

FAST-TRACK

Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*)

S. M. ROGERS and L. BERNATCHEZ

Département de biologie, Québec-Océan, Université Laval, Ste-Foy, Québec, G1K 7P4 Canada

Abstract

As natural selection must act on underlying genetic variation, discovering the number and location of loci under the influence of selection is imperative towards understanding adaptive divergence in evolving populations. Studies employing genome scans have hypothesized that the action of divergent selection should reduce gene flow at the genomic locations implicated in adaptation and speciation among natural populations, yet once 'outlier' patterns of variation have been identified the function and role of such loci needs to be confirmed. We integrated adaptive QTL mapping and genomic scans among diverging sympatric pairs of the lake whitefish (*Coregonus clupeaformis*) species complex in order to test the hypothesis that differentiation between dwarf and normal ecotypes at growth-associated QTL was maintained by directional selection. We found evidence of significantly high levels of molecular divergence among eight growth QTL where two of the strongest candidate loci under the influence of directional selection exhibited parallel reductions of gene flow over multiple populations.

Keywords: adaptive divergence, *Gasterosteus*, genome scan, Q_{ST} , QTL, selection, speciation

Received 8 August 2004; revision received 6 October 2004; accepted 6 October 2004

The role of natural selection in population divergence is fundamental towards a comprehensive understanding of biodiversity. Although directional selection is accepted as the predominant driving force of phenotypic diversification (Rieseberg *et al.* 2002), detecting the number and location of chromosomal regions under the influence of selection during adaptive divergence has been challenging (Black *et al.* 2001; Luikart *et al.* 2003). Implicit within this goal is an understanding of the genetic basis of adaptive phenotypic differentiation and the action of selection on this variation. Such knowledge will arguably become increasingly central towards our interpretations of population structure, speciation, and ultimately conservation.

Studies of genetic architecture (e.g. Via & Hawthorne 2002; Ungerer & Rieseberg 2003) and genome scans (e.g. Akey *et al.* 2002; Brumfield *et al.* 2003; Storz & Nachman 2003) have attempted to address these issues. By mapping the genome of a species through informative crosses, the number of loci associated with trait variation, i.e. QTL,

may be used to infer the genetic basis of adaptive traits underlying species differences. How selection has shaped genetic architecture has been increasingly documented in plants (e.g. Bradshaw *et al.* 1995; Rieseberg *et al.* 1996; Schemske & Bradshaw 1999; Juenger *et al.* 2000) yet rarely in animals (e.g. Nuzhdin *et al.* 1999; Slate *et al.* 2002). However, livestock and agricultural quantitative genetics have substantially contributed to our knowledge of genetic architecture along artificially selected lines (e.g. see Dekkers & Hospital 2002). Globally, it is evident that general patterns are more complicated than previously believed (Via & Hawthorne 2002; Pigliucci 2003), due partly to the unknown effect that selection may have on corresponding adaptive traits (Hoekstra & Nachman 2003) as well an inherent limit on how well genetic architecture of species' differences can predict the action of selection in nature (Rieseberg *et al.* 2003).

Genome scans partially circumvent these problems through the simultaneous study of a large number of loci in order to better understand the action of evolutionary forces on variation among populations at a locus-specific level (Black *et al.* 2001; Beaumont & Balding 2004). Such

Correspondence: Dr Louis Bernatchez. Fax: 418-656-2043; E-mail: Louis.Bernatchez@bio.ulaval.ca

studies hypothesize that the action of divergent selection should reduce gene flow at genomic regions implicated in adaptation and speciation. Few studies have attempted to characterize such loci among natural populations. Yet, empirical and simulation data strongly suggest that detecting loci subjected to directional selection is feasible on the basis of genetic differentiation estimates such as F_{ST} among loci (Beaumont & Nichols 1996; Wilding *et al.* 2001; Akey *et al.* 2002; Brumfield *et al.* 2003; Storz & Nachman 2003; Beaumont & Balding 2004; Campbell & Bernatchez 2004). In these cases, outlier loci exhibit patterns of differentiation significantly greater than expected under conditions of neutrality. However, genome scans must still confirm the function and role of outlier loci in order to demonstrate that the increased molecular divergence is due to selection (Luikart *et al.* 2003).

Previous efforts that have combined the study of genetic architecture with patterns of genetic diversity among natural populations have been prolific towards characterizing the genetic bases of population and species divergence (e.g. Rieseberg *et al.* 1999; Carney *et al.* 2000; Lexer *et al.* 2003; Shaw & Danley 2003; Streebman *et al.* 2003; Emelianov *et al.* 2004). A complementary approach combining the study of genetic architecture with genome scans in natural populations represents a potentially powerful tool to discover how the genetic bases of species differences at adaptive QTL are maintained by selection among natural populations. By identifying QTL linked to phenotype–environment associations, complementary genome scans employing QTL loci should be able to confirm which, if any, target QTL exhibit a significant reduction of gene flow among natural populations. Genome scans are also amenable to the possibility of identifying parallel trends of divergence through the analysis of multiple populations. Parallel evolution provides strong evidence for natural selection in trait evolution, as genetic drift is unlikely to produce predictable changes in independent lineages in correlation with the environment (Schluter & Nagel 1995).

The lake whitefish (*Coregonus clupeaformis*) species complex represents an ideal nonmodel organism to address these questions. Sympatric dwarf (limnetic) and normal (benthic) whitefish ecotypes coinhabit several lakes in Maine, USA, and Québec, Canada (Bernatchez & Dodson 1990). Dwarf ecotypes mature as early as one year and seldom exceed a body size of 20 cm and weight of 100 g while the normal ecotypes typically mature at two to five years of age and commonly exceed 40 cm and 1000 g. Under natural conditions, the food consumption rate of dwarf whitefish is 40–50% higher than that of normal ecotype, yet their respective conversion efficiency of these resources is two to three times lower than the normal ecotype (Trudel *et al.* 2001). These observations raised the hypothesis that competitive interactions within the derived dwarf ecotype may have led to a physiological cost associated with inhabiting

a limnetic environment (Bernatchez 2004). They also suggest that directional selection for growth is likely a major contributing factor underlying the genetic basis of ecotype divergence. Furthermore, a genome scan employing 440 AFLP loci has recently identified outlier loci among natural populations of four sympatric pairs of these ecotypes (Campbell & Bernatchez 2004). If the action of directional selection were responsible for the locus-specific reduction of gene flow observed at these loci, we would predict that the genetic basis underlying these loci should be associated with major adaptive traits such as selection for differential growth driving the divergence between ecotypes (Rogers *et al.* 2002; Bernatchez 2004).

In this paper we tested the hypothesis that differentiation between dwarf and normal ecotypes at growth-associated QTL has been maintained by directional selection. This was accomplished by first evaluating growth as a phenotype–environment association, detecting genetic associations with QTL mapping, screening natural populations for outlier levels of differentiation at QTL homologues, and finally assessing observed patterns of divergence for outliers over multiple environments.

Materials and Methods

Experimental families and growth phenotypes

The parental generation of dwarf and normal whitefish for laboratory crosses was sampled in 1996 from Témiscouata Lake (47°36' N, 68°45' W) and Aylmer Lake (45°50' N, 71°26' W), respectively. The F1 generation consisted of reciprocal pure dwarf, pure normal, and hybrid dwarf/normal crosses as previously described in Lu & Bernatchez (1998). Experimental groups were reared under identical temperature and photoperiod regimes that simulated natural lake conditions. Every six months between 1996 and 1999, a sample of F1 dwarf and normal fish (per group N age 0+ = 10, age 1+ = 20, age 2+ = 30) were anaesthetized with 0.001% Eugenol solution whereupon body weight was measured. During each measurement, biomass densities among tanks were equilibrated (≤ 20 g/L) and food rations were adjusted to maintain feeding at 0.8% body mass per day for all groups.

In November of 1999, a mapping family was generated by backcrossing an F1 hybrid female ($H_F D_M$) with an F1 dwarf male (denoted as $H_F D_M$) (Rogers *et al.* 2001). In September of 2001, $H_F D_M$ progeny were surgically implanted with passively integrated transponder (PIT) tags (Biomark, Inc.) for long-term individual identification while samples of adipose tissue were removed and stored in 95% EtOH for DNA extraction. A sample ($N = 50$) from this $H_F D_M$ family were weighed monthly starting in July of 2002 (age 2 years and over) until November of 2002 using the same protocol as used for the F1 measurements. In these fish, successive

body weight measurements permitted the calculation of individual absolute growth rate in order to determine which period yielded the highest rate of growth and thus likely represented the best candidate for QTL phenotypes.

Evaluating adaptive nature of growth differentiation

The overall hypothesis that differential growth between ecotypes is maintained by directional selection was first tested by comparing the extent of body weight differentiation (Q_{ST}) between the F1 pure dwarf and normal groups with that of neutral expectations (Spitze 1993). Q_{ST} was calculated as the proportion of among-population genetic variance in quantitative traits: $Q_{ST} = \sigma_{GB}^2 / (\sigma_{GB}^2 + 2\sigma_{GW}^2)$ where phenotypic variance was equal to twice the observational component of body weight variance for individuals within groups and used as a surrogate for $2\sigma_{GW}^2$ while body weight variance between groups was equated to the observational variance component for groups and used as a surrogate for σ_{GB}^2 . Variance components for body weight between F1 dwarf and normal groups were estimated from the samples measured at ages one, two, and three years such that overall 95% confidence intervals for Q_{ST} could be estimated. Genetic differentiation (θ) at neutral markers (for estimate of neutral expectation) between these dwarf and normal groups was assessed using six microsatellite loci as detailed in Rogers *et al.* (2002).

Genetic basis of growth differentiation

The AFLP mapping kit (Applied Biosystems, Inc.) was used according to the protocol of Vos *et al.* (1995) where 12 selective amplification primer combinations characterized informative loci in the $H_F D_M$ backcross family. Loci were scored using the software BINTHERE (Garnhart and Kocher, University of New Hampshire) that generates 1 bp size-specific bin spreadsheets allowing an objective scoring of presence/absence. An additional manual verification of all bins was performed in order to identify comigrating loci within the same bin as well as loci that were inadvertently of similar size to the 1 bp binning threshold. The average size and standard deviation of all loci was calculated to identify potential genotyping errors (denoted by elevated standard deviation) and to identify homologues between the mapping family and natural populations. AFLP locus notation consisted of the di-nucleotide extensions representative of their selective primer combination (*EcoRI* Axx: *MseI* Cxx) followed by the locus size in base pairs.

Sex-specific linkage groups and marker orders for 119 informative AFLP loci were previously determined in this $H_F D_M$ backcross family (Rogers *et al.* 2001) with MAPMAKER/EXP 3.0 (Lander *et al.* 1987). For QTL associations, we genotyped the 50 $H_F D_M$ individuals employed in the current growth experiments at these 119 loci in order to determine

loci potentially associated to growth rate. While this initial QTL sample is limited, previous studies attempting to detect associations between genotype and phenotype (Beldade *et al.* 2002; O'Malley *et al.* 2003; Schadt *et al.* 2003; Somorjai *et al.* 2003) suggest that given sufficient linkage between markers, such limited sampling would still be sufficient to detect QTL (Falconer & Mackay 1996), although certain QTL may go undetected while, the effect for those detected would certainly be overestimated (Beavis 1994). Map distances between loci for $H_F D_M$ QTL individuals were verified against the first map and in the event of significant changes between observed distances, markers were removed from the analysis until more samples could be added. Growth rate phenotypes were not normally distributed and were thus log-transformed. A maximum-likelihood interval analysis, performed in MAPMAKER/QTL 1.0 (Lander & Botstein 1989), estimated the likelihood of QTL from intervals spanning 2 cM along each linkage group between linked markers. As our objective was to identify markers flanking QTL among intervals for subsequent genome scans rather than focusing on the architecture of single QTL, occasional type-I errors were considered much less problematic than repeated type-II errors (Peripato *et al.* 2004; Verhoeven *et al.* 2004). We therefore employed a nominal significance threshold that attempted to maintain a 5% table-wide error rate considering the number of independent markers (M) analysed [$LOD = \log_{10}(1/(\alpha/M))$], corresponding to an LOD significance threshold of 2.89 (Lander & Botstein 1989). All loci flanking intervals that revealed QTL above this LOD threshold were selected as potential candidates for subsequent genome scans among the natural populations.

Genome scan of homologous growth QTL in natural populations

The possible role of selection driving observed patterns of divergence between dwarf and normal ecotypes was tested by comparing patterns of genetic differentiation (F_{ST}) for growth QTL in four sympatric pairs of whitefish. All sympatric pairs among lakes as well as the mapping family are known to share a recent, post-Wisconsinian glacial origin in a zone of secondary contact (Bernatchez *et al.* 1999). A total of 440 polymorphic AFLP were recently analysed in these four populations where it was established that sympatric pairs among lakes had established equilibrium and as a result it was possible to compare genome scan results among lakes (discussed in Campbell & Bernatchez 2004; but see extended methodology). Homology between QTL and loci among populations was inferred when loci from the same selective primers between samples yielded markers of the same molecular weight among populations, consistent with observations in other systems where sequencing AFLP markers of the same

molecular mass generated by the same primers among individuals were homologous with an error rate of 0.008 (Parsons & Shaw 2001; Emelianov *et al.* 2004).

Distribution of F_{ST} values for adaptive QTL outliers among sympatric pairs

Genetic differentiation estimates at homologous QTL loci between dwarf and normal sympatric pairs among the four lakes was estimated using the F_{ST} of Nei (1977) after Nei & Chesser (1983). The outcome of differential selection is to generate an excess of loci with high outlier F_{ST} values relative to neutrality (Akey *et al.* 2002; Luikart *et al.* 2003), thus distributions of F_{ST} were divided into 11 categories (from -0.1–1.0) where loci falling in the tail of the distribution (high F_{ST} values) were interpreted in terms of the action of directional selection. In order to target the most likely candidate QTL under the effect of directional selection, a threshold based on the expected level of differentiation under neutrality was defined. This threshold was established from simulations under a neutral model of evolution implementing an algorithm for dominant loci (Wilding *et al.* 2001), and modified for sympatric pairs of lake whitefish (Campbell & Bernatchez 2004). The 95% quantile of this expected distribution of differentiation under neutrality was employed as the threshold to infer QTL under selection (i.e. when F_{ST} estimates > 95% quantile). Thus, growth QTL were considered potentially under the effect of directional selection (outliers) when there was less than 5% chance of observing corresponding F_{ST} estimates under neutrality.

Given the frequency at which previous empirical estimates of 440 loci resulted in outlier loci among lakes in Campbell & Bernatchez (2004), we performed a χ^2 test assessing the probability of observing more QTL under the influence of selection than expected by chance alone. The expected probability of QTL outliers was calculated using the expected proportions over 440 loci in each lake (East Lake: nine observed out of 440 = 0.0205; Webster Lake: 11 observed out of 440 = 0.0250; Indian Lake: 19 out of 440 = 0.0432; Cliff Lake: 16 out of 440 = 0.0364). Expected numbers of outlier QTL loci per lake were therefore obtained by multiplying these probabilities by the absolute number of QTL outlier loci observed per lake.

Parallel trends in divergence among candidate QTL outliers

A χ^2 test was performed in order to assess if the number of QTL loci exhibiting parallel trends of divergence were observed more or less than expected under conditions of neutrality. A parallel trend in divergence for loci between ecotypes was inferred when the allelic frequency of presence (dominant allele) was consistently higher within

one ecotype among lakes. The probabilities of expected parallel and nonparallel trends under neutrality were previously estimated from 392 nonoutlier loci [expected probability of parallel divergence = 0.06, expected probability of nonparallel divergence = 0.94, see Campbell & Bernatchez (2004) for details or refer to extended methodology]. Expected numbers of parallel and nonparallel trends in the QTL loci (outlier and neutral) were thus obtained by multiplying these probabilities by the absolute number of QTL loci observed in each category.

An extended and more detailed version of the methodology applied in this study is available in the supplementary material for this paper (please refer to the end of the paper for the link).

Results

Adaptive growth differentiation between ecotypes

Significant differences in growth between dwarf and normal fish were observed for each age class. At age 1+, the dwarf ecotype (\bar{x} = 20.82 g, σ = 8.68) was significantly smaller (ANOVA, F = 7.13, P = 0.016) than the normal ecotype (\bar{x} = 33.32 g, σ = 12.12). These differences in size between ecotypes became more apparent by age 3+ (dwarf, \bar{x} = 180.1 g, σ = 63.7; normal, \bar{x} = 455.3 g, σ = 91.0), when even under the same environmental conditions the dwarf ecotype weighed on average 2.5 times less than the normal groups strongly suggesting a genetic basis for these differences (Fig. 1). Overall, these differences translated into higher

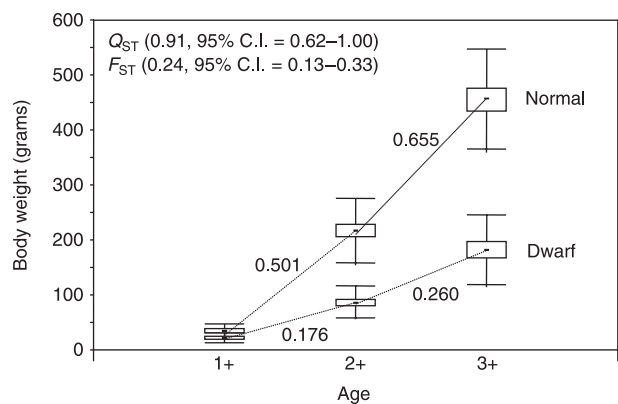


Fig. 1 Box plots (mean, standard error, standard deviation) of body weight estimated for F1 dwarf and normal groups reared under identical laboratory conditions from age 1+ to 3+ 0. Values along intersections between weight measurements represent average absolute growth rate during the respective growth period. Q_{ST} – F_{ST} relationship indicates that the proportion of among-population genetic variance as inferred by phenotypic growth (weight at age) is higher than expected under neutrally evolving expectations (inferred from F_{ST}) thus indirectly suggesting growth is under the influence of directional selection (see methods for details).

estimates of absolute growth rate in the normal (0.58 g/day) than the dwarf fish (0.22 g/day). The level of neutral genetic differentiation (θ) between these dwarf and normal ecotypes estimated from microsatellite loci was 0.24 (95% C.I. = 0.13–0.33). The corresponding Q_{ST} estimate for growth differentiation between ecotypes (0.91; 95% C.I. = 0.62–1.00) was significantly higher than F_{ST} .

In the $H_F D_M$ backcross, a significant change in weight was observed over the experimental period (ANOVA, $F = 8.31$, $P < 0.001$) with average weight of 182.35 g ($\sigma = 56.77$) at the end of the trial. Absolute growth rate was on average 0.37 g/day ($\sigma = 0.27$), with the highest growth rate and variance ($\bar{x} = 0.43$ g/day, $\sigma = 0.65$) observed between July and August, concordant with observations for whitefish in nature during this period. These July to August absolute growth rates were therefore representative of the best phenotypic estimates for subsequent QTL analysis.

Genetic basis of growth differentiation

Linkage analysis with QTL individuals was consistent with results obtained by Rogers *et al.* (2001). Out of 119 markers segregated into the first preliminary linkage map, 102 markers exhibited similar pairwise recombination frequencies resulting in only a slight change in average map distance between markers, 17.92 cM compared to 15.21 cM previously observed (Rogers *et al.* 2001). Cumulatively, the current linkage map resulted in 25 sex-specific linkage groups where distances should be seen as 'best estimates' until additional markers can be added to achieve total genome coverage over the 40 linkage groups present in lake whitefish. Exigent recombination frequencies measured in these individuals suggested that observed distances between markers were sufficient towards detection of QTL (Liu 1998). QTL analysis revealed significant associations to growth over 11 linkage groups in total (Table 1). LOD values were well above the threshold of 2.89 and ranged between 3.5 (Lg1) and 8.7 (Lg7 and Lg11). Overall, a total of 35 loci flanking significant QTL intervals were potentially useful for further genome scans among the natural populations (Table 1). Comparison of the size (bp) and standard deviations with the loci scored in natural populations by Campbell & Bernatchez (2004) revealed homology for 27 of these 35 loci (Table 1). In all cases, standard deviations of locus sizes were under 0.3 bp in both the backcross family and natural populations, suggesting limited size-shift migrations for AFLP loci. In cases where no homologous loci were found in the natural populations, genotypic data from the natural populations suggested that ambiguous comigrating loci of similar molecular weight forced the elimination of the loci from the analysis or the loci were monomorphic in one or more of the sympatric pairs.

Genome scan comparing outlier QTL against neutrality

Genome scans estimating locus-specific measures of differentiation between ecotypes for these 27 growth-QTL among four lakes identified eight QTL with F_{ST} values higher than the 95% quantile, which were therefore considered as candidates under the effect of directional selection (Table 2). Globally, this proportion of outlier QTL (8 out of 27, 30%) among lakes was significantly higher ($P < 0.0001$) than an overall expected proportion of 11% of loci under the influence of directional selection by chance alone based on previous observations for AFLP loci (48 outlier loci out of 440, Campbell & Bernatchez 2004). These outlier QTL were not randomly distributed on the map, with six of the eight outliers segregating to only three linkage groups (Table 2).

Distribution of F_{ST} values for adaptive QTL outliers among sympatric pairs

In East and Webster lake, only 3.7% of QTL exhibited outlier F_{ST} values which was not significantly different from neutral expectation (Fig. 2a–b). In contrast, both Indian and Cliff lakes revealed a significant skew of F_{ST} outlier distributions, with 18.5% (five loci observed vs. one expected, $P = 0.0003$) and 14.8% (four loci observed vs. one expected, $P = 0.0019$) of growth QTL under the effect of directional selection. Within these two lakes there were therefore more outlier QTL than expected by chance alone.

Parallel trends in divergence among adaptive QTL outliers

Among QTL with F_{ST} values lower than the (presumably neutral) 95% quantile, the number of loci exhibiting parallel divergence was not significantly different than expected under neutrality (two QTL out of 17 vs. 1.1 expected, $P = 0.375$). Conversely, for QTL with F_{ST} values higher than the 95% quantile, significantly more loci exhibited parallel trends of divergence among lakes than expected under neutrality (two QTL out of eight vs. 0.46 expected, $P = 0.0198$) (Table 2). These loci therefore represented the best candidate QTL loci linked to genes associated with differential growth and under the influence of directional selection among sympatric pairs inhabiting different lakes. Of these two, the strongest candidate growth QTL for parallel selection was CATA-104 which was an outlier in Cliff Lake ($F_{ST} = 0.617$), and was the only QTL possibly under the influence of selection in both East ($F_{ST} = 0.450$) and Webster ($F_{ST} = 0.369$) lakes (Table 2).

Discussion

Our aim was to test the hypothesis that differentiation between dwarf and normal ecotypes at growth-associated

Marker interval	LG	LOD	Mapping family		Natural populations	
			Size (bp)	SD	Size (bp)	SD
GGTG199	1	4.8	200.18	0.18	200.29	0.07
GGTG061			61.21	0.15	61.74	0.04
GGTG104	1	3.5	105.03	0.17	105.06	0.11
GGTG107			106.55	0.28	106.39	0.14
—	1	3.5				
GGTG148			148.57	0.04	148.64	0.06
CATA108	4	5.8	108.85	0.07	108.87	0.05
CATA104			104.48	0.07	104.50	0.06
—	4	7.8				
CATA144			144.55	0.05	144.58	0.03
—	4	7.7				
CATA170			171.03	0.04	171.07	0.03
—	4	5.5				
CATA119			120.31	0.05	120.26	0.06
CGTC137	5	5.3	137.36	0.08	137.39	0.04
CGTC123			123.63	0.26	123.71	0.25
ACTA135	6	5.4	135.82	0.07	n/a	
CTTC220			220.54	0.07	n/a	
CATA131	7	4.8	131.87	0.04	131.96	0.22
CATA142			143.25	0.04	143.19	0.06
—	7	5				
CATA120			121.37	0.05	121.30	0.06
—	7	5.6				
CATA082			82.73	0.22	n/a	
—	7	8.6				
CATA083			83.50	0.24	83.64	0.21
—	7	8.7				
CATA061			61.87	0.29	61.94	0.29
GGTG189	8	6.4	189.71	0.06	189.66	0.06
CCTC122			122.86	0.10	122.85	0.06
—	8	4.4				
CGTC060			61.84	0.25	61.67	0.07
CATA135	11	8.7	136.37	0.05	136.32	0.15
CATA087			87.79	0.09	n/a	
—	11	8.7				
CATA062			62.70	0.15	62.76	0.19
AGAC126	12	3.4	127.13	0.05	n/a	
AGAC175			176.04	0.05	n/a	
CATA126	13	7.2	127.42	0.05	127.41	0.04
CATA084			84.57	0.14	84.52	0.05
—	13	5.3				
CATA086			86.46	0.27	n/a	
CAAG176	24	8.1	176.76	0.05	176.65	0.24
GGTG119			120.29	0.41	120.03	0.21
CAAG053	26	7.1	53.55	0.21	53.23	0.17
CATA057			58.54	0.08	n/a	
Number of QTL loci			35		27	

Table 1 Summary of QTL interval analysis for loci that revealed significant associations (log-likelihood denoted by LOD) to the absolute growth rate phenotype to linkage groups (LG). Locus size and standard deviation illustrate potentially homologous QTL loci between the mapping family and natural populations where n/a infers loci were either monomorphic or not scored in natural populations

QTL has been maintained by directional selection among natural populations. This objective consisted of evaluating growth as a phenotype–environment association, determining its genetic basis with QTL mapping,

screening natural populations for outlier levels of differentiation at QTL homologues, and finally assessing observed patterns of divergence for outliers over multiple environments.

Locus	LG	F_{ST}			
		East $Q_{95} = 0.256$	Webster $Q_{95} = 0.362$	Indian $Q_{95} = 0.196$	Cliff $Q_{95} = 0.456$
CATA-104**	4	0.450*	0.369*	0.032	0.617*
CATA-108	4	< 0.001	0.248	0.199*	< 0.001
CCTC-122	8	< 0.001	0.005	0.043	0.595*
CGTC-060**	8	0.045	0.044	0.229*	0.865*
CATA-120	7	0.094	0.201	0.258*	< 0.001
CATA-142	7	< 0.001	0.225	0.271*	0.184
GGTG-104	1	< 0.001	0.162	< 0.001	0.865*
CATA-135	11	0.079	0.248	0.339*	0.103

Table 2 Summary of loci associated to growth rate QTL exhibiting significant outlier levels of genetic differentiation ($> Q_{95}$) between sympatric pairs among four lakes. LG; refers to linkage group in mapping family

*Outlier: estimate of level of differentiation (F_{ST}) was higher than expected under neutrality ($> 95\%$ Quantile), therefore possibly under the effect of directional selection.

**Locus exhibited parallel pattern of divergence inferred from allelic frequency differentials between ecotypes among all lakes tested.

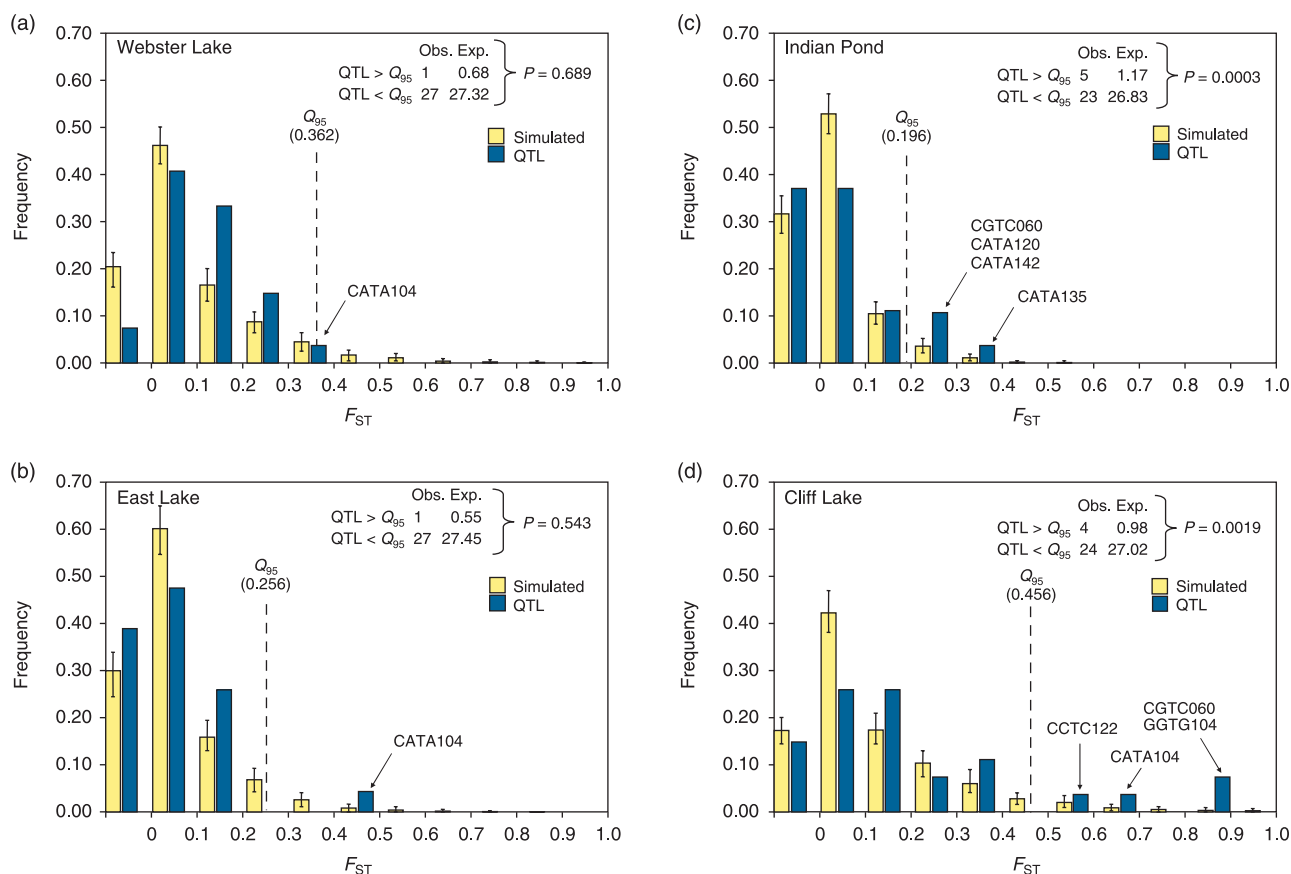


Fig. 2 Genome scans of differential gene exchange for growth associated QTL (blue bars) among four natural sympatric pairs of dwarf and normal lake whitefish (a) Webster Lake (b) East Lake (c) Indian Pond, and (d) Cliff Lake. Distributions represent the categorical frequency of F_{ST} values. Growth QTL were compared to the distribution of differentiation under neutrality (yellow bars, see Campbell & Bernatchez 2004 for details on simulated distributions). The 95% quantile of simulated distributions delineates outlier growth QTL potentially under directional selection (i.e. $F_{ST} > 95\%$ quantile). P -values were generated from a χ^2 test that tested the hypothesis that QTL loci associated to growth were observed more or less often than expected by chance (see methods).

First, highly significant differences in body size between ecotypes reared under laboratory conditions were indicative of a genetic basis for this trait, which translated into a higher level of quantitative divergence than expected under neutrality. Under a framework of ecological divergence (Schluter 2000), these results provided a priori evidence that observed growth differences between dwarf and normal whitefish ecotypes fulfil the adaptive criterion and have been driven by the effect of directional selection. Second, exploring the genetic basis of this differentiation identified 27 loci possibly associated to growth that were homologous to loci among natural populations (Table 1), although we warn that substantially more individuals will be needed in order to satisfy interpretations of genetic architecture. Third, a genome scan screening natural populations identified eight adaptive QTL exhibiting a more pronounced reduction in gene flow between whitefish ecotypes than expected under neutrality, a number that was higher than expected by chance alone.

By encompassing four independent sympatric pairs, parallel patterns of divergence for two QTL under directional selection (CATA-104 and CGTC-60, see Fig. 2 and Table 2) provided further evidence for the action of directional selection in restricting gene flow at these loci. These results are concordant with recent studies in sticklebacks where the genetic basis of parallel evolution was inferred from biased productions of genetic variance in close relatives (Schluter *et al.* 2004), and where trait evolution appears to be under the control of the same genes among independent populations (Colosimo *et al.* 2004; Cresko *et al.* 2004; Shapiro *et al.* 2004).

The observation that only eight (and perhaps as little as two) out of the 27 QTL were identified as potentially under the effect of directional selection does not strictly rule out the possibility that other growth QTL may be implicated in the adaptive divergence but were not detected in QTL analysis. A relatively sparse map covering only 62.5% of presumed extant linkage groups (25 out of 40) was used with limited samples. These results suggest, however, that the current map includes the major linkage groups associated with growth differentiation between ecotypes. Within these regions, gene flow has likely been constrained by selection against less 'fit' hybrid recombinant genotypes at as little as 2% of all mapped loci (potentially representative of only 1.2% of the genome once all 40 linkage groups are accounted for). This suggests that only a small proportion of genes might have been under the effect of directional selection during population divergence of lake whitefish, consistent with previous empirical genome scan studies implicating few loci (Wilding *et al.* 2001; Akey *et al.* 2002; Vigouroux *et al.* 2002; Campbell & Bernatchez 2004). Admittedly, however, it also begs for explanations as to why most adaptive QTL exhibit neutral levels of divergence among populations.

From the perspective of limits associated with genetic architecture, adaptive growth differences are clearly polygenic, whereby it is possible that selection on growth may be 'diluted' over many loci such that most QTL appear to behave neutrally (Lande 1976; McKay & Latta 2002). Adaptive QTL differentiation may therefore be constrained in the face of diversifying selection on trait values (Latta 1998), potentially due to the influence of epistatic interactions which may have reduced the resolution of association between the population structure of the QTL and the genetic basis of the adaptive traits (Fenster *et al.* 1997; McKay & Latta 2002). Moreover, QTL detection is unavoidably dependent on alleles with significantly different trait effects segregating in the mapping family. Thus, differential effects of QTL may be observed across genetic backgrounds even in the absence of selection (Tanksley & Hewitt 1988; Leips & Mackay 2000). However, recent evidence in *Arabidopsis thaliana* found that alleles for fitness-related traits still responded similarly to selection despite different genetic backgrounds (Ungerer & Rieseberg 2003). Overall, genetic background may limit the resolution of detecting selection among QTL, whereby informative parents used to generate variation within crosses are assumed to represent the variation that would be observed under natural conditions.

From an ecological perspective, there may be confounding demographic, spatial, or local effects on adaptive divergence and, consequently, F_{ST} among environments (Storz & Nachman 2003; Beaumont & Balding 2004). It is equally possible that selection for similar adaptive traits across environments may not necessarily implicate the same genetic basis. For example, novel polymorphisms within New Mexico pocket mice (*Chaetodipus intermedius*) melanism genes (*Mc1r*) showed no association to coat colour unlike their Arizona counterparts indicating that similar phenotypic change arose more than once, yet had a different genetic basis (Hoekstra & Nachman 2003). As selection acts indirectly on genes through phenotypic expression, parallel evolution of adaptive phenotypes may not necessarily be accompanied by parallel evolution of a given genotype. This exemplifies the importance of testing the action of selection on adaptive divergence across multiple environments.

Despite these limits, our results demonstrate the importance of addressing how selection operates on the genomic basis of phenotypic variation across environments once candidate QTL in natural populations have been identified. Such inference emphasizes the importance that ecological factors likely have towards shaping selection regimes during evolutionary divergence. For instance, sympatric pairs of lake whitefish vary with respect to the level of genetic and phenotypic divergence among lakes, strongly suggesting a role for the extent of niche segregation within respective limnetic and benthic habitats (Bernatchez *et al.* 1999). Our results revealed that gene flow for the two outlier

growth QTL most likely under the effect of parallel selection among sympatric pairs (CATA-104 and CGTC-60) was more restricted in lakes where morphological differentiation is more pronounced (Indian and Cliff) relative to the other two lakes (East and Webster) (Lu & Bernatchez 1999). This raises the hypothesis that locus-specific outcomes of selection on genetic architecture may be in part predictable. In this respect, subsequent studies will be needed to better characterize candidate QTL architecture towards gaining a better understanding of the action of selection and its relationship to the patterns of genetic heterogeneity across environments (Lexer *et al.* 2003).

In summary, characterizing the effect of selection at growth QTL among sympatric pairs of the lake whitefish species complex resulted in a better understanding of genetic and phenotypic basis of adaptive variation among lake whitefish populations. This reinforces the potential of a complementary QTL/genome scan approach towards studies of adaptive divergence, consistent with previous studies where complementary approaches have led to better interpretations of the evolutionary processes shaping population divergence (e.g. Rieseberg *et al.* 1999; Carney *et al.* 2000; Lexer *et al.* 2003; Strelman *et al.* 2003; Emelianov *et al.* 2004; Lexer *et al.* 2004). Embracing a genomic perspective will equally improve inferences into the understanding of adaptive evolution and population genetics, as even the initial identification of outlier loci represents a substantial gain towards understanding numerous facets of evolutionary history for a given species (Black *et al.* 2001; Wilding *et al.* 2001; Luikart *et al.* 2003).

Acknowledgements

We would like to thank S. Higgins and LARSA personnel for their whitefish rearing assistance. We are truly indebted to R. St-Laurent and D. Basley for their enormous field support and G. Côté for assisting with the whitefish experiments. We are grateful to D. Campbell, S. McCairns, D. Veliz, Bob Wayne and two anonymous reviewers for their insight into the manuscript. S. Rogers' research was supported by a Richard-Bernard and Société Provancher scholarship. L. Bernatchez's research program on the evolution and conservation of northern fishes is supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada, as well as by the Canadian Research Chair for in conservation genetics of aquatic resources.

References

Akey J, Zhang G, Zhang K, Jin L, Shriver M (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, **12**, 1805–1814.

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 genetic analysis of population structure. *Proceedings of the Royal Society London B*, **263**, 1619–1626.

Beavis WD (1994) *The power and deceit of QTL: experiments: lessons from comparative QTL studies. Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference*. 250–266. American Seed, Trade Association, Washington, D.C.

Beldade P, Brakefield PM, Long AD (2002) Contribution of Distal-less to quantitative variation in butterfly eyespots. *Nature*, **415**, 315–318.

Bernatchez L (2004) Ecological Theory of Adaptive Radiation: An empirical assessment from Coriginine fishes (Salmoniformes). In: *Evolution Illuminated: Salmon and Their Relatives* (eds Hendry AP, Stearns S), pp. 176–207. Oxford University Press, Oxford, U.K.

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

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.

Black W, Baer C, Antolin M, DuTeau N (2001) Population genomics: Genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441–469.

Bradshaw HD, Wilbert SM, Otto KG, Schemske DW (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature*, **376**, 762–765.

Brumfield RT, Beerli P, Nickerson DA, VES (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, **18**, 249–256.

Campbell D, Bernatchez L (2004) Genomic 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.

Carney SE, Gardner KA, Rieseberg LH (2000) Evolutionary changes over the fifty-year history of a hybrid population of sunflowers (*Helianthus*). *Evolution*, **54**, 462–474.

Colosimo PF, Peichel CL, Nereng K *et al.* (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLOS Biology. Public Library of Science*, **2**, e109.

Cresko WA, Amores A, Wilson C *et al.* (2004) Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Science of USA*, **101**, 6050–6055.

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics*, **3**, 22–32.

Emelianov I, Marec F, Mallet J (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society of London Series B-Biology Sciences*, **271**, 97–105.

Falconer DS, Mackay TF (1996) *Introduction to Quantitative Genetics*, 4th edn. Pearson Education Limited, Essex, England.

Fenster CB, Galloway LF, Chao L (1997) Epistasis and its consequences for the evolution of natural populations. *Trends in Ecology and Evolution*, **12**, 282–286.

Hoekstra HE, Nachman MW (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology*, **12**, 1185–1194.

Juenger T, Purugganan M, Mackay TFC (2000) Quantitative trait loci for floral morphology in *Arabidopsis thaliana*. *Genetics*, **156**, 1379–1392.

Lande R (1976) The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical Research*, **26**, 221–235.

- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
- Lander ES, Green P, Abrahamson J *et al.* (1987) MAPMAKER, an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174–181.
- Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *American Naturalist*, **151**, 283–292.
- Leips J, Mackay TFC (2000) Quantitative trait loci for life span in *Drosophila melanogaster*: Interactions with genetic background and larval density. *Genetics*, **155**, 1773–1788.
- Lexer C, Lai Z, Rieseberg LH (2004) Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytologist*, **161**, 225–233.
- Lexer C, Randell RA, Rieseberg LH (2003) Experimental hybridization as a tool for studying selection in the wild. *Ecology*, **84**, 1688–1699.
- Liu BH (1998) *Statistical Genomics: Linkage, Mapping, and QTL Analysis*. CRC Press, New York.
- Lu G, Bernatchez L (1998) Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of lake whitefish (*Coregonus clupeaformis* Mitchell). *Proceedings of the Royal Society of London Series B-Biology Sciences*, **265**, 1025–1030.
- 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.
- Luikart G, England PR, Tallman D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–993.
- McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology and Evolution*, **17**, 285–291.
- Nei M (1977) *F*-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics*, **41**, 225–233.
- Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**, 253–259.
- Nuzhdin SV, Dilda CL, Mackay TF (1999) The genetic architecture of selection response. Inferences from fine-scale mapping of bristle number quantitative trait loci in *Drosophila melanogaster*. *Genetics*, **153**, 1317–1371.
- O'Malley KG, Sakamoto T, Danzmann RG, Ferguson MM (2003) Quantitative trait loci for spawning date and body weight in rainbow trout: Testing for conserved effects across ancestrally duplicated chromosomes. *Journal of Heredity*, **94**, 273–284.
- Parsons YM, Shaw KL (2001) Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphisms. *Molecular Ecology*, **10**, 1765–1772.
- Peripato AC, De Brito RA, Matioli SR *et al.* (2004) Epistasis affecting litter size in mice. *Journal of Evolutionary Biology*, **17**, 593–602.
- Pigliucci M (2003) Selection in a model system: Ecological genetics of flowering time in *Arabidopsis thaliana*. *Ecology*, **84**, 1700–1712.
- Rieseberg LH, Church SA, Morjan CL (2003) Integration of populations and differentiation of species. *New Phytologist*, **161**, 59–69.
- Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM (1996) Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science*, **272**, 741–745.
- Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.
- Rieseberg LH, Widmer A, Arntz AM, Burke J (2002) Directional selection is the primary cause of phenotypic diversification. *Proceeding of the National Academy of Science of USA*, **99**, 12242–12245.
- Rogers SM, Campbell D, Baird SJE, Danzmann RG, Bernatchez L (2001) Combining the analyses of introgressive hybridisation and linkage mapping to investigate the genetic architecture of population divergence in the lake whitefish (*Coregonus clupeaformis*, Mitchell). *Genetica*, **111**, 25–41.
- Rogers S, Gagnon V, Bernatchez L (2002) Genetically based phenotype–environment association for swimming behavior in lake whitefish ecotypes (*Coregonus clupeaformis*, Mitchell). *Evolution*, **56**, 2322–2329.
- Schadt EE, Monks SA, Drake TA *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297–302.
- Schemske DW, Bradshaw HD (1999) Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proceedings of the National Academy of Science of USA*, **96**, 11910–11915.
- Schluter D, Nagel LM (1995) Parallel speciation by natural selection. *American Naturalist*, **146**, 292–301.
- Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, New York.
- Schluter D, Clifford EA, Nemethy M, McKinnon JS (2004) Parallel evolution and inheritance of quantitative traits. *American Naturalist*, **163**, 809–822.
- Shapiro MD, Marks ME, Peichel CL *et al.* (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717–723.
- Shaw KL, Danley PD (2003) Behavioral genomics and the study of speciation at a porous species boundary. *Zoology*, **106**, 261–273.
- Slate J, Visscher PM, MacGregor S *et al.* (2002) A genome scan for quantitative trait loci in a wild population of red deer (*Cervus elaphus*). *Genetics*, **162**, 1863–1873.
- Somorjai IML, Danzmann RG, Ferguson MM (2003) Distribution of temperature tolerance quantitative trait loci in Arctic charr (*Salvelinus alpinus*) and inferred homologies in rainbow trout (*Oncorhynchus mykiss*). *Genetics*, **165**, 1443–1456.
- Spitze K (1993) Population Structure in *Daphnia obtusa*: Quantitative Genetic and Allozymic Variation. *Genetics*, **135**, 367–374.
- Storz JF, Nachman MW (2003) Natural selection on protein polymorphism in the rodent genus *Peromyscus*: Evidence from interlocus contrasts. *Evolution*, **57**, 2628–2635.
- Streelman JT, Albertson RC, Kocher TD (2003) Genome mapping of the orange blotch colour pattern in cichlid fishes. *Molecular Ecology*, **12**, 2465–2471.
- Tanksley SD, Hewitt J (1988) Use of Molecular Markers in Breeding for Soluble Solids Content in Tomato — a Re-Examination. *Theoretical and Applied Genetics*, **75**, 811–823.
- 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 Sciences*, **58**, 394–405.
- Ungerer MC, Rieseberg LH (2003) Genetic architecture of a selection response in *Arabidopsis thaliana*. *Evolution*, **57**, 2531–2539.
- Verhoeven KJF, Vanhala TK, Biere A, Nevo E, van Damme JMM (2004) The genetic basis of adaptive population differentiation:

- a quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. *Evolution*, **58**, 270–283.
- Via S, Hawthorne DJ (2002) The genetic architecture of ecological specialization: Correlated gene effects on host use and habitat choice in pea aphids. *American Naturalist*, **159**, S76–S88.
- Vigouroux Y, McMullen M, Hittinger CT *et al.* (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 9650–9655.
- Vos P, Hogers R, Bleeker M. *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research*, **23**, 4407–4414.
- Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.
-
- This study is part of S. Rogers' doctoral research in the Bernatchez laboratory that aims to understand the genetic basis of adaptation and the reproductive isolation in the lake whitefish. Both authors share interests in the comprehension of genotype-environment interactions, particularly in cases of populations that have diverged to exploit distinct ecological resources.
-

