Predominance of *cis*-regulatory changes in parallel expression divergence of sticklebacks.

Jukka-Pekka Verta [1, 2, *], Felicity C. Jones [1, *]

1. Friedrich Miescher Laboratory of the Max Planck Society, Max-Planck-Ring 9, 72076 Tübingen, Germany
2. Organismal and Evolutionary Biology, University of Helsinki, Viikinkaari 9, 00790, Helsinki, Finland

*Correspondence: fcjones@tue.mpg.de, jukka-pekka.verta@helsinki.fi

Keywords: parallel evolution, adaptation, selective sweeps, allele specific expression, genetic assimilation, dominance, additivity/mode of inheritance, Haldane's sieve, epistasis

Abstract

Regulation of gene expression is thought to play a major role in adaptation but there are conflicting predictions for the relative importance of *cis*- and *trans*-regulatory mechanisms in the early stages adaptive divergence. Parallel evolution of marine and freshwater threespine stickleback fish provides an excellent opportunity to dissect whether the same molecular mechanisms underlie repeated adaptive divergence in gene expression. Using RNA-seq of four marine-freshwater ecotype pairs from Scotland and Canada, we first identified genes with parallel divergence in expression and show that these are found near previously reported adaptive loci and show a molecular signature of selection centered around the transcription start site. With allele-specific expression assays in F1 hybrids we next show that expression divergence is predominantly driven by *cis*-regulatory control in all four river systems, a pattern that is enriched in parallel divergently expressed genes. In particular, for genes whose expression is up-regulated in parallel among freshwater fish the quantitative degree of *cis*- and *trans*-regulation is also highly correlated, suggesting a shared genetic basis across populations. This stands in contrast to genes up-regulated in parallel in marine fish, whose degree of *cis*- and *trans*-regulation is less correlated and predictable. This observed asymmetry in parallelism in how genes are up-regulated in marine and freshwater fish can be explained by differences in the evolutionary contexts of the diverging ecotypes. Finally, we show that *cis*-regulation is predominantly additive and shows greater robustness to different in genetic backgrounds and environmental conditions. We argue that these features make *cis*-regulation well-poised for rapid adaptive divergence of gene expression under conditions of on-going gene flow. Combined our study highlights how natural selection on dispersed *cis*-regulatory elements can shape the adaptive landscape of the genome.

Summary

Adaptive evolution of gene expression may proceed via mutations influencing *cis*-regulatory elements such as promoters or enhancers or *trans*-acting factors such as transcription factors. There are conflicting predictions for the relative roles of *cis*- and *trans*-regulation in the early stages adaptive divergence-with-gene-flow. Populations that have evolved in parallel provide a powerful opportunity to look for general patterns and rules governing the evolution of gene expression. Here we show that parallel expression divergence in threespine stickleback fish is primarily driven by differences in *cis*-acting (allele-specific) gene regulation such as mutations in enhancers and promoters. We show that *cis*-regulation is predominantly additive and stable across genetic backgrounds and water salinities enabling *cis*-regulatory mutations to act as a motor for the evolution of traits in the early stages of divergence-with-gene-flow.

Introduction

The ability of organisms to rapidly adapt to new environments can be both facilitated and constrained by the underlying molecular basis and mechanisms operating at the genomic level. Significant advances in our understanding of the genomic basis of adaptive evolution include that adaptation is often polygenic and involves loci that are predominantly intergenic and putatively regulatory (1-4). Regulation of gene expression can be controlled by *cis*-acting regulatory elements that are linked to the transcript whose expression they regulate (e.g. promotors, enhancers) and act in an allele-specific manner, and *trans*-acting factors such as transcriptional regulator proteins that interact with both copies of the gene they regulate and are typically unlinked. Both theory and empirical data suggest (5) that adaptive evolution can favor *trans*-acting mechanisms, where a rapid change in phenotype can be driven by selection favoring changes in the expression or function of one *trans*-acting factor that in turn alters expression levels of numerous transcripts within a gene regulatory network (6, 7). Alternatively, selection may favor *cis*-acting mechanisms if it is more beneficial to change the expression of single transcripts at a time (8). *Cis*-regulatory evolution has been postulated to play an important role in adaptation: fine tuning the spatial and/or temporal expression of genes avoids the potentially deleterious effects of pleiotropic coding changes (8).

Cis- and *trans*-regulatory mechanisms are not mutually exclusive and adaptation is expected to promote co-evolution between *cis* and *trans*-acting mechanisms so that optimal gene expression levels are reached and maintained (9). Interdependence of *cis*- and *trans*-regulatory mechanisms has been hypothesized to act as a barrier for gene flow and contribute to incipient speciation: incompatible regulatory factors fail to promote optimal gene expression levels in hybrid progeny (10, 11). There is conflicting predictions and support for the relative importance of *cis*-, *trans*- and co-regulation in the early stages of intraspecific adaptive divergence. In this context, divergent adaptation to local environments often occurs in the face of ongoing gene flow. The shuffling effects of recombination will tend to dissociate co-evolved factors and selection will be more efficient on those that can maintain linkage-disequilibrium. Thus the advantage of a rapid adaptive response mediated via a small number of *trans*-regulatory mutations in a gene regulatory network, may shift to favor *cis*-regulatory architecture where co-evolved mutations are more closely linked to each other and the gene whose expression they regulate.

Parallel evolution provides a powerful context to explore the relative importance of *cis*- and *trans*-regulation in the early stages intraspecific adaptive divergence. Phenotypic parallelism arises when species or ecotypes evolve under similar selection pressures. As independent biological replicates of the evolutionary process it is possible to ask whether the same phenotype has evolved via the same or different molecular underpinnings. There are numerous examples of parallel evolution at the phenotypic level that also share parallelism at the genetic level: melanism has evolved across animal clades through changes in melanocortin signaling (12); climate adaptation in forest trees is associated with parallel genetic changes in few key genes (13); parallel evolution of life history traits in Pacific salmonid fish involve same regions of the genome (14). While regulatory changes seem to predominate in adaptation of natural populations we know little about the extent and parallelism in gene expression and its *cis*- and *trans*-regulation.

The threespine stickleback fish is an excellent system to address these questions. Following the retreat of the Pleistocene icesheet 10-20k years ago the parallel evolution of freshwater ecotypes from ancestral marine forms has occurred repeatedly and independently in thousands of populations across the Northern Hemisphere (15). Whole genome sequencing of marine and freshwater fish from multiple populations reveals that parallel freshwater adaptation has a genetic basis that is partly shared across populations, and involves reuse of standing genetic variation that is largely non-coding, suggesting that parallel evolution proceeds predominantly through changes in gene regulation (16). Forward mapping and functional dissection of two major morphological traits provide further support for the importance of gene regulation in stickleback adaptation. The parallel loss of bony armor plates in independent freshwater populations has evolved via reuse of an haplotype carrying a 3' regulatory mutation at the EDA locus (17, 18), while the parallel loss of the pelvic girdle has evolved via independent de novo deletions of a regulatory element at the Pitx1 locus (19, 20).

During adaptation to their divergent environments marine and freshwater sticklebacks have evolved differences in numerous morphological, physiological and behavioral traits. Two key divergent traits include their anadromous (migratory marine) versus resident-freshwater life histories and the associated ability to tolerate salt and freshwater water respectively. For this adaptation, the gill with multifunctional roles in osmoregulation and respiration is likely to be particularly important. Regulation of ion exchange genes in the teleost gill is required for maintaining homeostasis in different water salinities (21). In saline water, fish counteract water loss and ion gain by ion exclusion. In freshwater, fish compensate against ion loss and water gain by ion uptake. Expression changes in osmoregulatory genes has been linked to freshwater adaptation from anadromous ancestors in sticklebacks and other fish (22, 23). The genetic basis for these regulatory changes are not known.

We have previously shown that repeated parallel evolution of freshwater sticklebacks from marine ancestors involves reuse of preexisting genetic variation at ~81 predominantly intergenic and putatively regulatory loci across the genome (16), comprising 500kb and ~0.1% of the genome in total. These genomic regions of parallel and highly divergent marine and freshwater haplotypes have the potential to harbor gene regulatory elements and provide a possible mechanism for divergent adaptation via parallel gene expression changes. This would be detectable as parallel divergence in gene expression levels and gene-regulation in independently evolved marine and freshwater ecotypes regardless of their geographic origin. In addition, previous studies have shown that freshwater adaptation in sticklebacks and other fish is associated with changes in gene expression plasticity (24-26) consistent with genetic assimilation. The mechanisms by which loss of plasticity is coded as heritable genetic variation remain unclear but one predictable feature of these adaptations would be regulatory control of expression that is environmentally insensitive.

Here we study the evolution of gene expression and its *cis*- and *trans*-regulation in the gills of threespine sticklebacks as a model for regulatory evolution during early stages of parallel adaptive divergence with gene flow. Using laboratory bred strains of independently evolved freshwater-resident and anadromous marine sticklebacks from rivers in Scotland and the Pacific coast of Canada we ask to what extent parallel

divergent adaptation to marine and freshwater environments involves heritable and parallel divergence in gene expression in the gills when reared divergent ecotypes are reared under the same standard laboratory conditions. We explore the genomic distribution of these differentially expressed genes with respect to the previously described high-resolution genome-wide map of adaptive loci, and whether natural selection around differentially expressed genes has influenced the genomic landscape of variation and left detectable molecular signatures of selection. We then dissect the *cis*- and *trans*-regulatory basis of gene expression differences using allele specific expression analysis of marine-freshwater F1 hybrids and their parents. We ask whether *cis*- or *trans*-regulatory changes predominate in the early stages of adaptive divergence with gene flow, and, by comparisons across marine and freshwater ecotypes from four independently evolving river systems examine the degree of sharing in *cis*- and *trans*-architecture. Finally, we explore the extent to which the *cis*-regulation of divergently expressed genes is influenced by water salinity.

Our results indicate that parallel transcriptomic divergence involves a small subset of genes that are involved in physiological adaptation to water osmolarity. Genes with divergent expression levels in the gill are linked to genome regions showing parallel genetic divergence and have stronger molecular signatures of selection than non-differentially expressed loci confirming the gill organ to be an important target of selection in the divergent adaptation of marine and freshwater sticklebacks to their respective environments. Importantly, marine freshwater expression divergence has a predictable genetic basis with a predominance of *cis*-regulatory control in all four river systems examined and an enrichment of *cis*-regulation of parallel divergently expressed genes. At genes with strong parallel upregulation in freshwater fish we find both *cis*- and *trans*-regulation to be positively correlated while genes with strong parallel upregulation in marine have less correlated *cis*- and *trans*-regulation among populations. Our results further suggest that for genes with a plastic expression response to differences in salinity the degree of *cis*-regulation is not dependent on the environment. We propose that the additivity and environmental stability of *cis*-regulation makes this mechanism particularly suitable in adaptive evolution and hypothesise this to be a key mechanism underlying genetic assimilation of plastic expression responses. Combined our study highlights how natural selection on dispersed *cis*-regulatory elements can shape the adaptive landscape of the genome.

Results

Stickleback gill transcriptome assembly from four replicate marine and freshwater stickleback strains. We utilised marine and freshwater strains of sticklebacks derived from each of four independent river systems in Scotland and Canada (Fig 1a, Table S1). Gills of mature and reproductively active female and male fish were dissected and their transcriptomes analyzed using strand-specific RNA-seq. We built a reference-guided assembly (27) of the stickleback gill transcriptome based on RNA-seq reads from 10 freshwater and 10 marine fish from 4 marine and 4 freshwater strains (Table S2). The stickleback gill transcriptome contains 29295 transcribed loci, 17304 of which are multi-transcript loci with 171620 different transcripts combined (considerably more than Ensembl gene build 90: 22456 loci, 29245 transcripts). The number of transcripts per locus is highly skewed with a median of 3 and a small number of loci, including genes with immune function, with very high numbers of splice forms (Fig S1). The transcriptome includes 7147 novel transcribed

loci that do not overlap with any transcript in Ensembl gene build v90. Of these novel loci, candidate coding regions with complete open reading frames and likelihood scores >20 were identified for 1018 using TransDecoder (28). At the locus level our assembly shows very high sensitivity (Sn=81%) with few false negatives (fSn=100%) and moderate specificity reflective of the appreciable number of novel coding regions we detected relative to the Ensembl gene build 90 (Sp=59, fSp=69). Considering the raw FPKM data 21399 (73%) of loci were expressed at FPKM>=1 in at least one of the 20 marine and freshwater fish analyzed, and 16195 (55%) in more than ten individuals. Hierarchical clustering of expression levels revealed that the gill transcriptome can be characterized by five major groups of loci according to their average expression level (Fig S2) with the most highly expressed group of genes showing strong enrichment for biological processes with the respiratory function of the gill including mitochondrial respiration, ATP synthesis coupled proton transport and cytoplasmic translation (Supplementary note).

Detecting parallelism in marine-freshwater transcriptome divergence in a largely non-parallel evolving transcriptome. We hypothesized that selection for freshwater adaptation through gene expression changes would influence multiple freshwater strains independently of their geographic origin, and would be observed as parallel divergence in gene expression levels. To test this hypothesis, we first investigated the major sources of covariation in the freshwater and marine transcriptomes using Principal Component Analysis (PCA).

While the first major axis of variation separates individuals by river system (24% variation explained Fig S3), we identified PC2 and PC5 as major axes of variation that when combined capture divergence in the transcriptomes of freshwater and marine ecotypes (14.5% and 6.3 PVE respectively, Fig 1c). Similar to the proportionally small amount of parallel divergence observed at the genomic DNA level, this analysis indicates that under common environmental conditions parallel gene expression divergence represents only a small proportion of transcriptome variation.

Using loadings on PC2 and PC5 we defined a composite PC that captures the parallel dimension of freshwater-marine expression divergence (Fig 1c), and is highly correlated with mean freshwater-marine expression fold change in Tyne and Little Campbell (Fig S4). We identified loci with the highest contribution towards parallel marine-freshwater expression divergence (composite PC loadings falling in top or bottom 1%, N=586 transcripts), which we refer to as '*parallel diverged loci*'.

In addition to PCA we performed parametric tests for loci with differential expression contrasting combined freshwater and combined marine individuals from Litc and Tyne strains and identified 120 loci (Fig S5, FDR 20%; N=47 at FDR<1.8%) that showed both significant differences in mean and consistent direction (sign) of divergence between ecotype strains from different river systems. The differentially expressed genes tended to have large composite PC loadings (Fig S5). Top ranking genes include Na-Cl cotransporter (slc12a10), Basolateral Na-K-Cl Symporter (slc12a2), cation proton antiporter 3 (slc9a3.2), Potassium Inwardly-Rectifying Voltage-Gated Channel (kcnj1a.3), potassium voltage-gated channel (KCNA2), Epithelial Calcium Channel 2 (trpv6), Sodium/Potassium-Transporting ATPase (atp1a1.4), aquaporin 3a (aqp3a) — genes known to play a role in osmoregulation in fish and other organisms. This set of differentially expressed loci also include a microRNA (mir-182), 31 loci that

are annotated in previous Ensembl gene builds but have unknown function inferred from protein homology to other organisms, and 30 entirely novel loci that have no overlap with gene annotations from Ensembl gene build 90.

Both PCA and parametric test identified parallel evolving transcripts that were overrepresented in gene ontology processes and molecular functions associated with gill ion exchange, osmoregulation and blood traits (Table S3). This is consistent with our hypothesis that adaptive expression divergence influences physiological functions of the gill associated with a transition to permanent freshwater environment. Among the overrepresented categories were multiple processes involved in transmembrane transport, suggesting that the parallel diverged transcripts function in regulating osmolarity through ion exchange.

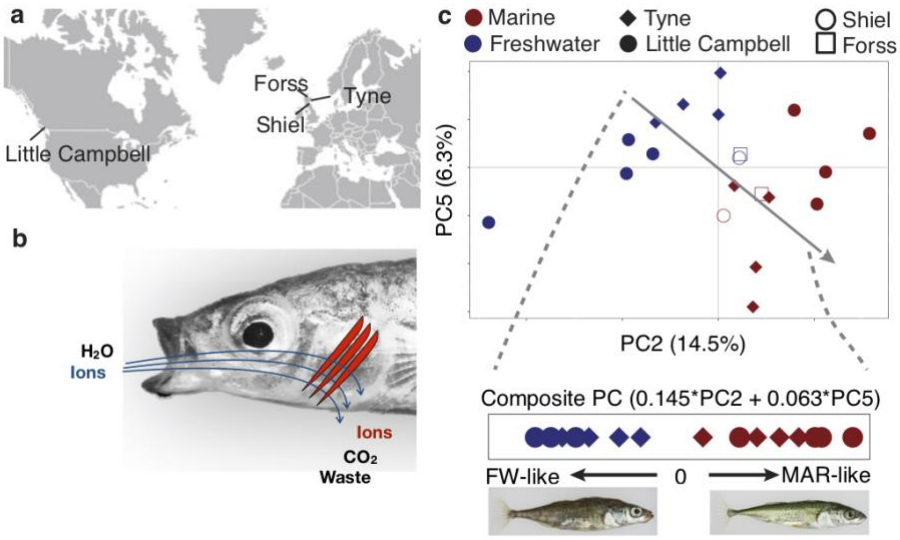

Figure 1. Freshwater and marine sticklebacks show parallel expression divergence among a largely non-parallel evolving transcriptome. (**a**) Marine and freshwater strains sampled from 4 different river systems. (**b**) The gill is a multifunctional organ with roles in osmoregulation, respiration and waste excretion. In freshwater gills uptake ions (blue), while in saline water ions are excreted (red). (**c**) Principal components analysis of normalized expression levels separates marine (red) from freshwater (blue) ecotypes along a composite PC axis (grey line). PCA is calculated based on a sample-size balanced set of Tyne and Little Campbell samples (solid symbols), onto which Forss and Shiel individuals are projected (open symbols).

Natural selection on parallel expression divergence. To explore the role of natural selection in parallel expression divergence, we used adaptive loci identified in a previously published study (2) and analysed newly generated whole genome sequence data of six unrelated fish of each ecotype from both Tyne and Little Campbell for molecular signatures of selective sweeps.

Transcripts with parallel expression divergence are distributed across all 21 chromosomes (Fig 2a), and show significant proximity to genomic regions undergoing parallel marine-freshwater adaptive divergence at the DNA sequence level (based on adaptive loci identified from the whole genome sequencing of 21 marine and freshwater fish from across the Northern Hemisphere, (2); randomization test, $P<<0.025$; Fig2b-c). More than 13 percent of parallel evolving transcripts were situated

within 10kb of regions of parallel genetic divergence, with a steady increase in association with increasing distance from the TSSs (Fig 2c), consistent with the hypothesis that the non-coding genome region contains regulatory elements that drive the parallel divergence in expression. We found similar results when we calculated CSS based on the whole genome sequences generated in this study (Supplementary note, Fig S6).

Natural selection is expected to reduce the effective population size of a local genomic region leaving detectable molecular signatures of selection such as increased genetic divergence (F_{st}) and reduced diversity (Pi) around adaptive loci (29). We calculated genetic divergence and nucleotide diversity genome-wide, and for 400 kilobase (kb) regions centered on transcription start sites (TSSs).

In Little Campbell genomes we observed a reduction in nucleotide diversity (Pi) around the transcription start sites (TSSs) of transcripts with parallel expression divergence (Fig 2d-f). Reduced within-population diversity was accompanied by increased genetic divergence (F_{st}, relative to a genome wide average of 0.34) that was centered on TSSs with a slight upstream bias (Fig 2d). In contrast, although we observed a slightly increased F_{st} around the TSSs of transcripts with parallel expression divergence in the Tyne, we did not detect a reduction in nucleotide diversity relative to other transcripts (Fig S7), possibly due to comparatively softer and/or younger sweeps in this population. Analysis of parallel differentially expressed transcripts defined using parametric tests gives largely consistent results (Fig S8).

Combined, these results provide compelling evidence for natural selection acting on parallel differentially expressed transcripts in the gill. This confirms the gill organ to be an important target of selection in the divergent adaptation of marine and freshwater sticklebacks to their respective environments and is consistent with gene regulation contributing to increasing genome-wide adaptive divergence of marine and freshwater sticklebacks.

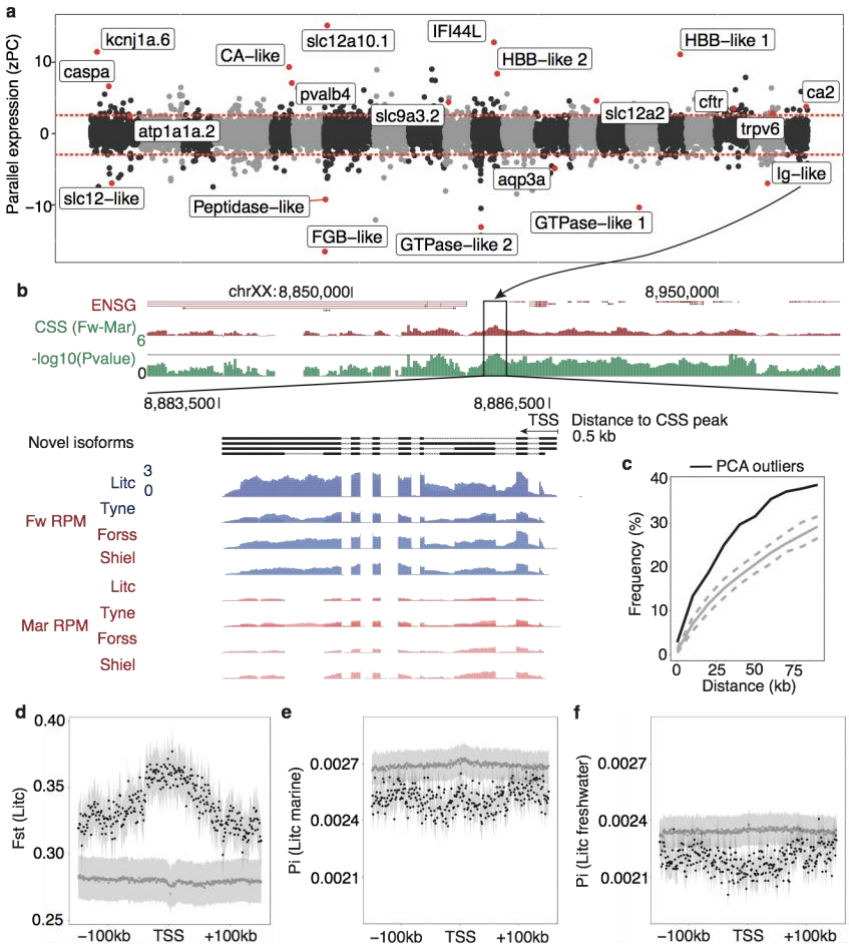

Figure 2. Transcripts with parallel marine freshwater expression divergence have osmoregulatory functions and molecular signatures of natural selection. (**a**) Composite PC loadings for each analyzed transcript. Consecutive chromosomes are depicted by alternating black and grey points. Dashed line corresponds to PC loadings falling in 1% highest or lowest quantiles. PC loadings are represented as z-scores (zPC). Highlighted genes are putatively involved in ion exchange (*kcnj1a.3*, *slc12-like, atp1a1.2, CA-like, ca2, pval4b, slc12a10.1, slc9a3.2, slc12a2, cftr, trpv6*), GTP signalling (*GTPase-like 1, GTPase-like 2*), respiration (*HBB-like 1, HBB-like 2*), blood clotting (*FGB-like, Peptidase-like*), jaw and gill morphogenesis (*caspa*), gill water permeability (*aqp3a*) and the immune system (*IFI44L, Ig-like*). (**b**) Ig-domain containing novel transcript that shows parallel expression divergence between freshwater and marine ecotypes in all four analyzed river-systems and is closely linked to a region of parallel genetic divergence. (**c**) Percentage of overlap between parallel evolving transcript loci and nearest repeatedly freshwater-marine diverged loci (freshwater-marine CSS (2); black) as a function of genomic distance. Solid grey line represents mean of all other transcripts (control), dashed grey lines represent 95% confidence intervals from 100 random samples of 586 transcripts. (**d-f**) Marine-freshwater genetic divergence (F_{st}) and nucleotide diversity (Pi) for 400kb region surrounding transcription start sites of parallel diverged transcripts in Little Campbell River. Black points represent average values of 10kb windows and grey shading the standard error of the mean across all parallel diverged transcripts. Grey points represent average values of 10kb windows of control loci (expressed loci showing non-parallel expression divergence). TSS; transcription start site.

Expression divergence between stickleback ecotypes is largely *cis*-acting. We hypothesized that the observed molecular signatures of selection around genes with divergent expression results from natural selection acting on *cis*-regulatory elements controlling gene expression levels. To investigate the role of *cis*-regulation, we compared the level of gene expression divergence in the gill transcriptomes of marine

and freshwater parents to the level of allele-specific expression (ASE) in their reproductively mature F1 hybrid offspring using laboratory reared freshwater and marine strains from four independent river systems (Tyne, Forss, Shiel and Little Campbell).

Since F1's carry a marine and freshwater copy of each chromosome within a common *trans*-nuclear environment, any marine or freshwater allele-specific bias in transcript expression can be attributed to *cis*-acting regulatory elements, rather than *trans*-acting factors. We whole-genome sequenced and identified polymorphisms across the genome where the parents of a given cross were homozygous for different alleles and then intersected these with the transcriptome assembly to identify loci that contained fully informative SNPs for allele-specific dissection (Fig S9, Fig S10). We defined regulatory divergence as *cis*-acting (allele-specific), *trans*-acting (non allele-specific), or a combination of the two, based on the allele-specificity and magnitude of expression divergence in F1 hybrids compared to freshwater and marine parents (Fig 3a-c, Table S4) (10, 30). Since we are interested in genetic variation natural populations, we used first generation wild derived parents and analyzed four F1 offspring per parental pair in order to account for the different genetic backgrounds of each offspring.

3807 (13%), 4472 (15%), 7716 (26%) and 10102 (34.5%) non-sex-biased genes could be assayed for *cis*- vs *trans*-mediated expression in at least one F1 in Tyne, Shiel, Forss and Little Campbell, respectively. In all ecotype-pairs, freshwater-marine expression divergence was predominantly *cis*-acting (Fig 3b-c, Fig S11), with the degree of *cis*-regulated expression divergence scaling positively with the degree of genetic divergence (measured as number of fixed differences between parents) between marine and freshwater ecotypes from each river systed. Whereas purely *cis*-divergence influenced the expression of ~350-900 transcripts in the Scottish ecotypes (~10-12% of analyzed genes), over 2000 transcripts were purely *cis*-diverged in Little Campbell (~20% of analyzed genes). We verified that the results were robust towards sequencing depth by downsampling to 30 million RNA-seq reads per sample (Fig S12).

The rate at which adaptive regulatory alleles rise into a high frequency in a population is influenced by the extent to which regulatory variation produces expression phenotypes that are additive or dominant/recessive (31). Selection will act more efficiently on additive trait variation because it is not obscured by dominance or recessivity of single allele copies. To investigate this, we used normalized read counts over informative SNPs to calculate the dominance/additivity ratio following Gibson et al (32). Treating each F1 offspring separately, additivity was calculated as the absolute value of half the difference between the parental mean transcript read count and dominance as the difference between the observed F1 transcript abundance and the midpoint of the two parents. *Cis*-regulatory divergence showed the strongest level of additivity among regulatory divergence types (Fig S13), consistent with evolutionary potential for fast allele-frequency changes at *cis*-regulatory elements. These findings are consistent with previous studies showing that *cis*-divergence is linked to higher additivity of between-species expression differences (31, 33).

Adaptation is also expected to favor non-epistatic alleles that confer a stable phenotype irrespective of the genetic background (5). We compared the level of

epistasis in different classes of regulatory divergence by analysing the reproducibility of regulatory divergence across siblings of the same F1 family. Each F1 is genetically unique because the parental alleles will have recombined in unique combinations. Epistatic regulatory divergence will therefore tend to be unique to each F1 individual, while non-epistatic regulatory divergence will be independent of the genetic background and observed in multiple F1's. *Cis*-regulatory divergence tended to be most stable across genetic backgrounds (Fig S14), indicating that *cis*-acting divergence was least influenced by epistasis. The results were consistent in both Tyne and Little Campbell crosses, indicating that our approach to test epistasis was robust to significant variation in genetic backgrounds.

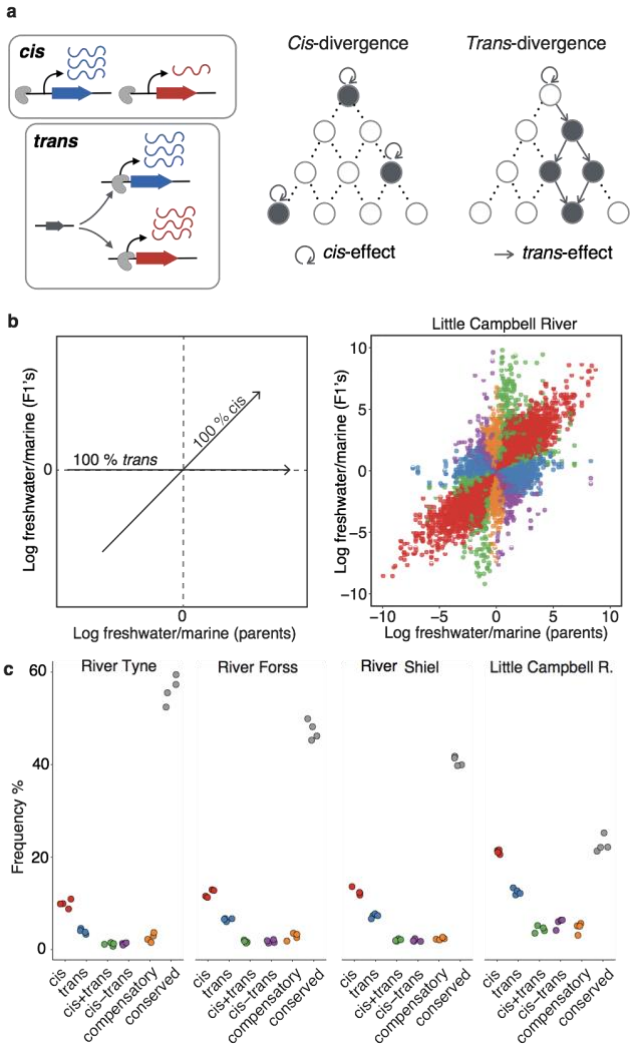

Figure 3. The genetic architecture of expression divergence between ecotype pairs is predominantly *cis*-acting. (**a**) In F1 hybrids, allele expression is measured in the same *trans*-acting background, allowing dissection of expression variation due to effects that co-segregate with the allele (such as enhancers and promoters) from effects that influence both alleles and can thus be attributed to *trans*-acting sources (such as transcription factors). *Cis* and *trans*-acting divergence is represented as effects that are contained within the nodes of a gene regulatory network (*cis*) or diffused from upstream regulators (*trans*). (**b**) Regulatory divergence is categorized in six categories by comparison of allele expression level and ratio in parents versus F1s (34). Individual data-points correspond to allele-specific expression values (y-axis) for each gene in each of four F1s relative to their parents (x-axis). Genes are colored according to their classification into different genetic architectures of expression divergence (Red *cis*, blue *trans*, green *cis+trans*, purple *cis-trans*, gold *compensatory*, grey

conserved. "*Ambiguous*" and "*conserved*" classes are omitted for clarity.) (**c**) Overall frequency of divergence architectures, relative to the number of analyzed transcripts. "*Ambiguous*" class is omitted for clarity.

Parallel expression divergence is mainly *cis*-regulatory. In order to test whether particular types of regulatory variants were more likely to contribute to parallel freshwater-marine expression divergence, we first investigated the genetic basis of expression divergence in transcripts that contributed most to parallel expression divergence on PCA, and that had informative SNPs that permitted distinguishing the parent alleles in F1's. Because the analysis was based on single individuals instead of a population sample as in our PCA analysis, we additionally required the marine-freshwater expression difference between the parents to be in the same direction between transcriptome sequenced pure marine and freshwater strains from the Tyne and Little Campbell. After filtering, 181 and 68 parallel evolving transcripts were testable in Little Campbell and Tyne, respectively.

Parallel evolving transcripts showed a significant excess of divergence architectures assigned as *cis*-acting and *cis*+*trans*-acting in Tyne and Little Campbell, compared to random expectations obtained by 1000 random draws of 528 genes from all genes analysed for allele specific expression (Fig 4). The most overrepresented regulatory type was *cis*-acting (18-29%, 20-22% above average, depending on the F1 individual). The overrepresentation of *trans*-acting divergence was lower, between 8-18% and 6-10%, respectively, and not different from random expectation in all F1's. Analysis of the overrepresentation of regulatory types using parallel differentially expressed genes defined through a parametric approach were consistent with our analysis based on PCA (Fig S15).

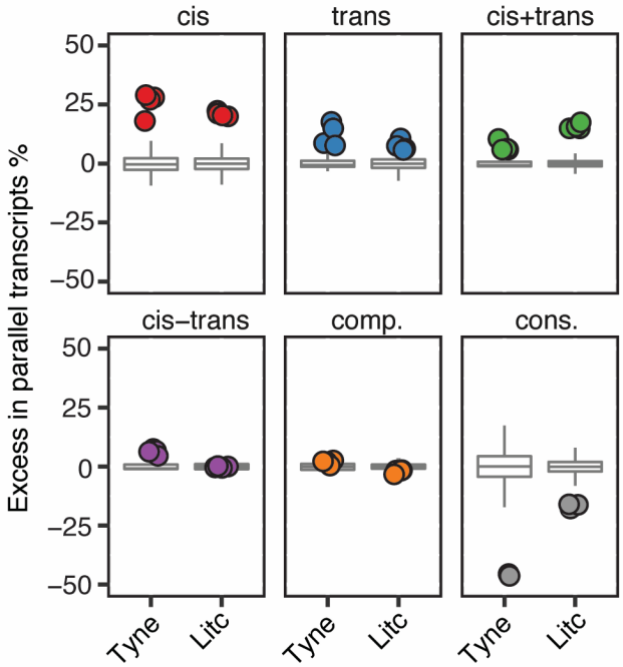

Figure 4. Parallel expression divergence is over-represented in *cis*-regulation. Overrepresentation of genetic architectures associated with parallel evolving transcripts, compared to the overall frequency of each regulatory divergence type. Points represent the observed values in four F1's per cross. Box-plots in background denote random expectation based on 1000 bootstrap samples of the corresponding number of transcripts. Horizontal bar in box-plot corresponds to median, box includes 50% and whiskers 99.3% of random expectation.

Both ecotype-pairs showed a significant excess of *cis*- and *trans*-regulatory divergence acting in the same direction (divergence in *cis*+*trans* co-regulation) in parallel diverging transcripts. Theory and experiments predict that directional selection on expression levels favors the accumulation of *cis*- and *trans*-regulatory changes that act in a manner in which they amplify expression divergence (9, 35). Between diverging lineages this is seen as *cis* and *trans*-regulatory divergence having collinear direction of effect. An excess of *cis*+*trans*-regulation, as seen in the transcripts that contribute to parallel expression divergence, is consistent with directional selection and inconsistent with the accumulation of regulatory differences through genetic drift, which should influence amplifying (*cis*+*trans*) and cancelling (*cis*-*trans*) divergence equally (9). Notably, the candidate genes slc12a2, cftr, atp1a1.2 and trpv6 that are necessary for sustaining teleost ion homeostasis as well as the aquaporin gene aqp3a were among the parallel diverged transcripts influenced by *cis*+*trans*-acting regulatory divergence.

We next explored the degree of parallelism in the strength of *cis*- and *trans*-regulation among the four populations at loci contributing to parallel expression divergence. For each F1 individual at each locus we used the F1 allele-specific expression ratio as a quantitative measure of *cis*-regulation, and subtracted the F1 allele-specific expression ratio from the parental expression ratio to obtain a quantitative measure of *trans*-regulation (36). For each river system we then averaged the *cis*-regulatory and *trans*-regulatory values for each locus across each of the four offspring to obtain a measure of the mean degree of *cis*- and *trans*-regulation per locus. We split this data into subsets of loci with increasing extremes of contribution to parallel marine freshwater expression divergence as measured by the loadings on the composite principal component axis. Parallel divergent loci that are upregulated in freshwater show strong positive correlations among populations in their quantitative extent of both *cis*- and *trans*-regulation (Fig 5a-b). This correlation is lost in loci that do not contribute to parallel expression divergence (composite PC loadings close to zero) and is less consistent among parallel divergent loci upregulated in marine fish. Finally, we used the absolute values of quantitative *cis*- and *trans*-regulation as an indication of the magnitude of *cis*- and *trans*-regulatory effects and calculated the mean magnitude for subsets of loci binned according to the degree of parallel expression divergence (composite PC loading). Divergently expressed genes upregulated in freshwater fish are regulated by *cis*- and *trans*-components of large effect size compared to divergently expressed genes upregulated in marine fish (Fig 5c, Fig S16).

Combined, these results support the hypothesis that natural selection has favored the same underlying *cis*- and *trans*-gene regulatory architecture when sticklebacks undergo parallel adaptation to freshwater environments. They suggest the predictability of the mechanisms regulating gene expression may depend on the environment and evolutionary context and may be related to differences in regulatory effect size controllling divergent gene expression in marine and freshwater fish.

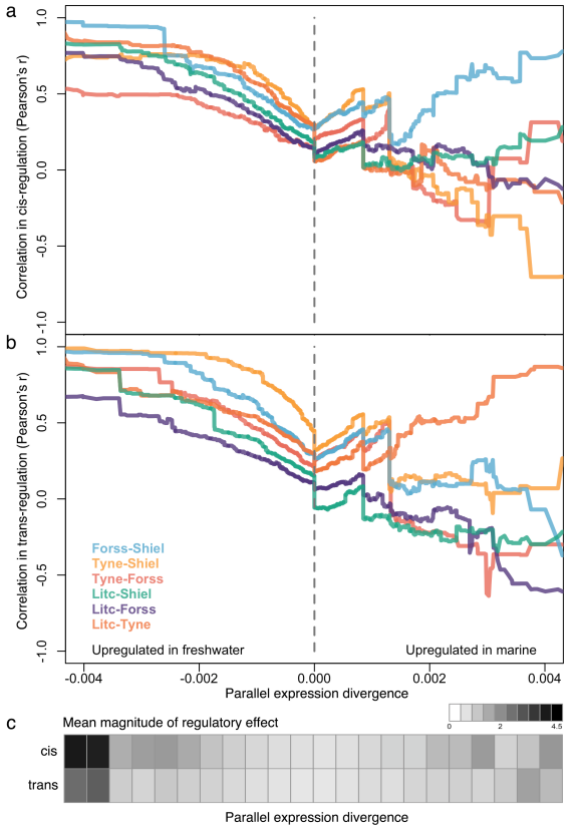

Figure 5. Degree of parallelism in *cis*- (a) and *trans*-regulation (b) at loci with increasingly parallel divergent expression (composite PC loadings). Parallel divergent genes that are upregulated in freshwater (strong negative loading on composite PC) show high correlation in their degree of *cis*- and *trans*-regulation among populations. In contrast, loci that have not diverged in parallel (PC loadings close to zero) and loci that are parallel upregulated in marine are not highly correlated in the magnitude of *cis*- and *trans*-regulation. For each population pair, Pearson correlation coefficients, r, were calculated for subsets of loci defined by an increasingly extreme positive or negative threshold on the composite PC loading scores. (c) The mean absolute magnitude of *cis*- and *trans*-regulatory effect across populations in subsets of loci defined by bins of parallel expression divergence (composite PC). For both *cis*- and *trans*-regulation higher magnitude effects (darker grey shades) are seen at parallel divergent loci that are upregulated in freshwater. Means and standard errors of effect size per population are shown in Fig S16.

***Cis*-regulatory effects are insensitive to environmental conditions.** In order to evaluate the possible mechanisms that may underlie *cis*-regulation as a predominant source of parallel regulatory differences, we investigated gene regulatory responses to water salinity, which is a major environmental contrast between freshwater and marine ecosystems. We first considered changes in gene expression in the gill transcriptome in response to water salinity because freshwater adaptation has been reported to associate with changes in gene expression plasticity (22, 23). We analyzed transcriptomic and *cis*-regulatory responses to water salinity using F1 hybrids which possess both freshwater and marine regulatory complements. A clutch of Tyne marine x freshwater F1 siblings were raised in standard laboratory 3.5 ppt salinity until 3 months old, separated into three and acclimated to 0.2 ppt (freshwater), 3.5 ppt (approximately equivalent to 10% sea water) and 35 ppt (marine salinity) over four months. At reproductive maturity we analyzed the gill transcriptomes of four F1's from each salinity using RNA-seq.

A total of 2542 transcripts were differentially expressed between at least two water

salinity treatments (FDR 1%), consistent with major changes in gill structure and function upon salinity acclimation (37). Differences between salinity treatments were mostly driven by freshwater acclimation (0.2 ppt), which elicited a response markedly different from both standard husbandry conditions (3.5 ppt) and typical marine salinity (35 ppt) (Fig S17). We observed that an unexpectedly high proportion of parallel diverged transcripts also showed an effect of salinity on expression level (97 of 586, hypergeometric test P<4.3e-6), supporting our conclusion that a large proportion of the parallel diverged transcripts in the gill are involved in physiological responses to water salinity.

Plasticity in expression of a given genotype in different environments can be mediated by either *cis* or *trans*-regulatory mechanisms. Focusing on the 97 transcripts with parallel expression divergence that showed a salinity response, we asked whether the regulatory control of these loci is sensitive to salinity conditions. We first calculated an expression profile for each F1 sibling by scaling the FPKM expression of each transcript to the average expression level across all individuals and then compared the profiles using Spearman correlation to capture the strength and direction of correlation in expression profiles among salinity treatments. F1's raised in similar salinities tended to have high positive correlation coefficients while the expression profiles of individuals raised in freshwater tended to be negatively correlated (opposite expression) with expression in individuals raised in salt (Fig 6, lower triangle).

Then, using allele-specific expression as a proxy for *cis*-regulation we investigated the stability of *cis*-regulation of expression across the salinity treatments. Despite the observed plastic response of gene expression (Fig 6, lower triangle), the degree of *cis*-regulation of salinity-responsive parallel diverged genes was highly correlated across salinities (Fig 6, upper triangle), indicating that the observed salinity response was not caused by *cis*-regulatory elements but largely due to *trans*-acting regulation that influenced the expression of both alleles in similar magnitude. Based on these results we hypothesise that *cis*-regulatory changes provide a mechanism for genetic assimilation of plastic responses into heritable variation (38) where the effects of regulatory alleles are independent of the environment.

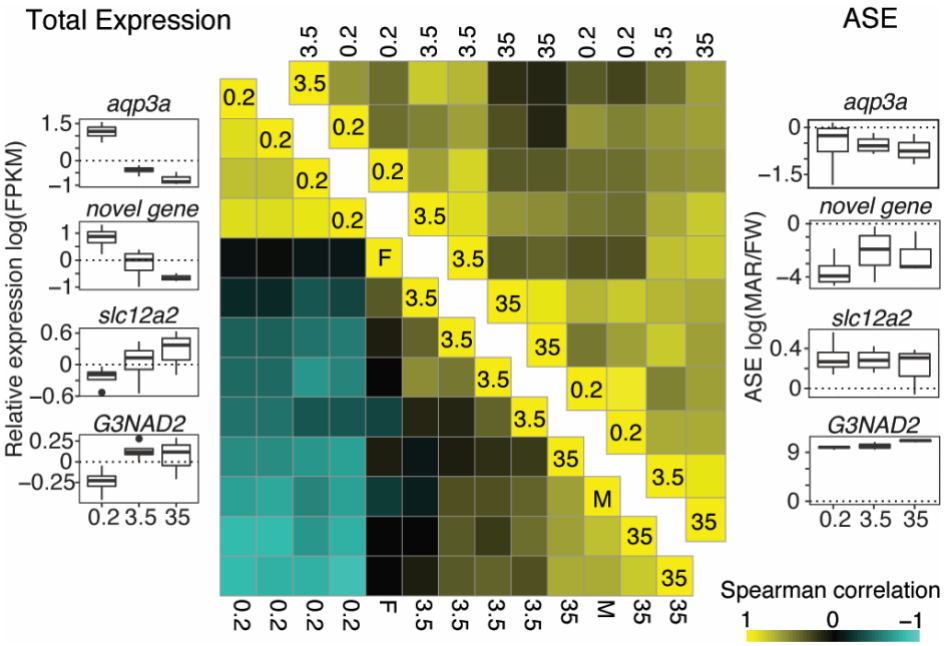

Figure 6. Parallel diverged transcripts respond to water salinity through *trans*-acting mechanisms. (**lower triangle**) Pairwise Spearman's correlation (*rho*) in expression levels of parallel diverged loci in marine x freshwater F1 siblings acclimated to different salinity conditions. A negative correlation indicates salinity-responsive genes were up-regulated in one salinity and down-regulated in another (such as *aqp3a* expression in 0.2 ppt samples and 35 ppt samples) while a positive correlation indicates that expression profiles were similar e.g. up-regulated in both (such as *aqp3a* expression in 3.5 ppt samples and 35 ppt samples. Examples of loci showing changes in expression in response to salinity are shown in the left panel and their salinity insensitive *cis*-regulation in the right panel. (**upper triangle**) Pairwise Spearman's correlation (*rho*) in *cis*-regulation of gene expression in parallel diverged salinity-responsive transcripts shows *cis*-regulation is insensitive to salinity conditions. For both upper and lower heatmaps, rows and columns are ordered based on euclidean distance. M, F refer to marine and freshwater parents. 0.2, 3.5 and 35 refer to freshwater, standard husbandry and marine salinity conditions (in parts per thousand).

Discussion

Regulation of gene expression is thought to play a major role in adaptation, yet relatively little is known about the patterns and predictability of adaptive regulatory mechanisms in the early stages of intra-specific adaptive divergence that evolve in the face of on-going gene flow. We characterised parallel expression divergence in the gill transcriptome - an organ with important respiratory and osmoregulatory functions - and dissected the *cis*- and *trans*-regulatory architecture in four marine-freshwater ecotype pairs in order to infer the rules and patterns shaping genome evolution and influencing rapid adaptation in natural populations.

Similar to the small proportion of the genome show parallel adaptive divergence at the DNA sequence level (2) we found that when sticklebacks are reared under the same standard husbandry conditions parallel transcriptomic divergence involves only a few hundred genes each with relatively small effect on expression divergence. Parallel divergently expressed genes are close to parts of the genome showing parallel divergence at the sequence level (identified in (2)), supporting a role for the reuse of ancient standing genetic variation in the parallel adaptive divergence of gene expression. We also observed a large proportion of loci with marine-freshwater divergent expression that was unique to a local river system (see Supplementary note)

indicating there is room for drift and/or local adaptation to play a major role shaping the evolution of divergent gene regulation and gene expression in any given river system. The transcripts showing strongest parallel expression divergence are associated with genes involved in gill osmoregulation, a key physiological trait in marine-freshwater divergence, in agreement with studies in other fish (22). Through population genomic analysis we show how natural selection on divergently expressed genes is shaping the evolution of the genome - strong molecular signatures of selection (elevated Fst, reduced Pi) were detected around the transcription start sites of parallel divergently expressed genes. We note that elevated Fst was markedly higher in at least 400 kb surrounding transcripts with parallel expression divergence, suggesting that transcripts showing parallel expression divergence is associated with genomic regions showing overall elevated genetic divergence between ecotypes.

Through analysis of allele-specific expression in F1 hybrids relative to their marine and freshwater parents, we found overwhelming evidence for the importance of *cis*-regulation of marine-freshwater divergent gene expression. The predominance of *cis*-regulation was observed in marine-freshwater ecotypes from four independent river systems, in analysis of all informative transcripts genome wide, and enriched in the set of loci identified as having evolved parallel marine-freshwater divergence in expression across rivers.

In animals and yeast, *cis*-regulatory differences contribute strongly to divergence in gene regulation over long evolutionary divergence scales (5, 31, 39-41). The stickleback ecotype pairs studied here evolved within the last 10-20000 years following the retreat of the Pleistocene ice sheet, indicating that strong *cis*-regulatory divergence can also evolve in relatively short timescales. While adaptation via re-use of ancient standing genetic variation is important in sticklebacks and may explain some part of the predominance of *cis*-regulatory changes, we also found that the extent of genomic divergence varied substantially between the parallel evolving stickleback ecotype-pairs we studied, and the proportion of *cis*-regulatory divergence scaled positively with this genomic divergence. This suggests that much of the genetic differences that accumulate in the early stages of adaptive divergence with gene flow translate to *cis*-regulatory differences. Mutation-accumulation experiments have shown that genetic drift, which promotes random regulatory changes, is biased towards *trans*-acting divergence due to a larger *trans*-mutational target size (42). Since most expression changes between freshwater and marine ecotypes however tended to be regulated in *cis*-, this points towards a stronger contribution of selection rather than drift.

Parallel adaptive divergence of marine and freshwater ecotypes provides biological replicates of the evolutionary process which we can interrogate to identify common patterns governing the molecular basis of adaptation. We not only identified parallelism across marine-freshwater ecotype pairs in the predominance of *cis*-regulation of divergent gene expression, but also parallelism in the quantitative extent of both *cis*- and *trans*-regulation of divergently expressed loci. This strong parallelism was particularly notable in parallel divergently expressed genes that are upregulated in freshwater, and considerably less strong and less predictable in parallel divergently expressed genes that are upregulated in marine. Similar to the underlying shared genetic basis of freshwater adaptation due to the parallel reuse of standing genetic variation, we can not only predict that freshwater populations are likely to carry the

same alleles at adaptive loci across the genome, but also that the magnitude and extent of *cis*- and *trans*-regulation of divergently expressed genes is likely to be shared among freshwater populations. Coinciding with this stronger parallelism in genes upregulated in freshwater we observed a tendency for the *cis*- and *trans*-regulatory effect size to be of larger magnitude than the *cis*- and *trans*-regulatory effects of genes upregulated in marine. It is plausible that this larger effect size in freshwater upregulated genes makes selection more efficient and contributes to the stronger parallelism in *cis*- and *trans*-regulation of freshwater upregulated genes. Further, the marine ecotype is thought to be the ancestral state in threespine sticklebacks and has been evolving in a comparatively stable marine environment for millions of years. Under this evolutionary context, freshwater populations have a smaller effective population size with reduced access to preexisting adaptive standing genetic variation, and have been subject to more recent and potentially stronger selection pressures than their marine ecotype counterparts who may have more opportunity to evolve via soft sweeps on standing genetic variation. Combined the evolutionary context is likely to have influenced which molecular mechanisms of gene expression regulation are seen, and efficiently respond to selection, resulting in relatively few paths to evolve adaptive gene expression in freshwater populations and comparatively more diverse mechanisms to achieve adaptive gene expression phenotypes in the marine population.

We note that our results contrast with a recent study by Hart et al (43) who reported a predominance of, and parallelism in the *trans*-regulatory control of marine-freshwater divergent gene expression in the pharyngeal tooth plate. One possible biological explanation for this might be differences in the multifunctional (pleiotropic) functional roles of the gill and its likely complex genetic architecture compared to dental tissue with a less pleiotropic functional role and more simple genetic architecture involve a few large effect loci upstream of other factors (44).

Our results indicate that parallel evolving divergence may converge on *cis*-regulation driven in part by a higher level of additivity and lower level of epistasis of *cis*-regulatory factors. Known as the effect of "*Haldane's sieve*", beneficial alleles sweep to high frequency fastest, and thus are most likely to become fixed, if they confer dominant or additive phenotypes (45). When alternative alleles are favored in their respective environments, as seem to be the case in sticklebacks, additive alleles have the highest likelihood of becoming fixed in both populations as both alleles can rise quickly to a high frequency. *Cis*-regulatory variation also tended to have a lower lever of epistasis within populations compared to other types of regulatory effects. Epistatic regulatory alleles tend to have different phenotypic effects depending on the genetic background, therefore inducing unpredictable fluctuation in expression levels. Hence, parallel evolution of gene expression seems to favor non-epistatic regulatory alleles that have similar effects on expression levels independently of the genetic background.

Our results further indicate that reinforcing *cis* and *trans*-acting regulatory variation that act in a manner to amplify one another has an important contribution to early ecotype divergence with gene flow. Overrepresentation in *cis*+*trans*-coevolution seemed to grow with increasing genetic and expression divergence between lineages, which is notable in the context of incipient ecological speciation. Given enough evolutionary time, the accumulation of amplifying *cis*+*trans*-regulatory divergence independently in diverging populations may lead to the evolution of genetic (Bateson-

Dobzhansky-Muller) incompatibilities. Because recombination between coevolved *cis* and *trans*-regulatory factors disrupts their combined phenotypic effects, it is expected that selection would favor linkage disequilibrium between the coevolved regulatory factors (46). Association between increasing *cis*+*trans*-regulation and genome-wide divergence as seen here suggests that selection against recombination between coevolved regulatory factors may contribute to increased genetic divergence as adaptation proceeds and thus shape the genomic landscape of incipient speciation.

Cis-regulation may play a particularly important role in the evolution of plasticity. Since plasticity is known to play a role in the divergent adaptation of marine and freshwater sticklebacks we investigated the sensitivity of gene expression and its *cis*-regulation to environmental salinity. Genetic assimilation involves the heritable encoding and loss of plasticity of a once plastic trait (38). Selection for reduced plasticity is predicted to evolve in response to strong selective pressures (5) and previous studies have found evidence for genetic assimilation in gene expression evolution between stickleback ecotypes (24). While we found a strong component of plasticity in gene expression among siblings raised in different salinities, the *cis*-regulation of this gene expression was stable and insensitive to differences in the environmental conditions. From this we infer that the observed plasticity in expression is likely mediated via *trans*-regulation, and hypothesise that the stable *cis*-regulatory component may serve as a mechanism for genetic assimilation.

The importance of *cis*-regulation in the early stages of adaptive divergence has implications for our understanding of genome function in natural populations. We have shown that *cis*-regulation predominates expression divergence, is additive, robust to differences in salinity, and potential epistasis caused by differences in genetic backgound. These features poise *cis*-regulation to be particularly responsive to selection and may explain parallelism in the predominance and quantitative extent of *cis*-regulation acoss populations. Combined our study highlights how natural selection can shape the adaptive landscape of the genome.

Data access

Data will be deposited to the Sequence Read Archive following manuscript acceptance. All scripts used in data analysis will be made available at https://github.com/jpverta/ following manuscript acceptance.

Acknowledgements

The authors acknowledge the Semiahmoo First Nation and Dolph Schluter for collaboration with sampling Little Campbell River sticklebacks and Saad Arif, Cecilia Martinez, Frank Chan, Paloma Medina and Li Ying Tan with stickleback sampling. We thank Stanley Neufeld, Enni Harjunmaa and Muhua Wang for helpful comments on the manuscript and Vrinda Venu for assitance in DNA library preparation and sequencing. FCJ is funded by ERC Consolidator Grant, Deutsche Forschungsgemeinschaft and the Max Planck Society.

Author contribution

JPV and FCJ designed the study. JPV conducted the experiments and analyzed the data. FCJ contributed to data analysis. JPV and FCJ wrote the manuscript.

Materials and Methods

Sample material. We captured freshwater-resident and anadromous marine sticklebacks from Little Campbell River, Canada, and Rivers Tyne, Forss and Shiel, Scotland, with wire-mesh minnow traps. Sampling locations are given in table S1. We identified freshwater resident and anadromous marine ecotypes based on their lateral plates. Most freshwater-resident populations of sticklebacks are plate-less, whereas anadromous marine forms exhibit complete lateral plating (15). Consistently, in most cases the large majority of fish captured in freshwater were plate-less and anadromous marine fish captured near the mouth of the river/lake were completely plated. We generated within-ecotype crosses via *in vitro* fertilisation of gravid females with males within ecotypes and transported the fertilized eggs to a common-garden environment at the Max-Planck campus in Tübingen in reverse-osmosis water supplemented with Instant Ocean salt to 3.5 ppt (~10% sea water salinity). The Max Planck Society holds neccessary permits to capture and raise sticklebacks. All animal experiments were done in accordance to EU and state legislation and avoiding unneccessary harm to animals.

The fish strains were raised in individual 100 liter tanks with 40-50 fish per tank in 3.5 ppt salinity and alternating light cycle of 16 hour light and 8 hour darkness of 6 months. No significant mortality occurred during transfer or captivity. Fish were raised with a diet of fry (freshly hatched artemia), juvenile (artemia Daphnia and cyclops) and adult food (bloodworm, white mosquito larvae, artemia, mysis shrimp and Daphnia) until adults. We selected four unrelated fish from independent field crosses per ecotype from Little Campbell River and River Tyne strains, and one fish per ecotype from Shiel and Forss strains. Exception to this was one Tyne freshwater male fish that was the progeny of unrelated lab-raised freshwater parents, which was included to complete the sampling. We in-vitro crossed one marine female and one freshwater male for each strain, and raised the F1 individuals in identical conditions (each cross in individual tank) as the parents until they were reproductively mature.

For the Tyne cross, we separated the F1 clutch into three at 3 months of age and transferred the F1s into separate 100 liter tanks with 3.5 ppt water. We added either 0.2 ppt or 35 ppt water in increments of 20 liters at a time twice a week over the course of one month to acclimate fish in two of the tanks to different water salinities. After one month of acclimation all water in the two tanks was changed to either 0.2 or 35 ppt. We raised the fish in 0.2, 3.5 or 35 ppt water for additional three months before harvesting tissue.

Sample preparation and RNA-sequencing. We harvested gill tissue for all strains and F1's, all staged as adults and reproductively active (gravid females and males exhibiting mating coloring). We flash-froze gills on liquid nitrogen and stored in -80 degrees C until used for mRNA-extraction. We disrupted gill tissue with a pestle on liquid nitrogen and extracted mRNA using Dynabeads mRNA direct kit (Invitrogen) and following manufacturer's instructions, followed by DNase treatment with the Turbo DNA-free kit (Ambion). We verified mRNA quality using Agilent BioAnalyzer.

We used 150 nanogram of mRNA to construct strand-specific RNA-seq libraries using the TruSeq Stranded RNA-seq kit (Illumina). We verified library yield using Qubit and size distribution using BioAnalyzer. We optimized library construction protocol to result

in mRNA insert size distribution centered on 290 base pairs. We pooled the libraries in equimolar amounts and sequenced in pools of eight samples on a HiSeq-3000 instrument, producing 150 bp paired-end reads (Table S2). We included replicate sequencing libraries in different lanes of the same run and different runs of the same instrument in order to measure the effect of batch on final data (none observed).

Gill transcriptome assembly from RNA sequencing. We verified read quality with *FastQC* software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed the reads of sequencing adapters using TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

We aligned RNA-seq reads from pure strains to the UCSC stickleback genome reference ("gasAcu1") with *STAR* aligner (47). We opted for running *STAR* in two-pass mode, gathering novel splice junctions from all pure-strain samples for the second alignment pass. After experimenting with alternative parameters, we opted for the following: --*outFilterIntronMotifs* RemoveNoncanonicalUnannotated --*chimSegmentMin* 50 --*alignSJDBoverhangMin* 1 --*alignIntronMin* 20 --*alignIntronMax* 200000 --*alignMatesGapMax* 200000 --*limitSjdbInsertNsj* 2000000.

We followed the *Cufflinks2.2* pipeline (27) for reference-guided transcriptome assembly and transcript and isoform expression level testing. We used *Cufflinks2* to assemble aligned RNA-seq reads into transcripts, using Ensembl gene models for stickleback (version 90) as guide and the following parameters: --*frag-bias-correct* gasAcu1.fa --*multi-read-correct* --*min-isoform-fraction* 0.15 --*min-frags-per-transfrag* 20 --*max-multiread-fraction* 0.5. We produced a single merged transcriptome assembly based on all pure strains using *CuffMerge* and used this in all subsequent analyses for all samples.

Principal Component Analysis. We summarized read counts over transcript models using the *Cufflinks* function *CuffQuant with 'fr-firststrand' strand-specific RNAseq library type and other settings as default* and normalized read counts to total library sizes using *CuffNorm*. We subsequently transformed the read data with the *DESeq2* (48) function *varianceStabilizingTransformation* so that the variance in read counts was independent of the mean, following the steps outlined in the *DESeq2* manual. We used the *R* (49) function *prcomp* with the option *scale=FALSE* to calculate PCA on expression level co-variances using data from all transcripts. We calculated the PCA based on a balanced set of four freshwater and four marine ecotypes from both Tyne and Little Campbell, and projected the single ecotypes from Forss and Shiel to principal components 2 and 5 using the *R* function *scale*. We verified the absence of batch effects in the RNA-seq data by PCA analysis of a replicate sequencing library sequenced on different lanes of the same run and on different runs. Technical variation was much smaller than biological variation.

As described in the main text, a combination of principal components 2 and 5 best described freshwater-marine divergence in transciptomes in our dataset. We therefore defined a composite principal component by summing principal components 2 and 5, weighing each with the percentage of variation explained. Finally, we extracted principal component 2 and 5 loadings for each transcript and used the identical approach to calculate transcript loadings on the composite principal component. This procedure produced a loading value for each transcript that described the importance

of that transcript in parallel freshwater-marine expression divergence.

Differential expression. For parametric testing of parallel expression differences between ecotypes, we combined samples from Little Campbell and Tyne and tested for differential expression using *CuffDiff,* specifying the parameter *-dispersion-method* per-condition. We imported the results tables into *R* and selected transcripts that were tested and where the q-value was less than 0.2. We additionally required that the mean expression difference between freshwater and marine ecotypes was of the same sign in Tyne and Little Campbell.

We used *CuffDiff* to test expression differences between male and female fish from freshwater and marine ecotypes of Tyne and Little Campbell river (combined). Transcripts with sex-dependent expression at FDR 10% (N=278) were excluded from analysis of genetic architecture of expression divergence (see below), but included in all other analyses where the number of male and female fish were balanced across the experimental contrast.

Gene Ontology analysis. Tests for enrichment of genes involved in specific biological processes, molecular functions and cellular components among top ranking differentially expressed genes was performed using GOrilla (50). Genes were sorted by CuffDiff differential expression q-values or by composite PC loading score and, because stickleback genes are largely unannotated for gene ontologies, were mapped to mouse orthologs (the vertebrate with the highest GO annotation quality (ref: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2241866/) using a REST command to access the Ensembl precomputed ortholog database. We tested for significant enrichment of gene ontologies (biological processes, molecular functions, and cellular compartments) with p-values less than $1x10^{-5}$. A similar approach using both human and zebrafish orthologs revealed enrichment in many of the same gene ontologies (not shown).

DNA preparation and sequencing. We extracted genomic DNA from fin clips of six unrelated fish per ecotype (six females of each ecotype from Little Campbell, three males and three females of each ecotype from Tyne) using standard lysis buffer and proteinase-K digestion, followed by SPRI bead extraction with Ambion magnetic beads. We verified DNA quality on agarose gel and quantified DNA concentration using Qubit.

We fragmented 700 ng of gDNA with Covaris instrument and selected 300-500 bp DNA fragments for library construction using double-sided SPRI selection. We constructed DNA-sequencing libraries using a custom protocol that includes DNA end-repair, A-tailing and Illumina TruSeq adapter ligation, followed by 6 cycles of PCR amplification. We verified library fragment size using BioAnalyzer and quantified library concentrations using Qubit. Libraries were sequenced with Illumina HiSeq-3000 instrument to an estimated whole-genome coverage of 10-40X.

DNA-sequencing read processing, alignment and SNP discovery. We verified DNA-sequencing read quality using FastQC and trimmed adapter sequences using TrimGalore. We aligned DNA-seq reads to the stickleback reference genome sequence (“BROAD S1” (16)) using *BWA mem* (51).

We used the Broad Institute Genome Analysis Tool Kit (*GATK*) (52) to call Single Nucleotide Polymorphisms (SNPs) in genomic resequencing data, following the DNA-seq best practices (as in June 2016). We ran *GATK HaplotypeCaller* individually for each sample and defining parameters *-stand_call_conf* 30 *-stand_emit_conf* 10 *--emitRefConfidence* GVCF *-variant_index_type* LINEAR *-variant_index_parameter* 128000. The step was followed by joint genotyping using *GenotypeGVCFs* after which we excluded indels from the analysis. The final step was *VariantQualityScoreRecalibration* (*VQSR*). Samples from each ecotype-pair as well as each controlled cross were analyzed together (but separately from other ecotype-pairs or controlled crosses) from *GenotypeGVCFs* -step onwards. Because stickleback lacks a set of known variant sites, we opted for using a hard-filtered set of SNPs as "true" set of SNPs (with *prior*=10). We used the *GATK SelectVariants* tool to extract a training set that fulfilled the following thresholds: *QD*>30, *FS*<60, *MQ*>40, *MQRankSum*>-12.5 and *ReadPosRankSum*>-8. After inspection of *VQSR* tranche plots, we selected the 99.9% quality tranche for downstream analysis, which captured 1.66M and 3.52M SNPs with transition/transversion ratios of 1.14 and 1.18 for Tyne and Little Campbell population genomic analyses respectively, and 2.27M, 1.62M, 2.75M, 1.54M SNPs with transitition/transversion ratios ranging from 1.16-1.17 for parents of the four crosses (strains from Forss, Tyne, Little Campbell, Shiel) used in allele-specific expression analysis.

Population genetic analyses. Population genetic analyses were based on a set of six freshwater and six marine fish from both Tyne and Little Campbell River. From the GATK variant calling analysis (described above) we identified over 3.5 million SNPs in the Little Campbell populations and over 1.6 million SNPs in Tyne. We calculated per-site statistics for Weir & Cockerham's Fst (53) and average pairwise-nucleotide diversity (Pi) genome-wide, and for 400 kilobase (kb) regions centered on transcription start sites (TSSs) using *VCFtools* (version 0.1.14) (54) allowing for a maximum of 4 missing genotypes per SNP for calculation of Fst and a maximum of 2 missing genotypes per SNP for calculation of Pi, corresponding to a maximum of 20 % missing genotypes in each case. Negative Fst values were rounded to zero. CSS score was calculated based on Pi and following (16) in 10kb non-overlapping windows across the genome. We used the 10kb windows to assign genome regions as having strongest level of parallel genetic divergence between Tyne and Little Campbell freshwater and marine ecotypes, keeping the top 1% windows with the highest CSS score.

We used custom R scripts to calculate Fst and Pi in 1kb windows centered on transcription start sites (TSSs, as reported by *CuffLinks*). We used custom R scripts and the R package *GenomicRanges* to compare the genomic cordinates of transcripts showing parallel expression divergence to the cordinates of the genomic windows showing high CSS values. We calculated the average distance between CSS outlier windows and transcripts in increments of 10kb and compared the average distance to distances calculated based on 1000 randomized sets of transcripts.

Allele-Specific Expression analysis. We defined a set of high-confidence SNPs for Allele-Specific Expression (ASE) analysis, based on genomic resequencing of parent fish used in controlled crosses (above). We then used *GATK FastaAlternateReferenceMaker* to mask the stickleback reference genome in the corresponding position with N's in order to avoid preferential mapping of reference

SNP alleles. We aligned RNA-seq reads from each F1 and the parents onto the N-masked reference genomes using *STAR* and parameters as described above, with the exception that we allowed for only uniquely mapping reads (*--outFilterMultimapNmax* 1). No significant preferential mapping of reference SNPs was observed after these steps.

We selected SNPs where the genotypes of the parents were covered by at least 10 DNA sequencing reads in each parent and where the genotypes of the parents were homozygous for different alleles. These SNP positions were assigned as informative for allele-specific analysis in each cross. Expression levels for allele-specific analyses were represented as read counts overlapping informative SNP positions. We generated allele-specific read counts for F1's and parents with the *GATK ASEReadCounter* tool and enabling default filters. We verified that parents had more than 99% counts assigned to right genotypes and excluded the few SNPs where the RNA-seq reads indicated that both alleles were expressed in a parent that should be homozygous. We combined all individuals in each cross (parents and F1's) in one data frame and normalized read counts between individuals to the total library size using *DESeq2* function *estimateSizeFactors* in order to have equal power across F1s. We then filtered for SNP positions covered by more than 10 reads in at least one F1 in order to avoid underpowered tests of allele-specific expression at loci showing no or very low expression. Finally, we intersected the SNP-based results with transcript models from our reference-guided assembly using the R package *GenomicRanges*, assigning each informative SNP position to an expressed locus and exon.

We tested for ASE in each informative SNP position using a binomial exact test in *R* and an FDR level of 10% (55). Normally the null hypothesis of binomial test is 0.5, in our case meaning that 50% of the RNA-seq reads represented either marine or freshwater parent alleles. Our approach takes into account a possible residual effect of preferential mapping of reference allele reads by calculating the null-hypothesis for the binomial test based on the ratio of all reference reads over all alternative reads per each F1 following (56). The null hypothesis calculated this way was between 0.5 and 0.52, indicating that the residual effect of preferential mapping of reference alleles, if detected, was small. Finally the results for the ASE test were converted from reference allele versus alternative allele format into marine parent versus freshwater parent by comparing to the genotypes of the parents.

Following ASE testing we tested for analogous expression difference between parents in the corresponding SNP positions, again using binomial exact test and FDR of 10%. We tested for difference in allele-expression ratio versus parental expression ratio with Fisher's exact test. We then compared ASE significance, ASE sign and ASE magnitude to parental expression difference in order to dissect parental expression differences into divergence classes following (10), outlined in table S4.

Given that the median number of assembled transcript isoforms per gene (locus) is 3, and the mean number of ASE informative SNPs tagging a given gene (locus) range from 2.0–6.2, we concluded that the level of evolutionary divergence between marine and freshwater stickleback strains used in our study was insufficient to dissect the *cis*- vs trans- genetic architecture underlying expression divergence at the transcript (isoform) level (see Supplemental Note for more information). We instead analyzed allele-specific expression at the gene (locus) level, classifying *cis*-/trans- architecture

of each gene into one divergence class based on the SNP that showed the highest statistical support (see above).

We performed concordance analysis for validating the reproducibility of allele expression levels and divergence classes on SNPs assigned to the same exon and the same transcript (Supplemental Note). For the final classification of transcipts to divergence types, we ordered informative SNPs per transcribed locus per F1 by the product of the *P*-values of the three tests (above) and selected the SNP that had the lowest product of *P*-value as a representative SNP for that transcript, per F1. We selected this approach among alternatives after taking into account the high concordance of allele expression levels, relatively complex isoform expression patterns, analysis sensitivity, statistically balanced approach, and parsimony in biological explanation of expression divergence (Supplementary Note).

A small number of SNPs showed monoallelic expression in F1's where RNA-seq reads overlapping one of the parent alleles were not observed (e.g. 1129 SNPs in Little Campbell River cross). Analysis of RNA-seq read coverage indicated that these SNP positions had lower coverage specifically of the alternative allele, and this effect was not observed when evaluated based on all SNPs or subsets of SNPs e.g. assigned as *cis*-diverged. This indicated that the cases of monoallelic expression are likely caused by mappability issues, and that the issue was specific to SNPs showing monoallelic expression. Although including these SNPs did not impact the results in a significant way we decided to exclude SNPs showing monoallelic expression in F1's or parents from the analysis.

We summarized the frequencies of transcripts assigned to divergence classes for each F1 using custom R scripts, excluding transcripts that showed differential expression between sexes at FDR 10%. We then assigned transcripts to two classes according to whether or not they showed parallel expression divergence in PCA and differential expression analysis. We tested for over-representation in divergence classes using a randomization test. We draw 1000 sets of 586 randomly chosen transcripts from the whole dataset, calculated the frequencies of divergence classes and compared the random expectation to our set of parallel diverged transcripts.

Salinity response. For testing the effect of salinity acclimation on gene expression in F1 gills, we estimated transcript-level expression using *CuffQuant* and normalized counts for each sample to total library sizes using *CuffNorm*. We then imported the gene count tables into *R* and tested for expression differences between salinity treatments using constrasts and an FDR level of 1%, as implemented in *DESeq2*.

For analysis of expression profiles, we imported FPKM values from *CuffLinks* into R. The FPKM values were highly correlated with normalized expression values from *DESeq2 VarianceStabilizingTransformation*, and allow for a more intuitive interpretation. We subsetted the data to only include the transcripts showing differential expression between at least one contrast and parallel expression divergence (N=97). We then log-transformed and normalized the expression of each transcript to the average expression level across all samples to produce an expression profile that represent expression in a given sample relative to others (for that transcript). We then compared the profiles of samples with Spearman correlation. A correlation approaching 1 indicates that expression profiles tended to be similar

relative to other samples. In contrast, a correlation approaching -1 indicated that sample profiles were mirror-images of one another. Finally, we clustered the samples based on euclidean distance and visualized the sample similarity profiles using the *pheatmap R* function.

For analysis of allele-specific expression in salinity treatments, we imported allele-specific counts over informative SNPs (as defined above) into R and transformed the counts into log-fold change of marine over freshwater allele. We identified one of the samples acclimated to 35 ppt as having outlier allelic expression levels very different from all other samples and excluded the sample from further analysis. We intersected the SNPs with transcripts that exhibited salinity response and parallel expression divergence (N=11 transcripts). We measured the similarity of fold-change expression differences between alleles across samples with Spearmans correlation, analogous to FPKM counts.

Supplementary Note

Hierarchical clustering of expression. The gill transcriptome can be characterized by 5 major groups of loci (Fig S2) comprising 766 highly expressed genes (2.6%) showing strong enrichment for biological processes including mitochondrial respiration, ATP synthesis coupled proton transport and cytoplasmic translation, 4337 moderately expressed genes (14.8%) enriched for metabolic processes such as mRNA processing and RNA splicing, and functions such as cadherin mediated cell adhesion, 6950 lowly expressed genes (23.7%) enriched with genes involved in protein modification and chromatin modification processes, 5565 low or partially expressed genes (19.0%) enriched in genes involved in developmental processes and regulation of developmental processes, and 11677 predominantly unexpressed genes (39.9%).

CSS based on Tyne and Little Campbell population data. We calculated a Cluster Separation Score (CSS) in 10kb non-overlapping windows across the genome following (2). The CSS score reflects parallel genetic divergence between freshwater and marine fish irrespective of their geographic origin. We assigned the genomic windows with the extreme 0.5% CSS values as regions showing the strongest signal of parallel genetic divergence (449 windows). More than 26 percent of the transcripts evolving in parallel between Tyne and Little Campbell were situated within 10kb of regions of parallel genetic divergence (randomization test, $P<<0.025$, Fig S6), which is two times more than what was observed for the global set of regions showing parallel genetic divergence (main text).

Ecotype-specific expression divergence. We used *CuffDiff* for testing of ecotype-specific expression divergence specifying the same parameters as in the main text. For this analysis, freshwater-marine expression differences were tested separately for Tyne and Little Campbell and the results were compared using in *R*. Transcripts that were differentially expressed in both ecotype-pairs with FDR 20% and where the ecotype difference was of the same sign were assigned as "parallel", whereas if the signs were opposite the transcripts were assigned as "anti-parallel".

Overall, 719 differentially expressed transcripts (FDR 20%) were identified using a parametric analysis, the majority of which (N=515) had marine-freshwater differential expression unique to Little Campbell compared to N=157 uniquely differential in Tyne strains. Consistent with largely river system-specific ecotype expression divergence only four percent of loci (N=29) show parallel expression divergence in both rivers (significant differential expression and identical sign of expression difference in both Tyne and Little Campbell) while 2.5% (N=18) show anti-parallel expression divergence (significant differential expression and opposite sign of expression difference).

SNP concordance analysis. We performed SNP concordance analysis to validate the reproducibility of allele expression levels between SNPs assigned to same exons and to different exons of the same transcript. Our assumption for this analysis was that individual SNPs assigned to the same exons should show correlated levels of ASE as well as concordant class of genetic divergence (*cis, trans* etc.) when compared within the same F1 individual. It is worth to note that the divergence class also depend on expression levels assigned to the SNP position in parents, which we ignored for simplicity in this analysis.

SNPs assigned to same exons and showing identical type of genetic divergence (concordant SNPs) tended to have strongly correlated ASE levels (Figure S18a). SNPs assigned to same exons but showing different classes of genetic divergence (discordant SNPs) were almost 50% rarer compared to concordant SNPs and, as expected, showed lower level of correlation in ASE and overall smaller allelic differences (Figure S18b). Concordant and discordant SNPs within exons showed overall similar distribution among divergence classes, indicating that discordant calls influenced all classes equally (not shown).

SNPs assigned to the same transcript but different exons had a correlated level of ASE in cases where the divergence class assigned to both exons were the same (Figure S18c). Cases where the exons showed different divergence class showed a lower correlation in ASE (S18d). These results are consistent with previous studies demonstrating variable levels of ASE along genes and between exons (57). Different exons of the same transcript that show different levels of ASE suggest that ASE effects are specific to single isoforms rather than all isoforms assigned to the same transcript. Through our RNA-seq analysis we identified over 162000 known and new isoforms distributed to 29296 transcribed loci (on average over 5 isoforms per transcribed locus).

In our final analysis, we opted to classify each transcript into one divergence class, based on the SNP that showed the highest statistical support (see Methods). The justification for this choice was based on the following considerations:

1) Individual SNPs assigned to the same exon tended to show similar ASE levels and divergence types, indicating that dissection of divergence architecture was generally robust to different SNPs within exons. Discordance in divergence types for SNPs assigned to the same exon influenced all divergence types equally and therefore is not expected to bias the results.

2) SNPs assigned to the same transcript but different exons showed different divergence types in roughly half of the cases, and the levels of ASE on the SNP loci were less correlated. This suggests that different exons may experience varying levels of ASE, likely because of alternative isoform expression, as has been demonstrated before (57). Visual inspection of expression tracks in candidate genes for variable ASE identified multiple instances of putative alternative isoform expression (example in Fig S19). Any procedure that would not distinguish different exons, for example averaging expression levels across SNPs, would therefore suffer from low sensitivity as loci not showing ASE would cancel the signal from loci showing ASE.

3) We discarded the option of averaging ASE levels for SNPs assigned to the same exon because different transcripts and ecotype-pairs showed markedly different densities of SNPs. Averaging would therefore influence transcripts and ecotype-pairs unequally and bias

References

1. Turner TL, Bourne EC, Wettberg von EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nat Genet* 42(3):260–263.

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

3. Grossman SR, et al. (2013) Identifying recent adaptations in large-scale genomic data. *Cell* 152(4):703–713.

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

5. Stern DL, Orgogozo V (2009) Is genetic evolution predictable? *Science* 323(5915):746–751.

6. Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in Escherichia coli. *PNAS* 100(3):1072–1077.

7. Chanut-Delalande H, Fernandes I, Roch F, Payre F, Plaza S (2006) Shavenbaby couples patterning to epidermal cell shape control. *PLoS Biol* 4(9):e290.

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

9. Fraser HB, Moses AM, Schadt EE (2010) Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc Natl Acad Sci USA* 107(7):2977–2982.

10. Landry CR, et al. (2005) Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. *Genetics* 171(4):1813–1822.

11. Landry CR, Hartl DL, Ranz JM (2007) Genome clashes in hybrids: insights from gene expression. *Heredity* 99(5):483–493.

12. Mundy NI (2005) A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proc Royal Soc Biol* 272(1573):1633–1640.

13. Yeaman S, et al. (2016) Convergent local adaptation to climate in distantly related conifers. *Science* 353(6306):1431–1433.

14. Prince DJ, et al. (2017) The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci Adv* 3(8):e1603198.

15. Bell MA, Foster SA (1994) *The evolutionary biology of the threespine stickleback* (Oxford University Press, Oxford ; New York).

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

17. Colosimo PF, et al. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307(5717):1928–1933.

18. O'Brown NM, Summers BR, Jones FC, Brady SD, Kingsley DM (2015) A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *eLife* 4. doi:10.7554/eLife.05290.

19. Shapiro MD, et al. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428(6984):717–723.

20. Chan YF, et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327(5963):302–305.

21. Evans DH, Piermarini PM, Choe KP (2005) The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Phys Rev* 85(1):97–177.

22. Velotta JP, et al. (2016) Transcriptomic imprints of adaptation to fresh water: parallel evolution of osmoregulatory gene expression in the Alewife. *Mol Ecol* 26(3):831–848.

23. Gibbons TC, Metzger DCH, Healy TM, Schulte PM (2017) Gene expression plasticity in response to salinity acclimation in threespine stickleback ecotypes from different salinity habitats. *Mol Ecol* 34(10):185–2725.

24. McCairns RJS, Bernatchez L (2010) Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution* 64(4):1029–1047.

25. Whitehead A (2010) The evolutionary radiation of diverse osmotolerant physiologies in killifish (Fundulus sp.). *Evolution* 64(7):2070–2085.

26. Gunter HM, Schneider RF, Karner I, Sturmbauer C, Meyer A (2017) Molecular investigation of genetic assimilation during the rapid adaptive radiations of East African cichlid fishes. *Mol Ecol* 5(3):457.

27. Trapnell C, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562–578.

28. Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494–1512.

29. Nielsen R, et al. (2005) Genomic scans for selective sweeps using SNP data. *Genome Res* 15(11):1566–1575.

30. Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in cis and trans gene regulation. *Nature* 430(6995):85–88.

31. Lemos B, Araripe LO, Fontanillas P, Hartl DL (2008) Dominance and the evolutionary accumulation of cis-and trans-effects on gene expression. *PNAS* 105(38):14471–14476.

32. Gibson G, et al. (2004) Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. *Genetics* 167(4):1791–1799.

33. McManus CJ, et al. (2010) Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res* 20(6):816–825.

34. Landry CR, et al. (2005) Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. *Genetics* 171(4):1813–1822.

35. Orr HA (1998) Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* 149(4):2099–2104.

36. Wittkopp PJ, Haerum BK, Clark AG (2008) Regulatory changes underlying expression differences within and between Drosophila species. *Nat Gen* 40(3):346–350.

37. Hwang P-P, Lee T-H, Lin L-Y (2011) Ion regulation in fish gills: recent progress in the cellular and molecular mechanisms. *AJP: Regulatory, Integrative and Comparative Physiology* 301(1):R28–R47.

38. Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150(3811):563–565.

39. Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. *PNAS* 104(suppl 1):8605–8612.

40. Emerson JJ, et al. (2010) Natural selection on cis and trans regulation in yeasts. *Genome Res* 20(6):826–836.

41. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ (2014) Tempo and mode of regulatory evolution in Drosophila. *Genome Res* 24(5):797–808.

42. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL (2007) Genetic properties influencing the evolvability of gene expression. *Science* 317(5834):118–121.

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

44. Cleves PA, et al. (2014) Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. *PNAS* 111(38):13912–13917.

45. Haldane JBS (1927) A mathematical theory of natural and artificial selection, part V: selection and mutation. *Mat Proc Camb Phil Soc* 23:838–844.

46. Verta J-P, Landry CR, Mackay J (2016) Dissection of expression-quantitative trait locus and allele specificity using a haploid/diploid plant system – insights into compensatory evolution of transcriptional regulation within populations. *New Phytol* 211(1):159–171.

47. Dobin A, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21.

48. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.

49. Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. *J Comput Graph Stat* 5(3):299–314.

50. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10(1). doi:10.1186/1471-2105-10-48.

51. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997

52. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.

53. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358.

54. Danecek P, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

55. Storey JD (2002) A direct approach to false discovery rates. *J Royal Stat Soc B* 64(3):479–498.

56. Buil A, et al. (2015) Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* 47(1):88–91.

57. Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM (2011) A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res* 21(10):1728–1737.