkage disequilibrium (LD) may arise through two different processes. Selection can strengthen prezygotic isolation through reinforcement to reduce the production of unfit hybrids (Servedio and Noor 2003; Bolnick and Fitzpatrick 2007), and selection can promote associations (i.e., coupling) between independent incompatibilities to increase the mean fitness by increasing the variance in compatibility, which can strengthen both pre- and postzygotic isolation (Barton and de Cara 2009). As LD builds up between selected loci, the efficiency of selection increases because each locus indirectly cumulates the selection coefficients of other loci in addition to its own selective effects (Barton 1983; Kruuk et al. 1999). As a result, the barrier to gene flow generated by each selected locus is stronger than what would happen if selection was acting with the same strength on each locus alone without LD (Barton and Bengtsson 1986). The reduction of effective gene flow around selected loci may then facilitate secondary divergence which eventually strengthens the barrier to gene flow and completes speciation (Feder et al. 2012b; Via 2012).

Currently, there is debate about the type of genome-wide patterns of divergence that emerge during the course of speciation-with-gene-flow (Feder et al. 2012a; Nosil and Feder 2012). Recent developments in genotyping-by-sequencing technologies have considerably enhanced the power of genome scan studies to provide genome-wide maps of genetic divergence between closely related taxa (Hohenlohe et al. 2010; Deagle et al. 2011; Ellegren et al. 2012; Jones et al. 2012a; Nadeau et al. 2012a; Roesti et al. 2012a; Stölting et al. 2012). These studies have found extensive variation in the level of genetic differentiation across the genome, with some regions showing significantly enhanced levels of divergence among populations compared to the genome-wide average. Highly differentiated regions likely reflect local reductions in the effective migration rate (m_e) relative to the gross migration rate (m), and are therefore of particular interest to determine the genetic architecture of population divergence. This includes the number of genomic regions under divergent selection, their distribution across the genome, their effect size, and association with ecologically important phenotypic differences revealed via quantitative trait loci (QTL), association, or admixture mapping approaches.

The chromosomal extent of effective gene flow reduction around selected loci remains a critical question in speciation genomics. Small "islands of genetic differentiation" extending to at most a few hundred kilobases have been described in some empirical examples of speciation-with-gene-flow (Turner et al. 2005; Wood et al. 2008; White et al. 2010; Ellegren et al. 2012; Nadeau et al. 2012a, b; Smadja et al. 2012), whereas in other models, the decay of genetic differentiation away from the selected sites was found to occur within regions spanning several megabases (Harr 2006; Rogers and Bernatchez 2007; Via and West 2008; Hohenlohe et al. 2010; Hohenlohe et al. 2012; Renaut et al. 2012; Via 2012). Variation in island size among studies may reflect not only methodological and demographic characteristics unique to each system (Nadeau et al. 2012a), but also structural aspects of species divergence. Indeed, patterns of genomic divergence can be affected by both fine and large-scale variations in recombination rate, with extended islands of divergence commonly occurring in areas of reduced recombination including inversions, centromeric regions, or chromosome centers (Feder et al. 2003; Navarro and Barton 2003; Hoffmann and Rieseberg 2008; Noor and Bennett 2009; Roesti et al. 2012a). In addition, the wide reported range of island size could be influenced by the degree of reproductive isolation, because empirical studies have relied on a variety of taxa presumably at different stages of the speciation-with-gene-flow mechanism (Feder et al. 2012a). To evaluate this latter hypothesis, genome-wide patterns of genetic differentiation must be compared among multiple population pairs showing variable degrees of reproductive isolation within a single species complex. However, few natural systems offer the conditions required for such comparison (but see Nadeau et al. 2012b; Roesti et al. 2012a).

Lake whitefish species pairs (*Coregonus clupeaformis*) provide such a valuable natural model to compare genome-wide patterns of divergence between replicate sympatric pairs of limnetic *dwarf* and benthic *normal* populations occurring in North American postglacial lakes. In the St. John River drainage (Maine, USA and Québec, Canada), there is a continuum of morphological, ecological, and genetic differentiation spanning from widespread introgression to near complete reproductive isolation across lakes that differ in their degree of niche divergence and potential for competitive interactions (Campbell and Bernatchez 2004; Rogers and Bernatchez 2007; Renaut et al. 2011; Renaut et al. 2012). *Dwarf* whitefish exclusively occur in sympatry with *normal* whitefish and therefore probably evolved through ecological character displacement in response to divergent selection on standing genetic variation in the postglacial *normal* population (Landry et al. 2007; Bernatchez et al. 2010). However, both

intrinsic and extrinsic barriers presently contribute to reproductive isolation between *dwarf* and *normal* whitefish, reflecting the combined effects of historical, demographic, and ecological factors during divergence (Rogers and Bernatchez 2006; Renaut et al. 2009; Landry and Bernatchez 2010). Integrating genetic mapping and population genomic approaches revealed that the genomic islands of divergence were preferentially associated with the regions underlying quantitative phenotypic traits, and that the size and number of islands increased with increasing morphological and genetic divergence (Rogers and Bernatchez 2007; Renaut et al. 2012). However, these measures could be interpreted as preliminary due to the low mapping resolution offered by the amplified fragment length polymorphism (AFLP) map. Recently, the construction of a high-density restriction site associated DNA (RAD) linkage map comprising 3438 markers allowed the detection of phenotypic, expression, and transmission ratio distortion (TRD) QTL with a 20-fold enhanced resolution (Gagnaire et al. 2013).

Here, we used single-nucleotide polymorphisms (SNPs) from this new linkage map to document the genetic architecture of reproductive isolation in five replicate pairs of *dwarf* and *normal* whitefish populations. We determined the number, size, and distribution of genomic islands of differentiation in each lake, and measured the extent of LD between them. Patterns of haplotype structure were used to evaluate possible reasons for incomplete parallelism in the divergence. Mean genetic differentiation was compared between QTL and the remainder of the genome to assess the role of selection on locally adaptive traits and genes causing hybrid dysfunction. Finally, the barrier strength was derived from measures of genetic differentiation to separate the confounding effects of selection, recombination, and demography on the extent of gene flow reduction around selected loci.

Materials and Methods

RAD GENOTYPING-BY-SEQUENCING

We compared genome-wide patterns of genetic differentiation between sympatric *dwarf* and *normal* whitefish species pairs across the five lakes from the St. John River basin: Témiscouata Lake, East Lake, Webster Lake, Indian Pond, and Cliff Lake, where these forms still exist. These lakes lie along a continuum of ecological niche divergence in parallel with increasing morphological divergence (Landry et al. 2007; Landry and Bernatchez 2010), increasing genetic differentiation and reproductive isolation between species pairs (Renaut et al. 2011; Renaut et al. 2012).

A total of 20 *normal* and 20 *dwarf* whitefish were sampled from each lake during June 2010 (Evans et al. 2012). Whole genomic DNA was isolated from frozen liver tissue using the DNeasy Tissue Kit (Qiagen, Valencia, CA), and subsequently digested with *Sbf*I-HF (New England Biolabs, Beverly, MA), the same restriction enzyme already used for building the lake whitefish RAD linkage map (Gagnaire et al. 2013). RAD library construction was performed by Floragenex, Inc. (Eugene, OR) according to the protocol described in (Baird et al. 2008). Individual RAD libraries labeled with unique 6 bp barcodes were pooled in equimolar proportions (13 individuals per lane) and sequenced on an Illumina HiSeq2000 (Genomic Core Facility, University of Oregon) that generates single reads of 101 bp. All sequence reads of natural population samples were deposed in the NCBI Short Read Archive SRA066094.

We used the *Stacks* software system (Catchen et al. 2011) to identify polymorphisms and call individual genotypes. We first removed low-quality reads as well as reads presenting ambiguous barcodes from the raw sequencing data of natural population samples. All remaining reads were trimmed to 80 bp by removing the last 21 bases such that data were fully compatible with those generated previously to build the RAD whitefish linkage map (Gagnaire et al. 2013). Sequence reads from the two *hybrid* and *dwarf* parents used for genetic map construction were then included in the dataset, and natural population samples were declared as being their progeny. For each individual, all perfect alignments of at least three exactly matching reads were identified and used to infer putative alleles that were subsequently merged into loci using a maximum within-individual distance of 3% divergence between alleles (i.e., two mismatches) to avoid paralog problems. The different sets of loci assembled from natural population samples were then matched against the catalog of loci obtained from the mapping parents to retain loci already detected in the mapping family. For each variable nucleotide position in each locus, the most likely genotype was inferred for each individual using a likelihood ratio test that takes into account possible sequencing errors (Hohenlohe et al. 2010). RAD loci including two SNPs were scored as haplotypes. Individual genotypes were only called for the 3438 RAD markers included in the *hybrid* × *dwarf* linkage map (comprising 40 linkage groups for an average resolution of 0.89 cM), although many more RAD loci could be identified in natural populations (data not shown). Despite the recent tetraploidization of the salmonid genome that occurred approximately 60 million years ago (MYA) (Crête-Lafrenière et al. 2012), this genotype matrix was expected to be free of false paralogous SNPs, because all mapped markers were already checked for Mendelian segregation using the backcross mapping family (Gagnaire et al. 2013).

DETECTION OF GENOMIC ISLANDS OF DIFFERENTIATION

The evolutionary history of sympatric *dwarf* and *normal* whitefish includes a period of allopatric divergence in two separate glacial refugia (Acadian and Atlantic-Mississippian) during approximately 60,000 years (Jacobsen et al. 2012) followed by a postglacial sympatric episode starting less than 12,000 years ago

(Bernatchez et al. 2010). The current genetic architecture of reproductive isolation results from the accumulation of incompatibilities during geographical isolation and subsequent genetic introgression combined with ecological divergence in sympatry (Lu and Bernatchez 1998; Rogers and Bernatchez 2006; Renaut et al. 2009). This complex evolutionary history challenges the use of a realistic model for simulating the null distribution of genetic differentiation under neutrality, and appropriate sequence data are still lacking to estimate model parameters. Thus, we detected regions exhibiting significantly elevated genetic differentiation relative to the genome-wide average using an empirical approach instead of using model-based simulations, which is a suitable approach in such contexts (Akey et al. 2010; Kolaczkowski et al. 2011; Cheng et al. 2012).

We used *Arlequin v3.5* (Excoffier and Lisher 2010) to calculate F_{ST} values (Weir and Cockerham 1984) between *dwarf* and *normal* whitefish in each lake for each RAD marker. Genome-wide distributions of F_{ST} along the whitefish linkage map were generated using a moving window Gaussian kernel smoothing technique (Hohenlohe et al. 2010). At each window position, we calculated the weighted average differentiation index ($\overline{F_{ST}}$), weighting F_{ST} values by the exponential of the squared distance from the window center divided by the squared window length. The window length was set to 1 cM, which is close to the 0.89 cM average resolution of the whitefish linkage map. The statistical significance of F_{ST} at each window position (i.e., every centi-Morgan) was tested by randomly resampling across the genome 10,000 times the number of loci included in the corresponding window. This procedure was meant to generate a distribution under the null hypothesis that $\overline{F_{ST}}$ equals the genome-wide average F_{ST}, as expected in the absence of divergent selection. Window positions exhibiting significantly elevated genetic differentiation relative to the genome-wide average were called "outlier genomic positions". To identify larger regions exhibiting enhanced genetic differentiation (i.e., >1cM), chromosomal intervals containing consecutive outlier genomic positions separated by less than 2 cM were merged into "outlier genomic regions". Genomic islands of differentiation were defined as regions of the genome where $\overline{F_{ST}}$ exceeded the neutral background value at the 5% genome-wide significance level, and therefore included isolated outlier genomic positions and outlier genomic regions separated from each other by at least 2 cM.

MEASURING ISLAND SIZE AND BARRIER STRENGTH

We analyzed the decay of F_{ST} around each island of differentiation using the most highly differentiated SNP as a central reference position, from which all neighboring SNPs toward the ends of their linkage group were assigned to discrete distance classes depending on their proximity to the reference position. When a given linkage group comprised more than one island of differentiation, the decay of F_{ST} was analyzed around each genomic island separately. For each lake, we finally combined the data from all islands and performed quantile regression analyses to examine the average relationship between F_{ST} and genetic (recombination) distance from the center of the genomic islands of differentiation. Because RAD markers represent random samples of existing genetic variation, many of the SNPs that are located in divergent regions are likely to correspond to mutations that have occurred after divergence and consequently do not exhibit the historical signatures of drift and selection (Roesti et al. 2012b). To avoid the influence of such polymorphisms, we used the 90th percentile value of F_{ST} as a measure of the F_{ST} value locally observed in the chromosomal neighborhood of the genomic islands of differentiation.

The F_{ST} value at equilibrium in a two-island model with migration (Crow and Aoki 1984) can be written as:

$$F_{ST} = \frac{1}{16N_e m_e + 1},$$

where m_e is the effective migration rate between populations, which varies among loci as a complex function of the gross (i.e., genome-wide) migration rate (m), the genetic (recombination) distance from a selected locus (r), and a selection coefficient (S) (Barton and Bengtsson 1986; Charlesworth et al. 1997). Because selection against hybrids acts simultaneously on several loci involved in reproductive isolation, the selection coefficient for a given locus (S) cumulates the indirect effects of selection on other loci in addition to its own.

To approximate the complex relationship between F_{ST} and r, we considered the linear regression model of the form:

$$\frac{1}{F_{ST}} \cong K \times r + \frac{1}{F_{ST_0}},$$

where K is a constant term depending on N_e, m, and S (i.e., the island shape parameter is $1/K$), and $1/F_{ST_0}$ is the value of $1/F_{ST}$ observed at the central reference position (i.e., the island height parameter is F_{ST_0}). This regression model was fitted in each lake to the 90th percentile value of F_{ST} as a function of r, forcing the intercept through $1/F_{ST_0}$. For comparative purposes among lakes, the mean size of the genomic islands of divergence was empirically determined as the recombination distance where the regression line of the 90th percentile intercepts the genome-wide 90th percentile of F_{ST}.

We described the degree of local gene flow reduction relative to the gross migration rate by estimating the strength of the genetic barrier between *dwarf* and *normal* whitefish at variable recombination distances from the center of genomic islands. The barrier strength (b, Barton and Bengtsson 1986) was calculated

using:

$$b = \frac{m}{m_e} = \frac{(1 - \mu_{GW}) \times F_{ST}}{(1 - F_{ST}) \times \mu_{GW}},$$

where μ_{GW} is the genome-wide average F_{ST} value. The regression of b against r was simply derived from the regression of the 90th percentile value of F_{ST} as a function of r.

DETECTING REGIONS OF PARALLEL GENETIC DIFFERENTIATION

To detect genomic regions exhibiting peaks of extremely high F_{ST} occurring at the same position in multiple lakes, we defined a "weighted mean multiple correlation index" (*WMR*) as follows:

$$WMR = \frac{1}{n} \times \sum_{i=1}^{n} R_i(\overline{F_{ST_1}} - \mu_{GW_i}),$$

where the coefficient of multiple determination R_i measures the degree to which the F_{ST} profile in lake i can be predicted using the F_{ST} profiles of all the other lakes within a given chromosome interval. For each lake, R_i was weighted by the local departure of $\overline{F_{ST}}$ from its genome-wide average, and weighted values were subsequently averaged across lakes ($n = 5$). The *WMR* statistic was calculated every centi-Morgan in a 6 cM window. Highly positive values of *WMR* exceeding the 95th percentile were used to detect regions of parallel genetic differentiation.

GENOMIC PATTERNS OF HAPLOTYPE STRUCTURE

Differences in haplotype structure between *dwarf* and *normal* whitefish were investigated in each lake at each genomic position using the cross-population extended haplotype homozygosity (XP-EHH) test (Sabeti et al. 2007). Only loci with a global minor allele frequency above 0.05 (2296 RAD markers) and individuals genotyped at more than 75% of the loci (CD: 19; CN: 17; ID: 20; IN: 20; WD: 16; WD: 20; ED: 20; EN: 20; TD: 19; TD: 18) were considered in this analysis. Individual genotypes were first statistically phased using the program FastPHASE (Scheet and Stephens 2006), with 25 random starts of the EM algorithm, each characterized by 50 iterations and 5000 haplotypes sampled from the posterior distribution. All of the 189 individuals retained were processed simultaneously. For each linkage group, the best number of clusters (locally representing common haplotypes) was searched between two and 10 using the cross-validation procedure, and individual haplotypes were inferred with the attempt to minimize the switch error.

The extended haplotype homozygosity (*EHH*, Sabeti et al. 2002), which measures the degree of haplotype identity at a given distance from a core SNP position for a given allele, was then estimated for each *dwarf* and *normal* whitefish population at each single SNP. We integrated *EHH* with respect to genetic distance for both alleles and calculated the log ratio of their integrated haplotype homozygosity (*iHH*) to obtain the integrated haplotype score (*iHS*, Voight et al. 2006). Standardized *iHS* values were then obtained by normalizing $|iHS|$ values to a mean of zero and a variance of one in 10 equally sized allele frequency classes. The *XP-EHH* score was finally calculated as the log ratio between *dwarf* and *normal* whitefish standardized *iHS* values. In each lake, we used the first and 99th percentiles of the *XP-EHH* distribution to identify outliers. Because the *XH-EHH* is directional, a positive outlying *XP-EHH* score is suggestive of positive selection in *dwarf* whitefish, whereas negative outlying values indicate positive selection in *normal* whitefish.

LONG DISTANCE LD

We used the most highly differentiated SNP of each genomic island to evaluate the extent of LD between selected loci within each lake. Nonrandom association between alleles at two loci was measured by both D and D', the latter corresponding to the ratio of D to its maximum possible absolute value given the allele frequencies in the sample. Pairwise measures of LD were calculated between all possible pairs of divergent SNPs, and the average LD was estimated as the mean of all pairwise LD values included in the triangular matrix of each lake.

GENETIC DIFFERENTIATION AT QTL COMPARED TO THE REMAINDER OF THE GENOME

Dwarf and *normal* whitefish differ in a number of traits that were fine mapped in a recent QTL mapping study (Gagnaire et al. 2013). Genomic positions of these QTL were used to compare genetic differentiation at quantitative loci to the baseline level of differentiation in the remainder of the genome. Three types of QTL were considered, including 31 QTL for ecologically important phenotypic traits (pQTL), 26 hotspots of expression QTL associated with differential gene expression of four or more genes (eQTL hotspots), and 17 TRD QTL. Several cases of overlap between pQTL, eQTL hotspots, and TRD QTL were observed on the RAD linkage map. Phenotypic QTL were distributed into four different categories, including behavioral (two pQTL for burst swimming, eight for directional change, six for depth selection, and three for activity), physiological (four pQTL for growth rate), morphological (three pQTL for the number of gill rakers), and life history (three pQTL for maturity and two for the gonadosomatic index). All phenotypes were obtained from Rogers and Bernatchez (2007), whereas gene expression data were from Derome et al. (2008), St-Cyr et al. (2008), and Whiteley et al. (2008).

For each type of QTL, we compared the mean genetic differentiation at markers associated with QTL (F_{STQ}) with the mean genetic differentiation at markers that are not (F_{ST}). Only strictly positive F_{ST} values were used for this analysis to exclude uninformative polymorphisms (Roesti et al. 2012bb), which did not change the relative differences among lakes (i.e., the

correlation between μ_{GW} and the mean genetic differentiation calculated using strictly positive F_{ST} values was highly significant: $R^2 = 0.999$, $P = 2.10^{-5}$). Four comparisons between F_{ST} and different measures of F_{STQ} were performed for each lake, with F_{STQ} being successively calculated from (1) the nearest RAD locus associated to each pQTL, (2) all markers included in the 1.5 LOD unit of support interval of each pQTL, (3) markers associated with eQTL hotspots, and (4) markers associated with TRD QTL. Differences between F_{STQ} and F_{ST} were tested using one-sided *t*-test for single lake comparisons or two-way analysis of variance (ANOVA) for multiple comparisons using lakes as replicates.

Results

GENOME-WIDE PATTERNS OF GENETIC DIFFERENTIATION

A total of 1.7×10^9 reads passing quality filters were retrieved after demultiplexing samples using individual barcode tags. We excluded five individuals with a low sequencing success, resulting in an average number of 8.7×10^6 reads per individual in 195 whitefish (Supplementary Table S1). All of the 3438 RAD markers included in the lake whitefish linkage map were identified and genotyped from natural population samples, yielding an overall genotype call rate of 85.1%. A total of 2734 mapped loci (79.5%) were polymorphic in at least one of the five lakes and were thus retained for downstream analyses. The mean observed heterozygosity within *dwarf* population samples was not significantly different than within *normal* whitefish across the five lakes (*dwarf* $H_o = 0.279$, *normal* $H_o = 0.259$, $P = 0.191$; Table 1).

The genome-wide average F_{ST} value between *dwarf* and *normal* whitefish (μ_{GW}) gradually increased from 0.008 in Témiscouata to 0.029 in East, 0.049 in Webster, 0.105 in Indian, and a maximum value of 0.216 in Cliff Lake. This baseline level of differentiation was positively correlated with the morphological differentiation index based on meristic and morphometric measures ($R^2 = 0.862$, $P = 0.023$) taken from Lu and Bernatchez (1999). Thus, the genome-wide level of gene flow was negatively related to the degree of phenotypic divergence between species pairs. The genome-wide variance in F_{ST} (σ_{GW}) also increased significantly with increasing baseline levels of differentiation ($R^2 = 0.991$, $P < 0.001$), and the right tail of the L-shaped frequency distribution of F_{ST} across loci progressively extended from 0.543 in Témiscouata to 1 in Indian Pond ($n = 2$ diagnostic loci) and Cliff Lake ($n = 16$ diagnostic loci; Table 1, Fig. 1).

The number of outlier genomic positions did not vary significantly among lakes ($\chi^2_{4df} = 7.008$, $P = 0.136$), as well as the number of outlier genomic regions ($\chi^2_{4df} = 0.738$, $P = 0.947$) and the number of genomic islands of differentiation ($\chi^2_{4df} = 3.014$,

Table 1. List of whitefish population samples for the five lakes analyzed in this study, including location, number of *dwarf* and *normal* individuals, number of polymorphic restriction site associated DNA (RAD) markers, and the morphological differentiation index based on meristic and morphometric measures taken from Lu and Bernatchez (1999). Population genomic statistics include the genome-wide average F_{ST} value (μ_{GW}) and variance (σ_{GW}), the 90th percentile and maximum values of F_{ST}, the number of outlier genomic positions and regions, the number of genomic islands of differentiation, and the number of islands falling within regions of parallel divergence. The three last columns provide estimated values for the island shape (1/*K*), height (F_{ST_0}), and size parameters.

Lake	Spatial coordinates	No. of individuals (*dwarf*/*normal*)	No. of polymorphic loci	Morphological differentiation index	μ_{GW}	σ_{GW}	90th F_{ST}	Maximum F_{ST}	No. of outlier positions	No. of outlier regions	No. of genomic islands	No. of islands within peaks of *WMR*	Island shape parameter (1/*K*)	Island height parameter (F_{ST_0})	Mean island size (cM)
Témiscouata	47°36′00″N 68°45′00″W	*20/18*	2595	0.796	0.008	0.054	0.077	0.543	114	35	57	5	1.473	0.338	16
East	47°11′00″N 69°33′00″W	*20/20*	2435	0.217	0.029	0.077	0.127	0.684	134	38	49	11	2.413	0.455	14.5
Webster	46°09′19″N 69°04′48″W	*20/20*	2355	1.435	0.049	0.107	0.189	0.890	129	32	48	15	6.272	0.709	25
Indian	46°15′25″N 69°18′05″W	*20/20*	2363	2.089	0.105	0.162	0.327	1	105	34	45	16	8.731	1	19
Cliff	46°23′51″N 69°15′05″W	*20/17*	2034	3.003	0.216	0.252	0.606	1	114	32	53	31	45.541	1	30.5

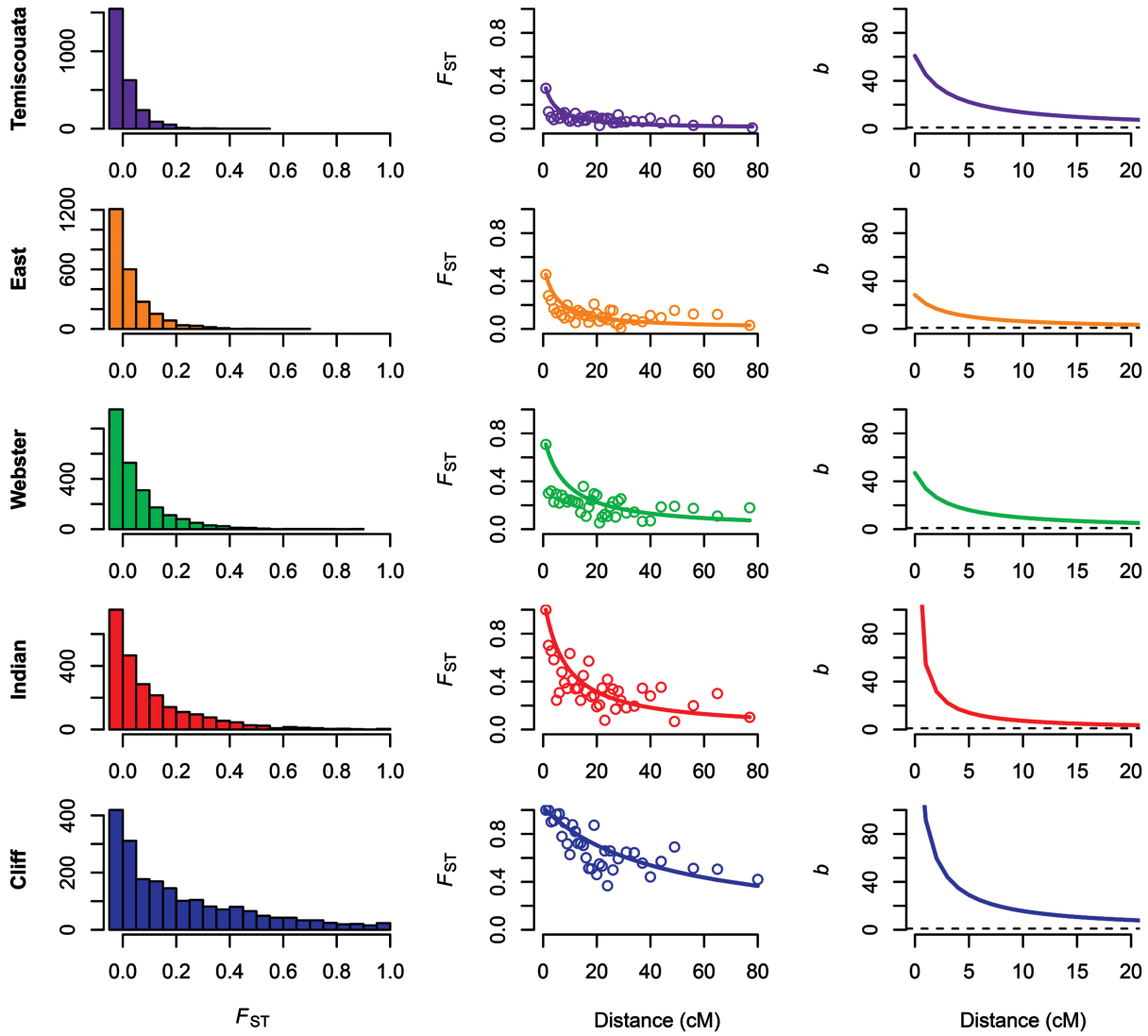

Figure 1. **(Left panels) Genome-wide distribution of F_{ST} values in the five lakes ordered by increasing magnitude of phenotypic divergence between *dwarf* and *normal* whitefish (from top to bottom). (Middle panels) Mean relationship between the 90th percentile F_{ST} value and the recombination distance from the center of the nearest genomic island on the chromosome. (Right panels) Mean relationship between the barrier strength parameter ($b = m/m_e$) and the recombination distance from the center of the nearest genomic island (note the change in scale for the recombination distance axis).**

$P = 0.555$). There was on average 50.4 genomic islands of differentiation in each lake (Table 1), which were randomly distributed across linkage groups (Supplementary Fig. S1; Témiscouata: $\chi^2_{39df} = 26.02$, East: $\chi^2_{39df} = 44.25$, Webster: $\chi^2_{39df} = 49.68$, Indian: $\chi^2_{39df} = 50.35$, Cliff: $\chi^2_{39df} = 46.31$, all $P > 0.05$). The positive correlation between the morphological differentiation index and the baseline level of differentiation was not influenced by the most divergent regions, and remained significant with the outliers removed ($R^2 = 0.870$, $P = 0.021$).

Genomic regions showing peaks of extremely high F_{ST} occurring at the same position in multiple lakes were found in 29 linkage groups out of 40 (Fig. 2). Every peak of *WMR* exceeding the 95th percentile of the genome-wide distribution was associated with at least one island of divergence in single lake genome scans. One region located on linkage group (LG) V was detected as outlier in all five lakes, four regions in three lakes (all being different combinations of lakes) and 17 in two lakes (of those seven between Indian and Cliff and six between Webster and Cliff). These partially overlapping divergence maps thus showed incomplete parallelism in genetic divergence across lakes. The number of genomic islands coinciding with significant peaks of parallel genetic differentiation gradually increased from Témiscouata ($n = 5$) to Cliff Lake ($n = 31$) and was significantly positively correlated with the baseline level of differentiation ($R^2 = 0.942$, $P = 0.006$) and the morphological differentiation index ($R^2 = 0.767$, $P = 0.052$).

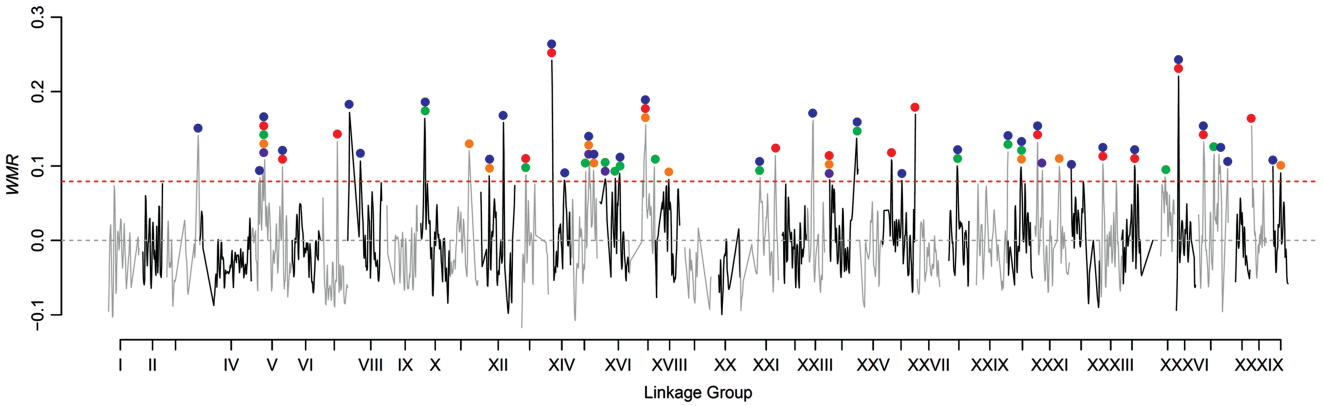

Figure 2. Genome-wide plot of the weighted mean multiple correlation index (*WMR*), showing regions of increased parallel differentiation across the 40 linkage groups represented on the x-axis according to their relative length. Peaks of *WMR* exceeding the 95th percentile of the genome-wide distribution (upper dotted line) were associated with islands of divergence detected in single lake genome scans, as indicated by colored dots (Témiscouata: purple; East: orange; Webster: green; Indian: red; Cliff: blue).

ISLAND SIZE AND BARRIER STRENGTH

All the quantile regressions of $1/F_{ST}$ against the recombination distance to the center of the genomic islands of differentiation were statistically significant (Fig. 1; Témiscouata: $R^2 = 0.582$, $P = 1.6\times10^{-8}$; East: $R^2 = 0.232$, $P = 1.9\times10^{-3}$; Webster: $R^2 = 0.572$, $P = 2.6\times10^{-8}$; Indian: $R^2 = 0.623$, $P = 2.4\times10^{-9}$; Cliff: $R^2 = 0.746$, $P = 5.8\times10^{-12}$). The island shape parameter ($1/K$) gradually increased from 1.473 in Témiscouata to 45.541 in Cliff Lake and was positively correlated with the baseline level of differentiation ($R^2 = 0.913$, $P = 0.011$). The mean island size, defined as the recombination distance where the quantile regression of $1/F_{ST}$ reaches the genome-wide value, was the smallest in the least phenotypically differentiated lake (East: 14.5 cM) and the largest in the lake exhibiting the strongest phenotypic divergence between *dwarf* and *normal* whitefish (Cliff: 30.5 cM). Also, the correlations between the phenotypic differentiation index and the island shape parameter ($R^2 = 0.722$, $P = 0.068$) or the mean island size ($R^2 = 0.723$, $P = 0.068$) were marginally significant.

To take into account that each lake may present different demographic characteristics in terms of population size and gross migration rate, the barrier strength (b) between *dwarf* and *normal* whitefish was derived from the quantile regressions of $1/F_{ST}$ at increasing recombination distances from the center of the genomic islands divergence (Fig. 1). As suggested by the measures of F_{ST}, the effective migration rate at the core of genomic islands was more strongly reduced in the two lakes where *dwarf* and *normal* whitefish are the most phenotypically divergent (Cliff and Indian). However, the barrier strength parameter decayed over a shorter recombination distance than did F_{ST} in all five lakes. Thus, neutral gene flow was strongly reduced mainly in the close chromosomal neighborhood of the loci under selection, that is, within the first 5 cM. In all lakes, the degree of local gene flow reduction relative to the gross migration rate was close to 1 beyond a 20 cM distance from the center of the islands of divergence.

PATTERNS OF LD

Genomic regions showing patterns of haplotype structure departing from neutral expectations between *dwarf* and *normal* whitefish were found across most linkage groups (Fig. 3; Supplementary Fig. S2). Some linkage groups exhibited overlapping or closely adjacent outlying values of *XP-EHH* that were detected in multiple lakes. These local excesses of haplotype structure often occurred near genomic regions showing parallel genetic differentiation among lakes (Fig. 3). Namely, highly positive peaks of *WMR* coinciding with outlying values of *XP-EHH* in all five lakes were detected on LGs XV, XXXI, and XXXVII. However, the directionality of haplotype structure was often inconsistent across lakes, with both significantly positive and negative *XP-EHH* scores being found in those genomic regions. Therefore, parallel genetic differentiation patterns were generally not associated with fully parallel haplotype structures.

We also found evidence for significant LD between genomic islands located on different linkage groups within each lake. Genomic islands were in nearly perfect LD in Lake Cliff, with an averaged D' value of 0.912 calculated over all the islands (Fig. 4). The mean D' value progressively decreased to 0.748 in Indian, 0.560 in Webster, 0.439 in East, and 0.414 in Témiscouata. A strong significant positive correlation was found between the averaged D' value and the extent of phenotypic differentiation ($R^2 = 0.937$, $P = 0.007$), as well as between the averaged D' value and the baseline level of differentiation ($R^2 = 0.946$, $P = 0.005$). These correlations remained significant using averaged $|D|$ values instead of D' (phenotypic differentiation: $R^2 = 0.893$, $P = 0.015$; baseline genetic differentiation: $R^2 = 0.986$, $P < 0.001$).

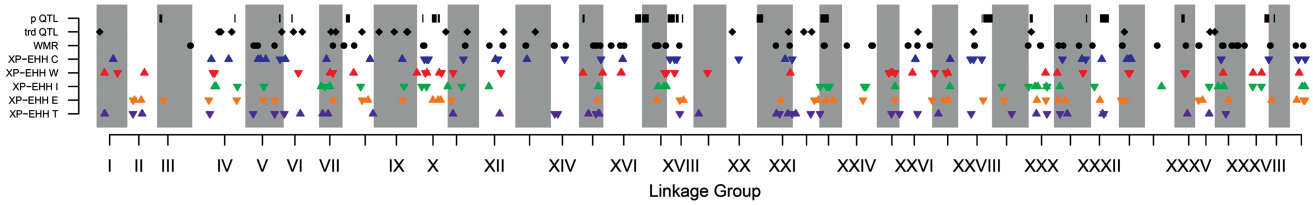

Figure 3. Genome map showing the locations of phenotypic quantitative trait loci (QTL) and transmission ratio distortion QTL detected in Gagnaire et al. (2013), regions of increased parallel differentiation revealed by *WMR*, and regions showing significant patterns of haplotype structure between *dwarf* and *normal* whitefish. Positive outlying *XP-EHH* scores, indicating positive selection in *dwarf* whitefish, are shown by upward triangles whereas negative outlying values indicating positive selection in *normal* whitefish appear as downward triangles.

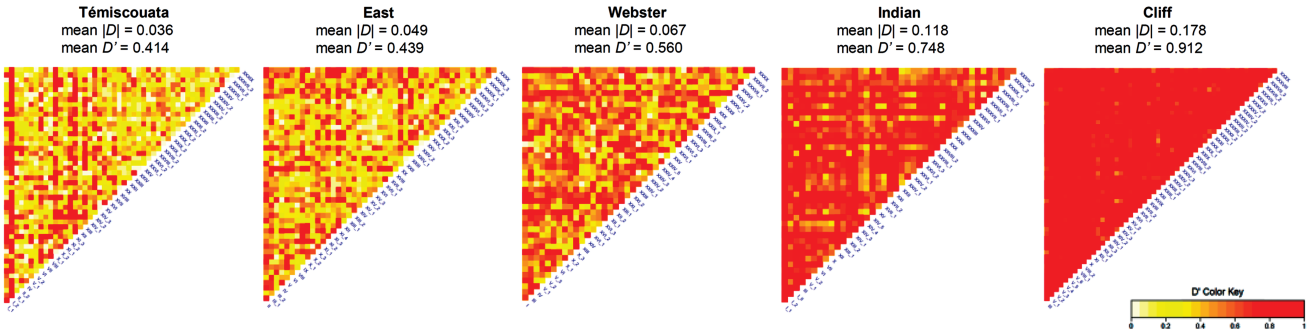

Figure 4. Heatmap of pairwise linkage disequilibrium measured by *D′* in each lake between all possible pairs of the most divergent SNP found within each genomic island of differentiation. The mean value of |D| and D′ is given for each lake.

GENETIC DIFFERENTIATION AT QTL

The mean F_{STQ} calculated from the nearest RAD locus associated to each pQTL was higher than the mean F_{ST} in all lakes except Témiscouata (Fig. 5), but was only significantly higher in Webster. The comparison using all five lakes as replicates was only marginally significant (Table 2), however, the level of significance increased when excluding Témiscouata from the comparison ($P = 0.016$), and was even higher when the analysis was restricted to East and Webster ($P = 0.0001$). The mean F_{STQ} calculated from markers included in the 1.5 LOD unit of support interval of each pQTL remained above the mean F_{ST} from non-pQTL regions in East, Webster, and Cliff, whereas still being significant in Webster. Moreover, the comparison restricted to East and Webster was again highly significant ($P = 0.0008$). In contrast, the mean F_{STQ} calculated from markers associated with eQTL hotspots was always below the mean F_{ST}, and reached statistical significance in East and Webster lakes. However, none of the comparisons using replicates was significant. Finally, no significant difference was detected between the mean F_{STQ} calculated from markers associated with TRD QTL, although again, the mean F_{STQ} lied above the mean F_{ST} in East and Webster lakes. Overall, this analysis revealed that the main contrast between F_{STQ} and F_{ST} was found in the lakes showing intermediate degrees of phenotypic and genetic divergence.

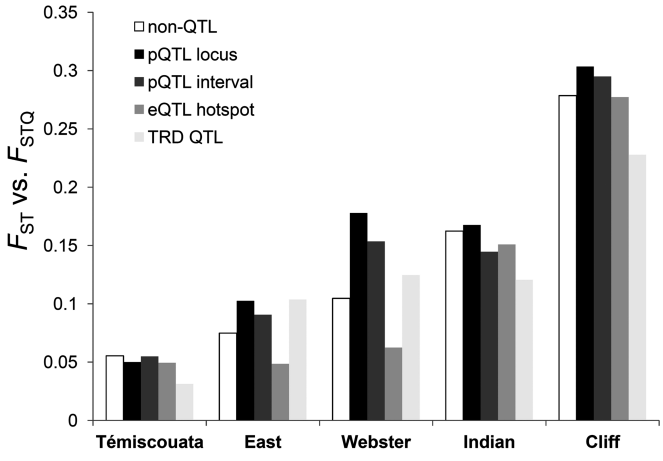

Figure 5. Comparison of the mean genetic differentiation between markers associated with quantitative trait loci (QTL; F_{STQ}) and those that are not associated with QTL (F_{ST}), as presented in Table 2.

Discussion

How the process of speciation progresses at the genomic level remains an unresolved question in evolutionary biology. Here, we investigated genome-wide patterns of divergence in five sympatric pairs of *dwarf* and *normal* whitefish populations spanning a continuum from weak phenotypic divergence and high global gene

Table 2. Genetic differentiation compared between markers associated with quantitative trait loci (QTL; F_{STQ}) and those not associated with QTL (F_{ST}). The mean F_{STQ} was either calculated from the nearest restriction site associated DNA (RAD) locus associated to each pQTL, all markers included in the 1.5 LOD unit of support interval of each pQTL, markers associated with eQTL hotspots, or markers associated with transmission ratio distortion (TRD) QTL. Only positive genetic differentiation values were used for this analysis. Differences between F_{STQ} and F_{ST} were tested independently for each lake as well as in multiple comparisons using five, four, or two lakes as replicates (last three columns).

		Témiscouata		East		Webster		Indian		Cliff		*T/E/W/I/C*		*E/W/I/C*		*E/W*	
		Mean *F*	*P* (*t*-test)	Mean *F*	*P* (*t*-test)	Mean *F*	*P* (*t*-test)	Mean *F*	*P* (*t*-test)	Mean *F*	*P* (*t*-test)	Mean *F*	*P* (ANOVA)	Mean *F*	*P* (ANOVA)	Mean *F*	*P* (ANOVA)
pQTL (nearest locus)	F_{STQ}	0.0501	0.6421	0.1025	0.1810	0.1779	**0.0493**	0.1676	0.4375	0.3034	0.3158	0.1780	0.0805	0.2190	**0.0157**	0.1780	**0.0001**
Non-pQTL	F_{ST}	0.0553		0.0749		0.1047		0.1624		0.2785		0.1460		0.1630		0.0904	
pQTL (1.5 LOD interval)	F_{STQ}	0.0550	0.5204	0.0907	0.1035	0.1535	**0.0042**	0.1446	0.8388	0.2950	0.2895	0.1610	0.1110	0.1790	0.1210	0.1220	**0.0008**
Non-pQTL	F_{ST}	0.0553		0.0744		0.1038		0.1632		0.2780		0.1460		0.1620		0.0898	
eQTL hotspot	F_{STQ}	0.0494	0.4258	0.0486	**0.0391**	0.0624	**0.0239**	0.1509	0.3496	0.2773	0.4907	0.1320	0.5200	0.1420	0.4150	0.0555	0.1400
Non-eQTL hotspot	F_{ST}	0.0553		0.0753		0.1060		0.1626		0.2788		0.1470		0.1630		0.0913	
TRD QTL	F_{STQ}	0.0313	0.9844	0.1036	0.1613	0.1247	0.2823	0.1206	0.8442	0.2279	0.7971	0.1280	0.4060	0.1460	0.5100	0.1130	0.2830
Non-TRD QTL	F_{ST}	0.0555		0.0749		0.1056		0.1627		0.2792		0.1470		0.1630		0.0909	

flow to highly pronounced phenotypic and genetic differentiation. We found that the overall level of genetic differentiation was positively correlated with that of phenotypic divergence across this continuum. However, the level of genetic differentiation between *dwarf* and *normal* whitefish was heterogeneous across the genome and highly divergent regions overlapped only partially among lakes. Thus, divergence patterns were not fully parallel at the genomic level. The number of genomic islands of divergence was not significantly different among lakes, but there was variation in island size reflecting the combined effects of selection, LD, and local demography. Increased genetic differentiation at phenotypic QTL relative to non-QTL markers was more pronounced in the lakes where effective migration was reduced locally (East and Webster) and was attenuated in the lakes where effective migration was also reduced globally (Cliff and Indian). The level of LD between genomic islands of divergence was strongly associated with the magnitude of phenotypic differentiation and the baseline genetic differentiation, raising the hypothesis that the buildup of associations between selected alleles has been important in the transition from divergence to genome hitchhiking in whitefish. Our results thus emphasize the importance of integrating linkage information with measures of differentiation in multiple population pairs to more fully understand genome-wide patterns of divergence during speciation-with-gene-flow.

THE POSSIBLE REASONS FOR INCOMPLETE PARALLELISM AT THE GENETIC LEVEL

Our genome scan for detecting enhanced levels of genetic differentiation occurring at the same position in multiple lakes confirmed the previous finding that fully parallel patterns of genetic differentiation between lakes are not frequent across the whitefish genome (Campbell and Bernatchez 2004; Rogers and Bernatchez 2007; Renaut et al. 2011), as also increasingly reported in other systems (Elmer and Meyer 2011). Nevertheless, it also revealed several genomic regions where parallelism was found between at least two lakes (i.e., incomplete parallelism), with only one case of complete parallelism on linkage group V (Fig. 2). This result contrasts with the much stronger parallelism observed at the morphological (Lu and Bernatchez 1999), physiological (Evans et al. 2012), and to a lower extent at the transcriptomic level (Derome et al. 2006; St-Cyr et al. 2008). In whitefish, parallel phenotypic divergence is thought to reflect differential adaptation to similar limnetic/benthic habitats across lakes, and consistent with this hypothesis, the extent of phenotypic divergence has been associated with the degree of ecological niche divergence and the potential for competitive interactions within lakes (Landry et al. 2007; Landry and Bernatchez 2010; Evans et al. 2012). Whether similar phenotypes observed in different lakes are achieved through the same allelic combinations depend on the genetic architecture of the phenotypic traits under divergent selection. In contrast,

with some famous phenotypic traits known to be controlled by a few genes of major effect (Nachman et al. 2003; Colosimo et al. 2004), most ecologically important phenotypic traits in lake whitefish (e.g., growth rate, depth selection, activity) are quantitatively distributed and have a polygenic basis involving multiple genes of moderate to small effect (Rogers and Bernatchez 2007; Gagnaire et al. 2013). Although the genetic basis of polygenic traits is more difficult to detect, it is likely to be more typical in nature than single gene traits (Rockman 2012). Adaptive divergence of polygenic traits usually occurs from standing variation and is therefore expected to produce partial divergence parallelism at the genomic level because similar phenotypes can be caused by different allelic combinations. Moreover, adaptive changes that occur through the evolution of multiple phenotypic traits controlled by a large number of loci usually involve small allele frequency changes that may not be detected in genome scans (Mackay et al. 2009; Pritchard and Di Rienzo 2010; Le Corre and Kremer 2012). This is consistent with the moderate increase in genetic differentiation at markers associated with pQTL (F_{STQ}) relative to non-QTL markers (F_{ST}). Also, the new QTL included in the high-density RAD linkage map may be of smaller effect than those already detected on the AFLP map, because the power to detect small effect QTL increases with increasing map resolution. This most likely explains why the average F_{STQ} was slightly lower than previously reported (Rogers and Bernatchez 2007; Renaut et al. 2012).

In addition to polygenic adaptation from standing variation, a second mechanism related with the evolutionary history of whitefish can produce incomplete divergence parallelism. When divergent populations (e.g., glacial lineages) come into secondary contact, historical genetic incompatibilities accumulated in allopatry may either scatter (i.e., barrier breakdown) or associate (i.e., barrier coupling) with each other (Barton and de Cara 2009). The efficiency of the coupling process depends on the strength of selection against incompatible genotypes, the symmetry of incompatibilities, as well as demographic effects related to genetic drift and migration rate. Only a subset of the genetic incompatibilities that are initially brought together may ultimately participate to the species barrier because some associations are lost and gene swamping occurs. Therefore, independent outcomes of the coupling process following secondary contact of the same two genetic backgrounds (Acadian and Atlantic Mississippian) in different lakes is expected to produce partially different genetic architectures of postzygotic isolation. This hypothesis clearly predicts that some barriers will be maintained in all lakes, whereas other barriers will be maintained or vanish depending on lakes. As secondary introgression subsequently erases historical genetic differentiation at neutral loci unlinked to any selected locus, the system eventually converges to a genome-wide divergence pattern that is similar to what is expected under a model of recent independent adaptation from standing genetic variation in the context of primary intergradation (Bierne et al. 2013).

Experimental evidence that both intrinsic and extrinsic barriers jointly contribute to postzygotic isolation between *dwarf* and *normal* whitefish (Rogers and Bernatchez 2006; Rogers and Bernatchez 2007; Renaut et al. 2009; Renaut and Bernatchez 2011) suggest that both the aforementioned mechanisms could be actually responsible for incomplete parallelism at the genetic level. Here, RAD mapping combined to population genomics approaches further revealed candidate regions harboring intrinsic incompatibilities around peaks of *WMR* associated with TRD QTL (Fig. 3: LG VII, XII, XXI, XXVI, and XXVIII), as well as candidate regions for loci involved in phenotypic adaptation around peaks of *WMR* associated with pQTL (Fig. 3: LG X, XVIII, and XXIII). However, signatures of selection on standing genetic variation were not clearly apparent from the analysis of haplotype structure. Repeated selection of the same standing variant at low frequency across independent lakes is expected to produce parallel signals of haplotype structure, which was not observed here in whitefish. Alternatively, endogenous incompatibilities may also display detectable signatures of haplotype structure because incompatible alleles often evolve under positive selection (Presgraves 2010). However, because the sign of association between independent incompatibilities can be either positive or negative (Barton and de Cara 2009), independent realizations of the coupling process may result in conflicting directionality in haplotype structure between *dwarf* and *normal* whitefish across lakes, as observed on LGs XV, XXXI, and XXXVII. Future studies will have to test whether—and if so, to what extent—divergence in regions showing signals of long-range haplotype in opposite directions predates secondary contact. This will require resequencing efforts within the islands of genomic divergence to compare alternative models of divergence history.

THE SIZE OF GENOMIC ISLANDS IS INFLUENCED BY BOTH SELECTION AND DEMOGRAPHY

Contrasting measures of island size among studies may reflect substantial variation in the strength of reproductive isolation across the range of species studied in speciation genomics (Feder et al. 2012a). However, it remains unclear what proportion of this variance can be attributed to demographic characteristics (in terms of N_e and m) unique to each species, the type of data that differed across studies, as well as to methodological approaches that have yet to be standardized.

In previous lake whitefish studies involving low resolution mapping and genome scan, large islands have been found around the outlier loci, some of which being associated to pQTL (Rogers and Bernatchez 2007; Renaut et al. 2012). The range of demographic and selective parameters found in whitefish (Lu and Bernatchez 1998; Campbell and Bernatchez 2004) were also

roughly consistent with those required for producing comparably large islands of differentiation at equilibrium in simulations, including small migration rates and population sizes ($m = 0.001$ and $N_e = 1000$), and strong divergent selection ($S = 0.5$; Feder and Nosil 2010; Via 2012). However, the two-fold variation in island size observed between the least and the most differentiated lakes could not be directly related to variation in selection intensity (Renaut et al. 2012). Indeed, island size is typically derived from empirical measures of F_{ST}, which integrate the complex interplay between the effects of population size, gross migration rate, recombination distance from the nearest locus under selection, strength of selection, number of selected loci, and the extent of LD among them (Feder and Nosil 2010).

Here, our approach relied on the assumption that the distribution of F_{ST} values at a given recombination distance from a selected locus reflects a local outcome of the effective migration-drift equilibrium, which was previously verified through simulations (Campbell and Bernatchez 2004). This model provided a strong fit to the observed data and confirmed the two-fold difference in island size across lakes, thus, establishing a link between island size and the extent of phenotypic differentiation. However, the contribution of selective effects was still confounded by demography. We thus used the barrier strength parameter ($b = m/m_0$) to estimate the degree of neutral gene flow reduction at increasing recombination distances from a selected locus, which reflects the combination of direct and indirect selective effects independently from demographic variables.

Island size measured by b was smaller than when assed with F_{ST}, although remaining relatively large (i.e., ≈5 cM). Differences between measures of F_{ST} and b can be explained by the nonlinear relationship between F_{ST} and N_em at equilibrium. Depending on the gross genome-wide N_em product, a small amount of reduction in the local effective migration rate can result in a more or less pronounced increase in F_{ST}. Thus, if the product N_em varies across lakes due to different population sizes and opportunities for gene flow, the observed correlation between the baseline level of differentiation (which approximates N_em) and the island shape parameter (which depends on N_e, m, and S) could be predicted for a fixed value of S. However, as m_e becomes globally reduced with multiple unlinked markers under divergent selection (i.e., genome hitchhiking, Feder et al. 2012), the baseline level of differentiation may not be a good approximation of the product N_em. Therefore, our estimates of the barrier strength at variable recombination distances from selected loci support that the barrier is much stronger in the most phenotypically differentiated lakes within the first 5 cM from targets of selection. These results highlight the importance of correcting the confounding effects of local demography before comparing the topography of the genomic islands of divergence among independent pairs of populations.

Admittedly, a higher density of markers would be required for a precise evaluation of the barrier strength parameter for each island separately. Nevertheless, combining measures of F_{ST} around different islands of divergence provides an average measure which remains informative for comparisons across lakes. Indeed, the variance in island size across the genome is supposed to be small if island size is principally determined by the transmission of selective effects through high-order LD compared to first-order LD. However, new mutations falling within already established islands may additionally contribute to the building of significantly larger islands (Rogers and Bernatchez 2007; Via and West 2008; Hohenlohe et al. 2010, 2012; Renaut et al. 2012; Via 2012). Ultimately, estimating the barrier strength parameter separately for each island will help to dissociate island-specific effects (direct selection) from the indirect selective effects transmitted through LD with other islands, which will probably require the type of data produced by whole genome resequencing projects (Ellegren et al. 2012; Jones et al. 2012b) or very dense genetic maps built using both a very large number of markers and hundreds if not thousands of progeny.

THE ROLES OF ECOLOGY IN WHITEFISH SPECIATION

Previous ecological studies in whitefish showed that the lakes exhibiting the highest degree of ecological niche divergence and potential for competitive interactions are associated with increased phenotypic divergence among *dwarf* and *normal* whitefish (Campbell and Bernatchez 2004; Rogers and Bernatchez 2007; Bernatchez et al. 2010; Renaut et al. 2011; Renaut et al. 2012). Here, we confirmed that the degree of phenotypic divergence was also positively related with the baseline level of genetic differentiation, thus, indicating that the observed gradient of genetic differentiation was associated with the gradient of ecological niche divergence. To understand how reproductive isolation builds up at the genomic level, we must now consider how the genetic architecture of reproductive isolation varies across this divergence continuum.

A general trend has previously been observed toward an increase in the number and size of genomic islands of divergence from the least to the most differentiated species pair (Renaut et al. 2012). Here, using a much higher density of SNP markers, we found no significant difference in the number of genomic islands among lakes. However, the number of islands falling within regions of parallel divergence was significantly positively correlated with the genome-wide level of differentiation, suggesting that the loci underlying differential adaptation to limnetic and benthic habitats are under stronger divergent selection in the lakes where ecological niche divergence is stronger, and/or that selection has maintained stronger associations between incompatibilities. Obviously, these two hypotheses are not independent because ecologically based selection facilitates both divergence

and the building of LD. Incomplete divergence parallelism and the identical number of genomic islands of divergence across lakes suggest that partially overlapping sets of incompatibilities were independently maintained in each lake following secondary contact between the Acadian and Atlantic-Mississippian genetic backgrounds. Thus, our finding that the level of LD between islands is stronger in the lakes with the highest ecological niche divergence strongly supports the role of divergent selection in building associations between incompatibilities. Theory predicts that for a given locus under selection, the cumulative effects of selection acting on other selected loci will be more efficiently transmitted when high-order LD is strong (Barton 1983; Kruuk et al. 1999). Thus, the extent of local gene flow reduction could depend on the degree of ecological divergence for two reasons: (1) increased selection intensity on local adaptation genes facilitates the building or high-order LD (i.e., coupling), (2) which in turn allows a more efficient transmission of indirect selective effects in the chromosomal neighborhood of every gene involved in the barrier. The consequence is an extended reduction in the effective migration rate compared to an identical genetic architecture of reproductive isolation without LD.

In conclusion, our results emphasize that the strength of associations between selected alleles from neighboring and distant genomic regions is an important parameter during speciation-with-gene-flow. We showed that the number of islands was roughly similar across lakes but that both first- and higher-order LD increased with the extent of phenotypic divergence between *dwarf* and *normal* whitefish. Considering that intrinsically incompatible alleles have similar fitness effects in each lake, ecologically based divergent selection on QTL underlying differential adaptation to limnetic/benthic habitats may be the main factor explaining differences in the extent of LD across lakes. As LD builds up, each selected locus involved in the barrier cumulates indirect selective effects from other selected loci more efficiently in addition to its own selection coefficient, which further strengthens the reproductive barrier between populations. At equilibrium, the resulting island size and shape integrate the combined effects of selection (S), demography (N_e and m), and recombination distance between neutral and selected loci (r). Our empirical estimates of islands' parameters show that the divergence patterns characterizing the successive phases of the speciation-with-gene-flow continuum proposed by Feder et al. (2012a) can be obtained under a more complex divergence history than pure primary divergence. The significant positive correlation between the island shape parameter and the baseline level of genetic differentiation suggests that genome hitchhiking is acting to some extent in all lakes, but more so in lakes of highest phenotypic and genetic divergence. Finally, divergence and genome hitchhiking can facilitate secondary divergence within already established islands through physical linkage between selected loci (Feder et al. 2012b; Via 2012), but this could not be specifically assessed here. A more detailed investigation of each island separately will be needed to investigate those aspects, which will be achieved using many additional RAD markers that were not included in the RAD linkage map, and that will be directly mapped on the reference genome that is currently being assembled.

ACKNOWLEDGMENTS

We are grateful to G. Côté, M. L. Evans, and W. Adam for sampling whitefish. We also thank J. Boone and T. Atwood from Floragenex, as well and D. Turnbull from the University of Oregon for their precious technical help with the development and sequencing of RAD libraries. We also thank associate editors P. Nosil and J. Feder, as well as three anonymous referees for their useful comments. This research was financially supported by a postdoctoral fellowship from the Ministry of Canadian foreign affairs to P-AG, a Natural Sciences and Engineering Research Council of Canada Discovery grant and a Canadian Research Chair to LB.

LITERATURE CITED

Akey, J. M., A. L. Ruhe, D. T. Akey, A. K. Wong, C. F. Connelly, J. Madeoy, T. J. Nicholas, and M. W. Neff. 2010. Tracking footprints of artificial selection in the dog genome. Proc. Natl. Acad. Sci. 107:1160–1165.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376.

Barton, N. H. 1983. Multilocus clines. Evolution 37:454–471.

Barton, N. H., and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. Heredity 57:357–376.

Barton, N. H., and M. A. R. de Cara. 2009. The evolution of strong reproductive isolation. Evolution 63:1171–1190.

Bernatchez, L., S. Renaut, A. R. Whiteley, N. Derome, J. Jeukens, L. Landry, G. Lu, A. W. Nolte, K. Ostbye, S. M. Rogers, et al. 2010. On the origin of species: insights from the ecological genomics of lake whitefish. Philos. Trans. R. Soc. B Biol. Sci. 365:1783–1800.

Bierne, N., P. A. Gagnaire, and P. David. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. Curr. Zool. 59:72–86.

Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: models and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 38:459–487.

Campbell, D., and L. Bernatchez. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Mol. Biol. Evol. 21:945–956.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 (Bethesda) 1:171–182.

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

Cheng, C., B. J. White, C. Kamdem, K. Mockaitis, C. Costantini, M. W. Hahn, and N. J. Besansky. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline in Cameroon: a population resequencing approach. Genetics 190:1417–1432.

Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro, D. Schluter, and D. M. Kingsley. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol. 2:e109.

Crête-Lafrenière, A., L. K. Weir, and L. Bernatchez. 2012. Framing the salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. Plos ONE 7, e46662.

Crow, J. F., and K. Aoki. 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proc. Natl. Acad. Sci. 81:6073–6077.

Deagle, B. E., F. C. Jones, Y. F. Chan, D. M. Absher, D. M. Kingsley, and T. E. Reimchen. 2011. Population genomics of parallel phenotypic evolution in stickleback across stream—lake ecological transitions. Proc. R. Soc. B Biol. Sci. 279:1277–1286.

Derome, N., P. Duchesne, and L. Bernatchez. 2006. Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. Mol. Ecol. 15:1239–1249.

Derome, N., B. Bougas, S. M. Rogers, A. R. Whiteley, A. Labbe, J. Laroche, and L. Bernatchez. 2008. Pervasive sex-linked effects on transcription regulation as revealed by expression quantitative trait loci mapping in lake whitefish species pairs (Coregonus sp., Salmonidae). Genetics 179:1903–1917.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, et al. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature 491:756–760.

Elmer, K. R., and A. Meyer. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. Trends Ecol. Evol. 26:298–306.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton Univ. Press, Princeton, NJ.

Evans, M. L., K. I. M. PrÆBel, S. Peruzzi, and L. Bernatchez. 2012. Parallelism in the oxygen transport system of the lake whitefish: the role of physiological divergence in ecological speciation. Mol. Ecol. 21:4038–4050.

Excoffier, L., and H. E. L. Lisher. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

Feder, J. L., and P. Nosil. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64:1729–1747.

Feder, J. L., J. B. Roethele, K. Filchak, J. Niedbalski, and J. Romero-Severson. 2003. Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. Genetics 163:939–953.

Feder, J. L., S. P. Egan, and P. Nosil. 2012a. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Feder, J. L., R. Gejji, S. Yeaman, and P. Nosil. 2012b. Establishment of new mutations under divergence and genome hitchhiking. Philos. Trans. R. Soc. B Biol. Sci. 367:461–474.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution 35:124–138.

Gagnaire, P. A., E. Normandeau, S. A. Pavey, and L. Bernatchez. 2013. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the lake whitefish (*Coregonus clupeaformis*). Mol. Ecol. In press.

Harr, B. 2006. Genomic islands of differentiation between house mouse subspecies. Genome Res. 16:730–737.

Hoffmann, A. A., and L. H. Rieseberg. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu. Rev. Ecol. Evol. Syst. 39:21–42.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in Threespine stickleback using sequenced RAD tags. PLoS Genet 6:e1000862.

Hohenlohe, P. A., S. Bassham, M. Currey, and W. A. Cresko. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos. Trans. R. Soc. B Biol. Sci. 367:395–408.

Jacobsen, M. W., M. M. Hansen, L. Orlando, D. Bekkevold, L. Bernatchez, E. Willerslev, and M. T. P. Gilbert. 2012. Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (Coregonus spp.). Mol. Ecol. 21:2727–2742.

Jones, F. C., Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady, A. M. Southwick, D. M. Absher, R. M. Myers, T. E. Reimchen, B. E. Deagle, et al. 2012a. A Genome-wide SNP Genotyping Array Reveals Patterns Of Global And Repeated Species-Pair Divergence In Sticklebacks. Curr. Biol. 22:83–90.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, et al. 2012b. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Kolaczkowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun. 2011. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics 187:245–260.

Kruuk, L. E. B., S. J. E. Baird, K. S. Gale, and N. H. Barton. 1999. A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. Genetics 153:1959–1971.

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

Landry, L., W. F. Vincent, and L. Bernatchez. 2007. Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. J. Evol. Biol. 20:971–984.

Le Corre, V., and A. Kremer. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Mol. Ecol. 21:1548–1566.

Lu, G., and L. Bernatchez. 1998. Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of lake whitefish (*Coregonus clupeaformis* Mitchill). Proc. R. Soc. Lond. B Biol. Sci. 265:1025–1030.

———. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. Evolution 53:1491–1505.

Mackay, T. F. C., E. A. Stone, and J. F. Ayroles. 2009. The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10:565–577.

Nachman, M. W., H. E. Hoekstra, and S. L. D'Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. Proc. Natl. Acad. Sci. 100:5268–5273.

Nadeau, N. J., S. H. Martin, K. M. Kozak, C. Salazar, K. K. Dasmahapatra, J. W. Davey, S. W. Baxter, M. L. Blaxter, J. Mallet, and C. D. Jiggins. 2012a. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Mol. Ecol. 22:814–826.

Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W. Baxter, M. A. Quail, M. Joron, R. H. ffrench-Constant, M. L. Blaxter, et al. 2012b. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philos. Trans. R. Soc. B Biol. Sci. 367:343–353.

Navarro, A., and N. H. Barton. 2003. Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. Science 300:321–324.

Noor, M. A. F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439–444.

Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. B Biol. Sci. 367:332–342.

Presgraves, D. C. 2010. The molecular evolutionary basis of species formation. Nat. Rev. Genet. 11:175–180.

Pritchard, J. K., and A. Di Rienzo. 2010. Adaptation—not by sweeps alone. Nat. Rev. Genet. 11:665–667.

Renaut, S., and L. Bernatchez. 2011. Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (Coregonus spp. Salmonidae). Heredity 106:1003–1011.

Renaut, S., A. W. Nolte, and L. Bernatchez. 2009. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (Coregonus spp. Salmonidae). Mol. Biol. Evol. 26:925–936.

Renaut, S., A. W. Nolte, S. M. Rogers, N. Derome, and L. Bernatchez. 2011. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (Coregonus spp.). Mol. Ecol. 20:545–559.

Renaut, S., N. Maillet, E. Normandeau, C. Sauvage, N. Derome, S. M. Rogers, and L. Bernatchez. 2012. Genome-wide patterns of divergence during speciation: the lake whitefish case study. Philos. Trans. R. Soc. B Biol. Sci. 367:354–363.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? Evolution 47:1637–1653.

Rockman, M. V. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66:1–17.

Roesti, M., A. P. Hendry, W. Salzburger, and D. Berner. 2012a. Genome divergence during evolutionary diversification as revealed in replicate lake—stream stickleback population pairs. Mol. Ecol. 21:2852–2862.

Roesti, M., W. Salzburger, and D. Berner. 2012b. Uninformative polymorphisms bias genome scans for signatures of selection. BMC Evol. Biol. 12:94.

Rogers, S. M., and L. Bernatchez. 2006. The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). J. Evol. Biol. 19:1979–1994.

Rogers, S., and L. Bernatchez. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonus sp. Salmonidae) species pairs. Mol. Biol. Evol. 24:1423–1438.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837.

Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E. H. Byrne, S. A. McCarroll, R. Gaudet, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918.

Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78:629–644.

Servedio, M. R., and M. A. F. Noor. 2003. The role of reinforcement in speciation: theory and data. Ann. Rev. Ecol. Evol. Syst. 34:339–364.

Smadja, C. M., and R. K. Butlin. 2011. A framework for comparing processes of speciation in the presence of gene flow. Mol. Ecol. 20:5123–5140.

Smadja, C. M., B. Canbäck, R. Vitalis, M. Gautier, J. Ferrari, J.-J. Zhou, and R. K. Butlin. 2012. Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. Evolution 66:2723–2738.

St-Cyr, J., N. Derome, and L. Bernatchez. 2008. The transcriptomics of life-history trade-offs in whitefish species pairs (Coregonus sp.). Mol. Ecol. 17:1850–1870.

Stölting, K. N., R. Nipper, D. Lindtke, C. Caseys, S. Waeber, S. Castiglione, and C. Lexer. 2012. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Mol. Ecol. 22:842–855.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3:e285.

Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos. Trans. R. Soc. B Biol. Sci. 367:451–460.

Via, S., and J. West. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. 17:4334–4345.

Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:e72.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

White, B. J., C. Cheng, F. Simard, C. Costantini, and N. J. Besansky. 2010. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. Mol. Ecol. 19:925–939.

Whiteley, A. R., N. Derome, S. M. Rogers, J. St-Cyr, J. Laroche, A. Labbe, A. Nolte, S. Renaut, J. Jeukens, and L. Bernatchez. 2008. The phenomics and expression quantitative trait locus mapping of brain transcriptomes regulating adaptive divergence in lake whitefish species pairs (Coregonus sp.). Genetics 180:147–164.

Wood, H. M., J. W. Grahame, S. Humphray, J. Rogers, and R. K. Butlin. 2008. Sequence differentiation in regions identified by a genome scan for local adaptation. Mol. Ecol. 17:3123–3135.

Associate Editor: J. Feder

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Genomic distribution of differentiation between *dwarf* and *normal* whitefish for each linkage group.

Figure S2. Genomic distribution of the cross-population extended haplotype homozygosity score (*XP-EHH*) in each lake.

Table S1. Detail of Illumina read counts per individual after quality filtering and demultiplexing.