

*Review*

# On the origin of species: insights from the ecological genomics of lake whitefish

Louis Bernatchez<sup>1,\*</sup>, Sébastien Renaut<sup>1</sup>, Andrew R. Whiteley<sup>1,2</sup>,  
Nicolas Derome<sup>1</sup>, Julie Jeukens<sup>1</sup>, Lysandre Landry<sup>1,3</sup>,  
Guoqing Lu<sup>1,4</sup>, Arne W. Nolte<sup>1,5</sup>, Kjartan Østbye<sup>1,6</sup>,  
Sean M. Rogers<sup>1,7</sup> and Jérôme St-Cyr<sup>1</sup>

<sup>1</sup>*Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada G1V 0A6*  
<sup>2</sup>*Department of Natural Resources Conservation, University of Massachusetts, Amherst, MA 01003, USA*  
<sup>3</sup>*ISMER, 310, allée des Ursulines, C.P. 3300, Rimouski, Québec, Canada G5L 3A1*  
<sup>4</sup>*Department of Biology, University of Nebraska at Omaha, Omaha, NE 68182, USA*  
<sup>5</sup>*Max Planck Institute for Evolutionary Biology, August-Thienemann-Str. 2, 24306 Plön, Germany*  
<sup>6</sup>*Department of Biology, University of Oslo, Postboks 1066 Blindern, 0316 Oslo, Norway*  
<sup>7</sup>*Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4*

In contrast to the large amount of ecological information supporting the role of natural selection as a main cause of population divergence and speciation, an understanding of the genomic basis underlying those processes is in its infancy. In this paper, we review the main findings of a long-term research programme that we have been conducting on the ecological genomics of sympatric forms of whitefish (*Coregonus* spp.) engaged in the process of speciation. We present this system as an example of how applying a combination of approaches under the conceptual framework of the theory of adaptive radiation has yielded substantial insight into evolutionary processes in a non-model species. We also discuss how the joint use of recent biotechnological developments will provide a powerful means to address issues raised by observations made to date. Namely, we present data illustrating the potential offered by combining next generation sequencing technologies with other genomic approaches to reveal the genomic bases of adaptive divergence and reproductive isolation. Given increasing access to these new genomic tools, we argue that non-model species studied in their ecological context such as whitefish will play an increasingly important role in generalizing knowledge of speciation.

**Keywords:** speciation; ecological genomics; 454 sequencing; microarrays; selection; stickleback

## 1. INTRODUCTION

One hundred and fifty years after the publication of *The Origin of Species* (Darwin 1859), elucidating the processes of population divergence and species diversification remains a central tenet of evolutionary biology. It is also increasingly recognized that understanding the evolutionary mechanisms of adaptive divergence is crucial to predicting how organisms will respond to global crises such as climate change, habitat loss, resistance to antibiotics or the spread of invasive species, to name a few (Bernatchez & Tseng 2008). As for any scientific discipline, progress in this field will be best achieved if studies are embedded within a strong, predictive, theoretical framework. The theory of adaptive radiation relates directly to Darwin's (1859) vision on the role of natural selection and

adaptation driven by ecological processes in the course of diversification: it holds that both phenotypic divergence and ultimately speciation are the outcomes of divergent natural selection stemming from environment heterogeneity and competitive interactions. In revising this theory in the light of studies that have been conducted since its formulation, Schluter (2000) showed that it remains one of the most successful conceptual frameworks ever advanced in evolutionary biology. Studies that have accumulated since 2000 also convincingly demonstrated that knowledge on the ecology of adaptive radiation has made substantial leaps and allowed identification of the most pressing research areas.

In contrast to the large amount of ecological information supporting the role of divergent natural selection as a main cause of population divergence and speciation, a thorough understanding of the genomic basis underlying these evolutionary processes remains in its infancy (Mitchell-Olds *et al.* 2007; Schluter 2009; Via 2009) being limited to traditional model species for which knowledge of their ecology

\* Author for correspondence (louis.bernatchez@bio.ulaval.ca).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2009.0274> or via <http://rstb.royalsocietypublishing.org>.

One contribution of 11 to a Theme Issue 'Genomics of speciation'.

is generally very limited (e.g. *Drosophila* species, see Noor & Feder 2006). Over the last few years, however, identifying genes and QTL (quantitative trait loci) implicated in adaptive divergence and reproductive isolation has become one of the most active research areas in evolutionary biology. Remarkable progress has been accomplished in an increasing number of organisms, both in plants (e.g. Schemske & Bierzychudek 2007; Kane *et al.* 2009) and animals (e.g. Hoekstra *et al.* 2006; Joron *et al.* 2006; Storz *et al.* 2007; Barrett *et al.* 2008; other papers in this issue). Yet, identification of causative genetic polymorphisms underlying phenotypic divergence and elucidating the effect of natural selection on the molecular genetic architecture of adaptive traits and reproductive isolation remains a daunting task, particularly in organisms lacking extensive sequence information and other genomic resources.

Combining different research approaches and targeting various functional and biological levels (e.g. variation at the DNA, gene expression and phenotypic levels) represents the best strategy for deciphering the genetic basis of evolutionary change and diversification driven by natural selection (Vasemagi & Primmer 2005; Stinchcombe & Hoekstra 2008). Moreover, progress towards this goal will be best accomplished by the comparative study of evolutionarily young and ecologically distinct lineages. Thus, genetic changes contributing to early steps of adaptive divergence and reproductive isolation can be studied before they become confounded by additional genetic differences between species that accumulate after speciation is complete and which erase traces of genomic islands of speciation (Schluter 2000; Via 2009).

In this paper, we review the main findings of a long-term research programme that we have been conducting on the ecological genomics of sympatric species pairs of whitefish (*Coregonus* spp.). We present this system as an example of how applying a combination of multiple approaches under the conceptual framework of the theory of adaptive radiation has yielded substantial insight into the genetic basis and evolutionary processes of divergence and reproductive isolation in a non-model species for which limited genomic resources are available. To this end, we adopted a strategy that consists of four main steps:

- Identifying phenotypic traits most likely to be adaptive.
- Elucidating the genetic bases of these phenotypes.
- Finding evidence for natural selection at the genome level in the wild.
- Identify mechanisms of reproductive isolation and elucidating their molecular basis.

## 2. WHITEFISH BIOLOGICAL ATTRIBUTES AND BACKGROUND ON EVOLUTIONARY HISTORY AND ECOLOGY

Many biological attributes make the genus *Coregonus* highly relevant for studying the genomic basis of adaptive divergence and reproductive isolation in an ecological context. It is the most speciose genus within the family Salmonidae, with a conservative

estimate of 28 recognized taxa distributed throughout the Northern Hemisphere (Reshetnikov 1988) and the actual number of biological species could vastly exceed this figure (Kottelat & Freyhof 2007). Since the Pleistocene, members of this genus have evolved to occupy a wide array of habitats which lead to substantial phenotypic variation both among taxonomically recognized species and among populations within taxa. Of particular relevance is the occurrence both in North America (taxon *C. clupeaformis*) and Eurasia (*C. lavaretus*) of lacustrine forms of whitefish that live in sympatry. Phylogeographic studies confirmed the young age of these sympatric forms that evolved post-glacially, less than 15 000 yr BP (Pigeon *et al.* 1997; Østbye *et al.* 2006) and population genetics surveys showed that they remain reproductively isolated despite ongoing gene flow and hybridization (Lu & Bernatchez 1999; Lu *et al.* 2001; Østbye *et al.* 2005a). Moreover, the life history and ecology of sympatric whitefish populations have been well documented (e.g. Lu & Bernatchez 1999; Amundsen *et al.* 2004; Harrod & Kahilainen 2006; Hudson *et al.* 2007; Kahilainen *et al.* 2007) such that the genomics of phenotypic divergence and reproductive isolation can be interpreted in a natural, ecological context (reviewed in Bernatchez 2004; Østbye *et al.* 2005b; Landry *et al.* 2007a; Vonlanthen *et al.* 2009).

The natural setting that we most extensively investigated is located in the St. John River drainage (northeastern US and southeast Quebec) where sympatric forms of whitefish have been reported in six lakes. Geographical isolation during the Pleistocene caused genetic divergence between whitefish populations inhabiting distinct glacial refugia east and west of this area, but without distinctive phenotypic divergence between glacial races in allopatry (Bernatchez & Dodson 1990, 1991). Secondary contact of these evolutionary lineages subsequently occurred 12 000 yr BP within these lakes. Both ecological opportunity and competitive interactions translating into character displacement have contributed to the evolution of a limnetic dwarf form, which has diverged in sympatry from the benthic normal form (Bernatchez 2004; Landry *et al.* 2007a). The weight of evidence thus indicates that some accumulation of genetic differences during allopatric geographical isolation as well as subsequent ecological divergence in sympatry have led to reproductive isolation between dwarf and normal whitefish (Lu & Bernatchez 1998; Rogers & Bernatchez 2006; Renaut *et al.* 2009).

## 3. IDENTIFYING ADAPTIVE PHENOTYPIC TRAITS

Dwarf and normal sympatric whitefish use limnetic and epibenthic habitat, respectively (Bernatchez *et al.* 1999; Landry 2009), and exhibit phenotype-environment associations with their respective niches. The most discriminating morphological trait between them is the gill-raker apparatus whereby dwarf whitefish typically have more numerous and less separate gill-rakers than normal whitefish, resulting in a more efficient retention of smaller, planktonic prey (reviewed in Bernatchez 2004; Kahilainen *et al.* 2007). Dwarf whitefish are further characterized by a

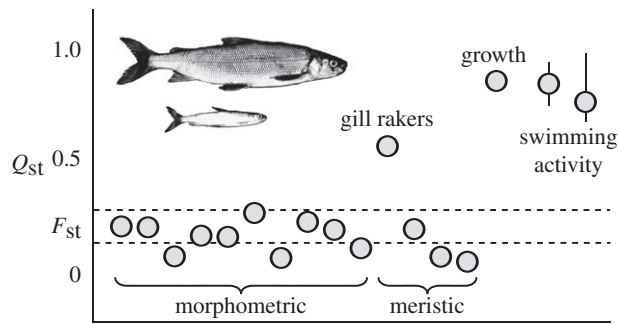


Figure 1.  $Q_{st}$  estimates for 10 morphological traits (in order: preorbital length, orbital length, trunk length, dorsal fin length, caudal peduncle length, maxillary width, maxillary length, body depth, head depth and interorbital width), four meristic traits (suprapelvic scales, dorsal ray count, pectoral ray count and gill-raker count), growth rate, and two swimming behaviour parameters (depth selection and burst swimming), with 95% CIs (CIs denote swimming activities). The 95% CI of neutral expectations ( $F_{st}$  value of 0.24) is delineated by dotted lines (modified from Rogers *et al.* 2002).

smaller size at maturity, slower growth rate, younger age at maturation and reduced fecundity relative to normal whitefish (Fenderson 1964; Rogers & Bernatchez 2005). In some lakes at least, limnetic (dwarf) fish tend to also suffer more predation than normal whitefish (Kahilainen & Lehtonen 2002). Normal and dwarf whitefish differ in heritable swimming behaviour: dwarf whitefish typically swim more actively and prefer higher positioning in the water column in experimental laboratory conditions (Rogers *et al.* 2002). Dwarf whitefish also have a higher metabolic rate, partly associated with the cost of more active swimming behaviour, and lower bioenergetic conversion efficiency (growth rate/consumption rate ratio; Trudel *et al.* 2001). Further evidence that natural selection has played an important role in the phenotypic divergence between dwarf and normal whitefish was provided by means of a  $F_{st}-Q_{st}$  analysis (Spitze 1993; figure 1). By comparing the extent of divergence of numerous phenotypic traits ( $Q_{st}$ ) with that measured at neutral microsatellite markers ( $F_{st}$ ) under both natural and experimental conditions, it was found that gill rakers, growth and swimming behaviour were the most likely traits to have evolved under the effect of divergent selection and therefore were identified as the most relevant phenotypic targets for further investigation (Rogers *et al.* 2002).

#### (a) Identifying adaptive phenotypic traits: insights from the transcriptome

Ever since the pioneering work of Britten & Davidson (1971) and King & Wilson (1975) on the mechanisms and evolutionary role of regulatory processes, there has been increasing recognition that variation in levels of gene expression represents a major source of evolutionary novelty, which in turn can lead to phenotypic divergence by natural selection (Oleksiak *et al.* 2002; Wittkopp 2007; Wray 2007; Fay & Wittkopp 2008). The development of microarray

technologies, allowing the simultaneous detection of expression modulations at thousands of genes, offers a powerful means of assessing the importance of evolutionary change in gene regulation involved in adaptive population divergence (Bochdanovits *et al.* 2003; Ranz & Machado 2005; Matzkin *et al.* 2006). Moreover, microarrays represent an efficient tool for exploring the molecular mechanisms of life history trade-offs (Roff 2007).

Over the last few years, we have intensively used gene expression studies to investigate the molecular basis of adaptive divergence between dwarf and normal whitefish. Some of the specific questions addressed were: (i) is parallel phenotypic evolution observed between independently evolved dwarf and normal whitefish pairs accompanied by parallelism in expression of the genes potentially underlying phenotypic divergence? (Derome & Bernatchez 2006; Derome *et al.* 2006; Whiteley *et al.* 2008; Jeukens *et al.* 2009), (ii) what are the molecular mechanisms underlying life-history trade-offs (including sexual conflicts) potentially involved in the adaptive divergence of dwarf and normal whitefish? (Derome *et al.* 2008; St-Cyr *et al.* 2008), and (iii) what proportion of the genome is differentially regulated and expressed at different life history stages (Nolte *et al.* 2009; Renaut *et al.* 2009).

These microarray studies were conducted both in controlled (laboratory) environments and two natural lakes, involved three life stages (embryos, juveniles and adults), and three tissues (white muscle, liver and brain). Given the non-availability of a specific whitefish microarray, total RNA extracts reverse-transcribed to cDNA were probed on either 3557 and later 16 006 features cDNA microarrays developed for the Atlantic salmon (*Salmo salar*) by the cGRASP (consortium for Genomic Research on All Salmon Project; Rise *et al.* 2004; von Schalburg *et al.* 2005). Studies have shown that this microarray platform performs well for closely related salmonid species, including lake whitefish (Rise *et al.* 2007) and that hybridization levels were comparable in both normal and dwarf whitefish (Derome *et al.* 2006).

Table 1 summarizes the general patterns of differentially expressed genes across three tissues at the adult stage. It first shows that several hundred genes were differentially expressed between adult dwarf and normal whitefish. Secondly, the proportion of differentially expressed genes out of the total number that were expressed was similar across tissues, varying from 11.1 per cent (brain) to 13.8 per cent (white muscle) and up to 15.9 per cent (liver). Out of those differentially expressed genes, a parallel pattern in level of gene expression between dwarf and normal whitefish from two natural lakes was observed for both white muscle and liver, with a higher proportion of parallel genes observed for the latter (2.39% versus 1.35%). This is of key interest, because parallel evolution among independent, closely related lineages is a predictable consequence of natural selection (Harvey & Pagel 1991; Schluter 2000). In addition to natural lakes, experiments conducted on fish (liver tissue) kept in controlled conditions revealed that 34 out of the 92

Table 1. Summary of differentially expressed genes across three tissues at the adult stage between dwarf and normal whitefish. Number refers to the total number of genes for which significant differences in level of expression were observed. Parallel refers to the number of genes that showed similar patterns of differences in expression between pairs of dwarf and normal whitefish from two natural lakes. (%) refers to either the proportion of differentially expressed genes (for number) or the proportion of parallel genes out of the total number of genes that were significantly expressed in a given tissue. Upregulated refers to the percentage of parallel genes that showed a higher level of expression in dwarf relative to normal whitefish. Variance ( $P/N$ ) is the ratio of inter-individual variance of expression for parallel genes over the variance for genes with non-parallel pattern of expression. n.a., not available. Note that the analyses of liver and brain were based on the cGRASP 16 006 cDNA feature microarray whereas the analysis of white muscle used the earlier 3557 feature version.

tissue	number (%)	parallel (%)	upregulated (in dwarf; %)	variance ( $P/N$ )	references
white muscle	164 (13.8)	16 (1.35)	62	0.11	Derome <i>et al.</i> (2006)
liver	502 (15.9)	92 (2.39)	71	0.57	St-Cyr <i>et al.</i> (2008)
brain	201 (11.1)	n.a.	n.a.	n.a.	Whiteley <i>et al.</i> (2008)

parallel genes observed in natural lakes showed the same differential patterns of gene expression. A heat-map presented in figure 2 shows how the differential pattern of expression at parallel genes unambiguously clusters dwarf and normal whitefish separately. Interestingly, the ratio of inter-individual variance of expression for genes showing parallelism of expression over the variance for genes with non-parallel pattern of expression showed that variance of expression was much smaller for the former than the latter in both muscle and liver (table 1). Moreover, for parallel genes observed in muscle and liver, the proportion of upregulated genes was higher in dwarf compared with normal whitefish, in accordance with the general observation of a more active metabolism in dwarf whitefish.

The hundreds of genes that showed differential patterns of transcription between dwarf and normal whitefish across the three tissues were classified into at least 30 different functional groups using information from various sources (e.g. cGRASP website, Swiss Prot/TrEMBL database, NCBI GO browser, KEGG Pathway database, SOURCE Database, EMBL Bioinformatic harvester, completed with references from the literature; see the electronic supplementary material, table S1). Of particular interest are those functional groups that were over-represented in terms of number of parallel genes showing differences between dwarf and normal whitefish relative to the total number of genes that were expressed in both forms for each functional group. Dwarf whitefish consistently showed significant over-expression of genes potentially associated with survival through enhanced activity (energy metabolism, muscle contraction, homeostasis, lipid metabolism and detoxification) whereas genes associated with growth (protein synthesis, cell cycle and cell growth) were generally upregulated in normal relative to dwarf whitefish (figure 2).

Gene expression was also screened using the 16 006 cDNA microarray to identify regulatory changes at embryonic and juvenile stages and compare them with previous results at the adult stage (table 2). This revealed that 16-week-old juvenile fish have 14 times more genes displaying significant regulatory divergence than embryos, for which very few genes showed significant differences in expression between dwarf and normal whitefish. Moreover, regulatory changes in 16-week-old juvenile fish matched patterns

in adult fish, which suggests that gene expression divergence is established early in juvenile fish and persists throughout the adult phase. For example, we found that at least 26 genes in juvenile fish were also candidate genes involved in adaptive divergence in adult fish. Finally, real-time polymerase chain reaction was used to document transcriptional divergence for specific candidate genes identified from the above microarray experiments among North American (*C. clupeaformis*) and European (*C. lavaretus*) populations (Jeukens *et al.* 2009). Parallelism in expression of candidate genes was observed across whitefish limnetic and benthic pairs, including opsin genes (vision function) that were not screened by microarrays, which provided indirect strong support for the hypothesis that divergent natural selection has been acting similarly on the same genes in the adaptive radiation of whitefish both in North American and European lakes.

#### 4. ELUCIDATING THE GENETIC BASES OF PHENOTYPES

##### (a) *Insights from pQTL (phenotype QTL) mapping*

Mapping genetic regions underlying the expression of phenotypes represent an essential step towards identifying the genomic locations involved in adaptive divergence. The role of these genomic locations as potential barriers to gene flow can then be assessed (Rieseberg *et al.* 1999). This approach involves documenting the genetic architecture, i.e. the number, location and effect of genomic locations contributing to differentiation within and among populations or species (Rieseberg 1998; Orr & Turelli 2001). Such a genome-wide perspective is also essential for a complete understanding of the functional genomic response that occurs as evolutionary processes influence populations as they diverge (Wu & Ting 2004). In order to elucidate the genetic bases of phenotypic divergence between dwarf and normal whitefish, we have used linkage mapping to document the number and effects of QTL involved in controlling adaptive traits differing between them. We first assembled genetic linkage maps for two backcross ( $F_1 \times$  limnetic,  $F_1 \times$  benthic) hybrid families wherein over 900 AFLP and microsatellite loci were genotyped and positioned in 336 progeny (Rogers *et al.* 2007). The homology of mapped loci between families resolved 34 linkage

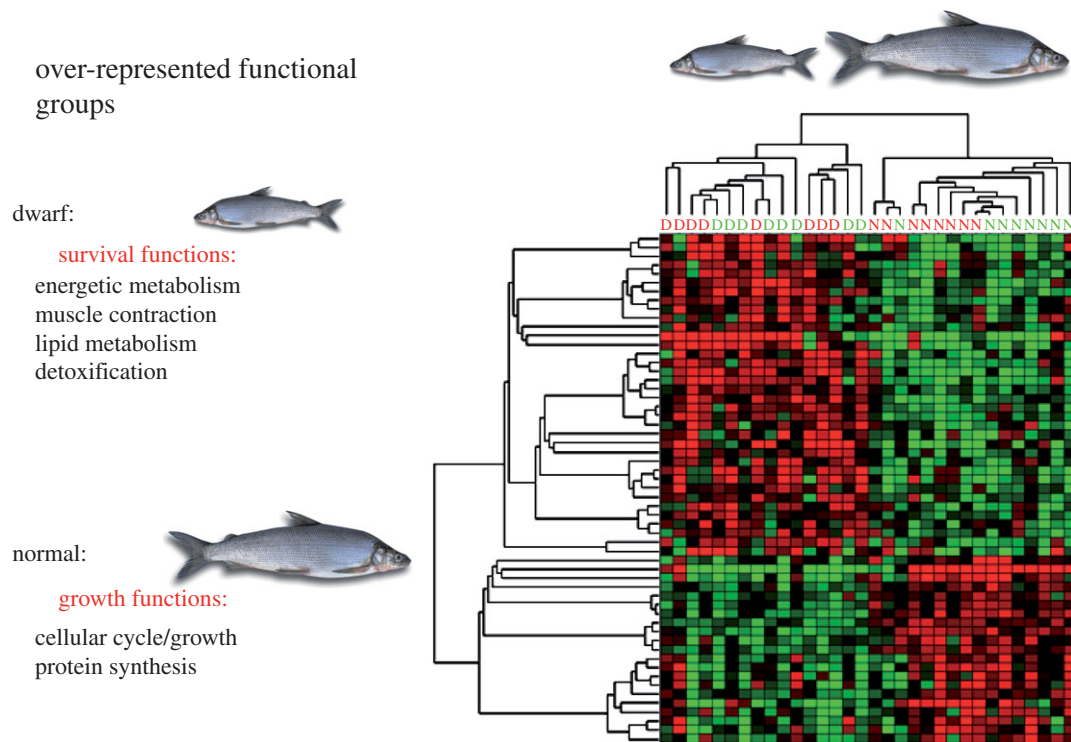


Figure 2. Heat-map illustrating parallel patterns of regulation between sympatric dwarf and normal whitefish from the two natural lakes (modified from St-Cyr *et al.* 2008). The dendrogram on top clusters individuals based on similar patterns of transcription for genes. Individuals are designated by the letters N and D for ‘normal’ and ‘dwarf’ whitefish (red, Cliff Lake; green, Indian Pond). The dendrogram on the left clusters genes with similar patterns of expression among individuals. Expression values shown in red represent genes that are upregulated in dwarf on average whereas those in green are upregulated in normal on average. To the left of the heat-map are listed functional groups that were over-represented in terms of number of parallel genes showing differences between dwarf and normal whitefish relative to the total number of genes that were expressed in both forms for each functional group (see details in the main text).

Table 2. Summary of the total number of differentially expressed genes across three life stages between dwarf and normal whitefish. Number refers to the total number of genes for which statistically significant differences in level of expression were observed. Proportion refers to the fraction of differentially expressed genes out of the total number of genes that were significantly expressed at a given life stage.

life stage	number	proportion	references
embryos	33	0.008	Nolte <i>et al.</i> (2009) and Renault <i>et al.</i> (2009)
juveniles	502	0.112	Nolte <i>et al.</i> (2009) and Renault <i>et al.</i> (2009)
adults	976	0.136	Derome <i>et al.</i> (2006), St-Cyr <i>et al.</i> (2008) and Whiteley <i>et al.</i> (2008)

groups that exhibited 83 per cent colinearity among linked loci between these two families. These maps represented the basis for assessing the genetic architecture of eight adaptive traits in the two hybrid backcross families where the phenotypes of individual offspring were determined throughout their life history (Rogers & Bernatchez 2005, 2007). A total of 34 pQTL linked to these traits and distributed over 13 linkage groups were identified using interval mapping (figure 3).

### (b) Insights from eQTL (expression QTL) mapping

The analysis of the genetic architecture of transcriptome variation is increasingly considered as a powerful way to further our understanding of the mechanistic basis of adaptive divergence (Roff 2007; Mackay *et al.* 2009). QTL analyses of gene expression traits (eQTL) treat transcript abundance as a quantitative trait and apply traditional QTL mapping techniques to localize genetic determinants of gene expression (Gibson & Weir 2005; Gilad *et al.* 2008). In addition, it is possible to compare genomic eQTL locations with pQTL to locate candidate genes underlying phenotypic traits (Wentzell *et al.* 2007). This ‘phenomics’ approach provides a solid framework for examining transcriptional underpinnings of phenotypic divergence.

We used microarray analyses to localize the genetic determinants of gene expression for progeny of a backcross whitefish family previously used to construct the linkage (and pQTL) map (Derome *et al.* 2008; Whiteley *et al.* 2008). The analyses were performed for two tissues on a subset of 66 (white muscle tissue) and 57 (brain tissue) of the same individual progeny used to build the genetic and pQTL map. We attempted to elucidate the genomic distribution of muscle and brain eQTL and to determine the extent to which genes controlling transcriptional variation may underlie adaptive divergence in dwarf and normal whitefish. Overall, we identified hundreds of significant eQTL in both brain and muscle that were

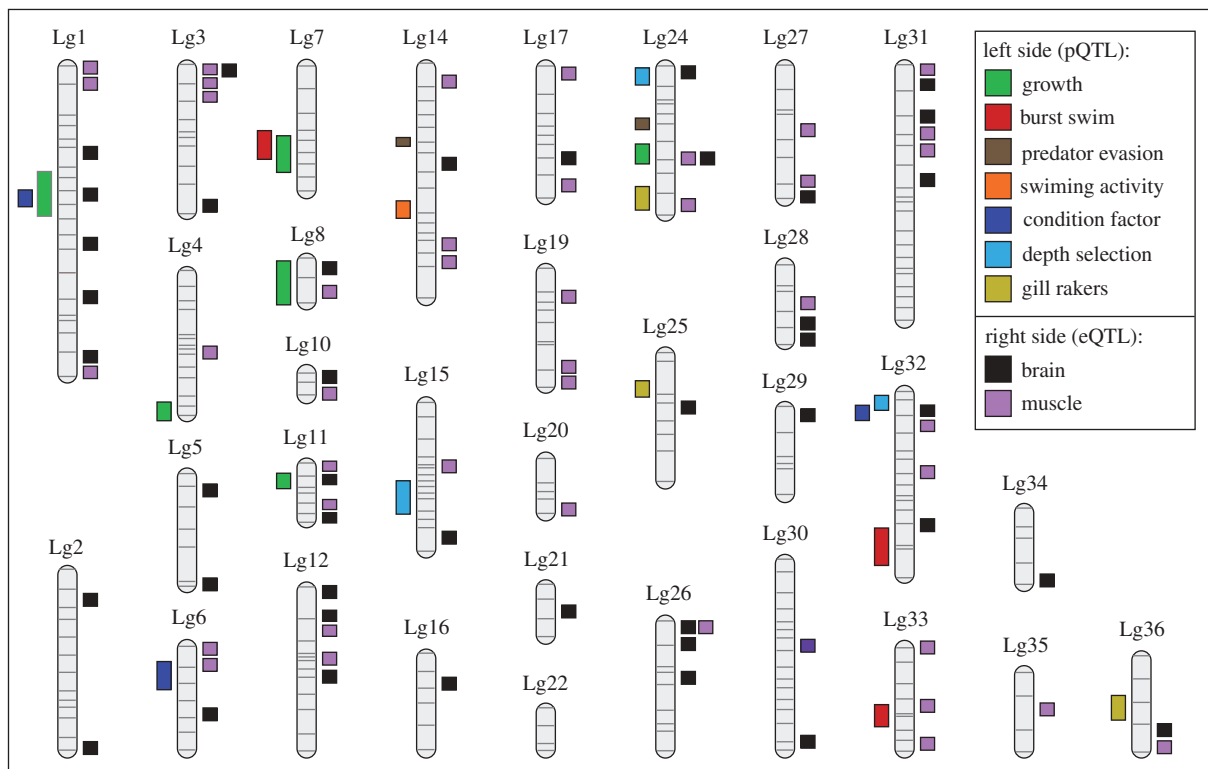


Figure 3. Genome locations of significant pQTL and eQTL observed both in white muscle and brain transcriptome. Locations of eQTL are illustrated on the right-hand side of linkage groups (white muscle, purple square; brain, black squares) and locations of pQTL are illustrated by various colours on the left side. Data for eQTL locations were compiled from combined and sex-specific eQTL maps constructed by Whiteley *et al.* (2008) (brain) and Derome *et al.* (2008) (white muscle). Each square represents the location of at least one eQTL. Data for pQTL are from Rogers & Bernatchez (2007).

distributed over numerous linkage groups with several overlapping with pQTL (figure 3).

More specifically, results for white muscle revealed 261 eQTL involving 138 unigenes (for a mean of about 1.9 eQTL per gene) distributed over 24 linkage groups and consisting of 15 eQTL that co-localized with pQTL. Results obtained for brain revealed a similar number of eQTL (249) but involving many more unigenes than for muscle (237 for a mean of 1.05 per gene) distributed over 28 linkage groups with up to 50 eQTL that co-localized with pQTL. This allowed identification of candidate genes for species pair divergence involved in various functions, namely energetic metabolism, protein synthesis, cell growth, muscle contraction and neural development, on the basis of co-localization of eQTL for genes expressed in both tissues with previously identified adaptive pQTL. For instance, 88 per cent of eQTL-pQTL co-localization involved growth rate and condition factor QTL, two traits central to the adaptive divergence of whitefish species pairs (figure 3). We found several cases of co-localization between eQTL and pQTL that could potentially have a functional association, although this remains to be rigorously investigated. For example, an eQTL regulating the expression of an ATP synthase  $\beta$  chain gene in white muscle involved in energy metabolism overlapped with a pQTL underlying condition factor on linkage group 32. In muscle also, an eQTL for troponin which plays a role in fast muscle contraction overlapped with a pQTL underlying burst swimming on linkage group 33.

The most prominent feature of both brain and muscle eQTL maps was the non-random distribution of eQTL, which translated into a prevalence of 'eQTL hotspots' characterized by the co-localization on the same chromosomal region of many eQTL associated with genes of very distinct functions. This is of prime interest to decipher the genetic architecture of adaptive divergence since regions with a high number of distal eQTL might harbour master regulator genes (e.g. transcription factors; Gilad *et al.* 2008). Defining a hotspot as the co-localization of four eQTL or more, 47 per cent of all 249 eQTL identified in brain were associated with 12 hotspots distributed over eight linkage groups (Whiteley *et al.* 2008). These hotspots contained nine eQTL on average, but up to 32 eQTL on the same hotspot were observed. Interestingly, 95 per cent of all eQTL grouping at a given hotspot showed the same directionality in additive effect. An even more pronounced non-random pattern of eQTL distribution was observed in white muscle with 41 per cent of all eQTL grouping into six hotspots distributed across four linkage groups (Derome *et al.* 2008). Each hotspot contained an average of 18 eQTL that co-localized to a same position. As for the brain results, the majority (80%) of all eQTL grouping at a given hotspot showed the same directionality in additive effect.

The most extreme case of non-random eQTL distribution was observed for muscle on linkage group 25, where two tightly linked hotspots, respectively, contained 34 and 53 eQTL. Each of these contained

eQTL regulating genes with numerous functions. The most represented functions were protein synthesis, cell growth, energy metabolism and muscle contraction. Transcripts involved in protein synthesis and cell growth exhibited systematically opposed additive effects to those associated with energy metabolism and muscle contraction (Derome *et al.* 2008), which is consistent with the observed trade-off in life-history traits and differential gene expression distinguishing dwarf and normal whitefish (St-Cyr *et al.* 2008). Directional predominance of additive effects provides strong evidence for the role of directional selection in shaping genome architecture (Orr 1998). Therefore, these results suggest that directional selection has been acting on patterns of gene regulation associated with these two tightly linked chromosomal regions. Moreover, such a concentration of eQTL belonging to diverse functional groups suggests that single master regulators located at these positions could control their expression. Indeed, these hotspots contained eQTL for genes known to play fundamental roles in transcription regulation (e.g. DPY-30-like protein gene, Histone H1 homologue protamine, small ubiquitin-like modifier (SUMO)). Putative master regulatory genes were also observed at other eQTL hotspots, such as zinc-finger proteins, both for brain and muscle. These could represent mutations related to gene regulation and be interpreted as a form of early substitution of large effect theoretically predicted during adaptive divergence (Orr 2005), and hypothetically play a central role in the divergence of dwarf and normal whitefish.

Another prominent feature of both eQTL maps was a pronounced sex bias in transcriptional genetic architecture observed for both brain and white muscle, with more eQTL observed in males in the case of brain and vice versa for white muscle (Derome *et al.* 2008; Whiteley *et al.* 2008). In addition, we detected differentially expressed transcripts associated with eQTL segregating in sex-specific datasets and mostly belonging to functional groups that differentiate dwarf and normal whitefish in natural populations.

## 5. FINDING EVIDENCE FOR NATURAL SELECTION AT THE GENOME LEVEL IN THE WILD

Understanding the response of genetic variation to natural selection is required to demonstrate that the mechanism causing divergence between populations or species at the genome level is adaptive. To this end, genome scans have been increasingly used to identify genomic regions resisting the homogenizing influence of gene flow and therefore likely to be under divergent selection (Nosil *et al.* 2009a,b). This approach relies on the principle that following hybridization, recombinant genotypes with maladaptive combinations of loci will be removed, resulting in a differential pattern of introgression across the genome. Loci (or linked loci) incurring reduced fitness in recombinant genotypes will be characterized by a greater genetic differentiation relative to neutral or selectively advantageous loci (Barton 2001). Therefore, identifying loci linked to genes implicated in differential adaptation and/or reproductive isolation

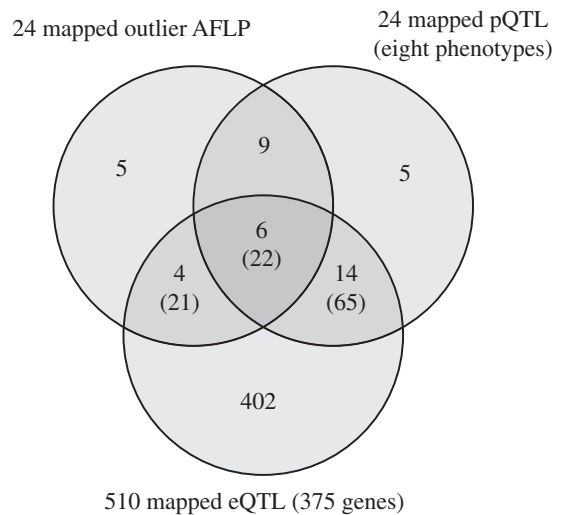


Figure 4. Venn diagram showing the number of significant mapped outlier AFLP identified in a genome scan of natural populations (Campbell & Bernatchez 2004), the number of significant mapped pQTL for eight different phenotypes (Rogers & Bernatchez 2007), and the number of significant eQTL locations (with number of associated unigenes in parentheses) combined over both studies on brain (Whiteley *et al.* 2008) and white muscle tissue (Derome *et al.* 2008). Numbers are presented for each separate group, as well as for the number of loci shared between any of the two or all three groups.

can be achieved by quantifying patterns of genetic differentiation between populations that are still exchanging genes (Beaumont & Nichols 1996; Beaumont & Balding 2004; Foll & Gaggiotti 2008). This requires a 'population genomics' approach wherein a large number of loci (hundreds or more) are genotyped to accurately estimate the expected level of genetic differentiation under neutrality, and the proportion of loci linked to genes implicated in adaptation and reproductive isolation. Genome scans can also reveal parallel trends of divergence through the analysis of multiple populations, consequently offering stronger support for the role of natural selection in adaptive trait evolution. When applied to populations in an early stage of divergence, genome scans reveal the mosaic nature of the genome whereby markers in divergently selected genomic regions will reflect the evolutionary history of adaptive divergence and ecologically based reproductive isolation (Via & West 2008).

We performed a genome scan by which the pattern of genetic differentiation obtained using 440 AFLP loci was compared with that expected under neutrality estimated by simulations in four sympatric pairs of dwarf and normal whitefish (Campbell & Bernatchez 2004). A total of 48 loci with an average of 14 loci per lake showed restricted gene flow relative to neutral expectation, suggesting a role of directional selection on their divergence. Further insight into the identity of these 'anonymous' AFLP markers was gained by combining information obtained from population genomics with that of pQTL and eQTL mapping (Rogers & Bernatchez 2007; Derome *et al.* 2008; Whiteley *et al.* 2008). Thus, homology between markers used in mapping and population genomics was found for 180 markers and for 24 of the 48 outlier loci. As illustrated in figure 4, 19 of these 24 outliers

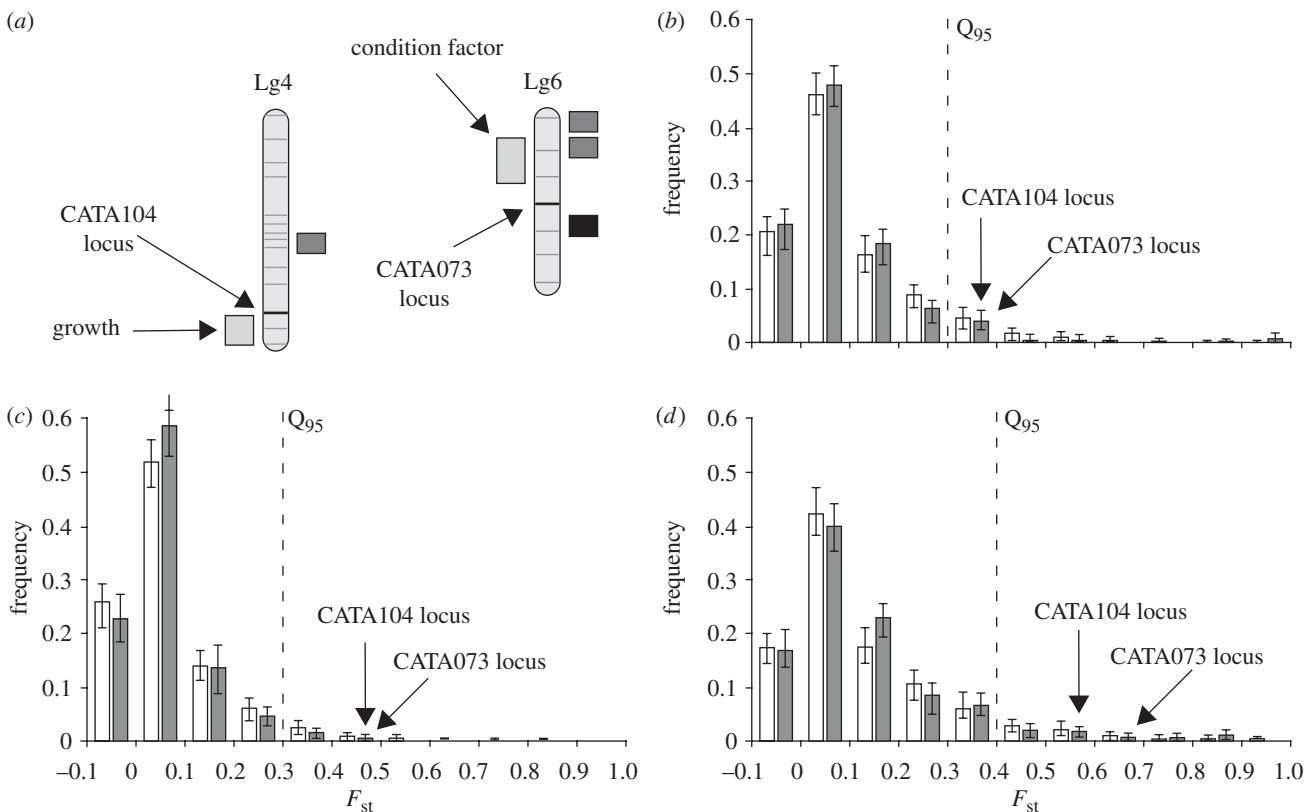


Figure 5. (a) Linkage groups showing location of AFLP markers linked to a growth pQTL (CATA104) on linkage group 4 and to an eQTL for cytochrome *c* oxidase subunit VI (CATA073), which is also a possible candidate for hitchhiking divergence with a pQTL for condition factor on linkage group 6 (grey-filled box, muscle eQTL; filled black box, brain eQTL). (b–d) Comparison of simulated (open bars) and observed (filled bars) distributions of  $F_{st}$  (with 95% CIs) for sympatric dwarf and normal whitefish from three different lakes (Webster, East, Cliff, respectively; Campbell & Bernatchez 2004), and position of  $F_{st}$  values observed for outliers CATA104 and CATA073 in each case.

corresponded to either pQTL ( $n = 9$ ), eQTL (10 positions involving eQTL for 21 genes), or both (six positions involving 22 genes).

For pQTL, loci exhibiting a signature of selection were associated with QTL relative to other regions of the genome more often than expected by chance alone (Rogers & Bernatchez 2007). Among the 15 such pQTL, eight were associated to growth and two of these showed patterns of parallelism among independent lakes. Figure 5 illustrates the case of pQTL CATA104 associated with growth on linkage group 4 and showing highly reduced gene flow between dwarf and normal whitefish in three different lakes. Moreover, these two parallel QTL outliers exhibited segregation distortion in mapping families, supporting the hypothesis that adaptive divergence contributing to parallel reductions of gene flow among natural populations may also cause genetic incompatibilities (see below, Rogers & Bernatchez 2007).

Table 3 presents the list of the 10 AFLP outliers distributed over 10 different linkage groups that were also associated with either white muscle or brain eQTL, among which six also co-localized with pQTL. These represent a total of 43 genes involved in numerous functional groups, protein synthesis and energy metabolism being the most represented. Most genes falling into these two functional groups were also previously identified as candidates in the adaptive divergence between dwarf and normal whitefish on the basis of

parallelism in transcription. One example of an eQTL outlier is CATA073; located on linkage group 6. This locus showed highly reduced gene flow between dwarf and normal whitefish in three lakes, is linked with an eQTL for cytochrome *c* oxidase subunit VI, a gene involved in the energy-producing oxidative phosphorylation (Whiteley *et al.* 2008), and is a possible candidate for divergence hitchhiking (*sensu* Via 2009) with a pQTL for condition factor (Rogers & Bernatchez 2007; figure 5). Evidence for parallelism in transcription or significant difference in expression in at least one dwarf/normal pair was also found for the majority of genes falling in other functional groups (table 3). It is also noteworthy that three out of 10 AFLP outliers also corresponded to eQTL hotspots, all of them containing a putative master regulatory gene.

## 6. MECHANISMS OF REPRODUCTIVE ISOLATION AND THEIR GENETIC BASIS

The ecological theory of adaptive radiation predicts that reproductive isolation evolves as a by-product of divergent natural selection, which shares a common genetic basis with adaptation through pleiotropic interactions (Schluter 2000). However, the underlying genetic basis of reduced hybrid fitness (including inviability) that causes reproductive isolation in cases of adaptive divergence has remained elusive for most



Table 3. List of the 10 AFLP outliers identified by genome scans of natural populations that were also associated with either white muscle (plain characters) or brain (in italics) eQTL, among which six also overlapped with pQTL and that were distributed over 10 different linkage groups (Lg). These comprised a total of 43 genes involved in numerous functional groups (function). Most genes associated with these eQTL were also previously identified as candidates in the adaptive divergence between dwarf and normal whitefish on the basis of either transcription parallelism in more than one species pair (P) or significant differences when only one pair was analysed (D). Other function refers to genes for which annotation could not be confirmed.

outlier mapped loci	Lg	pQTL	eQTL	function	candidate criterion
GGTG105.0	1	growth condition factor	<i>40S ribosomal protein SA</i> <i>60S ribosomal protein L10A</i>	protein synthesis protein synthesis	P P
CATA104.5	4	growth	collagen $\alpha 2(I)$ chain precursor	blood and transport	
CATA073.7	6	condition factor	<i>cytochrome c oxidase polypeptide VIa</i> <i>60S ribosomal protein L31, L35a</i>	energy metabolism protein synthesis	P P
CATA134.6	11	growth	aspartic acid-rich protein aspolin2	other function	
CTAG061.0	12	—	<i>transferrin gene, partial 5' sequence</i> <i>12S ribosomal RNA,</i> <i>Ataxin-2</i> <i>vacuolar protein sorting-associated protein 28 homologue</i> <i>vacuolar ATP synthase subunit B, brain isoform</i> <i>mitochondrial import inner membrane translocase subunit</i> <i>protein S100-A1</i>	iron homeostasis protein synthesis transcription regulation protein degradation energy metabolism energy metabolism other function neural plasticity	P D P D D
CAAG116.1	14	directional changes			D
CCTC051.3	19	—	acidic ribosomal protein P1 60S ribosomal protein L39 60S ribosomal protein L19 MHC class I LIM domain binding 3 F-actin capping protein $\beta$ subunit SPARC precursor 40S ribosomal protein S3 actin, cytoplasmic 2 Zinc finger protein 35 triosephosphate isomerase $\alpha$ actinin skeletal muscle isoform 3	protein synthesis protein synthesis protein synthesis immunity other function germ line formation muscle contraction protein synthesis muscle contraction transcription regulation energy metabolism muscle contraction	P P P P P P D P
GGTG052.9	24	gill rakers growth	40S ribosomal protein S5 40S ribosomal protein S19 autophagy protein 12-like <i>vigilin (high density lipoprotein-binding protein)</i> <i>Cyp19b-I gene for P450aromB-I glutathione S-transferase</i> <i>ATP-binding cassette transporter 1</i> <i>Ferritin-H</i> <i>hCG1982388, transcript variant 5</i> <i>TNF-<math>\alpha</math>2 gene</i> <i>Zinc finger protein 161</i> <i>tetratricopeptide repeat protein 35</i> <i>Reticulon 4 (Neurite outgrowth inhibitor)</i> <i>CA061 074 WD repeat protein 41</i>	protein synthesis protein synthesis protein metabolism other function immunity neuroprotection transport protein iron binding and homeostasis other function immune system transcription regulation other function neural development other function	P P P P P P P P D
CATA158.6	27	—	40S ribosomal protein S4, X isoform ubiquitin carboxyl-terminal hydrolase isozyme L3	protein synthesis other function	P P
CGTC060.7	28	—	<i>ATP synthase b-chain</i> glutathione S-transferase $\mu 5$	energy metabolism energy metabolism	P P

species (Coyne & Orr 2004). By examining the genetic basis of post-zygotic isolation during the early stages of population divergence, it is possible to determine the factors accounting for a reduction in hybrid fitness and to gain insight into the evolutionary forces that

lead to the formation of these barriers. This includes concurrently investigating intrinsic and extrinsic factors influencing reproductive isolation (Rice & Hostert 1993). Intrinsic post-zygotic isolation is independent of the environment and arises from genomic

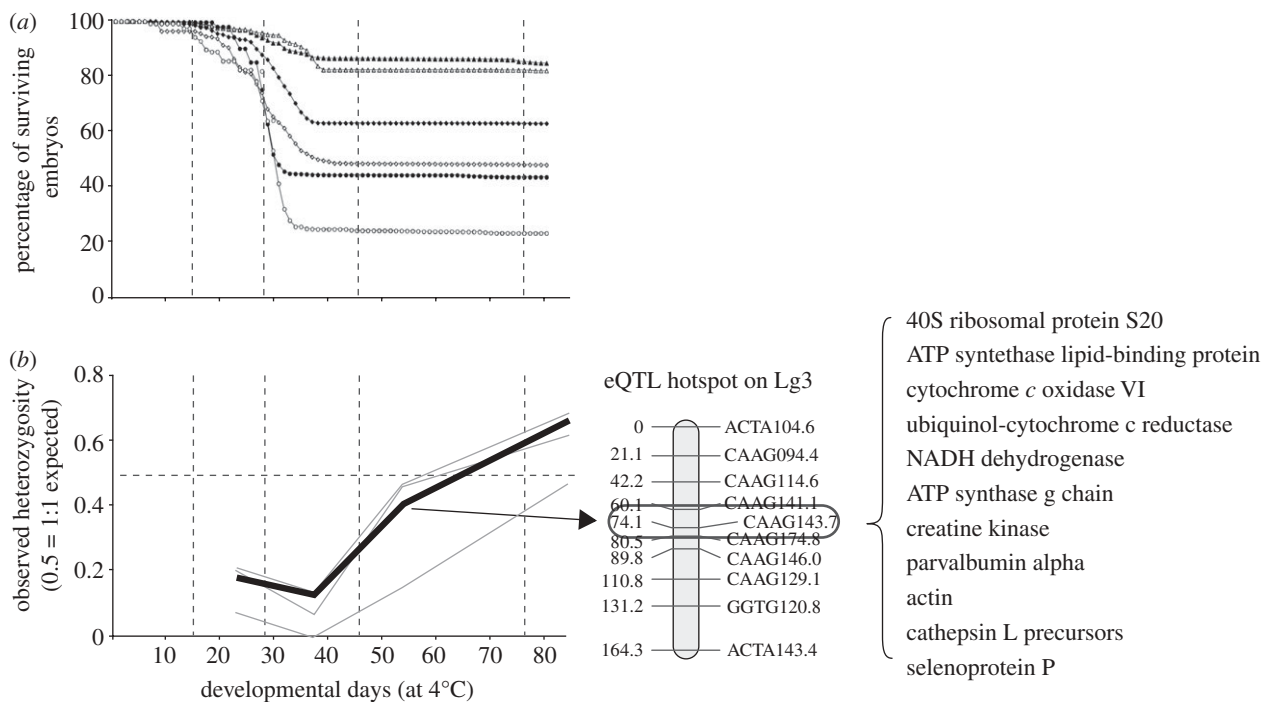


Figure 6. (a) Percentage of surviving embryos as a function of developmental time for dwarf, normal, F<sub>1</sub> hybrids and backcrosses hybrids (dwarf and hybrid X normal). Vertical lines delineate distinct developmental stages (modified from Rogers & Bernatchez 2006). (b) Genotypic response of embryonic mortality as a function of linked intrinsic genetic incompatibilities during development on linkage group 3 (Rogers & Bernatchez 2006). Developmental time on the x-axis is proportional to (a) for comparison with embryonic mortality rates. Four loci exhibited a significant shift in segregation ratios during development. Locus CAAG143.7 is identified by the dark line and its position on linkage group 3 is shown (Rogers *et al.* 2007), where it overlaps with a hotspot comprising eQTL for 11 genes involved in various biological functions (Derome *et al.* 2008; see main text for details).

incompatibilities caused by divergent developmental systems. On the other hand, extrinsic post-zygotic isolation is environmentally dependent and associated with lower hybrid fitness in specific environments. Investigating the genetic basis of intrinsic and extrinsic factors was particularly relevant in dwarf and normal whitefish since both the accumulation of genetic differences in allopatry and ecological divergence in sympatry following secondary contact may have contributed to their reproductive isolation. This was achieved through laboratory experiments and linkage mapping.

#### (a) *Intrinsic post-zygotic isolation*

Evidence for intrinsic post-zygotic isolation between dwarf and normal whitefish has come from two sources. However, observations were obtained from two mapping families only and should be interpreted accordingly. First, reduced sperm performance observed in backcross individuals relative to parental and F<sub>1</sub> hybrids was consistent with genomic incompatibilities that create a range of negative fitness effects in post-F<sub>1</sub> hybrids (Whiteley *et al.* 2009). Second, we observed a pronounced pattern of segregation distortion in the two backcross mapping families for nearly one third of mapped markers across 28 linkage groups (Rogers *et al.* 2007). Moreover, there was a significant correlation between the percentage and direction of segregation distortion for 64 homologous loci between both maps. During embryonic development, such pronounced locus-specific deviation from

Mendelian segregation is likely to reflect genomic incompatibilities associated with intrinsic hybrid inviability (Vogl & Xu 2000). In order to test this hypothesis, we analysed patterns of Mendelian segregation at loci with known linkage associations among backcross embryos sampled at critical periods of development against larvae within the same family that successfully survived to emergence (Rogers & Bernatchez 2006). Our data provided evidence for the role of embryonic mortality as a strong intrinsic reproductive isolation barrier (figure 6). At its peak mortality, the mortality rate of backcross hybrids was 2.5 to 3.7 times greater than that observed previously in F<sub>1</sub> hybrids (Lu & Bernatchez 1998) and 5.3 to 6.5 times greater than observed in pure crosses. When comparing embryos that died at different times against larvae that successfully hatched, several loci distributed along different linkage groups exhibited temporal deviation in Mendelian segregation ratios, which ultimately lead to significant segregation distortion in the surviving progeny. On linkage group 3 in particular, a positive relationship between four tightly linked loci suggested that increased embryonic mortality during development was a function of incompatibilities on this segment of the genome between dwarf and normal genomes (figure 6).

Very interestingly, combining information from the white muscle eQTL map with that from linkage mapping revealed that one of these severely distorted markers (CAAG143.7) overlapped with an eQTL hotspot on Lg3. This hotspot contained 11 genes involved

in many biological functions, the most important being energy metabolism (cytochrome *c* oxidase subunit VI, creatine kinase, ATP synthase, ubiquinol-cytochrome *c* reductase, NADH dehydrogenase, selenoprotein P), but also protein synthesis (40S ribosomal protein S20), muscle contraction (parvalbumin, actin), and reproduction (cathepsin L precursor; figure 6). It is also noteworthy that parallelism in patterns of gene expression was observed for the majority of those genes in microarray studies (Derome *et al.* 2006; St-Cyr *et al.* 2008) and that several loci linked to Lg3 had previously been found to be significantly resistant to introgression between whitefish from distinct glacial races (Rogers *et al.* 2001).

### (b) Evidence for genetically determined extrinsic post-zygotic isolation

Extrinsic, ecologically based, post-zygotic isolation was evidenced by a comparative analysis of the timing and synchronicity of larval emergence between pure and hybrid crosses, a trait that is critical to the survival of fishes in nature (Cushing 1990). We also tested for QTL associated with time to emergence to determine whether extrinsic post-zygotic isolation barriers may contribute to hybrid inviability (Rogers & Bernatchez 2006). Under controlled environmental conditions, the dwarf and normal experimental families exhibited a synchronicity similar to what was observed in the natural populations (Chouinard & Bernatchez 1998). In the F<sub>1</sub> hybrids, mean hatching time was intermediate to the dwarf and normal crosses, offering evidence that developmental time to emergence was under additive genetic control. In contrast, the time to emergence in backcrosses was highly asynchronous: time to emergence was delayed approximately 11 days while the total hatching span was 30 days longer than observed in pure crosses. Moreover, hybrid backcrosses exhibited increased variance with respect to time to emergence, which was 10 times higher when compared with natural dwarf and normal populations. In the linkage map, interval analyses provided a genetic basis for this trait as it revealed one significant and suggested another QTL associated with time to emergence, which, respectively, explained 29 and 27 per cent of the phenotypic variance. Additionally, the patterns of hatching in backcrosses were indicative of transgressive segregation. That is, variation in time to emergence in these hybrids exceeded the combined variation of both parental populations.

### (c) Molecular basis for reproductive isolation: further insights from the transcriptome

As exemplified above, both hybrid breakdown and transgressive segregation may explain the underlying basis of post-zygotic reproductive isolation (either intrinsic or extrinsic) between dwarf and normal whitefish. The manifestation of such extreme traits supposes non-additive, epistatic, gene interactions (Coyne & Orr 2004; but see also Rieseberg *et al.* 1999) causing abnormal gene regulation when mixing divergent regulatory elements into a common genetic background. For instance, Landry *et al.*

(2007b) showed how the regulation of coevolved *cis* regulatory regions and *trans* transcription factors could be disrupted and lead to phenotypic novelties in hybrids. Genome-wide analyses of the transcriptome confirmed that gene misexpression may underlie reproductive isolation mechanisms in F<sub>1</sub> hybrids between species that diverged many hundreds of thousands or millions of years ago (e.g. Ranz *et al.* 2004; Landry *et al.* 2007b). However, patterns of gene expression in young species pairs such as whitefish, as well as post-F<sub>1</sub> hybrid generations have been little explored, such that the underlying transcriptomic basis of reproductive isolation mechanisms remains largely unknown.

In recent studies using the 16 006 features salmon cDNA microarray, Renaut *et al.* (2009) contrasted gene expression divergence between pure normal and dwarf whitefish with that of first generation hybrids and second-generation backcrosses at the same embryonic and juvenile stages investigated by Nolte *et al.* (2009). The goal was to identify genes misexpressed in hybrids that could potentially link to abnormal phenotypic variation observed during embryonic development in previous studies. We found that very few transcripts (five of 4950 expressed or 0.1%) differed in mean expression level between pure forms and hybrids at the embryonic stage, in contrast to 16-week-old juvenile fish for which 617 out of 5359 transcripts (11%) differed significantly. Of particular interest were six key metabolic genes (glyceraldehyde-3-phosphate dehydrogenase, fructose-bisphosphate aldolase A, betaenolase, trypsin-1 precursor, cytochrome *c* oxidase polypeptide VIa and nucleoside diphosphate kinase) that were divergent between normal and dwarf whitefish at both juvenile and adult stages (Nolte *et al.* 2009). More specifically, we found that, in F<sub>1</sub>-hybrid juveniles, those genes mostly showed an intermediate pattern of expression compared with parents. An intermediate level of gene expression for those metabolic genes in hybrids could contribute to an atypical physiological phenotype and to an inferior, ecologically maladapted, individual. As such, reproductive isolation of lake whitefish could be seen as a by-product of divergent selection acting on metabolic genes, in accordance with previous studies showing that environmentally driven natural selection may be key in explaining incipient population divergence (e.g. Gow *et al.* 2007; van der Sluijs *et al.* 2008).

As illustrated in figure 7, we also found clear evidence for severe gene misexpression whereby non-additive gene interactions explained a large fraction of hybrid inheritance patterns in backcross (54%) compared with F<sub>1</sub>-hybrids (9%). Inter-individual variance in the level of gene expression was also approximately twice as important in hybrids relative to pure crosses. This translated into a pronounced transgressive segregation of transcription variation, especially in backcross hybrids for which the expression of 2622 (embryos) and 2316 (juveniles) genes exhibited exaggerated variance, where the range of backcross expression extended outside the range of both parents values. Such a high number of transgressive genes significantly exceeded that

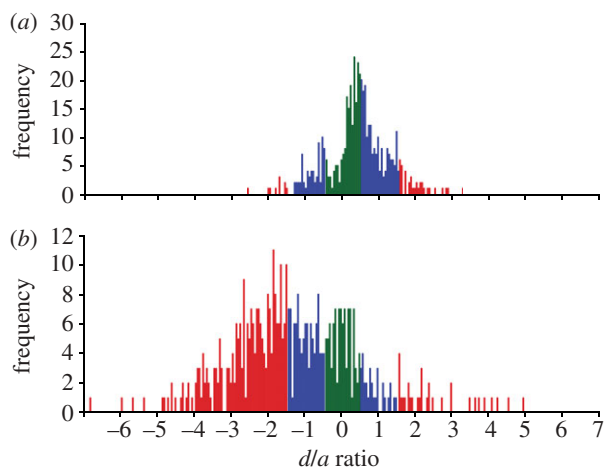


Figure 7. Distribution of dominance effects ( $d/a$  ratio) for (a)  $F_1$ -hybrids and (b) backcrosses for 501 transcripts significantly different between pure dwarf and normal whitefish. A  $|d/a|$  ratio between 0.5 and 1.5 is considered as dominant,  $>1.5$  non-additive,  $<0.5$  additive (modified from Renaut *et al.* 2009) (red bar, non-additivity; blue bar, dominance; green bar; additivity).

expected by chance alone (Renaut *et al.* 2009). In particular, five of the nine most transgressive transcripts closely matched with three different homologues (Immunoglobulin binding protein (protein folding); translation elongation factor alpha 1 (mRNA translation) and 40S ribosomal protein s11 (mRNA translation)) identified as essential for early embryonic development of zebrafish, *Danio rerio* (Amsterdam *et al.* 2004). Knockdown mutants in *D. rerio* for those genes show visible embryonic defect and almost invariably die prior to, or soon after hatching. Transgressive segregation may underlie post-zygotic isolation mechanisms given that transgressive hybrids often suffer a highly reduced survival (Barton 2001; Coyne & Orr 2004). Therefore, it is plausible that the overall patterns of transgressivity we observed, and particularly misexpression of several key developmental genes, may contribute to abnormal hybrid development and increased embryonic mortality identified as a significant intrinsic post-zygotic reproductive isolation mechanism between dwarf and normal whitefish (Lu & Bernatchez 1998; Rogers & Bernatchez 2006).

## 7. SUMMING UP THE EVIDENCE FOR THE CASE OF NATURAL SELECTION DRIVING ADAPTIVE DIVERGENCE AND REPRODUCTIVE ISOLATION IN WHITEFISH

Integrating a combination of multiple ecological and genomic approaches under the conceptual framework offered by the theory of adaptive radiation has proven highly efficient for highlighting the role of natural selection in driving the adaptive divergence and reproductive isolation between dwarf and normal whitefish, as has yield substantial insight into the genomic basis of these evolutionary processes.

First, combining information obtained both from a comparison of phenotypic and genome wide transcriptomic differentiation identified the phenotypic

(including physiological) traits most likely to be adaptive and therefore of most pertinent focus for further investigation of the genetic basis of adaptive divergence. These data also provided strong support for the mechanistic hypothesis that the adaptive divergence and evolution of distinct life-history strategies between dwarf and normal whitefish involve differential trade-offs between growth (and correlated fecundity) and survival which, in turn, is mediated through the higher energetic cost of occupying the limnetic relative to the benthic trophic niche. Thus, higher metabolic rate, more active swimming associated with planktonic feeding and predator avoidance, as well as reduced bioenergetic conversion efficiency may constrain available energy for growth, fecundity and reproduction at older ages in dwarf whitefish. In addition, the analysis of transcriptomic divergence between dwarf and normal whitefish revealed that their differentiation involves at least several hundred genes (conservatively between 1 and 3 per cent of the genome), thus pointing out that it may be misleading to draw general conclusions about the genetic basis of adaptive divergence based on the investigation of single or few candidate genes. Moreover, parallelism in patterns of gene transcription, as well as patterns of inter-individual variance in expression, provided strong evidence for the role of natural selection in the evolution of differential regulation of genes involving a vast array of physiological processes. This provided a mechanistic, genomic basis for the observed trade-off in life-history traits distinguishing dwarf and normal whitefish. Finally, the high number of differentially expressed genes at the juvenile and adult stages may provide a broad genomic basis for extrinsic post-zygotic isolation (Nosil *et al.* 2009a,b).

Second, the combined, integrated use of linkage, phenotypic and gene expression mapping provided an efficient means to elucidate the genetic architecture underlying the most probable adaptive phenotypic traits differentiating dwarf and normal whitefish. It also revealed several associations between specific gene and phenotypic expression and identified key genomic regions potentially harbouring genes (master regulators) of high pleiotropic effects on gene expression. The prevalence of ubiquitous directionality in additive affects was consistent with the action of natural selection on single master regulators controlling expression of a suite of numerous genes, which belong to diverse functional groups involved in the adaptive divergence of whitefish and ultimately modulate the organismal phenotype. Finally, combining pQTL and eQTL mapping provided the identity of numerous candidate markers for which the influence of natural selection in wild populations can be methodically tested.

Third, integrating pQTL and eQTL mapping with population genomics of multiple pairs of dwarf and normal whitefish efficiently identified genomic regions resisting the homogenizing influence of gene flow and therefore likely to be under divergent selection. Thus, the AFLP genome scan indicated that a relatively small proportion of the genome (approximately 1–2%, but still representing hundreds of loci over the whole genome) might be linked to genes implicated in the

adaptive divergence and reproductive isolation of dwarf and normal whitefish. The majority (19 out of 24 or 80%) of 'anonymous' AFLP outliers could be associated with either a pQTL, eQTL or both. This showed that outliers were mainly associated with either growth, swimming behaviour phenotypes or expression of genes associated with these functions (protein synthesis and energy metabolism). Most eQTL outliers also corroborated previously identified candidate genes on the basis of parallelism in transcription. These observations provided compelling evidence for the role of natural selection in restricting effective gene flow in regions modulating the expression of genes underlying contrasting whitefish life-history strategies that evolved for adapting to distinct environments and resources.

Fourth, combining experimental studies, pQTL and eQTL mapping with a comparative analysis of genome-wide transcription patterns identified several mechanisms of reproductive isolation and provided insight into their genetic basis. This first revealed that genetically based asynchronous emergence resulting from the admixture of distinct genomes in hybrids provides an underlying basis for extrinsic, ecologically based post-zygotic isolation. This phenotype would most certainly be selectively unfavourable in nature, since asynchronous emergence has been commonly correlated to increased risk of starvation, decreased opportunities for optimal growth and increased risk of predation (e.g. Cushing 1990). Furthermore, we demonstrated that highly reduced sperm performance and survival during embryonic development in back-cross hybrids translated into intrinsic post-zygotic isolation. Incompatibility in specific genomic regions, perhaps initiated during divergence in allopatry and reinforced by subsequent sympatric ecological specialization of dwarf and normal populations most probably underlies this intrinsic post-zygotic isolation mechanism. These observations also raised the hypothesis that hybrid inviability is associated with misregulation of genes with essential functions. This hypothesis was confirmed by genome-wide analysis of the transcriptome which revealed abnormal patterns of expression at numerous genes in hybrids, including key genes involved in embryonic development as well as genes playing central roles in energy metabolism and involved in the adaptive divergence of dwarf and normal whitefish. Gene misexpression in hybrids thus adds to other intrinsic and extrinsic reproductive isolation factors between dwarf and normal whitefish.

Overall, evidence for reduced sperm performance, and genetically based increased embryonic mortality in hybrids followed by asynchronous emergence indicates that intrinsic and extrinsic mechanisms of reproductive isolation are not mutually exclusive; both processes are jointly involved in the ongoing speciation process in whitefish. Moreover, since we substantiated that geographical isolation as well as ecological divergence that subsequently occurred in sympatry have contributed to the origins of these forms, it indicates that both processes of ecological and non-ecological speciation (*sensu* Rundell & Price 2009), such as mutation-order speciation, defined as the evolution of reproductive isolation by the chance

occurrence and fixation of different alleles (Schluter 2009), have been involved in the origin of dwarf and normal whitefish in eastern North America. However, this may not be generalized to all cases of sympatric divergence in *Coregonus* since there is no evidence for an initial phase of allopatry for Eurasian sympatric whitefish populations despite similar and in some cases, more pronounced patterns of phenotypic divergence observed among those (Douglas *et al.* 1999; Østbye *et al.* 2006).

## 8. FUTURE DIRECTIONS: STEPPING INTO THE NEXT-GENERATION SEQUENCING ERA

The combined use of gene mapping, population genomics and transcriptomics allowed the identification of nearly 500 genes representing a vast array of physiological functions that are probably candidates for the adaptive divergence and reproductive isolation of dwarf and normal whitefish (see the electronic supplementary material, table S1). The precise role and implication of these genes still need to be confirmed in whitefish, which will be the main focus of our research programme in the following years. Indeed, the identification of candidate genes does not constitute an end in itself, but rather the beginning of a new set of evolutionary relevant questions (Stinchcombe & Hoekstra 2008). In the specific case of the whitefish research programme, some of the key questions or issues to be addressed in future studies are: What are the size, distribution and nature of genomic islands of speciation, that is chromosomal regions that remain highly differentiated between populations despite ongoing gene flow (Turner *et al.* 2005)? This issue is of particular interest since recent studies suggested that surprisingly large genomic regions around divergently selected loci can be protected from recombination by divergence hitchhiking (Via 2009). If proven general, this process could neutralize one of the longstanding criticism of sympatric isolation, that is; the difficulty of maintaining linkage disequilibrium between genes involved in ecological divergence and those causing assortative mating (Coyne & Orr 2004). In the same way that it is invoked in the process of ecological speciation between dwarf and normal whitefish, can natural selection be strong enough to maintain species cohesiveness by favouring the spread of alleles that are advantageous within a same form (dwarf or normal; Morjan & Rieseberg 2004)? Are adaptations to similar ecological niches that evolved independently owing to the same mutations (Linnen *et al.* 2009)? Do alleles conferring their specificity to dwarf and normal whitefish emerge from standing genetic variation or new mutations? What is the role of positive selection in maintaining functional divergence at the sequence level of those genes? What are the relative contributions of regulatory and structural (protein) changes to the speciation process of whitefish? What is the relative role of genomic mechanisms (e.g. *cis*- versus *trans*- acting regulation, allele-specific differential expression, differential alternative splicing, expression of different isoforms, epigenetic interactions) in the differential pattern of transcription we

observed at hundreds of genes? Transcriptomics studies showed that energy metabolism is the main biological function involved in the divergence between dwarf and normal whitefish. Given the fundamental role of mitochondria as the main source of cellular energy production which involves networks of interactions between nuclear (slow evolving) and mitochondria (fast evolving) encoded genes, and since historical isolation led to the divergence of the mitochondrial genome of the two glacial races of whitefish (Bernatchez & Dodson 1990; Lu *et al.* 2001), can breakdown or mis-regulation of mitochondrial bioenergetics functions in hybrids be a major player in the speciation process of dwarf and normal whitefish, as revealed recently in other systems (Ellison & Burton 2008; Gershoni *et al.* 2009)? There is mounting evidence that selection has been acting more strongly on dwarf than normal whitefish (Bernatchez 2004; Derome *et al.* 2006; St-Cyr *et al.* 2008). Given the pronounced transcriptomic divergence observed between them, does natural selection favour canalization (defined as the evolved ability to produce a consistent expression of a same phenotype in different environments) more importantly in dwarf than normal whitefish (Landry 2009)? Similarly, is canalization of expression more pronounced for genes identified as candidates for the adaptive divergence of whitefish than other, 'non-candidate' genes?

Addressing those issues will necessitate the continued application of the same strategy we have used over the last years. That is, combining different research approaches and targeting various functional and biological levels, including the integration of the most recent technological developments. Namely, next generation sequencing technologies, which are rapidly revolutionizing evolutionary biology research will be of paramount interest (Rokas & Abbot 2009). These approaches have already showed their efficiency in various applications, including rapid SNP (single nucleotide polymorphism) discovery (Barbazuk *et al.* 2007), genome sequencing of ecologically important models (Vera *et al.* 2008) as well as accurate gene expression analysis (Torres *et al.* 2007; Lipson *et al.* 2009). High throughput pyrosequencing (454 Life Sciences, Margulies *et al.* 2005) is particularly relevant to the study of non-model organisms such as whitefish since it yields longer sequencing reads than any other method, allowing more accurate de novo sequence assemblies in the absence of reference genomes.

As an illustration of its usefulness to our research programme, we recently performed a 454 pyrosequencing analysis, which allowed assembling, de novo, over 130 mb of non-normalized cDNA using data from dwarf and normal whitefish (Renaut *et al.* 2010). Our main objectives with this first 'next-gen' sequencing application was to gather a large dataset of SNP markers to be used in further mapping and genome scan studies of candidate genes and to gain further insight into transcriptional divergence between dwarf and normal whitefish, which included quantifying allele-specific expression, something not possible with previous cDNA microarray analyses. We identified 6094 putative SNPs in 2674 contigs (mean size: 576 bp, range (101–6000 bp)), comprising 1540

synonymous and 1734 non-synonymous mutations for a genome-wide non-synonymous to synonymous substitution rate ratio ( $p_N/p_S$ ) of 0.37. About 90 SNPs that showed highly significant divergence in allele frequencies between normal and dwarf whitefish were identified. Among those, SNPs in genes annotated to energy metabolic and protein synthesis functions were the most abundant. This corroborated previous evidence that genes involved in these two functions are prime candidates underlying the adaptive divergence of normal and dwarf whitefish.

One example of such candidate SNP was found in a Triosephosphate isomerase gene which plays an important role in glycolysis and is essential for efficient energy production energy metabolism. This gene was screened between dwarf and normal whitefish from Cliff Lake by means of MALDI-TOF mass spectrometry (Sequenom MassARRAY system, Ehrich *et al.* 2005; Nolte *et al.* 2009, unpublished results; figure 8). The  $F_{st}$  value between dwarf and normal for this gene was 0.685, making it an outlier in the distribution of  $F_{st}$  values observed in the AFLP genome scan (Campbell & Bernatchez 2004). At the same time, this gene was previously identified as differentially expressed between dwarf and normal whitefish (Nolte *et al.* 2009). Furthermore, an AFLP marker (CCTC051), itself an outlier with a  $F_{st}$  value of 0.871, was linked to an eQTL for Triosephosphate isomerase on linkage group 19 (Derome *et al.* 2008). This eQTL was located within an eQTL hotspot comprising other genes involved in various functions, including protein synthesis (growth), muscle contraction, immune system as well as a putative transcription factor (Zinc finger protein 35; figure 8). Although preliminary and requiring rigorous confirmation, these results suggest that the genomic region on linkage group 19 may correspond to an important genomic island of divergence that is maintained by natural selection despite ongoing gene flow between sympatric dwarf and normal whitefish. These results also suggest that both structural and regulatory divergence may be involved in this adaptive differentiation. This example clearly illustrates the potential and efficiency offered by the combined use of the latest genomic approaches in elucidating the genomic bases of adaptive divergence and reproductive isolation in a non-model organism, as well as deciphering the role of natural selection in driving these evolutionary processes.

Quite clearly, we are witnessing an exciting era during which many gaps of knowledge about mechanisms linking genes and selection during the course of speciation will be filled and which in turn, will further strengthen Darwin's theory of evolution by means of natural selection. Undoubtedly, non-model species, such as whitefish, studied in their ecological context will have an increasingly important role to play in expanding our knowledge in this field. There have never been more stimulating times to train the 'next-gen' of young evolutionary biologists in ecological genomics.

We thank Hans Ellegren for kindly conveying the senior author to contribute to this special issue, Eric Taylor for his constructive and most useful comments, and Eric Normandeau for his skilful text, table and figure editing. This research programme on whitefish ecological genomics

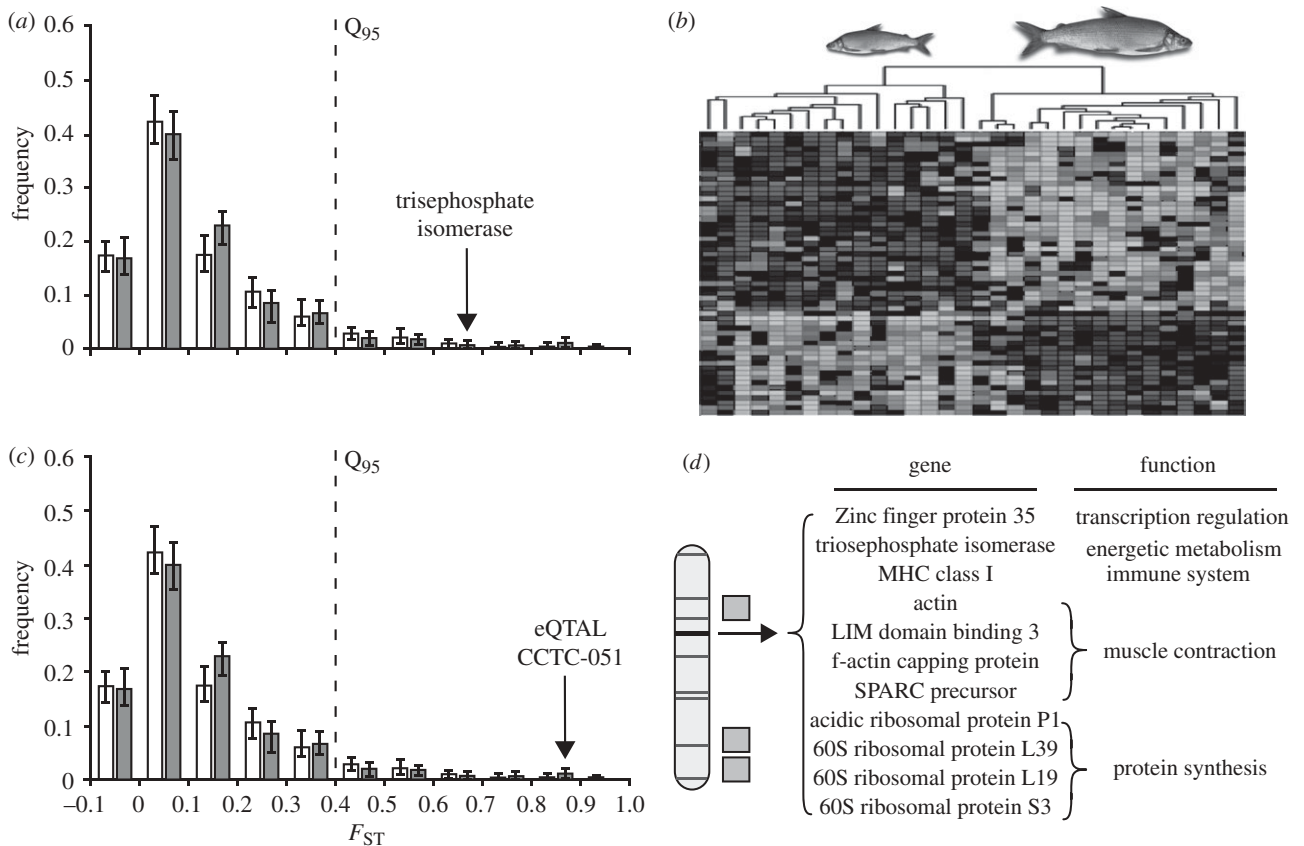


Figure 8. Position of  $F_{ST}$  value ( $F_{ST} = 0.658$ ) for a SNP of the Triosephosphate isomerase gene between dwarf and normal whitefish in Cliff Lake (open bar, simulated; filled bar, mapped loci) (a), which was also differentially expressed between dwarf and normal whitefish (b). Position of  $F_{ST}$  value ( $F_{ST} = 0.871$ ) for AFLP marker CCTC-051 between dwarf and normal whitefish which was linked to an eQTL for Triosephosphate isomerase (c), itself located within an eQTL hotspot on linkage group 19 comprising other genes with various physiological functions (d). AFLP Genome scan data are from Campbell & Bernatchez (2004), and those from eQTL mapping are from Derome *et al.* (2008).

has been supported by grants to L.B. from the Natural Sciences and Engineering Research Council of Canada (NSERC), a E.W.R. Steacie Fellowships (NSERC) as well as by the Canadian Research Chair in Genomics and Conservation of Aquatic Resources.

## REFERENCES

- Amsterdam, A., Nissen, R. M., Sun, Z., Swindell, E. C., Farrington, S. & Hopkins, N. 2004 Identification of 315 genes essential for early zebrafish development. *Proc. Natl Acad. Sci. USA* **101**, 12 792–12 797. (doi:10.1073/pnas.0403929101)
- Amundsen, P.-A., Bøhn, T. & Vågå, G. H. 2004 Gill raker morphology and feeding ecology of two sympatric morphs of European whitefish (*Coregonus lavaretus*). *Ann. Zool. Fenn.* **41**, 291–300.
- Barbazuk, W. D., Emrich, S. J., Chen, H. D., Li, L. & Schanble, P. S. 2007 SNP discovery via 454 transcriptome sequencing. *Plant J.* **51**, 910–918. (doi:10.1111/j.1365-313X.2007.03193.x)
- Barrett, R. D. H., Rogers, S. M. & Schluter, D. 2008 Threespine stickleback natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255–257. (doi:10.1126/science.1159978)
- Barton, N. H. 2001 The role of hybridization in evolution. *Mol. Ecol.* **10**, 551–568. (doi:10.1046/j.1365-294x.2001.01216.x)
- Beaumont, M. A. & Balding, D. J. 2004 Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **13**, 969–980. (doi:10.1111/j.1365-294X.2004.02125.x)
- Beaumont, M. A. & Nichols, R. A. 1996 Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B* **263**, 1619–1626. (doi:10.1098/rspb.1996.0237)
- Bernatchez, L. 2004 Ecological theory of adaptive radiation: an empirical assessment from coregonine fishes (Salmoniformes). In *Evolution illuminated: salmon and their relatives* (eds A. P. Hendry & S. C. Stearns), pp. 175–207. Oxford, UK: Oxford University Press.
- Bernatchez, L. & Dodson, J. J. 1990 Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial DNA restriction analysis. *Evolution* **44**, 1263–1271. (doi:10.2307/2409287)
- Bernatchez, L. & Dodson, J. J. 1991 Phylogeographic structure in mitochondrial DNA of the Lake Whitefish (*Coregonus clupeaformis*) in North America and its relationships to Pleistocene glaciations. *Evolution* **45**, 1016–1035. (doi:10.2307/2409706)
- Bernatchez, L. & Tseng, M. 2008 Editorial. *Evol. Appl.* **1**, 1–2. (doi:10.1111/j.1752-4571.2008.00016.x)
- Bernatchez, L., Chouinard, A. & Lu, G. 1999 Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, *Coregonus*, as a case study. *Biol. J. Linn. Soc.* **68**, 173–194. (doi:10.1111/j.1095-8312.1999.tb01165.x)
- Bochdanovits, Z., van der Klis, H. & de Jong, G. 2003 Covariation of larval gene expression and adult body size in natural populations of *Drosophila melanogaster*.

- Mol. Biol. Evol.* **20**, 1760–1766. (doi:10.1093/molbev/msg179)
- Britten, R. J. & Davidson, E. H. 1971 Repetitive and non-repetitive DNA sequences and a speculation on origins of evolutionary novelty. *Q. Rev. Biol.* **46**, 111–138. (doi:10.1086/406830)
- 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. *Mol. Biol. Evol.* **21**, 945–956. (doi:10.1093/molbev/msh101)
- Chouinard, A. & Bernatchez, L. 1998 A study of trophic niche partitioning between larval populations of reproductively isolated whitefish (*Coregonus* sp.) ecotypes. *J. Fish Biol.* **53**, 1231–1242.
- Coyne, J. A. & Orr, H. A. 2004 *Speciation*. Sunderland, MA: Sinauer Associates.
- Cushing, D. H. 1990 Plankton production and year class strength in fish populations—an update of the match mismatch hypothesis. *Adv. Mar. Biol.* **26**, 249–293. (doi:10.1016/S0065-2881(08)60202-3)
- Darwin, C. 1859 *On the origin of species by means of natural selection*. London, UK: John Murray.
- Derome, N. & Bernatchez, L. 2006 The transcriptomics of ecological convergence between two limnetic coregonine fishes (*Salmonidae*). *Mol. Biol. Evol.* **23**, 2370–2378. (doi:10.1093/molbev/msl110)
- Derome, N., Duchesne, P. & Bernatchez, L. 2006 Parallelism in gene transcription among sympatric lake whitefish ecotypes (*Coregonus clupeaformis* Mitchell). *Mol. Ecol.* **15**, 1239–1250. (doi:10.1111/j.1365-294X.2005.02968.x)
- Derome, N., Bougas, B., Rogers, S. M., Whiteley, A. & Bernatchez, L. 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. (doi:10.1534/genetics.107.086306)
- Douglas, M. R., Brunner, P. C. & Bernatchez, L. 1999 Do assemblages of *Coregonus* (Teleostei: Salmoniformes) in the Central Alpine region of Europe represent species flocks? *Mol. Ecol.* **8**, 589–603. (doi:10.1046/j.1365-294x.1999.00581.x)
- Ehrich, M., Bocker, S. & van den Boom, D. 2005 Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS. *Nucleic Acids Res.* **33**, e38. (doi:10.1093/nar/gni038)
- Ellison, C. K. & Burton, R. S. 2008 Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. *Proc. Natl Acad. Sci. USA* **105**, 15 831–15 836. (doi:10.1073/pnas.0804253105)
- Fay, J. C. & Wittkopp, P. J. 2008 Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* **100**, 191–199. (doi:10.1038/sj.hdy.6801000)
- Fenderson, O. C. 1964 Evidence of subpopulations of lake whitefish, *Coregonus clupeaformis*, involving a dwarfed form. *Trans. Am. Fish. Soc.* **93**, 77–94. (doi:10.1577/1548-8659(1964)93[77:EOSOLW]2.0.CO;2)
- Foll, M. & Gaggiotti, O. 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993. (doi:10.1534/genetics.108.092221)
- Gershoni, M., Templeton, A. R. & Mishmar, D. 2009 Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* **31**, 642–650. (doi:10.1002/bies.200800139)
- Gibson, G. & Weir, B. 2005 The quantitative genetics of transcription. *Trends Genet.* **21**, 616–623. (doi:10.1016/j.tig.2005.08.010)
- Gilad, Y., Rifkin, S. A. & Pritchard, J. K. 2008 Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408–415. (doi:10.1016/j.tig.2008.06.001)
- Gow, J. L., Peichel, C. L. & Taylor, E. B. 2007 Ecological selection against hybrids in natural populations of sympatric threespine sticklebacks. *J. Evol. Biol.* **20**, 2173–2180. (doi:10.1111/j.1420-9101.2007.01427.x)
- Harrod, C. & Kahilainen, K. 2006 Multiple tissue stable isotope analysis show resource segregation between whitefish morphs is maintained across different temporal scales. *J. Fish Biol. Suppl. C* **69**, 249–250.
- Harvey, P. H. & Pagel, M. D. 1991 *The comparative method in evolutionary biology*. Oxford, UK: Oxford University Press.
- Hoekstra, H. E., Hirschmann, R. J., Bunday, R. A., Insel, P. A. & Crossland, J. P. 2006 A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101–104. (doi:10.1126/science.1126121)
- Hudson, A. G., Vonlanthen, P., Mueller, R. & Seehausen, O. 2007 Review: the geography of speciation and adaptive radiation in coregonines. Biology and management of coregonid fishes. *Adv. Limnol.* **60**, 111–146.
- Jeukens, J., Bittner, D., Knudsen, R. & Bernatchez, L. 2009 Candidate genes and adaptive radiation: insights from transcriptional adaptation to the limnetic niche among coregonine fishes (*Coregonus* sp., *Salmonidae*). *Mol. Biol. Evol.* **26**, 155–166. (doi:10.1093/molbev/msn235)
- Joron, M. *et al.* 2006 A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol.* **4**, e303. (doi:10.1371/journal.pbio.0040303)
- 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 a subarctic Lake Muddusjarvi. *Ecol. Freshw. Fish* **11**, 158–167. (doi:10.1034/j.1600-0633.2002.t01-2-00001.x)
- Kahilainen, K. K., Malinen, T., Tuomaala, A., Alajaervi, E., Tolonen, A. & Lehtonen, H. 2007 Empirical evaluation of phenotype–environment correlation and trait utility with allopatric and sympatric whitefish, *Coregonus lavaretus* (L.), populations in subarctic lakes. *Biol. J. Linn. Soc.* **92**, 561–572. (doi:10.1111/j.1095-8312.2007.00856.x)
- Kane, N. C., King, M. G., Barker, M. S., Raduski, A., Karrenberg, S., Yatabe, Y., Knapp, S. J. & Rieseberg, L. H. 2009 Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution* **63**, 2061–2075. (doi:10.1111/j.1558-5646.2009.00703.x)
- King, M. C. & Wilson, A. C. 1975 Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116. (doi:10.1126/science.1090005)
- Kottelat, M. & Freyhof, J. 2007 *Handbook of European freshwater fishes*. Cornol, Switzerland: Publications Kottelat.
- Landry, C. R. 2009 Systems biology spins off a new model for the study of canalization. *Trends Ecol. Evol.* **24**, 64–66.
- Landry, L., Vincent, W. F. & Bernatchez, L. 2007a Parallel evolution of lake white fish dwarf ecotypes in association with limnological features of their adaptive landscape. *J. Evol. Biol.* **20**, 971–984. (doi:10.1111/j.1420-9101.2007.01304.x)
- Landry, C. R., Hartl, D. L. & Ranz, J. 2007b Genome clashes in hybrids: insights from gene expression. *Heredity* **99**, 483–493. (doi:10.1038/sj.hdy.6801045)
- Linnen, C. R., Kingsley, E. P., Jensen, J. D. & Hoekstra, H. E. 2009 On the origin and spread of an adaptive allele in deer mice. *Science* **325**, 1095–1098. (doi:10.1126/science.1175826)



- Lipson, D. *et al.* 2009 Quantification of the yeast transcriptome by single-molecule sequencing. *Nat. Biotechnol.* **27**, 652–659. (doi:10.1038/nbt.1551)
- Lu, G. & Bernatchez, L. 1998 Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of lake whitefish (*Coregonus clupeaformis* Mitchell). *Proc. R. Soc. Lond. B* **265**, 1025–1030. (doi:10.1098/rspb.1998.0394)
- 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. (doi:10.2307/2640895)
- Lu, G., Basley, D. J. & Bernatchez, L. 2001 Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*): relevance for speciation. *Mol. Ecol.* **10**, 965–985. (doi:10.1046/j.1365-294X.2001.01252.x)
- Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. 2009 The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* **10**, 565–577. (doi:10.1038/nrg2612)
- Margulies, M. *et al.* 2005 Genome sequencing in microfabricated high-density picoliter reactors. *Nature* **437**, 376–380. (doi:10.1038/nature03959)
- Matzkin, L. M., Watts, T. D., Bitler, B. G., Machado, C. A. & Markow, T. A. 2006 Functional genomics of cactus host shifts in *Drosophila mojavensis*. *Mol. Ecol.* **15**, 4635–4643.
- Mitchell-Olds, T., Willis, J. H. & Goldstein, D. B. 2007 Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat. Rev. Genet.* **8**, 845–856. (doi:10.1038/nrg2207)
- Morjan, C. L. & Rieseberg, L. H. 2004 How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol. Ecol.* **13**, 1341–1356. (doi:10.1111/j.1365-294X.2004.02164.x)
- Nolte, A. W., Renaud, S. & Bernatchez, L. 2009 Divergence in gene regulation at young life history stages of whitefish (*Coregonus* sp.) and the emergence of genomic isolation. *BMC Evol. Biol.* **9**, 59. (doi:10.1186/1471-2148-9-59)
- Noor, M. A. F. & Feder, J. L. 2006 Speciation genetics: evolving approaches. *Nat. Rev. Genet.* **7**, 851–861. (doi:10.1038/nrg1968)
- Nosil, P., Funk, D. J. & Ortiz-Barrientos, D. 2009a Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**, 375–402. (doi:10.1111/j.1365-294X.2008.03946.x)
- Nosil, P., Harmon, L. J. & Seehausen, O. 2009b Ecological explanations for (incomplete) speciation. *Trends Ecol. Evol.* **24**, 145–156. (doi:10.1016/j.tree.2008.10.011)
- Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. 2002 Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261–266. (doi:10.1038/ng983)
- Orr, H. A. 1998 Testing natural selection versus genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* **149**, 2099–2104.
- Orr, H. A. 2005 The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–127. (doi:10.1038/nrg1523)
- Orr, H. A. & Turelli, M. 2001 The evolution of postzygotic isolation: accumulating Dobzhansky–Muller incompatibilities. *Evolution* **55**, 1085–1094.
- Østbye, K., Bernatchez, L., Naesje, T. F., Himberg, M. & Hindar, K. 2005a Evolutionary history of European whitefish (*Coregonus lavaretus* L.) as inferred from mtDNA phylogeography and gill-raker numbers. *Mol. Ecol.* **14**, 4371–4388.
- Østbye, K., Naesje, T. F., Bernatchez, L., Sandlund, O. T. & Hindar, K. 2005b Morphological divergence and origin of sympatric populations of European whitefish (*Coregonus lavaretus* L.) in Lake Femund, Norway. *J. Evol. Biol.* **18**, 683–702. (doi:10.1111/j.1420-9101.2004.00844.x)
- Østbye, K., Amundsen, P.-A., Bernatchez, L., Klemetsen, A., Knudsen, R., Kristoffersen, R., Naesje, T. F. & Hindar, K. 2006 Parallel evolution of eco-morphological traits in European whitefish *Coregonus lavaretus* (L.) during postglacial times. *Mol. Ecol.* **15**, 3983–4001. (doi:10.1111/j.1365-294X.2006.03062.x)
- 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. (doi:10.2307/2410973)
- Ranz, J. M. & Machado, C. A. 2005 Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol.* **21**, 29–37. (doi:10.1016/j.tree.2005.09.002)
- Ranz, J. M., Namgyal, K., Gibson, G. & Hartl, D. 2004 Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res.* **14**, 373–379. (doi:10.1101/gr.2019804)
- Renaud, S., Nolte, A. W. & Bernatchez, L. 2009 Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol. Biol. Evol.* **26**, 925–936. (doi:10.1093/molbev/msp017)
- Renaud, S., Nolte, A. W. & Bernatchez, L. 2010 Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* sp.). *Mol. Ecol.* **19**(Suppl. 1), 115–131.
- Reshetnikov, J. S. 1988 Coregonid fishes in recent conditions. *Finn. Fish. Res.* **9**, 11–16.
- Rice, W. R. & Hostert, E. E. 1993 Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* **47**, 1637–1653. (doi:10.2307/2410209)
- Rieseberg, L. H. 1998 Genetic mapping as a tool for studying speciation. *Molecular systematics of plants 2* (eds D. E. Soltis, P. S. Soltis & J. J. Doyle), pp. 459–787. New York, NY: Chapman & Hall.
- Rieseberg, L. H., 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.
- Rise, M. L. *et al.* 2004 Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Res.* **14**, 478–490. (doi:10.1101/gr.1687304)
- Rise, M. L., von Schalburg, K. R., Cooper, G. A. & Koop, B. F. 2007 Salmonid DNA microarrays and other tools for functional genomics research. In *Aquaculture Genome Technologies* (ed. Z. Liu), pp. 369–411. Oxford, UK: Blackwell Publishing.
- Roff, D. A. 2007 Contributions of genomics to life-history theory. *Nat. Rev. Genet.* **8**, 116–125. (doi:10.1038/nrg2040)
- Rogers, S. M. & Bernatchez, L. 2005 Integrating QTL mapping and genomic scans towards the characterization of candidate loci under parallel directional selection in the lake whitefish (*Coregonus clupeaformis*). *Mol. Ecol.* **14**, 351–361. (doi:10.1111/j.1365-294X.2004.02396.x)
- Rogers, S. M. & Bernatchez, L. 2006 The genetic basis of intrinsic and extrinsic postzygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus Clupeaformis*). *J. Evol. Biol.* **19**, 1979–1994. (doi:10.1111/j.1420-9101.2006.01150.x)
- Rogers, S. M. & 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. *Mol. Biol. Evol.* **24**, 1423–1438. (doi:10.1093/molbev/msm066)

- Rogers, S. M., Campbell, D., Baird, S., Danzmann, R. G. & Bernatchez, L. 2001 Combining the analyses of introgressive hybridization and linkage mapping to investigate the genetic architecture of population divergence in the lake whitefish (*Coregonus clupeaformis*, Mitchill). *Genetica* **111**, 25–41. (doi:10.1023/A:1013773600304)
- Rogers, M. R., 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, S. M., 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.
- Rokas, A. & Abbot, P. 2009 Harnessing genomics for evolutionary insights. *Trends Ecol. Evol.* **24**, 192–200. (doi:10.1016/j.tree.2008.11.004)
- Rundell, R. J. & Price, T. D. 2009 Adaptive radiation, non-adaptive radiation, ecological speciation and nonecological speciation. *Trends Ecol. Evol.* **24**, 394–399. (doi:10.1016/j.tree.2009.02.007)
- Schemske, D. W. & Bierzychudek, P. 2007 Spatial differentiation for flower color in the desert annual *Linanthus parryae*: was Wright right? *Evolution* **61**, 2528–2543. (doi:10.1111/j.1558-5646.2007.00219.x)
- Schluter, D. 2000 *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.
- Schluter, D. 2009 Evidence for ecological speciation and its alternative. *Science* **323**, 737–741. (doi:10.1126/science.1160006)
- Spitze, K. 1993 Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* **135**, 367–374.
- St-Cyr, J., Derome, N. & Bernatchez, L. 2008 The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Mol. Ecol.* **17**, 1850–1870. (doi:10.1111/j.1365-294X.2008.03696.x)
- Stinchcombe, J. R. & Hoekstra, H. E. 2008 Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158–170. (doi:10.1038/sj.hdy.6800937)
- Storz, J. F., Sabatino, S. J., Hoffmann, F. G., Gering, E. J., Moriyama, H., Ferrand, N., Monteiro, B. & Nachman, M. W. 2007 The molecular basis of high-altitude adaptation in deer mice. *PLoS Genet.* **3**, 448–459.
- Torres, T., Metta, M., Ottenwalder, B. & Schlotterer, C. 2007 Gene expression profiling by massively parallel sequencing. *Genome Res.* **18**, 172–177. (doi:10.1101/gr.6984908)
- 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*). *Can. J. Fish. Aquat. Sci.* **58**, 394–405. (doi:10.1139/cjfas-58-2-394)
- Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. 2005 Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* **3**, e285. (doi:10.1371/journal.pbio.0030285)
- van der Sluijs, I., van Dooren, T. J. M., Seehausen, O. & van Alphen, J. J. M. 2008 A test of fitness consequences of hybridization in sibling species of Lake Victoria cichlid fish. *J. Evol. Biol.* **21**, 480–491. (doi:10.1111/j.1420-9101.2007.01495.x)
- Vasemagi, A. & Primmer, C. R. 2005 Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol. Ecol.* **14**, 3623–3642. (doi:10.1111/j.1365-294X.2005.02690.x)
- Vera, J. C., Wheat, C. W., Fescemyer, H. W., Frilander, M. J., Crawford, D. L., Hnski, I. & Marden, J. H. 2008 Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol. Ecol.* **17**, 1636–1647. (doi:10.1111/j.1365-294X.2008.03666.x)
- Via, S. 2009 Natural selection in action during speciation. *Proc. Natl Acad. Sci. USA* **106**, 9939–9946. (doi:10.1073/pnas.0901397106)
- Via, S. & West, J. 2008 The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* **17**, 4334–4345. (doi:10.1111/j.1365-294X.2008.03921.x)
- Vogl, C. & Xu, S. Z. 2000 Multipoint mapping of viability and segregation distorting loci using molecular markers. *Genetics* **155**, 1439–1447.
- Vonlanthen, P., Roy, D., Hudson, A. G., Largiader, C. R., Bittner, D. & Seehausen, O. 2009 Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *J. Evol. Biol.* **22**, 498–514. (doi:10.1111/j.1420-9101.2008.01670.x)
- von Schalburg, K. R., Rise, M. L., Cooper, G. A., Brown, G. D., Gibbs, A. R., Nelson, C. C., Davidson, W. S. & Koop, B. F. 2005 Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics* **6**, 126. (doi:10.1186/1471-2164-6-126)
- Wentzell, A. M., Rowe, H. C., Hansen, B. G., Ticconi, C., Halkier, B. A. & Kliebenstein, D. J. 2007 Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genet.* **3**, 1687–1701.
- Whiteley, A. R. *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. (doi:10.1534/genetics.108.089938)
- Whiteley, A. R., Persaud, K. N., Derome, N., Montgomerie, R. & Bernatchez, L. 2009 Reduced sperm performance in backcross hybrids between species pairs of whitefish (*Coregonus clupeaformis*). *Can. J. Zool.* **87**, 566–572. (doi:10.1139/Z09-042)
- Wittkopp, P. J. 2007 Variable gene expression in eukaryotes: a network perspective. *J. Exp. Biol.* **210**, 1567–1575. (doi:10.1242/jeb.002592)
- Wray, G. A. 2007 The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216. (doi:10.1038/nrg2063)
- Wu, C. I. & Ting, C. T. 2004 Genes and speciation. *Nat. Rev. Genet.* **5**, 114–122. (doi:10.1038/nrg1269)