RNA-seq Reveals Transcriptomic Shock Involving Transposable Elements Reactivation in Hybrids of Young Lake Whitefish Species

Anne-Marie Dion-Côté,*[1] Sébastien Renaut,[2] Eric Normandeau,[1] and Louis Bernatchez[1]

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

[2]Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, Canada

***Corresponding author:** E-mail: anne-marie.dion-cote.1@ulaval.ca.

Associate editor: Doris Bachtrog

Abstract

Identifying the molecular basis of reproductive isolation among diverging lineages represents an essential step toward understanding speciation in natural populations. Postzygotic barriers can lead to hybrid breakdown, a syndrome that has been documented in several systems, potentially involving the reactivation of transposable elements. In northeastern North America, two lake whitefish lineages have repeatedly colonized postglacial lakes ~12,000 years ago, and a dwarf limnetic species has evolved multiple times from the normal benthic species. Reproductive isolation is incomplete between them; viable hybrids can be generated in the laboratory but significant mortality occurs and is associated with a malformed phenotype in backcross embryos, thus revealing a hybrid breakdown syndrome. By means of RNA-seq analyses, the objective of this study was to determine which genes were misregulated in hybrids and rigorously test the hypothesis of transposable element reactivation. We compared the transcriptomic landscape in pure embryos, F1-hybrids, and healthy and malformed backcrosses at the late embryonic stage. Extensive expression differences consistent with previously documented adaptive divergence between pure normal and dwarf embryos were identified for the first time. Pronounced transcriptome-wide deregulation in malformed backcrosses was observed, with over 15% of transcripts differentially expressed in all comparisons, compared with 1.5% between pure parental forms. Convincing evidence of transposable elements and noncoding transcripts reactivation in malformed backcrosses is presented. We propose that hybrid breakdown likely results from extensive genomic incompatibilities, plausibly encompassing transposable elements. Combined with previous studies, these results reveal synergy among many reproductive barriers, thus maintaining divergence between these two young whitefish species.

Key words: **speciation, RNA-sequencing, transposable element, transcriptomic, whitefish, hybrid.**

Article

Introduction

Biodiversity is generated and maintained in different geographical settings through the process of speciation. In recent years, considerable efforts have been deployed toward the understanding of ecological sympatric speciation, a particular case of speciation where populations diverge, while still potentially freely interbreeding (Rundle and Nosil 2005; Schluter 2009; Nosil 2012). At the other end of the continuum, allopatric speciation occurs when populations diverge while geographically separated and is considered by modern evolutionary biologists to be the null hypothesis for speciation (Coyne and Orr 2004). Theory describes well the buildup and maintenance of accumulated divergence, with the Bateson–Dobzhansky–Muller (BDM) model of genetic incompatibilities at the heart of the process. Still, little is known about the specific molecular targets most likely maintaining "young" diverging lineages apart following secondary contact in natural populations, the "ultimate test" of allopatric speciation (Coyne and Orr 2004). Hence, key questions regarding the number and type of genes most likely to be involved in the speciation process remain. Accordingly, a better understanding of the mechanisms underlying recent speciation events are needed, especially in natural populations (Nosil and Schluter 2011; Marie Curie SPECIATION Network 2012).

Classically, reproductive barriers have been divided into pre- and postzygotic. Intrinsic postzygotic barriers, as described by the BDM model of genetic incompatibilities, result in hybrid sterility or inviability. These incompatibilities arise from accumulated genetic divergence, either neutral (Lynch 2007) or adaptive (Rundle and Nosil 2005). When secondary contact occurs, recombination can disrupt coadapted alleles occurring in parental forms and, according to the BDM model of genetic incompatibilities, this can lead to a breakdown of epistatic interactions, particularly so in post-F1-hybrid generations. Ultimately, this may results in strong postzygotic intrinsic reproductive barriers (Dobzhansky 1937; Muller 1942; Coyne and Orr 2004). Clearly, determining the genetic basis of intrinsic postzygotic reproductive barriers will help unravel how biodiversity is maintained as such barriers invariably prevent further gene flow among lineages and thus complete the speciation process (Muller 1939; Muller 1942). The best-documented examples of the molecular mechanisms underlying postreproductive barriers mostly involve one or

© The Author 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

a few genes in "old" species (i.e., several hundreds of thousands to a few millions of years), where speciation is nearly complete (Wittbrodt et al. 1989; Schartl et al. 1994; Barbash et al. 2003; Presgraves et al. 2003; but see Christie and Macnair 1984; Wright et al. 2013). Moreover, the combined action of time and linkage disequilibrium may eventually hinder the discovery of the causative mutations and associated postzygotic barriers (Via and West 2008). Therefore, investigating the molecular mechanisms underlying reproductive isolation in recently diverged natural populations may provide a better comprehension of the speciation process.

Recent work has provided evidence regarding the role of gene regulatory networks, epigenetics, and transposable elements in speciation, leading to a broader interpretation of the BDM model (reviewed in Abbott et al. 2013). In particular, numerous studies have suggested a crucial role for transposable elements in speciation (Rieseberg et al. 1995; O'Neill and Graves 1998; Labrador et al. 1999; Ungerer et al. 2006; Symonová et al. 2013). These studies generally support McClintock's hypothesis of a "genomic shock," which posits that transposable elements reactivation in hybrids results from the stress triggered by the admixture of two diverged genomes in a single individual (McClintock 1984). Early studies in *Drosophila melanogaster* have shown that a DNA transposon (*P element*) can lead to sterility in F1-progeny of dysgenic crosses (Kidwell et al. 1977; Rubin et al. 1982), likely as a result of the reactivation of multiple types of transposable elements (Khurana et al. 2011). Similarly, a recent study has shown the transcriptional reactivation of transposable elements in interspecific *Drosophila* hybrids (Kelleher et al. 2012). In marsupials, amplification of retroelements has been associated with genome-wide under-methylation and chromosome remodeling in hybrids (O'Neill and Graves 1998). Hence, evidence supporting the hypothesis that transposable element reactivation represents an important genetic mechanism underlying postzygotic hybridization barrier is rapidly mounting.

Lake whitefish represents an excellent model to study early postzygotic barriers because reproductive isolation is recent and remains incomplete in many populations. Natural populations from northeastern North America experienced a recent allopatric period during the Pleistocene glaciation (~60,000 years BP), followed by secondary contact in newly formed lakes once the ice sheet retreated (Bernatchez et al. 2010; Jacobsen et al. 2012). Lakes from the St. John River basin (Maine, USA and Québec, Canada) were colonized by two glacial lineages, so-called Atlantic and Acadian, the former evolving rapidly and repeatedly into a dwarf limnetic species from the normal benthic species (Pigeon et al. 1997). Early transcriptomic studies using heterologous cDNA microarrays have documented gene expression divergence linked to phenotypic divergence between dwarf and normal whitefish, both at the juvenile and adult stage, yet not at the embryonic one (Derome et al. 2006; St-Cyr et al. 2008; Whiteley et al. 2008; Nolte et al. 2009; Renaut et al. 2009; Jeukens et al. 2010). Given that hybrids can be generated and maintained under laboratory conditions, earlier studies have found strong postzygotic barriers between dwarf and normal whitefish, including an increased mortality of hybrid embryos, patterns of transgressive segregation comprising incoherent hybrid hatching times, as well as signs of segregation distortion in backcrosses, suggestive of BDM incompatibilities (Lu and Bernatchez 1998; Rogers and Bernatchez 2006; Rogers and Bernatchez 2007; Renaut et al. 2009; Renaut and Bernatchez 2011; Gagnaire et al. 2013). A proportion of backcross embryos (30–50% depending on crosses) has also been characterized by a unique malformed phenotype and displays extensive gene expression variance and underexpression of essential development genes (Renaut and Bernatchez 2011). Altogether, these observations suggest extensive genomic incompatibilities between normal and dwarf whitefish.

In the current study, the molecular phenotype of malformed backcross embryos was further characterized to disentangle the molecular mechanisms underlying post-zygotic hybrid incompatibilities. It was hypothesized that transposable elements would be reactivated in malformed backcross embryos and would be accompanied by the downregulation of key metabolic pathways in response to genomic stress. To test this hypothesis, gene expression was measured by means of RNA-sequencing in pure, F1-hybrids, healthy backcross, and malformed backcross embryos. Our results suggest a severe genomic shock involving transposable elements reactivation in the hybrid progeny, despite the very young age of these incipient species.

Results

Transcriptome Sequencing and Assembly

A total of 1.26×10^9 100-bp paired-end reads were sequenced, totalizing 1.26×10^{11} nucleotides. On average, $5.24 \times 10^7 \pm 6.23 \times 10^6$ paired-end reads were sequenced per library. Filtered individual reads (9.30×10^8 remaining reads) were assembled into 77,697 contigs with a minimum size of 200 bp using ABySS first for each library, then all libraries together, and finally with CLC genomics workbench (0.7 overlap, 98% similarity). The final transcriptome also included 1,350 contigs smaller than 200 bp and annotated as transposable elements, for a total of 79,047 contigs (total assembly size: 67,143,916 bp). This corresponds to roughly 2% of the expected whitefish genome size (3 Gb, Booke 1968; Hardie and Hebert 2003). The smallest contig was 94 bp and the largest 10,610 bp, with an average size of 565 bp and a median of 849 bp. A Blast of the assembled transcriptome against itself (BlastN, % identity >96%, overlap >100 bp) gave very few hits (n = 5,907), suggesting the assembly of numerous alternative transcripts, closely related paralog assembly, splitting of exons, or more likely a combination of the three.

Transcriptome Annotation

Approximately half of the contigs (n = 36,395, 46%) were successfully annotated by TBlastX (fig. 1). Among these annotations, 21,180 were unique (57.3%), consistent with

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

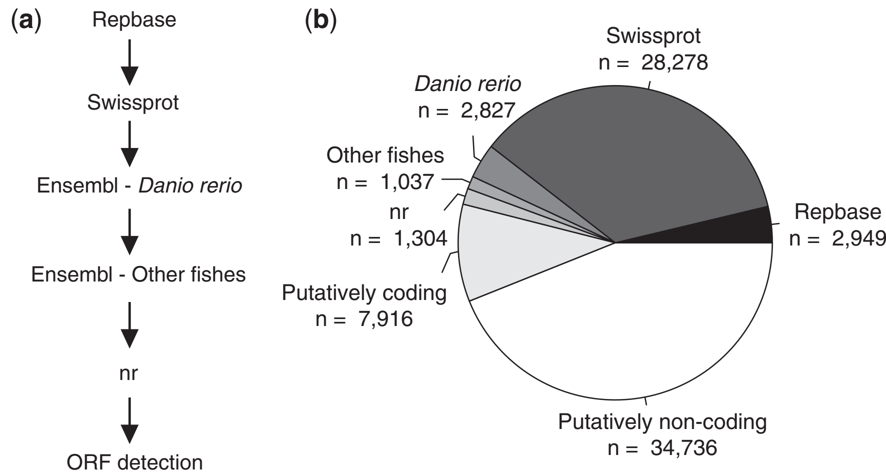

Fig. 1. Annotation summary of the assembled transcriptome. (*a*) Schematic view of the annotation procedure. (*b*) Pie chart showing the contribution of each database to the annotation. "Other fishes" category includes the combined Ensembl databases for *Gasterosteus*, *Oryzias*, *Takifugu*, and *Tetraodon*. Nonannotated contigs were classified as putatively coding if an ORF ≥ 300 bp was detected or putatively noncoding if an ORF ≥ 300 bp could not be detected.

the assembly of alternative transcripts, exon splitting, or closely related paralogs. An open reading frame (ORF) of size ≥300 bp was found for 71.3% of the contigs in the annotated portion (27,546 contigs out of 36,395), whereas an ORF was found for only 19.6% of the unannotated portion (7,917 contigs out of 42,652). Hence, contigs for which we found no TBlastX hit and no ORF ≥ 300 bp were annotated as "putatively noncoding" transcripts.

Annotations in figure 2 show that expressed transposable elements are largely composed of DNA transposons (*n* = 2,122, 72%) relative to retrotransposons (*n* = 823, 28%). Among transposable element superfamilies, the Mariner/Tc1 DNA transposons were the most prevalent (*n* = 2,003, 68%). The second most abundant are gypsy elements (*n* = 205, 7.0%), a class of LTR retrotransposons, followed by CR1 (*n* = 190, 6.5%), which are non-LTR retrotransposons. Other superfamilies (*n* = 550, 18.7%) individually account for less than 5% of the annotated expressed transposable elements.

Transcriptional Divergence between Pure Normal and Dwarf Embryos

To provide an overview of the transcriptomic landscape and reduce the number of dimensions of this large data set, a principal component analysis (PCA) was performed with the *prcomp* function in the R environment, based on normalized read counts obtained from edgeR. While normal and dwarf replicates closely cluster on the two first PCs (fig. 3*a*), PC3 shows that the dwarf transcriptomes form a distinct and more cohesive group as compared with the normal transcriptomes (fig. 3*b*). However, PC3 explained a relatively small proportion of the overall variance found in the data set (<7.5%).

Table 1 provides a summary of the number of differentially expressed transcripts between all groups for all contigs (below diagonal) and transposable elements (TE, above diagonal). Briefly, 1,166 transcripts were differentially expressed between dwarf and normal whitefish embryos, including 314 by at least

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

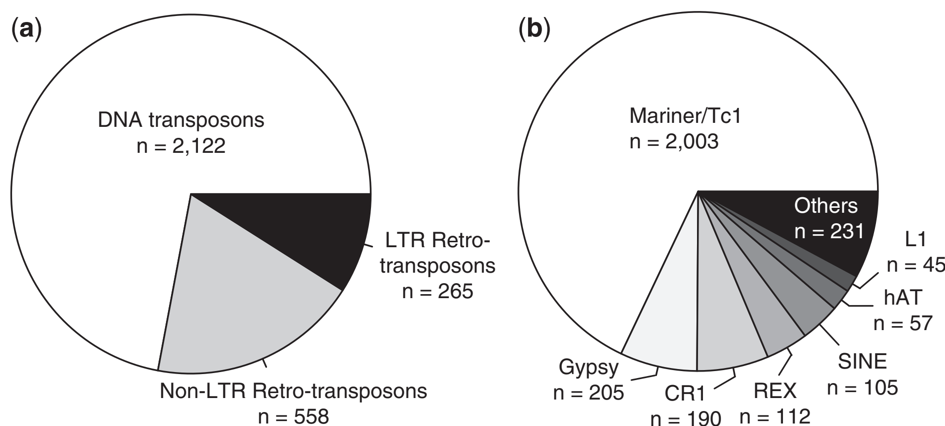

Fig. 2. Transposable elements annotation. (*a*) Transposable element type representation (*n* = 2,945 contigs). (*b*) Transposable element superfamily representation according to Kapitonov and Jurka (2008).

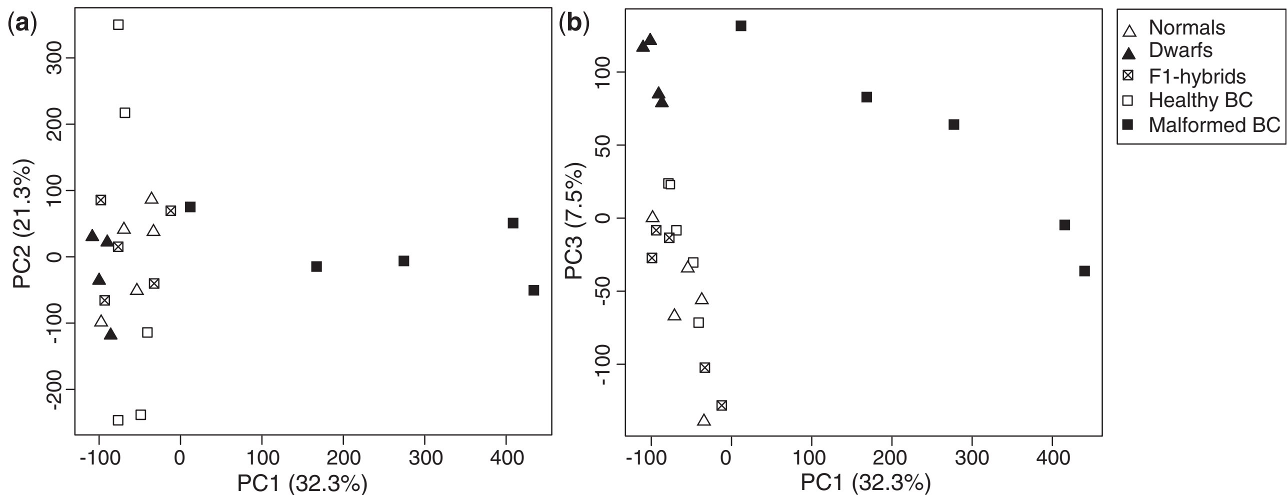

Fig. 3. Unique transcription profile observed in backcrosses and malformed backcrosses in particular. PCA as performed by the R function *prcomp*. (*a*) PC1 (32.3% of the variance) and PC2 (21.3 % of the variance). (*b*) PC1 (32.3% of the variance) and PC3 (7.5% of the variance). Normal whitefish embryos are depicted by empty triangles, dwarfs by filled triangles, F1-hybrids by checked squares, healthy backcrosses by empty squares, and malformed backcrosses by filled squares. BC: backcrosses.

Table 1. Differential Expression Summary.

	TE				
All Contigs	**Normal**	**Dwarf**	**F1-Hybrid**	**Healthy BC**	**Malformed BC**
Normal		83	22	26	350
Dwarf	1,166		31	64	400
F1-Hybrid	108	653		40	354
Healthy BC	248	967	436		317
Malformed BC	13,217	15,815	13,710	13,455	

Note.—Number of differentially expressed contigs for each comparison (FDR < 0.01, FC > 2). Above the main diagonal (shadowed) are the number of contigs annotated as transposable elements (TE) and below are all contigs. BC: backcross.

5-fold and 28 by more than 1000-fold (supplementary fig. S2, Supplementary Material online). Enriched Gene Ontology (GO) biological processes included terms related to muscle development and vision, in addition to cardiac development (table 2), and the term "Mitotic chromosome condensation." Phosphofructokinase, a key enzyme in glycolysis that catalyzes an irreversible reaction, was upregulated in dwarf as compared with normal whitefish (supplementary fig. S5, Supplementary Material online).

Gene Expression in Healthy Hybrids

As depicted in figure 3*a*, PC1 did not differentiate F1-hybrids and healthy backcrosses from pure embryos. However, healthy backcrosses displayed extensive transcriptional variability, comprising a range of expression largely outside the scope of the pure embryos on PC2. Also, PC3 showed that both F1-hybrids and healthy backcrosses were more similar to normal than dwarf embryos, possibly reflecting maternal effects (see Discussion). Differential expression analysis confirmed observations based on PCA. More transcripts were differentially expressed in F1-hybrids when compared to dwarfs ($n = 653$) than to normals ($n = 108$). Similarly, healthy backcrosses had more differentially expressed transcripts when compared to dwarf ($n = 967$) than to normal ($n = 248$).

Transcriptome-Wide Deregulation in Malformed Backcrosses

As shown in figure 3*a*, PC1 clearly distinguishes malformed backcrosses from all other healthy embryos. In addition, malformed backcrosses show pronounced heterogeneity among replicates relative to other cross types. A large number of transcripts were differentially expressed in malformed backcrosses compared with pure crosses, ranging from 17.5% ($n = 13{,}217$, compared with normal) to 21.0% ($n = 15{,}815$, compared with dwarf) of the expressed transcripts.

The proportions of overexpressed and underexpressed transcripts in malformed backcrosses are presented in table 3. The "all contigs" column shows that the proportion of over- vs. underexpressed transcripts does not significantly differ from a 1:1 ratio in all comparisons. Among overexpressed transcripts in malformed backcrosses relative to all other groups ("Common" column, $n = 4{,}379$), 25 enriched GO terms were found, most of which grouped into broader categories defined as protein synthesis, inflammatory response, and stress response (supplementary table S1, Supplementary Material online). In addition, enzymes from core metabolic pathways were downregulated in malformed embryos. For example, many downregulated enzymes were found in the carbohydrate metabolism pathway (supplementary fig. S5, Supplementary Material online). Similar trends of downregulation of housekeeping metabolic pathways were also observed for the TCA cycle, oxidative phosphorylation, fatty acid elongation, and purine and pyrimidine metabolism (supplementary figs. S6–S10, Supplementary Material online).

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

Table 2. GO Enrichment (biological processes, FDR < 0.05) for Differentially Expressed Transcripts between Pure Normal and Dwarf Whitefish Embryos (FDR < 0.01, FC > 2).

GO-ID	Term	FDR	No. Contigs
GO:0048769	Sarcomerogenesis	<0.0001	13
GO:0030240	Skeletal muscle thin filament assembly	<0.0001	13
GO:0030241	Skeletal muscle myosin thick filament assembly	<0.0001	13
GO:0055003	Cardiac myofibril assembly	<0.0001	14
GO:0048739	Cardiac muscle fiber development	<0.0001	13
GO:0045214	Sarcomere organization	<0.0001	14
GO:0055008	Cardiac muscle tissue morphogenesis	<0.0001	14
GO:0043056	Forward locomotion	<0.0001	9
GO:0051592	Response to calcium ion	<0.0001	13
GO:0007512	Adult heart development	0.0001	9
GO:0018298	Protein-chromophore linkage	0.0001	5
GO:0007076	Mitotic chromosome condensation	0.0008	6
GO:0090212	Negative regulation of establishment of blood-brain barrier	0.0011	3
GO:0007186	G-protein coupled receptor signaling pathway	0.0036	18
GO:0015949	Nucleobase-containing small molecule interconversion	0.0045	3
GO:0045859	Regulation of protein kinase activity	0.0049	21
GO:0060041	Retina development in camera-type eye	0.0066	9
GO:0001701	In utero embryonic development	0.0138	13
GO:0006956	Complement activation	0.0155	4
GO:0006565	L-Serine catabolic process	0.0156	2
GO:0001580	Detection of chemical stimulus involved in sensory perception of bitter taste	0.0156	2
GO:0009586	Rhodopsin mediated phototransduction	0.0156	2
GO:0046724	Oxalic acid secretion	0.0156	2
GO:0002576	Platelet degranulation	0.0174	6
GO:0006775	Fat-soluble vitamin metabolic process	0.0207	5
GO:0009268	Response to pH	0.0344	3
GO:0048871	Multicellular organismal homeostasis	0.0353	6
GO:0030049	Muscle filament sliding	0.0353	6
GO:0019265	Glycine biosynthetic process, by transamination of glyoxylate	0.0375	2
GO:0006941	Striated muscle contraction	0.0381	7
GO:0072521	Purine-containing compound metabolic process	0.0484	27
GO:0006813	Potassium ion transport	0.0500	7

Table 3. Overexpression of Transposable Elements and Noncoding Transcripts in Malformed Backcrosses.

	All Contigs	Transposable Elements	Noncoding
Normal	13,217 (50.6%/49.4%)	350 (70.6%/29.4%)***	6,007 (65.5%/34.5%)***
Dwarf	15,815 (51.8%/48.2%)	400 (70.3%/29.7%)***	7,111 (71.3%/28.7%)***
F1-hybrid	13,710 (50.6%/49.4%)	354 (63.6%/36.4%)***	6,312 (67.3%/32.7%)***
Healthy BC	13,455 (50.5%/49.5%)	317 (66.6%/33.4%)***	6,315 (62.8%/37.2%)***
Common	8,096 (54.1%/45.9%)	177 (70.6%/29.4%)***	3,553 (71.3%/28.7%)***

NOTE.—Number of differentially expressed transcripts for all contigs, transposable elements only and noncoding transcripts only in malformed backcrosses compared with other groups (FDR < 0.01, FC > 2). The last line shows the number of transcripts that are differentially expressed in all four comparisons. Percentages of transcripts over- or underexpressed in each category are given within parentheses. BC: backcross.

***All P values < 0.0001 (Fisher's exact test).

Transposable Elements and Noncoding RNA Reactivation in Malformed Backcrosses

In contrast to the even expression ratio observed among all contigs, transposable elements showed a highly significant excess of overexpression ($P < 0.0001$ in all comparisons, Fisher's exact test), with an approximate 2:1 ratio in all comparisons (table 3). Depending on the comparison, several transcripts were overexpressed by a factor of 10 or more relative to other embryo groups, including one transposable element showing a 232-fold increase compared with dwarf (supplementary fig. S3, Supplementary Material online). Among the overexpressed transposable elements in malformed backcrosses that were common to all comparisons, retrotransposons were overrepresented ($n = 51$ out of 125 overexpressed transposable elements [40.8%] vs. $n = 823$ out of 2,945 annotated transposable elements [27.9%]; $P = 0.032$, Fisher's exact test). However, only two superfamilies of retrotransposons (L1, $n = 45$ contigs out of 2,945 annotated transposable elements [1.5%] vs. $n = 3$ contigs out of 125 commonly overexpressed

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

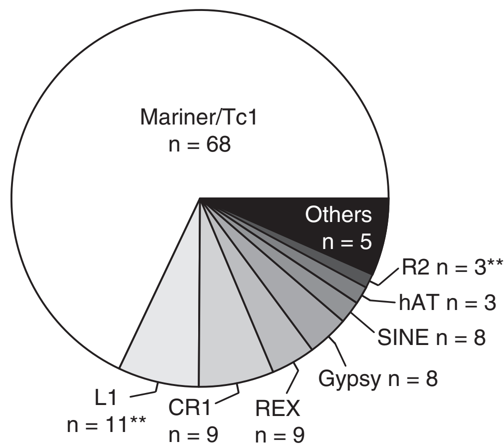

Fig. 4. Overexpressed transposable element superfamilies in malformed backcrosses in all comparisons ($n = 125$, FDR < 0.01, FC > 2). L1 and R2 superfamilies are significantly overrepresented as compared with the whole annotated transcriptome (**$P < 0.005$, Fisher's exact test).

transposable elements [2.4%]; R2, $n = 11$ out of 2,945 annotated transposable elements [3.7%] vs. $n = 3$ out of 125 commonly overexpressed transposable elements [2.4%]) were significantly overrepresented ($P = 1.81e^{-05}$ and $P = 0.0033$, respectively, Fisher's exact test) (fig. 4).

Similarly to transposable elements, putatively noncoding RNA displayed a highly significant excess of overexpression in malformed backcrosses, again with an approximate 2:1 expression ratio compared with all other cross types (table 3). Finally, a contig homologous to the DNA methyltransferase DNMT1 was significantly underexpressed in malformed backcrosses in all comparisons (fold change [FC] = 0.43, 0.40, 0.41, and 0.43 vs. normal, dwarf, F1-hybrids, and healthy backcrosses, respectively; all false discovery rates [FDRs] ≤ 0.00001). This gene is of particular interest because it has been shown to be involved in the transcriptional repression of transposable elements (see Discussion).

Discussion

The goal of this study was to identify the transcriptional mechanisms underlying the genomic shock previously described in hybrids of dwarf and normal whitefish. Specifically, we tested the hypothesis that transposable elements reactivation may be linked to the genomic shock observed in malformed backcross embryos. Key transcriptional differences were identified between normal and dwarf embryos, consistent with well-described phenotypes in adult normal and dwarf whitefish. Transcriptome-wide deregulation observed in malformed backcross embryos was associated with a shutdown of core metabolic pathways. In particular, observations supporting transposable elements and noncoding transcripts reactivation in malformed backcrosses are presented, giving a new angle to our understanding of the genetic mechanisms underlying postzygotic barriers in lake whitefish.

Transcriptome Assembly and Annotation

We assembled the most comprehensive and complete transcriptome published to date for a coregonid species (Jeukens et al. 2010; Renaut et al. 2010), providing a useful tool to the salmonids research community. The number of contigs assembled was relatively high ($n = 79{,}047$), which can be explained in several ways. First, all individuals were used to assemble the reference transcriptome, including hybrids, in order to consider "biologically aberrant" transcripts that may be expressed only in hybrids. In addition, salmonids are pseudo-tetraploids, having undergone a recent whole-genome duplication (~60 Ma; Crête-Lafrenière et al. 2012), potentially increasing the number of assembled contigs. Evidence also suggests exon splitting among assembled contigs (data not shown), but investigating the details of this phenomenon was beyond the scope of the current study. Considering the difficulties regarding gDNA assembly with next-generation DNA sequencing data encountered in whitefish and also another salmonids (Quinn et al. 2008; Hebert et al. 2013), we believe our assembly efforts are satisfactory for the questions being addressed here.

Noncoding transcripts can account for up to ±50% of the polyadenylated fraction of the transcriptome in well-annotated species such as human and yeast (Nagalakshmi et al. 2008; Djebali et al. 2012). Here, putatively noncoding transcripts accounted for 43.9% of the assembled contigs. This suggests that the vast majority of the contigs in the unannotated portion are indeed noncoding RNAs and reflects substantial expression of noncoding RNAs, as previously found in different model species (Clark et al. 2011).

Transcriptional and Regulatory Divergence between Dwarf and Normal Whitefish Embryos

We present the first significant transcriptional differences between dwarf and normal whitefish embryos. Specifically, we found that about 1.5% of the transcriptome was differentially expressed ($n = 1{,}166$), consistent with regulatory divergence between these young species very early in their ontogeny. In contrast, previous work using heterologous microarrays developed for Atlantic salmon (*Salmo salar*) identified few differentially expressed genes between normal and dwarf embryos (Renaut et al. 2009, $n = 5$; Renaut and Bernatchez 2011, $n = 2$).

In a study comparing RNA-sequencing to gene expression arrays in human, Marioni et al. (2008) detected more differentially expressed transcripts by RNA-seq than microarrays (but see Kogenaru et al. 2012). Hence, the discrepancy between this study and previous work could be caused by a combination of factors. First, RNA-seq has a marked increased dynamic range compared with microarrays (Nagalakshmi et al. 2008). Other factors inherent to microarray technology used by Renaut et al. (2009) and Renaut and Bernatchez (2011) include probe absence on the microarray itself or divergence with the studied species, loop design, and multiple comparisons (Kerr and Churchill 2001). Very few studies, if any in nonmodel species, have made comparisons of both methods with similar samples, making it difficult to definitely

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

conclude on the causes of the observed discrepancy. Nevertheless, our work suggests a significantly increased power of RNA-sequencing in detecting specific differentially expressed genes compared with microarrays in nonmodel species.

While common garden experiments have previously shown that phenotypic divergence is partially genetically determined in lake whitefish adults (Rogers et al. 2002), our results now illustrate that genetic divergence also has a transcriptional impact very early in ontogeny. GO analysis among differentially expressed transcripts revealed that skeletal muscle development genes were enriched between dwarf and normal embryos. This result is consistent with previous transcriptomic studies on adult white muscle (Derome et al. 2006; St-Cyr et al. 2008; Jeukens et al. 2010), phenotypic differences in adult size, and growth rate in dwarf and normal whitefish (Bernatchez et al. 2010). A new candidate physiological process potentially involved in phenotypic divergence between dwarf and normal embryos was also identified, namely cardiac development. This is of particular interest, as Trudel et al. (2001) found that dwarf whitefish have lower food conversion efficiency than normal whitefish, suggesting a higher dwarf metabolic rate. As cardiac output is tightly linked to metabolic demand in fishes, cardiac physiological changes are thus expected in lake whitefish (Farrell and Jones 1992). Along these lines, previous work has shown a trend toward increased ventricle size in dwarf compared with normal lake whitefish (Evans et al. 2013). Our finding that genes associated with cardiac development are differentially expressed between dwarf and normal embryos thus suggests that changes in heart development may contribute to metabolic divergence.

Alternatively, gene expression differences observed between dwarf and normal embryos could be the result of different stages of development or tissue compositions rather than being due to genetic differences. However, the number of genes that were differentially expressed between pure dwarf and normal whitefish is larger ($n = 1,166$) than the total number of genes that have been reported to differ in expression across "all" major embryonic developmental stages in zebrafish ($n = 732$, Mathavan et al. 2005). While the results on zebrafish and ours cannot be directly compared as they were obtained with different methods, they certainly provide an indication that most of the genes differentially expressed between dwarf and normal embryos are indeed the result of genetic differences rather than different stage of development. In addition, the fact that embryos of both forms used in this study corresponded to a well-described phenotypically developmental stage, and that genes underlying these differences are in agreement with the adaptive phenotypic divergence between these two forms again argues against the sole effect of tissue composition or development. Hence, we do not rule out a possible effect of tissue composition or development, potentially amplifying observed patterns. However, we argue that this is not the main explanation for such important gene expression divergence and misregulation.

Gene Expression in Healthy Hybrids

The hypothesis that gene expression patterns in hybrids are more similar to normal than dwarf embryos was verified, as expected based on previous observations in juveniles (Nolte et al. 2009). This is likely the consequence of the genetic background of backcrosses (~75% normal background) and maternal effects on gene expression in F1-hybrids, as found in hybrids of another salmonid (brook char, *Salvelinus fontinalis*, Bougas et al. 2010). In backcrosses, this could also result from segregation distortion disfavoring dwarf alleles in a normal genetic background, and this phenomenon has been previously documented in lake whitefish hybrids (Rogers and Bernatchez 2006; Gagnaire et al. 2013). Results presented here are supported by a previous study of inheritance patterns in juvenile lake whitefish (supplementary fig. S4, , Supplementary Material online; Renaut et al. 2009). This study revealed primarily additive patterns of gene expression in F1-hybrids, while backcross hybrids were characterized by extensive nonadditive gene expression and considerable gene expression variance (Renaut et al. 2009; Renaut and Bernatchez 2011). This could result from a combination of dominance effects and disruption of coadapted alleles through recombination events, ultimately leading to gene misexpression (Landry et al. 2007). On the other hand, genes with large pleiotropic effects, generally assumed to be transcription factors, could also lead to extensive patterns of gene misexpression. Along those lines, previous eQTL mapping studies in whitefish have localized several mapping hotspots (up to 53 eQTL mapping to a single locus), consistent with master regulator localization (Derome et al. 2008; Whiteley et al. 2008). Admittedly, our study design was not suitable for direct testing of these alternative hypotheses. Considering the large number of differentially expressed genes in hybrids, multiple genetic mechanisms are likely contributing to these gene expression profiles, as observed in hybrids of another salmonid (brook char, *S. fontinalis*; Bougas et al. 2010).

Evidence for a Transcriptome-Wide Shock in Malformed Backcrosses

Hybrid incompatibilities, reflected by the disruption of gene expression pathways, can lead to increased mortality and transgressive phenotypes in recombinant backgrounds (Dobzhansky 1940; Muller 1942). This phenomenon has been documented before in lake whitefish backcrosses through the characterization of a malformed phenotype unique to recombinant hybrids and by transcriptomic studies using microarrays (Renaut et al. 2009; Renaut and Bernatchez 2011). It was thus hypothesized that malformed backcross hybrids would show extensive transcriptional variance and differential expression. The observations of this study have confirmed this pattern.

Given that both healthy and malformed backcross embryos come from the same mother, our results cannot be attributed to maternal effects specific to malformed embryos. In addition, the same malformed phenotype was observed in two completely independent backcrosses, although in

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

variable proportions (normal female X F1-hybrid male, dwarf female X F1-hybrid male, Dion-Côté A-M, unpublished data). The PCA showed a unique transcriptional profile in malformed backcrosses and high heterogeneity among different pools of embryos (fig. 3). Differential expression with pure parental crosses also revealed extensive transcriptional deregulation in these embryos. In addition, overexpressed transcripts in malformed backcrosses were enriched in biological processes associated with stress and inflammatory responses, protein synthesis, and lipid metabolism, consistent with a genomic shock response (supplementary table S1, Supplementary Material online). We also observed that genes in core metabolic pathways were downregulated in these malformed backcrosses, complementing previous observation of downregulation of essential developmental genes (Renaut and Bernatchez 2011). Finally, the majority of the transcripts analyzed ($n = 799$ out of 1,166, 68.5%) showed a nonadditive level of expression in malformed backcrosses ($|$d/a ratio$| > 1.5$, supplementary fig. S4, Supplementary Material online). Altogether, these results provide new insights into the malformed phenotype observed in backcrosses by showing that the genomic shock involves a displacement of cellular activity from basal metabolism to stress response. In summary, our results are in line with genome-wide transcriptional BDM incompatibilities but could alternatively result from a few major regulatory genes impacting the whole genome (Landry et al. 2007; Maheshwari and Barbash 2011). However, as discussed earlier, it appears more likely that multiple intertwined genetic mechanisms generate these complex transcriptional patterns, including transposable elements reactivation, as discussed later.

An alternative explanation could be that the transcriptome-wide deregulation in malformed backcrosses results from different stages of development or tissue composition. Again, we do not rule out the possible contribution of these factors; however, we argue that they are not the main explanation for the observed pattern of transcriptome-wide deregulation. The total number of genes found to be differentially expressed across all major embryonic developmental stages in zebrafish ($n = 732$, Mathavan et al. 2005) is one order of magnitude lower than observed here ($n = 13,217$ to $n = 15,815$). In addition, a previous study found evidence of transcriptional deregulation, including transposable element reactivation, in individually dissected tissues from adult backcrosses (Renaut et al. 2010). Moreover, we found evidence of increased variance and elevated transgressivity in malformed backcrosses as compared with other groups, consistent with genetic incompatibilities. For all these reasons, we believe that the magnitude of the changes observed, in terms of number of transcripts differentially expressed and the fold change involved, cannot be explained solely by differential tissue distribution or sampling stage.

Transposable Elements and Noncoding Transcripts Reactivation in Hybrids

A preliminary study in lake whitefish has presented evidence that transposable elements were reactivated in adult backcrosses (Renaut et al. 2010). These elements must be tightly regulated in order to prevent their mobilization, which can lead to several deleterious effects such as chromosomal rearrangements (Lönnig and Saedler 2002), epigenetic alterations (Slotkin and Martienssen 2007), and the generation of DNA damage and apoptosis (Belgnaoui et al. 2006; Khurana et al. 2011). Transposable element regulation is achieved typically through cytosine methylation mediated by DNA methyltransferases (DNMT) and by RNA-based mechanisms, mainly by PIWI protein and PIWI-interacting RNA (piRNA) in the gonads, leading to their stable repression in the progeny (Slotkin and Martienssen 2007). Our results clearly show that transposable elements and noncoding transcripts are reactivated in malformed embryos, often by a factor of 10 or more, and as much as 232-fold. More than 6,000 noncoding transcripts are differentially expressed in all comparisons, with over 62% of these being overexpressed; such differences are very unlikely to be the mere result of different tissue composition, as discussed earlier. Moreover, the reactivation of transposable elements is expected to lead to gene deregulation because they can influence epigenetic marks in their genetic neighborhood (Slotkin and Martienssen 2007). We believe that transposable elements could be a key component of postzygotic isolation mechanisms in the lake whitefish system given that 1) around 14,000 transcripts are differentially expressed in malformed backcrosses, regardless of the comparison, 2) among these, more than 8,000 transcripts are commonly differentially expressed in all comparisons, 3) many transposable elements are reactivated by over 10-fold, including one by as much as 232-fold, and finally 4) a very strict cut-off was used to identify these (FC > 2, FDR 0.01). Hence, we propose that transposable elements represent very strong alternative candidates, in addition to genes involved in classical BDM incompatibilities, to explain the genome-wide gene expression disruption (Whitelaw and Martin 2001) in malformed hybrids.

Reactivation of transposable elements in malformed backcrosses has been observed previously in other interspecific crosses (Michalak 2009), but to our knowledge, this has not been demonstrated in such young vertebrate species (~12,000 years). We have shown here that many superfamilies of transposable elements were reactivated in malformed hybrids. Similar to the *P*-element system in *D. melanogaster*, this could suggest that a given lineage possesses a transposable element for which the other lineage is naive, leading to piRNA pathway shutdown and consequent reactivation of multiple superfamilies of transposable elements (Khurana et al. 2011). On the other hand, our results also support those obtained for *Drosophila* interspecific crosses where transposable element reactivation was shown to result from piRNA pathway protein divergence (Kelleher et al. 2012). Unfortunately, at this stage of our research program, it is impossible to distinguish between these two hypotheses. Interestingly, the DNA methyltransferase DNMT1, partly responsible for repression of repetitive elements, including transposable elements, through DNA methylation (Jones 2012), was downregulated in malformed backcrosses. DNMT1 underexpression may therefore also contribute to transposable elements and

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

noncoding transcripts reactivation. Altogether, our data argue for a severe genomic shock between recently diverged normal and dwarf whitefish resulting from a combination of divergence in regulatory mechanisms, in addition to a role for transposable elements and noncoding transcripts reactivation in hybrids.

Multiple Factors Contributing to Speciation in Lake Whitefish

The present work adds to a long-term research program that has highlighted multiple pre- and postzygotic barriers between dwarf and normal whitefish, some of which can be indirectly linked to the short period of geographic isolation (Bernatchez et al. 2010). Postzygotic barriers include incoherent hatching time in hybrids (Rogers and Bernatchez 2006), decreased sperm performance in backcrosses (Whiteley et al. 2009), genome-wide segregation distortion (Rogers and Bernatchez 2006; Gagnaire et al. 2013), and increased hybrid mortality throughout development (Lu and Bernatchez 1998; Rogers and Bernatchez 2006; Renaut and Bernatchez 2011). Here, major gene expression disruption previously documented in hybrids was confirmed (Renaut and Bernatchez 2011), and new insights regarding core metabolic pathways downregulation and transposable element reactivation in hybrids were provided. These transcriptional mechanisms appear to be related to a malformed phenotype and premature death in hybrid embryos.

Despite their young evolutionary age, dwarf and normal whitefish are at an advanced (albeit incomplete) stage along the speciation continuum, most likely because multiple mechanisms are acting synergistically to maintain and promote their divergence. This is consistent with the multifarious hypothesis advanced by Nosil et al. (2009), suggesting that speciation is likely to proceed faster and further where selection is acting on multiple traits. As argued by us and others, ~60,000 years of allopatry (~15,000–20,000 generations), followed by secondary contact and ecological divergence ~12,000 years ago (~3,000–4,000 generations) have generated sufficient genomic divergence between normal and dwarf whitefish to lead to one of the most complete speciation event involving young sympatric species pairs of North temperate freshwater fishes (Hendry 2009; Bernatchez et al. 2010).

Other systems provide compelling evidences that multiple barriers can act synergistically to promote speciation. Studies on the walking stick *Timema cristinae* have revealed that both geographical and ecological components are likely to have played a role (Nosil 2007). Nevertheless, speciation is less advanced in *Timema* as compared with dwarf and normal whitefish, as suggested by a lack of evidence for hybrid inviability, despite longer divergence time (~1–2 My vs. 60,000 years; Nosil 2007). Likewise, many reproductive barriers have also been documented in the *Ficedula* flycatcher system, including strong intrinsic postzygotic barriers, despite little ecological divergence (Saetre and Saether 2010). Similarly to the lake whitefish, *Ficedula* flycatchers have experienced allopatric isolation during the Pleistocene glaciation (~1.5–2 Ma), followed by secondary contact in central Europe. As portrayed by the *Ficedula*, *Timema*, and lake whitefish systems, mixed geographic modes of divergence tend to accelerate the divergence process (Nosil 2012). Where multiple factors, either adaptive or not, act synergistically, divergence is likely to be more pronounced and speciation to proceed faster (Nosil 2007). Clearly, none of these factors should be neglected, whether ecological or nonadaptive, to realize a holistic understanding of the mechanisms underlying incipient speciation. As White (1978) stressed more than 35 years ago, "speciation is the result of the combined action and interaction of many processes, and any model that relies exclusively on a single process is bound to be simplistic."

Material and Methods

Crosses and Sampling

Protocols for crosses and rearing were described previously (Nolte et al. 2009; Renaut et al. 2009; Renaut and Bernatchez 2011). While individuals used in the current study come from the same crosses generated during Fall 2006, biological samples are distinct and represent an older developmental stage than the one previously investigated. Briefly, pure half-sib families were created by crossing one female with five males. F1-hybrids were obtained by crossing one normal female (the same as pure normal) to five dwarf males (the same as pure dwarf). Backcrosses were made with a F1-hybrid laboratory female crossed to five normal males. Due to sexual products availability, it was impossible to do the complementary crosses. However, earlier studies have shown similar mortality and phenotypes for both types of crosses (Lu and Bernatchez 1998; Rogers and Bernatchez 2006). Embryos were individually sampled and inspected between 62 and 69 days postfertilization (between 310 and 345 degree-days, Renaut and Bernatchez 2011) and stored in RNAlater at $-20\,^{\circ}$C. This corresponds to the postphylotypic stage, where organs are well formed, heartbeat is visible, and eyes and dorsal line are pigmented. Malformed embryos showed strong deformities, with tail generally curved and no heartbeat detectable visually, as in Renaut and Bernatchez (2011). This phenotype does not correspond to any earlier stage of development observed in normally developing embryos, while still depicting characteristics of the phylotypic stage, including eye and dorsal line pigmentation. Malformed embryos can be easily distinguished from dead ones. To confirm viability, a comparative group of malformed embryos was followed for several weeks, and continued development was observed (see Renaut and Bernatchez 2011).

RNA Extraction, Library Preparation and Sequencing

Whole embryos were homogenized using a TissueLyser II (Qiagen, Hilden, Germany), and total RNA was extracted using the PureLink RNA Mini Kit following the manufacturer's instructions (Ambion, Life Technologies, Carlsbad, CA, USA). RNA was quantified with a NanoDrop2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA), and 2 µg of total RNA was digested with 4U of DNase I (Invitrogen, Life technologies, Carlsbad, United States) for 15 min at room temperature. DNaseI was inactivated by the addition of

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

2.5 mM EDTA (final concentration) and incubated at 65 °C for 10 min. RNA quality was assessed using the Experion RNA StdSens kit (Bio-Rad, Hercules, CA, USA). Only high-quality samples (intact rRNA and no detectable trace of gDNA) were kept for subsequent steps. RNA concentration was measured with Quant-iT RiboGreen RNA Assay Kit (Invitrogen, Life Technologies, Carlsbad, CA, USA) and RNA stored at −80 °C.

Each sequencing library was prepared from a pool of 400 ng of total RNA from three embryos, for a total of 1,200 ng. Five libraries were prepared for each group (normal, F1-hybrid, healthy backcross, and malformed backcross), except dwarf ($n = 4$, due to sample limitation). Libraries were individually tagged using the TruSeq RNA sample preparation kit V2 (Illumina, San Diego, CA, USA) following the manufacturer's instructions with one minor modification. Briefly, RNA was eluted, fragmented, and primed at 94 °C for 2 min in order to get the appropriate size range (300–700 bases). Library size and concentration were evaluated using the Experion RNA HighSens kit (Bio-Rad, Hercules, CA, USA). Sequencing was performed on the Illumina HiSeq 2000 platform for 100 cycles with paired-ends (6 libraries per lane for a total of 4 lanes), at the McGill University and Genome Quebec Innovation Centre, Montreal, Canada.

De Novo Assembly and Annotation

Tags and adaptor sequences were removed (Genome Quebec Innovation Centre) and reads were trimmed with a quality threshold of 2% error rate per base using CLC genomics workbench (CLCbio, Aarhus, Denmark). The first assembly was done with cleaned reads of ≥ 90 bp and comprising no more than one ambiguous base (1N) using ABySS 1.3.3 (Simpson et al. 2009), one library at a time, with k-mer sizes of 34, 44, 54, and 64 bp (abyss-pe q = 10 e = 5 c = 5 s = 200 d = 50 j = 4). Scaffolds were removed as they represented less than 1% of the contigs. Contigs >80 bp were kept and segmented in 100 and 200 bases fragments with 75% overlap. Fragmented contigs from all the libraries were then combined into a single file and contigs <80 bp were again removed. All contigs were reassembled with k-mer sizes of 34, 44, 54, and 64 bases. To reduce redundancy, contigs of ≥200 bases and k-mer coverage ≥500× were kept and reassembled together (70% overlap, 98% identity) using CLC Genomics Workbench. Contigs of <200 bases but with k-mer coverage ≥1,000× were also kept for subsequent analysis of repetitive elements.

Transposable elements were annotated using TBlastX against the Repbase database (Jurka et al. 2005). Then, BlastX was used against 1) Swissprot (http://www.uniprot.org, last accessed April 11, 2012), 2) Ensembl database for *Danio rerio* (Danio Zv9), 3) the combined Ensembl databases for *Gasterosteus* (BROADS1), *Oryzias* (HdrR), *Takifugu* (FUGU 4.0), and *Tetraodon* (TETRAODON 8.0) (http://www.ensembl.org, last accessed April 11, 2012), and finally 4) nr (http://www.ncbi.nlm.nih.gov, updated May 23, 2012). The best Blast result with a minimal *e*-value = 10^{-6} was kept for each contig (fig. 1). Finally, the function *getorf* from the EMBOSS package was used in order to detect ORFs (Rice et al. 2000). The longest ORF of ≥300 bp for each contig was kept for subsequent analysis.

Read Mapping and Differential Expression Analysis

Burrows-Wheeler Aligner (BWA, v0.6.2-r126) was used to map reads back to assembled contigs (Li and Durbin 2009). Default parameters were used with the following exceptions: -n 3 (maximum edit distance), -e 3 (maximum number of gap extensions), -l 16 (seed size), -R 30 (number of equally best hits to proceed to suboptimal alignment).

For differential expression analyses, edgeR (Robinson et al. 2010), an R Bioconductor package (http://www.bioconductor.org/, last accessed February 22, 2014), was used. Library sizes were normalized and only contigs for which we had ≥1 count per million reads in at least three samples were kept, leaving 75,431 contigs to be analyzed for differential expression out of 79,047 that were assembled (95.4%, 74,273 contigs ≥200 bp and 1,158 contigs <200 bp repetitive elements). This is because weakly expressed genes are more susceptible to be called as differentially expressed (Nagalakshmi et al. 2008). All groups were compared with each other in a pairwise manner, and differentially expressed contigs were determined with a strict FDR <0.01 and FC >2. GO enrichment was performed using Blast2GO with the default parameters (Conesa et al. 2005). In order to ascertain computed FC by edgeR, we compared them with the results of DEGseq (Wang et al. 2010). Fragments per thousand base pairs values were used to calculate FC with DEGseq using the samwrapper function (Tusher et al. 2001). Pearson correlation coefficient was >0.95 between FC computed with edgeR and DEGseq for all comparisons (data not shown).

Supplementary Material

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

Acknowledgments

The authors wish to acknowledge the contribution of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for RNA-sequencing. The authors thank Jean-Christophe Therrien and Serge Higgins (LARSA) for their technical support during the crossing and maturation of embryos. The authors also thank Anne C. Dalziel, François-Olivier Hébert, Scott A. Pavey, and Christopher Sauvage for inspiring discussions and comments. The manuscript was also improved by constructive comments of three anonymous referees. This work was supported by a Natural Science and Engineering Research Council of Canada (NSERC) discovery grant and Canadian Research Chair in Genomics and Conservation of Aquatic Resources to L.B., a NSERC postgraduate scholarship to A.-M.D.-C., and a NSERC postdoctoral fellowship to S.R. A.-M.D.-C. also received a stipend from Québec-Océan. This project is a contribution to the research program of Québec-Océan.

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and speciation. *J Evol Biol.* 26:229–246.

Barbash DA, Siino DF, Tarone AM, Roote J. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc Natl Acad Sci U S A.* 100:5302–5307.

Belgnaoui SM, Gosden RG, Semmes OJ, Haoudi A. 2006. Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int.* 6:13.

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

Booke HE. 1968. Cytotaxonomic studies of coregonine fishes of the Great Lakes, USA: DNA and karyotype analysis. *J Fish Res Board Can.* 25:1667–1668.

Bougas B, Granier S, Audet C, Bernatchez L. 2010. The transcriptional landscape of cross-specific hybrids and its possible link with growth in brook charr (*Salvelinus fontinalis* Mitchill). *Genetics* 186:97–107.

Christie P, Macnair MR. 1984. Complementary lethal factors in two North American populations of the yellow monkey flower. *J Hered.* 75:510–511.

Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, Ponting CP, Stadler PF, Morris KV, Morillon A, et al. 2011. The reality of pervasive transcription. *PLoS Biol.* 9:e1000625.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.

Coyne J, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates, Inc.

Crête-Lafrenière A, Weir LK, Bernatchez L. 2012. Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLoS ONE* 7:e46662.

Derome N, Bougas B, Rogers SM, Whiteley AR, Labbe 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.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. *Nature* 489:101–108.

Dobzhansky T. 1937. Genetics and the origin of species. New York: Columbia University Press.

Dobzhansky T. 1940. Speciation as a stage in evolutionary divergence. *Am Nat.* 312–321.

Evans ML, Chapman LJ, Mitrofanov I, Bernatchez L. 2013. Variable extent of parallelism in respiratory, circulatory, and neurological traits across lake whitefish species pairs. *Ecol Evol.* 3:546–557.

Farrell AP, Jones DR. 1992. The heart. In: S HW, J RD, Farrell AP, editors. Fish physiology. Vol. XIIA. San Diego (CA): Academic Press. p. 1–88.

Gagnaire P-A, Normandeau E, Pavey SA, Bernatchez L. 2013. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). *Mol Ecol.* 22:3036–3048.

Hardie DC, Hebert PDN. 2003. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* 46:683–706.

Hebert FO, Renaut S, Bernatchez L. 2013. Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric lake whitefish species pair (*Coregonus clupeaformis*). *Mol Ecol.* 22:4896–4914.

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

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

Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L. 2010. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol Ecol.* 19:5389–5403.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 13:484–492.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110:462–467.

Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet.* 9: 411–412.

Kelleher ES, Edelman NB, Barbash DA. 2012. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol.* 10: e1001428.

Kerr MK, Churchill GA. 2001. Statistical design and the analysis of gene expression microarray data. *Genet Res.* 77:123–128.

Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD, Weng Z, Theurkauf WE. 2011. Adaptation to *P* element transposon invasion in *Drosophila melanogaster*. *Cell* 147: 1551–1563.

Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 86:813–833.

Kogenaru S, Yan Q, Guo Y, Wang N. 2012. RNA-seq and microarray complement each other in transcriptome profiling. *BMC Genomics* 13:629.

Labrador M, Farré M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of *Osvaldo*. *Mol Biol Evol.* 16: 931–937.

Landry CR, Hartl DL, Ranz JM. 2007. Genome clashes in hybrids: insights from gene expression. *Heredity* 99:483–493.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Lönnig W-E, Saedler H. 2002. Chromosome rearrangements and transposable elements. *Annu Rev Genet.* 36:389–410.

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

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates Inc.

Maheshwari S, Barbash DA. 2011. The genetics of hybrid incompatibilities. *Annu Rev Genet.* 45:331–355.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18:1509–1517.

Marie Curie SPECIATION Network. 2012. What do we need to know about speciation? *Trends Ecol Evol.* 27:27–39.

Mathavan S, Lee SGP, Mak A, Miller LD, Murthy KR, Govindarajan KR, Tong Y, Wu YL, Lam SH, Yang H, et al. 2005. Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet.* 1:e29.

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226:792–801.

Michalak P. 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity* 102:45–50.

Muller HJ. 1939. Reversibility in evolution considered from the standpoint of genetics. *Biol Rev.* 14:261–280.

Muller HJ. 1942. Isolating mechanisms, evolution and temperature. *Biol Symp.* 6:71–125.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–1349.

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014

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. 2007. Divergent host plant adaptation and reproductive isolation between ecotypes of *Timema cristinae* walking sticks. *Am Nat.* 169:151–162.
Nosil P. 2012. Ecological speciation. Oxford: Oxford University Press.
Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends Ecol Evol.* 24:145–156.
Nosil P, Schluter D. 2011. The genes underlying the process of speciation. *Trends Ecol Evol.* 26:160–167.
O'Neill MJ, Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393:68–72.
Pigeon D, Chouinard A, Bernatchez L. 1997. Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution* 51: 196–205.
Presgraves DC, Balagopalan L, Abmayr SM, Orr HA. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* 423:715–719.
Quinn NL, Levenkova N, Chow W, Bouffard P, Boroevich KA, Knight JR, Jarvie TP, Lubieniecki KP, Desany BA, Koop BF, et al. 2008. Assessing the feasibility of GS FLX Pyrosequencing for sequencing the Atlantic salmon genome. *BMC Genomics* 9:404.
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 A, Bernatchez L. 2009. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. salmonidae). *Mol Biol Evol.* 26:925–936.
Renaut S, Nolte AW, Bernatchez L. 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol Ecol.* 19(Suppl 1), 115–131.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16:276–277.
Rieseberg LH, Vanfossen C, Desrochers AM. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375:313–316.
Robinson MD, Mccarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
Rogers S, 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.
Rogers SM, Bernatchez L. 2006. The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). *J Evol Biol.* 19:1979–1994.
Rogers 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.
Rubin GM, Kidwell MG, Bingham PM. 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* 29:987–994.
Rundle HD, Nosil P. 2005. Ecological speciation. *Ecol Lett.* 8:336–352.
Saetre G-P, Saether SA. 2010. Ecology and genetics of speciation in *Ficedula* flycatchers. *Mol Ecol.* 19:1091–1106.
Schartl A, Dimitrijevic N, Schartl M. 1994. Evolutionary origin and molecular biology of the melanoma-inducing oncogene of *Xiphophorus*. *Pigment Cell Res.* 7:428–432.
Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323:737–741.
Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19:1117–1123.
Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 8:272–285.
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.
Symonová R, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Rábová M, Ráb P. 2013. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol Biol.* 13:42.
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 Aquat Sci.* 58: 394–405.
Tusher VG, Tibshirani R, Chu G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A.* 98:5116–5121.
Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr Biol.* 16:R872–R873.
Via S, West J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol.* 17:4334–4345.
Wang L, Feng Z, Wang X, Wang X, Zhang X. 2010. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26:136–138.
White M. 1978. Modes of speciation. San Francisco (CA): W. H. Freeman and Company.
Whitelaw E, Martin DI. 2001. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat Genet.* 27:361–365.
Whiteley AR, Derome N, Rogers SM, St-Cyr J, Laroche J, Labbe 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.
Whiteley AR, Persaud KN, 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.
Wittbrodt J, Adam D, Malitschek B, Mäueler W, Raulf F, Telling A, Robertson SM, Schartl M. 1989. Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in *Xiphophorus*. *Nature* 341:415–421.
Wright KM, Lloyd D, Lowry DB, Macnair MR, Willis JH. 2013. Indirect evolution of hybrid lethality due to linkage with selected locus in *Mimulus guttatus*. *PLoS Biol.* 11:e1001497.

Downloaded from http://mbe.oxfordjournals.org/ at Universite Laval on July 22, 2014