Gene Coexpression Networks Reveal Key Drivers of Phenotypic Divergence in Lake Whitefish

Marie Filteau,[1] Scott A. Pavey,[1] Jérôme St-Cyr,[1] and Louis Bernatchez*[,1]

[1]Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada

***Corresponding author:** E-mail: louis.bernatchez@bio.ulaval.ca.

Associate editor: Douglas Crawford

Abstract

A functional understanding of processes involved in adaptive divergence is one of the awaiting opportunities afforded by high-throughput transcriptomic technologies. Functional analysis of coexpressed genes has succeeded in the biomedical field in identifying key drivers of disease pathways. However, in ecology and evolutionary biology, functional interpretation of transcriptomic data is still limited. Here, we used Weighted Gene Co-Expression Network Analysis (WGCNA) to identify modules of coexpressed genes in muscle and brain tissue of a lake whitefish backcross progeny. Modules were connected to gradients of known adaptive traits involved in the ecological speciation process between benthic and limnetic ecotypes. Key drivers, that is, hub genes of functional modules related to reproduction, growth, and behavior were identified, and module preservation was assessed in natural populations. Using this approach, we identified modules of coexpressed genes involved in phenotypic divergence and their key drivers, and further identified a module part specifically rewired in the backcross progeny. Functional analysis of transcriptomic data can significantly contribute to the understanding of the mechanisms underlying ecological speciation. Our findings point to bone morphogenetic protein and calcium signaling as common pathways involved in coordinated evolution of trophic behavior, trophic morphology (gill rakers), and reproduction. Results also point to pathways implicating hemoglobins and constitutive stress response (HSP70) governing growth in lake whitefish.

Key words: **ecological speciation, *Coregonus cupleaformis*, coexpression networks, microarray, phenotype, gill raker.**

Article

Introduction

Understanding the mechanism of ecological speciation, that is, the process by which organisms evolve new adapted phenotypes under divergent natural selection and become reproductively isolated, has been central to evolutionary biology (Schluter 2000; Butlin et al. 2012; Nosil 2012). Evolution of gene expression plays a key role in speciation (Wolf et al. 2010). However, compared with progress in understanding genomic divergence (Feder et al. 2012; Via 2012), its role has been given much less attention in recent literature (Pavey, Collin, et al. 2010). As changes in gene expression may underlie many of the phenotypic differences between species (Brawand et al. 2011), studying transcriptomic divergence of organisms in the early stages of speciation may shed light about the initial genetic targets of natural selection.

In North America, sympatric lake whitefish (*Coregonus cupleaformis*) populations have phenotypically diverged into a benthic (normal) and a limnetic (dwarf) ecotypes (Bernatchez et al. 1999) making it the most advanced case of divergence and reproductive isolation along the continuum of ecological speciation in freshwater fishes (Hendry 2009). In addition to their differential trophic specialization, normal and dwarf whitefish differ in life-history traits, morphology, behavior, and physiology (Rogers and Bernatchez 2007; Evans and Bernatchez 2012; Evans et al. 2012).

Several microarray approaches have been used to explore differences in gene expression between these divergent ecotypes (Derome et al. 2006, 2008; St-Cyr et al. 2008; Whiteley et al. 2008; Nolte et al. 2009; Renaut et al. 2009). Despite the substantial findings of these studies (reviewed in Bernatchez et al. 2010), the evolutionary consequences of specific gene-expression differences is difficult to interpret because some of them may evolve neutrally and therefore have little functional consequence (Khaitovich et al. 2004). However, a Weighted Gene Co-Expression Network Approach (WGCNA) has been proposed to compare expressed genes in terms of coexpression connectivity, to help identify key drivers of evolutionary changes (Oldham et al. 2006). In previous studies, gene connectivity has been shown to be a measure of functional relevance, in yeast in particular (Carlson et al. 2006). For instance, identifying coexpressed gene modules and the position of different genes in such networks, such as peripheral versus hub genes (i.e., the most highly connected genes) that are linked to variation in adaptive phenotypes could reveal important targets of evolution in fishes as well (Olson-Manning et al. 2012). Divergence in connectivity between species may reflect not only change in gene expression but also other types of evolutionary change, including divergent splicing, mRNA stability, or protein-coding sequence (Oldham et al. 2006).

The WGCNA approach provides a functional interpretation that is biologically significant (Miller et al. 2008) and performs well in constructing global network structures while being the most straightforward for exploring gene coexpression networks (Allen et al. 2012). It has been successfully

applied to identify functionally enriched modules implicated in complex diseases (Fuller et al. 2007; Miller et al. 2008; Plaisier et al. 2009; Kadarmideen et al. 2011; Rosen et al. 2011; Winden et al. 2011). However, its application to nonmodel organisms has been limited thus far (but see Ficklin and Feltus 2011; Weston et al. 2008; Kumar et al. 2010). However, this approach offers tremendous opportunities for transcriptome-wide analysis of nonmodel organisms to better understand the process of ecological speciation, such as in lake whitefish. This analytical framework also bypasses the multiple testing problems when relating gene expression to phenotypic traits and does not require an a priori gene annotation (Langfelder and Horvath 2008). The latter point is especially relevant because genes with unknown functions are plentiful in nonmodel organisms (Pavey et al. 2012).

Regulatory networks can be genotype dependent (Plaisier et al. 2009; Langfelder et al. 2012). Therefore, coexpression module conservation in multiple populations is indicative of the central functional importance of the module (Langfelder et al. 2011, 2012) and of its potential relevance for explaining parallel phenotypic divergence. Modules defined in a segregating population of hybrid crosses are presumed to reflect biological networks present in the parental forms along with some hybrid-specific coexpression patterns (Landry et al. 2007). Although this strategy cannot assess whether coexpression modules present in the pure populations were disrupted in the hybrid populations, modules in hybrids not preserved in parental populations would be indicative of such backcross novel (rewired) coexpression phenotype. By identifying misexpressed genes in hybrids in their functional context, this approach has the potential to overcome the problem that misexpressed genes might be downstream targets of genes that actually cause reproductive isolation (Butlin et al. 2012).

The main objectives of this study were to identify modules of coexpressed genes associated with phenotypes known to be involved in the ecological speciation of dwarf and normal whitefish, and to identify their key drivers. Because highly correlated traits may share common genetic factors, a strategy summarizing multiple phenotypic measurements of complex traits such as principal component analysis (PCA) might yield more insight into its underlying regulatory mechanism (He et al. 2008). Therefore, we used a segregating population of lake whitefish to generate principal component (PC) -based phenotypic gradients of 19 different phenotypic traits of known importance to whitefish divergence. Then we used WGCNA to define phenotype-correlated modules, and module preservation was measured in dwarf and normal ecotypes of different lakes. This approach provides a systems biology framework to evaluate the potential impact of evolutionary changes in a module, while simultaneously identifying candidate genes underlying known adaptive phenotypes. Moreover, we looked for evidence of backcross-specific rewiring to identify gene potentially implicated in intrinsic reproductive isolation. Finally, we tested the module preservation in a different species, sockeye salmon (*Oncorhynchus nerka*), to assess the phylogenetic conservation of our modules. The different network concepts and statistics used are presented in table 1.

Results

Phenotypic Gradients

The PCA on the 19 correlated phenotypic traits (listed in fig. 1) extracted three phenotypic gradients PC1, PC2, and PC3 representing 5%, 27%, and 15% of explained variance, respectively. Figure 1 shows that the PC1 was mainly explained by measurements linked to reproduction (gonad weight and sex-normalized gonadosomatic index). Although PC1 only explained 5% of the variation, it is still important to investigate what gene network may be involved in controlling the expression of sexual maturity, given the general lack of knowledge of such relationships in fishes and the importance of this trait in the parallel evolution of dwarf and normal whitefish species pairs. Figure 1 also shows that the PC2 was mainly explained by growth-related traits (weight and length, before and at maturity) and PC3 by behavioral traits (depth selection, burst swimming, and directional change). Also, the gill raker number was unexpectedly positively and negatively correlated to PC1 and PC3, respectively. Thus, to verify whether gender was a confounding factor, we looked separately at males and females. Gill raker correlation with PC1 was stronger in males ($r = 0.66$, $P < 0.0001$) than in females ($r = 0.48$, $P = 0.007$), so was the correlation with PC3 (males: $r = -0.37$, $P < 0.02$, females: $r = -0.29$, $P = 0.12$), but the directionality of the correlations was consistent, indicating that gender was not a confounding factor.

Network Construction

A total of 14 and 17 modules were defined in the brain and muscle networks (table 2). Of all genes in the networks, 2,788 were assigned to a module in brain network and 3,520 in muscle. The remaining genes that did not fit elsewhere were assigned to the same "grey" module (see supplementary figs. S1 and S2, Supplementary Material online, for a global view of the networks). The module quality statistic, Zsummary.qual. (table 1), indicated that all brain and muscle modules were highly reproducible (Zsummary.qual $\geq$ 12).

Functional Enrichment Analysis of Coexpression Modules

Overall, 46% of unique probes of the microarray were annotated with gene ontology (GO) terms. The brain network was composed of 46%, whereas the muscle network comprised 55% annotated genes. In both networks, only modules with a minimum of 67% annotated genes had enrichment terms with false discovery rate (FDR) < 0.05. Most significant and largest significant GO annotations are reported for each module in table 2. To ease interpretation throughout the rest of the article, each module's color name is followed by the most significant GO term in parentheses. Complete functional enrichment analysis of each module (Fisher's exact test) is presented in supplementary file S1, Supplementary Material online. Overall, there was no significant difference in module membership (table 1) between annotated versus

Table 1. Definition of Network Concepts.

Term	Description
PCs	Orthogonal composite variables (scores) from a PCA. Here, PCs represent uncorrelated phenotypic gradients across the whitefish hybrid population.
WGCNA	A method that constructs a gene network based on the similarity of expression profiles among samples. WGCNA defines modules of consistently coexpressed genes and correlates them to a trait.
Nodes	Nodes correspond to genes.
Edges	Connections between genes in the coexpression network. In WGCNA, the Pearson correlation between gene expression levels is transformed with a power function and used as a continuous edge attribute (weight).
Module	Cluster of highly interconnected nodes. In WGCNA, modules are defined in such a way that genes in a module tend to share the same neighbors in the network.
Module eigengene (ME)	Variable summarizing the expression level of a module across individuals. It is analogous to the first PC of the module.
Hub genes	Highly connected nodes, here corresponds to genes with the highest module membership.
GS	In general, the Pearson correlation between a gene and a trait. In this study, because we have multiple measured phenotypes in several groups, we correlate genes with trait principle components.
MS	Pearson correlation between the module eigengene and the phenotypic PCs.
AGS	The average of GS in each module.
Connectivity	In a weighted network, the connectivity of a node is the sum of weights of all its edges. In the context of modules, connectivity can be approximated by the module membership (see later).
Module membership (MM) (also known as K_{ME})	Correlation between the expression levels of a gene and the module eigengene. This continuous measure reflects the connectivity of a gene with other genes in the module and was used to define hubs.
Zsummary	A composite statistic based on a permutation test that takes into account the preservation of both connectivity and density in a module. Zsummary < 2 implies no evidence for module preservation, $2 <$ Zsummary < 10 weak to moderate evidence, and Zsummary > 10 strong evidence for module preservation (Langfelder et al. 2011). This measure tends to increase with module size, therefore its interpretation is only suited for comparing preservation of a given module in different populations.
MedianRank	An alternative composite statistic based on the ranking of several connectivity and density criteria. This measure is used to compare preservation between modules in a given population.
Zsummary.qual	This value compares the density and connectivity of a module to a random module of a thousand genes. It is a measure of module quality, indicative of how well-defined modules are in the reference set.

nonannotated genes (FDR-corrected Wilcoxon), meaning that nonannotated genes had the same opportunity to occupy hub positions as frequently has annotated genes.

Modules Correlated to PCs

Visual inspection of the gene dendrogram with module color and association with PCs (fig. 2) revealed that genes whose expression was correlated with PCs tended to cluster together in the same branches of the tree, indicating that these genes follow a modular organization. Two quantitative measures were considered when screening for trait-related modules: the module significance (MS) (table 1) (fig. 2) and the average gene significance (AGS) in the module (supplementary fig. S3, Supplementary Material online). In terms of MS, PCs generally had higher values than individual traits (supplementary fig. S4, Supplementary Material online). Two brain modules were associated with PC1 (reproduction), the $\text{cyan}_{\text{brain}}$ module (small molecule metabolic process) and the $\text{grey60}_{\text{brain}}$ module (reproductive process). Also, the $\text{black}_{\text{brain}}$ module (intracellular organelle) was correlated with PC2 (growth). As for muscle, the $\text{greenyellow}_{\text{muscle}}$ (catalytic activity), $\text{yellow}_{\text{muscle}}$ (cytoplasmic part), and $\text{brown}_{\text{muscle}}$ (regulation of phosphorus metabolic process) modules correlated positively with PC1 (reproduction). The $\text{blue}_{\text{muscle}}$ (translation) and $\text{red}_{\text{muscle}}$ (mitochrondrial inner membrane) modules correlated negatively with PC1. Only the $\text{purple}_{\text{muscle}}$ module (hemoglobin complex) was marginally associated with PC2 (growth). Finally, PC3 (behavior) was associated with the $\text{pink}_{\text{muscle}}$ module (RNA binding).

Overlap between Brain and Muscle Modules

Out of the 3,600 considered for our analysis, only 1,864 genes intersected between both tissues, hence preservation analysis between tissues was limited to a contingency analysis. Supplementary figure S5, Supplementary Material online, shows that only a few brain modules significantly intersect with muscle modules. However, it is noteworthy that part of the $\text{black}_{\text{brain}}$ (intracellular organelle) and $\text{yellow}_{\text{muscle}}$ (cytoplasmic part) modules overlap because both are correlated to PC2 (growth) and PC1 (reproduction). $\text{Grey60}_{\text{brain}}$ module (reproductive process) also overlaps slightly with $\text{red}_{\text{muscle}}$ module (mitochrondrial inner membrane), and they are both negatively correlated to PC1.

Brain Module Preservation in Pure Ecotypes

The composite Zsummary and MedianRank statistics (table 1) were used to assess module preservation between three dwarf and three normal populations (fig. 3*a*). Because Zsummary is size dependent, the complementary MedianRank statistic is used to assess relative preservation of

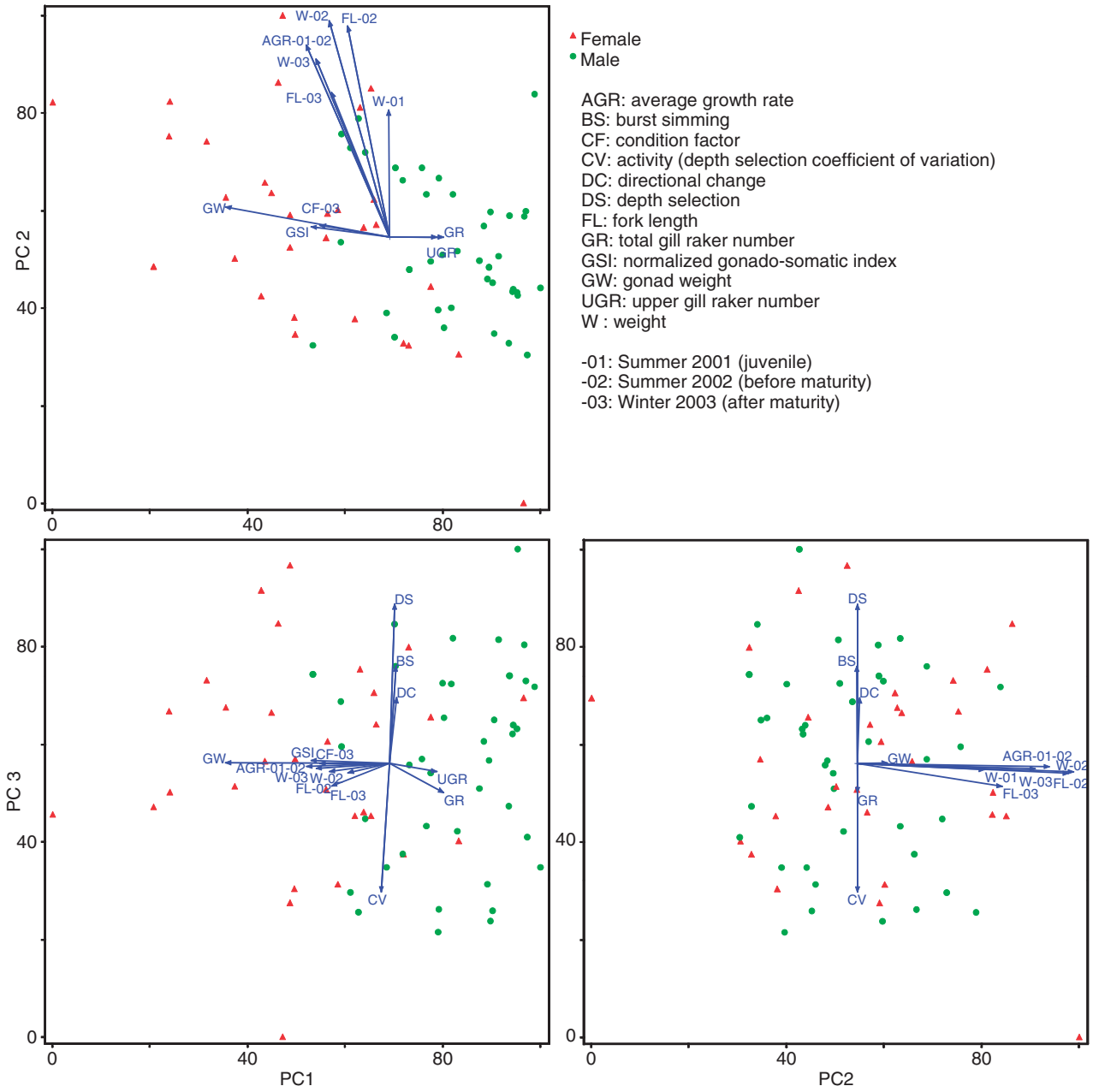

Fig. 1. Rotated PCA on correlation between phenotypic traits of backcross hybrids showing three phenotypic gradients (PC1, PC2, and PC3). Variance represented is 5%, 27%, and 15% for PC1, PC2, and PC3, respectively. For clarity, only traits (blue vectors) with coefficient of determination $r^2 > 0.2$ are shown. Traits acronyms are detailed on the upper right panel. Red triangles and green dots represent female and male samples, respectively.

modules (fig. 3*b*). On one hand, the backcross brain modules were moderately preserved in dwarf and normal populations, including their pure parental populations. The level of grey60$_{brain}$ module (reproductive process) preservation varied across populations but consistently ranked among the first three most preserved modules. The black$_{brain}$ module (intracellular organelle) also ranked high, yet its Zsummary preservation levels varied between ecotypes. Relative to normal whitefish from the same environment, preservation was higher in dwarf in the controlled environment and from Indian Pond but higher in normal than in dwarf whitefish in Cliff Lake, representing a nonparallel pattern of preservation. The cyan$_{brain}$ module was only weakly preserved in half of the populations. When visually comparing the connectivity pattern of less preserved modules in the combined pure dwarf and normal populations with the pattern in the backcross individuals, the red$_{brain}$ module (intracellular nonmembrane-bounded organelle) stood out: A noticeable group of genes was highly coexpressed in backcrosses but not in pure ecotypes (fig. 4).

Muscle Module Preservation in Sockeye Salmon

About half of the modules showed low to moderate preservation in muscle tissue of sockeye salmon populations (supplementary fig. S6, Supplementary Material online). Preservation varied between populations, but phenotype-correlated modules defined in whitefish were also generally more

Table 2. Brain and Muscle Functional Module Gene Enrichment Characterization Including Number of Genes and Percent Annotated, as well as Attributes of the Most Significant and Largest Enriched GO Terms.

Network Module	Number of Genes in Module	Percentage Annotated	Most Significant GO Enrichment			Largest Significant GO		
			GO Term	Number of Genes in GO Term[a]	P^b	GO Term	Number of Genes in GO Term[a]	P^b
Brain module								
Black	120	52	Intracellular organelle	55	3E-04	Cell part	59	1E-03
Blue	710	46	Regulation of developmental process	21	2E-03	Cellular response to stimulus	81	5E-02
Brown	181	44	Amine metabolic process	13	2E-02	Amine metabolic process	13	2E-02
Cyan	38	29	Small molecule metabolic process	6	3E-02	Small molecule metabolic process	6	3E-02
Green	133	68	Cellular biosynthetic process	59	1E-06	Cell part	87	2E-04
Greenyellow	58	41	Kinase activity	5	8E-04	Catalytic activity	14	3E-02
Grey60	27	63	Reproductive process	10	7E-05	Binding	17	7E-03
Lightcyan	130	32	—	—	—	—	—	—
Magenta	88	32	Extracellular region	6	1E-02	Extracellular region	6	1E-02
Midnightblue	35	34	Nuclear part	5	5E-02	Nuclear part	5	5E-02
Purple	86	22	Purine ribonucleoside triphosphate metabolic process	5	1E-03	Regulation of cellular process	8	3E-02
Red	122	42	Intracellular nonmembrane-bounded organelle	29	5E-04	Primary metabolic process	38	2E-02
Coral	42	50	Viral infectious cycle	7	8E-03	Macromolecular complex	15	3E-02
Tan	51	41	Structural molecule activity	9	3E-03	Intracellular part	20	1E-02
Turquoise	799	49	Postsynaptic density	5	8E-04	Signaling	90	4E-02
Yellow	168	35	Lymphocyte activation	5	9E-04	Biological regulation	34	1E-02
Muscle module								
Black	113	75	Viral infectious cycle	32	1E-12	Cytoplasmic part	54	9E-05
Blue	319	67	Translation	83	1E-11	Cellular process	193	2E-04
Brown	258	53	Regulation of phosphorus metabolic process	9	2E-03	Primary metabolic process	88	3E-02
Cyan	29	52	Cellular metabolic process	14	5E-02	Cellular metabolic process	14	5E-02
Green	156	79	Mitochondrial part	89	1E-52	Cell part	119	6E-04
Greenyellow	40	38	Catalytic activity	8	2E-02	Catalytic activity	8	2E-02
Magenta	64	48	Enzyme binding	7	2E-04	Regulation of biological process	19	3E-03
Pink	67	51	RNA binding	11	2E-02	Intracellular organelle	29	3E-02
Purple	50	78	Hemoglobin complex	18	8E-21	Cytosol	30	1E-08
Red	135	68	Mitochondrial inner membrane	25	1E-05	Intracellular part	87	4E-05
Coral	34	50	—	—	—	—	—	—
Tan	38	47	Transferase activity	5	7E-03	Regulation of cellular process	11	2E-02
Turquoise	2,005	49	Oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	14	4E-05	Biological regulation	433	9E-03
Yellow	212	66	Cytoplasmic part	107	1E-04	Cell part	125	4E-02

[a]Only GO categories with at least five genes in the test set were reported. Complete results are reported in supplementary files S1 and S2, Supplementary Material online.
[b]P values corresponding to FDR < 0.05 are underlined.

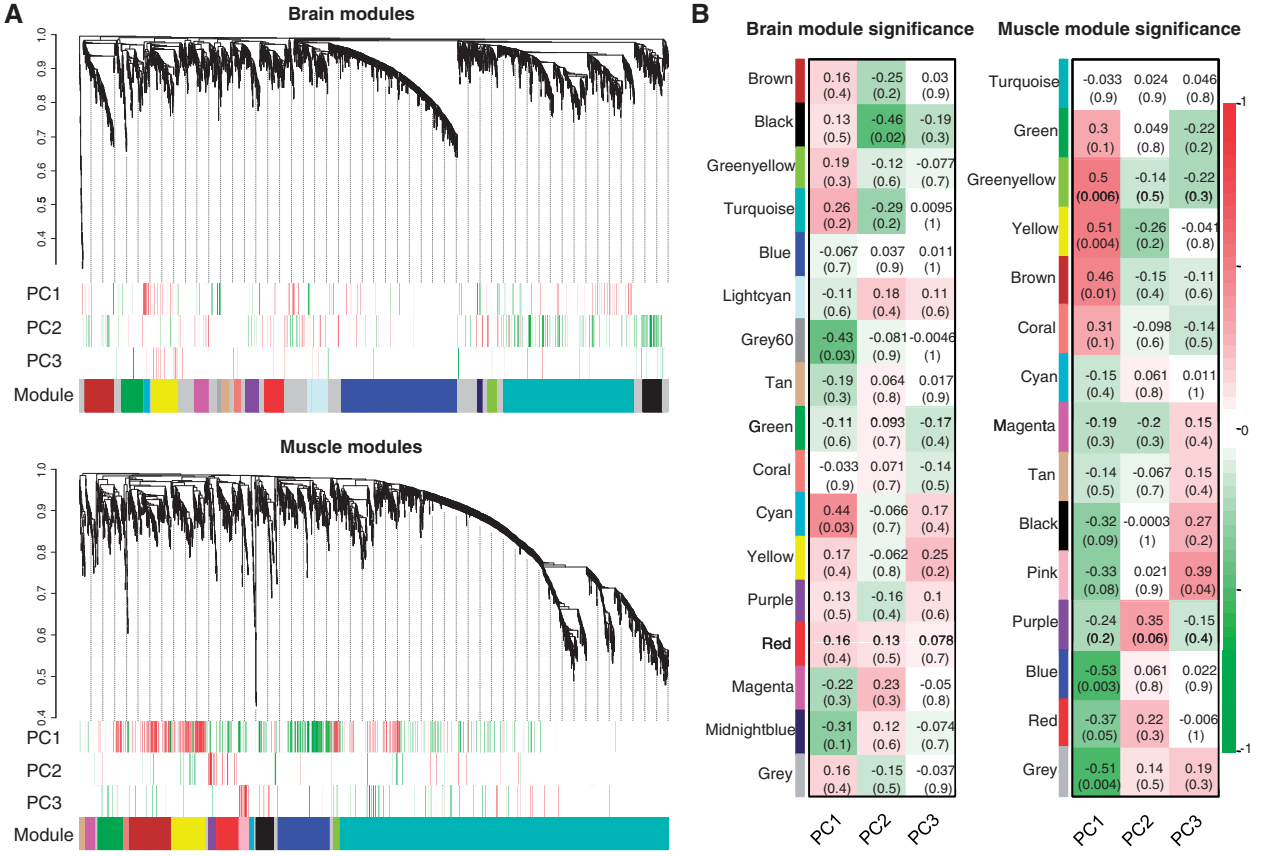

FIG. 2. Correlation between brain and muscle module and the phenotypic gradients. (*A*) Average linkage clustering tree (dendrogram) based on topological overlap distance in gene expression profile in both the brain and muscle data sets separately. Branches of the dendrogram correspond to modules, shown in the "module" colorbar below the dendrogram. Other color rows indicate GS ($P < 0.05$) for each phenotypic gradients (PCs). Green color indicates negative correlation and red color indicates positive correlation. (*B*) Correlation between brain and muscle module eigengenes and the phenotypic gradients. Each row corresponds to a module identified on the left side by its color. Each column corresponds to a PC. Each cell reports the Pearson correlation between the module eigengene and PCs using complete pairwise option along with uncorrected *P* value in parenthesis ($N = 29$ in muscle and $N = 26$ in brain). Cells are color coded using the correlation value according to color scale on the right; positive correlations are denoted red and negative correlation in green. Note that module colors are attributed in decreasing order of module size in each tissue and therefore do not correspond to the same gene sets in both tissues.

preserved in sockeye salmon than other modules. Moreover, $yellow_{muscle}$ (cytoplasmic part) and red_{muscle} (mitochrondrial inner membrane) modules that overlapped with preserved phenotype-correlated modules in brain in whitefish were also moderately preserved in at least one salmon population. Muscle-specific modules, such as the $pink_{muscle}$ (RNA binding), $blue_{muscle}$ (translation), and $purple_{muscle}$ (hemoglobin complex) also showed moderate evidence of preservation in at least one salmon population. In fact, the $purple_{muscle}$ was overall the most preserved module, and length measurement available for 27 of these salmon samples was positively correlated with $purple_{muscle}$ module eigengene (MS = 0.43, $P = 0.025$), a result that is concordant with the marginal MS of the $purple_{muscle}$ module for the growth gradient (PC2) in whitefish.

Candidate Genes Associated with Phenotypic Divergence

Several lines of evidence pointed toward regulatory genes as candidate for explaining complex traits divergence, as they were hub genes of modules of particular interest. First, the $grey60_{brain}$ module (reproductive process) was correlated negatively with reproduction (PC1), meaning that this module was expressed at higher levels in mature individuals, which corresponds to the dwarf phenotype at this age. This module also had 11 genes intersecting with the red_{muscle} module (mitochrondrial inner membrane) whose correlation with PC1 was also negative. The consensus hub of this subset is the fk506 binding protein 1a (expressed sequence tag [EST] accession: CB497859) gene, which is a ubiquitous abundant protein that is involved in many functions (supplementary fig. S7, Supplementary Material online). It regulates cell cycle by down-regulation of TGF-beta receptor signaling and binds tightly to intracellular calcium release channels (Aghdasi et al. 2001). It has also been implicated in osteogenic differentiation by interacting with the immunosuppressant drug FK506 via activation of bone morphogenetic protein (BMP) receptors (Kugimiya et al. 2005). The six annotated genes in this 11 gene subset were involved in developmental process (GO:0032502, $P = 8.1E\text{-}4$), and two were involved in cell recognition (GO:0008037, $P = 1.3E\text{-}4$) and single fertilization (GO:0007338, $P = 1.9E\text{-}4$), which is in line with the functional

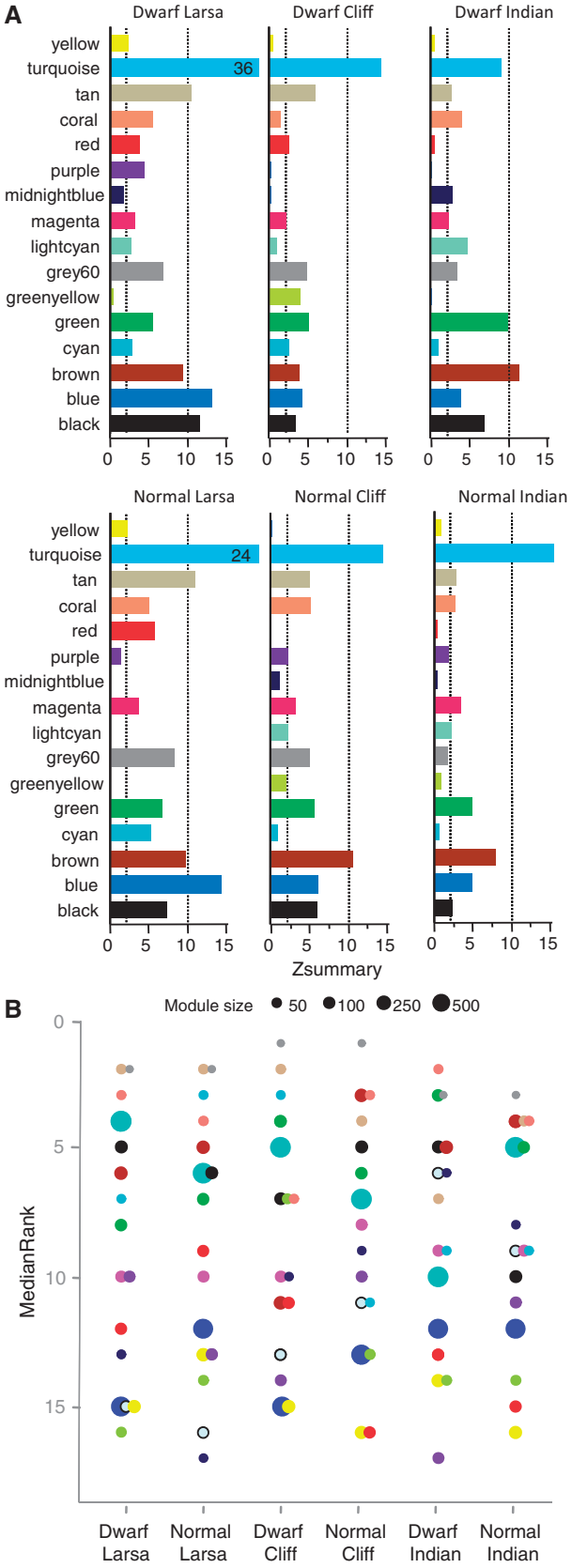

Fig. 3. Backcross brain module preservation in pure normal and dwarf populations. (*A*) Histogram of the Zsummary statistic comparing module preservation between populations. The dotted lines represent the interpretation threshold for no preservation (<2) and high preservation (>10). Values larger than 20 are shown directly on the bars.

enrichment in reproductive process of the grey60$_{brain}$ module (table 2).

Second, 32 genes in the preserved black$_{brain}$ module (intracellular organelle) overlapped with the yellow$_{muscle}$ module (cytoplasmic part) and correlated negatively with growth (PC2) in both tissue meaning that these genes are expressed at higher levels in smaller fish (supplementary fig. S7, Supplementary Material online). Overall, this subset of genes is involved in interspecies interactions between organisms (GO:0044419, $P = 2.0E-3$) and cellular metabolic process (GO:0044237, $P = 2.1E-2$), mostly cellular protein metabolism process (GO:0044267, $P = 1.4E-2$). The heat shock cognate 70 gene (*hsp70*, EST accession: CB498852) was the subset hub in both brain and muscle.

Third, the pink$_{muscle}$ module (RNA binding) was the only significant module for behavior (PC3) in both brain and muscle networks. Individuals with higher expression levels of this module were characterized by typical dwarf whitefish behaviors, exhibiting frequent burst swimming and preference for a high position in the water column (Rogers et al. 2002). The hub gene of the pink$_{muscle}$ module was annotated as the zinc finger protein 22 (*znf22*, EST accession: CA061373) (fig. 5), a transcription factor notably expressed during craniofacial development and implicated in tooth formation (Gao et al. 2003). Two other genes involved in craniofacial development were present in this module, namely genes encoding calmodulin (*calm*, two different transcripts, EST accession: CB501671 and CK990427) and BMP4 (*bmp4*, EST accession: CA056395). The pink$_{muscle}$ module comprised a total of 10 genes involved in organ development (GO:0048513) and 11 genes involved in response to stimulus (GO:0050896). One of these was the dynein light chain roadblock-type 2 gene (*dlrb2*, EST accession: CB500003), which was annotated with visual behavior.

Discussion

Gene regulatory networks provide a systematic understanding of molecular mechanisms underlying biological processes (Ihmels et al. 2002; Stuart et al. 2003). Moreover, understanding individual gene's properties within the network may be as important as understanding its function in isolation (Barabási and Oltvai 2004). As knowledge of gene function is largely based on biomedical research of the few traditional model organisms (Pavey et al. 2012), defining gene regulatory networks of emerging ecological models is a promising avenue for determining the ecological function of genes. In particular, analysis of the regulation of sets of genes comprising functional networks is crucial to decipher the genomic basis of ecological speciation (Prud'homme et al. 2007). Here, we

Fig. 3. Continued
(*B*) Bubble plot of the MedianRank statistic compares relative preservation of modules across populations. The bubbles represent modules by their color, and size is proportional to the number of genes they comprise. They are positioned according to their rank, first being the most preserved module.

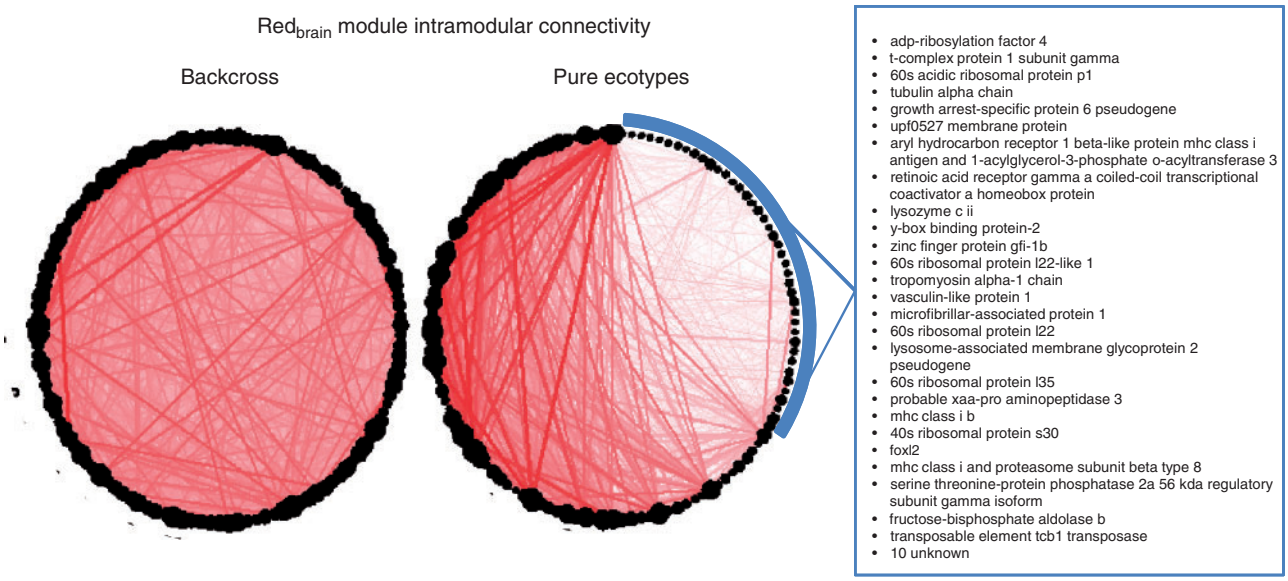

Fig. 4. Red$_{brain}$ module coexpression pattern in backcross and pure individuals. The red lines reflect the connectivity between the genes (black dots), and line thickness represents the degree of coexpression. The size of the dots reflects the total intramodular connectivity of each gene. The text box on the right lists the genes that are highly wired in the backcross population only.

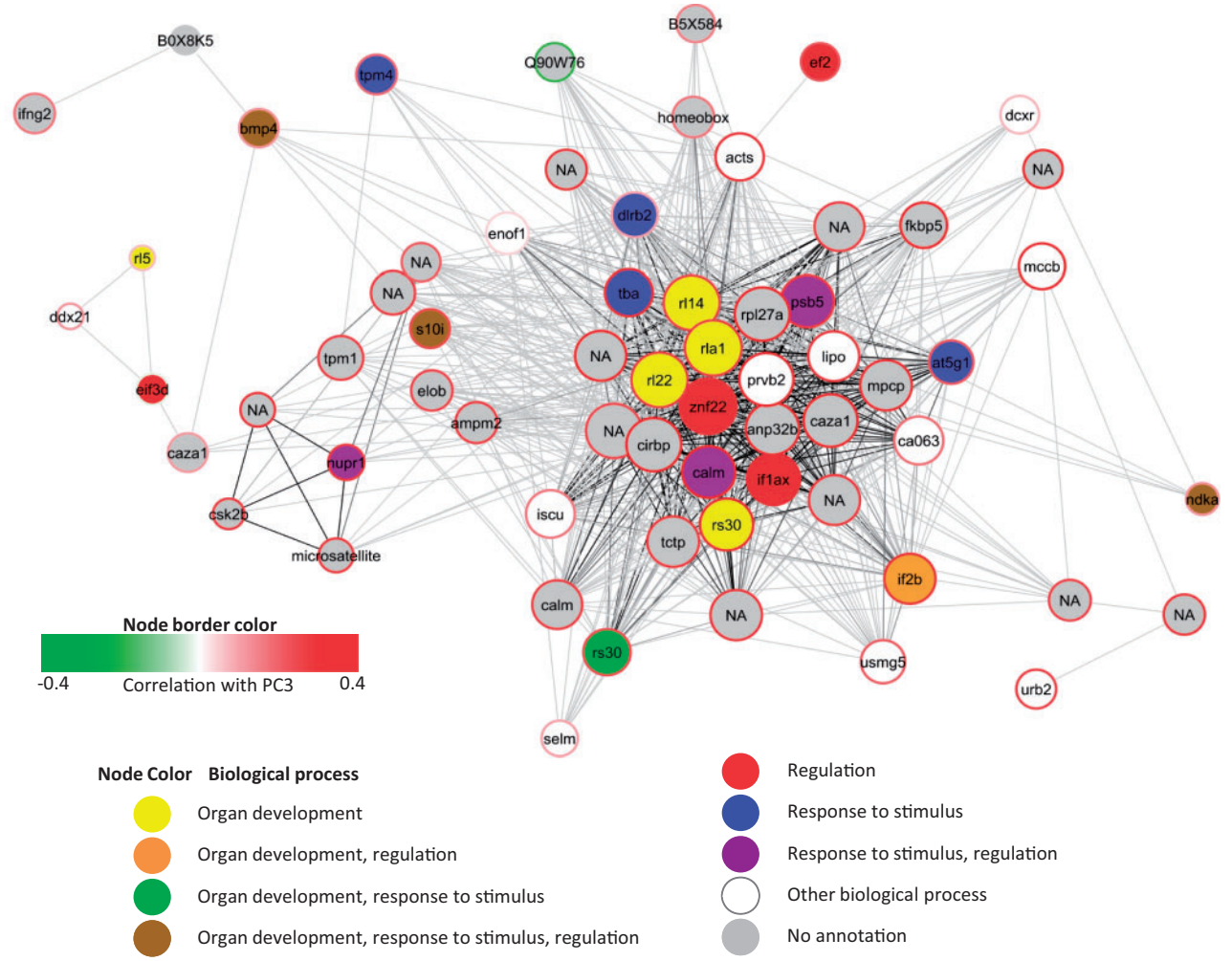

Fig. 5. Network view of the pink$_{muscle}$ module. Node are labeled with protein symbols when available. Nodes are colored according to GO biological process. Node size corresponds to module membership, and correlation with PC3 (behavior) is reflected by the line border color (green = negative, red = positive, white = no correlation). Edge line width reflects coexpression between genes in muscle tissue. NA, not available; gray, not annotated.

defined gradients of phenotypic variation encompassing reproductive, growth, morphological, and behavioral traits that diverged between dwarf and normal whitefish during the process of ecological speciation (Rogers and Bernatchez 2007). We further identified modules of coexpressed genes related to these phenotypic gradients and their key drivers, revealing potential primary targets of natural selection. Additionally, we tested the module preservation in parental and natural populations as well as in a different species, sockeye salmon (*Oncorhynchus nerka*), a first step toward untapping the phenotype–genotype map of nonmodel organisms. Later, we discuss how the functional analysis of transcriptomic data using a weighted gene coexpression network approach contributes to the understanding of mechanisms underlying ecological speciation in lake whitefish.

The Gill Raker Puzzle

Multivariate analysis of phenotypic traits revealed a high degree of concordance between single traits pertaining to sexual maturation, growth, and behavior, as they formed distinct gradients. The fact that these gradients were mostly independent suggests that traits composing each gradient may share some common genetic factors, whereas gradients remain largely independent. Thus, the phenotype-correlated modules identified here may represent set of genes that have coevolved in controlling the expression of phenotypic traits involved in the adaptive divergence of sympatric benthic and limnetic whitefish. However, the gill raker number followed an unexpected pattern. Gill rakers have been shown to be one of the most important traits under selective adaptive divergence in lake whitefish, as dwarf ecotypes have more gill rakers, which is associated with smaller prey selection (Rogers and Bernatchez 2007). Two modules following this trait were of particular interest. Specifically, backcross individuals at the higher end of the $pink_{muscle}$ (RNA binding) gene expression spectra were associated with a swimming behavior closer to the dwarf phenotype but a number of gill raker closer to the normal phenotype, contrary to expectation. Similarly, the expression of the $grey60_{brain}$ module (reproductive process) was associated with an earlier sexual maturation as seen in dwarfs and again with lower gill raker number. Gill raker and behavior traits have been mapped to a common linkage group in females of a hybrid x normal family of backcross but not in males (Rogers and Bernatchez 2007). Thus, the $pink_{muscle}$ and the $grey60_{brain}$ modules each represent candidate coadapted modules that could affect the development of gill rakers with pleiotropic effects, and our results suggest that this coadaptation would have been disrupted in this backcross family. The recombination of gill raker and behavior or maturity traits could represent extrinsic postzygotic reproductive barriers in specific environmental contexts under the hypothesis that the breakdown of these coadapted traits decreases fitness. $Pink_{muscle}$ and $grey60_{brain}$ gene annotation confirm the presence of genes playing a role in reproduction and behavior, respectively. Strikingly, both subsets of candidate genes implicated the BMP pathway and calcium signaling. In fact, the $pink_{muscle}$ module included particularly interesting genes, namely *bmp4* and *calmodulin* (Pavey, Collin, et al. 2010), whose expression during embryonic development have been implicated in Darwin's finches beak width and depth as well as in cichlid jaw morphogenesis (Parsons and Albertson 2009). Interestingly, gill arches are the structures from which jaws most likely evolved in the first place (Kimmel et al. 2001).

Behavior is thought to be the first trait to respond to selection in divergence toward limnetic habitats with respect to new resource utilization, followed by a divergence in prey size and type (Schluter 2000). Here, the fact that genes involved in trophic morphology (gill raker) are coexpressed in a module correlated with behavior responsible for habitat specialization (depth selection) brings an alternative hypothesis to the sequential events previously proposed (Schluter 2000; Rogers and Bernatchez 2007) as it suggests a partially common regulation mechanism coordinating behavior and morphology. The $pink_{muscle}$ module hub, the zinc finger protein 22, is a good candidate for this potential mechanism, because of its function in transcription regulation and bone structure development (Gao et al. 2003). This gene is expressed during craniofacial development, but also in adult muscle where its role is still unknown. *bmp4* has previously been implicated in larval swimming behavior in the cavefish *Astyanax mexicanus* (Pottin et al. 2010). Calmodulin-dependant signaling is also critical for synaptic function, hence its role in behavior phenotype (Wei et al. 2011). *Calmodulin* and *bmp4* are members of signaling pathways that may interact with each other (Parsons and Albertson 2009), yet they do not regulate each other's expression in chicken embryonic beak (Mallarino et al. 2011). Clearly, our results support the relevance of performing further detailed experiments including multiple life stages toward elucidating the role of these genes in the complex mechanisms governing the associations between coadapted swimming behavior and gill rakers.

A Role for Gene Expression Networks in Adaptive Divergence

The implication of the gene balance hypothesis for quantitative traits is that selection will occur by accumulation of subtle changes in many genes, because high variation in regulatory gene expression is selected against (Birchler and Veitia 2010). Here, PCs can be considered as quantitative traits that evolved under divergent selection in the process of whitefish ecological speciation (Bernatchez 2004). Following the gene balance hypothesis, it is expected that gene significance (GS, table 1) will be low to moderate for many genes if mutations of strong effect are selected against. Hence, applying a multiple testing threshold such as Bonferroni correction and even FDR would likely exclude many true positives. In our interpretation, we used both the AGS and the GS and analyzed multiple data sets to reinforce the biological significance of our results instead of applying a false discovery rate to account for multiple testing. It has been reported in the literature that biological signal can sometime be masked by too stringent threshold (Williams and Haines 2011). Nevertheless, when analyzed as coexpression network, such moderately

significant genes may be organized in a relatively small number of modules, making the modules biologically relevant. Indeed, it is expected that methods taking into account the relationship between genes would be superior to methods that assume independence among genes outside of their functional contexts (Minguez and Dopazo 2011). A good example here is the fact that the purple$_{muscle}$ module (highly enriched for hemoglobin complex) is preserved in salmon populations and significantly correlated to a measure related to growth, whereas its uncorrected *P* value in the backcross was only 0.06. However, its AGS was clearly higher than for the other modules to PC2, meaning that genes correlated to growth were concentrated in this module. We feel that testing for the preservation of modules in a completely different data set and species is a better demonstration of the generality of our findings than applying multiple testing corrections and represents a good balance between type I and type II errors. However, because other module-trait associations were not validated, they should be treated as suggestive.

The importance of the purple$_{muscle}$ module was further supported by the highly significant GO enrichment for hemoglobin complex. A role for regulatory genes is suggested as this module also comprised a homolog of the krueppel-like factor 11, which is a transcription factor of the globins gene promoter and repressor of promoters containing SP1-like binding sites inhibiting cell growth (Asano et al. 1999). Hemoglobin gene upregulation has been reported in brain tissue of dwarf lake whitefish (Evans et al. 2012); however, the expression of the purple$_{muscle}$ module was associated with larger (normal like) individuals in this study. Despite this discrepancy, this result supports the conclusion that metabolic traits involved in oxygen transport differ between ecotypes (Evans et al. 2012).

The black$_{brain}$ module (intracellular organelle) is an example of module rewiring between ecotypes. The conserved HSP70 gene family is known to respond to environmental stresses such as heat shock, UV, infection, and chemical exposure, which affect growth and physiological conditions in aquatic environments (Yamashita 2010). Also, transcripts of this gene family are differentially expressed between ecotypes and highly transgressive in hybrid embryos (Nolte et al. 2009; Renaut et al. 2009). The black$_{brain}$ module is negatively correlated with growth rate and size, which indicates that fish with a greater response to stress would be closer to the dwarf ecotype in size and metabolism. This is in line with the stress associated with occupying the limnetic niche, where active swimming necessary for increased foraging and predator avoidance engages high energetic costs that translate into reduced energetic conversion efficiency, slower growth rate, and overexpression of survival functions in dwarf lake whitefish (Trudel et al. 2001, St-Cyr et al. 2008).

Insights into the Genomic Basis of Adaptive Divergence and Reproductive Isolation

Langfelder et al. (2012) found that module preservation varied markedly among mouse crosses underlying the importance of genotypic variation among populations for gene coexpression. Differences in regulatory networks may also exist between different environmental (growth) conditions (Zhu et al. 2012) and a combined impact of these two factors cannot be ruled out. Here, preservation of the purple$_{brain}$ (hemoglobin complex) and the black$_{brain}$ (intracellular organelle) modules defined in whitefish backcross varied between populations reared in a controlled environment and also in natural lakes, evidencing genotype-specific modules. In contrast, preservation of the tan$_{brain}$ (structural molecule activity), blue$_{brain}$ (regulation of developmental process), and grey60$_{brain}$ (reproductive process) modules varied across environments but not between ecotypes in the same environment, even for the parental lines originating from different glacial races (Rogers and Bernatchez 2007). This pattern therefore suggests the predominance of environmental effects on the coexpression of these modules. Finally, the brown$_{brain}$ module (amine metabolic process) exemplifies an interaction of genotype-environment effect as it is equally preserved between ecotypes in a controlled environment but differentially preserved between ecotypes in natural lakes. Taken together, these results underline the importance of understanding gene expression variation from a network perspective to gain insight into the respective role of adaptation and plasticity, which very few study have investigated so far (Weston et al. 2008).

Brain modules that were not well preserved in parental populations could represent rewired coexpression patterns in hybrids. For instance, the red$_{brain}$ module (intracellular nonmembrane-bounded organelle) showed a clear pattern of rewired coexpressed genes in the backcross family (fig. 4), as part of the module was clearly preserved in all dwarf and normal populations, whereas the rest was only coexpressed in H1D individuals. This result adds to the previous report of hybrid misexpression in lake whitefish embryos and juveniles (Renaut et al. 2009) and may represent the effects of intrinsic reproductive isolating barriers.

Limitations of This Study

Several limitations impede further detailed interpretation of the results. First, many of the genes in our networks have no informative annotation. We have now placed many of these in coexpression modules, some of which are correlated with important life-history traits. We hope this will start a process of "ecological association" of these genes, which may be a first step to further study and eventually determine their function (Pavey et al. 2012). Second, networks (and their association with phenotypes) were constructed using both genders because of sample size constraints; hence, potential gender-specific modules would not have been detected. Similarly, this common garden gene expression experiment with the backcross family is removed from natural environment variation, which could have veiled heritable coexpression patterns that require particular environmental cues to manifest. Third, as muscle modules could not be compared with parental populations, preservation and potential rewiring of these modules remain to be tested. This is especially important given that little is known about the extent to which the

genetic architecture generated in this backcross family reflects the variation present in natural populations (Rogers and Bernatchez 2007). However, the modules showing evidence of preservation in sockeye salmon population makes it reasonable to assume that these would likely be preserved in whitefish parental ecotypes. Fourth, despite the fact that great care was taken to consistently sample in the same anatomical location in the case of muscle and always sampled entire brains, organs can have different cell type compositions and distributions, and there is potential for this to affect microarray results. Thus, increased gene expression in a sampled tissue could be the result of very different biological processes: increased expression in all cell types sampled, changing composition to different cell types or subtypes that tend to expresses certain genes more, or a combination or the two. Finally, only healthy backcross adults could be included in this study preventing interpretation about the significance of the rewired patterns observed on hybrid fitness (Renaut and Bernatchez 2011). Thus, future work should include coexpression network analysis of multiple families at early life stages and include both healthy and moribund (sensu Renaut and Bernatchez 2011) to address these questions.

Materials and Methods

Whitefish Data Description

We leveraged a previously studied cross of lake whitefish between a normal x dwarf hybrid female and a dwarf male (referred to as the H1D backcross) for which phenotypic data as well as muscle ($N = 76$) and brain ($N = 55$) microarray data sets (GSE11378 and GSE12068) were available (Rogers and Bernatchez 2007; Derome et al. 2008; Whiteley et al. 2008) (see supplementary fig. S8, Supplementary Material online, for data overview). The H1D family was reared in a controlled environment (Laboratoire de Recherche en Sciences Aquatiques [LARSA]). All transcriptomic profiles were obtained using the consortium for Genomic Research on All Salmon Project (cGRASP) 16,006 probes microarray developed from cDNAs of Atlantic salmon and rainbow trout allowing direct comparisons (von Schalburg et al. 2005). Additionally, brain tissue of individuals from the two pure populations (dwarf: Témiscouata L., normal: Aylmer L., $N = 8$ each) used to make the above hybrid crosses and raised at LARSA and new samples from wild populations from two lakes (Cliff L. and Indian Pond) (doi: 10.5061/dryad.k8760) ($N = 8$ for dwarf and normal whitefish in each lake) were used to compare module preservation among ecotypes. Brain tissue sampling and microarrays were performed as described previously (St-Cyr et al. 2008; Whiteley et al. 2008). To test the taxonomic generality of the modules, we included unpublished muscle transcriptomic data from three populations of another salmonid, sockeye salmon (*Oncorhynchus nerka*, $N = 15$, 27, and 7) to assess interspecific muscle module reproducibility (GSE42985).

Spawning female sockeye salmon muscle was sampled in August 2006 and 2007 from the outlet and beach habitats of Surprise Lake, Aniakchak National Monument and Preserve, AK (Pavey et al. 2007, Pavey, Nielsen, et al. 2010). Methods for RNA isolation, cDNA labeling, hybridization, scanning, and quantification were identical to Pavey et al. (2011).

Whitefish Phenotypic Gradients

Phenotypic traits including life history (weight, length, condition factor, and growth rate), swimming behavior (burst swimming, directional change, depth selection, and activity level), morphology (gill raker number), and sexual maturity (gonad weight, gonadosomatic index, and maturity index) have been previously characterized for the H1D progeny (Rogers and Bernatchez 2007). PCA performed in PC-ORD 5.10 was used to summarize the information contained in 19 phenotypic measurements using 70 individuals (supplementary fig. S9, Supplementary Material online). The first PC was rotated to maximize the correlation with gender to isolate sex-biased traits. For ease of interpretation, the second PC was then rotated to maximize correlation with average growth rate. PCs retained good orthogonality (>83%) after rotation. Rotated scores of the three first PCs were used as PCs (table 1).

Microarray Data Processing

Starting from raw intensities, a threshold of expression was calculated for all the samples and fixed at two standard deviations above background mean, estimated from the expression level of blank spots on the array. Spots whose expression level was below the calculated thresholds in more than 50% of the samples were excluded for the purpose of building networks, as their expression level was close to detection limit. This criterion resulted in retaining 8,894 and 5,703 of the 16,006 probes present on the array for brain and muscle tissues, respectively, in the H1D backcross individuals. Normalization was performed using the Limma R package BetweenArrayNormalization function with the Quantile method (Smyth 2005).

Weighted Gene Coexpression Network Analysis

The R package WGNCA was used for network constructions (Langfelder and Horvath 2008). For nodes in the network to correspond to genes, redundant or highly similar probes (that formed contigs using >99% similarity over >70% of sequence length as assembly parameters in CLC Genomics Workbench 4.9) were collapsed using the collapseRow function in WGCNA (Miller et al. 2011) yielding 8,103 and 5,177 genes for brain and muscle backcross tissues, respectively. Because the module identification is computationally intensive, the top 3,600 most connected genes were used for constructing each network and for subsequent comparison with other data sets. This criterion retains the core of the modules, whereas genes that are only loosely attached to the network are excluded. Preliminary clustering of the gene expression profiles (supplementary fig. S9, Supplementary Material online) did not segregate males and females allowing a mixed gender network construction.

The various statistics computed by WGCNA and used in this study are defined in table 1. Briefly, WGCNA constructs networks using the absolute value of the Pearson's correlation

coefficient as the gene coexpression measure, which is raised to a power to create the adjacency matrix (supplementary fig. S10, Supplementary Material online). The topological overlap distance calculated from the adjacency matrix is then clustered with the average linkage hierarchical clustering. Our modules were defined using the cutTreeStatic function with a minimum module size of 25 genes and a cut height of 0.975 and 0.965 for muscle and brain network, respectively. A module eigengene distance threshold of 0.25 was also used to merge highly similar modules. These parameters allowed for detection of a minimum number of large modules while visually respecting the pattern of correlation with the phenotypic gradients. Module preservation statistics were computed using the modulePreservation function (500 permutations) implemented in WGCNA (Langfelder et al. 2011). Network module preservation statistics quantify how density and connectivity patterns of modules defined in a reference data set are preserved in a test data set without the need to define modules in the test data set. This strategy was applied to measure backcross module preservation (as reference data set) in pure dwarf and normal populations (test data sets) for brain tissue and in sockeye salmon populations for muscle tissue.

Gene Enrichment Analysis

We reannotated the cGRASP probes using Blast2GO software (Conesa et al. 2005), and probe sequences were blasted with *e*-value threshold of 10e-3 against the Swissprot database, then the nr database for the remaining unannotated genes. Fisher's exact test was performed for each module using the tissue-specific network as background with a *P* value threshold of 0.05.

Supplementary Material

Supplementary figures S1–S10 and files S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Sean Rogers and three anonymous reviewers for their comments. They are also grateful to Eric Normandeau for his bioinformatics assistance, helpful discussions, and suggestions. They thank Ben Koop, Trisha Lundrigan, Amber Messmer, Glenn Cooper, Kris von Schalburg, Adrienne Robb, and the Centre for Biomedical Research at the University of Victoria for laboratory space, advice, training, and discussion with laboratory techniques for the sockeye salmon data set. They thank Troy Hamon and the staff of Katmai National Park for the sockeye salmon fieldwork. The sockeye salmon field and laboratory work was supported by the National Park Service and the USGS. This work was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery grant program), the Canadian Research Chair in Genomics and Conservation of Aquatic Resources, Fonds de recherche du Québec - Santé (FRQS), and the research program of Québec Océan.

References

Aghdasi B, Ye K, Resnick A, Huang A, Ha HC, Guo X, Dawson TM, Dawson VL, Snyder SH. 2001. Fkbp12, the 12-kda fk506-binding protein, is a physiologic regulator of the cell cycle. *Proc Natl Acad Sci U S A.* 98:2425–2430.

Allen JD, Xie Y, Chen M, Girard L, Xiao G. 2012. Comparing statistical methods for constructing large scale gene networks. *PLoS One* 7: e29348.

Asano H, Li XS, Stamatoyannopoulos G. 1999. Fklf, a novel kruppel-like factor that activates human embryonic and fetal beta-like globin genes. *Mol Cell Biol.* 19:3571–3579.

Barabási AL, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 5:101–113.

Bernatchez L. 2004. Ecological theory of adaptive radiation: an empirical assessment from coregonine fishes (salmoniformes). In: Hendry AP, Stearns SC, editors. Evolution illuminated: salmon and their relatives. Oxford: Oxford University Press. p. 175–207.

Bernatchez L, Chouinard A, Lu GQ. 1999. Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, *Coregonus* sp., as a case study. *Biol J Linn Soc.* 68:173–194.

Bernatchez L, Renaut S, Whiteley AR, et al. (11 co-authors). 2010. On the origin of species: insights from the ecological genomics of lake whitefish. *Philos Trans R Soc Lond B Biol Sci.* 365:1783–1800.

Birchler JA, Veitia RA. 2010. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* 186:54–62.

Brawand D, Soumillon M, Necsulea A, et al. (18 co-authors). 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478:343–348.

Butlin R, Debelle A, Kerth C, et al. (26 co-authors). 2012. What do we need to know about speciation? *Trends Ecol Evol.* 27:27–39.

Carlson MR, Zhang B, Fang Z, Mischel PS, Horvath S, Nelson SF. 2006. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics* 7:40.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.

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

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

Evans ML, Bernatchez L. 2012. Oxidative phosphorylation gene transcription in whitefish species pairs reveals patterns of parallel and nonparallel physiological divergence. *J Evol Biol.* 25:1823–1834.

Evans ML, Praebel K, Peruzzi S, Bernatchez L. 2012. Parallelism in the oxygen transport system of the lake whitefish: the role of physiological divergence in ecological speciation. *Mol Ecol.* 16:4038–4050.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet.* 28:342–350.

Ficklin SP, Feltus FA. 2011. Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiol.* 156:1244–1256.

Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S. 2007. Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome.* 18:463–472.

Gao Y, Kobayashi H, Ganss B. 2003. The human KROX6/ZNF22 gene is expressed at sites of tooth formation and maps to the locus for permanent tooth agenesis (He-Zhao deficiency). *J Dent Res.* 82: 1002–1007.

He LN, Liu YJ, Xiao P, et al. (12 co-authors). 2008. Genomewide linkage scan for combined obesity phenotypes using principal component analysis. *Ann Hum Genet.* 72:319–326.

Hendry AP. 2009. Ecological speciation! Or the lack thereof? *Can J Fish Aqua Sci.* 66:1383–1398.

Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N. 2002. Revealing modular organization in the yeast transcriptional network. *Nat Genet.* 31:370–377.
Kadarmideen HN, Watson-Haigh NS, Andronicos NM. 2011. Systems biology of ovine intestinal parasite resistance: disease gene modules and biomarkers. *Mol Biosyst.* 7:235–246.
Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Pääbo S. 2004. A neutral model of transcriptome evolution. *PLoS Biol.* 2:E132.
Kimmel CB, Miller CT, Keynes RJ. 2001. Neural crest patterning and the evolution of the jaw. *J Anat.* 199:105–120.
Kugimiya F, Yano F, Ohba S, Igawa K, Nakamura K, Kawaguchi H, Chung UI. 2005. Mechanism of osteogenic induction by FK506 via BMP/Smad pathways. *Biochem Biophys Res Commun.* 338:872–879.
Kumar CG, Everts RE, Loor JJ, Lewin HA. 2010. Functional annotation of novel lineage-specific genes using co-expression and promoter analysis. *BMC Genomics* 11:161.
Landry CR, Hartl DL, Ranz JM. 2007. Genome clashes in hybrids: insights from gene expression. *Heredity* 99:483–493.
Langfelder P, Castellani LW, Zhou Z, Paul E, Davis R, Schadt EE, Lusis AJ, Horvath S, Mehrabian M. 2012. A systems genetic analysis of high density lipoprotein metabolism and network preservation across mouse models. *Biochim Biophys Acta.* 1821:435–447.
Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.
Langfelder P, Luo R, Oldham MC, Horvath S. 2011. Is my network module preserved and reproducible? *PLoS Comput Biol.* 7:e1001057.
Mallarino R, Grant PR, Grant BR, Herrel A, Kuo WP, Abzhanov A. 2011. Two developmental modules establish 3d beak-shape variation in Darwin's finches. *Proc Natl Acad Sci U S A.* 108:4057–4062.
Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, Horvath S. 2011. Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics* 12:322.
Miller JA, Oldham MC, Geschwind DH. 2008. A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci.* 28:1410–1420.
Minguez P, Dopazo J. 2011. Assessing the biological significance of gene expression signatures and co-expression modules by studying their network properties. *PLoS One* 6:e17474.
Nolte AW, Renaut 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.
Nosil P. 2012. Ecological speciation. New York: Oxford University Press.
Oldham MC, Horvath S, Geschwind DH. 2006. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci U S A.* 103:17973–17978.
Olson-Manning CF, Wagner MR, Mitchell-Olds T. 2012. Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat Rev Genet.* 13:867–877.
Parsons KJ, Albertson RC. 2009. Roles for Bmp4 and CaM1 in shaping the jaw: evo-devo and beyond. *Annu Rev Genet.* 43:369–388.
Pavey SA, Bernatchez L, Aubin-Horth N, Landry CR. 2012. What is needed for next-generation ecological and evolutionary genomics? *Trends Ecol Evol.* 27:673–678.
Pavey SA, Collin H, Nosil P, Rogers SM. 2010. The role of gene expression in ecological speciation. *Ann NY Acad Sci.* 1206:110–129.
Pavey SA, Hamon TR, Nielsen JL. 2007. Revisiting evolutionary dead ends in sockeye salmon (*Oncorhynchus nerka*) life history. *Can J Fish Aqua Sci.* 64:1199–1208.
Pavey SA, Nielsen JL, Hamon TR. 2010. Recent ecological divergence despite migration in sockeye salmon (*Oncorhynchus nerka*). *Evolution* 64:1773–1783.
Pavey SA, Sutherland BJ, Leong J, Robb A, von Schalburg K, Hamon TR, Koop BF, Nielsen JL. 2011. Ecological transcriptomics of lake-type and riverine sockeye salmon (*Oncorhynchus nerka*). *BMC Ecol.* 11:31.
Plaisier CL, Horvath S, Huertas-Vazquez A, Cruz-Bautista I, Herrera MF, Tusie-Luna T, Aguilar-Salinas C, Pajukanta P. 2009. A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genet.* 5:e1000642.
Pottin K, Hyacinthe C, Retaux S. 2010. Conservation, development, and function of a cement gland-like structure in the fish *Astyanax mexicanus*. *Proc Natl Acad Sci U S A.* 107:17256–17261.
Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution. *Proc Natl Acad Sci U S A.* 104:8605–8612.
Renaut S, Bernatchez L. 2011. Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Heredity* 106:1003–1011.
Renaut S, Nolte AW, Bernatchez L. 2009. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol Biol Evol.* 26:925–936.
Rogers SM, Bernatchez L. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* spp. Salmonidae) species pairs. *Mol Biol Evol.* 24:1423–1438.
Rogers SM, Gagnon V, Bernatchez L. 2002. Genetically based phenotype-environment association for swimming behavior in lake whitefish ecotypes (*Coregonus clupeaformis mitchill*). *Evolution* 56:2322–2329.
Rosen EY, Wexler EM, Versano R, et al. (11 co-authors). 2011. Functional genomic analyses identify pathways dysregulated by progranulin deficiency, implicating Wnt signaling. *Neuron* 71:1030–1042.
Schluter D. 2000. The ecology of adaptive radiation. New York: Oxford University Press.
Smyth GK. 2005. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and computational biology solutions using R and bioconductor. New York: Springer. p. 397–420.
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.
Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302:249–255.
Trudel M, Tremblay A, Schetagne R, Rasmussen JB. 2001. Why are dwarf fish so small? An energetic analysis of polymorphism in lake whitefish (*Coregonus clupeaformis*). *Can J Fish Aqua Sci.* 58:394–405.
Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos Trans R Soc Lond B Biol Sci.* 367:451–460.
von Schalburg KR, Rise ML, Cooper GA, Brown GD, Gibbs AR, Nelson CC, Davidson WS, Koop BF. 2005. Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics* 6:126.
Wei P, Blundon JA, Rong YQ, Zakharenko SS, Morgan JI. 2011. Impaired locomotor learning and altered cerebellar synaptic plasticity in pep-19/pcp4-null mice. *Mol Cell Biol.* 31:2838–2844.
Weston DJ, Gunter LE, Rogers A, Wullschleger SD. 2008. Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. *BMC Syst Biol.* 2:16.
Whiteley AR, Derome N, Rogers SM, St-Cyr J, Laroche J, Labbé A, Nolte A, Renaut S, Jeukens J, Bernatchez L. 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.
Williams SM, Haines JL. 2011. Correcting away the hidden heritability. *Ann Hum Genet.* 75:348–350.
Winden KD, Karsten SL, Bragin A, Kudo LC, Gehman L, Ruidera J, Geschwind DH, Engel J Jr. 2011. A systems level, functional genomics analysis of chronic epilepsy. *PLoS One* 6:e20763.
Wolf JBW, Lindell J, Backstrom N. 2010. Speciation genetics: current status and evolving approaches. *Philos Trans R Soc Lond B Biol Sci.* 365:1717–1733.
Yamashita M, Yabu T, Ojima N. 2010. Stress protein HSP70 in fish. *AquaBioSci Mono.* 3:111–141.
Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, Vu H, Tu Z, Brem RB, Bumgarner RE, Schadt EE. 2012. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol.* 10:e1001301.