MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 3036–3048 doi: 10.1111/mec.12127

Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish *(Coregonus clupeaformis)*

PIERRE-ALEXANDRE GAGNAIRE, ERIC NORMANDEAU, SCOTT A. PAVEY and LOUIS BERNATCHEZ

Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Biologie, Université Laval, Pavillon Charles-Eugène-Marchand, Québec G1V 0A6, Canada

Abstract

The evolution of reproductive isolation in an ecological context may involve multiple facets of species divergence on which divergent selection may operate. These include variation in quantitative phenotypic traits, regulation of gene expression, and differential transmission of particular allelic combinations. Thus, an integrative approach to the speciation process involves identifying the genetic basis of these traits, in order to understand how they are affected by divergent selection in nature and how they ultimately contribute to reproductive isolation. In the Lake Whitefish (*Coregonus clupeaformis*), *dwarf* and *normal* species pairs sympatrically occur in several North American postglacial lakes. The limnetic *dwarf* whitefish distinguishes from its *normal* benthic relative by numerous life history, behavioural, morphological and gene expression traits, in relation with the exploitation of distinct ecological niches. Here, we have applied the RAD-Sequencing method to a hybrid backcross family to reconstruct a high-density genetic linkage map and perform QTL mapping in the Lake Whitefish. The 3061 cM map encompassed 3438 segregating RAD markers distributed over 40 linkage groups, for an average resolution of 0.89 cM. We mapped phenotypic and expression QTL underlying ecologically important traits as well as transmission ratio distortion QTL, and identified genomic regions harbouring clusters of such QTL. A narrow genomic region strongly associated with sex determination was also evidenced. Positional and functional information revealed in this study will be useful in ongoing population genomic studies to illuminate our understanding of the genomic architecture of reproductive isolation between whitefish species pairs.

Keywords: eQTL, NGS, pQTL, RAD sequencing, sex determination loci, speciation, TRD QTL, whitefish

Received 2 July 2012; revision received 29 September 2012; accepted 6 October 2012

Introduction

Understanding the genetic basis of phenotypic adaptation and its relationships with the genomic architecture of species divergence is a central question in evolutionary biology (Rice *et al.* 2011). This issue has received much attention in the context of ecological speciation, *that is*, when reproductive isolation evolves as a consequence of divergent selection on traits underlying adaptation to different environments (Schluter 2001). Phenotypic traits relevant to ecological divergence have been mapped to regions of the genome that quantitatively contribute to phenotypic variation in natural populations (Colosimo *et al.* 2004; Steiner *et al.* 2007; Stinchcombe & Hoekstra 2007; Ellegren & Sheldon 2008; Rice *et al.* 2011), enabling the authors to determine the number and location of quantitative trait loci (QTL) involved in phenotypic adaptation (but see Rockman 2012 for a critique of this approach). Besides the primary goal of linking phenotype

Correspondence: Pierre-Alexandre Gagnaire, Fax: +1 418 656 7176; E-mail: pierre-alexandre.gagnaire.1@ulaval.ca

to genotype, genetic-mapping approaches also enabled researchers to estimate the size effect of QTL and assess possible genetic correlations between traits (Slate *et al.* 2010). However, the existence of a causal link between phenotypic adaptation and the evolution of reproductive isolation often remains difficult to establish (Rice *et al.* 2011), as different components of reproductive isolation can contribute to the genetic barrier between incipient species (Coyne & Orr 2004; Barton & de Cara 2009). Moreover, adaptive phenotypic traits do not necessarily evolve under positive Darwinian selection when the ancestral capacity to express alternative phenotypes is lost following specialization to a single environment (Hughes 2012). Therefore, there is a need to determine whether polymorphisms underlying ecologically important traits are accompanied by signatures of divergent selection.

In recent years, genetic-mapping studies aimed at identifying QTL for fitness-related traits have been combined with population-genomic approaches looking for genetic markers with extreme (and presumably nonneutral) levels of differentiation in natural populations (Stinchcombe & Hoekstra 2007). These studies have revealed that some outlier loci detected in genome scans cluster in genomic regions close to QTL for traits showing adaptive divergence, thus establishing a link between phenotype and genetic markers harbouring signatures of divergent selection (e.g. Rogers & Bernatchez 2007; Via & West 2008). Although these analyses lacked the power to fully resolve genotype/phenotype correlations, they opened new avenues for measuring the effects of local gene flow reduction in the chromosomal regions surrounding QTL using high-density genetic mapping. As additional phenotypes of ecological relevance may be revealed by differential gene expression patterns (Pavey *et al.* 2010), expression quantitative traits loci (eQTL) have been mapped in addition to phenotypic QTL (pQTL) to elucidate the genetic architecture of phenotypic adaptation through gene regulation (Derome *et al.* 2008; Whiteley *et al.* 2008). However, neither individual eQTL nor eQTL hotspots appeared to be associated with the genomic islands of divergence revealed by genome scans in a study that combined genetic-mapping and population-genomic approaches (Renaut *et al.* 2012). There are credible biological reasons to expect this result such as the near neutrality of individual gene expression phenotypes due to their diluted effect on the complex external phenotypes under selection. However, the power of mapping studies to capture such associations inevitably depends on the precision of mapping QTL and genomic islands of divergence on the same linkage map. Therefore, the resolution of genetic-mapping approaches is a critical factor to pinpoint the genomic architecture of species divergence and its relationships with phenotypic adaptation.

Recent advances in next-generation sequencing (NGS) technologies have dramatically improved the process of constructing dense genetic maps in nonmodel organisms. Sequencing of reduced representation libraries using NGS simultaneously allows high-density SNP discovery and genotyping of population samples without intensive marker development (Van Tassell *et al.* 2008). Among the existing methods, restriction-associated DNA sequencing (RAD-Seq) relies on sequencing the regions flanking restriction sites following whole-genome digestion with a specific enzyme (Baird *et al.* 2008; Davey *et al.* 2011). This technique has been successfully applied for genetic-mapping purposes (Baird *et al.* 2008; Amores *et al.* 2011; Baxter *et al.* 2011; Chutimanitsakun *et al.* 2011; Pfender *et al.* 2011), enabling *de novo* construction of genetic linkage maps with thousands of SNPs in nonmodel organisms. Population-genomic studies have also used RAD markers to map regions of genetic differentiation directly to a reference genome (Hohenlohe *et al.* 2010; Andersen *et al.* 2012; Roesti *et al.* 2012), but there has been no attempt to use RAD sequencing to combine genetic mapping and F_{ST} scan approaches in a nonmodel species. Here, we present the first steps of an integrative research program aimed at combining linkage mapping, QTL mapping, gene expression, genome scan and candidate gene–targeted resequencing to investigate the genomic architecture of species divergence in the Lake Whitefish, *Coregonus clupeaformis*.

Lake Whitefish species pairs from North American postglacial lakes consist of two sympatric ecotypes whose evolutionary history includes a period of allopatric divergence in separate glacial refugia (Acadian and Atlantic–Mississippian), followed by a postglacial sympatric episode starting less than 12 000 years ago (Bernatchez *et al.* 2010). The present genomic architecture of reproductive isolation likely results from the accumulation of incompatibilities during geographical isolation and subsequent genetic introgression combined with ecological divergence in sympatry (Lu & Bernatchez 1998; Rogers & Bernatchez 2006; Renaut *et al.* 2009). *Dwarf* and *normal* whitefish differ in a number of morphological, behavioural, physiological and life history traits. The *normal* benthic whitefish is characterized by rapid growth rate and older age at maturation, while the limnetic *dwarf* whitefish which matures faster shows a smaller size at maturity and a higher metabolic rate partly associated with a more active swimming behaviour (Bernatchez *et al.* 2010). *Dwarf* and *normal* whitefish are also distinguished by their number of gill rakers, with *dwarf* fish typically showing a denser gill raker apparatus ensuring a high-feeding efficiency on planktonic prey. These divergent traits are believed to reflect phenotypic

adaptations to the limnetic and benthic lacustrine environments, acting as components of reproductive isolation between *dwarf* and *normal* whitefish (Rogers *et al.* 2002; Rogers & Bernatchez 2007). The molecular basis of adaptive divergence has also been investigated using gene expression profiles in controlled and natural conditions, revealing hundreds of differentially expressed genes between *dwarf* and *normal* whitefish (Derome *et al.* 2008; St-Cyr *et al.* 2008; Whiteley *et al.* 2008). Upregulated genes in *dwarf* whitefish were preferentially associated with survival functions whereas those overexpressed in *normal* whitefish were generally involved in growth functions (St-Cyr *et al.* 2008). Low-resolution QTL mapping combining phenotypic and expression traits have revealed additional key features of adaptive divergence. This includes the identification of candidate genes underlying important phenotypic traits through the detection of eQTL overlapped with pQTL and the localization of putative master regulatory regions through the detection of eQTL hotspots (Bernatchez *et al.* 2010). Finally, QTL mapping was combined with genome scans of independently evolved whitefish species pairs to demonstrate the role of natural selection in driving adaptive divergence and reproductive isolation between *dwarf* and *normal* ecotypes (Rogers & Bernatchez 2007). Loci showing extreme levels of differentiation in natural populations were found more often than expected by chance in the genomic regions surrounding QTL, and some of these 'QTL outlier loci' exhibited segregation distortion in the mapping families, in addition to distorted markers in other chromosomal regions (Rogers & Bernatchez 2006, 2007). Altogether, these integrated analyses conducted in Lake Whitefish species pairs emphasized the importance of using genetic linkage information to disentangle the genetic architecture of adaptive divergence and reproductive isolation.

Previous mapping studies in whitefish relied on a low-density AFLP linkage map (Rogers *et al.* 2007). As the degree of precision of genetic-mapping approaches increases with growing marker density, we aimed at building a new high-density genetic map using RAD markers in order to improve QTL detection and mapping using already-published phenotypes and expression data. The enhanced resolution of the new RAD linkage map was supported by the identification of multiple QTL underlying ecologically important traits and a narrow genomic region involved in sex determination.

Materials and methods

Experimental genetic crosses

We performed genetic-mapping experiments with a backcross family derived from the hybrid × *dwarf* cross already used for AFLP linkage map construction (Rogers *et al.* 2007) and QTL mapping (Rogers & Bernatchez 2007). More precisely, parents were sampled from two allopatric glacial races (*dwarfs* were of Acadian glacial origin and *normals* of Atlantic–Mississippian glacial origin) and used to produce two first generation groups: a hybrid cross (♀ *normal* × ♂ *dwarf*) and a pure *dwarf* cross (♀ *dwarf* × ♂ *dwarf*) (Lu & Bernatchez 1998). A F_1 female hybrid was then backcrossed to a male *dwarf* (♀ hybrid × ♂ *dwarf*) to produce a progeny of 198 individuals, of which 102 individuals were available for the RAD linkage map construction.

RAD-Seq genotyping by sequencing

Genomic DNA was isolated from the adipose fin of the female hybrid and male *dwarf* parents as well as 102 backcross progeny using the standard phenol–chloroform protocol and was later digested with the restriction endonuclease *Sbf*I-HF (New England Biolabs), which recognizes an 8-bp sequence (CCTGCAGG). Individual RAD-tag libraries, each labelled with a unique 6-bp barcode, were constructed by Floragenex Inc. (Eugene, OR) according to the protocol described in (Baird *et al.* 2008). Libraries were pooled in equimolar proportions by groups of 13 individuals per lane and sequenced on the Illumina GAIIx (80-bp single reads) or HiSeq2000 (101-bp single reads) plateform at the University of Oregon Genomic Core Facility. All sequence reads were deposed in the NCBI Short Read Archive SRA059085.

We used the *Stacks* software system (Catchen *et al.* 2011) to group homologous Illumina short reads into loci and determine individual genotypes. Low-quality reads as well as reads presenting ambiguous barcodes were first removed from the raw data, and all 101-pb reads were trimmed to 80 bp by removing the last 21 bases. Putative haplotypes were built *de novo* for each individual using a *minimum stack depth* parameter of three identical reads. Highly repetitive RAD-tags were discarded at this step as they probably arose from the sequencing of repeated elements. A maximum *within-individual distance* of two nucleotides between stacks was used to allow the merging of putative alleles into a locus for each individual, in order to avoid merging stacks from paralogous loci. Identical loci from the two parents were then merged to build a catalogue, against which progeny individuals' loci were matched. Genotypes were called within a likelihood framework developed to differentiate real SNPs from sequencing errors (Hohenlohe *et al.* 2010). Markers suitable for linkage mapping were of either four types (Amores *et al.* 2011): (i) homozygous in the male *dwarf* and heterozygous in the hybrid female (type *nn*×*np*, segregating 1:1), (ii)

heterozygous in the male *dwarf* and homozygous in the hybrid female (type *lm*×*ll*, segregating 1:1), (iii) heterozygous in both parents with two alleles in common (type *hk*×*hk*, segregating 1:2:1) or (iv) heterozygous in both parents with only one allele in common (type *ef*×*eg*, segregating 1:1:1:1). Individual genotypes were finally exported in *JoinMap 4* format with population type CP (van Ooijen 2006), designed for crosses between heterogeneously heterozygous and homozygous diploid parents.

Genetic map construction

The Lake Whitefish RAD linkage map was constructed with the *JoinMap 4* program (van Ooijen 2006), using markers with a maximum of 51 missing genotypes out of 102 progeny. We allowed segregation ratios to deviate from Mendelian proportions with a maximal Chi-square value of 40. Below this threshold, the Bonferroni corrected *P*-value of the Chi-square statistic remained always higher than 0.05, indicating no significant segregation distortion. Retained markers were first assigned to linkage groups using a LOD threshold of 7, and remaining ungrouped loci were subsequently assigned to established linkage groups using the strongest cross-link information in *JoinMap*, with a LOD threshold of 5. Markers were then ordered using the regression-mapping algorithm, and map distances in centimorgans (cM) were calculated from recombination frequencies using Kosambi's mapping function. To evaluate the degree of conserved synteny with a fully sequenced fish genome and indirectly estimate the accuracy of the newly constructed genetic map, we mapped whitefish coding RAD markers to the zebra fish genome using a BLASTX approach.

Phenotypic traits and quantitative expression data

We used measures of nine ecologically important phenotypic traits for pQTL detection: four behavioural traits (including depth selection, directional changes, burst swimming and activity), one physiological trait (growth rate), two morphological traits (weight/length^3 as condition factor and number of gill rakers) and two life history traits (onset of maturity and the gonadosomatic index) (Rogers & Bernatchez 2007). The gender of each sexually mature individual was also used to search genomic regions associated with sex determination.

Quantitative expression data from 2255 transcripts expressed in muscle of 66 progeny (Derome *et al.* 2008) and 3563 transcripts expressed in brain tissue of 57 progeny (Whiteley *et al.* 2008) were used for the detection of eQTL. More specifically, we focused on transcripts that showed different expression profiles between *dwarf* and *normal* whitefish in controlled conditions, including 253 genes differentially expressed in muscle (Derome *et al.* 2008) and 201 in brain (Whiteley *et al.* 2008) and 92 genes showing parallel patterns of transcription in liver tissue in two natural lakes: Cliff lake and Indian Pond (St-Cyr *et al.* 2008).

Phenotypic, expression and transmission ratio distortion QTL mapping

Genotype proportions at 2956 mapped markers of segregation types *nn*×*np* and *lm*×*ll* were used to evaluate the extent and direction of departure from Mendelian inheritance proportions (1:1) using a standard χ^2-test with 1 d.f. We applied a standard Bonferroni correction to account for multiple testing ($\alpha = 0.05/2956 \approx 1.7 \times 10^{-5}$).

The interval-mapping algorithm implemented in the *MapQTL 5* software (van Ooijen 2004) was used for pQTL analysis using 1 cM increments. For each of the nine phenotypic traits, a significance LOD threshold was determined for each linkage group using 1000 permutations to test for segregating pQTL with a 5% significance level. Markers with LOD scores exceeding the empirical threshold were first selected as cofactors before automatic cofactor selection was performed for each linkage group with a pQTL. A multiple QTL-mapping analysis was finally performed using selected cofactors, and pQTL were identified when LOD peaks were higher than the 5% chromosome-wide LOD threshold determined by permutations.

We used quantitative expression data as phenotypes along with the RAD linkage map to perform a genome-wide detection of eQTL with the interval-mapping algorithm. LOD scores were calculated every cM, but we did not perform permutation tests for each gene to save computation time. Instead, we used a conservative LOD threshold of five to declare significance of eQTL. Putative QTL hotspots were defined as genome regions where four or more eQTL colocalize (see Results).

Finally, the mapping family contained 37 identified female and 46 male progeny (remaining fish were not mature enough to determine sex), enabling the search for genomic regions associated with sex, which was coded as a binary trait. Sex determination loci were mapped using the SNPassoc R package (González *et al.* 2007), considering four alternative genetic models: codominant, dominant, recessive and log-additive.

Results

The whitefish RAD linkage map

A total of 7.5 million Illumina sequence reads were obtained for the F_1 hybrid female parent and 7.1 million

for the male *dwarf* parent, which both received twice as much sequencing effort compared to progeny. On average, each of the 102 progeny's library provided 3.1 million sequence reads (Table S1, Supporting information). *Stacks* reconstructed about 20 000 loci, from which we detected 4887 mappable markers. A total of 136 markers exhibiting significant segregation distortion were excluded from subsequent analysis. *JoinMap* assigned 3438 of the remaining markers to 40 different linkage groups (Table 1, Table S2, Figure S1, Supporting information), corresponding to the haploid number of chromosomes in *Coregonus clupeaformis* ($2n = 80$, Phillips *et al*. 1996). The length of linkage groups ranged from 49.4 to 134.3 cM, for a total map length of 3061 cM and an average resolution of 0.89 cM. Among the 3438 mapped markers, 2309 were of segregation type $nn \times np$, 647 of type $lm \times ll$, 439 of type $hk \times hk$ and 43 of type $ef \times eg$, thus reflecting a 2.5-fold higher heterozygosity of

Table 1 Summary of the Lake Whitefish RAD linkage map. Black circles indicate the occurrence of significant pQTL, eQTL hotspots and sex determination QTL on each linkage group. Transmission distortion ratio QTL in favour of the hybrid parent genotype are designated by H and those biased towards the *dwarf* parent by D. Several linkage groups showed TRD in both directions (H/D)

Linkage group	Linkage group size (cM)	Number of RAD markers	RAD markers density per cM	pQTL ($\alpha = 0.05$)	eQTL hotspot	TRD QTL (hybrid/*dwarf*)	Sex determination QTL
1	76.0	91	1.20	—	•	H	—
2	74.2	75	1.01	—	•	—	—
3	86.5	60	0.69	•	—	—	—
4	134.3	199	1.48	•	•	H/D	—
5	94.7	193	2.04	—	•	H	—
6	88.2	155	1.76	—	—	H/D	—
7	57.6	47	0.82	—	•	H/D	—
8	78.8	97	1.23	•	—	D	—
9	106.8	55	0.51	—	—	D	—
10	78.8	108	1.37	•	•	H	—
11	77.2	98	1.27	—	—	D	—
12	90.8	95	1.05	—	•	D	—
13	87.6	94	1.07	—	—	D	—
14	81.0	77	0.95	—	—	—	—
15	70.5	69	0.98	—	—	—	—
16	96.4	63	0.65	—	•	—	—
17	63.2	61	0.97	—	—	—	—
18	66.5	58	0.87	•	•	—	—
19	81.1	37	0.46	—	•	—	—
20	76.8	26	0.34	—	•	D	—
21	88.2	109	1.24	—	—	D	—
22	66.2	84	1.27	—	—	H/D	—
23	57.7	38	0.66	•	—	—	—
24	86.9	111	1.28	—	—	—	—
25	62.2	64	1.03	—	—	—	•
26	81.0	92	1.14	—	—	D	—
27	64.6	74	1.15	—	—	—	—
28	84.3	78	0.93	•	—	H	—
29	90.2	75	0.83	—	—	—	—
30	63.7	117	1.84	—	—	D	—
31	91.6	101	1.10	•	—	—	—
32	69.2	96	1.39	•	•	—	—
33	70.5	96	1.36	—	•	D	—
34	73.6	92	1.25	—	—	—	—
35	50.9	82	1.61	•	—	—	—
36	49.5	88	1.78	—	•	H	—
37	75.3	83	1.10	—	—	D	—
38	58.4	83	1.42	•	—	—	—
39	52.1	64	1.23	•	—	—	—
40	58.3	53	0.91	—	—	—	—
Whole map	3061	3438	1.12				

the hybrid female parent compared to the male *dwarf* parent used to produce the progeny.

BLASTX search against the zebra fish reference genome revealed the existence of homologous chromosome regions arranged in duplicated syntenic blocks (Figure S2, Supporting information), which probably arose as a result of whole-genome duplication in the ancestral salmonid (Allendorf & Thorgaard 1984; Lien *et al.* 2011). The top hits of BLASTX searches against SwissProt and GenBank (nr) databases were used to annotate 248 RAD markers (Table S2, Supporting information).

Detection of pQTL

We detected at least one QTL for each phenotypic trait with the exception of the condition factor and the male gonadosomatic index (Table 2). A total of 19 QTL were

Table 2 Summary of pQTL detected for nine quantitative traits pertaining to behavioural, physiological, morphological and life history categories. For each QTL, *N* denotes the number of individuals with available quantitative data and LG refers to the linkage group within which the QTL was detected. QTL interval and position are given in centimorgan (cM). LOD significance level was determined by 1000 permutations. The 1.5 LOD unit of support interval is the interval in which the LOD score is within 1.5 units of its maximum

Trait	*N*	LG	Position (cM)	Nearest RAD locus	LOD	1.5 LOD unit of support (range in cM)
Behavioural						
Burst swimming	51	18	38.618	*5372*	3.95**	38.505–39.893
Burst swimming	51	6	20.071	*39521*	3.70†	20.071
Directional change	51	10	42.828	*104593*	5.43***	41.644–43.239
Directional change	51	10	45.942	*28638*	5.04**	45.942–47.097
Directional change	51	10	54.212	*150119*	4.30**	52.749–55.559
Directional change	51	10	44.980	*128612*	4.03*	41.644–48.073
Directional change	51	10	38.398	*67789*	4.01*	38.137–39.923
Directional change	51	10	14.232	*41688*	3.89*	14.232–15.232
Directional change	51	32	38.082	*145008*	4.06**	31.012–42.853
Directional change	51	32	27.712	*43386*	3.37*	23.247–42.853
Directional change	51	18	26.895	*7283*	3.64*	23.693–29.016
Directional change	51	6	20.071	*39521*	3.60*	20.071–20.649
Directional change	51	30	6.628	*110457*	3.34*	3.628–9.628
Depth selection	58	28	66.719	*13173*	4.58***	65.216–73.719
Depth selection	58	8	14.786	*4647*	4.19**	9.457–17.487
Depth selection	58	39	13.537	*149216*	3.75***	13.081–14.441
Depth selection	58	21	4.494	*87077*	3.74*	0.000–13.988
Depth selection	58	38	53.339	*112713*	3.57**	48.745–58.389
Depth selection	58	23	21.616	*34216*	3.24*	5.381–24.361
Activity	56	18	17.951	*35016*	5.01***	17.095–18.951
Activity	56	18	5.620	*154979*	3.56**	3.000–18.951
Activity	56	15	10.712	*92525*	3.52*	9.712–12.719
Activity	56	28	65.216	*66560*	3.35*	64.044–83.719
Physiological						
Growth rate	77	35	22.916	*8663*	3.72***	22.916–23.600
Growth rate	77	35	20.885	*140903*	3.58***	17.857–21.885
Growth rate	77	23	3.620	*102256*	3.45**	3.620–6.381
Growth rate	77	32	31.889	*90115*	3.26*	30.012–32.092
Morphological						
Gill raker	79	4	106.545	*267*	4.08**	105.830–106.956
Gill raker	79	8	64.232	*38660*	3.54*	63.906–65.239
Gill raker	79	31	79.860	*101722*	3.25**	79.712–81.252
Life history						
Maturity	81	3	6.411	*87937*	8.6***	6.411–9.099
Maturity	81	3	11.024	*45997*	7.16***	6.411–11.247
Maturity	81	5	86.808	*87891*	7.55*	84.975–86.808
Maturity	81	16	88.198	*46930*	7.43*	80.199–93.198
Gonadosomatic index ♀	28	28	59.025	*73460*	7.26***	59.025
Gonadosomatic index ♀	28	17	3.567	*39874*	5.76*	1.567–15.150

†Chromosome-wide $\alpha = 0.10$, *$\alpha < 0.10$, **$\alpha < 0.05$, ***$\alpha < 0.01$.

found at the 5% significance level with one to four QTL detected for each individual trait. Despite the limited size of the mapping population with available quantitative data ($N = 83$), nine QTL were supported by strong statistical significance and narrow confidence intervals. Moreover, marginal evidence for 17 additional QTL was obtained at the 10% chromosome-wide level. A total of 11 behavioural QTL mapped to seven different chromosomes, including one QTL located on LG 18 for burst swimming, four QTL on LG 10 and 32 for directional change, four QTL on LG 8, 28, 38 and 39 for depth selection and two QTL on LG 18 for activity. Three QTL mapping to LG 23 and 35 were detected for growth rate, and the number of gill rakers had two QTL localized on LG 4 and 31. Two life history traits were mapped, with 2 QTL on LG 3 for maturity and one QTL on LG 28 for the female gonadosomatic index.

Considering marginally supported QTL, we observed four cases of overlapping QTL confidence intervals for different traits, including burst swimming and directional change on LG 6, depth selection and growth rate on LG 23, depth selection and activity on LG 28, and directional change and growth rate on LG 32 (Table 2, Fig. 1).

Detection of eQTL

Among the 51 progeny analysed with available quantitative expression data, we detected a total of 366 eQTL associated with 150 transcripts (Table S3, Supporting information). The number of eQTL per transcript ranged from one to 25 for an average value of 2.44. The distribution of eQTL was highly heterogeneous across the 40 linkage groups, with a minimum of three eQTL detected for LG 13 and LG 14 and a maximum number of 26 eQTL localized within LG 4 (Fig. 1). Assuming a Poisson distribution, the probability of observing at least four eQTL at a given position on the map was < 0.00001. Thus, we considered the occurrence of four or more eQTL at a given map position as an eQTL hotspot. A total of 26 of such eQTL hotspots were found distributed over 14 linkage groups, among which 15 major hotspots contained five eQTL or more. The chromosomal region with the most pervasive influence on transcript regulation was located on LG 10 between 71.8 and 73.8 cM, with up to nine eQTL colocalized. The chromosomal distribution of eQTL hotspots appeared biased towards the linkage groups' extremities. Cases of overlap between pQTL and eQTL were observed for 15 linkage groups (Fig. 1).

Transmission ratio distortion QTL

Significant segregation distortion was observed in the backcross mapping family. On a genome-wide scale, departures from expected genotype proportions under Mendelian inheritance were generally biased towards the genotype of the male *dwarf* parent (Fig. 2). We observed 17 significant transmission ratio distortion (TRD) QTL at which genotype frequencies were enriched in favour of the male *dwarf* parent (negative peaks) and 10 TRD QTL showing a directional bias in favour of the female hybrid parent genotype (positive peaks).

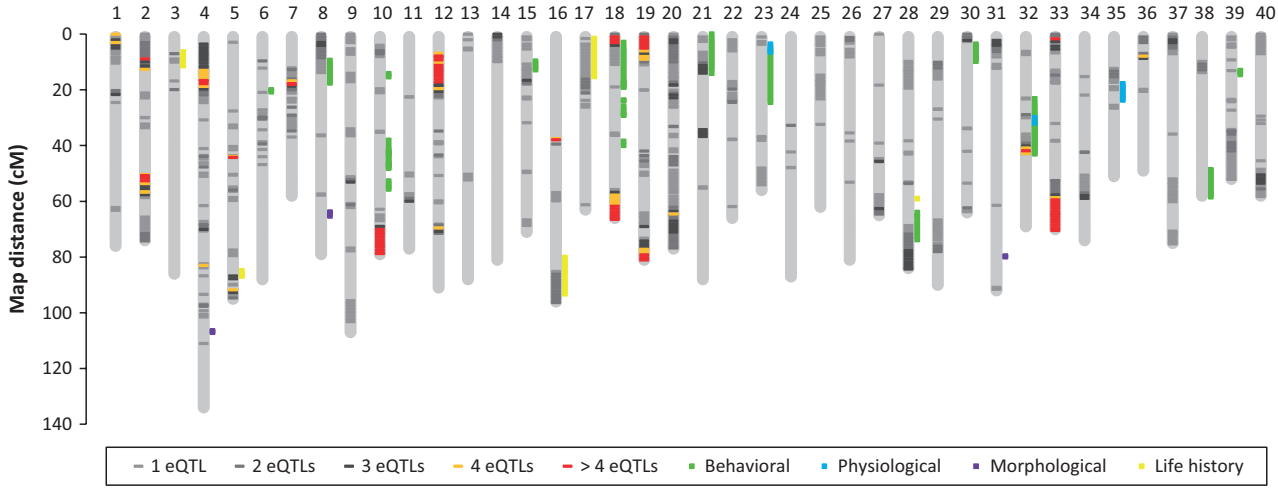

Fig. 1 Genome location of significant pQTL and eQTL. The position of significant eQTL is indicated within each linkage group, and confidence intervals of pQTL are illustrated on the right-hand side of linkage groups. The number of eQTL colocalizing at each genomic position and the category of each pQTL are indicated by the colour code defined in the inset. Hotspots of ≥ 4 eQTL colocalizing at the same genomic position appear in orange and red.

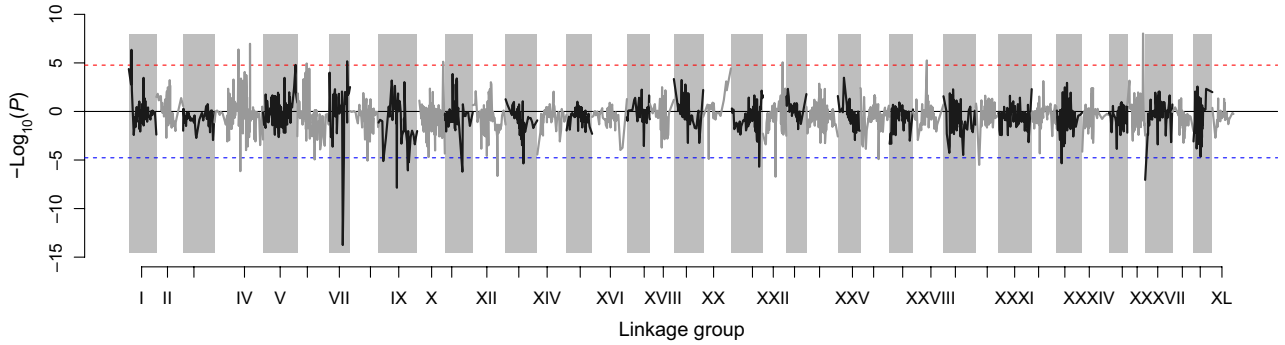

Fig. 2 Genome-wide detection of transmission ratio distortion QTL using 2956 mapped markers of segregation types *nn*×*np* and *lm*×*ll*. The significance level of the SNP-by-SNP χ^2 test with 1 d.f., evaluating the extent of deviation from the expected 1:1 ratio of genotype frequencies in the backcross family, is displayed on the *Y* axis. The values of $-\log_{10}$(*P*-value) were kept positive when distortion was in favour of the female hybrid parent and transformed into negative values when genotype frequencies were biased towards the genotype of the male *dwarf* parent. The 40 linkage groups are represented on the *X* axis according to their relative lengths. Horizontal dashed lines represent the Bonferroni-corrected significance threshold at the 5% level.

Close localization (<3 cM) between TRD QTL and eQTL hotspots was observed in three cases located within LG 5 (92 cM), LG 10 (72 cM) and LG 12 (69 cM) (Figs 1 and 2). The only pQTL that encompassed a TRD QTL within its confidence interval was associated with directional change (LG 30, 7 cM).

Genetic mapping of sex determination loci

Only one linkage group was significantly associated with sex after Bonferroni correction (Fig. 3). Five loci exceeding the significance threshold mapped to the interval between 0 and 25.2 cM in the proximal region of LG 25 under the dominant model, with the most strongly associated locus being located at 13.7 cM (locus *82008*, $P < 10^{-12}$). An additional significant peak was detected at position 38.9 cM of LG 25, and a minor peak that did not reach significance was found within linkage group 10 at position 23.8 cM. Genotypes at locus *82008* distributed as follows: 92.6% of the females were of genotype GG and 7.4% were GC, whereas 94.1% of the males were of genotype GC against 5.9% of genotype GG.

Discussion

We applied the RAD-Seq method in the Lake Whitefish to discover and genotype thousands of new SNP genetic markers in a backcross family (hybrid × *dwarf* cross) already used for linkage map construction (Rogers *et al.* 2007) and QTL mapping (Rogers & Bernatchez 2007) with AFLP markers. Our main objective was to construct a new high-density linkage map and identify QTL associated with various traits, including ecologically important phenotypic traits, gene expression data, segregation distortion and sex. Although some of the QTL detected here may be specific to the genomic backgrounds of the two parents used to

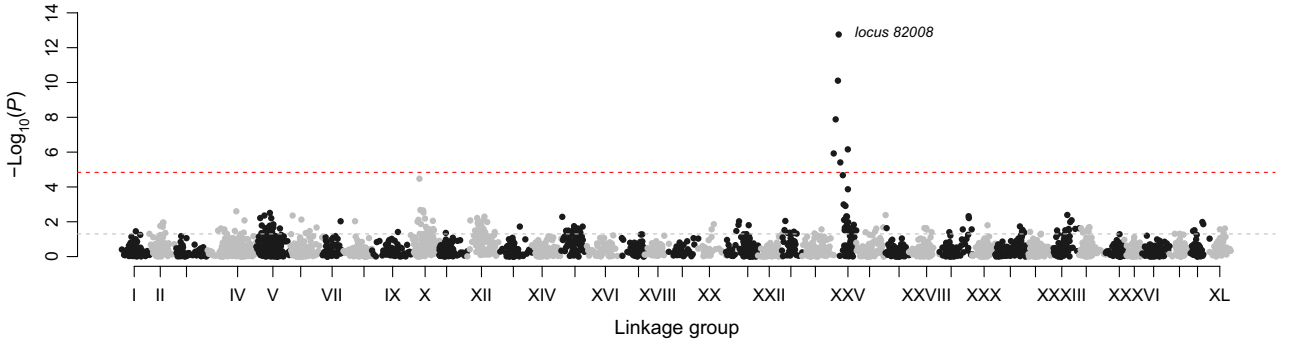

Fig. 3 Manhattan plot of the genome-wide association study of sex under the dominant model, for 3438 SNPs and 83 individuals. Genetic markers are represented on the *X* axis according to their chromosomal position on each of the 40 linkage groups. Significance of the SNP-by-SNP association test is displayed on the *Y* axis as $-\log_{10}$(*P*-value). The nominal and Bonferroni-corrected *P*-value threshold are respectively indicated by the lower and upper dashed lines. Only linkage group 25 was significantly associated with sex after Bonferroni correction. Five loci exceeding the significance threshold mapped to the interval between 0 and 25.2 cM in the proximal region of LG 25, with the most strongly associated locus (locus *82008*) being located at 13.7 cM.

produce the progeny, previous works based on the same mapping family support that the genetic basis of several traits is shared among different allopatric populations (Rogers & Bernatchez 2007). Therefore, the new whitefish RAD linkage map enriched with positional information for traits involved in ecological divergence will provide a useful genome-wide framework for improving our efforts towards elucidating the genetic basis of phenotypic adaptation and its relationships with the genomic architecture of reproductive isolation between *dwarf* and *normal* whitefish. This map will also provide an important tool for assembling the whitefish genome which is currently being sequenced.

The previous version of the whitefish linkage map included two sex-specific linkage maps assembled in each of two hybrid backcross families and comprising a total of 877 AFLP and 30 microsatellite markers (Rogers *et al.* 2007). Here, we focused on a single backcross family in which we reconstructed a single sex-averaged map, as sample availability limited our RAD-sequence effort. Also, our attempts to build a consensus map using both types of markers (dominant AFLP and codominant SNPs) were not successful due to the tendency of AFLP markers to cluster together in the combined linkage map. Thus, we could not directly compare our results with those based on the previous AFLP map, as we could not specifically address the differences in recombination rates between male and female maps, which have been commonly reported in several salmonid species, including whitefish (Gharbi *et al.* 2006; Rogers *et al.* 2007; Lien *et al.* 2011). However, our new sex-averaged RAD linkage map includes 3438 codominant markers distributed across the 40 chromosomes in *C. clupeaformis* (Phillips *et al.* 1996) and provides positional information for 248 annotated coding genes. Moreover, it has a 20 times enhanced average resolution of 0.89 cM compared to the average distance of 18 cM between loci in the previous AFLP linkage map that comprised only 34 linkage groups (Rogers *et al.* 2007). As such, this new RAD linkage map represents one of the most densely covered currently available for salmonid fishes (e.g. Lien *et al.* 2011; Hecht *et al.* 2012; Miller *et al.* 2012).

The recent whole-genome duplication experienced by the salmonid ancestor approximately 60 million years ago (Crête-Lafrenière *et al.* 2012) is a major challenge to developing genetic markers for genetic map reconstruction or population genomic surveys in salmonids (Seeb *et al.* 2011; Sauvage *et al.* 2012). In the present study, we attempted to adapt the parameters for read assembly and genotype inference in *Stacks* so as to minimize the inclusion of paralogous sequences in our SNP data set. Remaining artefactual polymorphisms related to the assembly of duplicated sequence reads were filtered out by excluding RAD markers with non-Mendelian segregation patterns in the mapping family. Thus, only high-quality genetic markers were used for map reconstruction ensuring the development of a reliable genetic map of the whitefish genome. This was confirmed by the comparison of synteny with the zebra fish genome using a BLASTX homology search of coding regions, which highlighted the whole-genome duplication pattern already detected in the Atlantic salmon (Lien *et al.* 2011).

Genetic mapping of pQTL and eQTL

Identifying the genomic regions involved in adaptive divergence remains a central objective of speciation genomic research (Rice *et al.* 2011). The development of high-density genetic-mapping approaches offers promising opportunities for detecting genomic regions underlying expressed phenotypes in nonmodel organisms for which adaptive traits can be measured, such as in the Lake Whitefish (Bernatchez *et al.* 2010). In complement to QTL-mapping approaches based on laboratory crosses, genome-wide association studies using unrelated individuals from natural populations now offer powerful alternatives to dissect the genetic basis of quantitative traits (e.g. Slate *et al.* 2010; Johnston *et al.* 2011). In the present study, the QTL approach allowed mapping eight phenotypic traits involved in the ecological speciation of *dwarf* and *normal* whitefish. Compared with the previous genetic mapping experiment based on the same family and the AFLP map (Rogers & Bernatchez 2007), the number of pQTL with nonoverlapping confidence intervals increased from 23 to 31, while the average length of confidence intervals decreased from 23 to 5.9 cM. Thus, the RAD linkage map provides an increased power and resolution for the detection of QTL. Although we could only use a limited number of backcross progeny from a single cross, several significant unlinked pQTL were detected for each trait, illustrating the polygenic nature of these phenotypes. Underlying genes of relatively medium to large effects may contribute together to a large fraction of the genetic variance associated with each trait and may therefore experience divergent selection in nature (Zhang & Hill 2005). However, because quantitative polygenic traits are often influenced by a high number of genes of small to moderate effects which are individually undetectable, we most likely detected only a small proportion of the QTL involved (Otto & Jones 2000; Rockman 2012). Recognizing the limited power of our study to detect small effect QTL, we also did not provide the phenotypic variance associated with significant QTL, as our data set could be subject to the Beavis effect which can be pronounced when less than 100 progeny are analysed (Xu 2003).

Simultaneous mapping of multiple phenotypes and expression QTL revealed interesting features of the genetic architecture of population divergence in the Lake Whitefish. As already observed in previous mapping studies (Derome *et al.* 2008; Whiteley *et al.* 2008; Bernatchez *et al.* 2010), the distribution of eQTL was not random across the genome, but was instead characterized by the colocalization of several eQTL within the same genomic regions. These eQTL hotspots may reflect the presence of putative master regulatory genes controlling the expression of several genes involved in the expression of *dwarf* and *normal* phenotypes (Gilad *et al.* 2008; Bernatchez *et al.* 2010). A new finding here was that eQTL hotpots mapped preferentially to the ends of linkage groups, suggesting that the genomic architecture of differential gene regulation between whitefish ecotypes could involve several subtelomeric regions. Concomitantly, most cases of overlap between eQTL and pQTL also occurred near the ends of linkage groups. Positional information for these remarkable features will be of prime importance to test whether genomic islands of differentiation between *normal* and *dwarf* whitefish colocalize with the genomic regions associated with divergent phenotypic and expression trait in the mapping family. Empirical evidence supports a causal link between phenotypic adaptation and local genetic barriers to gene flow based on AFLP markers (Rogers & Bernatchez 2007). However, associations between genetic divergence and quantitative trait loci were only significant for pQTL, but neither for individual eQTL nor for eQTL hotspots (Renaut *et al.* 2012). Because complex traits integrate multiple allelic effects, expression phenotypes of individual genes may be effectively neutral to natural selection. Moreover, substantial trait differentiation does not necessarily involve high frequency differences at the underlying loci due to the covariance of allelic effects, the relative contribution of which increases with the number of loci affecting the trait (McKay & Latta 2002; Le Corre & Kremer 2012). Finally, adaptive phenotypic traits may evolve without positive Darwinian selection when ancestral phenotypic plasticity is followed by specialization to a single environment and accompanied by the loss of capacity to express alternative phenotypes (Hughes 2012). Thus, upcoming genome scans in natural populations should aim to revisit the extent to which pQTL and eQTL are associated with signatures of divergent selection using the new genetic-mapping resources developed here.

TRD QTL may reveal genetic incompatibilities

Another major result from the previous genetic-mapping study in the Lake Whitefish was the large proportion of markers exhibiting significant segregation distortion (Rogers *et al.* 2007), consistent with the existence of genetic incompatibilities between *normal* and *dwarf* populations due to allopatric glacial separation preceding secondary contact (Lu & Bernatchez 1998; Rogers & Bernatchez 2006).

In the hybrid backcross family used for RAD-mapping analysis, we essentially focus on the segregation of *normal* whitefish alleles into the *dwarf* genomic background. The 2.5-fold higher heterozygosity of the F1 female parent compared to the male *dwarf* parent was an expected result, as the hybrid female was obtained from an interspecific cross between parents that were sampled from two allopatric whitefish populations belonging to distinct glacial races (Lu & Bernatchez 1998). On a genome-wide average, progeny exhibited a general bias towards the male *dwarf* genotype, consistent with the existence of distortion mechanisms directed against the introgression of *normal*-type alleles within the *dwarf* background. This was also illustrated by the asymmetrical distribution of significant TRD QTL, which were more frequently characterized by an enrichment of genotype frequencies in favour of the male *dwarf* parent. Although several mechanisms that influence the segregation of alleles may produce transmission ratio distortion (e.g. gamete competition, meiotic drive, male killing), the large proportion of markers exhibiting distorted ratio in the hybrid backcross family could be largely explained by negative selection acting against deleterious heterospecific allelic combinations in backcross individuals (Moyle & Graham 2006). Consistent with this hypothesis, chromosomal regions showing segregation distortion were previously found to be associated with differential embryonic mortality in hybrids during development time to emergence (Rogers & Bernatchez 2006). This, along with other results that showed strong patterns of gene misregulation (Renaut & Bernatchez 2011) and embryonic mortality (Lu & Bernatchez 1998) in hybrid crosses indicate that the 60 000 years of geographical isolation between the Acadian and Atlantic races (Jacobsen *et al.* 2012) were sufficient to cause genomic incompatibilities.

The existence of intrinsic postzygotic isolation mechanisms is relevant for interpreting the significance and nature of outlier regions found in genome scans (Bierne *et al.* 2011). Indeed, if TRD QTL are mainly attributable to endogenous incompatibilities between *normal* and *dwarf* whitefish, they probably cause local barriers to gene flow and consequently may translate into outliers in genome scans of natural populations. Moreover, because the number and selective effects of intrinsic incompatibilities also affect hybrid fitness in nature, such incompatibilities could contribute to the barrier strength and thus indirectly reduce the effective migration rate experienced by neutral loci (Barton &

Bengtsson 1986). For that reason, the existence of intrinsic postzygotic barriers coupled with ecological barriers to gene flow is expected to increase the size of the genomic islands of differentiation. Previous efforts to map regions of genomic divergence on AFLP genetic maps have highlighted the existence of relatively large genomic islands between *normal* and *dwarf* whitefish (Rogers & Bernatchez 2007; Renaut *et al.* 2012). However, the dominant nature of AFLP markers and the lower resolution of the previous linkage maps and genome scans have cast some doubts about conclusions regarding segregation distortion patterns and the size of genomic islands. Here, with a much higher-density genetic map, we confirmed several findings previously reported in mapping studies, including the large proportion of distorted markers (Rogers *et al.* 2007) and the colocalization between some TRD QTL and eQTL hotspots (Derome *et al.* 2008; Whiteley *et al.* 2008). Although we admit that the present study has its own limitations imposed by the relatively small number of progeny that were still available for genotyping, it offers new genetic-mapping resources to reassess the size of the genomic islands of differentiation with a significantly increased resolution.

Sex determination loci

Our results revealed the existence of a sex determination locus near position 13.7 cM on LG 25. According to the genotypes segregating at locus *82008* showing the strongest linkage association with sex, the nearby sex determination locus would be a male-dominant determinant in an XX/XY system similar to the one described in Medaka (*Oryzias latipes*), in which a sex-determining gene (*DMY/dmrt1bY*) has been identified (Matsuda *et al.* 2002; Nanda *et al.* 2002). In salmonids, the genetic mechanism of sex determination is known to rely on an XX/XY heterogametic system (Devlin & Nagahama 2002), with an additional potential influence of other determinants (such as genetic or environmental factors) on the sex phenotype. Genetic-mapping approaches of sex determination in several salmonid species have localized a sex determination locus, named *SEX*, at the end of a linkage group which differs from one species to the other (Woram *et al.* 2003). These results suggest that either the *SEX* locus has a different genetic basis in each species or that the same sex determination gene has moved to different chromosomes (Davidson *et al.* 2009). In the rainbow trout (*Oncorhynchus mykiss*), a male-specific sequence colocalized with the *SEX* locus has been recently identified (Yano *et al.* 2012). This raises the hypothesis that such a master sex-determining gene also exists in whitefish and that RAD locus *82008* may be physically closely linked to it. We are currently performing BAC library screening and sequencing in order to test this hypothesis.

Conclusions

We showed that using the RAD-Seq method greatly improves our ability to generate a high-density genetic linkage map in a nonmodel organism that has recently undergone whole-genome duplication. The whitefish RAD linkage map enabled us to map phenotypic and expression QTL as well as transmission ratio distortion QTL and thus revealed important positional information for ongoing genome scans of differentiation in natural populations. Mapping RAD markers to the next coming whitefish reference genome will be of prime interest for further elucidating the structural and functional aspects of species divergence.

Acknowledgements

We are grateful to Jason Boone and Tressa Atwood from Floragenex, as well as Doug Turnbull from the University of Oregon for their precious technical help with the development and sequencing of RAD libraries. We also thank Jérôme St-Cyr for his help with molecular experiments. Finally, we thank two anonymous referees and Subject Editor Paul Hohenlohe for their constructive comments and suggestions. This research was financially supported by a postdoctoral fellowship from the Ministry of Canadian foreign affairs to P-A.G., a Natural Sciences and Engineering Research Council of Canada Discovery grant and a Canadian Research Chair to L.B.

References

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (ed. Turner B), pp. 1–53. Plenum Press, New York.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, **188**, 799–808.

Andersen EC, Gerke JP, Shapiro JA *et al.* (2012) Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nature Genetics*, **44**, 285–290.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barton NH, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity*, **57**, 357–376.

Barton NH, de Cara MAR (2009) The evolution of strong reproductive isolation. *Evolution*, **63**, 1171–1190.

Baxter SW, Davey JW, Johnston JS *et al.* (2011) Linkage Mapping and Comparative Genomics Using Next-Generation RAD Sequencing of a Non-Model Organism. *PLoS ONE*, **6**, e19315.

Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of Lake

Whitefish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1783–1800.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci De Novo from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Chutimanitsakun Y, Nipper R, Cuesta-Marcos A *et al.* (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics*, **12**, 4.

Colosimo PF, Peichel CL, Nereng K *et al.* (2004) The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. *PLoS Biology*, **2**, e109.

Coyne JA, Orr HA (2004) *Speciation Sinauer Associates*, Sunderland, MA.

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.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Davidson WS, Huang TK, Fujiki K, von Schalburg KR, Koop BF (2009) The sex determining loci and sex chromosomes in the family Salmonidae. *Sexual Development*, **3**, 78–87.

Derome N, Bougas B, Rogers SM *et al.* (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.

Devlin RH, Nagahama Y (2002) Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture*, **208**, 191–364.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169–175.

Gharbi K, Gautier A, Danzmann RG *et al.* (2006) A linkage map for Brown Trout (Salmo trutta): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, **172**, 2405–2419.

Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics*, **24**, 408–415.

González JR, Armengol L, Solé X *et al.* (2007) SNPassoc: an R package to perform whole genome association studies. *Bioinformatics*, **23**, 654–655.

Hecht BC, Thrower FP, Hale MC, Miller MR, Nichols KM (2012) Genetic architecture of migration-related traits in rainbow and steelhead trout, oncorhynchus mykiss. *G3: Genes | Genomes | Genetics*, **2**, 1113–1127.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in Threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hughes AL (2012) Evolution of adaptive phenotypic traits without positive Darwinian selection. *Heredity*, **108**, 347–353.

Jacobsen MW, Hansen MM, Orlando L *et al.* (2012) Mito genome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (Coregonus spp.). *Molecular Ecology*, **21**, 2727–2742.

Johnston SE, McEwan JC, Pickering NK *et al.* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*, **20**, 2555–2566.

Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.

Lien S, Gidskehaug L, Moen T *et al.* (2011) A dense SNP-based linkage map for Atlantic salmon (Salmo salar) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics*, **12**, 615.

Lu G, Bernatchez L (1998) Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of Lake Whitefish (Coregonus clupeaformis Mitchill). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **265**, 1025–1030.

Matsuda M, Nagahama Y, Shinomiya A *et al.* (2002) DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, **417**, 559–563.

McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology and Evolution*, **17**, 285–291.

Miller MR, Brunelli JP, Wheeler PA *et al.* (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology*, **21**, 237–249.

Moyle LC, Graham EB (2006) Genome-Wide Associations Between Hybrid Sterility QTL and Marker Transmission Ratio Distortion. *Molecular Biology and Evolution*, **23**, 973–980.

Nanda I, Kondo M, Hornung U *et al.* (2002) A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, Oryzias latipes. *Proceedings of the National Academy of Sciences*, **99**, 11778–11783.

van Ooijen JW (2004) *MapQTL® 5 Software for the Mapping of Quatitative Trait Loci in Experimental Populations*. Kyazma B.V., Wageningen, the Netherlands.

van Ooijen JW (2006) *JoinMap® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Kyazma B.V., Wageningen, the Netherlands.

Otto SP, Jones CD (2000) Detecting the Undetected: Estimating the Total Number of Loci Underlying a Quantitative Trait. *Genetics*, **156**, 2093–2107.

Pavey SA, Collin H, Nosil P, Rogers SM (2010) The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, **1206**, 110–129.

Pfender W, Saha M, Johnson E, Slabaugh M (2011) Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *TAG Theoretical and Applied Genetics*, **122**, 1467–1480.

Phillips RB, Reed KM, Ráb P (1996) Revised karyotypes and chromosome banding of coregonid fishes from the Laurentian Great Lakes. *Canadian Journal of Zoology*, **74**, 323–329.

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). *Molecular Biology and Evolution*, **26**, 925–936.

Renaut S, Maillet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during speciation: the Lake Whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 354–363.

Rice AM, Rudh A, Ellegren H, Qvarnström A (2011) A guide to the genomics of ecological speciation in natural animal populations. *Ecology Letters*, **14**, 9–18.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, **66**, 1–17.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

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). *Journal of Evolutionary Biology*, **19**, 1979–1994.

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. *Molecular Biology and Evolution*, **24**, 1423–1438.

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

Rogers SM, Isabel N, Bernatchez L (2007) Linkage maps of the dwarf and Normal Lake Whitefish (Coregonus clupeaformis) species complex and their hybrids reveal the genetic architecture of population divergence. *Genetics*, **175**, 375–398.

Sauvage C, Vagner M, Derôme N, Audet C, Bernatchez L (2012) Coding gene single nucleotide polymorphism mapping and quantitative trait loci detection for physiological reproductive traits in Brook Charr, Salvelinus fontinalis. *G3: Genes | Genomes | Genetics*, **2**, 379–392.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372–380.

Seeb JE, Pascal CE, Grau ED *et al.* (2011) Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. *Molecular Ecology Resources*, **11**, 335–348.

Slate J, Santure AW, Feulner PGD *et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, **26**, 275–284.

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

Steiner CC, Weber JN, Hoekstra HE (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biology*, **5**, e219.

Stinchcombe JR, Hoekstra HE (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.

Van Tassell CP, Smith TPL, Matukumalli LK *et al.* (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, **5**, 247–252.

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

Whiteley AR, Derome N, Rogers SM *et al.* (2008) The phenomics and expression quantitative trait locus mapping of brain transcriptomes regulating adaptive divergence in Lake Whitefish species pairs (Coregonus sp.). *Genetics*, **180**, 147–164.

Woram RA, Gharbi K, Sakamoto T *et al.* (2003) Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Research*, **13**, 272–280.

Xu S (2003) Theoretical basis of the Beavis Effect. *Genetics*, **165**, 2259–2268.

Yano A, Guyomard R, Nicol B *et al.* (2012) An immune-related gene evolved into the master sex-determining gene in rainbow trout, Oncorhynchus mykiss. *Current Biology*, **22**, 1423–1428.

Zhang X-S, Hill WG (2005) Predictions of patterns of response to artificial selection in lines derived from natural populations. *Genetics*, **169**, 411–425.

This paper represents part of the post-doctoral work of P.-A.G. in L.B.'s Lab. P.-A.G. is a molecular population geneticist with particular interests in local adaptation and speciation genomics. E.N. is a bioinformatician developing programs and pipelines for NGS analyses. S.A.P. is an evolutionary biologist interested in the ecology and evolution of genes, gene networks and genomes in natural populations. L.B.'s research focuses on understanding patterns and processes of molecular and organismal evolution as well as their significance to adaptation, speciation and conservation.

Data accessibility

1. DNA sequences: Illumina RAD-tag Sequences have been submitted to the NCBI SRA database under project accession ID: SRA059085.
2. A complete data set including individual genotypes, morphological and expression data and input files for QTL-mapping analysis under *MapQTL5* is available in the online DRYAD archive entry doi:10.5061/dryad.sr42j.

Supporting information

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

Fig. S1 The whitefish RAD linkage map.

Fig. S2 Assessment of conserved synteny with the zebra fish genome. Only RAD markers mapping to protein coding regions identified through BLASTX search were used. Whitefish linkage groups were ordered according to their similarity with zebra fish chromosomes. Circle size is proportional to the number of mapped loci.

Table S1 Detail of Illumina read counts per individual after quality filtering and demultiplexing.

Table S2 Detail of the Lake Whitefish RAD linkage map including 3438 RAD markers.

Table S3 Detail of eQTL analysis.