Genome-Wide Genotype-Expression Relationships Reveal Both Copy Number and Single Nucleotide Differentiation Contribute to Differential Gene Expression between Stickleback Ecotypes

Yun Huang1,2,*, Philine G.D. Feulner3,4, Christophe Eizaguirre5, Tobias L. Lenz1, Erich Bornberg-Bauer 6, Manfred Milinski1, Thorsten B.H. Reusch7, and Frédéric J.J. Chain8,*

1Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Plön, Germany
2Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, ROC
3Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland
4Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Switzerland
5School of Biological and Chemical Sciences, Queen Mary University of London, United Kingdom
6Evolutionary Bioinformatics, Institute for Evolution and Biodiversity, Westfälische Wilhelms University, Münster, Germany
7Marine Evolutionary Ecology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany
8Department of Biological Sciences, University of Massachusetts Lowell, USA

*Corresponding authors: E-mails: yunhuang@gate.sinica.edu.tw; frederic_chain@uml.edu.
Accepted: July 10, 2019

Abstract
Repeated and independent emergence of trait divergence that matches habitat differences is a sign of parallel evolution by natural selection. Yet, the molecular underpinnings that are targeted by adaptive evolution often remain elusive. We investigate this question by combining genome-wide analyses of copy number variants (CNVs), single nucleotide polymorphisms (SNPs), and gene expression across four pairs of lake and river populations of the three-spined stickleback (Gasterosteus aculeatus). We tested whether CNVs that span entire genes and SNPs occurring in putative cis-regulatory regions contribute to gene expression differences between sticklebacks from lake and river origins. We found 135 gene CNVs that showed a significant positive association between gene copy number and gene expression, suggesting that CNVs result in dosage effects that can fuel phenotypic variation and serve as substrates for habitat-specific selection. Copy number differentiation between lake and river sticklebacks also contributed to expression differences of two immune-related genes in immune tissues, cathepsin A and GIMAP7. In addition, we identified SNPs in cis-regulatory regions (eSNPs) associated with the expression of 1,865 genes, including one eSNP upstream of a carboxypeptidase gene where both the SNP alleles differentiated and the gene was differentially expressed between lake and river populations. Our study highlights two types of mutations as important sources of genetic variation involved in the evolution of gene expression and in potentially facilitating repeated adaptation to novel environments.

Key words: habitat-specific adaptation, CNV, copy number variation, eSNP, cis-regulatory regions, expression differentiation, three-spined stickleback.

Introduction
Uncovering the genetic mechanisms underlying adaptive evolution is a major research focus in evolutionary biology (Barrett and Hoekstra 2011). Adaptive phenotypes can result from changes in amino acid sequences that affect protein structure and function (Hoekstra and Coyne 2007), as well as from alterations of gene expression patterns (Carroll 2008). Although gene expression can be plastic and respond to environmental stimuli (Gibson 2008), adaptive evolution of gene expression rests upon an inherited genetic basis.
Gene expression differences between populations and species often carry a significant heritable component and impact fitness, contributing to adaptation (Stamatoyannopoulos 2004; Whitehead and Crawford 2006b; Pavey et al. 2010). A growing body of evidence has linked the acquisition of adaptive phenotypes in new environments to gene expression changes, including elongated beaks in cactus finches (Abzhanov et al. 2006), camouflage pigmentation in deer mice (Linnen et al. 2009; Mallarino et al. 2017), convergent thick lips in cichlids (Colombo et al. 2013), and repeated pelvic loss in three-spined sticklebacks (Chan et al. 2010). If the differentiation in expression confers an adaptive advantage across independent population clines, it may lead to parallel evolution at the gene expression level. The parallel evolution of expression patterns can be directly inferred when, for example, heritable gene expression variation correlates with an environmental cline rather than by ancestry (Whitehead and Crawford 2006a; Lenz 2015). Parallel gene expression has been observed in a few cases of diverging ecotypes or species of adaptive radiations (Derome et al. 2006; Pavey et al. 2010; Colombo et al. 2013; Manousaki et al. 2013; Stutz et al. 2015; Zhao et al. 2015; Hanson et al. 2017). Yet, the genetic variants associated with these gene expression patterns remain understudied.

Genomic studies of recurring ecotypes have revealed a major contribution of regulatory elements to parallel genomic divergence (Jones et al. 2012; Brawand et al. 2014). Combining gene expression surveys with genome-wide sequence analysis allows evaluating the role of genetic variants on the evolution of expression differences between ecotypes. The genetic basis of expression differences may reside in close physical proximity of a gene (in cis) or far away (in trans) (Gilad et al. 2008). Genetic mutations altering the sequence of cis-regulatory elements can affect the binding affinity of transcription factors, whose effects are mainly limited to expression variation levels of neighboring genes, whereas mutations that affect trans-regulatory elements typically encode transcription factors that regulate multiple downstream genes (Wittkopp and Kalay 2011). Due to the local effects of cis-regulatory elements that confer a lower extent of pleiotropy compared with trans-, cis-regulatory elements have been suggested to be more important than trans-regulatory elements in the expression divergence between species (Wittkopp et al. 2008).

In addition to sequence changes in its regulatory region, the number of copies of a particular gene can affect its expression. Gene copy number can differ among individuals due to genetic deletions and duplications, giving rise to copy number variations (CNVs), which natural selection can act upon (Nguyen et al. 2006; Katju and Bergthorsson 2013). Copy number is generally positively correlated with expression levels (Haraksingh and Snyder 2013; Gamazon and Stranger 2015), producing a gene dosage effect (Zhang et al. 2009). Gene dosage effects are often detrimental to fitness as they can disrupt the stoichiometric balance in molecular networks (Papp et al. 2003; Veitia 2005; Birchler and Veitia 2012) and have been associated with diseases (Rice and McLysaght 2017). However, in some cases, dosage effects of CNVs have also been beneficial, such as the relationship observed between amylase gene copy numbers and starch diets in both humans and dogs (Perry et al. 2007; Axelsson et al. 2013), and the number of cytochrome P450 genes in insecticide-resistant populations of dengue mosquitoes (Faucon et al. 2015). Although variation in cis-regulatory elements and CNVs can both affect gene expression and contribute to adaptive phenotypes, their contribution to habitat-specific gene expression has not been systematically studied. Genotype-expression relationships become particularly interesting when divergence patterns across replicated populations independently adapted to different environments occur at both the genetic and expression levels, strongly suggesting a genetic basis underlying adaptive expression variation.

The three-spined stickleback (Gasterosteus aculeatus) is a powerful model species to investigate habitat-specific adaptation. After the last glaciation, marine three-spined sticklebacks repeatedly colonized different freshwater habitats, resulting in an adaptive radiation composed of habitat-specific ecotypes (McKinnon and Rundle 2002). In particular, recurrent adaptation to lakes and rivers (or streams) has given rise to distinct ecotypes across the northern hemisphere (Reusch et al. 2001), with morphological differences in body shapes and traits involved in foraging (Berner et al. 2008; Deagle et al. 2012; Kaeuffer et al. 2012; Ravinet et al. 2013; Lucek et al. 2014). Another profound difference between lake and river habitats is the distinct parasite community, in which lake fish generally suffer from a higher parasite burden than river fish, likely contributing to recurrent expression differences at both the phenotypic and genetic level (Kalbe et al. 2002; Eizaguirre et al. 2011; Feulner et al. 2015). Transcriptome analyses have revealed over a hundred genes with habitat-specific gene expression among wild-caught lake and river sticklebacks (Huang et al. 2016), some of which were also differentially expressed between lake and river sticklebacks in a laboratory-controlled parasite infection experiment (Lenz et al. 2013). Lake and stream sticklebacks raised in common garden conditions also exhibit parallel gene expression differences (Hanson et al. 2017). These results suggest a heritable component to habitat-specific gene expression, which is also supported by quantitative genetics analyses on pedigrees of sticklebacks (Leder et al. 2015). In sticklebacks, a greater contribution of cis-regulatory elements than trans-regulatory elements in expression variation and divergence between ecotypes has been suggested (Ishikawa et al. 2017; Pritchard et al. 2017; Verta and Jones 2019). However, unlike the parallel divergence observed between marine and freshwater sticklebacks at the sequence level (Jones et al. 2012) and in gene CNVs (Hirase et al. 2014), a
low degree of parallel genetic differentiation exists among repeatedly diverged lake and river ecotypes, both at the sequence level (Deagle et al. 2012; Roesti et al. 2012; Feulner et al. 2015; Stuart et al. 2017) and in copy numbers (Chain et al. 2014). This is despite habitat-specific patterns of gene expression (Huang et al. 2016; Hanson et al. 2017). Given the low degree of genomic parallelism, the genetic variation underlying the expression divergence between lake and river ecotypes remains elusive.

In this study, stickleback genomes and transcriptomes from the exact same individuals were used to study the molecular basis of habitat-specific adaptations between lake and river ecotypes. To identify candidate genes involved in adaptation to distinct parasite communities in lakes and rivers, we evaluated the relationships between gene expression variation in immune tissues and two types of variants, gene CNVs and single nucleotide polymorphisms (SNPs) in cis-regulatory regions. We tested for 1) associations between gene copy numbers or SNP genotypes and gene expression within and across individuals, 2) evidence of habitat-specific selection as inferred from different gene copy numbers between ecotypes or allele frequency differentiation of SNPs, and 3) differential gene expression between ecotypes. These serve as three pillars of evidence that genetic changes contribute to adaptive gene expression differences between ecotypes. In this way, we identified genetic variants that influence repeated differential expression between ecotypes, putatively contributing to habitat-specific adaptation.

Materials and Methods

Sampling Design
To study the genetic differentiation between lake and river stickleback ecotypes that underlie expression differentiation, we combined a whole genome and a whole transcriptome data set from a total of eight geographically widespread populations of three-spined sticklebacks that had been previously analyzed separately. The whole-genome sequence data set consisted of 48 fish from four parapatric population pairs; two independent drainages from Germany (G1 and G2), one from Norway (No), and one from Canada (Ca), with six individuals from each lake (L) and each river (R), respectively (Chain et al. 2014; Feulner et al. 2015; EBI Accession no: PRJEB5198; supplementary fig. S1 and table S1, Supplementary Material online). The final set of expression profiles consisted of 12,105 protein-coding genes remained, referred herein as “gene expressions”.

Identification of Gene eCNVs
CNV regions of the study populations were identified by Chain et al. (2014), where CNVs were assigned using consensus calls from the read depth approach implemented in the software CNVnator (Abyzov et al. 2011) and at least one other approach (paired-end and split-reads; for details see Chain et al. 2014). We identified genes with at least 95% length overlap with CNVs. Gene copy number was estimated using CNVnator and rounded to the closest integer. Genes showing no variation in estimated copy numbers among individuals of our study were excluded from copy number analyses. Genes with copy number estimates of zero but with detectable read depth >0 were removed from our analyses to avoid possible false deletion calls. A total of 832 autosomal protein-coding genes remained, referred herein as “gene CNVs.”

Using gene copy numbers and the corresponding gene expression from the same fish, we evaluated the association between gene copy number and expression level for each gene CNV in each individual, and for each tissue type separately. Using a linear mixed effect model, gene copy number was set as a fixed effect, and the population and sex were set as random effects (expression levels ~copy_number + (1|population) + (1|sex)). This approach makes use of the continuous nature of copy number genotypes and tests for dosage effects of CNVs, which is different from the approach of assessing copy number changes as discrete categories.
from the typical eQTL approach that associates expression variation to categorical genotypes. Benjamini–Hochberg’s multiple-test correction was applied to the $P$ values of the fixed effect of copy number (Benjamini and Hochberg 1995). Genes with corrected $P$ values <0.05 and positive correlation were considered as “gene eCNVs,” having statistically significant correlations between copy number and expression.

Identification of eSNPs

In addition to the evaluation of gene eCNVs, we mapped SNPs in cis-regulatory regions (eSNPs) to identify potential cis-regulatory elements that underlie gene expression variation. The eSNPs were determined for gene expression in head kidney and spleen separately, using SNPs within a 5 kb range of the transcription start sites (TSSs). We reasoned that the 5 kb upstream regions serve as a proxy for the location of potential cis-regulatory elements, based on empirical findings of cis-regulatory sequences in mouse (Shen et al. 2012). SNPs used in this study were extracted from a previous genome-wide survey (Feulner et al. 2015), excluding SNPs in CNV regions due to potential detection biases (Hartasánchez et al. 2018) and filtering SNPs for a minor allele frequency >0.05 in the four population pairs combined. Out of 12,105 and 12,451 genes expressed in the head kidney and in the spleen, respectively, 10,803 and 10,914 genes had a total of 815,341 and 841,063 SNPs, and jointly 870,917 SNPs that fulfilled our filtering criteria. For each expressed gene, we tested for a significant association between each SNP and expression levels in FastQTL v2.165 (Ongen et al. 2016) using the nominal pass and correcting for population stratification (population pairs and habitats) and sex. Two steps of multiple-testing correction (Benjamini–Hochberg) were applied on the $P$ value for each SNP: the $P$ values were first corrected for numbers of SNPs per gene and then for the total number of genes tested. SNPs with corrected $P$ values <0.05 were considered as eSNPs.

Expression Differentiation between Stickleback Ecotypes

Differential expression (DE) analyses implemented in the package EdgeR was previously used to identify significantly differentially expressed genes between ecotypes, indicative of habitat-specific gene expression (Huang et al. 2016). To complement this binary categorization, we quantified the extent of expression differentiation in a continuous manner by computing the variable $P_{CT}$, which evaluates the relative variance in expression between groups (here lake vs. river ecotypes) compared with the variance within groups. We calculated $P_{CT}$ between lake and river sticklebacks and accounted for expression variances among geographic population pairs and between sex using an ANOVA-based approach (methods adapted from Uebbing et al. 2016). $P_{CT}$ as a measure of relative differentiation in gene expression between lake and river ecotypes was calculated for each expressed gene and for the head kidney and spleen separately. Because the calculation of $P_{CT}$ is conceptually equivalent to the calculation of copy number differentiation ($V_{CT}$, see below) and nucleotide differentiation ($F_{CT}$, see below), the evaluation of expression differentiation is made directly comparable to that of genetic differentiation. To determine how likely each $P_{CT}$ value was obtained by chance, we recalculated $P_{CT}$ 1000 times for each gene after random permutations of the ecotype labels. The $P$ values were calculated as the fraction of permuted values that exceeded the observed value and were corrected by the Benjamini-Hochberg method for the numbers of genes tested (Benjamini and Hochberg 1995). Genes with adjusted $P$ values <0.05 for $P_{CT}$ were considered significant. For candidate genes, we also calculated $P_{CT}$ between each population pair separately, in the same way that $P_{CT}$ was calculated for all populations combined but without population stratification in the ANOVA model.

Copy Number Differentiation of Gene eCNVs

In order to investigate the contribution of gene eCNVs in expression differentiation, we evaluated copy number differentiation between ecotypes across all population pairs together. For each gene eCNV, we calculated $V_{CT}$ representing the relative variance in copy number between groups (here lake vs. river ecotypes) compared with the overall variance within groups, similarly to $P_{CT}$. $V_{CT}$ was calculated using all individuals from the four population pairs with an ANOVA-based approach, where lake and river ecotypes were treated as two comparison groups, while accounting for variance between population pairs (copy_number ~ ecotypes * population_pair). As we exclude CNVs in the sex chromosome for our analyses, we did not include sex as a factor in the model. $V_{CT}$ is different from $V_{ST}$, a measure of copy number differentiation between populations without a nested structure (Redon et al. 2006), which was previously calculated on the same data set but between each lake and river pair separately in Chain et al. (2014). Including all population pairs together to estimate copy number differentiation ($V_{CT}$) increases sensitivity to detect differences between ecotypes, as it does not require differentiation signals to be extreme in each pair. We applied 1000 permutations following the methods for $P_{CT}$ to identify gene eCNVs with significant $V_{CT}$, and $P$ values were corrected by the Benjamini-Hochberg method for the number of genes tested (Benjamini and Hochberg 1995). $V_{CT}$ with corrected $P$ values <0.05 were considered significantly differentiated between lake and river ecotypes. For candidate gene eCNVs, we also calculated $V_{CT}$ between each population pair separately, in the same way as $V_{CT}$ was calculated for all populations combined but without population stratification in the ANOVA model.
Allelic Differentiation of eSNPs

In addition to the evaluation of copy number differentiation, we calculated nucleotide differentiation between lake and river ecotypes for each SNP identified as eSNPs, evaluated as \( F_{CT} \) using the locus-by-locus AMOVA approach implemented in Arlequin (Excoffier and Lischer 2010). The \( F_{CT} \) was calculated as the percentage of variance between groups (lake vs. river ecotypes) relative to the total variance, using a hierarchical structure that groups lake and river ecotypes into four populations each. This AMOVA approach provides a more sensitive way to qualitatively evaluate habitat-specific patterns across replicated population pairs, compared with methods that scan for outlier regions in each population pair separately to identify parallel regions based on shared outliers (e.g., Feulner et al. 2015), for the same reason as mentioned above for \( V_{CT} \). We used permutation tests implemented in Arlequin to determine the significance of the \( F_{CT} \) values and identify eSNPs with significant \( F_{CT} \) values (\( P < 0.05 \) from 1,023 permutations).

Identifying Correlations between Expression and Genetic Differentiation

A genome-wide correlation between gene expression differentiation (\( P_{CT} \)) and genetic differentiation (\( V_{CT} \) and \( F_{CT} \)) was performed on all expressed genes. For this analysis, \( V_{CT} \) was calculated for each of 350 gene CNVs that had expression (not gene eCNVs), and \( F_{CT} \) was calculated for each of 11,935 autosomal protein-coding genes that had expression (not only for eSNPs), excluding genes in CNV regions. \( F_{CT} \) was evaluated for each gene based on SNPs in the 5 kb upstream regions, using the AMOVA approach implemented in Arlequin (Excoffier and Lischer 2010). With the resulting matrices of \( P_{CT} \), \( V_{CT} \), and \( F_{CT} \) of all genes expressed in the head kidney and/or spleen, the Spearman’s rank correlation was used to test for correlation in each tissue 1) between \( P_{CT} \) and \( V_{CT} \) and 2) between \( P_{CT} \) and \( F_{CT} \). All statistical analyses were carried out using the package R version 3.0.2 (R Development Core Team 2011) unless otherwise indicated.

Testing for Gene Ontology Enrichment in Genes with eSNPs and eCNVs

We tested for enrichment of gene ontology (GO) terms among the gene eCNVs, the genes with eSNPs, the gene eCNVs with significant \( V_{CT} \), and the genes with eSNPs with significant \( F_{CT} \). The enrichment tests were conducted with topGO (Alexa et al. 2006), based on Fisher’s exact tests applying Benjamini–Hochberg’s multiple-test correction. We used different background gene sets depending on the enrichment analysis: we compared gene eCNVs to all expressed genes in either tissue and to all gene CNVs that are expressed in either tissue; we compared genes with eSNPs to all genes that were included in the eSNP tests; we compared gene eCNVs with significant \( V_{CT} \) to the set of gene eCNVs; we compared genes with eSNPs with significant \( F_{CT} \) to all genes with eSNPs. Overrepresented GO terms were those with corrected \( P \) values < 0.05.

**Results**

We first evaluated genotype-expression relationships using CNVs and SNPs, and then investigated whether they contribute to expression divergence between ecotypes. Our overarching goal was to evaluate the relationship between genetic differentiation of the two variant types and gene expression differentiation between replicated pairs of lake and river three-spined stickleback ecotypes.

Gene Copy Numbers and Expression Levels Are Largely Positively Correlated

Out of a total of 19,782 protein-coding autosomal genes in stickleback genome, we identified 832 gene CNVs among our samples. Among these gene CNVs, 350 CNVs had available gene expression data, out of which 140 (40%) had a significant association between gene copy numbers and gene expression in at least one of the two immune tissues (corrected \( P \) values < 0.05). Five of these genes had a significant negative correlation between copy numbers and expression levels: WW domain binding protein 1 (WW domain binding protein 1, ENSGACG00000000318), slc47a1 (solute carrier family 47, member 1, ENSGACG00000020614) and two uncharacterized genes (ENSGACG00000020469 and ENSGACG00000012806) in head kidney samples, as well as cytochrome P450 family 3 subfamily A member 43 (cytochrome P450 family 3 subfamily A member 43, ENSGACG00000010952) in spleen samples. The other 135 genes (39% of all expressed gene CNVs) had a positive correlation in at least one of the two tissues and were considered “gene eCNVs” (fig. 1 and supplementary table S2, Supplementary Material online). Among these 135 gene eCNVs, 13 were only expressed and had a positive correlation in one tissue (5 in head kidney and 8 in spleen), 62 were expressed in both tissues but correlated with copy number in one tissue, while 60 were expressed and correlated in both tissues. Among the genes that were expressed in either the head kidney or spleen tissues, gene eCNVs were enriched for antigen processing and presentation (GO: 0019882, with 4 out of 28 genes), immune response (GO: 0006955, with 5 out of 72 genes), major histocompatibility complex (MHC) protein complex (GO: 0042611 with 4 of 27 genes), and MHC class I protein complex (GO: 0042612 with 4 of 18 genes). MHC immune genes were among functional categories that were previously reported as enriched among all gene CNVs in sticklebacks (Chain et al. 2014). When comparing gene eCNVs against all gene CNVs that were expressed in either tissue, there was no GO term enrichment observed.
Ten eCNVs Show Copy Number Differentiation between Ecotypes

As gene eCNVs are the putative genetic variants that affect gene expression, we evaluated differentiation in their gene copy numbers between ecotypes, which could contribute to gene expression divergence. We estimated $V_{CT}$ for each gene eCNV, which is the relative variance in gene copy numbers between ecotypes compared with the variance within ecotypes. Out of the total 135 gene eCNVs, 10 (7.4%) have a significant $V_{CT}$ ($FDR<0.05$, permutation test), with $V_{CT}$ values ranging from 0.144 to 0.578 (table 1). Of these ten genes, seven have higher average copy numbers in lake ecotypes than in river ecotypes, and three have higher copy numbers in river ecotypes. The 10 gene eCNVs with significant $V_{CT}$ are distributed across 6 of 20 stickleback autosomes (fig. 2a). The GO annotations of the ten $V_{CT}$ significant genes show that they are associated with various functions including ion binding, GTP binding, peptidase activity, diphosphatase activity, and transmembrane transport (table 1). But there was no functional enrichment of the ten gene eCNVs with significant $V_{CT}$ compared with all gene eCNVs.

An Abundance of Genes with SNPs in Cis Is Associated with Expression

In addition to the CNVs associated with gene expression, we also investigated SNPs that are associated with gene expression. Out of a total of 870,917 SNPs within 5kb range of the TSSs of 11,360 genes expressed in either tissues, 8,353 SNPs were found associated with expression of 1,351 genes in the head kidney, 4,261 SNPs associated with expression of 746 genes in the spleen, including 1,336 SNPs associated with expression of 232 genes in both tissue types (corrected $P$ values $<0.05$, supplementary table S3, Supplementary Material online). In total, 11,278 SNPs associated with 1,865 genes were determined as eSNPs that putatively contribute to gene expression differences among individuals. These eSNPs are symmetrically distributed across the 5 kb upstream and downstream range, with a slight peak within the 1kb range of the TSSs (supplementary table S3, Supplementary Material online). No GO term was enriched for the genes with eSNPs when compared with the joint set of 11,360 genes tested in the eSNPs analyses.

Fourteen eSNPs Show Allelic Differentiation between Ecotypes

For each eSNP, we evaluated the nucleotide differentiation, $F_{CT}$, between lake and river ecotypes. We found that 90.9% of eSNPs had negative or zero $F_{CT}$ values, indicating no differentiation between lake and river fish populations. Out of the 1,112 eSNPs with a positive $F_{CT}$, 14 were significantly differentiated ($P<0.05$, permutation test), with $F_{CT}$ values ranging from 0.120 to 0.378 (fig. 2a). These 14 eSNPs were associated with expression of 14 different genes. These 14 genes are annotated with various functions spanning mRNA splicing, DNA binding, rRNA methylation, signal transduction, ATP binding, and GTP binding (table 2), with no significant enrichment of GO categories compared with the set of genes with eSNPs.

One eSNP and Two eCNVs Display Expression Differentiation between Ecotypes

The eSNPs and the gene eCNVs that are differentiated between ecotypes putatively contribute to expression differentiation. Among 12,105 genes expressed in the head kidney and 12,451 genes in the spleen, we identified 115 and 88 genes with significant $P_{CT}$, respectively (corrected $P<0.05$, supplementary table S4, Supplementary Material online). Out of these genes, we found one gene with significant $P_{CT}$ (0.217) in the
head kidney that also had an eSNP with significant $F_{CT}$ (fig. 2). The $P_{CT}$ in the spleen was 0.142 (corrected $P_{CT} = 0.11$). The gene is dehydrogenase/reductase (SDR family) member 13a, duplicate 3 (dhrs13a.3, ENSGACG00000013614), a carboxypeptidase that catalyzes hydrolysis of peptide bonds (Uniprot entry: G3PTQ4). The SNP residing 630 bp upstream of the TSS of this gene had a $F_{CT}$ value of 0.204, and was significantly associated with gene expression in both tissues. We also found two genes with significant $P_{CT}$ that exhibited both differentiation in copy numbers (significant $V_{CT}$) and significant correlations between gene copy numbers and gene expression (gene eCNVs) in both tissues. The gene cathepsin A (ENSGACG00000015897) had significant $P_{CT}$ in spleen (0.289; $P_{CT}$ of 0.159 in head kidney) and the highest $V_{CT}$ (0.578) among all gene CNVs (fig. 2). The other gene, GTPase, IMAP family member 7 (GIMAP7, ENSGACG00000018877), had significant $P_{CT}$ identified in head kidney (0.245; $P_{CT}$ of 0.184 in spleen) and a $V_{CT}$ of 0.348 (fig. 2).

eSNP Regulating Expression Differentiation in dhrs13a.3

Examining the differentiation signals within each population pair, the gene dhrs13a.3 had higher expression levels in a subset of lake populations: in the head kidney of G1 ($P_{CT} = 0.648$) and G2 ($P_{CT} = 0.204$) but not in No (negative $P_{CT}$) and Ca ($P_{CT} = 0.076$) (fig. 3c); in the spleen of G1 ($P_{CT} = 0.305$) and No ($P_{CT} = 0.184$) but not in G2 ($P_{CT} = 0.076$) and Ca (negative $P_{CT}$). The genotypes of the eSNP residing 630 bp upstream of the TSS of dhrs13a.3 were significantly correlated with gene expression levels across individuals in both tissue types (corrected $P < 0.001$, fig. 3a showed in head kidney). This SNP was differentiated between lake and river ecotypes and had consistently higher allele frequency of the

### Table 1

| Gene ID          | Gene Name                      | GO Function                                                                 | Tissue of eCNV | Higher Copy Number | $V_{CT}$ | $P_{CT}$ |
|------------------|--------------------------------|------------------------------------------------------------------------------|----------------|--------------------|---------|---------|
| ENSGACG00000008264 | Novel Gene                     | Unknown                                                                      | Both River     | −0.025             | 0.045   |         |
| ENSGACG0000010952 | Cytochrome P450 family 3 sub-family A member 43 (CYP3A43) | Membrane; integral component of membrane | SP Lake        | 0.178              | 0.001   |         |
| ENSGACG0000012073 | Novel gene                     | Unknown                                                                      | SP Lake        | 0.578              | 0.289*  |         |
| ENSGACG0000015897 | Cathepsin A                   | Unknown                                                                      | Both Lake      | 0.197              | 0.148   |         |
| ENSGACG0000016770 | Deoxyuridine triphosphatase (dut) | dUTP diphosphatase activity                                                   | Both Lake      | 0.348              | 0.245*  | 0.184   |
| ENSGACG0000018877 | GTPase, IMAP family member 7 (GIMAP7) | GTP binding                                                                | Both Lake      | 0.210              | 0.123   | 0.072   |
| ENSGACG0000019933 | si: dkey-85k7.12              | Unknown                                                                      | Both Lake      | 0.178              | 0.032   | −0.001  |
| ENSGACG0000020614 | Solute carrier family 47 (scl47a1) | Membrane; integral component of membrane                                   | Both Lake      | 0.144              | 0.082   | 0.002   |
| ENSGACG0000008242 | Novel gene                     | Unknown                                                                      | Both Lake      | 0.215              | 0.186   | 0.136   |

**NOTE.**—NAs in $P_{CT}$, expression levels did not meet the filtering requirements and therefore $P_{CT}$ were not calculated.  
*Significant $P_{CT}$ (Benjamini-Hochberg corrected $P < 0.05$).
allele G in the lake populations (fixed in G1_L and G2_L, and 83.3% in No_L and Ca_L) and higher allele frequency of T in the river populations (25% in G1_R, and 41.7% in G2_R, No_R, and Ca_R, fig. 3b). Both alleles occur in all four population pairs, and we confirmed that both were also present in an adjacent marine population from the North Sea (Feulner et al. 2013), with a low frequency of the T allele (8.3%). This suggests that the T allele derives from standing genetic variation in the ancestral marine populations, and repeatedly increased in frequency among river populations possibly due to positive selection. However, no selective sweep was found based on nucleotide diversity ($p$) in the 50-kb flanking region of the SNP, which did not differ between lake and river populations (fig. 4a). The gene region of dhrs13a.3 harbors 51 SNPs across the four population pairs, with two synonymous and two nonsynonymous SNPs in the exons, and other SNPs in the introns. The nonsynonymous SNP, which substitutes a glycine with an arginine in the first exon, has the minor allele present in G1_L and G2_R with frequencies of 50% and 16.7%, respectively. The other nonsynonymous SNP, which substitutes a cysteine with a phenylalanine in the third exon, has the minor allele present in Ca_L with a frequency of 16.7%.

**eCNV Regulating Expression Differentiation in Cathepsin A**

The gene *cathepsin A* had higher expression levels in spleen among river sticklebacks in the two German population pairs G1 ($P_{CT} = 0.664$) and G2 ($P_{CT} = 0.409$; fig. 3f), but was not differentially expressed in No nor Ca (negative $P_{CT}$ values). In head kidney tissues, this gene also had higher expression in river sticklebacks in the population pairs of G1 ($P_{CT} = 0.797$) and G2 ($P_{CT} = 0.190$) and Ca ($P_{CT} = 0.112$) whereas in No it had higher expression in the lake fish ($P_{CT} = 0.521$). The consistent DE in the two German population pairs was accompanied by copy number differentiation. This gene was the most differentiated gene CNV between lake and river sticklebacks in the two German population pairs ($V_{CT}$ of 0.96 in G1 and 0.51 in G2) as previously reported (Chain et al. 2014), but not differentiated in No nor Ca ($V_{CT} = 0$) suggesting that the two German population pairs drive the overall habitat-specific signal (fig. 3e). We further identified *cathepsin A* as a gene eCNV, meaning that the gene copy numbers were significantly correlated with gene expression levels across individuals (corrected $P < 0.001$ in both tissue types, fig. 3d). To investigate whether the *cathepsin A* CNV is derived from standing genetic variation from an ancestral population, we searched...
| Gene ID | Gene Name | GO Function | Biological Process | Position in Relation to TSS | Distance to TSS (bp) | Tissue | $P_{CT}$ | eSNP | $P_{CT}$ | $P_{CT}$ |
|---------|-----------|-------------|--------------------|-----------------------------|---------------------|--------|---------|------|---------|---------|
| ENSGACG00000000642 | Novel Gene a&k homolog 6 (aktb6) | Unknown | Cellular Component | Up-stream | 1,637 | SP | 0.150 | NA | 0.073 | -0.005 |
| ENSGACG00000000741 | alkB homolog 6 (alkbh6) | Unknown | Oxidoreductase activity | Down-stream | 3,811 | HK | 0.378 | -0.021 | 0.005 |
| ENSGACG00000000382 | si: ch73-14h10.2 | Unknown | RNA binding | Up-stream | 1,611 | SP | 0.120 | NA | 0.075 | -0.006 |
| ENSGACG00000000425 | Activity-dependent neuroprotector homeobox b (adnpb) | Unknown | DNA binding | Down-stream | 2,315 | HK | 0.336 | 0.010 | -0.028 |
| ENSGACG00000000444 | Myocardin-related transcription factor b (mrtfba) | Unknown | Nucleic acid binding | Down-stream | 2,235 | HK | 0.336 | 0.010 | -0.028 |
| ENSGACG00000000484 | RAS related (rras) | Unknown | Nucleic acid binding | Down-stream | 2,235 | HK | 0.336 | 0.010 | -0.028 |
| ENSGACG00000000916 | Ly1 antibody reactive homolog (lyar) | Unknown | DNA binding | Down-stream | 1,808 | HK | 0.241 | 0.023 | -0.012 |
| ENSGACG00000000614 | Ly1 antibody reactive homolog (lyar) | Unknown | DNA binding | Down-stream | 1,808 | HK | 0.241 | 0.023 | -0.012 |
| ENSGACG00000000115 | Sulfotransferase family A member 1 (sst1a1) | Unknown | Unknown | Down-stream | 2,220 | SP | 0.279 | 0.320 | 0.002 |
| ENSGACG00000000126 | Sulfotransferase family A member 1 (sst1a1) | Unknown | Unknown | Down-stream | 1,970 | HK | 0.173 | -0.012 | 0.038 |
| ENSGACG00000000318 | None | None | Nucleic acid binding | Up-stream | 2,912 | HK | 0.226 | 0.039 | 0.047 |
| Gene ID | Description | Expression Type | Distance | Expression Levels |
|---------|-------------|-----------------|----------|------------------|
| ENSGACG00000013614 | Dehydrogenase/reductase (SDR family) member 13a, duplicate 3 (dhrs13a.3) | Up-stream | 630 | Both 0.205 0.217 * 0.142 |
| ENSGACG00000015279 | Downstream neighbor of SON (DONSON) | Up-stream | 1,525 | HK 0.174 0.009 0.268 |
| ENSGACG00000016707 | ATP-binding cassette, subfamily A (ABC1), member 4b (abca4b) | Membrane; integral component of membrane | Nucleotide binding; ATP binding; ATPase activity; ATPase activity, coupled to transmembrane movement of substances | Transmembrane transport | Down-stream | 2,738 | HK 0.206 0.008 0.002 |

*Significant PCT. (Benjamini-Hocherg corrected P < 0.05)

---

NOTE.—NAs in PCT, expression levels did not meet the filtering requirements and therefore PCT were not calculated.
for the presence of CNVs in the adjacent marine population from the North Sea (Feulner et al. 2013). The gene cathepsin A was not a CNV in the marine population, suggesting that the gene duplication occurred since the divergence of the freshwater populations (G1 and G2) from the marine population, or that the marine samples that were sequenced did not capture this variation. Note that the marine sampling only consists of six individuals, hence we lack power to detect variants at low frequency. A 5-kb region in the gene region of cathepsin A was depleted from SNPs in G1_R leading to a nucleotide diversity \( (\pi) \) of zero despite being duplicated compared to G1_L, suggesting a signature of background selection (fig. 4b). In the other German populations, the gene harbors 23 SNPs, with two synonymous and one nonsynonymous SNP. The nonsynonymous SNP, which substitutes a leucine by a phenylalanine in an alternatively spliced exon, has the minor allele present as heterozygous in three individuals in G1_L and in two individuals in G2_L, and as homozygous in one G2_L individual.

**eCNV Regulating Expression Differentiation in GIMAP7**

The gene GIMAP7 had overall higher expression levels in the head kidney among lake ecotypes, and comparisons within population pairs found consistent directional differences across population pairs (fig. 3). \( P_{CT} \) in the population pairs ranged from 0.11 in G1, to 0.19 in G2, and 0.39 in Ca whereas expression levels did not meet filtering criteria in No. The expression in spleen tissues displayed the same direction of expression changes between lake and river
sticklebacks as in the head kidney, but differentiation was less pronounced: \( P_{CT} \) of 0.05 in G1, 0 in G2, 0.68 in No, and 0.07 in Ca. The \( V_{CT} \) values were reasonably high in at least three population pairs: 0.53 in G1, 0.64 in No, and 0.70 in Ca (fig. 3h). As with cathepsin A, GIMAP7 was a gene eCNV (corrected \( P = 0.0074 \) in head kidneys and corrected \( P < 0.001 \) in spleen, fig. 3g). GIMAP7 was not detected as a CNV in the North Sea marine population. This suggests independent duplication and deletion events in the freshwater populations since they diverged from the marine ancestor or that this variant is at low frequency in the marine population. In the genomic regions adjacent to GIMAP7, we found no differences in the levels of nucleotide diversity among the eight freshwater populations (fig. 4c). The gene region harbors a total of 38 SNPs across the four population pairs, 24 of which are nonsynonymous. This suggests that duplication and deletion of this gene might also contribute to the amino acid sequence diversification across population pairs.

**Genome-Wide Correlation between Genetic Differentiation and Expression Differentiation**

Genome-wide, \( F_{CT} \) in cis-regulatory regions did not significantly positively correlate with \( P_{CT} \) in either head kidney or spleen (\( \rho = 0.011, P = 0.12, n = 10,671 \) in head kidney and \( \rho = 0.006, P = 0.24, n = 10,974 \) in spleen; one-sided Spearman rank correlation). \( V_{CT} \) had a significant positive correlation with \( P_{CT} \) in spleen but not in the head kidney (\( \rho = 0.166, P < 0.001 \) for spleen; \( \rho = 0.064, P = 0.064 \) for head kidney; one-sided Spearman rank correlation).

**Discussion**

The genetic underpinnings of expression differentiation in adaptive evolution remain a focus of intense research. In this study, we combined genome-wide genetic variation and transcriptomic data from repeatedly evolved ecotypes of the three-spined stickleback to better understand their relationships in the process of adaptation to distinct habitats.
We first report a prevalent dosage effect of CNV genes on gene expression and numerous SNPs in cis associated with expression. The prevalent association between genetic variants and expression levels might provide phenotypic variation that promotes adaptation to distinct lake and river habitats. We describe one gene with a differentiated SNP that is associated with expression differentiation between lake and river populations, and two genes with significant associations between copy number differentiation and expression differentiation. These findings provide evidence that both SNPs and CNVs contribute to gene expression differentiation between recently diverged ecotypes.

**Dosage Effects of CNVs Contribute to Expression Differentiation**

CNVs reflect components of genome architectures that vary in the number of copies of a sequence and have been proposed to have a greater impact on gene expression compared with sequence modifications (Sudmant et al. 2015; Huddleston and Eichler 2016). We found that 39% (135) of all expressed gene CNVs have a positive association with expression in at least one of the two tissues sampled, with 60 gene CNVs showing significant positive association in both tissues. These results demonstrate prevalent dosage effects on gene expression across tissue types. Similar number of genes show associations between CNVs and expression changes in humans (e.g., 110 genes in Schlattl et al. [2011] and 44–96 genes in Stranger et al. [2007]) and a similar proportion (42%) of genes in Drosophila (Cardoso-Moreira et al. 2016). Recently, the Genotype-Tissue Expression (GTEx) Project also found large effect sizes of structural variations on gene expression in humans (e.g., 110 and 44–96 genes in Stranger et al. 2007). Its isoforms have roles in MHC class II antigen presentation (Hsing and Rudensky 2005). More copies of the gene and therefore higher expression conceivably impact the immune response, whereas most of the gene region is depleted from variation despite the duplication in G1_R, suggesting background selection on the duplication. As river sticklebacks have lower MHC diversity compared with lake ecotypes (Eizaguirre et al. 2011), the higher copy number and expression of this gene potentially has a compensatory role and contributes to the defense against parasites specific to the river habitat. In contrast, lake ecotypes across population pairs were found to have higher copy numbers and higher expression of the gene GIMAP7, a GTPase that contains a domain AlG1-type G with immunity-associated functions (Krücken et al. 2004; Schwefel et al. 2010). The increase in GIMAP7 copy number is associated with higher expression, possibly contributing to higher immune competence in lake individuals, as the parasite pressure is more intense in lake habitats (Scharsack et al. 2007; Eizaguirre et al. 2011). The matching habitat-specific expression patterns of cathepsin A and GIMAP7 in immune tissues add to previous findings that CNVs are likely an important source of genetic variation that can help shape the host innate and adaptive immune response (Chain et al. 2014; Machado and Ottolini 2015). Our study on habitat-specific expression in immune tissues, which can potentially capture parasite-mediated
selection, has revealed two immune-related gene CNVs associated with expression differentiation, whereas other CNVs possibly contribute to habitat-specific adaptations in other tissues not sampled in our study. Previous investigation between marine and freshwater sticklebacks identified 24 gene CNVs consistent with parallel evolution, two of which were also found with DE between photoperiod treatments (APOL2 and ENSGACG00000003408, Hirase et al. 2014). These two genes were also gene CNVs in our population system, with ENSGACG00000003408 also marginally differentiated between our lake and river populations ($V_{CT} = 0.124, FDR = 0.053$), but neither gene was expressed in our transcriptome data. In addition to Hirase et al. (2014), our findings of two gene eCNVs with significant $V_{CT}$ and $P_{CT}$ highlight an important role of gene CNVs in adaptation to new environments in sticklebacks.

**eSNPs in cis Also Contribute to Expression Variation**

In addition to CNVs affecting gene expression, a total of 1,865 genes had SNPs in cis-regulatory regions identified as eSNPs putatively affecting gene expression. Though association tests between gene expression and SNPs do not necessarily reflect causal relationships, this result is consistent with previous studies that found abundant cis-eQTLs associated with expression divergence between stickleback ecotypes (Ishikawa et al. 2017; Pritchard et al. 2017; Kitano et al. 2019). Comparing marine and freshwater sticklebacks, Ishikawa et al. (2017) reported that about half of their local eQTLs resided in genomic regions of high divergence. Extending the comparison to multiple population pairs and between lake and river populations, we identified a gene differentiated between ecotypes both at the genetic level of an eSNP and in gene expression. The lake and river sticklebacks used in this study exhibit low parallel genomic divergence despite an isolation-by-adaptation signal (Feulner et al. 2018); genomic regions that most likely contribute to ecological divergence vary across different population pairs, suggesting the regulatory changes responsible for expression differentiation might also be population specific. As for dhrs13a.3, the homozygous T genotype of the eSNP 630 bp upstream of the TSS was associated with lower expression, and present in higher frequency in river populations where parasite abundance is generally much lower than in lakes (Scharsack et al. 2007; Eizaguirre et al. 2011). This allele is present in a detectable but low frequency (8.3%) in a source marine population (North Sea, Feulner et al. 2013) as well as in our lake populations, suggesting repeated increases in frequency in river habitats putatively due to habitat-specific adaptation.

Despite the abundance of genes with eSNPs, sequence differentiation of 5 kb upstream regions had an overall nonsignificant correlation with expression differentiation. This lack of genome-wide correlation between sequence-based differentiation in cis-regulatory regions and expression differentiation is consistent with other studies in whitefish, flycatcher, and Drosophila (Renaut et al. 2012; Zhao et al. 2015; Uebbing et al. 2016), and can be at least partly explained by the narrow transcriptomic snapshot analyzed. Sequence differentiation might still impact expression differentiation in other tissues or at different developmental times not captured in our data. We also cannot exclude the impact that environmental plasticity might play in shaping expression differentiation. Although trans-regulatory changes may also contribute to expression divergence (e.g., Hart et al. 2018), we focused on cis-regulatory changes, which were found to account for large parts of parallel expression changes between marine and freshwater sticklebacks (Verta and Jones 2019). Taken together, our results highlight examples of SNPs and CNVs that contribute to expression differentiation linked to adaptive divergence.

**Conclusion**

By combining genome and transcriptome data from the same individuals across independently evolved population pairs, we describe generalities of the genetic basis of gene expression differentiation between lake and river sticklebacks. We revealed numerous changes of nucleotides in cis-regulatory elements that are associated with expression variation and prevalent dosage effects of CNVs on gene expression, providing variation that can foster rapid adaptation to different environments. We report one SNP in cis and two CNVs linked to gene expression differentiation that likely contribute to divergence between repeatedly evolved ecotypes. Our findings highlight both SNPs and CNVs as sources of genetic variation that promote repeated adaptation via cis-regulatory effect or dosage effect on gene expression.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.

**Acknowledgments**

We thank the International Max Planck Research School for Evolutionary Biology for research support. We thank Prof. Tal Dagan for discussions on the study. We thank Derk Wachs Murdoch for computational assistance. We also thank Belinda Chang and three anonymous referees for their help in improving this article. This study was funded by the Max Planck Society.

**Author Contributions**

Y.H., F.J.J.C., and P.G.D.F. and E.B-B designed the analyses. Y.H. performed the analyses, and all authors contributed to discussions on research design and interpretation of the results. Y.H. drafted the article together with F.J.J.C. and P.G.D.F. All authors revised the article.
Literature Cited

Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21(6):974–984.

Abzhavanov A, et al. 2006. The calmodulin pathway and evolution of elongated beat morphology in Darwin’s finches. Nature 442(7102):563–567.

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22(13):1600–1607.

Axelsson E, et al. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495(7441):360–364.

Barrett RD, Hoeckstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet. 12(1):767–780.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B (Methodological). 57(1):289–300.

Berner D, Adams DC, Grandchamp AC, Hendry AP. 2008. Natural selection drives patterns of lake–stream divergence in stickleback foraging morphology. J Evol Biol. 21(6):1653–1665.

Birchler JA, Veitia RA. 2012. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci U S A. 109(37):14746–14753.

Bravand D, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513(7518):375–381.

Cardoso-Moreira M, et al. 2016. Evidence for the fixation of gene duplications by positive selection in Drosophila. Genome Res. 26(6):787–798.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134(1):25–36.

Chan FY, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Ptx1 enhancer. Science 327(5963):302–305.

Chiang C, et al. 2017. The impact of structural variation on human gene expression. Nat Genet. 49(5):692–699.

Ching T, Huang S, Garmire LX. 2014. Power analysis and sample size estimation for RNA-Seq differential expression. RNA 20(11):1684–1696.

Colombo M, et al. 2013. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. Mol Ecol. 22(3):670–684.

Deagle BE, et al. 2012. Population genomics of parallel phenotypic evolution in stickleback across stream–lake ecological transitions. Proc Biol Sci. 279(1732):1277–1286.

Dernet N, Duchesne P, Bernatchez L. 2006. Parallelism in gene transcription among sympatric lake whitefish (Coregonus clupeaformis Mitchell) ecotypes. Mol Ecol. 15(5):1239–1249.

Eizaguirre C, et al. 2011. Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. Evol Ecol. 25(3):605–622.

Eizaguirre C, Lenz TL. 2010. Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. J Fish Biol. 77(9):2023–2047.

Eizaguirre C, Lenz TL, Traulsen A, Milinski M. 2009. Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. Ecol Lett. 12(1):5–12.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 10(3):564–567.

Faucon F, et al. 2015. Identifying genomic changes associated with insecticide resistance in the dengue mosquito Aedes aegypti by deep targeted sequencing. Genome Res. 25(9):1347–1359.

Feulner PG, et al. 2013. Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. Mol Ecol. 22(3):635–649.

Gamazon ER, Stranger BE. 2015. The impact of human copy number variation on gene expression. Brief Funct Genomics. 14(5):352–357.

Gibson G. 2008. The environmental contribution to gene expression profiles. Nat Rev Genet. 9(8):575–581.

Gilad Y, Rifkin SA, Pritchard JK. 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 24(8):408–415.

González-Prendes R, et al. 2017. Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. Sci Rep. 7:39830.

Hanson D, Hu J, Hendry AP, Barrett R. 2017. Heritable gene expression differences between lake and stream stickleback include both parallel and antiparallel components. Heredity 119(5):339–348.

Haraksingh RR, Snyder MP. 2013. Impacts of variation in the human genome on gene regulation. J Mol Biol. 425(21):3970–3977.

Hart J, Ellis NA, Eisen MB, Miller CT. 2018. Convergent evolution of gene expression in two high-toothed stickleback populations. PLoS Genet. 14(6):e1007443.

Hartasánchez DA, Brós–Vives M, Heredia-Genestar JM, Pybus M, Navarro A. 2018. Effect of collapsed duplications on diversity estimates: what to expect. Genome Biol Evol. 10(11):2899–2905.

Henrichsen CN, Chaignat E, Reymond A. 2009. Copy number variants, diseases and gene expression. Hum Mol Genet. 18(R1):R1–R8.

Hirase S, Ozaki H, Iwasaki W. 2014. Parallel selection on gene copy number variations through evolution of three-spined stickleback genomes. BMC Genomics 15(1):735.

Hoeckstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. Evolution 61(5):995–1016.

Hsing LC, Rudensky AY. 2005. The lysosomal cysteine proteases in MHC class II antigen presentation. Immunol Rev. 207:229–241.

Huang Y, et al. 2016. Transcriptome profiling of immune tissues reveals habitat-specific gene expression between lake and river sticklebacks. Mol Ecol. 25(4):943–958.

Huddleston J, Eichler EE. 2016. An incomplete understanding of human genetic variation. Genetics 202(4):1251–1254.

Ishikawa A, et al. 2017. Different contributions of local- and distant-regulatory changes to transcriptome divergence between stickleback ecotypes. Evolution 71(3):565–581.

Jones FC, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484(7392):55–61.

Kaeuffer R, Peichel CL, Bolnick DI, Hendry AP. 2012. Convergence and non-convergence in ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. Evolution 66(2):402–418.

Kalbe M, Wegner KM, Reusch TBH. 2002. Dispersion patterns of parasites in 0+ year three-spined sticklebacks: a cross population comparison. J Fish Biol. 60(6):1529–1542.

Katju V, Bergthorsson U. 2013. Copy-number changes through evolution of three-spined stickleback genomes. Dev Growth Differ. 61(1):104–113.

Krücken J, et al. 2004. Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. Gene 341:291–304.

Leder EH, et al. 2015. The evolution and adaptive potential of transcriptional variation in sticklebacks—signatures of selection and widespread heritability. Mol Biol Evol. 32(3):674–689.
Lenz TL. 2015. Transcription in space – environmental vs. genetic effects on differential immune gene expression. Mol Ecol. 24(18):4583–4585.

Lenz TL, Ezaguirre C, Rotter B, Kalbe M, Milinski M. 2013. Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis. Mol Ecol. 22(3):774–786.

Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009. On the origin and spread of an adaptive allele in deer mice. Science 325(5944):1095–1098.

Lueck KE, Sivasundar A, Kristjánsson BK, Skúlason S, See Hansen O. 2014. Quick divergence but slow convergence during ecotype formation in lake and stream stickleback pairs of variable age. J Evol Biol. 27(9):1878–1892.

Machado LR, Ottolini B. 2015. An evolutionary history of defensins: a role for copy number variation in maximizing host innate and adaptive immune responses. Front Immunol. 6:115.

Manousaki T, et al. 2013. Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. Mol Ecol. 22(3):650–669.

Mallarino R, Linden TA, Linnen CR, Hoekstra HE. 2017. The role of isoforms in the evolution of cryptic coloration in Peronoscyces mice. Mol Ecol. 26(1):245–258.

Manousaki T, Hull PM, Kusche H, Machado-Schiaffino G, Franchini P, Harrod C, Elmer KR, Meyer A. 2013. Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. Mol Ecol. 22(3):650–669.

McKinnon JS, Randie HD. 2002. Speciation in nature: the threespine stickleback model systems. Trends Ecol Evol. 17:480–488.

Nguyen D-Q, Webber C, Ponting CP. 2006. Bias of selection on human copy-number variants. PLoS Genet. 2(2):e20.

Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. 2016. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 32(10):1479–1485.

Papp B, Pal C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene copy-number variants. PLoS Genet. 2(2):e20.

Perry GH, et al. 2007. Diet and the evolution of human amylase gene copy number. Nat Genet. 39(10):1217–1224.

Papp B, Pal C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene copy-number variants. PLoS Genet. 2(2):e20.

Perry GH, et al. 2007. Diet and the evolution of human amylase gene copy number. Nat Genet. 39(10):1217–1224.

Pritchard VL, et al. 2017. Regulatory architecture of gene expression variation across continents in Drosophila melanogaster. Mol Biol Evol. 33(5):1308–1316.

Schwefel D, et al. 2010. Structural basis of oligomerization in septin-like GT-Pase of immunity-associated protein 2 (GI-MAP2). Proc Natl Acad Sci U S A. 107(47):20299–20304.

Shen Y, et al. 2012. A map of the cis-regulatory sequences in the mouse genome. Nature 488(7409):116–120.

Stamatoyannopoulos JA. 2004. The genomics of gene expression. Genomics 84(3):449–457.

Stranger BE, et al. 2007. Population genomics of human gene expression. Nat Genet. 39(10):1217–1224.

Stuart YE, et al. 2017. Contrasting effects of environment and genetics generate a continuum of parallel evolution. Nat Ecol Evol. 1(6):158.

Stutz WE, Schmerzer M, Coates JL, Bolnick DI. 2015. Among-lake reciprocal transplants induce convergent expression of immune genes in threespine stickleback. Mol Ecol. 24(18):4629–4646.

Sudmant PH, et al. 2015. An integrated map of structural variation in 2,504 human genomes. Nature 526(7571):75–81.

Tanaka T, Yamauchi S, Muto S, Akiyama K, Yano K. 2015. Non-coding RNA sequences are enriched near regulatory sites in the human genome. Genome Biol. 16:133.

Timmur ZK, Akylidiz Demir S, Seyrantepe V. 2016. Lysosomal Cathespin A plays a significant role in the processing of endogenous bioactive peptides. Front Mol Biosci. 3:68.

Uebbing S, et al. 2016. Divergence in gene expression within and between two closely related flycatcher species. Mol Ecol. 25(9):2015–2028.

Veitia RA. 2005. Gene dosage balance: deletions, duplications and dominance. Trends Genet. 21(1):33–35.

Verta JP, Jones FC. 2019. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. eLife 8:e43785.

Whitehead A, Crawford DL. 2006a. Neutral and adaptive variation in gene expression: raw material for evolution. Mol Ecol. 15(5):1197–1211.

Wittkopp PJ, Haerum BK, Clark AG. 2008. Regulatory changes underlying expression differences between low and high latitude populations of Drosophila melanogaster. Mol Ecol. 17:480–488.

Wittkopp PJ, Kalbe M, Harrod C, Rauch G. 2007. Habitat-specific adaptation of immune responses of stickleback (Gasterosteus aculeatus) lake and river ecotypes. Proc Biol Sci. 274(1617):1523–1532.

Zhao L, Wit J, Svetec N, Begun DJ. 2015. Parallel gene expression differences between low and high latitude populations of Drosophila melanogaster and D. simulans. Mol Ecol. 24(18):4629–4646.