REGULATORY ARCHITECTURE OF GENE EXPRESSION VARIATION IN THE THREESPINE STICKLEBACK, *GASTEROSTEUS ACULEATUS*.

Victoria L. Pritchard [1], Heidi M. Viitaniemi [1], R.J. Scott McCairns [2], Juha Merilä [2], Mikko Nikinmaa [1], Craig R. Primmer [1] & Erica H. Leder [1].

[1] Division of Genetics and Physiology, Department of Biology, University of Turku, Turku, Finland
[2] Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, Helsinki, Finland.

Abstract

Much adaptive evolutionary change is underlain by mutational variation in regions of the genome that regulate gene expression rather than in the coding regions of the genes themselves. An understanding of the role of gene expression variation in facilitating local adaptation will be aided by an understanding of underlying regulatory networks. Here, we characterize the genetic architecture of gene expression variation in the threespine stickleback (*Gasterosteus aculeatus*), an important model in the study of adaptive evolution. We collected transcriptomic and genomic data from 60 half-sib families using an expression microarray and genotyping-by-sequencing, and located QTL underlying the variation in gene expression (eQTL) in liver tissue using an interval mapping approach. We identified eQTL for several thousand expression traits. Expression was influenced by polymorphism in both *cis* and *trans* regulatory regions. *Trans* eQTL clustered into hotspots. We did not identify master transcriptional regulators in hotspot locations: rather, the presence of hotspots may be driven by complex interactions between multiple transcription factors. Observed hotspots did not co-locate with regions of the genome known to be involved in adaptive divergence between marine and freshwater habitats, suggesting they do not play a role in this well-documented stickleback radiation.

Introduction

It is now known that much adaptive evolution is underlain by changes in regions of the genome regulating gene expression, rather than in the protein coding regions of the genes themselves (Pavey *et al.* 2010). Recent work has demonstrated that much variation in gene expression is heritable, and thus evolvable via selection (e.g. Ayroles *et al*. 2009, Powell *et al.* 2013, Leder *et al.* 2015). Correspondingly, studies using model species have found that the genetic polymorphisms underlying phenotypic variation are typically not within genes (Flint and Mackay 2009). Variation in gene expression has been shown to underlie several well-documented cases of phenotypic and/or adaptive divergence. These include plumage coloration and beak shape in birds (Mallarino *et al.* 2011; Poelstra *et al.* 2015), mimetic wing patterns in butterflies (Reed *et al.* 2011; Hines *et al.* 2012), and flower colour (Durbin *et al.* 2003). Further, differences in gene expression patterns have been found to correlate with adaptive divergence in multiple species (e.g. Bernatchez *et al.* 2010; Barreto *et al.* 2011). Dysregulation of gene expression due to interactions amongst regulatory loci has potential to cause reduced fitness of inter-population hybrids and thus contribute to reproductive isolation (Ellison and Burton 2008; Turner *et al.* 2014). However, it may also promote hybrid speciation by enabling hybrids to exploit new niches (Lai *et al.* 2006).

The genetic architecture of gene expression regulation can be investigated by treating expression variation as a quantitative trait and identifying the genomic locations associated with it (termed 'expression quantitative trait loci' or 'eQTL'). Such studies have shown that the expression of a gene can be regulated by multiple genomic regions, which are traditionally classified as either *cis* or *trans*. *Cis* regulators, including promoters that activate transcription and enhancers that influence transcription levels, are located close to the regulated gene(s). They contain binding sites for regulatory molecules (proteins or mRNA) that are produced by more distant, *trans*, regulators. As *cis* regulators are expected to affect only one or a few focal genes, while *trans* regulators may have pleiotropic effects on many genes, *cis* and *trans* regulators are subject to different evolutionary dynamics. *Cis* regulatory changes are expected to be important drivers of local adaptation (Steige *et al.* 2015), while intraspecific *trans* regulatory variation is considered more likely to be under purifying selection (Schaefke *et al.* 2013 but see also Landry *et al.* 2005 for discussion of *cis-trans* coevolution). Correspondingly, *trans* regulatory polymorphisms tend to affect gene expression less strongly than *cis* polymorphisms, and their effects are more likely to be non-additive (Zhang *et al.* 2011; Gruber *et al.* 2012; Schaefke *et al.* 2013; Meiklejohn *et al.* 2014). Nevertheless, work in multiple species has demonstrated an important role for both *cis* and *trans* polymorphism in shaping expression variation (Cubillos *et al.* 2012) and the role of *trans* variation may have been underestimated due to the higher statistical power required to detect it (Mackay *et al.* 2009; Clément-Ziza *et al.* 2014). Interactions involving *trans* regulators may be particularly important in reducing the fitness of inter-population hybrids (Turner *et al.* 2014). Supporting the pleiotropic role of *trans*

regulators, a ubiquitous feature of eQTL studies is the identification of '*trans* eQTL hotspots', genomic locations associated with expression variation in many distant genes which are thought to harbour one or more important *trans* regulators (Wu *et al.* 2008; Clément-Ziza *et al.* 2014; Meiklejohn *et al.* 2014).

The threespine stickleback (*Gasterosteus aculeatus*) is an important model in the study of adaptive evolution. Ancestral anadromous populations of threespine stickleback have repeatedly and independently colonized freshwater throughout the Northern Hemisphere (Taylor and McPhail 2000; Mäkinen *et al.* 2006). Sympatric and parapatric freshwater populations may exploit different habitats (Schluter and McPhail 1992; Roesti *et al.* 2012). The species is also distributed throughout semi-marine environments with large temperature and salinity gradients, such as estuaries and the brackish water Baltic Sea (McCairns and Bernatchez 2010; Guo *et al.* 2015; Konijnendijk *et al.* 2015). Successful colonization of these diverse habitats necessitates evolutionary adaptation to novel environmental conditions including changed temperature, salinity and predation regimens, a process that can occur rapidly (Barrett *et al.* 2011; Terekhanova *et al.* 2014; Lescak *et al.* 2015). Parallel adaptations between independently founded freshwater populations frequently involve the same regions of the genome and arise from pre-existing genetic variation in the marine population (Colosimo *et al.* 2005; Hohenlohe *et al.* 2010; Jones *et al.* 2012; Liu *et al.* 2014; Conte *et al.* 2015, but see DeFaveri *et al.* 2011; Leinonen *et al.* 2012; Ferchaud and Hansen 2016). Local adaptation in environmentally heterogeneous habitats such as the Baltic Sea (Guo *et al.* 2015) and lake-stream complexes (Roesti *et al.* 2015) has been shown to involve similar genomic pathways. Evidence suggests that much of this adaptation may be due to changes in gene regulation rather than protein structure (Jones *et al.* 2012). In addition, plasticity in gene expression in response to different environmental conditions may facilitate colonization of novel habitats in the first place (McCairns and Bernatchez 2010; Morris *et al.* 2014). Leder *et al.* (2015) recently demonstrated substantial heritability of gene expression variation within a brackish-water population of threespine stickleback, confirming that it is amenable to evolution. One well-documented locally adaptive trait, reduction of the pelvic girdle, is known to be underlain by variation in the *cis* regulatory region of the *Pitx1* gene (Chan *et al.* 2010). Recently, Di Poi et al (2016) showed that differences in behaviour and response to stress between marine and freshwater sticklebacks may be modulated by variation in the expression of hormone receptors. Otherwise, the architecture of gene expression regulation in the threespine stickleback and its role in adaptive evolution is only starting to be explored (Chaturvedi *et al.* 2014).

An understanding of the potential role of gene expression variation in facilitating local adaptation will be aided by an understanding of the regulatory architecture underlying that gene expression. Here, we perform the first genome-wide study of this regulatory architecture in the threespine stickleback, by mapping QTL underlying the variation in expression of several thousand genes in a population from the Baltic Sea.

Methods

Experimental crosses.

We used a multi-family paternal half-sib crossing design for QTL mapping. Crossing procedures have previously been detailed in Leinonen *et al.* (2011) and Leder *et al.* (2015). In short, 30 mature males and 60 gravid females were collected from the Baltic Sea for use as parents. Each male was artificially crossed with two females, producing 30 half-sib blocks each containing two full-sib families. Families were reared in separate 10L tanks with density standardized to 15 individuals per tank, temperature at 17 ± 1°C and 12:12h light/dark photoperiod. At the age of six months, ten offspring from each family (5 treated, 5 controls) were subject to a temperature treatment as part of a related experiment (Leder *et al.* 2015), and immediately euthanized for DNA and RNA collection.

RNA preparation, microarray design, and data normalization

RNA preparation, gene expression microarrays, hybridization, and normalization procedures are described in detail in Leder *et al.* (2009, 2015). Briefly, total RNA was isolated from offspring liver tissue using standard protocols. RNA that passed quality thresholds was labelled (Cy3 or Cy5) using the Agilent QuickAmp Kit, with equal numbers of individuals within family groups (control & temperature-treated; males & females) assigned to each dye. Labelled RNA was hybridized to a custom 8x15 microarray, with sample order randomized (Agilent Hi-RPM kit). Images of the arrays were acquired, image analysis was performed, and array quality was assessed as detailed in Leder *et al.* (2015). Post-processed signals were standardized across arrays using a supervised normalization approach, implemented in the package 'snm' for R/Bioconductor (Mecham *et al.* 2010; R Core Team 2015). Dye, array and batch (i.e. slide) were defined as 'adjustment variables'; sex, family and temperature treatment were defined as 'biological variables'. Following normalization, individual intensity values more than two standard deviations from their family-by-treatment mean, and probes with missing values for an entire family or >10% of individuals were removed. The final dataset contained 10,527 expression traits (10,495 genes plus 32 additional splice variants) and 563 individuals (158 control females; 125 control males; 152 treated females; 128 treated males).

Genotyping-by-Sequencing

For genotyping-by-sequencing of parents and offspring we used the method of Elshire *et al.* (2011) with an additional gel excision step to improve size selection. DNA was extracted from ethanol preserved fin tissue (parents) or frozen liver tissue (offspring) and DNA concentrations were measured using a NanoDrop ND-1000 spectrophotometer. DNA (80 ng) was digested with the restriction enzyme Pst1 1.5 U (New England Biolabs) and 1x NEB buffer 3, 1x bovine serum albumin (BSA) and dH_2O (3.3 µl) in a thermocycler (37°C, 2h; 75°C, 15min; 4°C, 10min). The digested DNA was ligated to adapters with T4-ligase 0.6x (New England Biolabs), 1x Ligase Buffer, 21 µl dH_2O and

50 nM of pooled forward and reverse adapters, which were prepared according to Elshire *et al.* (2011; ligation program: 22°C, 1h; 65°C, 30min; 4°C, 10min). Up to one hundred and four unique barcodes were used in each library to label individual samples. The ligation products were pooled into libraries and purified with a QIAquick PCR Purification Kit (Qiagen). The purified libraries were PCR amplified with the following components: Purified ligated library (20µl), reaction buffer 1x, $MgCl_2$ 1.5nM (Bioline), primer mix 0.5 µM, dNTPs (Fermentas) 0.4µM, BioTaq 0.05 U (Bioline) and dH_2O (20µl) (Amplification program: [72C°, 5min; 4 cycles [95°C, 30s; 95°C, 10s; 65°C, 30s; 70°C, 30s]; 11 cycles [95°C, 10s; 65°C, 30s; 72°C, 20s]; 72°C, 5min; 4°C, 10min). Lastly, we performed a manual size selection by loading 40 µl of the amplified library on a gel (MetaPhor [Lonza] 2.5 %, 150 ml, 100 V. 1.5 h) and cutting the 300-400 bp range from the resultant smear. The DNA was extracted from the gel with a QIAquick Gel Extraction Kit (Qiagen). The cleaned product was again separated on a gel, cut and cleaned.

Six hundred and fifty one individuals, multiplexed into ten separate libraries (maximum library size = 104 individuals), were sequenced with paired-end reading on the Illumina HiSeq2000 platform by the Beijing Genomics Institute (BGI). An additional 55 individuals (including duplicates) were paired-end sequenced on Illumina HiSeq platforms at the Finnish Institute for Molecular Medicine or at the University of Oslo.

Variant calling

Reads were split by barcode, and barcodes removed, using a custom perl script. Low quality bases were removed from the reads via window adaptive trimming using Trim.pl (available: http://wiki.bioinformatics.ucdavis.edu/index.php/Trim.pl, Illumina quality score ≤ 20). Sufficient numbers of reads were obtained for 626 of the 672 individuals sent for sequencing. Paired-end reads for each of these individuals were aligned to the BROAD S1 stickleback genome using BWA aln/sampe (v 0.6.2) with default parameters (Li and Durbin 2009). The threespine stickleback genome comprises 21 assembled chromosomes plus 1,823 un-placed genomic scaffolds. Unmapped reads, and reads with non-unique optimal alignments, pair-rescued alignments, or any alternative suboptimal alignments, were discarded from resulting SAM files. SAM files were converted to sorted BAM files using samtools 0.1.18 (Li *et al.* 2009) and variants were called within each paternal family using the samtools mpileup function with extended BAQ computation (options: -AED, max-depth 500), in combination with bcftools (Li *et al.* 2009). We did not degrade mapping quality for reads with large numbers of mismatches as we found this to reject high-quality reads due to fixed polymorphisms between our European stickleback samples and the North American stickleback genome. Indel and multi-allelic variants were discarded. Initial filters based on SNP quality and variability within and across families resulted in list of 26,290 candidate bi-allelic SNPs for further analysis. Samtools and bcftools, applied to each paternal family separately, were then used to call each individual for the

genotype at each of the 26,290 sites. Sites at which bcftools identified multiple variant types (SNPs, indels and multi-base polymorphisms) within and among families were removed, leaving 25,668 successfully called variant sites.

Genotype quality control

Vcftools (Danecek *et al.* 2011) was used to recode genotypes with a genotype quality phred score (GQ) < 25 or a sequencing depth (DP) < 8 or > 1000 to missing. Vcf files for all families were merged and the merged file converted to the input format for Plink 1.07 (Purcell *et al.* 2007). For SNPs on all autosomal chromosomes and the pseudoautosomal region of Chromosome 19 (see below), the following filters were applied in Plink: hwe (based on founders only) < 0.01, maximum missing genotypes = 0.25, minor allele frequency > 0.05, non-founders with > 70% missing data removed. Adjacent SNPs in complete linkage disequilibrium were manually consolidated into a single locus, with combined SNP information used to call genotypes.

Several approaches were used check for sample contamination or errors in barcode splitting and family assignment: in Plink, the *mendel* option was used to screen families for Mendelian errors, and sample relatedness was examined by graphically visualizing genome-wide IBD-sharing coefficients generated by *genome*; the program SNPPIT (Anderson 2012) was used to assign individuals to parents, based on five independent datasets of 100 SNPs; and 220 SNPs on Stratum II of Chromosome 19 (see below) were examined for their expected pattern in males and females (all heterozygous in males vs. all homozygous in females).

The stickleback Chromosome 19 is a proto-sex chromosome (Roesti *et al.* 2013; Schultheiß *et al.* 2015), with a normally recombining pseudo-autosomal domain (approximately 0-2.5mB), a non-recombining domain in the male version (Stratum I, approximately 2.5-12mB) and a domain largely absent in the male version (Stratum II, approximately 12-20mB). For Stratum I, parental and offspring genotypes were inspected manually in order to identify the male-specific allele and this was recoded to a unique allele code ('9') for the purposes of linkage map construction. Where the male-specific allele could not be identified, all genotypes within a family were re-coded as missing. Genotypes were also inspected manually for Stratum II, and any SNP found to be heterozygous in males was excluded. All remaining Stratum II SNPs were considered to be hemizygous in males, and one of the alleles was also recoded as '9'.

Linkage map construction

We constructed a linkage map using the improved version of Crimap (Green *et al.* 1990, available: http://www.animalgenome.org/tools/share/crimap/). Remaining Mendelian errors in the dataset were removed using the *set-me-missing* option in Plink. For each SNP, the number of informative meiosis

was examined using Crimap, and markers with < 150 informative meioses or within 500bp of one another were discarded.

The initial map build included 6,448 markers. Where applicable, SNPs were ordered according to the modified genome build of Roesti *et al.* (2013). We attempted to position all previously un-placed scaffolds containing at least two genotyped SNPs on to the map. Scaffolds were assigned to chromosome on the basis of LOD score using the Crimap function *two-point*, and then positioned using a combination of information from pilot Crimap *builds*, *chrompic*, and *fixed* together with known start and end points of previously assembled scaffolds (Roesti *et al.* 2013). Information from *chrompic* and *fixed* were also used to confirm the orientation of scaffolds newly placed by Roesti *et al.* (2013). Once all possible scaffolds had been placed, recombination distance between ordered SNPs was estimated using *fixed*. To refine the map, we iteratively removed SNP genotypes contributing to implied double crossovers within a 10 cM interval (presumed to be genotyping errors), and SNPs generating recombination distances of >1cM per 10,000 bp and recalculated distances using *fixed*. Remaining regions of unusually high recombination on the map were investigated by examining whether removal of individual SNPs altered map distance.

eQTL identification

Expression QTL (eQTL) were identified using an interval mapping approach (Knott *et al.* 1996) implemented in QTLMap 0.9. 0 (http://www.inra.fr/qtlmap; QTLMap option: -- *data-transcriptomic*). As we found that missing values in the expression trait file caused QTLMap to over-estimate the LRT statistic (see below), we eliminated these from the dataset by removing two individuals and 195 expression traits. Eighty-seven genotyped parents, 474 genotyped and phenotyped offspring (mean no. offspring per family = 15.8, mean proportion of missing genotypes in offspring = 0.11; maximum = 0.56), and 10,332 expression traits were included in the analysis. We applied linkage analysis assuming a Gaussian trait distribution (QTLMap option: --*calcul = 3*), and included dye, temperature treatment, and sex as fixed factors in the model. Due to the relatively small size of some of our half-sib families, we examined sire effects only, with a separate QTL effect estimated for each sire. A fast algorithm was used to identify phase and estimate transmission probabilities at each chromosomal location (Elsen *et al.* 1999, QTLMap option: --*snp*). Autosomes and the pseudoautosomal portion of the sex chromosome were scanned at 1cM intervals, and the presence of QTL on a chromosome was assessed using a likelihood ratio test (LRT) under the hypothesis of one versus no QTL. LRT significance thresholds for each trait on each chromosome were identified empirically, by permuting fixed effects and traits amongst individuals within families and recalculating LRT scores (5000 permutations). As the combination of 5000 permutations x 10,332 traits x 21 chromosomes was computationally prohibitive, we first performed permutations on a subset of 200 expression traits to identify a LRT threshold below which identified QTL were unlikely to be significant at chromosome-

wide $p < 0.05$ (LRT = 55), and then used permutations to assess significance of all QTL above this threshold. The non-pseudo-autosomal region of the female Chromosome 19 can be considered analogous to the X chromosome; identification of QTL in this region requires estimation of dam effects and was therefore not performed. The 95% confidence interval for each QTL was estimated using the drop-off method implemented in QTLMap 0.9.7, which returns flanking map positions plus their nearest marker.

Cis vs. trans eQTL

To discriminate *cis* vs. *trans* QTL, we compared inferred QTL location to the position of the expressed gene according to the BROAD *G. aculeatus* genome annotation v. 1.77 (available: http://ftp.ensembl.org/pub/release-77/gtf/gasterosteus_aculeatus/). All positions on the BROAD annotation were re-coded to positions on our modified chromosome assemblies. We considered a QTL to be in *cis* if the SNP closest to the upper or lower 95% confidence bounds of that QTL was within 5Mb of the regulated gene; all other QTL were considered *trans*-QTL. For genes on scaffolds un-anchored to our assembly, we also used information on scaffold position available in the recently published map of Glazer *et al.* (2015). Following Johnsson *et al.* (2015) we applied a local significance threshold (chromosome-wide $p < 0.01$) for evaluation of possible *cis*-QTL and a genome-wide significance threshold (genome-wide $p < 0.021$, = chromosome-wide threshold of 0.001 * 21 chromosomes) for evaluation of possible *trans*-QTL. Although this significance threshold is permissive, we considered it acceptable as our aim was to analyse the eQTL distribution across the genome rather than to identify individual QTL-locus associations. Similar significance thresholds have been used for eQTL detection in comparable studies (e.g. Whiteley *et al.* 2008).

To ask whether the effect of variation in *trans* regulatory sites was more often non-additive than the effect of variation in *cis* regulatory sites, we examined the narrow sense heritability (h^2) and dominance proportion of genetic variance (d^2) estimated for each expression trait by Leder *et al.* (2015) and provided in the Supplementary Data for that paper.

Genes with plastic vs. non-plastic expression

To investigate whether genes exhibiting an alteration in expression level in response to a temperature stress treatment (i.e. those exhibiting environmental plasticity) had a different underlying regulatory architecture to those not exhibiting such a response, we divided genes into a 'responding' and 'non-responding' group based on the results in Leder *et al.*(2015) and compared the frequency and position of *cis* and *trans* eQTL between the two groups.

Evaluation of eQTL hotspots

As all identified eQTL had wide 95% confidence intervals, meaning that physically close eQTL positions could be due to the effect of the same locus (see below), we evaluated potential eQTL hotspots by counting eQTL within 5cM bins across the genome ('hotspot size' = number of eQTL). Where the number of 1cM bins within a chromosome was not a simple multiple of 5, bin sizes at the start and/or end of the chromosome were increased to 6 or 7. To obtain an empirical significance threshold above which clusters of eQTL could be considered a 'hotspot', we simulated the expected neutral distribution of eQTL across the genome using a custom script. We performed 5000 simulations: for each, we assigned *n* eQTL (where *n* = relevant number of significant eQTL) randomly across the 3,062 1cM bins of the genome and then summed them into 5cM (or larger) bins as described above. Conservatively, we compared the size of hotspots in the real data to the size distribution of the largest hotspot observed over each of the 5000 simulations.

Association of eQTL with regions under selection

Hohenlohe *et al.* (2010), Jones *et al.* (2012), and Terekhanova *et al.* (2014) documented parallel regions of the genome divergent between marine and freshwater sticklebacks on Chromosomes 1, 4 (three regions), 7, 11 and 21, and clusters of QTL associated with morphological variation also occur on Chromosome 20 (Miller *et al.* 2014). We investigated whether these regions harboured important *trans* regulators that might contribute to marine/freshwater adaptation by comparing the location of these regions with the location of our identified *trans* eQTL hotspots. We also compared hotspot location to regions of the genome inferred by Guo *et al.* (2015) to be involved in adaptive differentiation amongst different stickleback populations in the Baltic Sea.

Ortholog identification

In order to maximize the functional information available, we identified human orthologues for *G. aculeatus* genes. As a first attempt, we used BioMart (Durinck *et al.* 2005; Smedley *et al.* 2009) to identify human orthologues and obtain the HGNC symbols for the human genes. When BioMart failed to return a human orthologue, protein BLAST searches were used to identify orthologues using the Ensembl human protein database. The identifier conversion tool, db2db, from bioDBnet (https://biodbnet-abcc.ncifcrf.gov/db/db2db.php) was used to convert between Ensembl identifiers and HGNC gene symbols when needed (Mudunuri *et al*. 2009).

Hotspot annotation

For functional annotation analysis of *G. aculeatus* genes, Human Ensembl IDs were used as input into AmiGO2 (Carbon et al. 2009) or the Database for Annotation, Visualization and Integrated Discovery (DAVID, Huang *et al.* 2009a; b). To identify regulatory genes physically associated with an eQTL hotpot, we defined hotspot confidence boundaries as being the most frequently observed 95% confidence limits of all significant eQTL centred in the hotspot. We identified the map markers

closest to the two boundaries (Table S3), and used AmiGO2 to search for intervening genes annotated with 'molecular function' or 'biological process' Gene Ontology (GO) terms that contained the words 'transcription' and 'regulation'. As an important transcriptional regulator generating a hotspot might itself be regulated by the hotspot rather than physically present within it, we repeated this analysis for all genes with eQTL mapped to the hotspot. We used DAVID to examine GO term enrichment for the sets of genes with significant eQTL mapping to each hotspot, using the 9,071 genes on the microarray with identified human orthologues as the background.

Upstream regulator and functional interaction analyses

To search for regulatory genes which may be responsible for the expression variation in genes with identified *trans* eQTL, we used the upstream regulator analysis in the Ingenuity Pathway Analysis (IPA) software (Qiagen). This analysis uses a Fisher's Exact Test to determine whether genes in a test dataset are enriched for known targets of a specific transcription factor. We used the human HGNC symbols as identifiers in IPA. First we examined all genes that that had a significant *trans*- eQTL mapping to any location at a genome-wide $p < 0.021$ (chromosome –wide $p< 0.001$). To investigate in more detail the upstream regulators potentially involved in generating eQTL hotspots, we lowered our stringency and also examined all genes with *trans* eQTL mapping to the hotspot locations at genome-wide $p < 0.057$ (chromosome-wide $p < 0.0027$).

Since transcription is typically initiated by a complex of genes rather than a single transcription factor, we examined functional relationships among the identified upstream regulators for each hotspot (Table S7b), the genes located within a hotspot, and the genes with significant eQTL mapping to that hotspot (Table S3; *cis* eQTL significant at chromosome-wide $p < 0.01$, *trans* eQTL significant at genome-wide $p < 0.021$), using STRING v10 (Jensen *et al.* 2009, http://string-db.org/). We searched for evidence of functional relationships from experiments, databases and gene co-expression, and applied a minimum required interaction score of 0.4.

Results

Genotyping by sequencing

For the 604 sticklebacks that we retained for analysis, we obtained a total of 583,032,024 raw paired reads (40,357 – 11,940,726 per individual, median=834,286). Approximately 67% of these reads remained aligned to the stickleback genome following removal of reads with non-unique optimal alignments, any alternative suboptimal alignments, or pair-rescued alignments (range 36.2% - 78.8%, median = 70.1%). Raw read and alignment statistics for each individual are provided in Table S0.

Linkage map construction

Following SNP calling and quality control steps 13,809 of the original 25,668 SNPs, genotyped in 604 individuals (mean number of offspring per family = 18), were available for linkage map construction. Following removal of markers with < 150 informative meioses or within 500bp, 6,448 SNPS were included in the initial map build. The final sex-averaged linkage map spanned 3,110 cM Kosambi (including the complete Chromosome 19) and included 5,975 markers, of which approximately 45% were located at the same map position as another marker (Figure 1, Figure S1, Table S1). Forty-three previously un-placed scaffolds (10.35 mB) were added to the chromosome assemblies of Roesti *et al.* (2012, Table S2). Thirty-five of these scaffolds were also recently added to the stickleback assembly in an independent study by Glazer *et al.* (2015). Although there were some differences in scaffold orientation, location of the new scaffolds was almost completely congruent between the two maps (Table S2). For QTL detection with QTLMap, the map was reduced to 3,189 SNPs with unique positions (average inter-marker distance = 0.98cM, Table S1).

Identification of cis and trans eQTL

At chromosome-wide $p < 0.01$, we identified 5,226 eQTL associated with 4,411 expression traits (42.7% of the 10,322 expression traits examined, Table S3). Based on our recoded gene positions, we classified 2,072 of these as *cis* eQTL, 2,988 as *trans* eQTL, and 165 as unknown – that is, the expressed gene was located on a scaffold that had not been assigned to a *G. aculeatus* chromosome by either this study or Glazer *et al.* (2015; Table S3, Table S4). Five hundred and eighty of the *trans* eQTL were significant at genome-wide $p < 0.021$. Of these, 68.3% mapped to a chromosome other than the one containing the regulated gene. After application of this genome-wide significance threshold for *trans* eQTL, 2,713 expression traits (26.3% of those examined) remained associated with one or more significant *cis* or *trans* eQTL. Of these, 74.3% were associated with a *cis* eQTL, 18.9% with one or more *trans* eQTL, 2.1% with both a *cis* and a *trans* eQTL and 4.7% with eQTL of unknown class (Table S3). The physical distribution across the genome of the 2,713 loci with significant *cis* or *trans* eQTL is shown in Figure S1. Mean 95% confidence interval of significant eQTL was 10.1 cM (range 1-74cM). Overall, *trans* regulated expression traits did not exhibit more dominance variance than *cis* regulated loci (*trans* regulated loci, mean $h^2 = 0.32$, mean $d^2 = 0.16$; *cis* regulated loci: mean $h^2 = 0.37$, mean $d^2 = 0.18$; values from Leder *et al.* 2015).

Trans eQTL hotspots

Trans eQTL (significant at genome wide $p < 0.021$) were not evenly distributed across the genome and we identified eight 5cM bins, located on six different chromosomes, as containing eQTL clusters (7 or more eQTL; $p < 0.05$ based on the largest hotspot observed in neutral simulations; Figure 1). A particularly large eQTL hotspot (38 *trans* eQTL within the 5cM bin) was identified close to one end of Chromosome 6, three hotspots (18, 10, and 9 *trans* eQTL) were present at separate locations on Chromosome 12, and remaining hotspots were located near the ends of Chromosomes 7, 8, 9 and 16

(13, 11, 7 and 9 *trans* eQTL). To eliminate the possibility that distant *cis* eQTL mis-classified as *trans* were contributing to observed hotspots, we repeated the analysis with the 396 *trans* eQTL that were on a different chromosome to their regulatory target: the same eight hotspots were identified (7 or more eQTL, $p < 0.004$). Physical hotspot boundaries were assigned from inspection of eQTL hits and 95% confidence intervals as follows: Chromosome 6, 111-116cM ('Chr6', 17,238,934-17,469,219bp); Chromosome 7, 5-12cM ('Chr7', 396,541-1,107,393bp); Chromosome 8, 134-139cM ('Chr8', 19,917,746-20,316,565bp); Chromosome 9, 165-174cM ('Chr9', 19,822,078-20,440,410bp); Chromosome 12, 0-1cM ('Chr12a', 0-337,849bp); Chromosome 12, 72-79cM ('Chr12b', 5,853,981-7,440,742bp); Chromosome 12, 109-119cM ('Chr12c', 15,551,555-17,229,387bp); Chromosome 16, 123-130cM ('Chr16', 17,658,526-18,257,571bp).

Genes with plastic vs. non-plastic expression

Following FDR correction, 4,253 genes were found by Leder *et al.* (2015) to exhibit a significant change in expression in response to a temperature treatment. We identified significant eQTL underlying 1,033 of these genes (Table S3; eQTL type: 76.0% *cis*, 18.0% *trans*, 2.2% both, 3.8% unknown). The distribution of the 216 significant *trans* eQTL across 5cM bins indicated five hotspots (5 or more eQTL, $p < 0.02$, Figure S2), four of which had been previously observed in the full dataset. The Chromosome 16 hotspot was greatly increased in relative importance, and a new hotspot was observed on Chromosome 18 (Chr 6: 12 eQTL; Chr16: 9 eQTL; Chr12a: 5 eQTL; Chr12b: 5 eQTL; Chromosome 18, 'Chr18': 5 eQTL, 96-102cM, 13,870,895-14,643,331bp).

Association of eQTL with regions under selection

None of our identified eQTL hotspots overlapped parallel regions of the genome divergent between marine and freshwater sticklebacks identified by Hohenlohe *et al.* (2010), Jones *et al.* (2012), and Terekhanova *et al.* (2014), or with the clusters of morphological QTL on Chromosome 20 (Miller *et al.* 2014, Table S5). However, one genomic region identified as divergent between marine and freshwater populations by Terekhanova *et al.* (2014) alone overlapped with the Chr12b eQTL hotspot. Only four of the 297 genes inferred by Guo *et al.* (2015) as being under selection amongst Baltic Sea populations experiencing different temperature and salinity regimens overlapped observed eQTL hotspots (Chr7 and Chr12b, Table S5).

Hotspot annotation

We identified human orthologues for 16,315 of the 20,787 protein-coding genes annotated on the Broad stickleback genome (78.5%, Table S4). There were 300 genes with human annotation physically located within the designated boundaries of the nine hotspots (Table S5). Of these, 41 had a GO term related to transcription regulation (Table 1, Table S6). In addition, 21 genes with significant *cis* eQTL or *trans* eQTL mapping to a hotspot had GO terms related to transcriptional

regulation (Table 1, Table S6). Following correction for multiple testing we found no significant GO term enrichment amongst any group of genes regulated by the same eQTL hotspot.

Upstream regulator and functional interaction analyses

When examining all 580 genes with *trans* eQTL significant at genome wide $p < 0.021$, 84 significantly enriched upstream regulators were identified (Table S7a). In total, these regulators had 244 of the genes in the dataset as known targets. Hepatocyte nuclear factor 4α (HNF4A) was identified as a particularly important regulator, with 73 (29.9%) of these genes as downstream targets. Other important regulatory factors were: tumor protein p53 (TP53; 40 genes; 16.4%); estrogen receptor 1 (ESR1; 38 genes; 15.6%); myc proto-oncogene protein (MYC; 30 genes; 12.3%) and huntingtin (HTT; 27 genes; 11.1%). The full list of 85 significant upstream regulators is in Table S7a.

To identify upstream regulators that could be contributing to the nine eQTL hotspots (including one only observed when examining genes with a plastic response to temperature), we further examined all genes that had *trans* eQTL mapping to the hotspots at genome-wide $p< 0.057$ (1120 genes). One hundred and fifty seven different enriched upstream regulators were identified for these genes (Table S7b). For genes with *trans* eQTL mapping to the Chr6, Chr12a, Chr12b, Chr12c and Chr18 hotspots, HNF4A remained an important regulator. Only two of the identified upstream regulators were physically located within a hotspot (serum response factor, SRF, Chr9; nuclear receptor subfamily 4, group A, member 1, NR4A1, Chr12b). Two had significant *trans* eQTL mapping to the Chr6 hotspot: catenin (cadherin-associated protein) beta (CTNNB1) and hypoxia inducible factor 1 alpha (HIF1A). One had a significant *trans* eQTL mapping to the Chr7 hotspot: junction plakoglobin (JUP), and one had a significant trans eQTL mapping to the Chr12b hotspot: Nuclear Receptor Subfamily 1, Group H, Member 4 (NR1H4; Table 1).

When the enriched upstream regulators, genes with cis eQTL mapping to a hotspot at chromosome-wide $p < 0.01$, and genes with trans eQTL mapping to a hotspot at genome wide $p < 0.021$ were examined in STRING, multiple protein-protein interactions were found (Figure 2, Figure S4). In particular for the Chr6 hotspot we found an interaction network that included two molecules *trans*-regulated by this hotspot (CTNNB1 and HIF1A), one molecule *cis*-regulated by the hotspot (C1D Nuclear Receptor Co-Repressor), and multiple molecules inferred as important upstream regulators by IPA (Figure 2a). Similarly, for the Chr12b hotspot, we observed a large network of interactions involving molecules *cis* and *trans* regulated by the hotspot, molecules produced by genes physically located in the hotspot, and inferred upstream regulators (Figure 2b).

Discussion

In this study we identified regions of the genome underlying variation in gene expression in a marine population of threespine stickleback from northern Europe. We used a genotyping-by-sequencing

approach to generate an improved linkage map, and applied interval mapping to identify eQTL. Our new map was independent of that recently constructed by Glazer *et al.* (2015), and the congruent placement of scaffolds between the two maps confirms the reliability of these new genome assemblies. Our map covered a substantially larger distance in cM than those of Roesti *et al.* (2013) and Glazer *et al.* (2015), probably due to differences in experimental design. Nevertheless, for our Baltic Sea population, we observe very similar patterns of recombination rate variation across and between chromosomes as found by Roesti et al. (2013) for freshwater sticklebacks from central Europe and Glazer *et al.* (2015) for marine-freshwater crosses from western North America, (Figure S1). Thus, the large scale pattern of recombination rate variation across the genome may impose, and/or be under, similar evolutionary constraints throughout the range of the species.

Using a chromosome-wide significance threshold for *cis* regulatory loci and a genome-wide threshold for *trans* loci, we identified eQTL for just over a quarter of the 10,332 expression traits examined. Because at least 74% of these expression traits exhibit significant heritable variation (Leder *et al.* 2015), and gene expression is commonly regulated by multiple eQTL, we expect that a much larger number of underlying eQTL remain undetected due to low statistical power. Despite expectations that *trans* regulatory regions might be under purifying selection due to their potentially pleiotropic effect, and that the effect of *trans* eQTL on expression will be weaker than that *cis* eQTL, we found many cases where gene expression was influenced by regulatory variation in *trans* but not in *cis*. This suggests that a frequently-used approach of detecting local selection by examining patterns of differentiation at markers linked to genes that are adaptive candidates (e.g. DeFaveri et al. 2011, Shimada et al. 2011) may fail to identify such selection as it is acting to change gene expression via *trans* regulatory regions. We did not observe any difference in additive vs dominance variance underlying genes found to be regulated in *cis* vs. those regulated in *trans*. However this may again be due to low statistical power to detect many of the underlying eQTL: genes are expected to be influenced by a large number of eQTL, meaning that the observed heritable variation is generated by a combination of additively and non-additively acting regulatory regions.

The *trans* eQTL that we detected were not randomly distributed across the genome but instead clustered into multiple eQTL hotspots. This observation is a ubiquitous feature of eQTL studies and is thought to indicate the existence of 'master regulators' acting in *trans* to influence many genes. However apparent eQTL hotspots may also arise as a statistical artefact as a result of many false positive QTL when testing thousands of expression traits in combination with spurious correlation between these traits due to uncorrected experimental factors (Wang *et al.* 2007; Breitling *et al.* 2008). Disentangling gene expression correlation that is due to common underlying regulatory architecture from that caused by experimental artefacts is a difficult analytical problem that we are unable to fully address here (Joo *et al.* 2014). Therefore, we caution that these hotspots should be verified using other stickleback populations and different approaches.

The parents for this study came from a genetically diverse marine population of threespine stickleback (DeFaveri *et al.* 2013). Local adaptation of threespine sticklebacks to freshwater has been demonstrated to arise, at least partly, from selection on standing genetic variation in the marine environment. Further, QTL underlying morphological divergence between marine and freshwater populations have been demonstrated to have pleiotropic effects (Rogers *et al.* 2012; Miller *et al.* 2014), and frequently co-localize with regions of the genome found to be under parallel selection amongst independent freshwater colonisations. One way in which these regions could exert such pleiotropic effects is by harbouring loci that influence the expression of many genes, i.e. eQTL hotspots. However, only one of the *trans* eQTL hotspots found in this study overlapped with genomic regions found to be associated with marine/freshwater divergence by Hohenlohe *et al.* (2010), Jones *et al.* (2012), or Terekhanova *et al.* (2014), indicating that they do not underlie the multiple parallel changes observed when sticklebacks colonize freshwater. It remains possible that regulatory hotspots acting in tissues or life stages that we did not examine do have a role in this freshwater adaptation.

To investigate the potential genetic mechanisms underlying the nine observed eQTL hotspots we searched for associated loci with known transcriptional regulatory functions, and performed upstream regulator analysis for the genes with eQTL in the hotspots. Although the pathways regulating transcription are still poorly characterized for most genes, particularly in non-mammalian species, these analyses can provide useful preliminary information. We found no evidence that eQTL hotspots were due to the presence of a single 'master' regulatory locus, or a cluster of regulatory genes, at the hotspot locations. Although many genes with roles in transcriptional regulation were present in, or regulated by, hotspots, finding such genes is not unexpected: approximately 10% of the human orthologues of BROAD stickleback genes are annotated with the GO terms that we used to identify transcriptional regulators. It is also possible that the regulatory elements generating such hotspots are not annotated coding genes: microRNAs and long non-coding RNAs are potentially important *trans* regulators (Vance and Ponting 2014) and not yet well characterized across the stickleback genome.

Our results suggest that, alternatively, these hotspots may be generated by a complex interaction of multiple transcription regulators. Several well-characterized regulatory proteins were identified as upstream regulators for numerous genes with eQTL within the hotspots. In particular, HNF4A was identified as a strongly enriched regulator for all genes with significant *trans* eQTL (Table S7a), and the subsets of genes with *trans* eQTL mapping to the hotspots on Chromosome 6, Chromosome 12, and Chromosome 18 (Table S7b). In mammals, HNF4A is known to be a master regulator of transcription in the liver (Odom *et al*. 2004). Although the gene is not physically located in any hotspot, and we were unable to identify any significant eQTL underlying its expression, it is less than 300 kb from hotspot Chr12b. HNF4A likely acts through direct and indirect interactions with other proteins to regulate transcription. Interacting molecules particularly of interest in respect to hotspot locations are HIF1A and CTNNB1 (*trans* regulated by the Chr6 hotspot, Fig. 2a) and NR4A1 (located

in the Chr12b hotspot, Fig. 2b): all of these are also identified as significantly enriched upstream regulators when examining genes with trans eQTL mapping to any of the nine hotspots (Table S7b). CTNNB1 is an important transcriptional coactivator in the cell nucleus (Willert & Jones 2006). NR4A1 along with its subfamily members NR4A2 (*trans* regulated by the Chr16 hotspot) and NR4A3 (not on microarray) are orphan nuclear receptors that interact with other regulators to influence transcription (Ranhotra 2015). From the point of view of local adaptation, HIF1A is particularly interesting. It is part of a transcriptional complex (HIF) that alters the expression of numerous genes in response to low oxygen conditions. HIF1A has been demonstrated to regulate responses to hypoxia in fishes (Nikinmaa and Rees 2005 Liu *et al*. 2013) and is also involved in inflammation and temperature adaptation (Rissanen *et al.* 2006; Liu *et al.* 2013). It has been investigated as a possible selective target for adaptation to low-oxygen conditions, such as those encountered in benthic habitats, in various fish species. Rytkönen *et al.* (2007) found no association between variation in the HIF1A coding region and adaptation to hypoxic conditions across species, and markers linked to HIF1A do not appear be under directional selection amongst Baltic Sea stickleback populations (Shimada *et al.* 2011). HIF1A is known to be transcriptionally regulated in fish (Liu *et al.* 2013), and the identification of a *trans* eQTL for HIF1A demonstrates that regulatory variation for this gene is present in Baltic Sea sticklebacks and could be an alternative target of selection. HNF4A has also been found to be an important regulator of hypoxia response (Xu *et al.* 2011).

HNF4A was not implicated in the regulation of genes with *trans* eQTL mapping to the Chr7, Chr8, Chr9 or Chr16 hotspots, suggesting that different regulatory complexes may be underlie these additional hotspots. Comparison of the regulatory architecture underlying genes exhibiting a plastic response to the temperature treatment to that underlying genes not responding indicates that the Chr16 and Chr18 eQTL hotspots are particularly strongly associated with this gene expression plasticity. These eQTL hotspots are both linked with the gene bone morphogenic protein 2 (BMP-2, Table 1), suggesting that this may have a role in mediating such plasticity, although we are unable to examine this further here.

In conclusion, we have performed the first genome-wide characterisation of the regulatory architecture of gene expression in *G. aculeatus.* We found that variation in gene expression was influenced by polymorphism in both *cis*-acting and *trans* acting regulatory regions. Trans-acting eQTLS clustered into hotspots, however these did not co-locate with regions of the genome known to be involved in adaptive divergence among marine and freshwater threespine sticklebacks. Hotspots locations appear to be mediated by complex interactions amongst regulator molecules rather than the presence of few 'master regulators'. Our broad-scale study suggests many avenues for finer-scale investigation of the role of transcriptional regulation in stickleback evolution.

Data accessibility

Raw and normalized microarray data, in addition to R scripts describing the normalization procedure, are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-3098. RAD sequence reads for each individual are to be deposited in the NCBI Sequence Read Archive. Input files and scripts will be deposited in a relevant archive on article publication.

Acknowledgements

This work was funded by grants from Academy of Finland to C.P., E.L., M. N. and J.M. (grant nos. 129662, 134728, 133875, 136464, 141231, 250435 and 265211). Tiina Sävilammi wrote the script for barcode splitting and we thank her for bioinformatics advice. We also thank the numerous people who helped in obtaining and maintaining sticklebacks. Generous computing resources were provided by the Finnish Centre for Scientific Computing (CSC-IT). Research was conducted under an ethical license from the University of Helsinki (HY121-06).

References

Anderson E. C., 2012 Large-scale parentage inference with SNPs: an efficient algorithm for statistical confidence of parent pair allocations. Stat. Appl. Genet. Mol. Biol. **11**.

Ayroles J. F., Carbone M. A., Stone E. A., Jordan K. W., Lyman R. F., Magwire M. M., *et al*. 2009 Systems genetics of complex traits in *Drosophila melanogaster*. Nat. Genet. **41**: 299–307.

Barreto F. S., Moy G. W., Burton R. S., 2011 Interpopulation patterns of divergence and selection across the transcriptome of the copepod *Tigriopus californicus*. Mol. Ecol. **20**: 560–72.

Barrett R. D. H., Paccard A., Healy T. M., Bergek S., Schulte P. M., Schluter D., Rogers S. M., 2011 Rapid evolution of cold tolerance in stickleback. Proc. Biol. Sci. **278**: 233–8.

Bernatchez L., Renaut S., Whiteley A. R., Derome N., Jeukens J., Landry L., *et al.* 2010 On the origin of species: insights from the ecological genomics of lake whitefish. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **365**: 1783–800.

Breitling R., Li Y., Tesson B. M., Fu J., Wu C., Wiltshire T., *et al*. 2008 Genetical genomics: spotlight on QTL hotspots. PLoS Genet. **4**: e1000232.

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., AmiGO Hub, Web Presence Working Group, 2009. AmiGO: online access to ontology and annotation data. Bioinformatics. 25: 288-9.

571 Chan Y. F., Marks M. E., Jones F. C., Villarreal G., Shapiro M. D., Brady S. D., *et al.* 2010 Adaptive
572 evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science
573 **327**: 302–5.

574 Chaturvedi A., Raeymaekers J. A. M., Volckaert F. A. M., 2014 Computational identification of
575 miRNAs, their targets and functions in three-spined stickleback (*Gasterosteus aculeatus*). Mol.
576 Ecol. Resour. **14**: 768–77.

577 Clément-Ziza M., Marsellach F. X., Codlin S., Papadakis M. A., Reinhardt S., Rodríguez-López M.,
578 *et al.* 2014 Natural genetic variation impacts expression levels of coding, non-coding, and
579 antisense transcripts in fission yeast. Mol. Syst. Biol. **10**: 764.

580 Colosimo P. F., Hosemann K. E., Balabhadra S., Villarreal G., Dickson M., Grimwood J., *et al.* 2005
581 Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles.
582 Science **307**: 1928–33.

583 Conte G. L., Arnegard M. E., Best J., Chan Y. F., Jones F. C., Kingsley D. M., *et al.* 2015 Extent of
584 QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. Genetics
585 **201**: 1189–200.

586 Cubillos F. A., Coustham V., Loudet O., 2012 Lessons from eQTL mapping studies: non-coding
587 regions and their role behind natural phenotypic variation in plants. Curr. Opin. Plant Biol. **15**:
588 192–8.

589 Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., DePristo M. A., *et al.* 2011 The variant
590 call format and VCFtools. Bioinformatics **27**: 2156–8.

591 DeFaveri J., Jonsson P. R., Merilä J., 2013 Heterogeneous genomic differentiation in marine
592 threespine sticklebacks: adaptation along an environmental gradient. Evolution **67**: 2530–2546.

593 DeFaveri J., Shikano T., Shimada Y., Goto A., Merilä J., 2011 Global analysis of genes involved in
594 freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*). Evolution **65**: 1800–
595 7.

596 Di-Poï C., Bélanger, D., Amyot, M. Rogers, S., Aubin-Horth, N. 2016. Receptors rather than signals
597 change in expression in four physiological regulatory networks during evolutionary divergence
598 in threespine stickleback. Mol. Ecol. (accepted).

599 Durbin M. L., Lundy K. E., Morrell P. L., Torres-Martinez C. L., Clegg M. T., 2003 Genes that
600 determine flower color: the role of regulatory changes in the evolution of phenotypic

adaptations. Mol. Phylogenet. Evol. **29**: 507–18.

Durinck S., Moreau Y., Kasprzyk A., Davis S., Moor B. De, Brazma A., Huber W., 2005 BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics **21**: 3439–40.

Ellison C. K., Burton R. S., 2008 Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. Proc. Natl. Acad. Sci. U. S. A. **105**: 15831–15836.

Elsen J.-M., Mangin B., Goffinet B., Boichard D., Roy P. Le, 1999 Alternative models for QTL detection in livestock. I. General introduction. Genet. Sel. Evol. **31**: 213.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., Buckler E. S., Mitchell S. E., 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One **6**: e19379.

Ferchaud A.-L., Hansen M. M., 2016 The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: threespine sticklebacks in divergent environments. Mol. Ecol. **25**: 238–295.

Flint J., Mackay T. F. C., 2009 Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res. **19**: 723–33.

Glazer A. M., Killingbeck E. E., Mitros T., Rokhsar D. S., Miller C. T., 2015 Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. G3 **5**: 1463–72.

Green, P., Falls, K., Crooks, S. 1990. Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St. Louis, MO.

Gruber J. D., Vogel K., Kalay G., Wittkopp P. J., 2012 Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. PLoS Genet. **8**: e1002497.

Guo B., DeFaveri J., Sotelo G., Nair A., Merilä J., 2015 Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. BMC Biol. **13**: 19.

Hines H. M., Papa R., Ruiz M., Papanicolaou A., Wang C., Nijhout H. F., *et al*. 2012 Transcriptome analysis reveals novel patterning and pigmentation genes underlying *Heliconius* butterfly wing pattern variation. BMC Genomics **13**: 288.

630 Hohenlohe P. A., Bassham S., Etter P. D., Stiffler N., Johnson E. A., Cresko W. A., 2010 Population
631 genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS
632 Genet. **6**: e1000862.

633 Huang D. W., Sherman B. T., Lempicki R. A., 2009a Systematic and integrative analysis of large
634 gene lists using DAVID bioinformatics resources. Nat. Protoc. **4**: 44–57.

635 Huang D. W., Sherman B. T., Lempicki R. A., 2009b Bioinformatics enrichment tools: paths toward
636 the comprehensive functional analysis of large gene lists. Nucleic Acids Res. **37**: 1–13.

637 Jensen L. J., Kuhn M., Stark M., Chaffron S., Creevey C., Muller J., *et al.* 2009. STRING 8 - a global
638 view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. **37**:
639 D412-6.

640 Johnsson M., Jonsson K. B., Andersson L., Jensen P., Wright D., 2015 Genetic regulation of bone
641 metabolism in the chicken: similarities and differences to Mammalian systems. PLoS Genet. **11**:
642 e1005250.

643 Jones F. C., Grabherr M. G., Chan Y. F., Russell P., Mauceli E., Johnson J., Swofford R., *et al*. 2012
644 The genomic basis of adaptive evolution in threespine sticklebacks. Nature **484**: 55–61.

645 Joo J. W. J., Sul J. H., Han B., Ye C., Eskin E., 2014 Effectively identifying regulatory hotspots while
646 capturing expression heterogeneity in gene expression studies. Genome Biol. **15**: r61.

647 Knott S. A., Elsen J. M., Haley C. S., 1996 Methods for multiple-marker mapping of quantitative trait
648 loci in half-sib populations. Theor. Appl. Genet. **93**: 71–80.

649 Konijnendijk N., Shikano T., Daneels D., Volckaert F. A. M., Raeymaekers J. A. M., 2015 Signatures
650 of selection in the three-spined stickleback along a small-scale brackish water - freshwater
651 transition zone. Ecol. Evol. **5**: n/a–n/a.

652 Lai Z., Gross B. L., Zou Y., Andrews J., Riesberg L. H., 2006 Microarray analysis reveals differential
653 gene expression in hybrid sunflower species. Mol. Ecol. **15**: 1213–1227.

654 Landry C. R., Wittkopp P. J., Taubes C. H., Ranz J. M., Clark A. G., Hartl D. L., 2005 Compensatory
655 cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of
656 *Drosophila*. Genetics **171**: 1813–22.

657 Leder E. H., McCairns R. J. S., Leinonen T., Cano J. M., Viitaniemi H. M., Nikinmaa M., *et al*. 2015
658 The evolution and adaptive potential of transcriptional variation in sticklebacks--signatures of
659 selection and widespread heritability. Mol. Biol. Evol. **32**: 674–89.

660 Leder E., Merilä J., Primmer C., 2009 A flexible whole-genome microarray for transcriptomics in
661 three-spine stickleback (*Gasterosteus aculeatus*). BMC Genomics **10**: 426.

662 Leinonen T., Cano J., Merilä J., 2011 Genetics of body shape and armour variation in threespine
663 sticklebacks. J. Evol. Biol. **24**: 206–218.

664 Leinonen T., McCairns R. J. S., Herczeg G., Merilä J., 2012 Multiple evolutionary pathways to
665 decreased lateral plate coverage in freshwater threespine sticklebacks. Evolution **66**: 3866–75.

666 Lescak E. A., Bassham S. L., Catchen J., Gelmond O., Sherbick M. L., Hippel F. A. von, Cresko W.
667 A., 2015 Evolution of stickleback in 50 years on earthquake-uplifted islands. Proc. Natl. Acad.
668 Sci. USA **112**: E7204–E7212.

669 Li H., Durbin R., 2009 Fast and accurate short read alignment with Burrows-Wheeler transform.
670 Bioinformatics **25**: 1754–60.

671 Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., *et al.* 2009 The Sequence
672 Alignment/Map format and SAMtools. Bioinformatics **25**: 2078–9.

673 Liu J., Shikano T., Leinonen T., Cano J. M., Li M.-H., Merilä J., 2014 Identification of major and
674 minor QTL for ecologically important morphological traits in three-spined sticklebacks
675 (*Gasterosteus aculeatus*). G3 **4**: 595–604.

676 Liu S., Zhu K., Chen N., Wang W., Wan H., 2013 Identification of HIF-1α promoter and expression
677 regulation of HIF-1α gene by LPS and hypoxia in zebrafish. Fish Physiol. Biochem. **39**: 1153-
678 1163

679 Mackay T. F. C., Stone E. A., Ayroles J. F., 2009 The genetics of quantitative traits: challenges and
680 prospects. Nat. Rev. Genet. **10**: 565–77.

681 Mäkinen H. S., Cano J. M., Merilä J., 2006 Genetic relationships among marine and freshwater
682 populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by
683 microsatellites. Mol. Ecol. **15**: 1519–34.

684 Mallarino R., Grant P. R., Grant B. R., Herrel A., Kuo W. P., Abzhanov A., 2011 Two developmental
685 modules establish 3D beak-shape variation in Darwin's finches. Proc. Natl. Acad. Sci. U. S. A.
686 **108**: 4057–62.

687 McCairns R. J. S., Bernatchez L., 2010 Adaptive divergence between freshwater and marine
688 sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of
689 candidate gene expression. Evolution **64**: 1029–47.

690 Mecham B. H., Nelson P. S., Storey J. D., 2010 Supervised normalization of microarrays.
691 Bioinformatics **26**: 1308–15.

692 Meiklejohn C. D., Coolon J. D., Hartl D. L., Wittkopp P. J., 2014 The roles of cis- and trans-
693 regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene
694 expression. Genome Res. **24**: 84–95.

695 Miller C. T., Glazer A. M., Summers B. R., Blackman B. K., Norman A. R., Shapiro M. D., *et al.*
696 2014 Modular skeletal evolution in sticklebacks is controlled by additive and clustered
697 quantitative trait loci. Genetics **197**: 405–20.

698 Morris M. R. J., Richard R., Leder E. H., Barrett R. D. H., Aubin-Horth N., Rogers S. M., 2014 Gene
699 expression plasticity evolves in response to colonization of freshwater lakes in threespine
700 stickleback. Mol. Ecol. **23**: 3226–40.

701 Mudunuri, U., Che, A., Yi, M., Stephens, R. M. 2009. bioDBnet: the biological database network.
702 Bioinformatics **25**:555-6.

703 Nikinmaa M., Rees B. B., 2005 Oxygen-dependent gene expression in fishes. Am. J. Physiol. Regul.
704 Integr. Comp. Physiol. **288**: R1079–90.

705 Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., *et al.* 2004.
706 Control of pancreas and liver gene expression by HNF transcription factors. Science **303**: 1378-
707 1381.

708 Pavey S. A., Collin H., Nosil P., Rogers S. M., 2010 The role of gene expression in ecological
709 speciation. Ann. N. Y. Acad. Sci. **1206**: 110–29.

710 Poelstra J. W., Vijay N., Hoeppner M. P., Wolf J. B. W., 2015 Transcriptomics of colour patterning
711 and colouration shifts in crows. Mol. Ecol. **24**: 4617–28.

712 Powell J. E., Henders A. K., McRae A. F., Kim J., Hemani G., Martin N. G., *et al.* 2013 Congruence
713 of additive and non-additive effects on gene expression estimated from pedigree and SNP data.
714 PLoS Genet. **9**: e1003502.

715 Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A. R., Bender D., *et al.* 2007 PLINK: a
716 tool set for whole-genome association and population-based linkage analyses. Am. J. Hum.
717 Genet. **81**: 559–75.

718 R Core Team 2015. R: a language and environment for statistical computing.Vienna, Austria: R
719 Foundation for Statistical Computing.

720 Ranhotra, H. S. 2015 The NR4A orphan nuclear receptors: mediators in metabolism and diseases.
721 Journal of Receptors and Signal Transduction 35, Iss. 2, 2015

722 Reed R. D., Papa R., Martin A., Hines H. M., Counterman B. A., Pardo-Diaz C., *et al.* 2011 Optix
723 drives the repeated convergent evolution of butterfly wing pattern mimicry. Science **333**: 1137–
724 41.

725 Rissanen E., Tranberg H. K., Sollid J., Nilsson G. E., Nikinmaa M., 2006 Temperature regulates
726 hypoxia-inducible factor-1 (HIF-1) in a poikilothermic vertebrate, crucian carp (*Carassius*
727 *carassius*). J. Exp. Biol. **209**: 994–1003.

728 Roesti M., Hendry A. P., Salzburger W., Berner D., 2012 Genome divergence during evolutionary
729 diversification as revealed in replicate lake-stream stickleback population pairs. Mol. Ecol. **21**:
730 2852–62.

731 Roesti M., Kueng B., Moser D., Berner D., 2015 The genomics of ecological vicariance in threespine
732 stickleback fish. Nat. Commun. **6**: 8767.

733 Roesti M., Moser D., Berner D., 2013 Recombination in the threespine stickleback genome--patterns
734 and consequences. Mol. Ecol. **22**: 3014–27.

735 Rogers S. M., Tamkee P., Summers B., Balabahadra S., Marks M., Kingsley D. M., Schluter D., 2012
736 Genetic signature of adaptive peak shift in threespine stickleback. Evolution **66**: 2439–50.

737 Rytkönen K. T., Vuori K. A. M., Primmer C. R., Nikinmaa M., 2007 Comparison of hypoxia-
738 inducible factor-1 alpha in hypoxia-sensitive and hypoxia-tolerant fish species. Comp. Biochem.
739 Physiol. Part D. Genomics Proteomics **2**: 177–86.

740 Schaefke B., Emerson J. J., Wang T.-Y., Lu M.-Y. J., Hsieh L.-C., Li W.-H., 2013 Inheritance of gene
741 expression level and selective constraints on trans- and cis-regulatory changes in yeast. Mol.
742 Biol. Evol. **30**: 2121–33.

743 Schluter D., McPhail J. D., 1992 Ecological character displacement and speciation in sticklebacks.
744 Am. Nat. **140**: 85–108.

745 Schultheiß R., Viitaniemi H. M., Leder E. H., 2015 Spatial dynamics of evolving dosage
746 compensation in a young sex chromosome system. Genome Biol. Evol. **7**: 581–90.

747 Shimada Y., Shikano T., Merilä J., 2011 A high incidence of selection on physiologically important
748 genes in the three-spined stickleback, *Gasterosteus aculeatus*. Mol. Biol. Evol. **28**: 181–93.

749 Smedley D., Haider S., Ballester B., Holland R., London D., Thorisson G., Kasprzyk A., 2009
750 BioMart--biological queries made easy. BMC Genomics **10**: 22.

751 Steige K. A., Laenen B., Reimegård J., Scofield D., Slotte T., 2015 The impact of natural selection on
752 the distribution of cis-regulatory variation across the genome of an outcrossing plant. BioRxiv.

753 Taylor E. B., McPhail J. D., 2000 Historical contingency and ecological determinism interact to prime
754 speciation in sticklebacks, *Gasterosteus*. Proc. Biol. Sci. **267**: 2375–84.

755 Terekhanova N. V, Logacheva M. D., Penin A. A., Neretina T. V, Barmintseva A. E., Bazykin G. A.,
756 *et al.* 2014 Fast evolution from precast bricks: genomics of young freshwater populations of
757 threespine stickleback *Gasterosteus aculeatus*. PLoS Genet. **10**: e1004696.

758 Turner L. M., White M. A., Tautz D., Payseur B. A., 2014 Genomic networks of hybrid sterility.
759 PLoS Genet. **10**: e1004162.

760 Vance K. W., Ponting C. P., 2014 Transcriptional regulatory functions of nuclear long noncoding
761 RNAs. Trends Genet. **30**: 348–55.

762 Wang S., Zheng T., Wang Y., 2007 Transcription activity hot spot, is it real or an artifact? BMC Proc.
763 **1 Suppl 1**: S94.

764 Whiteley A. R., Derome N., Rogers S. M., St-Cyr J., Laroche J., Labbe A., *et al.* 2008 The phenomics
765 and expression quantitative trait locus mapping of brain transcriptomes regulating adaptive
766 divergence in lake whitefish species pairs (*Coregonus* sp.). Genetics **180**: 147–64.

767 Wickham, H. 2009 *ggplot2: Elegant Graphics for Data Analysis*. Springer New York.

768 Willert K., Jones, K. A. 2006. Wnt signaling: Is the party in the nucleus? Genes Dev **20:**1394–1404

769 Wu C., Delano D. L., Mitro N., Su S. V, Janes J., McClurg P., *et al.* 2008 Gene set enrichment in
770 eQTL data identifies novel annotations and pathway regulators. PLoS Genet. **4**: e1000070.

771 Xu, H., Lu, A., Sharp, F. R. (2011) Regional genome transcriptional response of adult mouse brain to
772 hypoxia. BMC Genomics2011**12**:499

773 Zhang X., Cal A. J., Borevitz J. O., 2011 Genetic architecture of regulatory variation in *Arabidopsis*
774 *thaliana*. Genome Res. **21**: 725–33.

775

776 **Table 1:** Known transcriptional regulators associated with identified eQTL hotspots. Human orthologues of stickleback genes were identified using BioMart.
777 Location is as follows: 'Hotspot': annotated gene is in genomic region of hotspot; '*Cis*': gene is cis-regulated by hotspot at chromosome wide p<0.01;
778 '*Trans*': gene is trans-regulated by hotspot at genome-wide p<0.021.

Hotspot	Location	Stickleback Ensembl_ID	Human Ensembl_ID	Gene Name	Description
Chr06	*Cis*	ENSGACG00000012317	ENSG00000266412	NCOA4	Nuclear receptor coactivator 4
Chr06	*Cis*	ENSGACG00000001371	ENSG00000167380	ZNF226	Zinc finger protein 226
Chr06	Hotspot	ENSGACG00000011981	ENSG00000197223	C1D	C1D nuclear receptor co-repressor
Chr06	*Trans*	ENSGACG00000018659	ENSG00000112983	BRD8	Bromodomain containing 8
Chr06	*Trans*	ENSGACG00000004982	ENSG00000065883	CDK13	Cyclin-dependent kinase 13
Chr06	*Trans*	ENSGACG00000005983	ENSG00000168036	CTNNB1	Catenin (cadherin-associated protein), beta 1, 88kDa
Chr06	*Trans*	ENSGACG00000003088	ENSG00000116580	GON4L	Gon-4-like
Chr06	*Trans*	ENSGACG00000008525	ENSG00000100644	HIF1A	Hypoxia inducible factor 1, alpha subunit
Chr06	*Trans*	ENSGACG00000013704	ENSG00000096968	JAK2	Janus kinase 2
Chr07	*Cis*/Hotspot	ENSGACG00000018669	ENSG00000137462	TLR2	Toll-like receptor 2
Chr07	Hotspot	ENSGACG00000000325	ENSG00000135625	EGR4	Early growth response 4
Chr07	Hotspot	ENSGACG00000018606	ENSG00000109670	FBXW7	F-box And WD repeat domain containing 7, E3 ubiquitin protein ligase
Chr07	Hotspot	ENSGACG00000000304	ENSG00000170448	NFXL1	Nuclear transcription factor, X-box binding-like 1
Chr07	Hotspot	ENSGACG00000000370	ENSG00000164985	PSIP1	PC4 and SFRS1 interacting protein 1
Chr07	Hotspot	ENSGACG00000018586	ENSG00000074966	TXK	Tyrosine kinase

Chr07	*Trans*	ENSGACG00000000333	ENSG00000173801	JUP	Junction plakoglobin
Chr08	Hotspot	ENSGACG00000014457	ENSG00000162733	DDR2	Discoidin domain receptor tyrosine kinase 2
Chr08	Hotspot	ENSGACG00000014404	ENSG00000187764	SEMA4D	Sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (Semaphorin) 4D
Chr08	*Trans*	ENSGACG00000006033	ENSG00000125686	MED1	Mediator complex subunit 1
Chr08	*Trans*	ENSGACG00000017475	ENSG00000137699	TRIM29	tripartite motif containing 29
Chr09	Hotspot	ENSGACG00000019898	ENSG00000162961	DPY30	Dpy-30 histone methyltransferase complex regulatory subunit
Chr09	Hotspot	ENSGACG00000019915	ENSG00000132664	POLR3F	Polymerase (RNA) III (DNA directed) polypeptide F, 39 KDa
Chr09	Hotspot	ENSGACG00000020002	ENSG00000112658	SRF	Serum response factor
Chr12a	*Cis*	ENSGACG00000000816	ENSG00000126767	ELK1	ELK1, member of ETS oncogene family
Chr12a	Hotspot	ENSGACG00000000295	ENSG00000146109	ABT1	Activator of basal transcription 1
Chr12a	Hotspot	ENSGACG00000000248	ENSG00000106785	TRIM14	Tripartite motif containing 14
Chr12a	*Trans*	ENSGACG00000019625	ENSG00000164134	NAA15	N(Alpha)-acetyltransferase 15, NatA auxiliary subunit
Chr12a	*Trans*	ENSGACG00000001088	ENSG00000111581	NUP107	Nucleoporin 107kDa
Chr12b	Hotspot	ENSGACG00000011155	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
Chr12b	Hotspot	ENSGACG00000010943	ENSG00000110925	CSRNP2	Cysteine-serine-rich nuclear protein 2
Chr12b	Hotspot	ENSGACG00000011240	ENSG00000163349	HIPK1	Homeodomain interacting protein kinase 1
Chr12b	Hotspot	ENSGACG00000011086	ENSG00000101096	NFATC2IP	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2
Chr12b	Hotspot	ENSGACG00000010788	ENSG00000123358	NR4A1	Nuclear receptor subfamily 4, group A, member 1

Chr12b	Hotspot	ENSGACG00000010925	ENSG00000184271	POU6F1	POU class 6 homeobox 1
Chr12b	Hotspot	ENSGACG00000010990	ENSG00000079337	RAPGEF3	Rap guanine nucleotide exchange factor (GEF) 3
Chr12b	Hotspot	ENSGACG00000010838	ENSG00000181852	RNF41	Ring finger protein 41, E3 ubiquitin protein ligase
Chr12b	Hotspot	ENSGACG00000010929	ENSG00000135457	TFCP2	Transcription factor CP2
Chr12b	Hotspot	ENSGACG00000011135	ENSG00000182463	TSHZ2	Teashirt zinc finger homeobox 2
Chr12b	Hotspot	ENSGACG00000011187	ENSG00000204859	ZBTB48	Zinc finger and BTB domain containing 48
Chr12b	Hotspot	ENSGACG00000011128	ENSG00000020256	ZFP64	Zinc finger protein 64
Chr12b	Hotspot	ENSGACG00000011124	ENSG00000101115	SALL4	Spalt-like transcription factor 4
Chr12b	*Trans*	ENSGACG00000006074	ENSG00000185513	L3MBTL1	L(3)mbt-like
Chr12b	*Trans*	ENSGACG00000011682	ENSG00000162761	LIMX1A	LIM homeobox transcription factor 1, alpha
Chr12b	*Trans*	ENSGACG00000004938	ENSG00000012504	NR1H4	Nuclear receptor subfamily 1, group h, member 4
Chr12c	*Cis*/Hotspot	ENSGACG00000004839	ENSG00000188157	AGRN	Agrin
Chr12c	Hotspot	ENSGACG00000004256	ENSG00000101126	ADNP	Activity-dependent neuroprotector homeobox
Chr12c	Hotspot	ENSGACG00000004544	ENSG00000009307	CSDE1	Cold shock domain containing E1, RNA-binding
Chr12c	Hotspot	ENSGACG00000004732	ENSG00000101412	E2F1	E2F transcription factor 1
Chr12c	Hotspot	ENSGACG00000004740	ENSG00000078747	ITCH	Itchy E3 ubiquitin protein ligase
Chr12c	Hotspot	ENSGACG00000004213	ENSG00000197780	TAF13	TAF13 RNA Polymerase II, TATA box binding protein (TBP)-associated factor, 18kDa
Chr12c	Hotspot	ENSGACG00000004773	ENSG00000122691	TWIST2	Twist homolog 2
Chr12c	Hotspot	ENSGACG00000004763	ENSG00000111424	VDR	Vitamin D (1,25- dihydroxyvitamin D3) receptor
Chr12c	Hotspot	ENSGACG00000004662	ENSG00000197114	ZGPAT	Zinc finger, CCCH-type with G patch domain

Chr12c	Hotspot	ENSGACG00000004734	ENSG00000131061	ZNF341	Zinc finger protein 341
Chr12c	*Trans*	ENSGACG00000017068	ENSG00000104221	BRF2	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
Chr16	*Trans*	ENSGACG00000012487	ENSG00000125845	BMP2	Bone morphogenetic protein 2
Chr16	*Trans*	ENSGACG00000005831	ENSG00000153234	NR4A2	Nuclear receptor subfamily 4, group A, member 2
Chr18	Hotspot	ENSGACG00000012487	ENSG00000125845	BMP2	Bone morphogenetic protein 2
Chr18	Hotspot	ENSGACG00000012415	ENSG00000125812	GZF1	GDNF-inducible zinc finger protein 1
Chr18	Hotspot	ENSGACG00000012595	ENSG00000100811	YY1	YY1 transcription factor
Chr18	Hotspot	ENSGACG00000012744	ENSG00000165588	OTX2	Orthodenticle homeobox 2
Chr18	*Trans*	ENSGACG00000016702	ENSG00000103449	SALL1	Spalt-like transcription factor 1

779

780 **Figure Legend**

781 **Figure 1** Position of SNP markers along each chromosome (top) and location of *trans* eQTL hits for
782 all assayed genes (bottom). Black bars show the number of eQTL hits at each 1cM Kosambi interval
783 along the chromosome. Blue shading shows the number of eQTL with 95% confidence intervals
784 overlapping each 1cM interval. Arrows indicate the location of eight significant *trans* eQTL hotspots.
785 Figure created using ggplot2 (Wickham 2009) in R.

786 **Figure 2:** Networks of known protein-protein interactions inferred by String 10 for proteins
787 associated with a) Chr6 hotspot and b) Chr12b hotspot. 'Upstream Regulator': significantly enriched
788 upstream regulator identified when examining genes *trans*-regulated by the hotspot using IPA;
789 'Hotspot Location': protein is coded by a gene physically located in the hotspot; 'Trans regulated':
790 protein is *trans* regulated by an eQTL mapping to the hotspot and significant at genome-wide
791 $p<0.021$; Cis/Hotspot: both present in and significantly *cis* regulated by the hotspot. Interactions not
792 involving an identified upstream regulator are not shown.

793

Figure 1

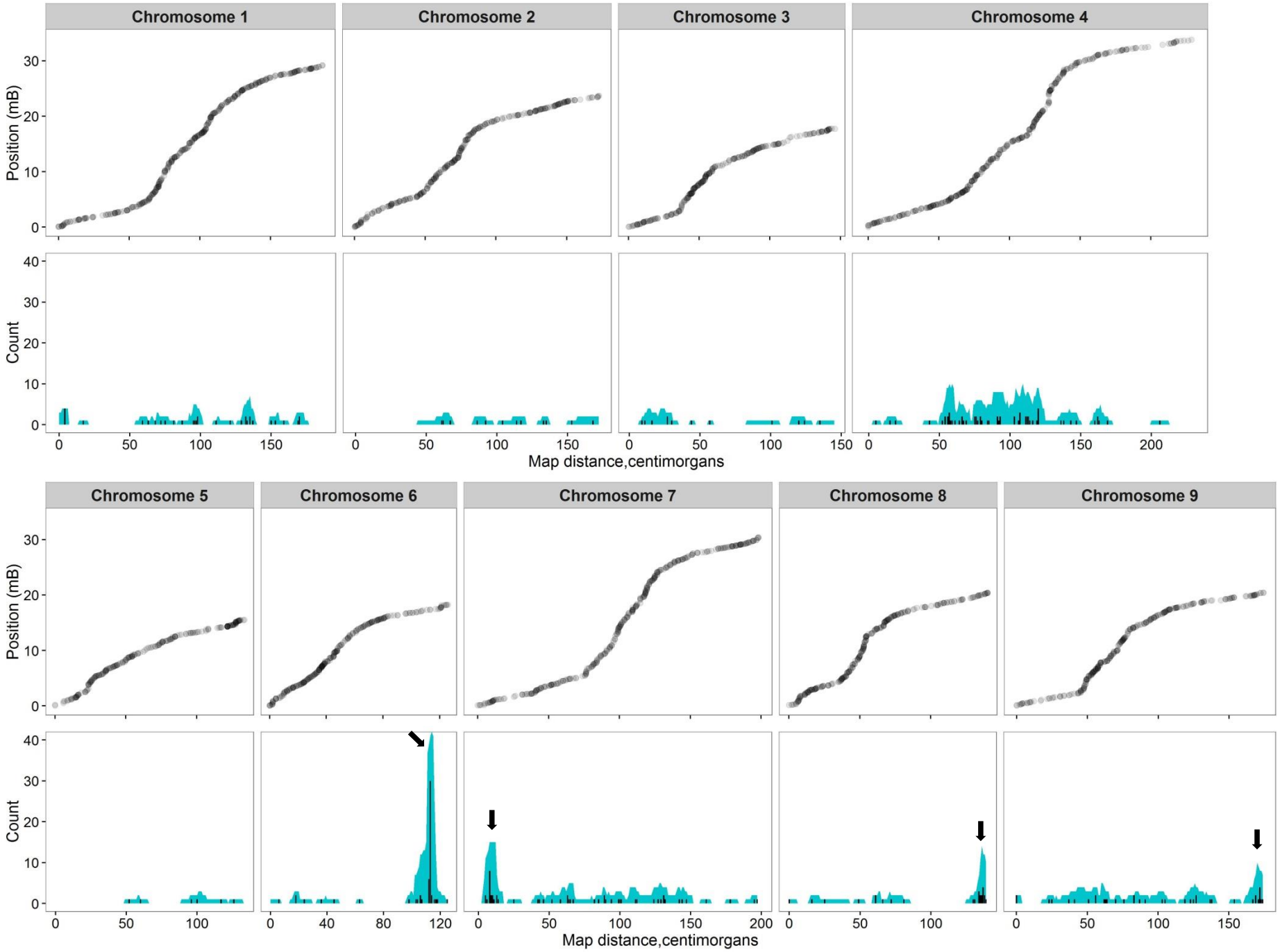
Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Position (mB)
Count
Map distance,centimorgans
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Chromosome 9
Position (mB)
Count
Map distance,centimorgans

Figure 1 (cont.)

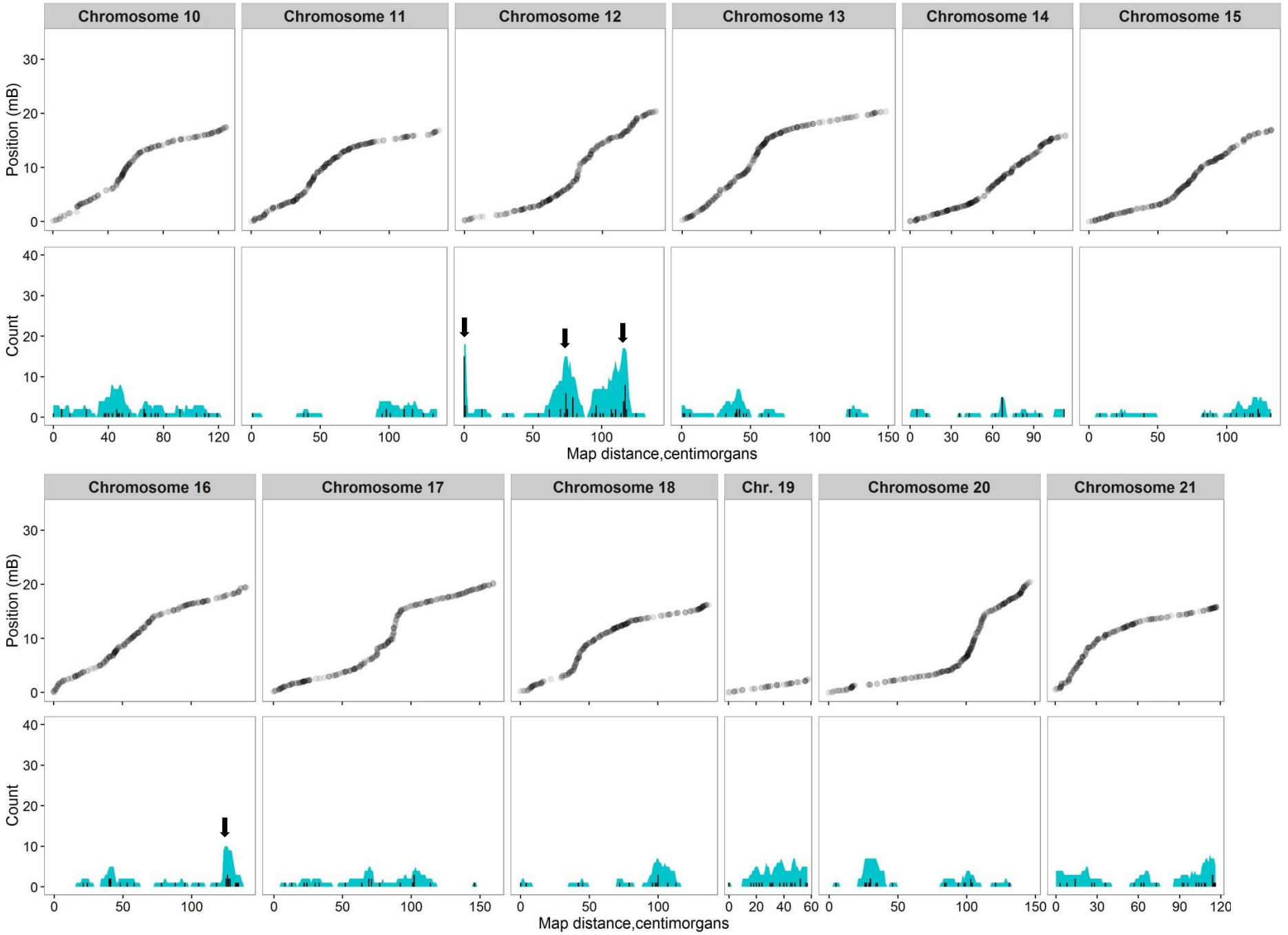

Figure 2a: Chr6

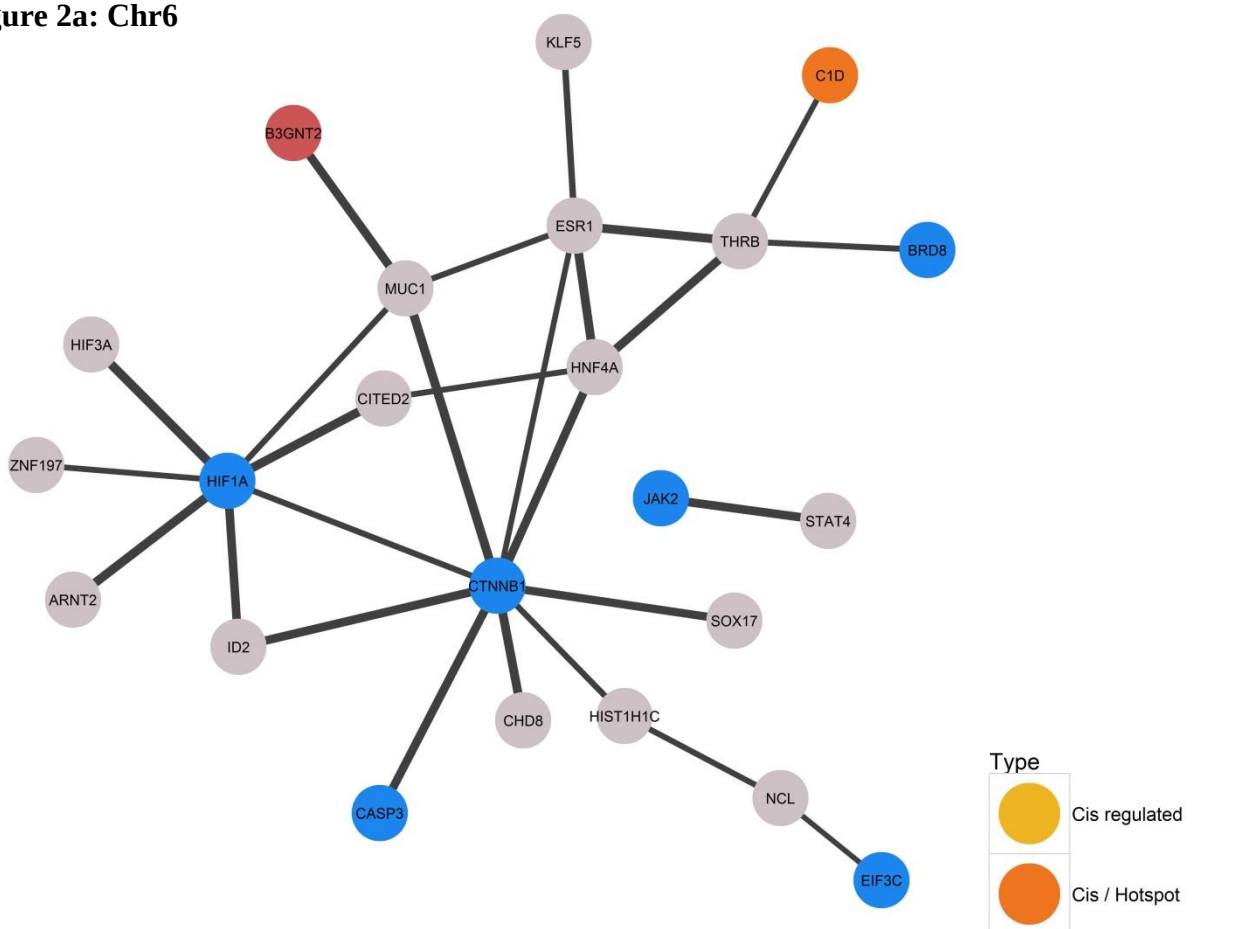

Figure 2b: Chr12b