Transcriptome profiling of immune tissues reveals habitat-specific gene expression between lake and river sticklebacks

YUN HUANG,* FRÉDÉRIC J. J. CHAIN,** MAHESH PANCHAL,*,‡ § CHRISTOPHE EIZAGUIRRE,¶ MARTIN KALBE,* TOBIAS L. LENZ,* IRENE E. SAMONTE,* MONIKA STOLL,** ERICH BORNBERG-BAUER,† † THORSTEN B. H. REUSCH,‡ ‡ MANFRED MILINSKI* and PHILINE G. D. FEULNER*§§

*Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany, †Department of Biology, McGill University, Montreal, QC, Canada, H3A 1B1, ‡Bioinformatics Infrastructures for Life Sciences (BILS), Uppsala Biomedicinska Centrum (BMC), Husargatan 3, 751 23 Uppsala, Sweden, §Institute of Medical Biochemistry and Microbiology, Uppsala Biomedicinska Centrum (BMC), Husargatan 3, 751 23 Uppsala, Sweden, ¶School of Biological and Chemical Sciences, Queen Mary University of London, E1 4NS London, UK, **Institute of Human Genetics, Genetic Epidemiology, Westfälische Wilhelms University, 48149 Münster, Germany, ††Institute for Evolution and Biodiversity, Evolutionary Bioinformatics, Westfälische Wilhelms University, 48149 Münster, Germany, ‡‡Evolutionary Ecology of Marine Fishes, GEOMAR Helmholtz Centre for Ocean Research Kiel, 24105 Kiel, Germany, §§Department of Fish Ecology and Evolution, Eawag Swiss Federal Institute of Aquatic Science and Technology, Center for Ecology, Evolution and Biogeochemistry, 6047 Kastanienbaum, Switzerland

Abstract

The observation of habitat-specific phenotypes suggests the action of natural selection. The three-spined stickleback (Gasterosteus aculeatus) has repeatedly colonized and adapted to diverse freshwater habitats across the northern hemisphere since the last glaciation, while giving rise to recurring phenotypes associated with specific habitats. Parapatric lake and river populations of sticklebacks harbour distinct parasite communities, a factor proposed to contribute to adaptive differentiation between these eco-types. However, little is known about the transcriptional response to the distinct parasite pressure of those fish in a natural setting. Here, we sampled wild-caught sticklebacks across four geographical locations from lake and river habitats differing in their parasite load. We compared gene expression profiles between lake and river populations using 77 whole-transcriptome libraries from two immune-relevant tissues, the head kidney and the spleen. Differential expression analyses revealed 139 genes with habitat-specific expression patterns across the sampled population pairs. Among the 139 differentially expressed genes, eight are annotated with an immune function and 42 have been identified as differentially expressed in previous experimental studies in which fish have been immune challenged. Together, these findings reinforce the hypothesis that parasites contribute to adaptation of sticklebacks in lake and river habitats.

Keywords: habitat-specific gene expression, immune genes, parasites, RNA-Seq, three-spined stickleback, transcriptomics

Received 1 April 2015; revision received 18 November 2015; accepted 10 December 2015

Introduction

The repeated occurrence of similar phenotypes associated with a distinct habitat is often attributed to the direct effect of natural selection (Elmer & Meyer 2011).
Parallel phenotypic evolution among populations from geographically distant but ecologically similar habitats, referred to here as habitat-specific phenotypes, is thought to reflect the advantages of those phenotypes in their respective habitat (Savolainen et al. 2013). Numerous examples have been documented including pharyngeal jaw and thick lips in cichlids (Albertson et al. 2005; Colombo et al. 2013), similar ecotype morphs of anolis lizards (Losos et al. 1998; Harman et al. 2005), habitat-specific pigmentation in isopods (Hargeby et al. 2004), repeated ecotypes with distinct shell sizes in the periwinkle snail (Butlin et al. 2014) and repeated differences of body depth and gill raker numbers between lake and stream sticklebacks (Berner et al. 2008; Kaeuffer et al. 2012; Lucek et al. 2014). Although phenotypic plasticity can contribute to such habitat-specific phenotypes (Muschick et al. 2012; Machado-Schiaffino et al. 2014; Moser et al. 2015), some of these traits have been shown to be genetically determined and under adaptive evolution (Hargeby et al. 2004; Albertson et al. 2005; Colombo et al. 2013). Adaptive genetic changes include those that result from polymorphisms that alter protein structures (Ffrench-Constant et al. 1993; Hoekstra et al. 2006; Protas et al. 2006) as well as those that influence phenotypes via regulation of gene expression (Rebeiz et al. 2009; Chan et al. 2010). Gene expression has been associated with adaptive changes in morphological and physiological changes (Rebeiz et al. 2009; Manceau et al. 2011; Harrison et al. 2012) and is believed to contribute to adaptive divergence in natural populations (Pavey et al. 2010).

As gene expression bridges the underlying genotype to the ultimate morphological and physiological phenotypes, it can be considered as an extended molecular phenotype (Ranz & Machado 2006). Hence, it is interesting to evaluate whether or not gene expression patterns differ between contrasting habitats and if so whether they hold across geographically distant populations. Such habitat-specific gene expression could arise due to several factors, such as genetically determined expression patterns among similar habitat types (ecotypes), as well plastic responses to extrinsic environmental conditions specific to a habitat. Aside from other mechanisms that might control regulation of transcription such as epigenetics, genetic studies have demonstrated variable degrees of heritability of gene expression and have for some phenotypes revealed the genetic basis underlying expression differences (Stamatoynopoulos 2004; Gibson & Weir 2005; Gilad et al. 2008). There are examples of mutations affecting cis- and trans-regulatory regions in the genome that silence or dramatically shift gene expression, including single nucleotide polymorphisms (SNPs) (Cheung & Spielman 2009; Fraser 2013), copy number variations (CNVs) (Haraksingh & Snyder 2013) and tandem repeats (Gemayel et al. 2010). Genomic changes in regulatory regions can alter the efficiency of transcription factors and thus affect expression of adjacent or remote genes. In sticklebacks, for example, frequent independent deletion events in the enhancer of Pitx1 suppress expression of the gene and result in repeated pelvic reduction in freshwater populations (Chan et al. 2010). Besides its heritable (genetic) component, gene expression is also a versatile phenotype that dynamically responds to changes in the environment (Gibson 2008) and holds the potential to facilitate plasticity to buffer against environmental changes (Franssen et al. 2011; Whitehead 2012; Morris et al. 2014). Despite the variability introduced by uncontrollable environmental factors, studies of gene expression in wild-caught populations offer the opportunity to estimate the physiological responses of organisms in their environment, potentially providing insight into the role of gene expression variation in adaptation and acclimation to environmental stresses through genetic or plastic changes (Cheviron et al. 2008).

The repeated and independent postglacial colonization history of the three-spined stickleback (Gasterosteus aculeatus) makes it a powerful study system to investigate habitat-specific phenotypic evolution. Sticklebacks inhabit various marine and freshwater habitats across the northern hemisphere (MacKinnon & Rundle 2002), a distribution likely attributable to rapid adaptation from extensive standing genetic variation (Barrett & Schluter 2008; Eizaguirre et al. 2012a). Genetically diverged but geographically adjacent lake and river population pairs exhibit consistent morphological differentiation across multiple pairs, such as divergence for body depth and gill raker number (Berner et al. 2008; Kaeuffer et al. 2012; Lucek et al. 2014). These lake and river populations are also often referred to as ecotypes (Reusch et al. 2001). Many ecological factors differ between lake and river habitats, such as flow regime, temperature, food resource and predator communities, all contributing to the differentiation of lake and river stickleback ecotypes, for example in foraging traits (Berner et al. 2010) and antipredator traits (Lucek et al. 2014). Another important ecological difference between lakes and rivers is the locally distinct parasite communities (Kalbe et al. 2002; Eizaguirre et al. 2011; Karvonen et al. 2015). Besides harbouring different species of parasites between ecotypes, lake fish commonly have a higher parasite load than river fish comparing parapatic population pairs (Eizaguirre et al. 2011), and higher immuno-competence (Scharsack et al. 2007). Lake fish also exhibit a higher diversity in the major histocompatibility complex (MHC) (Eizaguirre et al. 2011), believed to be a result of local adaptation (Eizaguirre et al.
et al. 2009, 2012b). Distinct immune expression patterns between lake and river individuals were detected upon multiple experimental parasite exposure of laboratory-bred sticklebacks (Lenz et al. 2013). Altogether, these studies suggest that parasites play an important role in the differentiation of lake and river ecotypes by shaping the diversity and expression patterns of immune-related genes. It is, however, not yet known whether the generality of these patterns holds in multiple lake-river systems under natural conditions.

In this study, we performed an extensive transcriptomic survey using an RNAseq approach across four parapatric lake and river stickleback population pairs to investigate patterns of habitat-specific gene expression. We used two major organs involved in immune response, the head kidney and the spleen. Differential expression analysis was performed between fish from lake and river habitats, and results were compared to the differentially expressed genes between laboratory-bred individuals in controlled parasite infection experiments (Lenz et al. 2013; Haase et al. 2014). Our study describes gene expression differences in an ecological framework, highlighting habitat-specific expression of genes that might be involved in adaptation.

Materials and methods

Sampling

Three-spined sticklebacks were sampled in 2010 for genomic studies (Chain et al. 2014; Feulner et al. 2015), from which four parapatric lake–river population pairs were used in this study. These included two independent drainages from Germany: Großer Plöner See lake (G1_L) and Malenter Au river (G1_R), Westensee lake (G2_L) and Eider river (G2_R), one pair from Norway: Skogseidvatnet lake (No_L) and Orraelva river (No_R), and one pair from Canada: Misty Lake (Ca_L) and Misty Stream (Ca_R) (See Table 1). All these lake-river population pairs are significantly differentiated from each other, with a mean genome-wide $F_{ST}$ ranging between 0.11 and 0.28 (for more detailed information about sampling sites and genetic differentiation between the populations, see Feulner et al. 2015). The two population pairs from Germany were sampled in May while the Norwegian and Canadian populations were sampled in September. About 20 individual fish per site were caught using dip nets or minnow traps and kept alive for a few hours in the water from where they were sampled until being euthanized using MS222 and dissection. For each population pair, the fish were treated identically after capture and lake fish and river fish were alternatively dissected. Fish standard length and weight were recorded and macroparasites screened following established procedures for three-spined sticklebacks (Kalbe et al. 2002) (Table S1, Supporting information). Immediately after euthanasia, the whole head kidneys and spleens were dissected out and preserved in RNAlater (Sigma-Aldrich) for later transcriptomic library preparation. These are the main immune organs in teleost fish and are commonly used for immunological studies (Press & Evensen 1999). Six individuals (three males and three females, except No_L with four males and two females) were selected for transcriptomic sequencing per sampling site. Fish selection was performed ignoring parasite screening results, but was nonrandom to ensure an equal sex distribution for each population and with a preference for larger fish to guarantee sufficient yield of RNA. Body weights of the selected fish suggest that all fish were older than 1 year (Table S1, Supporting information).

RNA library preparation and sequencing

Total RNA (using the entire tissue dissected) was extracted from preserved samples using NucleoSpin® RNA (Mackerey-Nagel) and reverse transcribed to cDNA using Omniscript RT kits (Qiagen). RNA was quantified with NanoDrop and Bioanalyzer, and ~1 μg of RNA in a concentration of 20 ng/μL was used for library construction. A few samples with poor RNA quality were excluded before constructing 77 libraries.
Therefore, sample sizes per population vary between 3 and 6 individuals (Table 1). TruSeq RNA sample preparation kit (Illumina) was used for paired-end library construction according to the manufacturer’s instructions. Each sample was barcoded with a unique sequence index tag, and pools of 12 different barcoded samples were loaded in 8 lanes of a single flow cell of Illumina HiScanSQ machine.

**Read filtering and mapping**

Raw reads were quality filtered before read mapping in the following steps. All raw reads output to fastq files were 101 base pairs (bp) in length. Sequencing adaptors were removed using `SEQPREP` 0.4 (https://github.com/jstjohn/SeqPrep). `PRINSEQ` 0.20.3 lite (Schmieder & Edwards 2011) was used to trim the read tails with a PHRED quality score below 20 as well as poly-A tails longer than 10 bp. We kept read pairs for which both reads were longer than 60 bp after trimming. After filtering, read lengths varied from 60 to 101 bp, with about 60% of the reads exhibiting the initial 101-bp length. Exact duplicates of both paired-ends were removed with `PRINSEQ`. The remaining quality-filtered reads were aligned against the stickleback reference genome from ENSEMBL version 68 (Flicek et al. 2012) using `TOPHAT2` v2.0.13 (Kim et al. 2013) with default settings. `HTSEQ` 0.5.4p5 (Anders et al. 2014) was used to quantify read count for each gene using ENSEMBL gene annotations (version 68) using default settings except for excluding reads with alignment quality below 5.

**Gene expression analyses**

Gene expression across all samples was evaluated with the Bioconductor package `EDGER` 3.4.2 (Robinson et al. 2010). First, weakly expressed genes were filtered out when they had <1 read per million in half (38) of the 77 samples (Anders et al. 2013). All libraries were then simultaneously normalized with the trimmed mean of M-value (TMM) method (Robinson & Oshlack 2010), implemented in the `EDGER` package. The TMM method computes the scaling factors as the weighted mean of log fold changes for the majority of genes between libraries, based on the assumption that the majority of genes are not differentially expressed. After applying the TMM method, most genes should have unified expression levels across individuals and the scaling factors for the libraries should be close to 1 (Dillies et al. 2012). Except for one head kidney library from G1_R with a scaling factor of 0.35, all other transcriptome libraries obtained scaling factors close to 1 (from 0.75 to 1.18, Table S2, Supporting information). The outlier library had fewer genes expressed compared to other libraries (12 769 vs. 15 735–17 341). This indicates a distinct expression profile likely dominated by technical artefacts, and therefore, this library was excluded from further analyses.

Next, the dispersion of the negative binomial distribution for the expression of each gene was estimated in `EDGER`. It represents the biological coefficient of variation of a gene’s expression. This was used to evaluate the expression variance where a high dispersion value indicates high variance of gene expression pattern among samples. A principal component analysis (PCA) was then performed in R 3.0.1 (R Development Core Team 2008) using `prcomp` function based on log-transformed normalized read counts of all 12 222 expressed genes (across both tissues and after filtering out weakly expressed genes as mentioned above) to assess differences in gene expression across libraries (Fig. 1).

To identify habitat-specific gene expression, that is the expression patterns that are similar within habitat types while significantly different between habitat types, we employed differential expression (DE) analyses that contrast lake and river fish from all four population pairs. On the basis of the PCA result (Fig. 1), DE analyses were performed separately for head kidney and spleen libraries in `EDGER`. Because the PCA results suggest that the Canadian populations are substantially diverged from the European populations, the DE analyses were also performed only among the three European population pairs (those results are presented in the Supplement only). Hence, four DE analyses were performed (comparing gene expression in the head kidney across all four population pairs, in spleen across all four population pairs, in head kidney across only the three European population pairs and in spleen across only the three European population pairs). Before conducting DE analyses, weakly expressed genes were filtered out to avoid bias in fold changes due to weak expression of some genes. Genes were filtered out from the DE analyses if they did not have at least 1 read per million in \( n \) of the samples, where \( n \) is the size of the smaller group (lake or river) in the DE comparisons (Anders et al. 2013). Libraries were renormalized within each comparison group with the TMM method in `EDGER`. A multi-factor design was used in a negative binomial generalized linear model, which accounts for the variation attributed to different population pairs as well as for the variation associated to the sex of the individuals (Expression–Habitat type + Population pair + Sex). The gene-wise dispersion was reestimated based on the generalized linear model within each comparison group. For each tissue, the distribution of dispersion values was left-skewed with long tails, indicating that most genes had uniform expression, with a small proportion of genes having highly variable expression across...
individuals being compared (Fig. S1, Supporting information). We calculated the Pearson correlation of gene expression between all possible pairs of individuals within biological replicates (individuals of the same habitat, population pair and sex) using count data in R. The overall average correlation of gene expression across all pairwise comparisons was 0.86 (first quartile: 0.81 and third quartile: 0.95). Likelihood ratio tests for the contrast coefficient (lake vs. river) were performed, and P-values were corrected for multiple testing using the Benjamini-Hochberg method (Benjamini & Hochberg 1995). Genes with corrected P-values smaller than 0.05 were categorized as differentially expressed genes (DE genes). In addition to performing all DE analyses in EDGER as described above, DE analyses were also performed with the default pipeline in the DESEQ2 package 1.0.19 (Love et al. 2014) giving similar results (Table S3, Supporting information).

Functional analyses

Out of 20 787 stickleback genes, 13 568 are annotated with Gene Ontology [GO, (Ashburner et al. 2000)] terms in ENSEMBL version 80. We complemented this with 13 044 gene annotations acquired from the Zebrafish Model Organism Database (ZFIN, Howe et al. 2013) genes associated with stickleback ENSEMBL IDs, with annotation information from ftp://ftp.geneontology.org/pub/go/gene-associations/gene_association.zfin.gz. After merging all annotations, a total of 17 081 of 20 787 stickleback genes were annotated with GO terms. We tested for the enrichment of GO terms in our DE gene sets with the BIOCONDUCTOR package TOPGO (Alexa et al. 2006; Alexa & Rahnenfuhrer 2010), based on Fisher’s exact tests. The gene pools against which we compared the DE gene sets were the genes having sufficient expression and entering the differential expression analyses (see gene expression analyses section above). Overrepresented GO terms were those with a multiple-test corrected P-value (Benjamini-Hochberg’s false discovery rate, FDR) smaller than 0.05. To infer the potential involvement of the habitat-specific expressed genes in parasite defence in nature, we identified our DE genes that were also differential expressed in two previous laboratory-controlled parasite exposure experiments (Lenz et al. 2013; Haase et al. 2014).

Results

Qualitative description of expression patterns

For each of the 77 transcriptome libraries, an average of 6.5 million read pairs of 101 bp were produced. After adapter cleaning, quality trimming, and duplicate- and length-filtering, 92.78% of the reads remained for analyses (Table S2, Supporting information). On average, 88.10% of the quality-filtered reads mapped to the reference genome and 2.71% of these mapped to multiple regions of the genome, which were subsequently excluded from further analyses. Out of a total of 22 456 genes annotated in the stickleback genome (ENSEMBL version 68), an average of 16 397 (±944) genes were found expressed. The median number of reads mapping back to each expressed gene was 60 read pairs (first quartile to third quartile: 13–166). The principal component analysis (PCA) clearly separated the two tissue types along the first principal component, which accounted for 41% of the variance observed in the data set (Fig. 1). Within the same tissue type, the second principal component (variance explained: 8%)
clearly separated European samples from the Canadian samples.

**Differential expression (DE) analyses**

After filtering out weakly expressed genes (see Methods), 12,105 genes expressed in head kidney and 12,451 expressed in spleen were contrasted between lake and river ecotypes across all four population pairs. A total of 139 genes showed significant differential expression after correction for multiple testing (Fig. 2). There were 73 DE genes in the head kidney, 74 DE genes in the spleen and 8 of these genes were shared between both tissues (Table S3, Supporting information). All 8 shared DE genes showed the same directional difference of expression between habitat types. A majority of the DE genes (75% in head kidney and 65% in spleen) showed higher expression in individuals from lakes than from rivers. Most of these same DE genes were identified using another commonly used software with default parameters (DESEQ2: 70 of 73 in the head kidney and 67 of 74 in the spleen, Table S3, Supporting information). Although the PCA analyses mentioned above suggested that the overall expression patterns of the European samples seemed distinct from the Canadian samples, a separate analysis of expression log fold changes between lake and river fish from the three European population pairs showed a strong positive correlation with that of all four population pairs together (linear regression, \( R^2 = 0.61, P < 0.001 \) for head kidney and \( R^2 = 0.82, P < 0.001 \) for spleen) and resulted in about half of the same DE genes (Table S4, Supporting information). The 5 DE genes with the smallest adjusted \( P \)-value in the head kidney across all lake-river comparisons include three genes that have higher expression in lake fish (leucine-rich repeat containing 17, ryanodine receptor 3 and colony-stimulating factor 1b) and two that have higher expression in river fish (cub and sushi multiple domains 3 and one uncharacterized protein-coding gene ENSGACG0000000187). The five genes with smallest adjusted \( P \)-values in the spleen include three that have higher expression in lake fish (solute carrier family 43, member 3b, actin binding LIM protein 1b and complement factor D) and two uncharacterized protein-coding genes (ENSGACG0000000187 and ENSGACG0000012387) that have higher expression in river fish (see Table S3, Supporting information for all 139 DE genes identified).

**Functional analyses of DE genes**

GO annotations from ENSEMBL and the ZFIN database were available for 105 of the 139 DE genes (Table S3, Supporting information). The DE genes in head kidney had no significant GO term enrichment, while the DE genes in spleen were enriched for collagen (GO:0005581, with 3 of 18 genes annotated with this term in the gene pool), extracellular region (GO:0005576, with 8 of 265 genes) and extracellular matrix part (GO:0044420, with 3 of 20 genes). Applying a less stringent cut-off for DE genes (FDR < 0.10) to test for enrichment of GO terms (FDR < 0.05), only extracellular region (GO:0005576) remained significant in the spleen, with no additional terms found in both tissues. The top 50 GO terms from the enrichment analyses of original DE gene sets (FDR < 0.05) are provided in the Dryad database (see Data Accessibility Section). To specifically investigate the differential expression of immune genes in the sampled immune-related tissues, a list of 1126 stickleback genes with putative immune functions was acquired from a previous study (Haase et al. 2014). Among the DE genes between lake and river fish, three of the 73 DE genes in the head kidney and five of the 74 DE genes in the spleen are putatively immune genes (Table 2). These included macrophage receptors, an interferon regulatory factor and a gene annotated with the functions of antigen processing and presentation and immune response.

While our analysis only detected very few immune function genes showing differential gene expression, the parasite survey of our sampled fish showed that lake fish harbour higher parasite loads than river fish (Table S1, Supporting information). This has already been demonstrated previously using a larger sample size (Fig. 1 in Feulner et al. 2015). To further investigate the role of parasite infection and potential resistance in driving differential gene expression between lake and river habitats, we compared our results with two laboratory-controlled parasite exposure experiments that assessed gene expression in sticklebacks from the same German populations as used in our study. Lenz et al. (2013) described the transcriptional responses of laboratory-bred lake and river sticklebacks under either controlled or parasite-challenged conditions. That study used three parasites that are found in the natural environment of those fish: Diplostomum pseudopathaceum, Anguillicola crassus and Camallanus lacustris. These parasites were also found in our sampled fish (see discussion and Table S1, Supporting information). Of 166 DE genes between twice parasite-exposed lake and river fish (Lenz et al. 2013), 51 and 73 genes showed the same directional differences of expression between habitat types in our study among all lake-river population pairs, in the head kidney and in the spleen, respectively. Some of the differences between the two studies are likely due to that the majority of DE genes in Lenz et al. 2013 were highly expressed in river fish as they
are exposed to equal dosage of parasites compared to lake fish, while in our study the majority of DE genes were highly expressed in lake fish as the river fish were exposed to less parasites in nature. Nevertheless, among those genes with same directional differences, one gene methyltransferase like 13 (mettl13) was also identified significantly differentially expressed in our study (Table 3, also see Discussion for more details). In addition, 10 of the 1057 DE genes between control and parasite-challenged fish (Lenz et al. 2013) overlapped with our set of DE genes (four in the head kidney and six in the spleen). In another recent parasite infection study, laboratory-bred lake sticklebacks (from the G1_L population) were challenged with the trematode *D. pseudospathaceum* (Haase et al. 2014), and DE was assessed in the head kidney and in the gill. Of 1060 DE genes
between control and challenged fish in the head kidney (Haase et al. 2014), six overlapped with the DE genes from our study (all in the spleen). Of 1415 DE genes in the gill (Haase et al. 2014), 25 overlapped with our set of DE genes (12 in the head kidney and 14 in the spleen, including 1 in both tissues, Table 3).

Discussion

Habitat-specific expression

This study investigated transcriptional profiles of threespined sticklebacks from contrasting lake and river habitats across a wide geographical scale. Physical and ecological differences between lake and river habitats, consisting of differences in flow regime, vegetation, food resources and parasite communities among others, can influence individual fitness, behaviour, life history, morphology and physiology. Studies contrasting lake and river sticklebacks have mainly focused on their morphology (Berner et al. 2010; Lucek et al. 2014) and genomic variation (Deagle et al. 2012; Roesti et al. 2012; Chain et al. 2014; Feulner et al. 2015). Here, we evaluated how lake and river ecotypes differ in gene expression profiles in their natural environments. We have identified habitat-specific gene expression patterns, that is differential expression between habitats across four lake-river pairs, three from European locations and one from Canada. For differentially expressed genes, fish from the same habitat have a similar expression, which is distinct from the expression in fish from the contrasting habitat. These habitat-specific expression patterns suggest that a part of the transcriptome (about 1%) is shaped by the global environmental contrast across all lake-river pairs, although a larger fraction may be affected by local habitat differences within a given population pair or expressed in other tissues or during a different season or ontogenetic stage. These findings add to the growing discussion of parallelism at the regulatory level between contrasting ecotypes and morphs (Derome et al. 2006; Pavey et al. 2011; Manousaki et al. 2013).

Plasticity and heritability of gene expression

A combination of evolutionary mechanisms could be shaping the habitat-specific expression patterns observed in this study. Freshwater sticklebacks likely possess the innate ability to regulate certain genes in acclimating to the different conditions in lakes and rivers (Stutz et al. 2015). This plasticity could result in habitat-specific expression patterns. Alternatively, differential expression across habitats might also reveal adaptive genetic differences between lake and river fish. These alternative explanations for habitat-specific patterns are by no means mutually exclusive and may both

Table 2 Differentially expressed genes between all lake and river populations with putative immune functions

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene name</th>
<th>GO term (biological process)</th>
<th>Tissue</th>
<th>Log fold-change*</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSGACG00000001509</td>
<td>Marco</td>
<td>Scavenger receptor activity (molecular function)</td>
<td>Head kidney</td>
<td>0.73</td>
<td>0.0053</td>
</tr>
<tr>
<td>ENSGACG000000016979</td>
<td>CMKLR1 (2 of 2)</td>
<td>G-protein coupled receptor signalling pathway</td>
<td>Head kidney</td>
<td>0.77</td>
<td>0.0070</td>
</tr>
<tr>
<td>ENSGACG000000015855</td>
<td>RAB27A</td>
<td>Nucleocyttoplasmic transport</td>
<td>Head kidney</td>
<td>0.56</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small GTPase mediated signal transduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Signal transduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intracellular protein transport</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSGACG00000010551</td>
<td>Mst1ra</td>
<td>Protein phosphorylation</td>
<td>Spleen</td>
<td>0.89</td>
<td>0.0030</td>
</tr>
<tr>
<td>ENSGACG00000012609</td>
<td>LGALS1 (2 of 3)</td>
<td>Carbohydrate binding (molecular function)</td>
<td>Spleen</td>
<td>0.73</td>
<td>0.0038</td>
</tr>
<tr>
<td>ENSGACG00000004966</td>
<td>IRF4 (2 of 2)</td>
<td>Regulation of transcription, DNA-templated</td>
<td>Spleen</td>
<td>−0.59</td>
<td>0.028</td>
</tr>
<tr>
<td>ENSGACG00000019291</td>
<td>irak3</td>
<td>Signal transduction</td>
<td>Spleen</td>
<td>0.42</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Protein phosphorylation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSGACG00000001978</td>
<td></td>
<td>Antigen processing and presentation</td>
<td>Spleen</td>
<td>−1.44</td>
<td>0.048</td>
</tr>
</tbody>
</table>

*Positive values represent higher expression in lake fish than in river fish and vice versa.
Table 3 Differentially expressed genes between lake and river populations also found as differentially expressed in previous parasite infection studies

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene name</th>
<th>Comparisons in Lenz et al. 2013*</th>
<th>Tissue in Lenz et al. 2013</th>
<th>Log fold-change in Lenz et al. 2013†</th>
<th>Tissue in this study</th>
<th>Log fold-change in this study‡</th>
<th>FDR in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSGACG00000011746</td>
<td>fyco1a FYVE and coiled-coil domain containing 1a sax7</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>−3.21</td>
<td>Head kidney</td>
<td>0.71</td>
<td>0.0096</td>
</tr>
<tr>
<td>ENSGACG00000010806</td>
<td>SRY-box containing gene 7</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>−2.58</td>
<td>Head kidney</td>
<td>0.74</td>
<td>0.038</td>
</tr>
<tr>
<td>ENSGACG00000013129</td>
<td>MRPLA9 (2 of 2) mitochondrial ribosomal protein L49 bno1</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>−2.69</td>
<td>Head kidney</td>
<td>−0.91</td>
<td>0.017</td>
</tr>
<tr>
<td>ENSGACG00000015653</td>
<td>LIM domain only 1 met113 Methyltransferase-like 13</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>−1.20</td>
<td>Head kidney</td>
<td>1.13</td>
<td>0.033</td>
</tr>
<tr>
<td>ENSGACG00000014705</td>
<td>Pancreatic progenitor cell differentiation and proliferation factor a</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>−4.4 and −2.69</td>
<td>Spleen</td>
<td>−0.63</td>
<td>0.028</td>
</tr>
<tr>
<td>ENSGACG00000011977</td>
<td>Pancreatic progenitor cell differentiation and proliferation factor a</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>1.03</td>
<td>Spleen</td>
<td>3.17</td>
<td>0.011</td>
</tr>
<tr>
<td>ENSGACG00000010923</td>
<td>N-6 adenine-specific DNA methyltransferase 2</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>1.56</td>
<td>Spleen</td>
<td>−0.88</td>
<td>0.025</td>
</tr>
<tr>
<td>ENSGACG00000004515</td>
<td>Cfd Complement factor D (adipsin)</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>−1.48</td>
<td>Spleen</td>
<td>1.16</td>
<td>0.00065</td>
</tr>
<tr>
<td>ENSGACG00000012609</td>
<td>LGALS1 (2 of 3) lectin, galactoside binding, soluble, 1</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>−5.09</td>
<td>Spleen</td>
<td>0.73</td>
<td>0.0038</td>
</tr>
<tr>
<td>ENSGACG00000011683</td>
<td>Solute carrier family 5, member 6</td>
<td>Control vs. infected</td>
<td>Head kidney</td>
<td>−2.46</td>
<td>Spleen</td>
<td>−0.45</td>
<td>0.045</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene name</td>
<td>Comparisons in Haase et al. 2014†</td>
<td>Tissue in Haase et al. 2014‡</td>
<td>Log fold-change in Haase et al. 2014†</td>
<td>Tissue in this study</td>
<td>Log fold-change in this study‡</td>
<td>FDR</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------</td>
<td>----------------------------------</td>
<td>-----------------------------</td>
<td>--------------------------------------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>ENSGACG0000003716</td>
<td>CASQ2 (1 of 2) calsequestrin 2</td>
<td>Control vs. clone XII</td>
<td>Head kidney</td>
<td>2.8</td>
<td>Spleen</td>
<td>-0.97</td>
<td>0.0080</td>
</tr>
<tr>
<td>ENSGACG0000017615</td>
<td>Smox</td>
<td>Control vs. clone I, control vs. XII and control vs. clone mix</td>
<td>Head kidney</td>
<td>4.79, 5.81 and 5.11</td>
<td>Spleen</td>
<td>0.65</td>
<td>0.011</td>
</tr>
<tr>
<td>ENSGACG0000011977</td>
<td>ppdpfa Pancreatic progenitor cell differentiation and proliferation factor a</td>
<td>Control vs. clone I</td>
<td>Head kidney</td>
<td>-4.08</td>
<td>Spleen</td>
<td>3.17</td>
<td>0.011</td>
</tr>
<tr>
<td>ENSGACG0000004966</td>
<td>IRF4 (2 of 2) interferon regulatory factor 4b</td>
<td>Control vs. clone I, control vs. XII and control vs. clone mix</td>
<td>Head kidney</td>
<td>1.41, 2.26 and 1.79</td>
<td>Spleen</td>
<td>-0.59</td>
<td>0.028</td>
</tr>
<tr>
<td>ENSGACG0000020628</td>
<td>angptl5 Angiopoietin-like 5</td>
<td>Control vs. clone mix</td>
<td>Head kidney</td>
<td>2.83</td>
<td>Spleen</td>
<td>0.71</td>
<td>0.028</td>
</tr>
<tr>
<td>ENSGACG0000017656</td>
<td>SVIL (2 of 2) supervillin apml</td>
<td>Control vs. clone mix</td>
<td>Head kidney</td>
<td>2.96</td>
<td>Spleen</td>
<td>0.59</td>
<td>0.040</td>
</tr>
<tr>
<td>ENSGACG0000008510</td>
<td>Actinoporin-like protein</td>
<td>Control vs. clone mix</td>
<td>Gill</td>
<td>-1.20</td>
<td>Head kidney and spleen</td>
<td>0.88 and 1.05</td>
<td>0.0093 and 0.0030</td>
</tr>
</tbody>
</table>

For the DE genes Haase et al. 2014 identified in gill, only the overlapped DE genes we identified in both head kidney and spleen are shown.

*Comparisons where the genes were previously identified as differentially expressed are indicated. In Lenz et al. 2013, DE gene sets between control naïve fish from lake and from river, between twice exposed fish from lake and from river (2nd infection), and between infected fish and control fish were compared to DE gene sets in this study. In Haase et al. 2014, DE gene sets between control fish and infected fish with different parasite clones were compared.

†In lake-river comparisons, positive log fold-change values represent higher expression in lake fish and vice versa. In control-infection comparisons, positive values represent upregulation with infection compared to control.

‡Positive log fold-change values represent higher expression in lake fish and vice versa.
contribute to shape the gene expression profiles of lake and river sticklebacks. Setting our study into the context of previous findings, we further evaluated these explanations. Using the same individuals from this study (as well as additional individuals), recent genomic studies have shown little evidence for sequence-based habitat-specific patterns using genome scan approaches with single nucleotide polymorphisms (SNPs; Feulner et al. 2015) and with copy number variations (Chain et al. 2014). Hence, from a genomic perspective, despite significant differentiation between lake and river sticklebacks at a regional scale and across a wider continental scale (Deagle et al. 2012; Roesti et al. 2012; Feulner et al. 2015), there is little evidence for parallel genetic differentiation between lake and river sticklebacks across the distribution area of the fish. In other words, genetic differences between freshwater ecotypes of sticklebacks are for the large part not shared across population pairs, whereas here we identified several genes with habitat-specific gene expression. This discrepancy is consistent with the observation that phenotypes are similar among lake-river populations while the genetic basis is different (Deagle et al. 2012; Kaeuffer et al. 2012; Feulner et al. 2015). Gene expression, which bridges the underlying genetic basis and the ultimate phenotypes, might contribute to the understanding of the discrepancy between phenotypes and genotypes. Habitat-specific expression patterns could be controlled by various trans-regulatory elements from different genomic sources in different populations. Another explanation is that pathways regulating expression might be triggered at different steps in signalling cascades and therefore leave distinct signatures in the genomes of different populations (Pritchard et al. 2010). Based on controlled laboratory studies, there is evidence that expression differences in sticklebacks can be largely heritable (Leder et al. 2014). In addition, a laboratory-controlled experiment in which laboratory-bred G1_L and G1_R sticklebacks exhibited different transcriptional responses to parasite exposure suggested that the genetic background plays an important role in differential gene expression between fish ecotypes (Lenz et al. 2013). It is interesting that this differentiation between lake and river fish was most pronounced in their adaptive immune response (triggered upon 2nd exposure) to parasites, most likely resembling the differences we are observing in nature, where the fish are very likely to have multiple encounters with parasites. In the light of these studies, adaptive genetic differences between lake and river sticklebacks appear to be a likely explanation for habitat-specific expression patterns. However, a reciprocal transplant experiment suggested that environmentally induced plasticity strongly affects the expression of some carefully selected immune genes (Stutz et al. 2015). Hence, plasticity in gene expression might have also shaped the habitat-specific expression pattern of some of the genes identified in this study.

Immunological relevance of DE genes

Large-scale observational studies such as the current one are complementary to experimental studies in general, and here to the stickleback system in particular. Previous studies on sticklebacks in German lake-river systems highlighted that lake fish harbour higher parasite loads than river fish in terms of intensity and species diversity (Kalbe et al. 2002; Eizaguirre et al. 2011, 2012b). This trend of contrasting parasite loads was further confirmed across a wide geographic range including all populations used in our study (Feulner et al. 2015). Experiments have established that lake and river sticklebacks have differences in immune-competence due to habitat-specific adaptation to the distinct parasite communities (Scharsack et al. 2007). It was further investigated that genetic differences in MHC genotypes between lake and river fish provide a basis for parasite-mediated local adaptation (Eizaguirre et al. 2011, 2012a) following the idea that parasite resistance could represent a magic trait involved in speciation (Eizaguirre et al. 2009). As the differences in parasite pressure between niches could be a force driving divergent adaptation in lake and river sticklebacks, we surveyed gene expression in immune tissues with a specific focus on genes involved in immune functions. Across the 139 candidate genes, we found three putative immune genes in the head kidney and five in the spleen with habitat-specific expression patterns (Table 2). We found that genes with an immune function were not overrepresented, which indicates that under natural conditions, other factors besides parasites and immunity also contribute to the differentiation between ecotypes. The overrepresented GO terms from these habitat-specific expressed genes suggest the gene products are often extracellular components, such as collagen-structured proteins. Given the generic GO terms, their contribution to habitat-specific adaptation is open to speculation. Nevertheless, a detailed examination of the DE genes showing most significant expression differences (with smallest adjusted P-values) between lakes and rivers revealed some associations with immune-related functions. One of the genes that is highly expressed in lake fish and differentially expressed in both the head kidney and in the spleen is colony-stimulating factor 1b (csf1b), which is involved in macrophage production and differentiation (Stanley et al. 1976). Another DE gene in the head kidney which is highly expressed in lake fish, leucine-rich repeat containing 17 (lrc17), regulates osteoclasts in mice cells (Kim et al. 2009). The
repeated domain of this gene is involved in a variety of protein-protein interactions, including binding to pathogen-associated molecular patterns and surface receptors and thus has been studied in pathogen–host interactions (Kedzierski et al. 2004). Some DE genes with putative immune functions are in contrast more highly expressed in river fish. For example, an uncharacterized protein-coding gene (ENSGACG00000000187) is differentially expressed in both head kidney and spleen, and its sequence is homologous to NOD-like receptor family CARD domain containing 3 (NLRC3). NLRC3 is a negative regulator of innate immune signalling (Zhang et al. 2014), which inhibits the activity of T cells (Conti et al. 2005) and Toll-like receptor (Schneider et al. 2012). Another DE gene that is highly expressed in river populations in the head kidney is cub and sushi multiple domains 3 (csmd3), reported to be associated with periodontal pathogen colonization in human (Divaris et al. 2012). The putative immune-related function of these candidate habitat-specific genes is consistent with the hypothesis that parasites act as important selective agents driving differentiation between river and lake sticklebacks (Wegner et al. 2003; Scharsack et al. 2007; Eizaguirre et al. 2011, 2012b; Feulner et al. 2015).

To investigate how differences in parasite load between lake and river populations may be reflected in gene expression in the wild, we compared the set of DE genes with the DE gene sets identified in two previous parasite infection experiments performed on G1 stickleback populations. Despite using different conditions, sequencing technologies and bioinformatic analyses to identify DE genes, this exercise provides information on immune-related functions of DE genes given their putative role in parasite defence based on experimental studies. The two laboratory-controlled parasite exposure experiments that we compared with our results used three-spined sticklebacks subjected to infection with parasites that are found in their natural environment: the three parasites Diplostomum pseudospathaceum, A. crassus and C. lacustris in a study by Lenz et al. (2013), and D. pseudospathaceum in a separate study by Haase et al. (2014). An independent parasite survey performed on our own transcriptome-sequenced fish (Table S1, Supporting information) showed that lake fish have a significantly higher abundance of Diplostomum sp. than river fish (negative binomial GLM, z = −4.87, P < 0.001, see Fig. S2, Supporting information), whereas A. crassus did not show a habitat-specific pattern (binomial GLM, z = −0.075, P = 0.94) and the lake-specific parasite C. lacustris (Eizaguirre et al. 2011) was only found in one G1_L fish in our samples. Lenz et al. (2013) assessed gene expression in the head kidney following parasite infection carried out with one of the European population pairs (G1_L and G1_R) used in our study. Among the DE genes found in that study, methyltransferase like 13 (mettl13) was expressed at lower levels in the parasite-challenged fish compared to controls, and in lake vs. river individuals after a 2nd parasite infection. In our study, this same gene was also differentially expressed with lower expression in the lake populations in the spleen. These results suggest that mettl13 expression is downregulated when the fish are challenged with more parasites, for example in lakes vs. rivers. mettl13 is therefore an interesting candidate for mediating a differential expression between lake and river sticklebacks shaped by the contrasting parasite environment. These comparisons to experimental studies demonstrate another way of inferring functional insights of candidate genes, which goes beyond functional annotations based on sequence similarity with model organisms. These transcriptomic results are in line with the hypothesis that parasite-mediated selection contributes to lake and river population differentiation; however, it does not act alone but in interaction with other factors under natural conditions.

Limits of the study

Even though we have been able to gain insight into the role of gene expression in population differentiation, various factors confound the analysis of wild-caught animals. For instance, temporal variation in expression, genetic background differences and stochastic environmental fluctuations introduce variation at the transcriptomic level (Harrison et al. 2012; Lenz 2015). Because our samples are derived from different regions and have been caught at different times of the year, geographical and seasonal factors influenced the observed expression patterns. An important biotic aspect with respect to this study is that fish accumulate parasites from spring to autumn, and their immune system responds differently to early and to late parasite infections (Rohlenová et al. 2011). Furthermore, our study focused on macroparasites, but we acknowledge that there are more pathogens and factors in the natural environment that affect fitness, physiology and immune response. For example, it was found that gut microbiota composition in lake sticklebacks might contribute to shape the genetic polymorphism of MHC class IIb genes (Bolnick et al. 2014), a known genetic basis that vary between fish populations (e.g. Eizaguirre et al. 2011). Hence, microparasites most likely also impact the gene expression of the fish in their natural environments.

In addition, factors like temperature and light condition can vary substantially across geographical regions and seasons. Environmental factors cannot be controlled for sampling on large geographical scale and add noise
to the data, reducing the ability to detect habitat-specific patterns. However, for each location, parapatric lake and river fish were processed at the same time and alternatively dissected, minimizing the variation between lake and river fish within sampling locations. Despite analysing wild-caught individuals, the majority of our samples showed reasonable correlations between replicated individuals (same habitat, population and sex), resulting in an average Pearson correlation of 0.86. Moreover, including multiple lake-river contrasts can help to overcome some of the variance among wild-caught samples, as it is unlikely that environmental fluctuations would produce habitat-specific expression patterns across multiple individuals and populations by chance. Therefore, our results are conservative estimates of habitat-specific gene expression across the replicated systems.

Having a single population pair from Canada might also affect some results. As the Canadian populations were rather distinct from the other populations, we also conducted DE analyses only on the three European population pairs for a comparison. However, differential expression between lake and river in the two data sets (with and without the Canadian population pair) was significantly positively correlated and about half of the DE genes are found in both data sets (Table S4, Supporting information). Therefore, including one geographically distant population pair from Canada allows identifying habitat-specific patterns on a more global scale. It provides an opportunity to examine which genes show consistent habitat-specific expression patterns in fish across continents, forming a subset of the DE genes from all four population pairs (asterisks in Fig. 2).

As we studied the transcriptomic profiles of wild-caught fish, a large number of replication in terms of individuals and populations is required to accommodate environmental variations. This results into trading off sample size and sequencing depth. The Encyclopedia of DNA Elements (ENCODE) consortium recommends 30 million pair-end reads of length >30 nucleotides, in which 20–25 million reads are mappable to the genome for evaluating transcriptional profiles. In our study, the sequencing depths are generally 5× lower than the recommendation, limiting our ability to detect genes with low expression. When we used a more stringent cut-off to filter out weakly expressed genes (at least two reads per million in half of the samples), 10 715 genes (compared to 12 183) in the spleen passed the filtering step. A total of 36 of 73 DE genes in the head kidney and 58 of 74 DE genes in the spleen remained with the higher cut-off, suggesting at least half of the detected DE results are robust against the low sequencing depth.

Conclusions and prospects

Despite some intrinsic shortcomings, studying gene expression in wild-caught animals provides a view on differential expression responses caused by both genetic and environmental factors. Our study provides additional evidence that environmental differences, which contrast lakes and rivers and among those the distinct parasite community, shape differential gene expression patterns in sticklebacks. We utilize results of previous laboratory-controlled experiments to explain the patterns we detected in the wild. This comparison suggests that among other factors the distinct parasite community is most likely an important explanatory factor causing expression differences between habitats. Our results add to previous laboratory results by examining the expression patterns of candidate genes under natural conditions. Those genes identified both here and in previous laboratory studies deserve special attention in potential follow-up studies.

Acknowledgements

We thank Derk Wachsmuth for computational assistance and MPI technicians, especially Anja Hasselmeyer for help with laboratory work. We thank Anika Witten for help with RNA-Seq library preparation and sequencing. We thank Ludovic Mallet for discussions on the study. We thank the International Max Planck Research School for Evolutionary Biology for research support. We also acknowledge the work and generosity of sample collectors including Andrew Hendry, Renaud Kaeuffer, Shahin Muttalib, Caroline Leblond, Noémie Erin, Per Jakobsen, Tom Klepaker and Hendrik Schultz. We thank Sean Rogers and five reviewers for their insightful and constructive comments on a previous version of this manuscript.

Reference


Kueffer R, Peichl CL, Bolnick DI, Hendry AP (2012) Parallel and nonparallel aspects of ecological, phenotypic, and
Lucek K, LeMoine M, Seehausen O (2014) Contemporary evolutionary divergence during a recent range expansion was facilitated by adaptive introgression. Journal of Evolutionary Biology, 27, 2233–2248.

© 2016 The Authors. Molecular Ecology Published by John Wiley & Sons Ltd.
and transcription factor NF-kappaB. *Nature Immunology, 13*, 823–831.


Zhang L, Mo J, Swanson KV et al. (2014) NLRC3, a member of the NLR family of proteins, is a negative regulator of innate immune signaling induced by the DNA sensor STING. *Immunity, 40*, 329–341.

M.M. funded the project. F.J.J.C., C.E. and M.K. organized and contributed to the sample collections and dissections. F.J.J.C., I.E.S., M.S. and P.G.D.F. prepared the RNA samples. Y.H. performed quality assessment of the sequencing data, transcriptome mapping and processed data for analysis. Y.H., F.J.J.C., M.P. and P.G.D.F. designed the differential expression analyses, and all authors contributed to discussions on research design and results interpretations. Y.H. drafted the manuscript together with F.J.J.C. and P.G.D.F. All authors revised the manuscript.

Data accessibility

The raw reads of RNA-Seq data (fastq format) and mapping files (bam format) are available through the European Nucleotide Archive (study Accession no.: PRJEB8677, URL: http://www.ebi.ac.uk/ena/data/view/PRJEB8677). HTseq read counts, EDGER results and topGO results are archived in Dryad (doi:10.5061/dryad.hq50s). Morphological and parasite data are included in Table S1 (Supporting information).

Supporting information

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

Fig. S1 Distribution of dispersion values in (a) head kidney samples and (b) spleen samples.

Fig. S2 Counts of *Diplostomum* sp. across stickleback populations.

Table S1 Morphological data and parasite loads of sequenced fish

Table S2 Summary of library statistics in sequence quality filtering, sequence mapping and library scaling factor.

Table S3 Differentially expressed genes between lake and river across four population pairs.

Table S4 Differentially expressed genes between lake and river in European populations.