**Title:** Evolution of spatial and temporal *cis*-regulatory divergence between marine and freshwater sticklebacks

Katya L. Mack[1], Tyler A. Square[2], Bin Zhao[1], Craig T. Miller[2], Hunter B. Fraser[1]

[1]Department of Biology, Stanford University, Stanford, CA, 94305

[2]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Correspondence to: hbfraser@stanford.edu

**Abstract**

*Cis*-regulatory changes are thought to play a major role in adaptation. Threespine sticklebacks have repeatedly colonized freshwater habitats in the Northern Hemisphere, where they have evolved a suite of phenotypes that distinguish them from marine populations, including changes in physiology, behavior, and morphology. To understand the role of gene regulatory evolution in adaptive divergence, here we investigate *cis*-regulatory changes in gene expression between marine and freshwater ecotypes through allele-specific expression (ASE) in F1 hybrids. Surveying seven ecologically relevant tissues, including three sampled across two developmental stages, we identified *cis*-regulatory divergence affecting a third of genes, nearly half of which were tissue-specific. Next, we compared allele-specific expression in dental tissues at two timepoints to characterize *cis*-regulatory changes during development between marine and freshwater fish. Applying a genome-wide test for selection on *cis*-regulatory changes, we find evidence for lineage-specific selection on several processes, including the Wnt signaling pathway in dental tissues. Finally, we show that genes with ASE, particularly those that are tissue-specific, are enriched in genomic regions associated with marine-freshwater divergence, supporting an important role for *cis*-regulatory differences in adaptive evolution of sticklebacks. Altogether, our results provide insight into the *cis*-regulatory landscape of divergence between stickleback ecotypes and supports a fundamental role for *cis*-regulatory changes in rapid adaptation to new environments.

## Introduction

Understanding how organisms adapt to new environments is a major goal in evolutionary biology. Central to this goal is understanding what genetic changes underlie adaptive traits. Threespine sticklebacks (*Gasterosteus aculeatus*) are a powerful model for studying the genetic basis of adaptation[1]. After the end of the last ice age, marine sticklebacks colonized thousands of freshwater habitats in the Northern Hemisphere [2]. In these freshwater environments, populations have rapidly evolved a number of traits that distinguish them from the ancestral marine form. While adaptation to each lake or stream is independent, several traits have evolved repeatedly across multiple freshwater systems either through parallel, convergent, or distinct genetic changes (e.g., changes in body shape, skeletal armor, dentition, behavior, and pigmentation)[2–6]. The repeated evolution of similar phenotypes in freshwater systems is strong evidence that these traits reflect local adaptation and provide a powerful platform for studying the genetic architecture of adaptive phenotypic evolution [7–9].

Mutations in *cis*-regulatory elements can change how nearby genes are regulated. Such mutations are thought to be an important substrate for adaptive evolution[10–12]. In contrast to protein-coding changes, *cis*-regulatory mutations can alter the expression of gene targets in tissue- or temporally- specific ways. As a consequence, *cis*-regulatory changes may be less constrained by the deleterious side-effects of negative pleiotropy, making this class of mutations important targets for natural selection [10,11]. *Cis*-regulation has been shown to be the major driver of local environmental adaptation in recent human evolution [13], and likewise plays a central role in the local adaptation of sticklebacks to freshwater environments. Genome scans have found that genomic regions associated with recurrent divergence between ecotypes are predominantly intergenic, suggesting parallel divergence may often involve the reuse of pre-existing gene regulatory variation 1. Pirinen M, Lappalainen T, Zaitlen NA, GTEx Consortium, Dermitzakis ET, Donnelly P, et al. Assessing allele-specific expression across multiple tissues from RNA-seq read data. Bioinformatics. 2015;31: 2497–2504. doi:10.1093/bioinformatics/btv074

2. Mitsiadis TA, Pagella P, Cantù C. Early Determination of the Periodontal Domain by the Wnt-Antagonist Frzb/Sfrp3. Frontiers in Physiology. 2017;8. Available: https://www.frontiersin.org/articles/10.3389/fphys.2017.00936
3. Neves VCM, Babb R, Chandrasekaran D, Sharpe PT. Promotion of natural tooth repair by small molecule GSK3 antagonists. Sci Rep. 2017;7: 39654. doi:10.1038/srep39654
4. Wang C, Wang Y, Wang H, Yang H, Cao Y, Xia D, et al. SFRP2 enhances dental pulp stem cell-mediated dentin regeneration in rabbit jaw. Oral Diseases. 2021;27: 1738–1746. doi:10.1111/odi.13698
5. Fjeld K, Kettunen P, Furmanek T, Kvinnsland IH, Luukko K. Dynamic expression of Wnt signaling-related Dickkopf1, -2, and -3 mRNAs in the developing mouse tooth. Dev Dyn. 2005;233: 161–166. doi:10.1002/dvdy.20285
6. Liu F, Chu EY, Watt B, Zhang Y, Gallant NM, Andl T, et al. Wnt/β-catenin signaling directs multiple stages of tooth morphogenesis. Developmental Biology. 2008;313: 210–224. doi:10.1016/j.ydbio.2007.10.016
7. Lin M, Li L, Liu C, Liu H, He F, Yan F, et al. Wnt5a regulates growth, patterning, and odontoblast differentiation of developing mouse tooth. Dev Dyn. 2011;240: 432–440.

doi:10.1002/dvdy.22550
8. Løes S, Luukko K, Hals Kvinnsland I, Salminen M, Kettunen P. Developmentally regulated expression of Netrin-1 and -3 in the embryonic mouse molar tooth germ. Developmental Dynamics. 2003;227: 573–577. doi:10.1002/dvdy.10317

[8]. C*is*-regulatory mutations have been implicated in specific morphological differences between marine and freshwater forms, including the loss of pelvic spines [14], bony armor plates [7], changes in pigmentation[4], and increased pharyngeal tooth number [6,15]. While these lines of evidence suggest an important role for gene regulatory evolution in stickleback adaptation, the global *cis*-regulatory landscape of marine-freshwater divergence remains poorly understood. Exploration of *cis*-regulatory changes between ecotypes has largely been limited to assaying individual gene targets in a small number of tissues (e.g., [6,7,16,17]). Transcriptome-wide *cis*-regulatory divergence between marine and freshwater fish has been characterized in two tissues so far: the gills[18] and ventral pharyngeal tooth plates [19]. Surprisingly, these two tissues showed highly divergent regulatory landscapes[18], suggesting tissue-specific regulatory architecture may play an important role in stickleback adaptation.

Here we survey global *cis*-regulatory divergence between marine and freshwater sticklebacks in seven tissues to understand the role of gene expression evolution in adaptive divergence. To characterize *cis*-regulatory changes between ecotypes, we crossed marine and freshwater fish to generate F1 hybrids. As F1 hybrids carry both a marine and freshwater copy of each chromosome, alleles from both parents are present in the same cellular environment (e.g., subject to the same *trans*-acting factors). Expression differences between the two parental alleles (i.e., allele-specific expression) can therefore only result from *cis*-regulatory changes [20,21]. We use this approach to examine a collection of tissues important for behavioral, physiological, feeding, and morphology differences between marine and freshwater forms (i.e., brain, liver, eyes, flank skin, dorsal and ventral pharyngeal tooth plates, and mandible). As morphological changes include especially dramatic changes to the craniofacial skeleton and dentition[6,22], likely reflecting adaptations to different diets in freshwater, we also examine a second developmental timepoint in dental tissues to characterize *cis*-regulatory modifications during development. We use these data to dissect the landscape of *cis*-regulatory divergence and then ask whether these changes are associated with genomic signals of selection. Overall, our results highlight the importance of the tissue- and developmental stage-specific *cis*-regulatory changes in marine-freshwater divergence and the importance of *cis*-regulatory variation in local adaptation.

## Results and Discussion

### *Extensive allele-specific expression across tissues in freshwater-marine hybrids*

To investigate *cis*-regulatory divergence between marine and freshwater individuals, we analyzed allele-specific expression in F1 hybrids between marine and freshwater fish (freshwater Paxton lake benthic [PAXB] x marine Rabbit Slough [RABS])(Figure 1A). Seven tissues were collected from F1 hybrids at the young adult stage (~35 millimeters [mm] standard length [SL]) (brain, eyes, liver, flank skin, ventral pharyngeal tooth plate [VTP], dorsal pharyngeal tooth plate [DTP], mandible) (Figure 1A-B). Additionally, three dental tissues (mandible, VTP, and DTP) were also collected from full-siblings at an earlier juvenile stage (15-20mm SL) for a temporal comparison of dental development (hereafter, "early" vs. "late" developmental timepoint). We sequenced mRNA from each tissue for two biological replicates, obtaining a median of 66.7 million reads per sample (Tables S1, S2). To phase heterozygous sites in F1s, we also performed whole-genome sequencing of the freshwater parent (PAXB) to an average coverage of ~30X (Figure S1).

Principal component (PC) analysis of gene-wise mRNA abundance and allele-specific expression (ASE) revealed tissue-type to be the primary driver of variation (Figures 1B and S2). Allele-specific expression values clustered largely by tissue of origin on PC1 and PC2 (PC1: 52% of the variation, PC2: 30% of variance). Dental tissues formed their own cluster to the exclusion of other tissues, as did eyes and brain. Flank skin, where bony lateral plates develop, also formed a group with dental tissues on PC1 (Figure 1B). PC analysis of allele-specific expression of dental tissue timepoints also separated samples based on developmental stage (early vs. late) on PC1 or PC2 (Figure S3).

Extensive ASE was found across tissues (Figure 1C-D). Nearly 33% of genes (4,411) were found to have significant ASE in at least one tissue or tissue-timepoint (DESeq2 Wald-test, FDR<0.05, see Table S3, Figure 1D; 13,551 genes tested). In each tissue, these ASE genes accounted for approximately 5-12% of genes surveyed. Dental tissues had the greatest number of ASE genes overall, particularly at the earlier developmental timepoint (Figure 1D). The lowest number of ASE genes was identified in the brain (714 genes, 5.6%). The number of ASE genes identified in a tissue was not related to differences in read depth between tissues (Figure S4).

Comparing overlap of genes with ASE between tissues, we found that the largest distinct groups were tissue-specific rather than shared, indicating largely tissue-specific *cis*-regulatory divergence between marine and freshwater fish (Figure 1E). Across the seven tissues sampled at the late timepoint (SL ~35mm), 1,660 genes showed ASE in only one tissue (48% of genes with ASE overall). In particular, the liver was found to have the greatest number of unique ASE genes (366 genes, 32% of genes with ASE in liver). Comparisons between tissues also revealed many genes with evidence for shared ASE (Figure 1E).

In particular, we found high overlap between eyes and brain (29% shared), between dental tissues (36%-44%), and between dental tissues and flank skin (32%-36%)(Table S4). In contrast, few genes (~2%) showed ASE across all tissues (78 genes across all tissues, 92 genes at the late developmental timepoint). For genes with ASE in multiple tissues, directionality was typically maintained, with only 236 genes showing a change in which parental allele was upregulated between tissues.

***Widespread heterogeneity in allele-specific expression across tissues in marine-freshwater hybrids***

Comparisons of ASE across tissues revealed abundant *cis*-regulatory divergence between marine and freshwater fish. To investigate variation in allele-specific expression between tissues in marine-freshwater hybrids, we employed a Bayesian approach to partition genes in a tissue into three states – no ASE, moderate ASE, and strong ASE – based on the numbers of reads supporting the marine and freshwater allele [23]. Tissues are further classified as showing ASE heterogeneity if the strength of ASE varied across tissues (e.g., ASE is present in some tissues but absent in others or varies in magnitude between different tissues). Finally, we consider a sub-state of ASE heterogeneity to be tissue-specificity, where the ASE state (i.e., moderate, strong ASE, or no ASE) is observed in only one tissue despite expression of the gene across multiple tissues. Consequently, tissue-specificity describes cases where ASE state is unique to a single tissue.

We found that heterogeneity in ASE between tissues was common. Comparing across the seven different tissues collected at our second timepoint, we found that 44% of genes with ASE were classified as having heterogeneous ASE at a posterior probability (PP)>0.9 (at PP>0.95, 38%)(Figure 2A; Full list in File S1). Nearly all the genes with ASE heterogeneity (99%) did not show ASE in at least one of the tissues surveyed, with the remaining 1% showing evidence for ASE of varying magnitudes across all tissues. Evidence of tissue-specificity was also found for 448 genes (PP>0.9, File S1)(Figure 2B). Liver harbored the greatest number of genes with tissue-specific ASE (141 genes), followed by the eyes (66 genes). Repeating this analysis to incorporate dental tissues from the early timepoint and late timepoint, we also identified 73 genes with developmental- and tissue- specific ASE in tissues from the early developmental stage (File S1). Overall, allele-specific expression across tissues was found to be highly heterogeneous, likely reflecting tissue-specific *cis*-regulatory differences between marine and freshwater individuals.

Several genes with tissue-specific ASE were of interest for their reported tissue-specific functions in other systems (File S1). For example, *Dgat2* was expressed in five tissues at the second timepoint but found to have liver-specific ASE (Figure 2C). *Dgat2* is involved in triglyceride synthesis and plays an important role in energy metabolism; in mammals and zebrafish, gene mutants are associated with fatty liver

[24,25]. In dental tissues, genes with tissue-specific expression include a number of genes involved in tooth and bone formation (e.g., *Spp1*, *Dlx1a, Odam, Sox2, Epha3*, *Ssuh2rs1, Tgfbr2b, Stc2a*)(File S1). *Stc2a*, which was found to have tissue-specific ASE in the early mandible, was also recently shown to underlie changes in pelvic spine length between stickleback populations [16].

***Temporal differences in allele-specific expression during dental development***

Marine and freshwater sticklebacks show a number of phenotypic differences associated with feeding morphology (e.g., larger jaws, more teeth), likely reflecting adaptations to larger prey found in the benthic zone of lakes [2,26]. Divergence in tooth number arises during late development, providing an opportunity to study *cis*-regulatory divergence in the context of developmental evolution [6,19,27]. To investigate *cis*-regulatory divergence during dental development, we examined ASE in three dental tissues at two developmental timepoints (Figure 3A).

Developmental stage was a major component of variation in allele-specific expression. Principal component analysis of marine-freshwater allelic $\log_2$ fold changes clustered tissues by timepoint, with late and early dental tissues forming separate clusters on PC2 (22% of the variance, Figure S5). PC analysis of allele-specific counts from individual tissues also clustered tissues based on developmental timepoint on either PC1 (mandible and DTP, 50% and 48% of variance, respectively) or PC2 (VTP, 33% of variance)(Figure S3).

In contrast to our comparison of more diverse tissues, dental ASE was often shared across tissues or developmental stages (Figures 3B, S6). Nearly 10% of genes with ASE in dental tissues (330/3471 genes) showed ASE in all three dental tissues and at both developmental timepoints. Overall, a greater proportion of genes with ASE were shared across dental tissues in late development compared to early development: 19% of ASE genes (564 genes) were shared across all three tissues in the early stage versus 26% (528 genes) in the late stage (Chi-square test, $P$=0.002). Examining stickleback orthologs of genes implicated in mammalian tooth development collected from the Bite-It and ToothCode databases (hereafter referred to as "BiteCode" genes [19]), we found that these genes were enriched for ASE (Fisher's exact test, $P$=0.0046)(Figure S5B). BiteCode enrichment is consistent with the conservation of regulatory networks regulating dental development in mammals and fish [28,29].

Next, we characterized developmental differences in *cis*-regulatory divergence by comparing ASE between timepoints. Across developmental stages, divergent ASE can reflect the activity of temporally-specific genes controlled by divergent *cis*-regulatory elements between marine and freshwater fish (Figure

3A). Comparing the ratio of marine to freshwater allelic counts between early and late development in hybrids, we observed widespread differential allele-specific expression between the two developmental stages for each tissue, accounting for 37-43% of genes with ASE at either timepoint (Figure 3C)(Fisher's exact tests, FDR<0.05; see Methods). The majority of differential ASE reflected ASE that was timepoint specific, meaning ASE was only observed at one developmental stage. However, roughly a quarter of differential ASE in each tissue was due to changes in the magnitude of ASE between timepoints.

More genes with differential ASE were found to have a larger *cis*-effect at the early stage than the late stage (i.e., |log2 fold change in early| > |log2 fold change in late|); Figure 3C), consistent with the greater proportion of genes with ASE at the early timepoint overall (Figure 1D). This result was surprising, as greater phenotypic divergence is observed between marine and freshwater fish in the pharyngeal tooth plates in late development [27]. Developmental differences were also typically tissue-specific: 67% of genes with developmental stage-bias ASE were unique to one tissue. Thus, *cis*-regulatory differences between marine and freshwater individuals are often specific to both tissue and tissue-developmental stage.

***Polygenic selection on cis-regulatory divergence between marine and freshwater sticklebacks***

*Cis*-regulatory changes between marine and freshwater sticklebacks are potentially interesting for their role in local adaptation [8]. However, the majority of *cis*-regulatory changes are expected to be neutral. To test for selection on *cis*-regulatory changes between marine and freshwater fish, we employed a gene-set approach based on the sign test framework [30,31]. Under neutrality, QTLs for any given trait are expected to be unbiased with respect to their directionality, assuming these QTLs are independent (i.e. caused by different genetic variants) [32]. In a marine/freshwater genetic cross, each allele would be expected to be equally likely to increase the trait value if that trait is not under lineage-specific selection. Similarly, if a gene set associated with a biological function shows a significant directional bias in ASE (with more *cis*-changes acting in the same direction than expected), this suggests lineage-specific selection on the *cis*-regulation of this gene set [30,31]. Applying the sign test to GO gene sets in individual tissues and in the combined dental tissue set, we identified multiple gene sets with evidence for biased directionality (full list in Tables S5, S6).

In the combined dental tissue set, we found biased directionality for the GO terms "canonical Wnt signaling pathway" (Permutation based *P*-value = 0.0078), "embryonic viscerocranium morphogenesis" (*P* = 0.0093), and "Inflammatory response" (*P* = 0.0095). Wnt signaling plays a critical and evolutionarily conserved role in tooth and bone development [33,34] and genes in this pathway with ASE

have been directly implicated in regulating dental development in other species (e.g., *Wnt5a*, *Sfrp2*, *Ctnnb1*, *Net1*)(Table S7). While the GO annotation term included both positive and negative regulators of Wnt signaling, the pathway "Negative regulation of canonical Wnt signaling" was also nominally significant for biased downregulation of freshwater alleles (7/7 genes, Fisher's exact test, $P$=0.0048), suggestive of biased Wnt inhibition in marine fish (Figure 3D,E). Only three positive regulators of canonical Wnt signaling had ASE in dental tissues, precluding a separate statistical test of their directionality. As disruption or inhibition of canonical Wnt signaling results in arrested/aberrant tooth formation, selection on this pathway could potentially reflect selection for increased tooth number or related changes in feeding morphology in freshwater fish. Consistent with this, genes in the Wnt signaling pathway were previously shown to be upregulated in the VTP in PAXB freshwater compared to marine fish [19].

The GO term "embryonic viscerocranium morphogenesis", which encompasses a set of genes involved in the generation and organization of the facial skeleton, also included ASE genes directly implicated in tooth and jaw formation (Table S8). For instance, *Dlx3b* and *Dlx1a*, genes encoding members of the Dlx family of homeodomain transcription factors [35], are involved in tooth and jaw patterning in mammals and fish [36,37]. Thus, biased directionality of this process category may reflect selection for morphological changes to the freshwater fish in the facial region related to feeding morphology.

We also found biased directionality for gene sets in individual tissues (Table S5). For instance, the GO category term "methyltransferase activity" ($P$=0.0018, 10/10 terms) showed biased upregulation of marine alleles in the eye and "endoplasmic reticulum" ($P$=0.0039, 38/50) showed biased upregulation of marine alleles in the flank skin. Since these gene sets are not yet associated with specific phenotypes, it is unclear what traits may have been impacted by their lineage-specific selection.

***Overlap between signatures of selection and genes with cis-regulatory divergence***

If *cis*-regulatory changes underlie adaptive divergence between freshwater and marine forms, we may expect genes with ASE to fall within or near regions with signatures of selection. To test this hypothesis, we utilized a recent whole-genome analysis of differentiation between marine and freshwater populations from the northeast Pacific basin [9] (the source of the freshwater PAXB population studied here), where genomic regions of repeated marine-freshwater divergence were identified through marine-freshwater cluster separation scores (CSS). A CSS score quantifies average marine-freshwater genetic distance after subtracting the genetic distance found within each ecotype for a genomic window [8,9].

We asked whether genes with ASE co-localized with genomic regions with greater evidence for marine-freshwater divergence (i.e., greater CSS Z-scores). As power to detect ASE is related to the number of variant sites, we compared median CSS Z-scores between ASE and background genes with similar SNP densities (see Methods, Figure S7). Genes with ASE were associated with greater Z-scores per SNP density bin (Figure 4A,B; Permutation $P<0.0001$), indicating an enrichment of ASE genes in genomic regions with greater evidence for marine-freshwater divergence. This pattern is consistent with the hypothesis that repeated marine-freshwater divergence may often involve changes in gene regulation [8,18,18].

Regions with significant CSS scores (EcoPeaks) overlapped 611 ASE genes (13.8% of ASE genes overall; 1.9-fold enrichment, Permutation test $P<0.001$)(Figure 4C, Table S9). Genes with evidence for ASE heterogeneity between tissues were enriched within EcoPeaks compared to all genes with evidence for ASE (Fisher's exact test, $P=0.01$), as were genes with evidence for tissue-specific ASE (Fisher's exact test, $P=0.018$).

Marine-freshwater EcoPeaks are clustered throughout the genome, which is thought to reflect selection on linked "supergene" complexes affecting multiple traits [9,38]. We also find that genes with ASE are enriched on particular chromosomes (Figures S8, S9; see Methods). EcoPeaks and QTL associated with phenotypic divergence are particularly concentrated on ChrIV and this chromosome also harbored the highest proportion of ASE genes over background (Permutation $P<0.001$) as well as a quarter of EcoPeak ASE genes (154 genes). A more modest enrichment of ASE genes was also found for chrXXI ($P=0.034$) and chrXI ($P=0.033$), which have been shown to harbor inversions between marine and freshwater fish [8]. We identified 56 and 48 ASE genes within EcoPeaks on these chromosomes, respectively.

To identify potential candidate genes for marine-freshwater divergence, we overlapped ASE genes identified in marine-freshwater peaks with QTL for dental and skeletal traits [6,22,39](Figure 4C,D). QTL for variation in dental traits between PAXB freshwater and marine fish (e.g., VTP or DTP tooth plate size and shape, tooth number, and jaw size and shape) overlapped 401 genes with ASE in relevant tissues (File S1). A small subset of these have previously been implicated in dental or craniofacial morphology in other species (Table 1), including several genes involved in the Wnt signaling pathway identified in the sign test (e.g., *Wnt5a*, *Sfrp2*, *Ctnnb1*, *Net1*).

**Conclusions**

Changes in gene expression regulation are thought to play a major role in evolutionary adaptation. Here, we surveyed allele-specific expression across tissues and developmental stages to understand the landscape of *cis*-regulatory divergence between marine and freshwater sticklebacks. We identified widespread ASE that was largely heterogeneous between tissue types and developmental stages. For a subset of these *cis*-regulatory changes, we found evidence for polygenic selection on particular processes/pathways with a sign test. Finally, we demonstrated that *cis*-regulatory changes are often associated with regions of marine-freshwater divergence, further supporting the role of *cis*-regulatory differences in adaptive evolution in sticklebacks [8,9].

Our results indicate the *cis*-regulatory divergence between marine and freshwater fish is often specific to an individual tissue or developmental stage. Gene expression differences that are spatially or temporally restricted may be important in the process of adaptation to new environments. Through context-specific expression regulation, *cis*-regulatory mutations can avoid negative pleiotropy associated with global changes in expression or protein structure. Thus, it is possible that *cis*-regulatory variation that introduces discrete changes in gene expression may be favored during adaptation. Interestingly, we found that genes with evidence for tissue-specific ASE in particular were enriched in regions of recurrent marine-freshwater divergence. Tissue- or context- specific *cis*-regulatory differences have previously been shown to underlie adaptive traits in sticklebacks [4,14] and other systems[40]. The tissue- and developmental-specificity of *cis*-regulatory changes we identified between marine and freshwater sticklebacks highlights the utility of studying gene regulation across multiple tissues and contexts in understanding regulatory adaptation.

Genes with ASE in regions of repeated marine-freshwater divergence may be interesting candidates for adaptive phenotypic differences between ecotypes. *Cis*-regulatory changes have been found to underlie a number of phenotypic differences between marine and freshwater forms. For example, *cis*-regulatory changes at *Bmp6* are associated with evolved tooth gain[6,15], and *cis*- regulatory changes at *Eda* and *GDF6* have been implicated in skeletal differences between marine and freshwater fish[7,41]. While we did not have sufficient expression to examine ASE at these genes specifically (e.g., *Eda, GDF6*, *Bmp6*) in our dataset, we identified a number of potentially interesting candidate genes within differentiated genomic regions with ASE. For example, *cis*-regulatory variation at *Stanniocalcin2a* (*Stc2a*) was recently associated with changes in spine length in freshwater sticklebacks [16]. We found that *Stc2a* also showed ASE in the early mandible timepoint. *Stc2a* falls within a marine-freshwater divergent region on ChrIV that overlaps several QTL, including the QTL with the largest effect on dentary size in crosses between PAXB freshwater fish and marine fish [22]. In mice, *Stc2a* modulates bone size and growth and

342 overexpression results in smaller mandibles[42,43], making this gene an exciting candidate for divergent
343 jaw morphology between marine and freshwater benthic fish. Our results also highlighted Wnt signaling
344 genes as potential candidates for divergence in feeding morphology. A sign test indicated evidence for
345 lineage-specific selection on *cis*-regulatory alleles involved in Wnt signaling, and several of these genes
346 were also found within QTL/EcoPeaks and involved in tooth development or craniofacial morphology
347 (see Table S1). For example, *Wnt5a,* associated with QTL for tooth plate and dentary shape, plays an
348 important role in facial and tooth development in mammals [44–46]. Our results establish the landscape
349 of stickleback *cis*-regulatory divergence across tissues and developmental stages; we look forward to
350 future studies that elucidate the roles that specific ASE genes have played in stickleback adaptation.
351
352 **Methods**
353 ***Stickleback husbandry***
354 All animal work was approved by UC Berkeley IACUC protocol AUP-2015–01-7117. Fish were raised in
355 aquaria at 18ºC in brackish water (3.5g/L Instant Ocean salt, 0.217mL/L 10% sodium bicarbonate) with 8
356 hours of light per day. Fry (SL < 10mm) were fed live Artemia, early juveniles (SL ~10-29 mm) were fed
357 live Artemia and frozen Daphnia. Fish above ~20mm were fed frozen bloodworms and Mysis shrimp. To
358 generate F1 hybrids, a freshwater Paxton Benthic (Paxton Lake, Canada) strain male was crossed with a
359 marine Rabbit Slough (Alaska) strain female. Individuals from these lineages have been maintained in the
360 lab for >10 generations. The resulting full-sibling fish were raised together in a common dish or tank until
361 sample collection. Female F1 hybrids were selected for dissection at two timepoints (15-20 mm SL and
362 35mm SL). Fish were euthanized individually via immersion in 250 mg/L MS-222. Tissue samples for
363 RNA-seq were immediately dissected on an ice-cold tray. Brain samples included all bilateral brain
364 regions from the olfactory bulb to the brain stem. Liver samples were derived from the anteriormost lobe
365 of the fish liver. Eye samples encompassed the entirety of the left eye of each fish, including the majority
366 of the optic nerve. Flank skin samples were taken by removing the majority of the skin covering the left
367 side of each fish, capturing a region that would normally be covered by lateral armor plates in adulthood
368 (anteriormost boundary at the level of the 1$^{st}$ dorsal spine, where the anteriormost armor plates had begun
369 ossification at the time of dissection, posterior boundary at the back of the dorsal fin where armor plates
370 were not yet ossified). Dorsal pharyngeal tooth plate samples included left and right DTP1 and DTP2, as
371 well as underlying epibranchial bones and surrounding soft tissues and teeth. Ventral pharyngeal tooth
372 plate samples included left and right ceratobranchial 5 and surrounding soft tissues and teeth. The
373 mandible consisted of the dentary bone and lower lip, and all associated soft tissues and teeth. Samples
374 were placed into 50 ul of TRIzol (Invitrogen), briefly agitated by shaking, and incubated on ice for 10
375 minutes. All samples from each timepoint were all prepared on the same day.

376

377 ***RNA extraction, library preparation, and sequencing***

378 Dissected tissues were kept in TRI reagent and stored in -80°C prior to RNA-extraction. Total RNA
379 extraction was performed as described previously [19]. Total RNA was quantified by Qubit Fluorometer,
380 and quality was checked by Agilent Bioanalyzer. Libraries were constructed with New England Biolabs
381 NEBNext Poly(A) mRNA Magnetic Isolation Module (E7490S), NEBNext Ultra II Directional RNA
382 Library Prep Kit (E7765S) and NEBNext Multiplex Oligos for Illumina (96 Unique Dual Index Primer
383 Pairs, E6440S) following the manufacturer's instructions. Library quality was analyzed on an Agilent
384 Bioanalyzer (Table S1). Libraries were pooled and sequenced on an Illumina HiSeq platform (2x150 bp
385 reads). We obtained a total of 1,439,700,457 reads across 20 samples (10 tissues x 2 replicates) (Table
386 S1).

387

388 ***Whole-genome re-sequencing of PAXB***

389 To phase RNA-seq reads, whole-genome resequencing was performed on the PAXB parent. DNA was
390 extracted from fin tissue. Library preparation and sequencing were performed by Admera Health (South
391 Plainfield, NJ). Libraries were sequenced on an Illumina HiSeqX platform (2x150 bp reads) to a depth of
392 ~30X (Figure S1). Coverage per site was calculated with Samtools depth [47] based on reads aligned to
393 the reference genome (described below).

394

395 ***Read mapping and SNP calling***

396 RNA-seq read quality was assessed using FastQC. Reads were trimmed for adaptor sequences with
397 Trimmomatic[48] and then mapped to the stickleback reference genome [49]. F1 hybrid RNA-seq reads
398 were mapped to the stickleback reference genome with STAR v2.7 [50]. Genomic reads from PAXB
399 were mapped with bowtie2 v2.3.4 (argument: --very-sensitive)[51].

400

401 SNP calling was then performed with the Genome Analysis Tool Kit (GATK)[52]. Duplicates were
402 marked with the Picard tool MarkDuplicates. Read groups were added with AddOrReplaceReadGroups.
403 For RNAseq reads, we used GATK tool SplitNCigarReads to split reads that contain Ns in their cigar
404 string (e.g., spanning splice events). GATK HaplotypeCaller and GenotypeGVCFs were used for joint
405 genotyping. SNP calls were subsequently filtered for low quality calls with VariantFiltration (QD < 2.0;
406 QUAL < 30.0; FS > 200; ReadPosRankSum < -20.0).

407

408 To assign allele-specific reads to the parent of origin (i.e., "freshwater" parent vs. "marine" parent), we
409 retained only variants where the PAXB parent was homozygous. Heterozygous sites for each F1

individual were used for separating allele-specific RNA-seq reads into freshwater and marine pools, as described below.

***Identifying allele-specific expression***

To identify allele-specific expression (ASE), reads from each library were then mapped again with STAR, implementing the WASP filter based on heterozygous calls [53]. WASP reduces mapping bias by identifying reads containing SNPs, simulating reads with alternative alleles at that locus, re-mapping these reads to the reference, and then flagging reads that do not map to the same location. Reads that do not map to the same location were discarded[53]. Parental origin for each allele was assigned based on PAXB (freshwater parent, see above). Reads were counted over marine-freshwater variants with ASEReadCounter [52] for individual heterozygous sites. To mitigate the effects of SNP calling errors and read mapping bias, we removed heterozygous sites with: 1) large ratio differences indicative of mapping bias ($\log_2$ fold changes of allelic counts > 10), or 2) no reads mapped to one of the parental alleles. Mapping was then repeated a second time based on the updated list of heterozygous sites. Analysis of ASE ratios in each library centered around a $\log_2$ ratio of zero, indicating approximately equal mapping to both parental alleles.

Gene-wise estimates of allele-specific expression were quantified by counting allele-specific reads overlapping exons using HTSeq [54] based on Ensembl annotations (BROAD S1) [8], with coordinates converted by LiftOver to the v4 stickleback assembly (https://stickleback.genetics.uga.edu/downloadData/)[49]. Total counts per parental allele per tissue are available in Table S2. Across tissues, we did not observe a consistent bias towards either of the parental alleles. To examine transcriptome-wide patterns of expression, we transformed expression values (allele-specific and total counts) using variance stabilizing transformation and assessed transcriptome-wide expression patterns via principal components analysis (PCA)(Figures 1B, S2).

DESeq2 [55] was used to identify ASE using the individual as a blocking factor and allele-specific expression ("marine" vs. "freshwater" allele) as the variable of interest (Wald test). As read counts from "marine" and "freshwater" alleles come from the same sequencing library, library size factor normalization was disabled by setting SizeFactors = 1. *P*-values were adjusted using the Benjamini-Hochberg method in DESeq2 for multiple comparisons. Genes were examined at FDR<0.05 and FDR<0.1 (Table S3). Comparing genes with ASE in VTP from the late timepoint with the results of Hart et al. [19], which also tested for ASE in crosses between PAXB and RABS in the VTP, we found highly significant overlap (Fisher's exact test, $P=8.83\text{x}10^{-292}$). Fifty-one percent of ASE genes identified here

were also identified in the previous analysis. Additionally, $\log_2$ fold changes of genes with ASE were found to be correlated (Pearson's Correlation, $r$ =0.64, $P$=2.47x10$^{-69}$).

Developmental stage differences in ASE were identified by comparing reads mapping to freshwater vs. marine alleles at both timepoints, summed across the two replicates. We compared marine and freshwater allelic ratios for genes with evidence of ASE in at least one of the two developmental stages with a Fisher's exact test. Resulting $P$-values were corrected using the Benjamini-Hochberg method.

***Assessing heterogeneity in allele-specific expression across tissues***

To identify heterogeneity in allele-specific expression across tissues, we adopted a Bayesian model comparison framework from Pirinen et al. [23]. In this approach, tissues are classified as no ASE ($\theta(\mathcal{N})$), strong ASE ($\theta(\mathcal{S})$), or moderate ASE ($\theta(\mathcal{M})$) based on freshwater and marine allelic counts summed across replicates per gene under a grouped tissue model [23]. Tissues are further classified as showing ASE heterogeneity if one tissue showed evidence for either strong or moderate ASE and at least one other tissue did not show ASE (HET0) or when all tissues showed some evidence for ASE but the magnitude differed (HET1). Finally, we consider a sub-state of ASE heterogeneity tissue-specificity, where ASE state (i.e., moderate, strong ASE, or no ASE) is observed in only one tissue [23].

The following priors were selected to describe groups:

$$\theta(\mathcal{N}) \sim \text{Beta}(2000,2000)$$

$$\theta(\mathcal{M}) \sim \tfrac{1}{2}\text{Beta}(80,36) + \tfrac{1}{2}\text{Beta}(36,80)$$

$$\theta(\mathcal{S}) \sim \tfrac{1}{2}\text{Beta}(80,7) + \tfrac{1}{2}\text{Beta}(7,80)$$

Densities of the prior distributions for the proportion of allelic counts are found ins Figure S9. Parameters for Beta distributions were chosen to clearly separate the three groups from each other to allow the classification of tissues to a particular group, following Pirinen et al. [23]. The "No ASE" condition dominates the region around 0.5 (0.47,0.53), allowing for some deviation for technical bias or noise [23,56]. Strong ASE dominates at extreme frequencies ([0.85,0.96],[0.3,0.15]) and moderate ASE dominants between these two groups (Figure S9). Genes expressed in at least two tissues at a minimum depth of 10 reads/allele were included in the analysis (15,477 genes). For each gene, we excluded tissues for which coverage was low (less than or equal to 10 reads per allele).

***Sign test on ASE***

To search for selection on cis-regulatory variation, we applied a sign test based on the directionality of ASE in a gene set [30,31]. Gene Ontology (GO) categories for zebrafish were obtained from ZFIN (https://zfin.org/downloads)[57] and mapped to stickleback orthologs based on Ensembl ortholog annotations. Genes with evidence for *cis*-regulatory divergence were divided into categories based on the upregulated allele (freshwater vs. marine). We excluded GO categories with fewer than 10 members in a test set with ASE. As test sets contain different proportions of upregulated marine vs. freshwater alleles, we tested for lineage-specific bias in each test set with a Fisher's exact test.

Because many GO categories were tested, we determined the probability of an enrichment by permuting gene category assignments, as described previously [30,31,58]. Gene assignments were shuffled and the test was repeated 10,000 times. Permutation-based *P*-values were determined by asking how often a result of equal or greater significance would be observed in permuted datasets [30,31,58]. Tests were performed on individual tissues and on the dental tissues together, as many genes with ASE are shared across these tissues. In the grouped tissue analysis, we looked for biased directionality across all genes with ASE in a tissue group. In the event that signs differ between tissues (i.e., freshwater allele is upregulated in tissue #1, marine allele is upregulated in tissue #2), the gene is discarded from the analysis. Changes in directionality across tissues was only seen for one gene associated with a significant GO term (dental tissue: "inflammatory response"). To ensure that biased directionality in our group analyses were robust to tissue-specific patterns, we performed a second test where we combined *P*-values for a tissue group from individual tissues with Fisher's method, as in [31]. We performed a Fisher's exact test for each category as described above for individual tissues. *P*-values for GO categories that are represented across all tested groups were then combined using the R package metap. We report on GO terms significant in both approaches, as these represent cases of biased directionality across tissue groups and robust to individual tissue patterns. Combined *P*-values are reported in Table S6.

***Identifying overlap with EcoPeaks***

Data from Kingman et al. [9] was downloaded from the UCSC Genome Browser Table Browser ([22]: https://sbwdev.stanford.edu/kingsleyAssemblyHub/hub.txt). Intervals were associated with overlapping genes using bedtools [59]. As power to detect ASE is related to the density of informative sites, we calculated SNP density per gene as the number of informative heterozygous sites divided by transcript length, based on BROAD S1 gene annotations [9]. To determine whether genes with ASE had higher average Z-scores than background genes, while controlling for the effect of SNP density on our power to identify ASE, we grouped genes with similar SNP densities into bins based on the distribution of SNP density values. We excluded bins for which there were fewer than five genes in each category

(ASE, no ASE) so as not to skew results based on bins with few observations (330 bins, average of 23 genes per bin)(Figure S8A). For each bin, we calculated the median Z-score for genes with ASE and for background genes (Figure S8B). We compared this result to a permuted dataset. Within a density bin, we shuffled gene category assignments and again calculated the median Z-score for each category in each bin. We repeated this 10,000 times. To obtain a permutation-based *p*-value, we compared how often the median difference in category Z-scores was as extreme or more extreme than in empirical data. This result was robust to varying bin sizes (Table S10).

To ask whether ASE genes were enriched on particular chromosomes, as with QTLs and EcoPeaks [9], we performed a resampling test to account for differences in SNP density between genes. We sampled random sets of genes (equal to the number of total ASE genes) with SNP densities matched to the ASE gene set (1000 times). The number of genes associated with each chromosome were counted for each permuted gene set and compared to the empirical data. *P*-values were calculated based on how often an equal or more extreme result was observed for permuted gene sets (Figure S9).

***Gene annotations and QTL overlap***

Stickleback genes were annotated to zebrafish and mouse orthologs based on Ensembl ortholog annotations. BiteCode genes were annotated as in Hart et al. [19], from the BiteIt database (http://bite-it.helsinki.fi/) and ToothCODE database (http://compbio.med.harvard.edu/ToothCODE/). Phenotype annotations for zebrafish were downloaded from ZFIN (https://zfin.org/downloads), mouse mutant phenotypes were downloaded from Ensembl and the Mouse Genome Database[60].

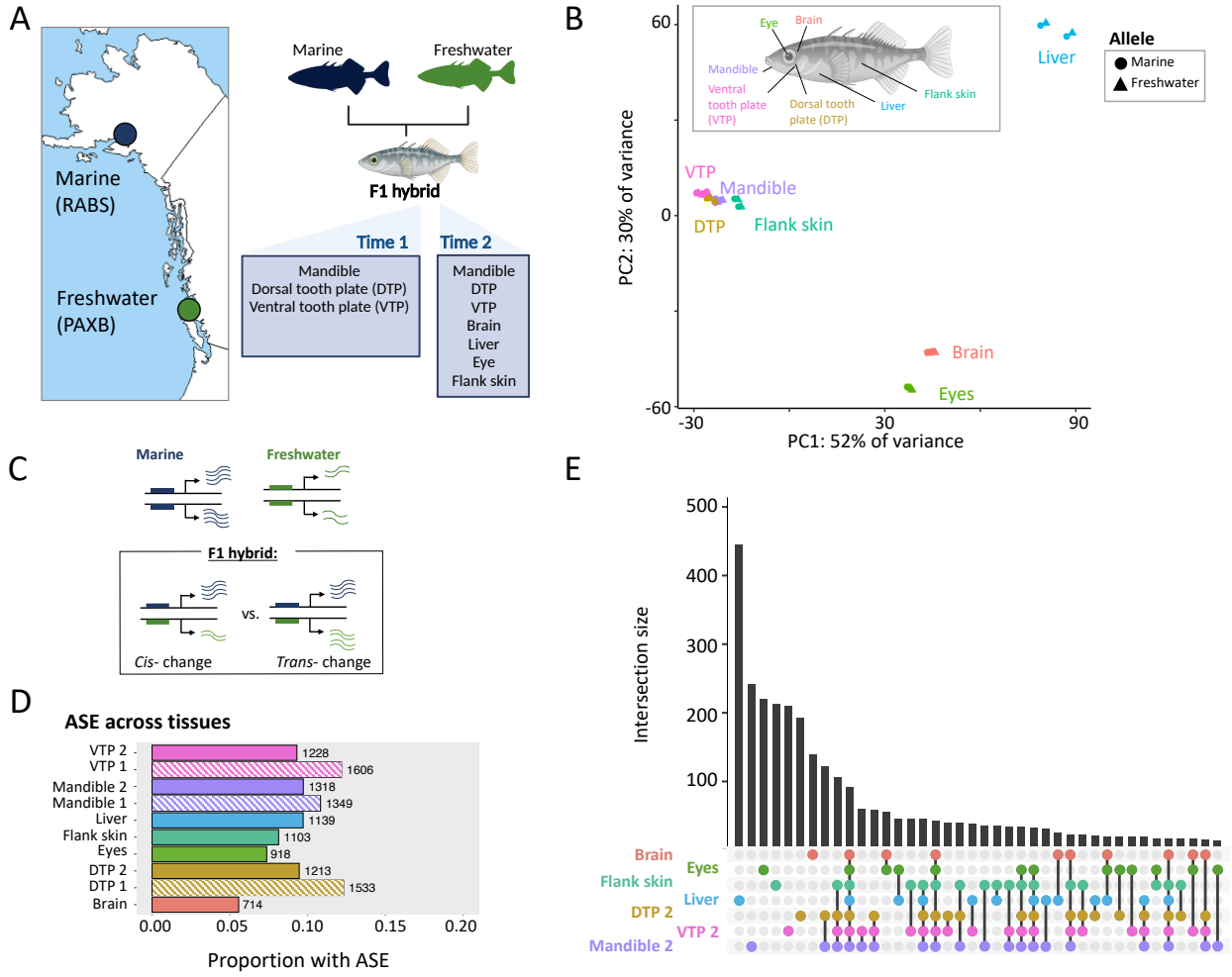
QTL coordinates for overlap are based on genomic coordinates in Marques and Peichel [61]. For overlap with ASE genes, we focused on QTL mapping studies utilizing crosses between PAXB freshwater and marine individuals. Dental QTL for Table S1 and File S1 were obtained from three studies [6,22,39]. Coordinates were converted by LiftOver to the v4 stickleback assembly for overlap with EcoPeaks. Genes of interest for Table 1 were identified based on intersections between these genes (ASE/EcoPeak/QTL) and phenotype/Gene Ontogony annotations or the BiteCode gene list. A full list of gene overlaps is available in File S1.

**Data Availability**

All sequence data generated in this study have been deposited to the National Center for Biotechnology Information Sequence Read Archive as a BioProject (SUB12080818). Supplemental datasets are available

545 in File S1. Scripts associated with this manuscript are available on GitHub
546 (https://github.com/katyamack-hub/SticklebackASE).
547
548 **Funding**
549 This work was supported by the National Institutes of Health (R01-GM097171 to H.B.F., DE021475 to
550 C.M. and DE027871 to T.S.) and the Ruth L. Kirschstein National Research Service Award Individual
551 Postdoctoral Fellowship(F32) to K.L.M.
552
553 **Supplemental Material**
554 Supplement Figures and Tables
555 Figues S1-S10
556 Tables S1-S10
557 Supplemental references
558
559 File S1: Supplemental gene lists

560 **Figures**




561

562 **Figure 1.** The *cis*-regulatory landscape of divergence between marine and freshwater stickleback. **A**. Marine
563 (RABS) and freshwater (PAXB) sticklebacks were crossed to produce F1 hybrids. Tissues were collected for
564 RNAseq across two developmental timepoints from full siblings. **B**. Principal component analysis of allelic counts
565 from marine-freshwater F1 hybrids. Allelic reads cluster by tissue of origin on PC1 and PC2. **C**. A schematic of how
566 regulatory divergence can be dissected with F1 hybrids. Here, a gene is upregulated in marine fish (wavy lines). In
567 an F1 hybrid, differential expression between the freshwater and marine allele (e.g., allele-specific expression, ASE)
568 indicates a *cis*-regulatory change. In contrast, equal expression of the two alleles indicates a *trans*- only change. **D**.
569 Numbers and proportions of genes with ASE across tissues. Striped bars indicate tissues from the early timepoint. **E**.
570 UpSet plot showing distinct intersections of genes with ASE across tissues from the late timepoint. A single dot in a
571 column indicates ASE specific to one tissue, whereas multiple dots connected by lines indicate ASE shared across
572 multiple tissues.

573

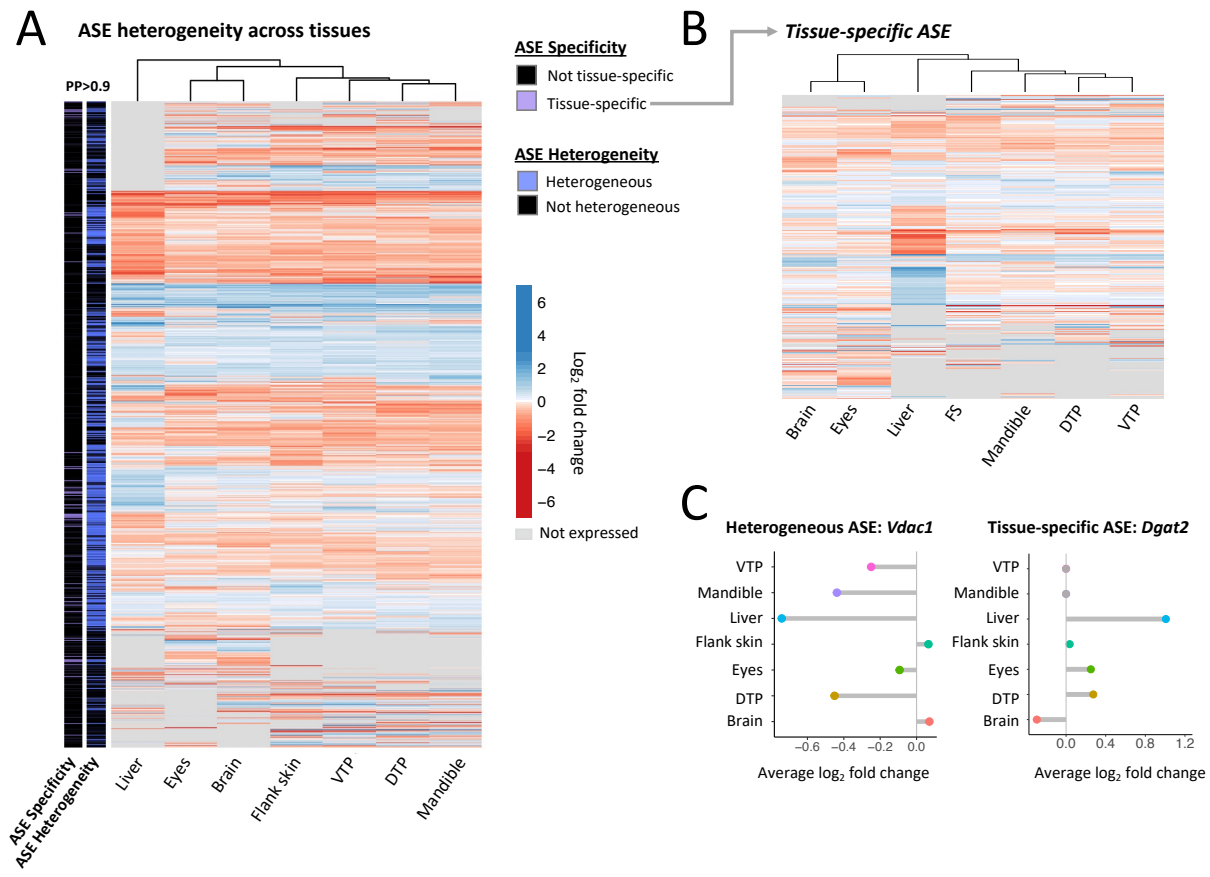



**Figure 2.** Heterogeneity of allele-specific expression across tissues. **A.** A heatmap of genes with evidence of allele-specific expression in late development. The left bars indicate genes where ASE varies across tissues (in presence or magnitude, "ASE Heterogeneity") and substate of ASE heterogeneity where ASE patterns are specific to one tissue (tissue-specific ASE, "ASE Specificity") at a posterior probability of >0.9. Genes are colored by average $\log_2$ fold change in each tissue. Gray panels indicate the gene is not expressed in a given tissue. **B**. Heatmap of genes with evidence for tissue-specific ASE. **C**. Examples of genes with heterogeneous ASE (*Vdac1,* left) and tissue-specific ASE (*Dgat2*, right). ASE was observed for voltage-dependent anion channel *Vdac1* in some tissues (e.g., liver, DTP2, mandible) but not others, where triglyceride synthesis gene *Dgat2* only shows evidence for ASE in liver.

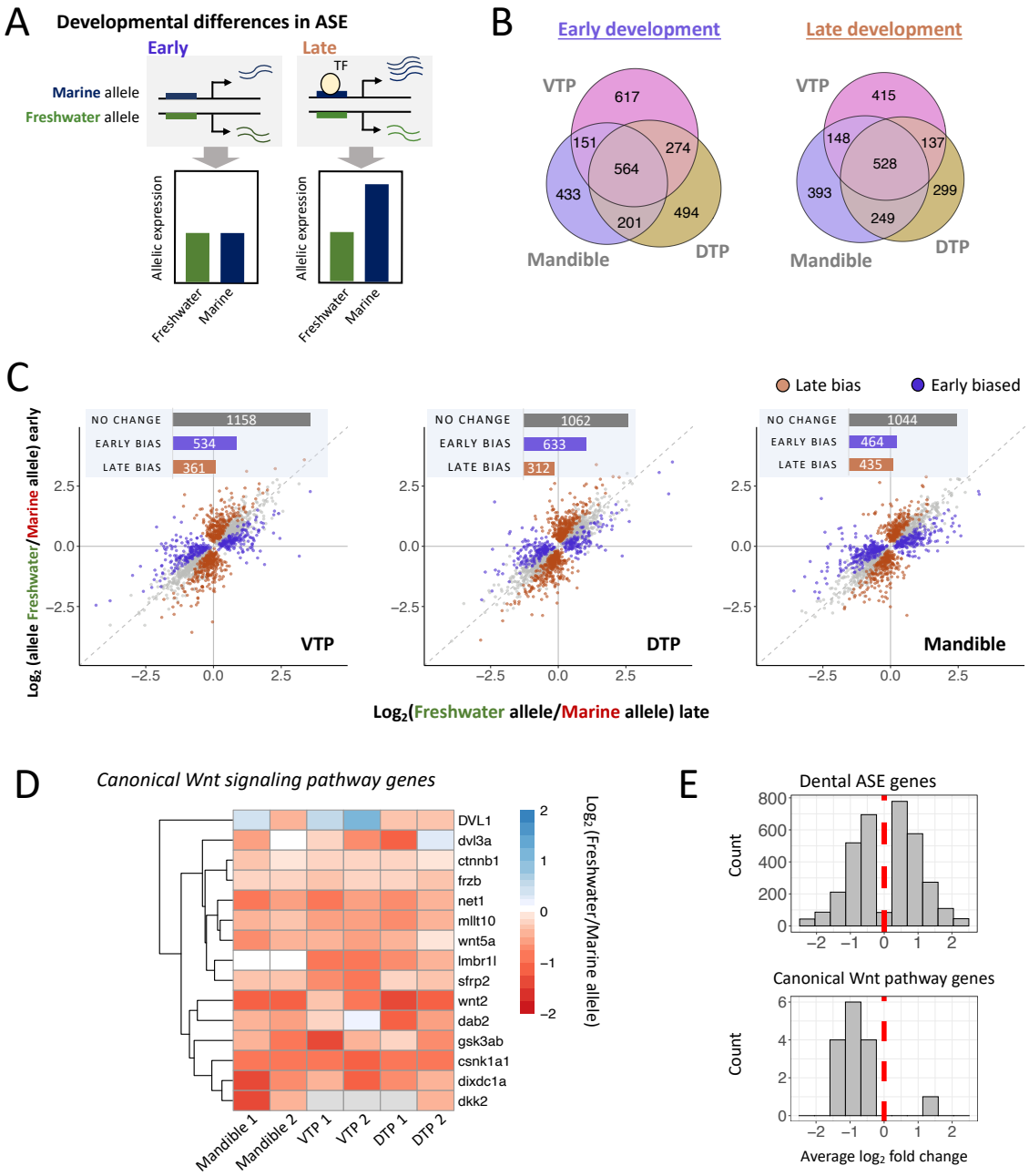



**Figure 3.** Developmental allele-specific expression in dental tissues. **A**. A schematic of differential ASE during development. In this example, sequence divergence between marine and freshwater sticklebacks at a *cis*-regulatory region results in allele-specific expression only in the presence of a context-specific transcription factor ("TF", circle) expressed during late development. This results in differential allele-specific between developmental stages, shown in the bar plots. **B**. Venn diagram of ASE between dental tissues at the early (left) and late (right) developmental timepoint. **C**. Temporal differences in ASE during development in three dental tissues (left to right: ventral tooth plate, dorsal tooth plate, mandible). Genes with differential allele-specific between timepoints are colored based on magnitude of ASE in timepoint 1 vs. 2. Genes with greater differences in allelic expression in early

development are shown in purple ("early bias"), genes with greater expression differences at the late timepoint are shown in terracotta ("late bias"). Gray points/bar ("no change") indicate genes without evidence for significant differential allele-specific expression between timepoints. **D-E**. Genes involved in Canonical Wnt signaling show evidence of polygenic *cis*- regulatory evolution in dental tissues. Here we show genes from two Wnt signaling GO terms with biased directionality (Canonical Wnt signaling [GO:0060070]; Negative regulation of canonical Wnt signaling [GO:0090090]) (Table S7). In (**D**), the heatmap shows $\log_2$ fold changes for genes associated with canonical Wnt signaling and with ASE in at least one dental tissue. In (**E**), histograms of average ASE gene $\log_2$ fold changes from all dental genes (top) and the canonical Wnt signaling gene set (bottom). For each gene, $\log_2$ fold changes are averaged across any dental tissues in which ASE was identified.

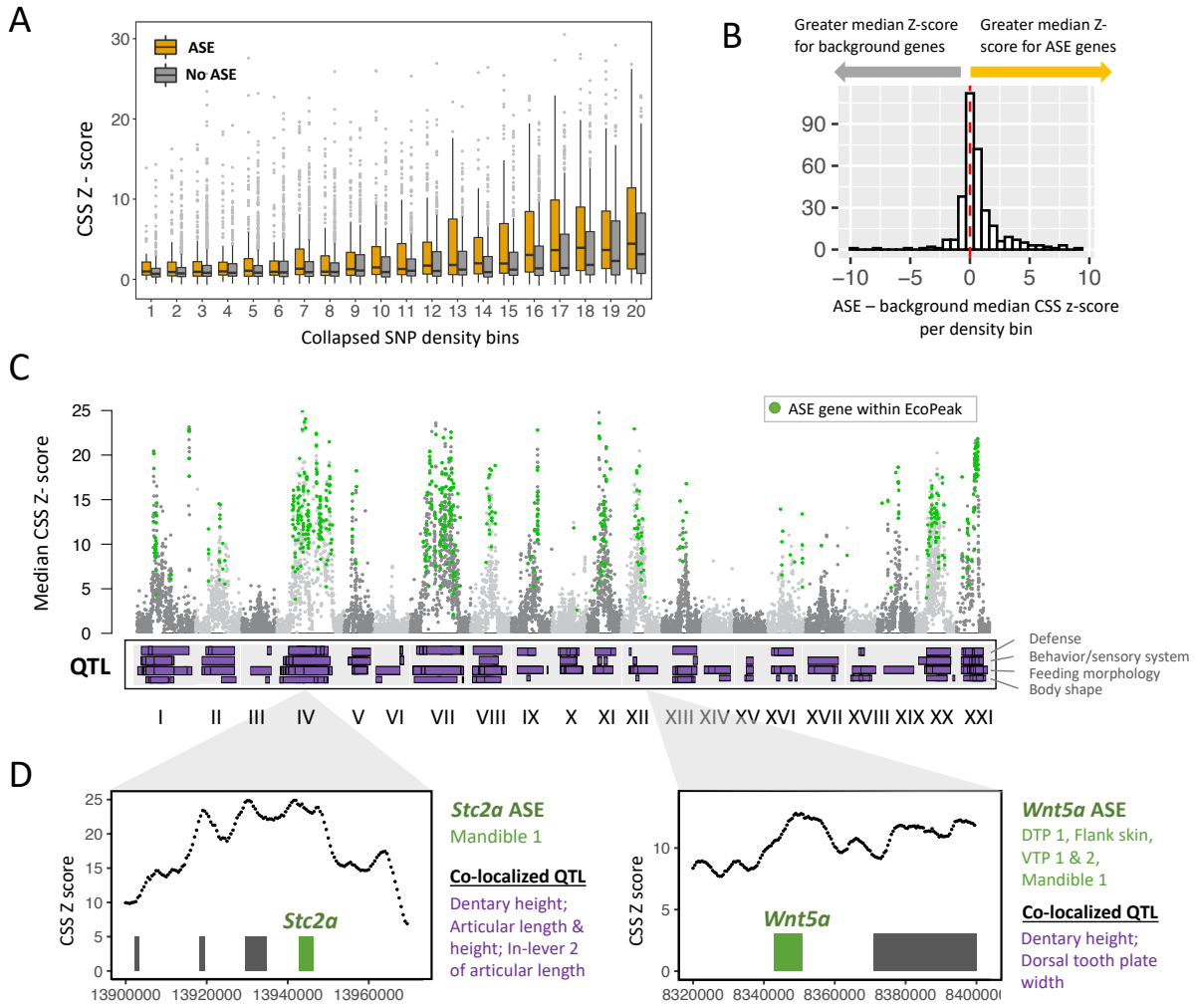



**Figure 4.** ASE genes are associated with regions of repeated marine-freshwater divergence. **A**. Average marine-freshwater cluster separation score (CSS) Z-scores for ASE genes and background genes binned by SNP density. Here, genes are separated into 20 density bins for visualization, with higher numbers corresponding to greater SNP density. **B**. ASE genes have higher average CSS Z-scores than background genes. The histogram shows median Z scores for ASE genes minus background genes for each SNP density bin. More density bins show positive values, indicating higher average Z-scores for ASE genes overall (Permutation $P<0.0001$). **C**. Manhattan plot of median gene CSS Z score vs. chromosome position. Highlighted in green are genes with ASE that overlap significant regions of recurrent marine-freshwater divergence in the northeast Pacific basin ("EcoPeaks"). Below we show locations of quantitative trait loci (QTLs) identified in previous genetic crosses between PAXB and marine fish. QTL are divided into three broad categories (from top to bottom: defense, behavior and sensory system, feeding morphology, and body shape). **D**. Two candidate ASE genes within regions of marine-freshwater divergence. ASE was observed for *Wnt5a* and *Stc2a* in one or more dental tissues and these genes co-localize with QTL related to feeding morphology. Bars in the panel indicate genes within these regions, with candidate genes *Wnt5a* and *Stc2a* highlighted in green. Tissue(s) in which ASE was identified (green) and relevant overlapping QTL (purple) are listed to the right of each gene panel.

**Table 1. Candidate ASE genes within differentiated regions with overlapping QTL**

| Gene | Marine-Freshwater EcoPeak | ASE tissues | QTLs[1] | Relevant functions |
|---|---|---|---|---|
| *Cldn4* | chrI:16985780-17009145 | DTP 1 & 2, VTP 1, Mandible 1, Liver | Tooth plate shape | Tooth development [62] |
| *Igfbp5a* | chrI:26093586-26548184 | DTP 2 | Tooth plate shape | Tooth development, bone development [63,64] |
| *Scube1* | chrIV:21368021-22019696 | DTP 1, VTP 1, Mandible 2, Eyes | Tooth number, jaw shape | Craniofacial development[65] |
| *Stc2a* | chrIV:13853040-14033725 | Mandible 1 | Jaw shape | Skeletal development [42] |
| *Net1* | chrIV:21368021-22019696 | DTP 1, VTP 1 & 2, Mandible 1 & 2, Flank skin | Tooth plate tooth number, jaw shape | Tooth development, bone development [66] |
| *Kdm5a* | chrIV:26410831-26967766 | VTP 1 | Tooth plate tooth number | Tooth development[67] |
| *Kdm6bb* | chrVII:19682650-19906375 | DTP 1 & 2, VTP 1 | Tooth plate area and shape, tooth plate tooth number, dentary shape | Tooth development, bone development [68,69] |
| *Cldnb* | chrVII:21853746-22015247 | DTP 1 & 2, VTP 2, Mandible 1 & 2, Flank skin | Jaw, dentary, and tooth plate shape, tooth plate tooth number, defense plates | Tooth development [62] |
| *Postna* | chrVII:22876973-23002851 | DTP 1 & 2, Flank skin, VTP 2, Mandible 1 & 2, eyes | Jaw shape, dentary shape, tooth plate shape, tooth plate tooth number, defense plates | Tooth development, bone development [70] |
| *Kdm6ba* | chrVII:8555721-8746343 | Mandible 1, Liver | Jaw shape | Bone development [69] |
| *Timp2b* | chrXI:9651474-9924679 | DTP 1 & 2, Flank skin, eyes, liver | Tooth plate tooth number | Tooth development[71] |
| *Mmp9* | chrXII:10684220-10754257 | DTP 2, Flank skin | Tooth plate shape | Tooth development, bone development [72,73] |
| *Itga5* | chrXII:7057481-7113347 | DTP 1 | Tooth plate shape | Tooth development [74] |
| *Wnt5a* | chrXII:8202228-8410911 | DTP 1, Flank skin, VTP 1 & 2, Mandible 1 | Tooth plate shape, dentary shape | Tooth development, facial development[44,45] |
| *Tgfbr1b* | chrXXI:3449938-3520071 | VTP 1, DTP 2 | Tooth plate area, tooth plate shape | Tooth development [75] |
| *Sulf1* | chrXXI:9696109-11646044 | DTP 1, VTP 1 & 2, Mandible 2, eyes | Tooth plate shape, jaw shape | Tooth development, skeletal development[76,77] |
| *Bmi1a* | chrXXI:9696109-11646044 | VTP 1 | Tooth plate tooth number, tooth plate shape | Tooth development [78] |
| *Mllt10* | chrXXI:9696109-11646044 | DTP 1 & 2, VTP 1 & 2 | Tooth plate shape, tooth plate tooth number | Craniofacial development[79] |

[1]QTL data: Miller et al. [22], Cleves et al. [6], Erickson et al. [39]

626 **References**

627 1. Reid K, Bell MA, Veeramah KR. Threespine Stickleback: A Model System For
628 Evolutionary Genomics. Annual Review of Genomics and Human Genetics. 2021;22: 357–
629 383. doi:10.1146/annurev-genom-111720-081402

630 2. Bell M, Foster S. The evolutionary biology of the threespine stickleback. Oxford, UK:
631 Oxford University Press; 1994.

632 3. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al.
633 Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles.
634 Science. 2005;307: 1928–1933. doi:10.1126/science.1107239

635 4. Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, et al. cis-Regulatory
636 Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks
637 and Humans. Cell. 2007;131: 1179–1189. doi:10.1016/j.cell.2007.10.055

638 5. Walker JA, Bell MA. Net evolutionary trajectories of body shape evolution within a
639 microgeographic radiation of threespine sticklebacks (Gasterosteus aculeatus). Journal of
640 Zoology. 2000;252: 293–302. doi:10.1111/j.1469-7998.2000.tb00624.x

641 6. Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, et al. Evolved
642 tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. Proceedings of
643 the National Academy of Sciences. 2014;111: 13912–13917. doi:10.1073/pnas.1407567111

644 7. O'Brown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. A recurrent regulatory
645 change underlying altered expression and Wnt response of the stickleback armor plates
646 gene EDA. Krumlauf R, editor. eLife. 2015;4: e05290. doi:10.7554/eLife.05290

647 8. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic
648 basis of adaptive evolution in threespine sticklebacks. Nature. 2012;484: 55–61.
649 doi:10.1038/nature10944

650 9. Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, et al. Predicting
651 future from past: The genomic basis of recurrent and rapid stickleback evolution. Science
652 Advances. 2021;7: eabg5285. doi:10.1126/sciadv.abg5285

653 10. Stern DL, Orgogozo V. Is genetic evolution predictable? Science. 2009;323: 746–751.
654 doi:10.1126/science.1158997

655 11. Prud'homme B, Gompel N, Carroll SB. Emerging principles of regulatory evolution.
656 Proceedings of the National Academy of Sciences. 2007;104: 8605–8612.
657 doi:10.1073/pnas.0700488104

658 12. Signor SA, Nuzhdin SV. The evolution of gene expression in cis and trans. Trends Genet.
659 2018;34: 532–544. doi:10.1016/j.tig.2018.03.007

660 13. Fraser HB. Gene expression drives local adaptation in humans. Genome Res. 2013;23:
661 1089–1096. doi:10.1101/gr.152710.112

14. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010;327: 302–305. doi:10.1126/science.1182213

15. Stepaniak MD, Square TA, Miller CT. Evolved Bmp6 enhancer alleles drive spatial shifts in gene expression during tooth development in sticklebacks. Genetics. 2021;219: iyab151. doi:10.1093/genetics/iyab151

16. Roberts Kingman GA, Lee D, Jones FC, Desmet D, Bell MA, Kingsley DM. Longer or shorter spines: Reciprocal trait evolution in stickleback via triallelic regulatory changes in Stanniocalcin2a. Proceedings of the National Academy of Sciences. 2021;118: e2100694118. doi:10.1073/pnas.2100694118

17. Howes TR, Summers BR, Kingsley DM. Dorsal spine evolution in threespine sticklebacks via a splicing change in MSX2A. BMC Biology. 2017;15: 115. doi:10.1186/s12915-017-0456-5

18. Verta J-P, Jones FC. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. de Meaux J, Tautz D, editors. eLife. 2019;8: e43785. doi:10.7554/eLife.43785

19. Hart JC, Ellis NA, Eisen MB, Miller CT. Convergent evolution of gene expression in two high-toothed stickleback populations. PLOS Genetics. 2018;14: e1007443. doi:10.1371/journal.pgen.1007443

20. Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. Detection of regulatory variation in mouse genes. Nat Genet. 2002;32: 432–437. doi:10.1038/ng992

21. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430: 85–88. doi:10.1038/nature02698

22. Miller CT, Glazer AM, Summers BR, Blackman BK, Norman AR, Shapiro MD, et al. Modular Skeletal Evolution in Sticklebacks Is Controlled by Additive and Clustered Quantitative Trait Loci. Genetics. 2014;197: 405–420. doi:10.1534/genetics.114.162420

23. Pirinen M, Lappalainen T, Zaitlen NA, GTEx Consortium, Dermitzakis ET, Donnelly P, et al. Assessing allele-specific expression across multiple tissues from RNA-seq read data. Bioinformatics. 2015;31: 2497–2504. doi:10.1093/bioinformatics/btv074

24. Choi CS, Savage DB, Kulkarni A, Yu XX, Liu Z-X, Morino K, et al. Suppression of Diacylglycerol Acyltransferase-2 (DGAT2), but Not DGAT1, with Antisense Oligonucleotides Reverses Diet-induced Hepatic Steatosis and Insulin Resistance*. Journal of Biological Chemistry. 2007;282: 22678–22688. doi:10.1074/jbc.M704213200

25. O'Hare EA, Yang R, Yerges-Armstrong L, Sreenivasan U, McFarland R, Leitch CC, et al. TM6SF2 rs58542926 impacts lipid processing in liver and small intestine. Hepatology. 2017;65: 1526–1542. doi:10.1002/hep.29021

698 26. Lavin PA, McPhail JD. Adaptive Divergence of Trophic Phenotype among Freshwater
699 Populations of the Threespine Stickleback (Gasterosteus aculeatus). Can J Fish Aquat Sci.
700 1986;43: 2455–2463. doi:10.1139/f86-305

701 27. Ellis NA, Glazer AM, Donde NN, Cleves PA, Agoglia RM, Miller CT. Distinct
702 developmental genetic mechanisms underlie convergently evolved tooth gain in
703 sticklebacks. Development. 2015;142: 2442–2451. doi:10.1242/dev.124248

704 28. Stock DW. Zebrafish dentition in comparative context. Journal of Experimental Zoology
705 Part B: Molecular and Developmental Evolution. 2007;308B: 523–549.
706 doi:10.1002/jez.b.21187

707 29. Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT. An ancient
708 gene network is co-opted for teeth on old and new jaws. PLoS Biol. 2009;7: e31.
709 doi:10.1371/journal.pbio.1000031

710 30. Bullard JH, Mostovoy Y, Dudoit S, Brem RB. Polygenic and directional regulatory
711 evolution across pathways in Saccharomyces. Proc Natl Acad Sci U S A. 2010;107: 5058–
712 5063. doi:10.1073/pnas.0912959107

713 31. Fraser HB, Babak T, Tsang J, Zhou Y, Zhang B, Mehrabian M, et al. Systematic Detection
714 of Polygenic cis-Regulatory Evolution. PLOS Genetics. 2011;7: e1002023.
715 doi:10.1371/journal.pgen.1002023

716 32. Orr HA. Testing natural selection vs. genetic drift in phenotypic evolution using
717 quantitative trait locus data. Genetics. 1998;149: 2099–2104.

718 33. Liu F, Chu EY, Watt B, Zhang Y, Gallant NM, Andl T, et al. Wnt/β-catenin signaling
719 directs multiple stages of tooth morphogenesis. Developmental Biology. 2008;313: 210–
720 224. doi:10.1016/j.ydbio.2007.10.016

721 34. Square TA, Sundaram S, Mackey EJ, Miller CT. Distinct tooth regeneration systems deploy
722 a conserved battery of genes. EvoDevo. 2021;12: 4. doi:10.1186/s13227-021-00172-3

723 35. Kraus P, Lufkin T. Dlx homeobox gene control of mammalian limb and craniofacial
724 development. Am J Med Genet A. 2006;140: 1366–1374. doi:10.1002/ajmg.a.31252

725 36. Pang L, Zhang Z, Shen Y, Cheng Z, Gao X, Zhang B, et al. Mutant dlx3b disturbs normal
726 tooth mineralization and bone formation in zebrafish. PeerJ. 2020;8: e8515.
727 doi:10.7717/peerj.8515

728 37. Jeong J, Li X, McEvilly RJ, Rosenfeld MG, Lufkin T, Rubenstein JLR. Dlx genes pattern
729 mammalian jaw primordium by regulating both lower jaw-specific and upper jaw-specific
730 genetic programs. Development. 2008;135: 2905–2916. doi:10.1242/dev.019778

731 38. Erickson PA, Baek J, Hart JC, Cleves PA, Miller CT. Genetic Dissection of a Supergene
732 Implicates Tfap2a in Craniofacial Evolution of Threespine Sticklebacks. Genetics.
733 2018;209: 591–605. doi:10.1534/genetics.118.300760

734 39. Erickson PA, Glazer AM, Killingbeck EE, Agoglia RM, Baek J, Carsanaro SM, et al.
735 Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks.
736 Evolution. 2016;70: 887–902. doi:10.1111/evo.12897

737 40. Hu CK, York RA, Metz HC, Bedford NL, Fraser HB, Hoekstra HE. cis-Regulatory changes
738 in locomotor genes are associated with the evolution of burrowing behavior. Cell Rep.
739 2022;38: 110360. doi:10.1016/j.celrep.2022.110360

740 41. Indjeian VB, Kingman GA, Jones FC, Guenther CA, Grimwood J, Schmutz J, et al.
741 Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins.
742 Cell. 2016;164: 45–56. doi:10.1016/j.cell.2015.12.007

743 42. Johnston J, Ramos-Valdes Y, Stanton L-A, Ladhani S, Beier F, Dimattia GE. Human
744 stanniocalcin-1 or -2 expressed in mice reduces bone size and severely inhibits cranial
745 intramembranous bone growth. Transgenic Res. 2010;19: 1017–1039. doi:10.1007/s11248-
746 010-9376-7

747 43. Gagliardi AD, Kuo EYW, Raulic S, Wagner GF, DiMattia GE. Human stanniocalcin-2
748 exhibits potent growth-suppressive properties in transgenic mice independently of growth
749 hormone and IGFs. Am J Physiol Endocrinol Metab. 2005;288: E92-105.
750 doi:10.1152/ajpendo.00268.2004

751 44. Lin M, Li L, Liu C, Liu H, He F, Yan F, et al. Wnt5a regulates growth, patterning, and
752 odontoblast differentiation of developing mouse tooth. Dev Dyn. 2011;240: 432–440.
753 doi:10.1002/dvdy.22550

754 45. Hosseini-Farahabadi S, Gignac SJ, Danescu A, Fu K, Richman JM. Abnormal WNT5A
755 Signaling Causes Mandibular Hypoplasia in Robinow Syndrome. J Dent Res. 2017;96:
756 1265–1272. doi:10.1177/0022034517716916

757 46. Yamaguchi TP, Bradley A, McMahon AP, Jones S. A Wnt5a pathway underlies outgrowth
758 of multiple structures in the vertebrate embryo. Development. 1999;126: 1211–1223.
759 doi:10.1242/dev.126.6.1211

760 47. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
761 Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.
762 doi:10.1093/bioinformatics/btp352

763 48. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
764 data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170

765 49. Peichel CL, Sullivan ST, Liachko I, White MA. Improvement of the Threespine
766 Stickleback Genome Using a Hi-C-Based Proximity-Guided Assembly. Journal of
767 Heredity. 2017;108: 693–700. doi:10.1093/jhered/esx058

768 50. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
769 universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21.
770 doi:10.1093/bioinformatics/bts635

771 51. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
772 2012;9: 357–359. doi:10.1038/nmeth.1923

773 52. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
774 Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA
775 sequencing data. Genome Res. 2010;20: 1297–1303. doi:10.1101/gr.107524.110

776 53. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for
777 robust molecular quantitative trait locus discovery. Nat Methods. 2015;12: 1061–1063.
778 doi:10.1038/nmeth.3582

779 54. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput
780 sequencing data. Bioinformatics. 2015;31: 166–169. doi:10.1093/bioinformatics/btu638

781 55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
782 RNA-seq data with DESeq2. Genome Biology. 2014;15: 550. doi:10.1186/s13059-014-
783 0550-8

784 56. Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM. A powerful and flexible
785 statistical framework for testing hypotheses of allele-specific gene expression from RNA-
786 seq data. Genome Res. 2011;21: 1728–1737. doi:10.1101/gr.119784.110

787 57. Bradford YM, Van Slyke CE, Ruzicka L, Singer A, Eagle A, Fashena D, et al. Zebrafish
788 information network, the knowledgebase for Danio rerio research. Genetics. 2022;220:
789 iyac016. doi:10.1093/genetics/iyac016

790 58. Artieri CG, Naor A, Turgeman-Grott I, Zhou Y, York R, Gophna U, et al. Cis-regulatory
791 evolution in prokaryotes revealed by interspecific archaeal hybrids. Sci Rep. 2017;7: 3986.
792 doi:10.1038/s41598-017-04278-4

793 59. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
794 features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

795 60. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ, et al. Mouse Genome
796 Database (MGD): Knowledgebase for mouse–human comparative biology. Nucleic Acids
797 Research. 2021;49: D981–D987. doi:10.1093/nar/gkaa1083

798 61. Peichel CL, Marques DA. The genetic and molecular architecture of phenotypic diversity in
799 sticklebacks. Philosophical Transactions of the Royal Society B: Biological Sciences.
800 2017;372: 20150486. doi:10.1098/rstb.2015.0486

801 62. Bello IO, Soini Y, Slootweg PJ, Salo T. Claudins 1, 4, 5, 7 and occludin in ameloblastomas
802 and developing human teeth. J Oral Pathol Med. 2007;36: 48–54. doi:10.1111/j.1600-
803 0714.2006.00497.x

804 63. Aizawa C, Saito K, Ohshima H. Regulation of IGF-I by IGFBP3 and IGFBP5 during
805 odontoblast differentiation in mice. J Oral Biosci. 2019;61: 157–162.
806 doi:10.1016/j.job.2019.07.001

64. Mukherjee A, Rotwein P. Insulin-like growth factor-binding protein-5 inhibits osteoblast differentiation and skeletal growth by blocking insulin-like growth factor actions. Mol Endocrinol. 2008;22: 1238–1250. doi:10.1210/me.2008-0001

65. Xavier GM, Sharpe PT, Cobourne MT. Scube1 is expressed during facial development in the mouse. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 2009;312B: 518–524. doi:10.1002/jez.b.21260

66. Abdullah A, Herdenberg C, Hedman H. Netrin-1 functions as a suppressor of bone morphogenetic protein (BMP) signaling. Sci Rep. 2021;11: 8585. doi:10.1038/s41598-021-87949-7

67. Li Q-M, Li J-L, Feng Z-H, Lin H-C, Xu Q. Effect of histone demethylase KDM5A on the odontogenic differentiation of human dental pulp cells. Bioengineered. 2020;11: 449–462. doi:10.1080/21655979.2020.1743536

68. Xu J, Yu B, Hong C, Wang C-Y. KDM6B epigenetically regulates odontogenic differentiation of dental mesenchymal stem cells. Int J Oral Sci. 2013;5: 200–205. doi:10.1038/ijos.2013.77

69. Zhang F, Xu L, Xu L, Xu Q, Li D, Yang Y, et al. JMJD3 promotes chondrocyte proliferation and hypertrophy during endochondral bone formation in mice. J Mol Cell Biol. 2015;7: 23–34. doi:10.1093/jmcb/mjv003

70. Romanos GE, Asnani KP, Hingorani D, Deshmukh VL. PERIOSTIN: Role in Formation and Maintenance of Dental Tissues. Journal of Cellular Physiology. 2014;229: 1–5. doi:10.1002/jcp.24407

71. Nunia K, Urs AB, Kumar P. Interplay Between MMP-9 and TIMP-2 Regulates Ameloblastoma Behavior and Tooth Morphogenesis. Appl Immunohistochem Mol Morphol. 2016;24: 364–372. doi:10.1097/PAI.0000000000000196

72. Vu TH, Shipley JM, Bergers G, Berger JE, Helms JA, Hanahan D, et al. MMP-9/Gelatinase B Is a Key Regulator of Growth Plate Angiogenesis and Apoptosis of Hypertrophic Chondrocytes. Cell. 1998;93: 411–422. doi:10.1016/S0092-8674(00)81169-1

73. Ni Q, Chen S. Assessing the Effect of Matrix Metalloproteinase-9 on the Growth of Mice Teeth by NMR. J Biol Pharm Chem Res. 2014;1: 192–204.

74. Wang H, Ning T, Song C, Luo X, Xu S, Zhang X, et al. Priming integrin α5 promotes human dental pulp stem cells odontogenic differentiation due to extracellular matrix deposition and amplified extracellular matrix-receptor activity. J Cell Physiol. 2019;234: 12897–12909. doi:10.1002/jcp.27954

75. Niwa T, Yamakoshi Y, Yamazaki H, Karakida T, Chiba R, Hu JC-C, et al. The dynamics of TGF-β in dental pulp, odontoblasts and dentin. Sci Rep. 2018;8: 4450. doi:10.1038/s41598-018-22823-7

843 76. Hayano S, Kurosaka H, Yanagita T, Kalus I, Milz F, Ishihara Y, et al. Roles of Heparan
844 Sulfate Sulfation in Dentinogenesis. J Biol Chem. 2012;287: 12217–12229.
845 doi:10.1074/jbc.M111.332924

846 77. Ratzka A, Kalus I, Moser M, Dierks T, Mundlos S, Vortkamp A. Redundant function of the
847 heparan sulfate 6-O-endosulfatases Sulf1 and Sulf2 during skeletal development.
848 Developmental Dynamics. 2008;237: 339–353. doi:10.1002/dvdy.21423

849 78. Yin Y, Zhou N, Zhang H, Dai X, Lv X, Chen N, et al. Bmi1 regulate tooth and mandible
850 development by inhibiting p16 signal pathway. J Cell Mol Med. 2021;25: 4195–4203.
851 doi:10.1111/jcmm.16468

852 79. Ogoh H, Yamagata K, Nakao T, Sandell LL, Yamamoto A, Yamashita A, et al. Mllt10
853 knockout mouse model reveals critical role of Af10-dependent H3K79 methylation in
854 midfacial development. Sci Rep. 2017;7: 11922. doi:10.1038/s41598-017-11745-5

855