MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 545–559 doi: 10.1111/j.1365-294X.2010.04952.x

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.)

SÉBASTIEN RENAUT,*†[1] ARNE W. NOLTE,‡[1] SEAN M. ROGERS,§ NICOLAS DEROME* and LOUIS BERNATCHEZ*

**Département de Biologie, IBIS (Institut de Biologie Intégrative et des Systèmes), Université Laval, Québec, QC, Canada G1V 0A6, †Department of Botany, University of British Columbia, 6270 University Blvd, Vancouver, BC, Canada V6T 1Z4, ‡Max-Planck-Institute for Evolutionary Biology, August-Thienemann-Str., 2, 24306, Plön, Germany, §Department of Biological Sciences, University of Calgary, 2500 University Drive N.W. Calgary, AB, Canada T2N 1N4*

Abstract

As populations adapt to novel environments, divergent selection will promote heterogeneous genomic differentiation via reductions in gene flow for loci underlying adaptive traits. Using a data set of over 100 SNP markers, genome scans were performed to investigate the effect of natural selection maintaining differentiation in five lakes harbouring sympatric pairs of normal and dwarf lake whitefish (*Coregonus clupeaformis*). A variable proportion of SNPs (between 0% and 12%) was identified as outliers, which corroborated the predicted intensity of competitive interactions unique to each lake. Moreover, strong reduction in heterozygosity was typically observed for outlier loci in dwarf but not in normal whitefish, indicating that directional selection has been acting on standing genetic variation more intensively in dwarf whitefish. SNP associations in backcross hybrid progeny identified 16 genes exhibiting genotype–phenotype associations for four adaptive traits (growth, swimming activity, gill rakers and condition factor). However, neither simple relationship between elevated levels of genetic differentiation with adaptive phenotype nor conspicuous genetic signatures for parallelism at outlier loci were detected, which underscores the importance of independent evolution among lakes. The integration of phenotypic, transcriptomic and functional genomic information identified two candidate genes (sodium potassium ATPase and triosephosphate isomerase) involved in the recent ecological divergence of lake whitefish. Finally, the identification of several markers under divergent selection suggests that many genes, in an environment-specific manner, are recruited by selection and ultimately contributed to the repeated ecological speciation of a dwarf phenotype.

Keywords: ecological speciation, F_{ST} outlier methods, genome scan, genotype–phenotype association, SNP, standing genetic variation

Received 18 August 2010; revision received 19 October 2010; accepted 27 October 2010

Introduction

One of the main objectives of evolutionary biology is to elucidate the genetic basis of adaptive phenotypic traits. As natural selection acts to shape phenotypic diversity, it modulates the underlying genomic architecture in intricate ways. Despite tremendous advances in genetic studies, a link between adaptive phenotypes and genotype has been made for only a small number of traits in an even smaller number of organisms (e.g. Colosimo *et al.* 2005; Hoekstra *et al.* 2006; Miller *et al.* 2007; Counterman *et al.* 2010). These demonstrations require the combination of different research approaches

Correspondence: Sébastien Renaut, Fax: 1 604 822 6089; E-mail: sebastien.renaut.1@ulaval.ca

[1]These authors contributed equally to this manuscript.

targeting various functional and biological levels (e.g. variation at the DNA, gene expression and phenotypic levels) and represent the best strategy for deciphering the genetic basis of evolutionary change and diversification driven by natural selection (Vasemägi & Primmer 2005; Stinchcombe & Hoekstra 2008; Storz & Wheat 2010). Moreover, progress towards this goal will be best accomplished by the comparative study of evolutionarily young and ecologically distinct lineages where genetic conflicts are not fully resolved and natural introgressive hybridization, albeit limited, is still possible (Mayr 1963; Schluter 2000; Via 2009; Presgraves 2010). As such, the genetic changes contributing to the early steps of adaptive divergence and reproductive isolation can be studied before they become confounded and erased by additional genetic differences that accumulate after speciation is complete.

According to the ecological theory of adaptive radiation, shifts of organisms to novel habitats are hypothesized to be adaptive, whereby populations should diverge for specific phenotypes and genotypes that influence survival and reproduction when exposed to different environments (Mayr 1963; Schluter 2000). As a consequence, divergent selection will create heterogeneous genomic differentiation by causing specific loci (and those physically linked to them) to flow between populations less readily than others. This will result in accentuated genetic divergence of regions affected by selection, while, on the contrary, the homogenizing effects of gene flow will preclude divergence in other regions (Lewontin & Krakauer 1973; Wu 2001; Nosil *et al.* 2009). As speciation takes its course, regions under the effect of divergent selection, expected to be originally rare, will tend to grow in size and number until eventually the whole genome becomes fully incompatible and true *sensus stricto* biological species are formed (Wu 2001; Wu & Ting 2004).

Several methods have been developed to identify regions of genetic divergence (F_{ST} outlier genome scan methods, Lewontin & Krakauer 1973; Beaumont & Nichols 1996; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier & Foll 2009). Nevertheless, any of these approaches has its limitations and may be biased towards identifying only markers under particularly strong selective pressure (Michel *et al.* 2010; Storz & Wheat 2010). Thus, genome scan methods should be complemented by other approaches towards linking the effect of selection with genetic and, ultimately, adaptive phenotypic divergence (Butlin 2008; Michel *et al.* 2010). Accordingly, demonstrating the effect of divergent selection for specific loci while simultaneously associating these same loci to adaptive characters known to influence assortative mating brings compelling evidence of the genetic basis of ecological speciation. Nevertheless, such demonstrations remain few and difficult to document (Noor & Feder 2006; Schluter 2009; Presgraves 2010). Towards this goal, Via & West (2008) showed that quantitative trait loci (QTL) for adaptive traits between pea aphid populations were, albeit weakly, linked to regions of higher genetic differentiation. Via (2009) further highlighted a similar scenario in lake whitefish (Rogers & Bernatchez 2007) and hypothesized that natural selection should create relatively large region of genetic differentiation as among-populations effective rate of recombination around these markers becomes highly reduced. Conversely, speciation can also be initiated by selection acting simultaneously on many physically unlinked loci. These alternative scenarios are not mutually exclusive as divergent selection may act strongly on individual genes, creating large regions of differentiation through the process of divergence hitchhiking while concurrently acting in a global, intricate manner (Feder & Nosil 2010; Michel *et al.* 2010).

Lake whitefish from the St. John River basin (southeastern Quebec, Canada and northeastern USA) are characterized by the occurrence of several lakes harbouring dwarf and normal sympatric whitefish (Bernatchez *et al.* 2010). They represent a rare illustration of a continuum of both morphological and genetic differentiation within a given taxon, spanning from complete introgression to near-complete reproductive isolation, depending on the history (Lu *et al.* 2001) or the unique ecological characteristics of each lake (Lu & Bernatchez 1999; Landry *et al.* 2007; Landry & Bernatchez 2010). Furthermore, mounting evidence has indicated that dwarf is the derived phenotype, evolved from a normal whitefish ancestor (Landry *et al.* 2007; Rogers & Bernatchez 2007; Landry & Bernatchez 2010). For instance, dwarf whitefish exclusively occur in sympatry with normal whitefish. In addition, lakes inhabited by sympatric pairs and isolated since the last glacial retreat about 12 000 BP indicate that dwarf whitefish have evolved in parallel more than once (Pigeon *et al.* 1997). At the genetic level, genome scans have provided evidence of markers under divergent selection while concurrently identifying limited parallel patterns of genetic differentiation between independent lakes (Campbell & Bernatchez 2004; Rogers & Bernatchez 2007). In addition, a genetic basis has been demonstrated through common garden experiments and QTL mapping for adaptive traits known to differ between both forms: namely swimming behaviour, growth, morphology and gene expression variation (Rogers *et al.* 2002; Rogers & Bernatchez 2007; Derome *et al.* 2008; Whiteley *et al.* 2008). These comprehensive studies using anonymous AFLP markers nevertheless beg the question as to the nature and functional identity of the genes underlying adaptive traits under selection.

Here, through the use of a set of over 100 informative single nucleotide polymorphism (SNPs) markers developed from lake whitefish coding regions, we aimed at complementing this largely anonymous genetic basis of adaptive divergence with a more functional ecogenomic approach by conducting both genome scans in five distinct lakes containing sympatric normal dwarf species pairs as well as genotype–phenotype associations. These five lakes, differentiated in their potential for competitive interactions, phenotypic and genetic divergence between normal and dwarf, represent a continuum of ongoing ecological speciation (Bernatchez *et al.* 1999; Lu *et al.* 2001; Campbell & Bernatchez 2004; Landry *et al.* 2007; Landry & Bernatchez 2010). As such, we predicted that lakes with a lower potential for competition (and associated weaker genetic and phenotypic differentiation) should exhibit fewer SNP markers affected by natural selection compared to lakes with potentially higher competitive environments. Following this, using the same set of SNP markers, we tested the statistical association between genetic variation and phenotypic traits known to underlie the differential adaptation of normal and dwarf lake whitefish. Then, we investigated the hypothesis that genetic markers showing elevated levels of genetic differentiation should also be more strongly associated with adaptive phenotypes than other neutral markers. Finally, through the integrated use of F_{ST} outlier genome scan, genotype–phenotype association and functional genomics, we identified candidate genes involved in the recent ecological divergence of lake whitefish normal and dwarf species pairs.

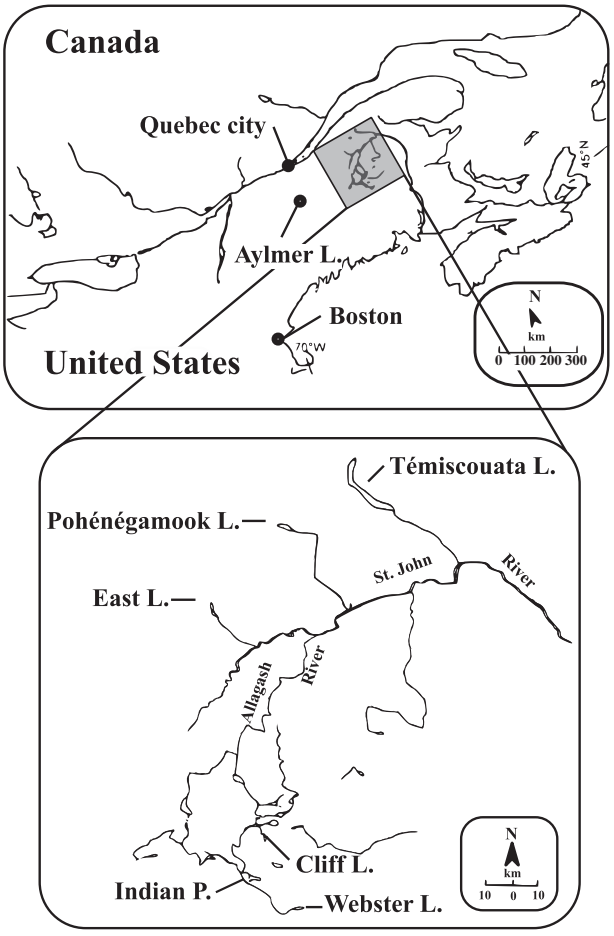

Fig. 1 Map of the study area, with locations of population samples in the St. John River basin.

Methods

Samples and study system

We used DNA samples previously collected (Campbell & Bernatchez 2004) from lakes harbouring sympatric populations of normal (N) and dwarf (D, Fig. 1): Cliff Lake (27 N, 30 D), Webster Lake (26 N, 22 D), Indian Pond (13 N, 28 D) and East Lake (24 N, 24 D) as well as material collected in 2007 from Témiscouata Lake (47°41′N, 68°47′W; 24 N, 24 D). The colonization history of all lakes except East involved a secondary contact between two independent evolutionary lineages isolated for 100 000–200 000 years BP (Acadian and Atlantic, Bernatchez & Dodson 1990). In addition, in these lakes, the dwarf phenotype most likely evolved from a normal phenotype of the Acadian lineage ancestry (Lu *et al.* 2001). In contrast, East Lake has been colonized only by the Acadian lineage (Pigeon *et al.* 1997). We also included samples collected from an allopatric normal population from Pohénégamook Lake, providing information regarding ancestral standing genetic variation that existed in the pure Acadian lineage prior to secondary contact. DNA for genotyping and sequencing was extracted using a standard proteinase K digest of tissues in SDS buffer and consecutive chloroform and high NaCl isolation.

For the association study, we used DNA samples previously extracted from 196 backcross [(Normal × Dwarf) × Dwarf] individuals (see Rogers *et al.* 2007 for details). These backcross individuals trace their origins back to Témiscouata Lake (dwarf, Acadian lineage) and Aylmer Lake (normal, Atlantic-Mississippian lineages). When necessary, we re-extracted DNA using standard proteinase K digest of tissues in SDS buffer and consecutive chloroform and high NaCl isolation from fin clips preserved in ethanol. For these individuals, nine different phenotypes previously measured and found to differentiate normal and dwarf lake whitefish were used (behavioural traits—depth selection, burst swimming, directional change, activity; physiological traits—growth rate, condition factor; morphological traits—gill rakers; life history traits—onset of sexual maturation,

gonadosomatic index, see Rogers & Bernatchez 2007 for details about phenotypic measurements).

SNP identification and genotyping

Two approaches were used for SNP discovery. First, we designed primers based on salmon ESTs that were used in microarray studies to detect differences in gene expression (Nolte *et al.* 2009a; Renaut *et al.* 2009). These primers were used to PCR-amplify fragments from the genomic DNA from pools of dwarf whitefish from Témiscouata Lake and normal whitefish from Aylmer Lake. PCR amplicons were Sanger-sequenced and then visually screened for putative SNPs. Additional SNPs were chosen from a 454 sequencing experiment of cDNA libraries derived from dwarf and normal whitefish from Cliff Lake as well as from the same backcross individuals used in the association study. Putative SNPs previously identified (see Renaut *et al.* 2010 for all details on SNP identification criteria) were visually inspected in an attempt to discard erroneous assemblies or low-quality SNPs, which may cause errors in primer design and amplification during genotyping. Briefly, regions 200-bp upstream and downstream of a SNP of interest and which contained two or more SNPs or indel in full linkage disequilibrium were discarded as they are likely to represent paralogous sequence variants. SNPs closer than 100 bp from the contig end were also discarded. All sequences were matched (BLASTn) against NCBI nr/nt database, and only the best hit for each amplicon was retained for annotation purposes.

Genotyping assays were designed and performed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) developed by Sequenom (San Diego, CA, USA) at the Genome Quebec Innovation Center (McGill University, Montreal, Canada). Genotyping assays were developed for 470 putative SNPs. Replicate genotyping of positive controls indicated a maximum error-calling rate of 4.10% (24 inconsistencies out of 586 genotypes).

We calculated pairwise F_{ST}, observed and expected heterozygosities for each whitefish population independently using MICROSATELLITE ANALYZER (Dierenger & Schlötterer 2003; Table S3, Supporting Information). For the purpose of identifying loci subject to selection, we used F_{ST} estimates from the five lakes harbouring sympatric populations. BAYESFST (Beaumont & Balding 2004) was used to test for the significance of outlier loci for all polymorphic SNPs within each lake (two populations defined, normal and dwarf). We interpreted evidence of positive selection as suggested by Beaumont & Balding (2004). For each locus, a positive value of the locus parameter (α_i) suggests that locus *i* is subject to divergent selection, whereas a negative value suggests that balancing selection tends to homogenize allele frequencies over populations. We calculated 10 000 values of α_i and defined α_i to be 'significant at level *P*' if *P* per cent of the values were positive (evidence of divergent selection). We ran simulations five times with different seed values for the algorithm to ensure reproducibility of probability values (Pearson's correlation coefficient >0.99). Note that BAYESFST deals with the problem of multiple hypotheses testing through the prior distribution of the regression parameter for the locus parameter α_i. Therefore, *P*-values calculated in BAYESFST are very conservative compared to frequentist method based on summary statistics (e.g. FDIST, Beaumont & Nichols 1996; Beaumont & Balding 2004).

A chi-square test was performed to verify whether parallel trends of genetic divergence between lakes were observed more often than expected at random. Expected numbers of loci showing parallel trends between two lakes were calculated from the product of the percentage of outliers in the first and second lake by the total number of markers surveyed (Campbell & Bernatchez 2004).

Normalized phenotypic data for the nine adaptive phenotypes (Rogers & Bernatchez 2007) were used to perform an association analysis in the R environment (v.2.10.1. The R Foundation for Statistical Computing®, 2010) with the package SNPassoc (v1.6, Gonzalez *et al.* 2007) using a codominant genetic model. We applied, for each SNP, a general linear model and a likelihood ratio test to obtain probability values. For each test, *P*-values were then corrected for multiple hypotheses testing using *q*-value correction (*q*-value package, Storey 2002).

Results

SNP identification and genotyping

After exclusion of failed marker assays, monomorphic markers and those markers exhibiting excess heterozygosity, a set of 112 (16 identified through Sanger sequencing, 96 through 454 sequencing) was retained for further analyses (see Table S1, Supporting Information (genome scan) and 2 (association) for lists of informative SNPs as well as the dbSNP database, http://www.ncbi.nlm.nih.gov/projects/SNP, ss270137550-270137661). Ninety-six of those SNPs were informative in the genome scan of natural populations (mean missing data rate of 7.5% across individuals, Table S1, Supporting Information) and 87 for genotype–phenotype associations (mean missing data rate of 16.1% of across individuals, Table S2, Supporting Information), such that 63% of all markers (70/112) were informative in both data sets.

Table 1 Estimates of genetic differentiation between dwarf and normal lake whitefish

Normal–dwarf pairwise F_{ST} (mean)	Cliff	Webster	Indian	East	Témiscouata
SNP*	0.28	0.11	0.06	0.02	0.01
AFLP†	0.22	0.17	0.04	0.11	NA
Microsatellite‡	0.26	0.14	0.08	0.06	0.04

*Estimates for SNP are based on 94 polymorphic nuclear loci, †440 AFLP loci (reported from Campbell & Bernatchez 2004) and ‡six microsatellite loci (reported from Lu & Bernatchez 1999). NA: Témiscouata Lake was not analysed using AFLP markers by Campbell & Bernatchez (2004).

Genome scan of sympatric normal and dwarf species pairs

Cliff Lake had the highest mean pairwise F_{ST} value (0.28), followed by Webster (0.11), Indian (0.06), East (0.02) and Témiscouata (0.01, Table 1). These values were highly concordant with F_{ST} values calculated previously from both microsatellite (Pearson's correlation coefficient = 0.99) and AFLP (Pearson's correlation coefficient = 0.83) markers as well as with the extent of phenotypic differentiation (Pearson's correlation coefficient = 0.86) previously observed between normal and dwarf whitefish (Lu & Bernatchez 1999; Campbell & Bernatchez 2004).

We identified outlier loci showing accentuated patterns of genetic differentiation (Fig. 2 and Table 2). Outliers were detected in every pairwise comparison, except for Témiscouata Lake. At a *P*-value of 0.2 calculated from the 10 000 iterations of the locus parameter (α_i) in BAYESFST, the number of outliers was 12, 7, 5, 3 and 0 for Cliff, Webster, Indian, East and Témiscouata, which represented 15, 8.4, 6.1, 3.7 and 0 per cent of the markers tested, respectively. Furthermore, the number and percentage of outliers were positively correlated with the extent of genetic divergence between normal and dwarf (Pearson's correlation coefficient = 0.96). The two mitochondrial markers (cytochrome c oxidase subunit 3 and NADH ubiquinone oxidoreductase chain 5) were completely fixed between normal and dwarf in Cliff Lake (F_{ST} = 1), as previously reported for the whole mitochondrial genome (Bernatchez & Dodson 1990). Glucose-6-phosphatase (F_{ST} = 0.94, Cliff Lake), a gene playing a key role in regulating glucose levels in the blood, was the nuclear gene showing the greatest F_{ST} value. Several other genes also had a high outlier F_{ST} value depending on lakes (e.g.: probable ubiquitin carboxyl terminal hydrolase with F_{ST} = 0.93 in Cliff Lake and F_{ST} = 0.41 in Indian Lake; heat-shock protein HSP 90 beta with F_{ST} = 0.71 in Webster Lake, cyclin I with F_{ST} = 0.21 in East Lake). Four genes (cyclin I, heat-shock 27-kDa protein, probable ubiquitin carboxyl terminal hydrolase and sodium/potassium-transporting ATPase

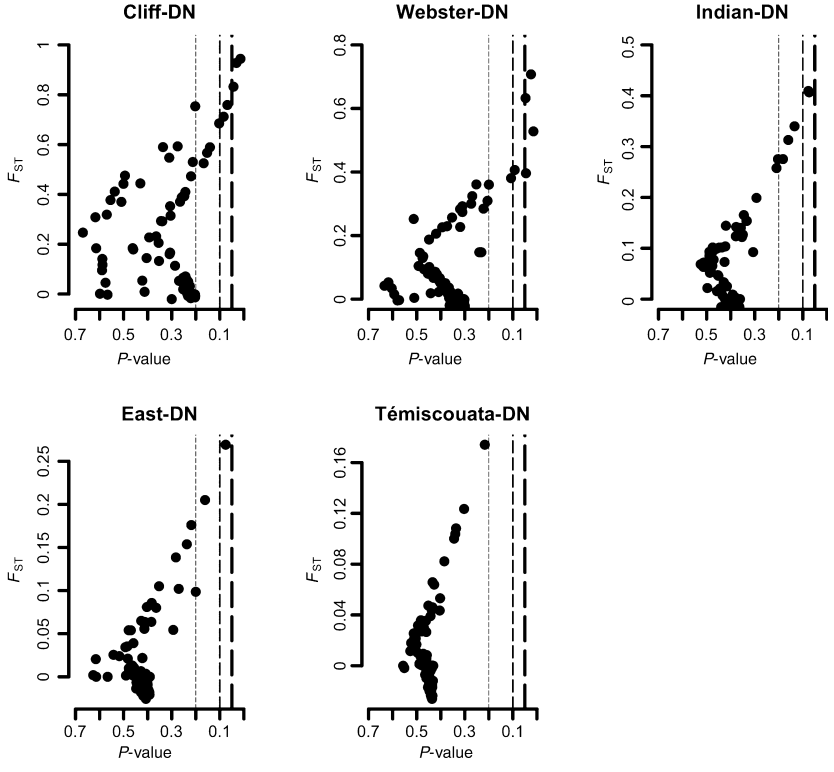

Fig. 2 Pairwise F_{ST} as a function of probability between sympatric dwarf and normal whitefish in five lakes. Dashed lines represent 0.2, 0.1 and 0.05 *P*-values for all five genome scans performed independently.

subunit alpha) were also identified as parallel F_{ST} outliers in more than one lake, although in only one comparison (Cliff–East Lake), was the number of outliers greater than expected by chance alone (chi-square test, *P*-value = 0.04, Table 3).

Selection acting on standing genetic variation in dwarf whitefish

In Cliff and Webster Lakes where dwarf and normal whitefish are the most genetically and phenotypically differentiated, observed heterozygosity was significantly reduced for outlier loci in dwarf whitefish ($H_{o\,(dwarf)}$ = 0.03 and 0.32 for outliers compared to 0.28 and 0.44 for all other markers in Cliff and Webster, respectively, *t*-test, *P*-value <0.05), but not in normal whitefish (Fig. 3). In Cliff Lake, ten of the twelve outliers that were polymorphic in normal were fixed in dwarf whitefish. This trend of reduced diversity at outlier loci in dwarf but not in normal whitefish was similar, but not significant in Indian and East Lakes, possibly owing to large variance estimates associated with a smaller number of outliers. In addition, heterozygosity values in Pohénégamook Lake (H_o = 0.28, pure Acadian lineage origin) and East Lake normal whitefish (H_o = 0.33, pure Acadian lineage origin) for loci identified as outliers in Cliff Lake confirmed that these SNPs were polymorphic in the ancestral Acadian lineage (*P*-value <0.01 & *P*-value <0.001 against dwarf H_o for outliers in Cliff lake).

Association study

As the fish used in the association study came from a backcross-like family [(H × D) × D], most informative polymorphic markers segregated in a 1:1 homozygous/heterozygous fashion (61 markers), while 26 informative markers segregated in a 1:2:1 fashion (26 markers). Thirty-one out of 87 markers (35%) showed evidence of segregation distortion (*q*-value <0.2, Table S2, Supporting Information), which is consistent with microsatellite and AFLP data (Rogers *et al.* 2007). One SNP genotype (myosin regulatory light chain 2 skeletal, *q*-value = 0.03) was associated with swimming activity (standard deviation of the depth preference of the individual divided by the mean depth observed, as defined by Rogers & Bernatchez 2007), three with condition factor (weight/length^3) and nine with number of gill rakers. The strongest association was observed for growth (grams of weight gained per day) where ten SNPs were associated with this phenotype (Table 4). In total, seven SNPs were associated with two phenotypes (aryl hydrocarbon receptor 2 alpha, EAP30 subunit of ELL complex a (Eap30a), ferritin middle subunit, *Gasterosteus aculeatus* clone VMRC26-150D01, PREDICTED:

Table 2 Summary of all F_{ST} outliers identified from genome scans between sympatric dwarf and normal whitefish in five lakes

SNP functional annotation	Cliff	Webster	Indian	East	Témiscouata
26S protease regulatory subunit 4	**0.83*****	0.04	0.10	0.00	−0.02
Antithrombin III precursor	0.39	0.02	**0.41****	0.08	−0.01
ATP-binding cassette subfamily E member 1	0.37	0.15	**0.28***	−0.02	−0.01
Glucose 6 phosphatase	**0.94*****	0.29	0.20	−0.02	0.00
Triosephosphate isomerase	**0.69****	0.06	−0.03	0.00	−0.02
Proteasome subunit beta type 8 precursor	**0.57***	0.36	0.14	−0.01	−0.02
Fibrinogen beta chain precursor	0.47	**0.38***	0.14	0.02	−0.02
T complex protein 1 subunit epsilon	0.16	**0.40****	−0.03	0.01	0.08
ATP synthase subunit e	0.14	**0.63****	0.17	−0.01	0.01
Multifunctional protein ADE2	0.10	0.13	**0.31***	0.01	−0.02
Probable ubiquitin carboxyl terminal hydrolase	**0.93*****	−0.01	**0.41****	0.04	0.00
Heat-shock 27-kDa protein	0.39	**0.41****	**0.28***	0.18	0.01
NADH ubiquinone oxidoreductase chain 5 (mitochondrial)	**1.00*****	−0.04	0.15	0.00	0.00
Cyclin I	**0.53***	**0.36***	0.08	**0.21***	0.02
Cytochrome c oxidase subunit 3 (mitochondrial)	**1.00*****	−0.02	0.15	0.00	0.00
Heat-shock protein HSP 90 beta	0.44	**0.71*****	0.05	−0.01	0.07
Complement component C9	**0.59****	−0.02	0.13	0.02	0.03
Angiotensinogen	0.41	**0.53*****	0.34	−0.03	−0.02
Salmo salar RED protein	**0.76****	0.31	0.14	0.09	−0.02
Sodium/potassium-transporting ATPase subunit alpha	**0.71****	0.00	NA	**0.10***	−0.02
Sodium/potassium-transporting ATPase subunit beta	**0.75***	0.32	0.09	0.06	0.05
Ribulose phosphate 3 epimerase	−0.01	0.07	−0.03	**0.27****	−0.01

Significant outliers are in bold. *P*-values calculated from BAYESFST outputs as described in MM: **P*-value <0.2, ***P*-value <0.1, ****P*-value <0.05. Genes underlined were also associated with adaptive phenotypes (see Table 3).

Table 3 Chi-square test to assess whether parallel trends for F_{ST} outlier loci were observed more often than expected by chance

	Cliff	Webster	Indian	East	Témiscouata
Cliff	X	Cyclin I	Ubiquitin carboxyl terminal hydrolase	Na-K transporting ATPase subunit alpha & Cyclin I*	0
Webster	1.21	X	Heat Shock 27 kDa protein	0	0
Indian	0.88	0.49	X	0	0
East	0.53	0.30	0.22	X	0
Témiscouata	0	0	0	0	X

Genes included in this table were those that were outliers in more than one lake. Below the diagonal are expected values; above the diagonal are observed values (outlier loci). *P-value = 0.04 in East–Cliff comparison (two genes), all other comparisons non-significant.

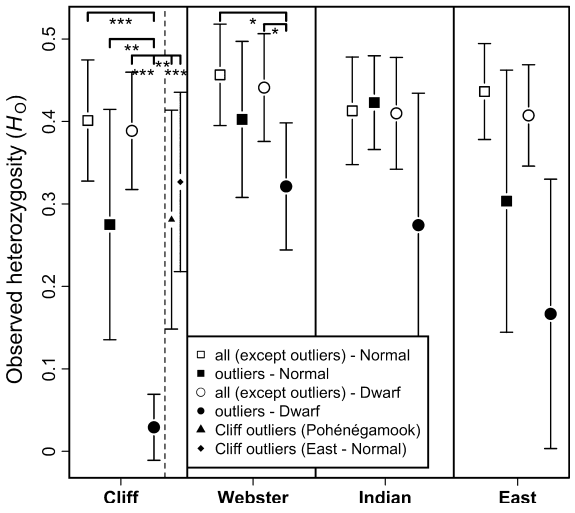

Fig. 3 Observed heterozygosity for outlier and non-outlier loci, separately for normal and dwarf populations. Heterozygosity was also compared for Cliff Lake outlier loci in Pohénégamook, as well as East Lake normal whitefish, because they represent the ancestral Acadian lineage from which Cliff dwarf whitefish evolved. *T*-tests comparing mean heterozygosity between each of the groups. *P-value <0.05, **P-value <0.01, ***P-value <0.001.

Danio rerio zinc finger protein 638-like, TY3 GYPSY-like LTR retrotransposon, zebrafish DNA from clone DKEY 16P21) and nine with one phenotype (ATP-binding cassette subfamily E, inosine monophosphate dehydrogenase 2, myosin regulatory light chain 2 skeletal, NA/K-transporting ATPase subunit alpha 1, putative ISG12(3) protein, T complex protein 1 subunit epsilon, tetraspanin 4, triosephosphaste isomerase, uncharacterized protein C21orf51).

Link between genome scan and association

There was no significant trend for SNP markers associated with adaptive phenotypes to show greater evidence of reduced gene flow (higher F_{ST}) compared to all other markers (P-value >0.05 for all five lakes, Wilcoxon rank sum test between candidate loci and all other markers, Fig. 4). Three genes (triosephosphate isomerase, T complex protein 1 subunit epsilon, NA/K-transporting ATPase subunit alpha) associated with an adaptive phenotype also exhibited outlier levels of divergence in Cliff, Webster and Cliff, and East Lake, respectively, but this was not more than expected at random (P-value >0.05, chi-square test for all five lakes, whereby the expected numbers of outlier loci associated with phenotypes were calculated by multiplying the percentage of outlier loci for a lake times the percentage of markers associated with phenotypes times the total number of markers surveyed).

Discussion

Genome scan of sympatric normal and dwarf species pairs

In a recent review, Nosil *et al*. (2009) reported that the proportion of outliers in the literature varied greatly between genome scan studies, ranging from 0.4% to 25%. According to the authors, this discrepancy between studies is best explained by variation in the number of populations and individuals examined, molecular markers employed, methods for estimating baseline neutral differentiation and criteria determining outlier status. While this evidently explains part of the variation, we believe that this may also reflect the specific intensity of divergent selection towards a new adaptive optimum and/or time since divergence varying between model systems. For instance, genome-wide SNP data for *Mus musculus musculus* and *Mus musculus domesticus* which have diverged for nearly 1 Myr revealed that genomic islands of differentiation represented 7.5% of autosomal regions and 90% of the X chromosome (Harr 2006). In comparison, thorough genome scan between M and S forms of *Anopheles gambiae*, where divergence is very recent and speciation ongoing, identified three islands of divergence representing <3% of the genome (White *et al*. 2010). While between-species comparisons will remain difficult owing to the

Table 4 SNP genotypes associated with adaptive phenotypes

Trait	SNP functional annotation	*N*	G1	G2	G3	Mean (95% CI) – G1	Mean (95% CI) – G2	Mean (95% CI) – G3	*P*-value	*q*-value
Behavioural										
Activity	Myosin regulatory light chain 2 skeletal	103	AA	AG	–	50.28 (43.61–56.94)	36.38 (31.71–41.04)	–	0.000	0.032
Physiological								–		
Growth	T complex protein 1 subunit epsilon	178	CC	CT	TT	540.33 (318.97–761.69)	527.3 (431–623.59)	232.37 (79.71–385)	0.004	0.057
Growth	Gasterosteus aculeatus clone VMRC26-150D01	178	GG	AG	–	319.15 (218.18–420.12)	549 (430.3–668.2)	–	0.004	0.057
Growth	TY3 GYPSY-like LTR retrotransposon	178	CC	CT	–	323.18 (219.71–426.65)	543.1 (426.46–659.74)	–	0.006	0.057
Growth	Aryl hydrocarbon receptor 2 alpha	178	GG	AG	–	543.1 (426.46–659.74)	323.18 (219.71–426.65)	–	0.006	0.057
Growth	Tetraspanin 4	178	GG	TG	–	323.24 (216.7–429.76)	546.05 (427.2–664.91)	–	0.007	0.057
Growth	PREDICTED: Danio rerio zinc finger protein 638-like	178	CC	CT	–	330.4 (233.25–427.56)	549.35 (426.22–672.48)	–	0.006	0.057
Growth	Uncharacterized protein C21orf51	178	AA	AG	GG	−116.7 (−1190.6 to −957.3)	532.63 (430.36–634.9)	328.15 (260.16–396.14)	0.001	0.057
Growth	EAP30 subunit of ELL complex a (Eap30a)	178	AA	AC	–	543.1 (426.46–659.74)	323.18 (219.71–426.65)	–	0.006	0.057
Growth	Ferritin middle subunit	178	TT	CT	–	319.38 (214.7–424.06)	543.1 (426.46–659.74)	–	0.006	0.057
Growth	Zebrafish DNA from clone DKEY 16P21	178	GG	AG	–	543.1 (426.46–659.74)	323.18 (219.71–426.65)	–	0.006	0.057
Morphological										
Gill rakers	ATP-binding cassette subfamily E	138	CC	AC	–	23.35 (23.05–23.65)	22.84 (22.49–23.2)	–	0.030	0.195
Gill rakers	Gasterosteus aculeatus clone VMRC26-150D01	138	GG	AG	–	23.3 (23–23.64)	22.84 (22.5–23.17)	–	0.031	0.195
Gill rakers	Putative ISG12(3) protein	138	GG	GT	TT	23.56 (23.06–24.05)	23 (22.75–23.24)	22.69 (22.06–23.32)	0.033	0.195
Gill rakers	TY3 GYPSY-like LTR retrotransposon	138	CC	CT	–	23.34 (23.03–23.65)	22.84 (22.52–23.16)	–	0.030	0.195
Gill rakers	Aryl hydrocarbon receptor 2 alpha	138	GG	AG	–	22.84 (22.52–23.16)	23.34 (23.03–23.65)	–	0.030	0.195
Gill rakers	PREDICTED: Danio rerio zinc finger protein 638-like	138	CC	CT	–	23.34 (23.04–23.64)	22.82 (22.48–23.15)	–	0.023	0.195
Gill rakers	EAP30 subunit of ELL complex a (Eap30a)	138	AA	CA	–	22.84 (22.52–23.16)	23.34 (23.03–23.65)	–	0.030	0.195
Gill rakers	Ferritin middle subunit	138	GG	AG	–	23.34 (23.03–23.65)	22.84 (22.52–23.16)	–	0.030	0.195
Gill rakers	Zebrafish DNA from clone DKEY 16P21	138	TT	CT	–	22.84 (22.52–23.16)	23.34 (23.03–23.65)	–	0.030	0.195
Condition factor	NA/K-transporting ATPase subunit alpha 1 precursor	182	AA	AC	CC	1.18 (1.13–1.22)	1.18 (1.167–1.2)	1.11 (1.06–1.16)	0.008	0.199
Condition factor	Triosephosphate isomerase	182	CC	AC	–	1.19 (1.17–1.21)	1.14 (1.11–1.17)	–	0.005	0.199
Condition factor	Inosine monophosphate dehydrogenase 2	182	CC	AC	–	1.14 (1.12–1.16)	1.19 (1.16–1.23)	–	0.005	0.199

N refers to the sample size. G1, G2, G3 are the genotypes of the SNP with the mean and 95% confidence interval for the phenotypic values corresponding to G1, G2 and, when applicable, G3 genotype.

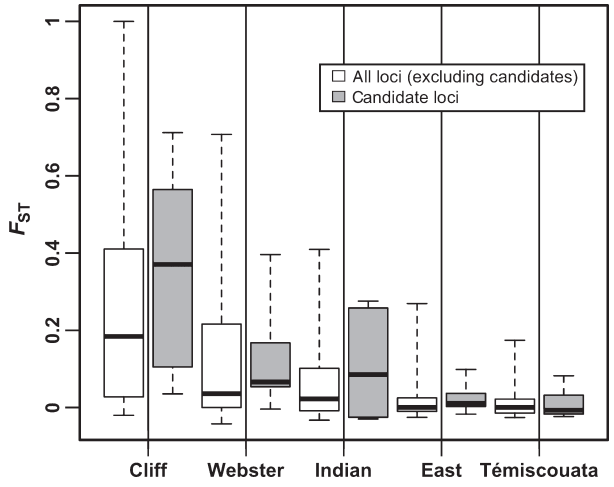

Fig. 4 Boxplot of F_{ST} in each of the five sympatric lakes for loci associated with adaptive phenotypes (candidate loci) and all other loci. Boxes represent 95% confidence intervals of the median (dark line), and whiskers extend to data extremes.

different analytical methods and markers used, cases of closely related species such as dwarf and normal whitefish permit evaluations of the effect of divergent selection according to a gradient of ongoing ecological speciation. Here, by comparing species pairs that evolved independently, we identified that the proportion of markers under divergent selection was associated with the previously recorded lake-specific potential for competition and phenotypic differentiation, as discussed below.

All five lakes studied belong to the same river basin and are situated at relatively the same altitude (from 201 m for Témiscouata Lake to 322 m for East Lake). Yet, none has direct connection, and the lakes are therefore physically isolated one from the other. Based on the hydrological history of the area, they most likely have been colonized and isolated from one another at around the same time period, during the isostatic rebound following the removal of the 2.2-km-thick ice sheet cover (12 000 BP, Castric *et al.* 2001). In contrast to similar colonization times, differences in abiotic and biotic characteristics translated into different ecological landscapes (Landry *et al.* 2007; Landry & Bernatchez 2010). For example, Landry *et al.* (2007) showed that the three lakes harbouring the most divergent sympatric populations (Cliff, Webster and Indian Lakes) were characterized by less habitat available owing to shallower mean depth and oxygen depletion below the thermocline during the growing season, less zooplanktonic prey biomass and smaller prey size range compared to the least divergent populations (East and Témiscouata lake), which had more habitat available and greater prey density. Such a prey structure has also previously been interpreted as evidence for increased potential for competition in other systems (Magnan 1988; Svanbäck & Persson 2004). Presented with this evidence, the authors concluded that resource limitation resulted in increased potential for competition and selective pressure towards optimal normal and dwarf adaptive peaks (Landry *et al.* 2007). Here, by means of genome scans, we showed that this increased potential for competition and intensity of selection at the phenotypic level is also reflected at the genetic level because more markers were identified as being potentially under the effect of divergent selection in Cliff, Webster and Indian (20 different SNPs) compared to East and Témiscouata Lakes (two different SNPs). As such, we experimentally confirmed one of the premises of the genic view of speciation (Wu 2001), whereas populations that are more representative of the early steps of ecological speciation in whitefish (Témiscouata and East Lakes) have fewer genetic markers under the effect of natural selection compared to more diverged species (Cliff, Webster and Indian Lakes). At this point, linkage information from the association family regarding the number or size of these islands of divergence is limited given the predicted haploid number of chromosomes (40). Nevertheless, given that these outliers are not in greater linkage disequilibrium than the rest of the genome (Table S4, Supporting Information), they should not represent a single large block in LD, but more likely, it suggests that outlier markers are situated on distinct linkage groups. In addition, the exact number of significant outliers will depend on the stringency of the multiple hypotheses testing correction, and the conclusions still hold with more stringent criteria (at P-value <0.1, number of outliers is 6, 5, 2, 1, 0 and again strongly correlated with the extent of genetic differentiation, see Table 2).

Parallel patterns of genetic differentiation

Previous studies have provided evidence for the role of parallel phenotypic evolution in lake whitefish speciation. For example, Lu & Bernatchez (1999) and Rogers & Bernatchez (2005) have documented strong parallelism for phenotypic traits varying between dwarf and normal whitefish in independent lakes. Parallelism was also observed at the transcriptome level, whereas genes were differentially expressed between normal and dwarf whitefish in independent populations more often than expected (Derome *et al.* 2006; St-Cyr *et al.* 2008). In contrast, less evidence for parallelism has been observed at the genetic level. For instance, modest, yet significant, parallelism between at least two lakes out of four was identified for only six out of 48 anonymous outlier AFLP markers (Campbell & Bernatchez 2004; see

also Rogers & Bernatchez 2007). Similarly, here we observed little congruence among lakes (Tables 2 and 3). Only four parallel SNPs out of 96 markers displayed parallel genetic differentiation, and this was not more than expected at random except in one lake pair comparison. Although these findings remain to be rigorously confirmed using a greater number of markers, they follow the same trend as previous AFLP genome scans. Consequently, in whitefish at least, parallelism at the phenotypic level (including gene expression) is not mirrored by mutations in coding regions.

Given that selection acts at the level of the phenotype, it is plausible that alternate evolutionary trajectories will be taken as selection recruits different mutations while, ultimately, leading whitefish to the same ecological normal (benthic) and dwarf (limnetic) niche space in the adaptive landscape. In beach mice for example, similar fur coloration evolved independently through alternative mutations (Steiner *et al.* 2009). Recent analyses of the factors that shape parallel hybrid zones in sculpins (*Cottus* spp) have also provided evidence that the genetic factors that underlie adaptive differentiation differ between populations (Nolte *et al.* 2009b). Another unequivocal case comes from Stanek *et al.* (2009). Here, the authors set up a simple selection experiment in *Escherichia coli* as a mean to assess the genetic basis of adaptation. While they identified a beneficial mutation rising to fixation and conferring a strong fitness advantage in one population, contrary to expectation they could not find any evidence of parallel adaptation in any of the other 11 replicate populations. As such, they concluded that even for simplistic evolutionary scenarios, the genetic basis of adaptation is highly unpredictable. In our current study, different sets of outlier genes were detected in each species pairs of whitefish, which supports our current working hypothesis that many genes associated with numerous biological functions are involved in the adaptive divergence of lake whitefish.

Selection acting on standing genetic variation in dwarf whitefish

Mounting evidence has revealed that, in the context of this adaptive radiation, directional selection acted more strongly on dwarf rather than normal whitefish (Bernatchez 2004). Namely, dwarf whitefish appear to be at an 'ecological disadvantage' relative to normal, in terms of growth, fecundity (Rogers & Bernatchez 2005) and survival (Fenderson 1964). Higher mortality rate in dwarf whitefish could be related to higher predation pressure (Kahilainen & Lehtonen 2002), while stunted dwarf growth is probably attributable energy trade-offs at the profit of higher swimming and metabolic activity (Trudel *et al.* 2001; Rogers *et al.* 2002). In addition, diversity of prey utilized by dwarf whitefish is also less than for normal, translating into a more specialized diet (Bernatchez *et al.* 1999) and more pronounced selection in dwarf acted on the number of gill rakers, a trait involved in prey selection (Bernatchez 2004).

Here, we identified strong reduction in heterozygosity for outlier loci under selection in dwarf whitefish only. This was especially true in Cliff, the lake harbouring the most divergent dwarf–normal pair, where the fixation of a single allele was observed for ten out of twelve outlier markers (Fig. 3). Given the strong effect of selection and low level of gene flow between dwarf and normal whitefish in Cliff Lake, we were able to assess the level of ancestral polymorphism in dwarf whitefish for these outlier SNPs. These loci were polymorphic and did not show any reduction in heterozygosity in Pohénégamook Lake as well as the normal whitefish from East Lake, two populations of pure Acadian ancestry closely related to the ancestors of the extant dwarf whitefish in Cliff Lake (Lu *et al.* 2001). The pattern observed in Cliff dwarf whitefish cannot either be explained by a population bottleneck that occurred during lake colonization because heterozygosity was not reduced for the genome as a whole. Similarly, a general reduction in heterozygosity at outliers in all dwarf populations cannot be explained by a single deterministic event prior to the colonization because dwarfs have evolved multiple times in independent lakes (Pigeon *et al.* 1997). Conversely, locus-specific reduction in genetic variability is often a telltale sign of selective sweep and positive selection (Nielsen 2005). Accordingly, our results imply that natural selection, by differentially sorting out standing genetic variation present prior to the recent ecological speciation of whitefish, has ultimately contributed to the independent evolution of a dwarf phenotype in each lake. As such, our results represent a clear case of the predominant role of selection acting on standing genetic variation, rather than *de novo* mutations, in driving adaptive divergence in the early steps of ecological speciation.

Association

Even for a relatively small set of 87 markers, 16 SNP genotypes were significantly associated with adaptive phenotypes. Seemingly, the exact number of significant association depends on the stringency of the multiple hypotheses correction. We also believe, as discussed in the previous section, that to identify specific targets of speciation, combining several lines of evidence is more informative than relying on a single rigid analysis, which may in any case still present biases (Butlin 2008; Stinchcombe & Hoekstra 2008). The strongest evidence

for genotype–phenotype association was with growth as ten SNPs were associated with this phenotype. This corroborates results of Rogers & Bernatchez (2007), showing that growth QTLs were the most common and, therefore, the slower growth of dwarf vs. normal (Trudel *et al.* 2001; Rogers & Bernatchez 2005) is likely to be under polygenic control. The gill raker apparatus is another common adaptive trait known to differentiate benthic (few gill rakers) and limnetic (many gill rakers) species pairs (McPhail 1993; Bernatchez 2004). Here, this trait also appears to be under polygenic control because it was associated with nine SNPs, although the higher *q*-values (0.19) imply a greater false discovery rate. Finally, we also found a genetic basis for the differences in swimming activity and condition factor, whereas dwarf whitefish are known to be more active swimmers and have a more slender body (smaller condition factor) compared to normal fish (Trudel *et al.* 2001).

Link between genome scan and association

There was no significant overall link between elevated rates of genetic divergence and association with adaptive phenotypes (Fig. 4). As suggested, confounding demographic, spatial or local effects on adaptive divergence may affect F_{ST} among environments (Beaumont & Balding 2004; Storz 2005). This appears to be corroborated by other studies looking at the relationship between selection at the genetic level and adaptive QTLs, which found either weak or no correlation between both (whitefish, Rogers & Bernatchez 2007; sunflowers, Yatabe *et al.* 2007; pea aphids, Via & West 2008; sticklebacks, Mäkinen *et al.* 2008). Here, variation owing to the lineage origin may influence the genetic architecture sparking ecological divergence and sculpted by selection, thus generating unique evolutionary scenarios in each lake. This, in turn, could dampen our power to detect adaptive traits under selection unless we had generated independent hybrid families in each lake, which, at this point at least, is not technically feasible. Lastly, because our study probably examined only a subset of adaptive phenotypic traits, truly outlier loci may nonetheless be associated with adaptive phenotypes not yet examined (enzyme production, parasite avoidance mechanisms, mating behaviours, etc.).

Moreover, relationships between adaptive phenotypes and regions of elevated genetic divergence may always be, at best, weak. Given the few recombination events in a hybrid backcross, large chromosomal regions are expected to be in linkage disequilibrium. Therefore, this can explain why, using few markers, we were able to find a relatively large number of associations in the backcross family. On the other hand, in natural populations, linkage disequilibrium around a selected locus can, in theory, be much smaller, thus distorting the association between divergent selection and adaptive phenotypes. This can explain why we did not find a significant link between divergent selection and adaptive phenotypes while using a greater number of markers in the same study system, Rogers *et al.* (2007) identified a small, yet significant, connection.

Integrating data towards the identity of candidate genes

Finding outlier loci also responsible for adaptive traits and being differentially expressed allows stronger inferences than the sole use of genome scans about the underlying genes associated with ecological divergence (Stinchcombe & Hoekstra 2008; Nosil *et al.* 2009). Here, we discuss this integrated approach for the two strongest candidates while being conscious that other candidates could also merit from a more detailed analysis (e.g. T complex protein 1 subunit epsilon or cyclin I and heat-shock 27-kDa protein as previously explained). Admittedly, we still do not have evidence that these SNPs are the mutations responsible for an adaptive phenotype or the direct target of selection. Nevertheless, they must at least be in strong linkage with a causative mutation nearby. Our ongoing work, involving screening and sequencing BAC libraries, may provide a more in-depth appreciation of the relative causative importance of regulatory and/or structural mutations for these two candidate genes.

The first case involves the sodium/potassium ATPase gene (Fig. 5a). This highly conserved protein complex is composed of two subunits in teleost fish (alpha and beta) and actively transports sodium and potassium ions in opposite directions across the plasma membrane (Lodish *et al.* 2008). As a result of its essential role, it is one of the single major users of ATP, responsible for 5%–40% steady-state cellular energy consumption (Ewart & Klip 1995). Furthermore, it has frequently been identified as involved in local adaptation in fish (e.g. McCairns & Bernatchez 2010). Therefore, as tradeoffs in energy allocation between high metabolic rate (dwarf) and increased growth (normal) are one of the main factors explaining the differentiation between normal and dwarf phenotypes (Trudel *et al.* 2001; Rogers & Bernatchez 2007; St-Cyr *et al.* 2008), these high ATP consumers' genes emerge as plausible candidates. Previous microarray experiments showed that the NaK-ATPase (alpha) gene is upregulated in normal whitefish compared to dwarf at the juvenile stage (Nolte *et al.* 2009a). Here, one SNP located in the 3′ UTR of the alpha subunit (Fig. 5a) was associated with a condition factor phenotype (CC genotype associated with smaller

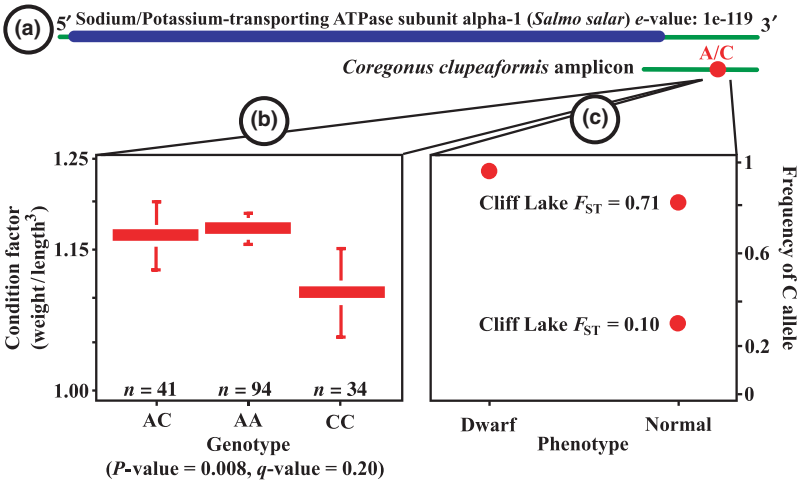

Fig. 5 (a) Genotypic characteristics of an SNP found in sodium/potassium ATPase subunit alpha gene of whitefish (BLAST *e*-value <1e-119). The SNP is located in the 3′ UTR (in green) according to the *Salmo salar* open reading frame (in blue). (b) SNP is associated with condition factor. (c) In natural populations, the SNP is an outlier in both Cliff and East Lakes.

condition factor, Fig. 5b). In the genome scan, this SNP was an outlier fixed for the same allele (C) in both Cliff and East Lakes (Fig. 5c). In fact, the C allele was statistically associated with dwarf whitefish in the association family, while the A allele with normal fish (Fig. 5c), and this corroborates the fact that dwarf whitefish are more slender (smaller condition factor relative to normal whitefish, Fig. 5b). Finally, another SNP coding for a synonymous mutation in the subunit *beta* was outlier in Cliff Lake and, in all other lakes, had an F_{ST} value above the mean F_{ST} for that lake (Table 2).

The second case involves the triosephosphate isomerase gene (Fig. 6a). TPI regulates the fifth step of glycolysis and is essential for efficient energy production (Lodish *et al.* 2008). Again, given the previously identified trade-offs in energy allocation between dwarf and normal, functional or regulatory changes in genes directly involved in energy production, through either glycolysis or oxidative phosphorylation, are predicted (Gershoni *et al.* 2009). Furthermore, TPI expression is upregulated in dwarf compared to normal whitefish at the juvenile stage (Nolte *et al.* 2009a) and downregulated at the adult stage (Derome *et al.* 2006). In this study, one SNP located in an intron of TPI (Fig. 6a) was associated with a condition factor phenotype (Fig. 6b). This SNP was also identified as an outlier (F_{ST} = 0.69) in Cliff Lake (Fig. 6c). Here, the genotype (CC) associated with a robust phenotype was more common in dwarf whitefish compared to the genotype (AC) associated with a slender phenotype and more common in normal whitefish (Fig. 6b, c). This counterintuitive result demonstrates again the complex relationship between selection and adaptation. Finally, previous studies identified an outlier AFLP marker with an F_{ST} value of 0.87, linked to an eQTL for TPI (Rogers & Bernatchez 2007; Derome *et al.* 2008). This eQTL for TPI was located within a regulatory hotspot comprising several genes involved in various functions. Therefore, the authors concluded that either the expression of TPI itself (*cis*-regulation) or a gene regulating TPI expression

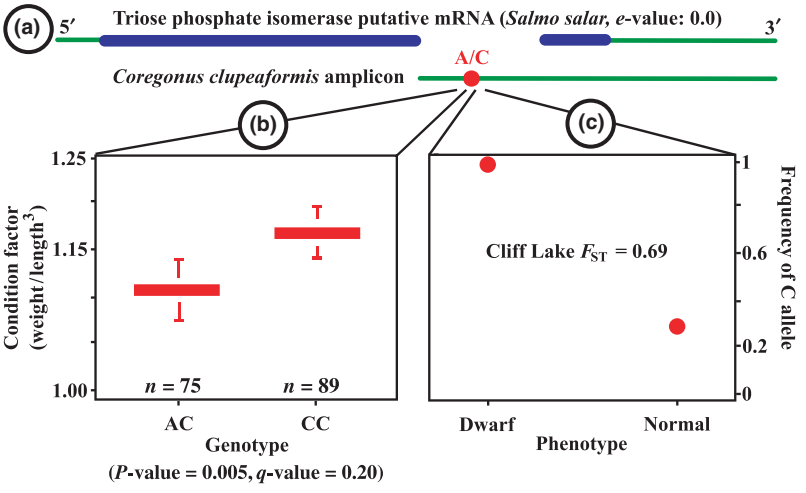

Fig. 6 (a) Genotypic characteristics of an SNP found in triosephosphate isomerase gene of whitefish (BLAST *e*-value = 0.0). The SNP is found in an intron (in green) according to *Salmo salar* open reading frame (in blue). (b) The SNP is associated with condition factor. (c) In natural populations, the SNP is an outlier in Cliff Lake.

(*trans*-regulation) was under divergent selection (Bernatchez *et al*. 2010).

In conclusion, the identification of several markers under divergent selection or linked to adaptive phenotypes suggests that many genes, in a lake-specific manner, are recruited by selection acting on standing genetic variation, during the adaptive divergence of lake whitefish. While we are accustomed to thinking of adaptive divergence and reproductive isolation being linked to single causal mutations (Colosimo *et al*. 2005; Hoekstra *et al*. 2006; Miller *et al*. 2007), this paradigm will probably soon shift towards a more intricate, yet more complete view of the genomics of speciation (Stern & Orgogozo 2008; Baxter *et al*. 2010; Counterman *et al*. 2010; Michel *et al*. 2010).

Acknowledgements

We thank V. Bourret, N. Isabel, B. Pelgas and C. Sauvage for insightful suggestions about the analysis, J. Laroche for help with bioinformatics and G. Côté with help in the laboratory. Special thanks to A. Belisle, A. Montpetit and P. Lepage at the Genome Quebec Innovation Center for genotyping troubleshooting. This research was funded by a Natural Science and Engineering Research Council of Canada (NSERC) and Canadian Research Chair in Genomics and Conservation of Aquatic Resources to LB, a NSERC postgraduate scholarship to SR, a postdoctoral research stipend from the DFG to AN and NSERC discovery grant to SMR. This study is a contribution to the research programme of Québec-Océan.

References

Baxter SW, Nadeau N, Maroja L *et al*. (2010) Genomic Hotspots for adaptation: population genetics of Mullerian mimicry in the *Heliconius melpomene* clade. *PLoS Genetics*, **6**, e794.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Bernatchez L (2004) Ecological theory of adaptive radiation: an empirical assessment from coregonine fishes (Salmoniformes). In:*Evolution Illuminated: Salmon and Their Relative* (eds Hendry AP, Stearns S), pp. 176–207. Oxford University Press, Oxford, UK.

Bernatchez L, Dodson JJ (1990) Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial DNA restriction analysis. *Evolution*, **44**, 1263–1271.

Bernatchez L, Chouinard A, Lu G (1999) Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, *Coregonus* sp., as a case study. *Biological Journal of the Linnean Society*, **68**, 173–194.

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

Butlin RK (2008) Population genomics and speciation. *Genetica*, **138**, 409–418.

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Castric V, Bonney F, Bernatchez L (2001) Landscape structure and hierarchical genetic diversity in the Brook Charr, *Salvelinus fontinalis*. *Evolution*, **55**, 1016–1028.

Colosimo PF, Hosemann KE, Balabhadra S *et al*. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.

Counterman BA, Araujo-Perez F, Hines HM *et al*. (2010) Genomic hotspots for adaptation: the population genetics of Mullerian mimicry in *Heliconius erato*. *PLoS Genetics*, **6**, e1000796.

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

Derome N, Bougas B, Rogers SM *et al*. (2008) Pervasive sex-linked effects on transcription regulation as revealed by eQTL mapping in lake whitefish species pairs (*Coregonus* sp, Salmonidae). *Genetics*, **179**, 1903–1917.

Dierenger D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Ewart HS, Klip A (1995) Hormonal regulation of the Na+-K+-ATPase: mechanisms underlying rapid and sustained changes in pump activity. *American Journal of Physiology. Cell Physiology*, **269**, C295.

Excoffier L, Foll M (2009) Detecting loci under selection in a hierarchically structured population Hierarchical test of selection. *Heredity*, **103**, 285–298.

Feder JL, Nosil P. (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.

Fenderson O (1964) Evidence of subpopulations of lake whitefish, Coregonus clupeaformis, involving a dwarfed form. *Transactions of the American Fisheries Society*, **93**, 77–94.

Foll M, Gaggiotti O (2008) Identifying the environmental factors that determine the genetic structure of Populations. *Genetics*, **174**, 875–891.

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

Gonzalez JR, Armengol L, Sole X *et al*. (2007) SNPassoc: an R package to perform whole genome association studies. *Bioinformatics*, **23**, 644–645.

Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Research*, **16**, 730–737.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101–104.

Kahilainen K, Lehtonen H (2002) Brown trout (*Salmo trutta* L.) and Arctic charr (*Salvelinus alpinus* (L.)) as predators on three sympatric whitefish (*Coregonus lavaretus* (L.)) forms in the subarctic Lake Muddusjärvi. *Ecology of Freshwater Fish*, **11**, 158–167.

Landry L, Bernatchez L (2010) Role of epibenthic resource opportunities in the parallel evolution of lake whitefish species pairs (*Coregonus* sp.). *Journal of Evolutionary Biology*, **23**, 2602–2613.

Landry L, Vincent WF, Bernatchez L (2007) Parallelism between limnological features and phenotypic evolution of lake whitefish dwarf ecotypes. *Journal of Evolutionary Biology*, **20**, 971–984.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Lodish H, Berk A, Kaiser CA *et al.* (2008). *Molecular and Cell Biology*, 6th edn. W.H. Freeman and Company, New York.

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

Lu G, Basley DJ, Bernatchez L (2001) Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, **10**, 965–985.

Magnan P (1988) Interactions between Brook charr, *Salvelinus fontinalis*, and non salmonid species-ecological shift, morphological shift, and their impact on zooplankton communities. *Canadian Journal of Fisheries and Aquatic Sciences*, **45**, 999–1099.

Mäkinen HS, Cano JM, Merilä J (2008) Identifying footprints of directional and balancing selection in marine and freshwater threespine stickleback (*Gasterosteus aculeatus*) populations. *Molecular Ecology*, **17**, 3565–3582.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, UK.

McCairns RJ, Bernatchez L (2010) Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution*, **64**, 1029–1047.

McPhail JD (1993) Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): origin of the species pairs. *Canadian Journal of Zoology*, **71**, 515–523.

Michel AP, Sima S, Powella THQ, Taylora MS, Nosil P, Feder JF (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences USA*, **107**, 9724–9729.

Miller CT, Beleza S, Pollen AA *et al.* (2007) cis-Regulatory changes in kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell*, **131**, 1179–1189.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Nolte AW, Renaut S, Bernatchez L (2009a) Divergence in gene regulation at young life history stages of whitefish (*Coregonus* sp.) and the emergence of genomic isolation. *BMC Evolutionary Biology*, **9**, 925–936.

Nolte AW, Gompert Z, Buerkle CA (2009b) Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.

Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. *Nature Reviews Genetics*, **7**, 851–861.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Pigeon D, Chouinard A, Bernatchez L (1997) Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution*, **51**, 196–205.

Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, **11**, 175–180.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Renaut S, Nolte AW, Bernatchez L (2009) Gene expression divergence and hybrid misexpression between Lake Whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Biology and Evolution*, **26**, 925–936.

Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology*, **19**(Suppl. 1), 115–131.

Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, **14**, 351–361.

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

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

Rogers SM, Isabel N, Bernatchez L (2007) Linkage maps of the dwarf and normal lake whitefish (*Coregonus clupeaformis*) species complex and their hybrids reveal the genetic architecture of population divergence. *Genetics*, **175**, 1–24.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, New York.

Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. *Science*, **323**, 727–741.

Stanek MT, Cooper TF, Lenski RE (2009) Identification and dynamics of a beneficial mutation in a long-term evolution experiment with *Escherichia coli*. *BMC Evolutionary Biology*, **9**, 302.

St-Cyr J, Derome N, Bernatchez L (2008) The transcriptomics of life-history trade-offs between whitefish species pairs (*Coregonus* sp.). *Molecular Ecology*, **17**, 1850–1870.

Steiner CC, Römpler H, Boettger LM, Schöneberg T, Hoekstra HE (2009) The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Molecular Biology and Evolution*, **26**, 35–45.

Stern DL, Orgogozo V (2008) The loci of evolution: how predictable is genetic evolution? *Evolution*, **62**, 2155–2177.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.

Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, **64**, 479–498.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Storz JF, Wheat CW (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution*, **64**, 2489–2509.

Svanbäck R, Persson L (2004) Individual diet specialization, niche width and population dynamics: implication for trophic polymorphisms. *Journal of Animal Ecology*, **73**, 973–982.

Trudel M, Tremblay A, Schetagne R, Rasmussen J (2001) Why are dwarf fish so small? An energetic analysis of polymorphism in lake whitefish (Coregonus clupeaformis). *Canadian Journal of Fisheries and Aquatic Science*, **58**, 394–405.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of USA*, **106**, 9939–9946.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

White BJ, Cheng C, Simard F, Costantini C, Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology*, **19**, 925–939.

Whiteley AR, Derome N, Rogers SM *et al.* (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.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Wu CI, Ting CT (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114–122.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, **175**, 1883–1893.

The authors are broadly interested in the nature of genetic changes that are associated with speciation. This study is part of SR's doctoral research in LB's laboratory, which aims at studying the genomic bases of adaptive divergence in the context of a recent ongoing speciation event in lake whitefish. AN is interested in the diversity of fishes and understanding the genetic basis of adaptation. SMR studies the evolutionary mechanisms for coping with environmental change by integrating ecological genomics and quantitative genetics with field studies of natural selection. ND and LB's research focuses on understanding the patterns and processes of molecular and organismal evolution as well as their significance to conservation.

Supporting information

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

Table S1 Summary of genotyping frequency, amplicon sequences and BLAST annotations for all 96 SNPs retained for analyses of genetic differentiation in natural populations of sympatric dwarf and normal whitefish

Table S2 Summary of genotyping frequency, amplicon sequences and BLAST annotations for all 87 SNPs retained for analyses of association with phenotypic traits

Table S3 Summary of observed and expected heterozygosities and F_{ST} for all 96 SNPs retained for analyses of genetic differentiation in natural populations

Table S4 Summary of linkage disequilibrium for all loci in the association family and loci identified as outliers in the genome scan

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.