Eda, gene pathways and alternative splicing: the genetic basis of adaptive evolution in threespine stickleback

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Carlos E. Rodríguez Ramírez

from México and Portugal

Supervisor of the doctoral thesis:
Prof. Dr. Catherine L. Peichel
Institute of Ecology and Evolution

Eda, gene pathways and alternative splicing: the genetic basis of adaptive evolution in threespine stickleback

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Carlos E. Rodríguez Ramírez
from México and Portugal

Supervisor of the doctoral thesis:
Prof. Dr. Catherine L. Peichel
Institute of Ecology and Evolution

Accepted by the Faculty of Science.

Bern, **26th of August** 2024

The Dean
Prof. Dr. Jean-Louis Reymond

Contents

For my family

For all your love and support

Even half-way across the world

You are always in my thoughts

Summary

Understanding the genetic basis of adaptation is one of the main goals of evolutionary biology. This entails understanding which and how many loci underlie adaptive phenotypes (the genetic architecture), but also entails understanding how the underlying genes and mutations work and through which mechanisms the create these adaptive phenotypes (the molecular mechanisms). While much progress has been made in the last decades in discovering the genetic architecture of phenotypic variation in the wild, understanding the molecular mechanisms linking genotype and phenotype remains challenging. Understanding this link however is essential to understand why certain genes are involved in adaptive evolution and others not. Answering these questions is essential not only to understand how individuals have adapted to their current environments, but also to one day be able to predict how they might adapt to new environments.

A model that is at the forefront of research in this field is the threespine stickleback (*Gasterosteus aculeatus*), a teleost fish widespread throughout the oceans of the Northern Hemisphere. After the Last Glacial Age, threespine stickleback repeatedly and independently adapted and colonized freshwater habitats across its distribution range. This interesting evolutionary history coupled with rich ecological knowledge and a wealth of genetic tools make threespine stickleback a great system to study the genetic basis of adaptation. One of the most characteristic traits found in freshwater sticklebacks is the repeated loss of defensive lateral plates. This phenotype has been mapped to a large effect gene, *Ectodysplasin A* (*Eda*), which is strongly under selection in freshwater populations. Interestingly, *Eda* is a pleiotropic gene, and besides the lateral plates it also causes changes in the patterning of the sensory lateral line and schooling behaviour. However, despite years of research on *Eda,* not much is known yet about the molecular mechanisms through which *Eda* causes these phenotypic effects.

In Chapter 1 of this thesis, I describe my work in looking for the causative mutation of the phenotypic effects of *Eda,* which remains unknown to this day. In this study my collaborators and I test a candidate 16 bp deletion in the first intron of *Eda* with a CRISPR-Cas9 manipulation experiment. Unfortunately, our results show that our candidate has no effect on the number of lateral plates. We conclude that the causative region of *Eda* needs to be narrowed down further before more candidate mutations are tested. These results also serve as a reminder of how seemingly strong candidate mutations might still turn out to have no effect on the phenotype under study.

In Chapter 2, I describe my work in exploring the downstream molecular effects of the *Eda* haplotype. This is a 16 kb region of the genome that includes *Eda* plus two neighbouring genes

(*Tnfsf13b* and *Garp*), which are linked in wild populations of stickleback. Using RNAseq I found that the *Eda* haplotype causes downstream effects in hundreds of genes in skin. These include genes involved in bone and neuronal development pathways, making them strong candidates to be the mediators of the phenotypic effects of *Eda* on the lateral plates and the lateral line. I also find evidence of an effect of the haplotype in immune genes and an immune organ (head kidney), which could be related to the presence of *Tnfsf13b* and/or *Garp*. Finally, I find that the haplotype effect on downstream genes happens not only through changes in gene expression levels, but also by changes in alternative splicing patterns. Interestingly, these two mechanisms affect mostly non-overlapping sets of genes that also present differences in their average expression and pleiotropy levels.

In Chapter 3 of this thesis, I described the results of a study inspired by our previous results on alternatively spliced genes. In this study I explore the role of alternative splicing in the marine-freshwater divergence of threespine stickleback. By comparing publicly available RNAseq data of marine and freshwater populations, I find over one hundred differentially spliced genes (DSGs) between the ecotypes. I also find that these genes are enriched in regions of the genome associated to phenotypic divergence and in regions of the genome with signatures of divergent selection between ecotypes. Furthermore, I find that different types of splicing seem to contribute differently to the divergence, with splicing of mutually exclusive exons being the most common in our DSGs and having the strongest signatures of genetic divergence. Taken together the results of this chapter suggest that alternative splicing might mediate some of the adaptive divergence between marine and freshwater sticklebacks.

I finalize this thesis with a general discussion, where I discuss how the results of this thesis give us interesting insights into the molecular mechanism mediating adaptive evolution in threespine stickleback and highlight the value and the need to keep studying these questions in our field.

General Introduction

1 The genetics of adaptation

One of the most amazing characteristics of life on Earth is its ability to colonize, adapt and thrive in almost every type of environment. From the bottom of the sea all the way to Antarctica, life can be found almost everywhere on Earth. This is possible thanks to the evolution of amazing adaptations that give organisms the ability to deal with the challenges posed by their environments. Thermophilic archaea and bacteria have evolved hyperthermostable proteins in order to survive the scorching temperatures found near the hydrothermal vents and hot springs they inhabit (Berezovsky and Shakhnovich 2005). Mangrove trees and shrubs have evolved salt secreting glands and other complex adaptations to live in salt water in intertidal regions (Naskar and Palit 2015). Polar fish have evolved anti-freeze proteins that allow them to survive the sub-zero water temperatures found in polar oceans (DeVries and Cheng 2005). Tardigrades have evolved the ability to enter cryptobiosis (a reversible ametabolic state) when faced with unfavourable environmental conditions like dehydration or lack of oxygen (Møbjerg et al. 2011). The examples are endless, each one more amazing than the last, and all of them were the result of the same process: adaptive evolution. This is the process through which, by natural selection, heritable phenotypes that increase the fitness of individuals in a given environment, spread through a population until it becomes locally adapted and its individuals better at surviving in this environment than before (Reeve and Sherman 1993; Kawecki and Ebert 2004; Barrett and Hoekstra 2011). Over millions of years, this process has helped life to evolve the complex and amazing adaptations that allowed it to colonize most environments in our planet.

Along with natural selection a second mechanism is essential to adaptive evolution: heritability. If the phenotypic variation that is favoured by natural selection cannot be inherited, adaptive evolution cannot happen. The rediscovery of Mendel's work on heritability and genes provided us with the missing mechanism to evolutionary theory (Bomblies and Peichel 2022). The development and integration of population genetics models by Fisher, Haldane and Wright, and their integration into evolutionary theory by creating the Modern Synthesis in the 1930s and 1940s, created an extremely powerful explanatory framework of evolution, that explained how adaptive evolution happens through natural selection acting on random genetic variation (Charlesworth et al. 2017). This led to the realization that to fully understand adaptive evolution, one needs to understand its genetic basis. This means understanding the number and genomic location of the genes underlying an adaptive trait (the genetic architecture), as well as their molecular functions and the molecular effects of the causative mutations (the molecular mechanisms) (Bomblies and Peichel 2022; Kitano et al. 2022). Research in this field has uncovered the genetic basis of many phenotypic changes in wild populations,

including changes in colouration in peppered moths (Van't Hof et al. 2016), deer mice (Linnen et al. 2009; Linnen et al. 2013) and cave tetra fish (Protas et al. 2006); changes in skeletal traits in stickleback (Shapiro et al. 2004; Colosimo et al. 2005; Cleves et al. 2014; Erickson et al. 2018) or beak size and shape in Darwin finches (Lamichhaney et al. 2015; Lamichhaney et al. 2016). These and other studies have given us valuable insights into major questions in adaptive evolution, like whether it happens in large or small steps; how repeatable it is and what kind of constrains it is subject to (Kitano et al. 2022). However, these questions are still far from solved, and more research on the molecular mechanisms underlying adaptive evolution is necessary to change that.

1.1 Approaches to studying the genetic basis of adaptation

Research on the genetic basis of adaptive phenotypes can be done through two complementary approaches: "forward genetics" (or "top-down genetics") and "reverse genetics" (or "bottom-up genetics") (Barrett and Hoekstra 2011; Bomblies and Peichel 2022). Forward genetics studies start by identifying a (often putatively) adaptive phenotypic change in a certain environment and then look for the main genomic locus controlling this change, seeking to identify the causative gene and mutation(s). This requires genotyping and phenotyping great number of genetically and phenotypically heterogeneous individuals to find genetic loci whose variation correlates with variation at the phenotype of interest (Barrett and Hoekstra 2011; Bomblies and Peichel 2022). Two popular methods used in forward genetics studies are quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS) (Mackay et al. 2009; Bomblies and Peichel 2022; Kitano et al. 2022). For QTL mapping individuals with opposite morphs of the phenotype of interest are crossed to generate offspring with more variation in the phenotype of interest and a mixed genetic background. These offspring are then investigated for loci associated with variation in the trait of interest. GWAS is used when crossing individuals is not possible and consists in sequencing and phenotyping large numbers of individuals in natural populations and then looking for genotype-phenotype associations across the genome.

The reverse genetics approach starts instead by first identifying regions of the genome with high genetic divergence between populations in different environments (ecotypes). The genetic distances are usually calculated based on some metric of bias of allelic frequencies in single-nucleotide polymorphisms (SNPs) between ecotypes or across an environmental cline. Peaks of divergence in the genome are then used to identify loci, and if possible, genes, putatively under divergent selection. To identify the phenotypes affected by the loci under selection, candidate phenotypes can be tested either by genetic manipulation of the loci under selection, or by association mapping studies testing the correlation between the genotype at the loci under selection and the candidate phenotype(s). This

second approach has the problem of being difficult to disentangle the effect of closely linked loci to the loci of interest, which is why genetic manipulation is considered the gold standard for determining the causality between genotype and phenotype (Barrett and Hoekstra 2011; Bomblies and Peichel 2022).

Both forward and reverse genetics have advantages and disadvantages. The main advantage of forward genetics is that it does not require prior knowledge of the genetic basis of the phenotype in question. However, it tends to be biased towards easily observable phenotypes, like colour or morphology (Bomblies and Peichel 2022). It is also often difficult to have enough resolution to identify the causative gene and mutation(s) due to linkage disequilibrium (Kitano et al. 2022). Reverse genetics on the other side has the advantage of not having a phenotype-observation bias and thus has the potential to uncover a greater range of adaptive phenotypes. However, figuring out what phenotypes are affected by the genetic locus is not straightforward, and even genetic manipulation might not help unless the right phenotypes are measured (e.g. if the locus under study affects physiology or immunity but the phenotyping is only done at the morphological level). Furthermore, reverse genetics studies are limited by other factors like: our ability to differentiate signatures of selection from demographic history; the challenge of detecting weakly selected loci; and our limited knowledge of gene functions, specially outside of model organisms (Barrett and Hoekstra 2011; Tiffin and Ross-Ibarra 2014; Bomblies and Peichel 2022). Because of this, it has been proposed that combining forward and reverse genetics is the best way to get a more comprehensive view of the genetic basis of adaptation (Barrett and Hoekstra 2011; Bomblies and Peichel 2022). By combining both approaches one can identify genes that underlie putative adaptive traits using forward genetics and the ask if they have signatures of natural selection using reverse genetics (Barrett and Hoekstra 2011). However, to date there are few systems where this has been done (Bomblies and Peichel 2022).

It is however also important to highlight that mechanisms like linkage disequilibrium or pleiotropy can confound the genotype-phenotype association (Barrett and Hoekstra 2011). Thus, both approaches require follow-up with genetic manipulation studies (e.g using CRISPR-Cas9) to empirically prove the causality of the genotype and phenotype under study (Barrett and Hoekstra 2011; Bomblies and Peichel 2022). While not always done, this challenging step is essential to ensure that we accurately identify the genes and mutations underlying adaptive phenotypes. This is, of course, essential to have a true understanding of the genetic basis of adaptation. Chapter 1 of this thesis describes one such experiment, where we seek to empirically identify the causative mutation of an important adaptive gene.

1.2 Pleiotropy

The role of pleiotropy in adaptive evolution is one of the questions for which the identification of the exact causative mutations is essential. Pleiotropy is the phenomenon of one gene or mutation affecting multiple traits in the organism (Orr 2000; Otto 2004). Gene pleiotropy is determined by the biological functions, protein-protein interactions, and position in regulatory networks of any given gene (He and Zhang 2006). Meanwhile, the pleiotropy of a mutation is a function of the gene pleiotropy, the location of the mutation within the gene (coding versus non-coding region), and the level of modularity on how information is coded in the gene (e.g regulatory modules or exons that are tissue-specific) (Stern and Orgogozo 2008). Pleiotropy has the potential to hinder or enhance adaptation through antagonistic or synergistic pleiotropy, respectively (Orr 2000; Otto 2004; Leiby and Marx 2014; McGee et al. 2016; Bomblies and Peichel 2022; Kitano et al. 2022). Antagonistic pleiotropy happens when mutations in pleiotropic genes promote an increase in fitness of the organisms through one trait, but cause a decrease in fitness through another trait, resulting in a trade-off (Orr 2000; Otto 2004). Synergistic pleiotropy is the opposite phenomenon, when a mutation on a pleiotropic gene improves simultaneously the fitness benefits of two or more phenotypes affected by that gene changed simultaneously and those changes increase the fitness of the organism (Leiby and Marx 2014; McGee et al. 2016).

Theoretical models predict antagonistic pleiotropy to be more common and to significantly slow the rate of adaptation (Orr 2000; Otto 2004). Since highly pleiotropic genes are more likely to suffer from antagonistic pleiotropy, and more complex organisms are expected to have higher levels of pleiotropy, this has led to the hypothesis of the “cost of complexity”, which proposes that adaptation is slower in more “complex” organisms due to their higher levels of pleiotropy (Orr 2000; Otto 2004). However, empirical data has identified pleiotropic mutations underlying phenotypic variation in *Arabidopsis*, *Drosophila,* humans and stickleback (Mckay et al. 2003; Kamberov et al. 2013; Nagy et al. 2018; Archambeault, Bärtschi, et al. 2020), suggesting that pleiotropy is not always a constraint (Bomblies and Peichel 2022). This has led to the hypothesis that the “cost of complexity” can be overcome if strong synergistic pleiotropy outweighs antagonistic pleiotropy, for example by promoting the co-inheritance of adaptive phenotypes and facilitating large steps towards a fitness optimum (Archambeault, Bärtschi, et al. 2020; Bomblies and Peichel 2022). Recent studies have identified genes with intermediate levels of pleiotropy involved in adaptation in *Arabidopsis thaliana* and threespine stickleback, prompting the authors to suggest that adaptation through genes with intermediate levels of pleiotropy could reduce antagonistic pleiotropy, thereby overcoming the “cost of complexity” through synergistic pleiotropy (Frachon et al. 2017; Rennison and Peichel 2022). Furthermore, Stern and Orgogozo (2008) also point out that theoretical models of the cost of complexity do not make a

clear distinction between pleiotropic genes and pleiotropic mutations and always assume null mutations. However, depending on the type of mutation, a mutation does not need to be as pleiotropic as the gene where it is found (Stern and Orgogozo 2008). This is supported by empirical data suggesting that cis-regulatory mutations (which are predicted to be less pleotropic than coding mutations) are more prevalent in morphological adaptive evolution (Stern and Orgogozo 2008). Taken together, this suggests that natural selection might be able to keep mutations that promote synergistic pleiotropy while removing mutations that cause antagonistic pleiotropy, thus maximizing the benefits of pleiotropy. However, pleiotropy is difficult to measure accurately and more empirical studies are necessary to improve our understanding on the role of pleiotropy in adaptive evolution (Bomblies and Peichel 2022).

1.3 Coding versus cis-regulatory evolution

The location of a mutation in a gene has been proposed to be one major factor determining its level of pleiotropy (Stern and Orgogozo 2008). Mutations in a gene can be classified as coding or cis-regulatory based on their location and effects. Coding mutations affect the region of the gene that code for its final product, usually a protein but can also be a non-coding RNA. In protein-coding genes, due to the nature of the genetic code, 76% of coding mutations change the amino acid sequence (non-synonymous mutations) while the remaining 24% do not (synonymous mutations) (Wilke 2004). Non-synonymous coding mutations can have considerable effect on the function of the protein by changing its amino acid content or even causing an early stop codon that truncates the protein. Synonymous mutations are generally considered to be mostly neutral, though they can affect mRNA stability and translation efficiency (Stern and Orgogozo 2008). Since coding mutations can directly affect the function of the protein, they have traditionally been the target of studies in the genetic basis of adaptation and there are many examples of phenotypic evolution involve coding mutations (Courtier-Orgogozo et al. 2020). These include melanic pigmentation evolution in vertebrates driven by coding sequence evolution at *Mc1r* (Hoekstra 2006); the evolution of an anti-freeze protein in Notothenioid fish from the trypsinogen gene (Chen et al. 1997); the increase in the frequency of charged amino acid residues in the proteins of thermophilic bacteria (Berezovsky and Shakhnovich 2005) or the increase in oxygen affinity of the haemoglobin of vertebrates adapted to high-altitude environments (Storz and Moriyama 2008).

However, evolution can also happen through cis-regulatory mutations. These are mutations located outside of the gene's coding region but still in the same strand of DNA (in *cis*) and which directly regulate its expression. This includes mutations in regions like gene enhancers and promoters, but also

introns and also the 5' and 3' untranslated regions (UTRs). However, it does not include mutations in regulatory proteins like transcription factors (trans-regulation) (Stern and Orgogozo 2008; Wittkopp and Kalay 2012). Since the discovery that human and chimpanzee proteins are mostly similar, changes in gene regulation have been proposed to be important for phenotypic divergence between species (King and Wilson 1975). Evidence has accumulated in support of this hypothesis, specifically for cis-regulatory variation (Stern and Orgogozo 2008; Courtier-Orgogozo et al. 2020). In threespine stickleback, for example, genomic scans between marine and freshwater populations found that most peaks of genetic divergence are located outside of coding regions, suggesting that cis-regulatory evolution plays an important role in the adaptive divergence of this ecotypes (Jones et al. 2012; Roberts Kingman et al. 2021). This is supported by studies looking at the genetic basis of phenotypic divergence in skeletal traits between ecotypes, which have found evidence that cis-regulatory mutations underlie changes in the number and size of lateral plates, in the length of the dorsal and pelvic spines, and in the number of pharyngeal teeth (Colosimo et al. 2005; Chan et al. 2010; Cleves et al. 2014; O'Brown et al. 2015; Indjeian et al. 2016; Howes et al. 2017; Archambeault, Bärtschi, et al. 2020). Examples in other systems include cis-regulatory changes in an enhancer of the ASIP gene underlying cryptic colouration evolution in beach mice (Wooldridge et al. 2022); mutations in an enhancer of the shavenbaby gene driving loss of ventral trichomes in *Drosophila sechellia* (Stern and Frankel 2013); and cis-regulatory mutations in the gene coding for flavonoid 3'-hydroxilase (F3'H) driving transitions from blue to red flower colour in the Mina lineage of morning glories (Des Marais and Rausher 2010). Pleiotropy is thought to be one of the main reasons why cis-regulatory mutations can be advantageous in phenotypic evolution. Cis-regulatory regions are often organized in modules that affect only a subset of the expression pattern of the gene. So even if a gene is pleiotropic, if its pleiotropic effects are regulated by different cis-regulatory elements, then a mutation in one of these regions will be less pleiotropic than the gene itself. This is in contrast to a coding mutation in that same gene, that would affect the protein sequence everywhere the gene is expressed (Stern and Orgogozo 2008) (but see the next section "Alternative splicing in adaptive evolution" for an exception to this).

Nonetheless, as mentioned previously, empirical evidence suggests that both coding and cis-regulatory are used for phenotypic and adaptive evolution (Hoekstra and Coyne 2007; Stern and Orgogozo 2008; Wessinger and Rausher 2012; Courtier-Orgogozo et al. 2020; Bomblies and Peichel 2022; Elkin et al. 2023). The use of one or another type of mutation seems to strongly depend on the molecular mechanisms mediating the formation of each specific phenotype (Hoekstra and Coyne 2007; Stern and Orgogozo 2008; Wessinger and Rausher 2012; Elkin et al. 2023), which again highlights the importance of studies seeking to understand the molecular mechanisms connecting genotype to phenotype. Chapter 2 of this thesis describes one such study in an important adaptive gene in

threespine stickleback. Finally, it is important to highlight that not all mutations fall neatly into this distinction of coding versus cis-regulatory. Many structural mutations like duplications, gene loss or re-arrangements, can affect both the coding and cis-regulatory region of a gene (Stern and Orgogozo 2008).

1.4 Alternative splicing in adaptive evolution

Alternative splicing (AS) mutations are another good example of mutations that can have both coding and regulatory effects on the gene. Gene splicing is the molecular mechanism that regulates the removal of introns and non-constitutive exons from the pre-mRNA to form the mRNA in eukaryotes. This is done through a ribonucleoprotein complex, the spliceosome, which detects splice site motifs at the exon-intron boundaries and branch point motif in the intermediate intron. However, these core splicing motifs are short and ubiquitous, so extra specificity of splice events is given by a suit of surrounding splicing enhancers and silencers which can be located both in introns or exons (Lee and Rio 2015; Wright et al. 2022). Depending on the strength of this core splicing motifs and the surrounding enhancers and silencers, splicing events can be either constitutive or alternative. Alternative splicing is a probabilistic event and its frequency depends, in part, on the strength of the splicing motifs at the exon-intron boundaries. When alternative splicing happens, non-canonical combinations of exons, and sometimes introns are included in the mature mRNA. These alternative mRNA isoforms can lead to the translation of multiple proteins from one gene. There are five types of alternative splicing events: 1) exon skipping (ES), when an alternative exon is not included in the mRNA; 2) mutually exclusive exons (MXE), when only one exon from a group of two or more is included in the mRNA at a time; 3) intron retention, when an intron is kept in the mRNA instead of being spliced out as usual; 4) alternative 3' splice sites (A3SS) and 5) alternative 5' splice sites (A5SS), when part of the 3' or 5' end of an exon is spliced out of the mRNA (Figure 1 from Chapter 3). These different alternative splicing events have the potential to cause changes to the protein function by changing the exons that are incorporated into the mRNA and can thus reveal or hide coding genetic variation in alternative exons. However, more often alternative splicing seems to have a regulatory role. This is achieved by changing the location or stability of the mRNA molecule, or by indirectly down-regulating gene expression levels through nonsense-mediated decay (NDM) of aberrant isoforms (Verta and Jacobs 2022).

Alternative splicing is thus a flexible mechanism that can both mediate changes in the protein structure or regulate a gene's expression. This has led to the question of whether this mechanism can play a role in adaptation. Alternative splicing is widespread in eukaryotic lineages. Between 20% to

80% of genes are alternatively spliced in most eukaryotes and this correlates with complexity as measured by cell-type diversity, with the highest percentages found in vertebrates (Chen et al. 2014). Since alternative splicing can change the sequence of the protein in one isoform, but still maintain it intact in another, it has proposed as a mechanism to reduce antagonistic pleiotropy while still allowing for new proteins to evolve (Singh and Ahi 2022; Verta and Jacobs 2022; Wright et al. 2022). Despite this, the literature on the role of cis-regulatory evolution in phenotypic evolution and adaptation has for the most part focused on changes in gene expression and until recently has mostly ignored alternative splicing. However, recently the evidence for the role of alternative splicing in adaptation has started to accumulate (Bush et al. 2017; Singh and Ahi 2022; Verta and Jacobs 2022; Wright et al. 2022). For example, infrared sensation in vampire bats has been linked to the skipping of one exon in the temperature receptor gene *Trpv1,* which truncates the protein and increases its sensitivity to temperature (Gracheva et al. 2011). Likewise, increased lipid accumulation in some populations of the cavefish *Astyanax mexicanus* has been linked to another exon skipping event that provokes a premature stop codon in the gene *per2,* a suppressor of lipid metabolism (Xiong et al. 2022). Furthermore, transcriptomic studies have identified many differentially spliced genes (DSGs) between diverging species or ecotypes in human lice, cichlids, charr, sunflower and *Arabidopsis* (Tovar-Corona et al. 2015; Singh et al. 2017; Smith et al. 2018; Wang et al. 2019; Jacobs and Elmer 2021).

However, despite the potential of alternative splicing to increase protein diversity while minimizing pleiotropy, relatively few studies have explored whether there is selection on alternative splicing during adaptation to new environments and how big of a role this mechanism plays (but see Manahan and Nachman 2024 and Jacobs and Elmer 2021). Furthermore, to my knowledge no study has looked at whether the five types of alternative splicing events contribute equally to adaptation. These questions will be explored in Chapter 3 of this thesis in the marine-freshwater adaptive divergence of threespine stickleback.

1.5 Outstanding questions on the genetics of adaptation

While notable advances in our understanding of the genetics of adaptation have been made in the last decades, there are many questions that remain to be solved. The causative genes and mutations behind many phenotypic variants in the wild remain unknown. And for those that we know, they are for the most part candidate loci that remain empirically untested. Thus, we still do not know the relative contributions to adaptation of ancestral standing variation versus *de novo* mutations; coding versus cis-regulatory mutations; small versus big effect size mutations; or antagonistic versus synergistic pleiotropy. There is evidence for a role of both components in these questions, however

their relative contribution to adaptation and in what scenario one is favoured over the other are still open questions (Bomblies and Peichel 2022; Kitano et al. 2022). Another topic which remains unknown in many cases are the molecular mechanism and gene pathways through which adaptive genes mediate their phenotypic effects. This is important to understand why some genes instead of others mediate adaptive phenotypic changes and thus to understand why evolution uses (or not) repeatedly the same genes. Understanding these questions is essential to understanding how organisms in nature have adapted to their current environments and how they might respond to future changes in the environment (Barrett and Hoekstra 2011; Bomblies and Peichel 2022; Kitano et al. 2022).

2. Threespine stickleback

2.1 Marine – freshwater divergence

A biological model that is well suited to explore many outstanding questions in the genetics of adaptation is the threespine stickleback (*Gasterosteus aculeatus*). The threespine stickleback is a small teleost fish distributed across the Northern Hemisphere (Bell and Foster 1994). Marine threespine stickleback independently and repeatedly adapted to freshwater environments approximately 15 000 years ago, when many new freshwater habitats became available after the Last Glacial Maximum. (Bell and Foster 1994). Marine and freshwater stickleback differ in several key aspects of ecology, physiology and morphology. Morphological phenotypes in particular are often observed to evolved repeatedly between independent freshwater populations, as it is the case of the loss of lateral plates and spine reduction (Bell and Foster 1994). This remarkable evolutionary history coupled with an amenability to lab rearing and experimental manipulation makes threespine stickleback a very powerful system to study questions related to adaptation, phenotypic evolution and the repeatability of evolution (Peichel & Marques, 2017). This motivated the development of many genetic and genomic tools that allowed researchers to explore the genetics of adaptation in this system (Peichel and Marques 2017). These efforts have led to the identification of several genes underlying phenotypic changes in freshwater sticklebacks, particularly from skeletal morphology. Examples of this are *Eda*'s role in the loss of lateral plates (Colosimo et al. 2005)*; Pitx1*'s effect in pelvic spine reduction (Chan et al. 2010); *Bmp6*'s role in the increase of pharyngeal teeth in benthic freshwater stickleback (Cleves et al. 2014); *GDF6's* role on lateral plate size (Indjeian et al. 2016); *Msx2a*'s role in dorsal spine reduction (Howes et al. 2017); or *Fads2*'s contribution to increased essential fatty acid production in freshwater through gene duplications (Ishikawa et al. 2019).

These and other studies on the genomic and phenotypic divergence of marine and freshwater sticklebacks have given us many insights into the genetic basis of adaptive evolution. These include: 1) evolution can be highly repeatable (Jones et al. 2012; Peichel and Marques 2017); 2) ancestral standing genetic variation is an important source for repeated phenotypic evolution (Jones et al. 2012; Peichel and Marques 2017); 3) Pleiotropic loci can mediate repeated adaptive divergence (Mills et al. 2014; Greenwood et al. 2016; Archambeault, Bärtschi, et al. 2020; Rennison and Peichel 2022); and 4) cis-regulatory evolution can play a major role in adaptive and phenotypic evolution (Colosimo et al. 2005; Chan et al. 2010; Jones et al. 2012; Cleves et al. 2014; Indjeian et al. 2016; Howes et al. 2017; Archambeault, Bärtschi, et al. 2020; Roberts Kingman et al. 2021). Taken together, it is clear that research in threespine stickleback has contributed greatly to improving our understanding of the

genetic basis of adaptation. However, just like in the rest of the field, there are still many questions about adaptation in threespine stickleback that remain open, including in its most emblematic and studied adaptive gene.

2.2 Loss of lateral plates and *Eda*

One of the most consistent phenotypic changes between marine and freshwater stickleback is the loss of bony lateral plates in freshwater populations (Bell and Foster 1994). Lateral plates in threespine stickleback provide protection against fish and bird predation in clear open-water environments (Reimchen 1992; Reimchen 2000; Kitano et al. 2008; Leinonen et al. 2011; Reimchen et al. 2013). Nonetheless, freshwater populations often experience a fast and strong selection for the loss of these lateral plates (Bell et al. 2004; Barrett et al. 2008; Gelmond et al. 2009; LE Rouzic et al. 2011; Bell and Aguirre 2013; Schluter et al. 2021: 202). Several hypotheses have been proposed for this selection against lateral plates in freshwater, like differences in ion concentration levels, temperature, swimming mobility requirements, and differences in predator pressures (Heuts 1947; Bell et al. 1993; Bourgeois et al. 1994; Smith et al. 2014; Smith et al. 2020). While there is some evidence supporting some of these hypotheses, no conclusive evidence has been found to date (Archambeault, Durston, et al. 2020).

However, more is known about the genetic basis of this phenotype. Gene mapping and transgenic studies found that the gene *Ectodysplasin A* (*Eda*) on chromosome IV of the stickleback genome, controls most of the variation in the number of lateral plates between marine and freshwater populations (Colosimo et al. 2004; Colosimo et al. 2005; Archambeault, Bärtschi, et al. 2020). *Eda* codes for a signalling protein that is part of a conserved pathway in vertebrates controlling the development of ectodermal appendages like hair, teeth, feathers and scales (Cui and Schlessinger 2006; Sadier et al. 2014). In humans, mutations in *Eda* are responsible for ahidrotic/hypohidrotic ectodermal dysplasia (HED), a disease that affects hair, teeth, nails and sweat glands, all ectodermal tissues. Furthermore, another gene of the pathway, the *Ectodysplasin A receptor* (*Edar*), has signatures of positive selection in East Asian populations and correlates with increased hair thickness and changes in tooth morphology (Sadier et al. 2014). In mouse, changes in the levels of the signaling protein coded by *Eda* have been linked to changes in tooth morphology (Sadier et al. 2014). In medaka fish, a mutation that leads to aberrant splicing of *Edar* causes complete loss of scales (Kondo et al. 2001) while a null mutation in *Eda* causes deformities in fin rays, scales, teeth and skull shape (Iida et al. 2014). In bearded dragons, a domesticated scaleless morph has been associated to a deletion in a highly conserved motif of the signaling protein coded by *Eda* (Di-Poï and Milinkovitch 2016). Even in

birds, signalling from the *Eda* pathway has been found to be critical for feather formation (Ho et al. 2019). Interestingly, loss of feathers in two flightless bird species, the ostrich and the emu, happen through different mechanisms: in ostrich it involves changes in the expression patterns of *Eda,* but in emu it is completely unrelated to the *Eda* pathway, instead involving changes in the migration of precursor cells necessary for feather formation (Ho et al. 2019). In threespine stickleback, years of research have revealed much about the role of *Eda* in the marine-freshwater adaptive divergence. *Eda* has been demonstrated to be under strong selection in freshwater populations (Barrett et al. 2008) and to controls between 75-94% of the variation in the number of lateral plates (Colosimo et al. 2005; Archambeault, Durston, et al. 2020). *Eda* has also been found to have pleiotropic affects in the patterning and number of the sensory neuromasts in the lateral line (Wark et al. 2012; Mills et al. 2014; Archambeault, Bärtschi, et al. 2020) and in schooling behaviour (Greenwood et al. 2013; Greenwood et al. 2016). Limited coding changes between the marine and freshwater alleles of *Eda* and fine-scale association mapping suggest that the causative mutation for the phenotypic effects of *Eda* is cis-regulatory (Colosimo et al. 2005; Archambeault, Bärtschi, et al. 2020). Furthermore, in natural populations, *Eda* is found in a 16 kb haplotype, i.e a region with fixed polymorphisms between (in this case) ecotypes that are inherited together. This *Eda* haplotype includes two other genes besides *Eda*: *Tnfsf13b* and *Garp*, both of which are predicted to have immune functions (Colosimo et al. 2005; Jones et al. 2012; O'Brown et al. 2015; Archambeault, Bärtschi, et al. 2020).

Nonetheless, there are still several open questions about the molecular mechanisms through which *Eda* operates in the marine-freshwater divergence in threespine stickleback. We still do not know what the causative mutation for the phenotypic effects of *Eda* is. One study found that a SNP on a downstream enhancer of *Eda* decreased its expression, prompting the authors to propose this to be the main causal mutation (O'Brown et al. 2015). However, a subsequent fine-scale association mapping found a 1.4 kb region in the first intron of *Eda* to have the strongest association to the number of lateral plates and neuromast patterning, and this SNP to have the weakest association in the *Eda* haplotype. The study, however, did not have enough resolution to identify the exact mutation (Archambeault, Bärtschi, et al. 2020). Chapter 1 of this thesis tests one candidate mutation within this region. We also do not know yet why *Eda* is linked to *Tnfsf13b and Garp.* While, the consistent maintenance of such a large haplotype suggests selection against recombinants, this has never been tested. There is however some evidence suggesting that the haplotype might have an effect in immunity, which could be related to these two genes (Robertson et al. 2017), however the study did not fully disentangle the effects of the *Eda* haplotype from the genetic background so more studies are necessary. Chapter 2 of this thesis uses an RNAseq experiment on the *Eda* haplotype to explore this among other questions. Finally, we still do not know much about the gene pathways that mediate the phenotypic effects of *Eda*. While

we know that *Eda* responds to *Wnt* signaling in threespine stickleback (O'Brown et al. 2015), its downstream effects in other pathways are completely unknown. Chapter 2 of this thesis also explores this question and identifies genes and pathways that are strong candidates to mediate the phenotypic effects of *Eda*

3. This thesis

This thesis describes my work (with the help of many collaborators) exploring the genetic basis of adaptation in the threespine stickleback. We explore unanswered questions about the molecular mechanisms mediating the marine-freshwater adaptive divergence in this species, with a particular focus on *Eda.*

Chapter 1 describes a genetic manipulation experiment using CRISPR-Cas9 to test a candidate causal mutation for the phenotypic effects of *Eda* in threespine stickleback. This candidate is a 16 bp deletion in the first intron of *Eda* and derived in the freshwater allele. Using CRISPR-Cas9 we delete this region in the marine allele of heterozygous individuals and assess its impact in the number of lateral plates. Unfortunately, our results show that by itself, this 16 bp deletion is not sufficient to affect the number of lateral plates, and we conclude that the causative mutation is either somewhere else, or is the combination of this and another mutation(s).

Chapter 2 describes a RNAseq experiment on the downstream effects of the *Eda* haplotype, which contains *Eda, Tnfsf13b* and *Garp.* We fix the different genotypes of the haplotype in the same marine genomic background so that we can look only at the changes caused by this region in the transcriptomes of skin and head kidney. We find that the *Eda* haplotype affects hundreds of genes, including some involved in bone development, neuronal development and immunity. We also find that this effect is not only through changes in the gene expression levels of those genes, but also in some cases through changes in their patterns of alternative splicing.

Chapter 3 describes a study exploring the role of alternative splicing in the marine-freshwater adaptive divergence of threespine stickleback. For this, we analyze two published gill RNAseq datasets with samples from marine and freshwater populations. We find not only the presence of over one hundred differentially spliced genes (DSGs) between marine and freshwater sticklebacks, but also find evidence that these genes are found more often than expected by chance in regions of the genome under divergent selection and regions of the genome mediating phenotypic divergence between the ecotypes. We also find evidence that different types of alternative splicing contribute differently to the divergence between marine and freshwater stickleback.

I finalize this thesis with a discussion on how these results contribute to our knowledge of the molecular mechanisms mediating adaptive evolution and what I believe would be interesting venues for future research.

References

Archambeault SL, Bärtschi LR, Merminod AD, Peichel CL. 2020. Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback Eda locus. *Evol. Lett.* 4:282–301.

Archambeault SL, Durston DJ, Wan A, El-Sabaawi RW, Matthews B, Peichel CL. 2020. Phosphorus limitation does not drive loss of bony lateral plates in freshwater stickleback (Gasterosteus aculeatus). *Evol. Int. J. Org. Evol.* 74:2088–2104.

Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nat. Rev. Genet.* 12:767–780.

Barrett RDH, Rogers SM, Schluter D. 2008. Natural selection on a major armor gene in threespine stickleback. *Science* 322:255–257.

Bell MA, Aguirre WE. 2013. Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evol. Ecol. Res.* 15:377–411.

Bell MA, Aguirre WE, Buck NJ. 2004. Twelve years of contemporary armor evolution in a threespine stickleback population. *Evol. Int. J. Org. Evol.* 58:814–824.

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

Bell MA, Ortí G, Walker JA, Koenings JP. 1993. Evolution of pelvic reduction in threespine stickleback fish: a test of competing hypotheses. *Evol. Int. J. Org. Evol.* 47:906–914.

Berezovsky IN, Shakhnovich EI. 2005. Physics and evolution of thermophilic adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 102:12742–12747.

Bomblies K, Peichel CL. 2022. Genetics of adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 119:e2122152119.

Bourgeois JF, Blouw DM, Koenings JP, Bell MA. 1994. Multivariate analysis of geographic covariance between phenotypes and environments in the threespine stickleback, *Gasterosteus aculeatus* , from the Cook Inlet area, Alaska. *Can. J. Zool.* 72:1497–1509.

Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO. 2017. Alternative splicing and the evolution of phenotypic novelty. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20150474.

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327:302–305.

Charlesworth D, Barton NH, Charlesworth B. 2017. The sources of adaptive variation. *Proc. R. Soc. B Biol. Sci.* 284:20162864.

Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. 2014. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol. Biol. Evol.* 31:1402–1413.

Chen L, DeVries AL, Cheng CH. 1997. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. U. S. A.* 94:3811–3816.

Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, Miller CT. 2014. Evolved tooth gain in sticklebacks is associated with a *cis* -regulatory allele of *Bmp6*. *Proc. Natl. Acad. Sci.* 111:13912–13917.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307:1928–1933.

Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, Kingsley DM. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol.* 2:e109.

Courtier-Orgogozo V, Arnoult L, Prigent SR, Wiltgen S, Martin A. 2020. Gephebase, a database of genotype–phenotype relationships for natural and domesticated variation in Eukaryotes. *Nucleic Acids Res.* 48:D696–D703.

Cui C-Y, Schlessinger D. 2006. EDA signaling and skin appendage development. *Cell Cycle* 5:2477–2483.

Des Marais DL, Rausher MD. 2010. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in Ipomoea. *Evol. Int. J. Org. Evol.* 64:2044–2054.

DeVries AL, Cheng C -H. C. 2005. Antifreeze proteins and organismal freezing avoidance in polar fishes. *Fish Physiology* 22:155–201.

Di-Poï N, Milinkovitch MC. 2016. The anatomical placode in reptile scale morphogenesis indicates shared ancestry among skin appendages in amniotes. *Sci. Adv.* 2:e1600708.

Elkin J, Martin A, Courtier-Orgogozo V, Santos ME. 2023. Analysis of the genetic loci of pigment pattern evolution in vertebrates. *Biol. Rev.* 98:1250–1277.

Erickson PA, Baek J, Hart JC, Cleves PA, Miller CT. 2018. Genetic dissection of a supergene implicates *Tfap2a* in craniofacial evolution of threespine Sticklebacks. *Genetics* 209:591–605.

Frachon L, Libourel C, Villoutreix R, Carrère S, Glorieux C, Huard-Chauveau C, Navascués M, Gay L, Vitalis R, Baron E, et al. 2017. Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nat. Ecol. Evol.* 1:1551–1561.

Gelmond O, von Hippel FA, Christy MS. 2009. Rapid ecological speciation in three-spined stickleback Gasterosteus aculeatus from Middleton Island, Alaska: the roles of selection and geographic isolation. *J. Fish Biol.* 75:2037–2051.

Gracheva EO, Cordero-Morales JF, González-Carcacía JA, Ingolia NT, Manno C, Aranguren CI, Weissman JS, Julius D. 2011. Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* 476:88–91.

Greenwood AK, Mills MG, Wark AR, Archambeault SL, Peichel CL. 2016. Evolution of schooling behavior in threespine sticklebacks is shaped by the *Eda* gene. *Genetics* 203:677–681.

Greenwood AK, Wark AR, Yoshida K, Peichel CL. 2013. Genetic and neural modularity underlie the evolution of schooling behavior in threespine sticklebacks. *Curr. Biol. CB* 23:1884–1888.

He X, Zhang J. 2006. Toward a molecular understanding of pleiotropy. *Genetics* 173:1885–1891.

Heuts MJ. 1947. Experimental studies on adaptive evolution in *Gasterosteus aculeatus* L. *Evolution* 1:89–102.

Ho WKW, Freem L, Zhao D, Painter KJ, Woolley TE, Gaffney EA, McGrew MJ, Tzika A, Milinkovitch MC, Schneider P, et al. 2019. Feather arrays are patterned by interacting signalling and cell density waves. *PLOS Biol.* 17:e3000132.

Hoekstra HE. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97:222–234.

Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995–1016.

Howes TR, Summers BR, Kingsley DM. 2017. Dorsal spine evolution in threespine sticklebacks via a splicing change in MSX2A. *BMC Biol.* 15:115.

Iida Y, Hibiya K, Inohaya K, Kudo A. 2014. Eda/Edar signaling guides fin ray formation with preceding osteoblast differentiation, as revealed by analyses of the medaka all-fin less mutant *afl*. *Dev. Dyn.* 243:765–777.

Indjeian VB, Kingman GA, Jones FC, Guenther CA, Grimwood J, Schmutz J, Myers RM, Kingsley DM. 2016. Evolving new skeletal traits by cis -regulatory changes in bone morphogenetic proteins. *Cell* 164:45–56.

Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A, et al. 2019. A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* 364:886–889.

Jacobs A, Elmer KR. 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol. Ecol.* 30:4955–4969.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.

Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, Tan L, Yang Y, Li S, Tang K, Chen H, et al. 2013. Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* 152:691–702.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecol. Lett.* 7:1225–1241.

King M-C, Wilson AC. 1975. Evolution at two Levels in humans and chimpanzees: their macromolecules are so alike that regulatory mutations may account for their biological differences. *Science* 188:107–116.

Kitano J, Bolnick DI, Beauchamp DA, Mazur MM, Mori S, Nakano T, Peichel CL. 2008. Reverse evolution of armor plates in the threespine stickleback. *Curr. Biol. CB* 18:769–774.

Kitano J, Ishikawa A, Ravinet M, Courtier-Orgogozo V. 2022. Genetic basis of speciation and adaptation: from loci to causative mutations. *Philos. Trans. R. Soc. B Biol. Sci.* 377:20200503.

Kondo S, Kuwahara Y, Kondo M, Naruse K, Mitani H, Wakamatsu Y, Ozato K, Asakawa S, Shimizu N, Shima A. 2001. The medaka rs-3 locus required for scale development encodes ectodysplasin-A receptor. *Curr. Biol.* 11:1202–1206.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–375.

Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT, Grant BR, Grant PR, Andersson L. 2016. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* 352:470–474.

LE Rouzic A, Østbye K, Klepaker TO, Hansen TF, Bernatchez L, Schluter D, Vøllestad LA. 2011. Strong and consistent natural selection associated with armour reduction in sticklebacks. *Mol. Ecol.* 20:2483–2493.

Lee Y, Rio DC. 2015. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* 84:291–323.

Leiby N, Marx CJ. 2014. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in Escherichia coli. *PLoS Biol.* 12:e1001789.

Leinonen T, Herczeg G, Cano JM, Merilä J. 2011. Predation-imposed selection on threespine stickleback (Gasterosteus aculeatus) morphology: a test of the refuge use hypothesis. *Evol. Int. J. Org. Evol.* 65:2916–2926.

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

Linnen CR, Poh Y-P, Peterson BK, Barrett RDH, Larson JG, Jensen JD, Hoekstra HE. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* 339:1312–1316.

Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10:565–577.

Manahan DN, Nachman MW. 2024. Alternative splicing and environmental adaptation in wild house mice. *Heredity* 132:133–141.

McGee LW, Sackman AM, Morrison AJ, Pierce J, Anisman J, Rokyta DR. 2016. Synergistic pleiotropy overrides the costs of complexity in viral adaptation. *Genetics* 202:285–295.

Mckay JK, Richards JH, Mitchell-Olds T. 2003. Genetics of drought adaptation in *Arabidopsis thaliana* : I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol. Ecol.* 12:1137–1151.

Mills MG, Greenwood AK, Peichel CL. 2014. Pleiotropic effects of a single gene on skeletal development and sensory system patterning in sticklebacks. *EvoDevo* 5:5.

Møbjerg N, Halberg KA, Jørgensen A, Persson D, Bjørn M, Ramløv H, Kristensen RM. 2011. Survival in extreme environments - on the current knowledge of adaptations in tardigrades. *Acta Physiol. Oxf. Engl.* 202:409–420.

Nagy O, Nuez I, Savisaar R, Peluffo AE, Yassin A, Lang M, Stern DL, Matute DR, David JR, Courtier-Orgogozo V. 2018. Correlated evolution of two copulatory organs via a single cis-regulatory nucleotide change. *Curr. Biol. CB* 28:3450-3457.e13.

Naskar S, Palit PK. 2015. Anatomical and physiological adaptations of mangroves. *Wetl. Ecol. Manag.* 23:357–370.

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:e05290.

Orr HA. 2000. Adaptation and the cost of complexity. *Evolution* 54:13–20.

Otto SP. 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc. R. Soc. Lond. B Biol. Sci.* 271:705–714.

Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20150486.

Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* 38:107–111.

Reeve HK, Sherman PW. 1993. Adaptation and the goals of evolutionary research. *Q. Rev. Biol.* 68:1–32.

Reimchen TE. 1992. Injuries on stickleback from attacts by a toothed predator (Oncorhychus) and implications for the evolution of lateral plates. *Evol. Int. J. Org. Evol.* 46:1224–1230.

Reimchen TE. 2000. Predator handling failures of lateral plate morphs in Gasterosteus aculeatus: functional implications for the ancestral plate condition. *Behaviour* 137:1081–1096.

Reimchen TE, Bergstrom* C, Nosil P. 2013. Natural selection and the adaptive radiation of Haida Gwaii stickleback. *Evol. Ecol. Res.* 15:241–269.

Rennison DJ, Peichel CL. 2022. Pleiotropy facilitates parallel adaptation in sticklebacks. *Mol. Ecol.* 31:1476–1486.

Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, et al. 2021. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv.* 7:eabg5285.

Robertson S, Bradley JE, MacColl ADC. 2017. Eda haplotypes in three-spined stickleback are associated with variation in immune gene expression. *Sci. Rep.* 7:42677.

Sadier A, Viriot L, Pantalacci S, Laudet V. 2014. The ectodysplasin pathway: from diseases to adaptations. *Trends Genet. TIG* 30:24–31.

Schluter D, Marchinko KB, Arnegard ME, Zhang H, Brady SD, Jones FC, Bell MA, Kingsley DM. 2021. Fitness maps to a large-effect locus in introduced stickleback populations. *Proc. Natl. Acad. Sci.* 118:e1914889118.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428:717–723.

Singh P, Ahi EP. 2022. The importance of alternative splicing in adaptive evolution. *Mol. Ecol.* 31:1928–1938.

Singh P, Börger C, More H, Sturmbauer C. 2017. The role of alternative splicing and differential gene expression in cichlid adaptive radiation. *Genome Biol. Evol.* 9:2764–2781.

Smith C, Spence R, Barber I, Przybylski M, Wootton RJ. 2014. The role of calcium and predation on plate morph evolution in the three-spined stickleback (Gasterosteus aculeatus). *Ecol. Evol.* 4:3550–3554.

Smith C, Zięba G, Spence R, Klepaker T, Przybylski M. 2020. Three-spined stickleback armour predicted by body size, minimum winter temperature and pH. *J. Zool.* 311:13–22.

Smith CCR, Tittes S, Mendieta JP, Collier-zans E, Rowe HC, Rieseberg LH, Kane NC. 2018. Genetics of alternative splicing evolution during sunflower domestication. *Proc. Natl. Acad. Sci.* 115:6768–6773.

Stern DL, Frankel N. 2013. The structure and evolution of *cis* -regulatory regions: the *shavenbaby* story. *Philos. Trans. R. Soc. B Biol. Sci.* 368:20130028.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62:2155–2177.

Storz JF, Moriyama H. 2008. Mechanisms of hemoglobin adaptation to high altitude hypoxia. *High Alt. Med. Biol.* 9:148–157.

Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* 29:673–680.

Tovar-Corona JM, Castillo-Morales A, Chen L, Olds BP, Clark JM, Reynolds SE, Pittendrigh BR, Feil EJ, Urrutia AO. 2015. Alternative splice in alternative lice. *Mol. Biol. Evol.* 32:2749–2759.

Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534:102–105.

Verta J-P, Jacobs A. 2022. The role of alternative splicing in adaptation and evolution. *Trends Ecol. Evol.* 37:299–308.

Wang X, Yang M, Ren D, Terzaghi W, Deng X, He G. 2019. *Cis* -regulated alternative splicing divergence and its potential contribution to environmental responses in Arabidopsis. *Plant J.* 97:555–570.

Wark AR, Mills MG, Dang L-H, Chan YF, Jones FC, Brady SD, Absher DM, Grimwood J, Schmutz J, Myers RM, et al. 2012. Genetic architecture of variation in the lateral line sensory system of threespine sticklebacks. *G3 Bethesda Md* 2:1047–1056.

Wessinger CA, Rausher MD. 2012. Lessons from flower colour evolution on targets of selection. *J. Exp. Bot.* 63:5741–5749.

Wilke CO. 2004. Molecular clock in neutral protein evolution. *BMC Genet.* 5:25.

Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13:59–69.

Wooldridge TB, Kautt AF, Lassance J-M, McFadden S, Domingues VS, Mallarino R, Hoekstra HE. 2022. An enhancer of *Agouti* contributes to parallel evolution of cryptically colored beach mice. *Proc. Natl. Acad. Sci.* 119:e2202862119.

Wright CJ, Smith CWJ, Jiggins CD. 2022. Alternative splicing as a source of phenotypic diversity. *Nat. Rev. Genet.*

Xiong S, Wang W, Kenzior A, Olsen L, Krishnan J, Persons J, Medley K, Peuß R, Wang Y, Chen S, et al. 2022. Enhanced lipogenesis through Pparγ helps cavefish adapt to food scarcity. *Curr. Biol.* 32:2272-2280.e6.

Chapter 1

A 16 bp deletion in the first intron of *Eda* is not responsible for the loss of lateral plates in threespine stickleback

A 16-bp deletion in the first intron of *Eda* is not responsible for the loss of lateral plates in threespine stickleback

Carlos E. Rodríguez-Ramírez, Sophie L. Archambeault[1], Matthew Josephson[1], Nicole Nesvadba[1], Anna Räss[1], Åsa Lind[1], Valentine C. Paquet[1], Yannick E. Jauslin[1], Melanie Hiltbrunner[1], Verena Saladin[1], Catherine L. Peichel[1]

[1]Division of Evolutionary Ecology, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

This chapter is in preparation for publication

Abstract

Identifying the mutations underlying the phenotypic effects of adaptive genes is essential to answer many fundamental questions in genetics of adaptation. While much progress has been made in identifying the genes underlying phenotypic variation in the wild, identifying the exact causal mutations remains a challenging undertaking. An example of this is the gene *Ectodysplasin A* (*Eda*), which mediates the loss of lateral plates in freshwater populations of the threespine stickleback (*Gasterosteus aculeatus*). Despite being a classic example in genetics of adaptation, the causative mutation mediating the phenotypic effects of *Eda* is still unknown. This is in part because of strong linkage disequilibrium in a 16 kb haplotype around *Eda*. However, a recent fine-mapping association study of a population with historical recombination within this 16 kb haplotype narrowed down the causative mutation to a 1.4 kb region in the first intron of *Eda.* In this study, we identify LP3621, a 16 bp freshwater deletion in this region, as a strong candidate causative mutation and test this hypothesis by deleting it from the marine allele of heterozygous fish using CRISPR-Cas9. Our results show that *LP3621* is not enough by itself to recapitulate the loss of lateral plates in threespine stickleback. This suggests that the causative mutation is either somewhere else or that *LP3621* interacts with another mutation in the *Eda* haplotype to cause the loss of lateral plates in threespine stickleback.

1 Introduction

Understanding the genetic basis of adaptive phenotypic variation in nature is important to tackle long-standing questions in evolutionary biology, like whether evolution happens in small or large steps, how constrained is evolution and how repeatable it is (Bomblies and Peichel 2022; Kitano et al. 2022). Answering these questions is important to better understand how organisms in nature have adapted to their current environments and how they might respond to future environmental changes (Kitano et al. 2022). It also has the benefit of providing fundamental knowledge about the functions of genes (Bomblies and Peichel 2022). Recent advances in genome sequencing and editing technologies have enabled the identification of the genes underlying phenotypic variation in nature (Courtier-Orgogozo et al. 2020; Bomblies and Peichel 2022; Kitano et al. 2022). However, in many cases there is still not empirical data supporting causality of the candidate mutation(s) in the phenotypes being studied. Factors like linkage disequilibrium and the effect size of the locus make it difficult to have enough resolution to identify strong candidate causative mutation(s) to test (Kitano et al. 2022). Even when strong candidate mutations are found, in most species it is not feasible to perform genetic editing experiments to test these candidate mutations due to technical (e.g embryo injection with CRISPR-Cas9 constructs), biological (e.g long generation times or unsuitability for lab rearing), or ethical (e.g genetic manipulation of humans or endangered species) limitations.

A powerful system to study the genetic basis of adaptive phenotypic variation is the threespine stickleback (*Gasterosteus aculeatus*). This small teleost fish is found in marine and freshwater systems across the Northern hemisphere. Marine threespine sticklebacks colonized newly formed freshwater environments approximately 12 000 to 15 000 years ago after the end of the Last Glacial Maximum (Bell and Foster 1994). This caused the independent but repeated evolution of freshwater populations, with many morphological traits evolving in parallel across different freshwater populations (Bell and Foster 1994). This interesting evolutionary history, coupled with its ameanability to lab rearing and experimental manipulation, makes the threespine stickleback a very interesting system to study the genetic basis of adaptation. Thus, many genetic and genomic tools were developed for threespine stickleback in the last decades (Colosimo et al. 2005; Chan et al. 2010; Jones et al. 2012; Cleves et al. 2014; Peichel and Marques 2017; Roberts Kingman et al. 2021) which have turned this small fish into a powerful system to study questions related to adaptation, phenotypic evolution and the repeatability of evolution (Peichel & Marques, 2017).

One of the most consistent phenotypic changes between marine and freshwater sticklebacks is the loss of lateral plates (Bell and Foster 1994). Lateral plates in sticklebacks protect against predation by fish and birds (Reimchen 1992; Reimchen 2000), and there is selection for the completely-plated

morph in clear open-water environments (Kitano et al. 2008; Reimchen et al. 2013). However, in many freshwater populations there is strong and fast selection against lateral plates, resulting in low-plated freshwater morphs (Bell et al. 2004; Barrett et al. 2008; Gelmond et al. 2009; LE Rouzic et al. 2011; Bell and Aguirre 2013; Reimchen et al. 2013; Schluter et al. 2021). While the selective pressure driving this loss of plates in freshwater is still not clear (Archambeault, Durston, et al. 2020), gene mapping and transgenic experiments have identified a gene on chromosome IV, *Ectodysplasin A* (*Eda*), as the main gene underlying this phenotypic change (Colosimo et al. 2004; Colosimo et al. 2005; Archambeault, Bärtschi, et al. 2020). Depending on the population, *Eda* can explain between 75% to 94% of the variation in the number of lateral plates (Colosimo et al. 2004; Kitano et al. 2008; Archambeault, Bärtschi, et al. 2020). *Eda* codes for a signaling protein that is part of a conserved pathway in vertebrates controlling the development of ectodermal appendages like hair, teeth, feathers and scales (Cui and Schlessinger 2006; Sadier et al. 2014). This gene has also been found to have pleiotropic effects on the patterning of the sensory neuromasts in the lateral line (Wark et al. 2012; Mills et al. 2014; Archambeault, Bärtschi et al 2020) and schooling behaviour (Greenwood et al. 2013; Greenwood et al. 2016) in threespine stickleback. Interestingly, in wild populations *Eda* is found in linkage with two neighboring putative immune genes (*Tnfsf13b* and *Garp*), in a 16 kb haplotype with multiple fixed polymorphisms between marine and freshwater populations (Colosimo et al. 2005; Jones et al. 2012; O'Brown et al. 2015; Archambeault, Bärtschi, et al. 2020).

A SNP in an enhancer downstream of *Eda* but still within the 16 kb haplotype was the first candidate causal mutation for the phenotypic effects of *Eda* (O'Brown et al. 2015). The *Eda* freshwater allele is expressed at a lower level than the marine allele both in heterozygous and homozygous individuals (O'Brown et al. 2015; Rodríguez-Ramírez et al. 2023). This SNP, named NAKA after the population where it was identified, was found to affect *Eda* expression early in development, before the formation of the plates, which led to the hypothesis that it could be the main causative mutation for the phenotypic effects of *Eda* (O'Brown et al. 2015). However, this hypothesis was not supported by a subsequent fine mapping study of the number of lateral plates and neuromast patterning across the 16 kb haplotype. In this study, a freshwater population with historical recombination between the marine and freshwater alleles of *Eda,* together with a set of over ten markers across the 16 kb haplotype, were used to narrow down the region within the 16 kb haplotype with the strongest association to the phenotypic effects of *Eda.* Strikingly, the authors found the NAKA SNP to have the lowest correlation of all markers within the haplotype to these phenotypes. Instead, the strongest association with the number of lateral plates and neuromast patterning was to a 1.4 kb region in the first intron of *Eda* (Figure 1), which contains multiple polymorphism between the marine and freshwater allele (Archambeault, Bärtschi, et al. 2020).

In this study we identify and test a candidate causative mutation for the loss of lateral plates within the 1.4 kb region in the first intron of *Eda.* This mutation *LP3621*, is a 16 bp deletion in the freshwater allele (hereafter L allele, for "low-plated") of *Eda* and overlaps a putative binding site for transcription factor TFAP4, which in mammals regulates important cellular processes like proliferation, differentiation, senescence and apoptosis (Wong et al. 2021) and whose family member TFAP2a affects stickleback craniofacial development (Erickson et al., 2018). We tested whether *LP3621* was the causative mutation for the loss of lateral plates by deleting it from the marine allele (hereafter C allele, for "completely-plated") of *Eda* heterozygous individuals using CRISPR-Cas9. In our study population (Puget Sound), heterozygous individuals at *Eda* are usually completely-plated (Archambeault, Bärtschi, et al. 2020). Thus, if *LP3621* is the main causal mutation, deleting it from the C allele should make it functionally equivalent to the L allele and induce a low-plated phenotype in *Eda* heterozygous individuals.

2 Methods

2.1 Ethics statement

Fish husbandry and all experimental procedures were approved by the Veterinary Service of the Department of Agriculture and Nature of the Canton of Bern (VTHa-Nr. BE4/16, G BE1/18, and BE82/17).

2.2 Identification of candidate causal mutation within the 1.4 kb locus

Following the identification of the 1.4 kb region by Archambault, Bärtschi et al 2020, a former member of the lab (Sophie Archambeault) scanned this region for candidate causal mutations. Polymorphisms in this region were evaluated based on three criteria: 1) whether they were differentiated between most marine and freshwater haplotypes; 2) whether they were derived in freshwater (this was evaluated using sequences of closely related species); and 3) whether they overlapped putative transcription factor binding domains. Using these criteria, one of the markers used in Archambeault et al., 2020 emerged as the strongest candidate for the causative mutation within the 1.4 kb region: *LP3621*, a 16bp deletion in the freshwater alelle of *Eda*. This locus: 1) is the marker most consistently associated with low-plated morphology in Archambeault et al., 2020; 2) is predicted to be derived in freshwater; and 3) overlaps a putative binding domain for transcription factor TFAP4.

2.3 CRISPR-Cas9 experiment and fish husbandry

A CRISPR-Cas9 transgenic experiment was designed to delete *LP3621* in the C allele of *Eda* heterozygous fish to test whether this was sufficient to cause a low-plated phenotype (Figure 3). In the Puget Sound population used for this study, heterozygous fish are completely-plated (Archambeault, Bärtschi, et al. 2020). Thus, if *LP3621* is the main causal mutation behind lateral plate loss in *Eda*, its deletion in the C allele of a heterozygous individual should cause a low-plated phenotype.

Lab fish descending from wild marine individuals caught either at Puget Sound, WA, USA or from the Pacific Ocean were genotyped at *Stn382*, a marker within the 16 kb haplotype that differentiates most marine and freshwater haplotypes (Figure 1 and Table S1) (Colosimo et al. 2005; Archambeault, Bärtschi, et al. 2020). We used *Stn382* to determine whether the individuals carried two marine alleles (CC), two freshwater alleles (LL), or were heterozygous (CL) for the *Eda* 16kb haplotype. A former

member of the lab (Matthew Josephson) made eleven crosses of *Eda* CC fish with *Eda* CL fish and injected the clutches of fertilized eggs with CRISPR-Cas9 protein and one of four combinations of sgRNAs targeting *LP3621* (Figure 2, Table S2). However, even when the editing was successful, the injected fish (hereafter F0 generation) were genetically mosaic, with a mix of edited and non-edited cells in their bodies. To identify mosaic fish with edited cells, we isolated DNA from fin clips of F0 individuals, amplified a 150 bp region around *LP3621* (Figure 2, Table S1), and sent the samples for Sanger sequencing at Mycrosynth AG (Balgach, Switzerland). F0 fish with evidence of editing at *LP3621* were crossed with unedited *Eda* LL fish to try to obtain fully edited F1 offspring and to evaluate their lateral plate morph (Figure 3). We prioritized the use of *Stn382* CL F0 fish, but in two cases *Stn382* CC F0's were used.

Mosaic F0 and their F1 offspring were all kept in 100L tanks at approximately 15 ^{o}C in 3.5 ppt salinity. Young fish were fed a diet of brine shrimp nauplii twice during workdays and once during weekends, and adult fish were fed brine shrimp in the mornings and frozen *Mysis* shrimp in the afternoons three times per week. Individuals were kept under a light cycle of 11 hours of daylight (3450 lumen) and 11h of moonlight (600 lumen) with 1h of sunset and sunrise in between.

2.4 Phenotyping of F1 offspring

We assessed the plate morph of all (n = 354) F1 individuals by staining the live fish with 0.2% calcein, a temporary non-toxic fluorescent stain for external bony structures, following Mills et al. (2014). Stained fish where then examined using fluorescence microscopy to identify their plate morph. Fish were classified as either completely-plated, partially-plated or low-plated. Fish with plate abnormalities, usually gaps in the lateral plates, where recorded whenever observed. F1 individuals were later euthanized with a lethal dose of 0.2% MS-222 (tricaine methane sulfonate). We removed and stored the right pectoral fin in 99% ethanol and then stained a subset of the fish (n = 237) with Alizarin Red, which is used to permanently stain bony structures in threespine sticklebacks, following Peichel et al. (2001). The stained fish were then stored in 37% isopropanol. We measured the standard length and counted the number of lateral plates on the left and right side of all individuals stained with Alizarin Red.

2.5 Genotyping of F1 offspring

We genotyped all F1 individuals (n = 335) at the *Stn382* marker and the *LP3621* indel through PCR amplification and gel electrophoresis (Table S1). Additionally, for the 237 individuals stained with

Alizarin Red we genotyped *LRR,* a sex marker in threespine stickleback (Table S1). For the *LP3621* genotyping, we also added some uninjected *Eda* CC, CL and LL individuals to use as reference. Any F1 individuals that revealed unusual bands after the gel electrophoresis were labelled as putatively edited. For a subset of the putatively edited individuals, we cloned the DNA fragments resulting from PCR amplification with *LP3621* primers into bacterial vectors and sent them for Sanger sequencing. Since the bacterial vectors only incorporate one strand of DNA this allowed us to sequence the C and L alleles separately and look for editing in the C allele, while using the L allele as a reference and a control.

3 Results

3.1 Editing at *LP3621* did not have an effect on plate morph

On 309 injected F0 fish, we looked for fish that had evidence of being mosaic for edits at *LP3621* (Figure 3). Besides having a mosaic distribution of unedited and edited cells, these fish potentially had different edits in different parts of the body. Thus, to obtain fully edited individuals with only one edit, we crossed eight F0 fish (six CL and two CC at *Stn382*) with evidence of editing with uninjected *Eda* LL fish and obtained 335 F1 offspring (Table S3a). Of these, 45 fish had evidence of editing in their genotype at *LP3621*, i.e bands with unusual fragment sizes (Table S3, Figure 4). Of these edited individuals, 40 had a CL genotype at *Stn382*, while interestingly 5 of them were LL, suggesting editing at the L allele of *LP3621* in the heterozygous parent. Of these 45 edited genotypes, 20 were deletions and 25 were insertions (Table S3). Strikingly, none of these individuals showed a discrepancy between *Stn382* genotype and the expected lateral plate phenotype (Figure 5, Table S3), suggesting either failure of the editing to fully delete *LP3621* in the C allele or a lack of effect of *LP3621*. Furthermore, while we found some plate abnormalities (e.g gaps in the lateral plates) in unedited fish, we found none in the fish with evidence of editing at *LP3621* (Figure S1, Table S3).

3.2 *LP3621* deletion in the marine C allele is not enough to cause a low-plated phenotype

While these results suggest *LP3621* has no effect on plate morphology, PCR genotyping is not enough to discern between targeted edits at *LP3621* and overlapping or slightly off-target edits within a 150 bp neighbourhood of *LP3621* (Figure 2). Thus, we sequenced 19 heterozygous edited F1 individuals to characterize their genotypes at *LP3621*. We found seven different insertions (Supplementary Figure S2), two of which were within *LP3621*. We also found five different deletions, of which two fully deleted *LP3621*, one a 38 bp deletion and another a 92 bp deletion (Figure 6). While we only sequenced one individual for each of these two *LP3621* complete deletions, based on the PCR genotypes, there are a total of four F1 individuals with the 38 bp deletion and 11 F1 individuals with the 92 bp deletion (Supplementary Figure S3). All 15 of these individuals had a completely-plated phenotype, showing that deletion of *LP3621* is not sufficient to induce a low-plated phenotype in threespine stickleback.

3.3 *LP3621* does not affect the number of lateral plates

While *LP3621* is not sufficient to cause the low-plated phenotype in threespine stickleback, it still possible that it has a minor effect in the number of lateral plates. To test whether this was the case, we did a second round of phenotyping and genotyping on a subset of the F1 fish (237 individuals, including 40 with edits)(Table S3b) but this time counting the number of lateral plates. However, consistent with our previous results, we found no effect of editing at *LP3621* in the number of lateral plates (linear model: plates_total ~ *LP3621*_edit*edited_parent+size; Stn382 LL fish: estimate = -1.07, standard error = 0.95, p-value = 0.26; Stn382 CL fish: estimate = -1.36, standard error = 1.24, p-value = 0.27) (Supplementary Figure S4). This was still true even in the 15 individuals with full deletions of *LP3621* (linear model: plates_total ~ full_knockout*edited_parent+size; estimate = -0.71, standard error = 1.58, p-value = 0.65) (Figure 7).

4 Discussion

Identifying the genetic basis of adaptation is a challenging task but essential if we want to have a complete understanding of adaptive evolution. In this study we tested a candidate causal mutation for the effect of *Eda* on the lateral plates of threespine stickleback. This mutation (*LP3621*) is a derived 16bp deletion in the freshwater allele of *Eda*, which is found within the 1.4 kb intronic region of *Eda* that had the strongest correlation to the number of lateral plates in an association mapping study conducted in a polymorphic freshwater population (Archambeault, Bärtschi, et al. 2020). *LP3621* overlaps a putative binding domain for transcription factor TFAP4, which controls cellular proliferation and differentiation in mammals (Wong et al., 2021). A related gene *Tfap2a* mediates differences in craniofacial development between marine and freshwater sticklebacks (Erickson et al., 2018). Together with the fact that the L allele of *Eda* is known to be down-regulated in skin compared to the C allele (O'Brown et al. 2015; Rodríguez-Ramírez et al. 2023), this made *LP3621* a strong candidate to be the main mutation driving the effect of *Eda* in the lateral plates. In this study we tested this hypothesis by deleting *LP3621* in the C allele of heterozygous individuals, which in the population that we worked with usually are completely plated. We found that *LP3621* is not sufficient to recreate the low-plated morph in heterozygous individuals, and thus is not the causal mutation for the effects of *Eda* in this phenotype. Likewise, we found no evidence that *LP3621* has even a minor effect in the number of lateral plates.

LP3621 is not sufficient to cause the loss of lateral plates in threespine stickleback

Our results show that *LP3621* is not sufficient to induce a change in the number of lateral plates of threespine stickleback. Using CRISPR-Cas9 we created 45 edited stickleback with deletions or insertions at or in the close proximity of *LP3621*. None of the transgenic sticklebacks had a plate phenotype different than expected based on their *Eda* genotype, including the two individuals that we confirmed through Sanger sequencing to be deletions of *LP3621*, one through a 38 bp deletion and another through a 92 bp deletion. Furthermore, these two individuals had a total of 13 siblings that had similar deletion bands in their genotypes and thus also likely had successful deletions of *LP3621*.

While these results show that *LP3621* is not sufficient to cause the low-plate morph by itself, they do not exclude that it might have a minor effect on the number of lateral plates by itself. Archambeault et al., 2020 found evidence that even outside of the 1.4 kb region, other loci within the 16 kb *Eda* haplotype can have a minor effect on the number of lateral plates. Thus, we tested whether *LP3621*

has a minor effect on the number of lateral plates by conducting a second round of genotyping and phenotyping in a subset of the initial F1 individuals we analysed. However, we found no evidence that editing at *LP3621* was correlated with any change in the number of lateral plates or with abnormal plate phenotypes (Figure 7 and Supplementary Figure S1). Taken together, our results suggest that while *LP3621* is found within the 1.4 kb region with the strongest association to variation with the number of plates within the *Eda* haplotype, it is either just linked with the causative mutation(s) or it requires interactions with other mutations to manifest its effects on the lateral plates (discussed below).

It is also important to highlight that we only tested the effect of *LP3621* on the number of lateral plates. However, *Eda* also affects the patterning and number of neuromasts in the lateral line. While the effects of *Eda* in these two phenotypes are thought to be closely linked and potentially the result of developmental constraints (Mills et al. 2014; Archambeault, Bärtschi, et al. 2020), we cannot discard the possibility that the causative mutations for the effects of *Eda* in these two phenotypes are different and that *LP3621* could be affecting the patterning of the lateral line. Future studies in our samples assessing the pattern and number of neuromast in the lateral line will be required to test this possibility.

The causative mutation for the loss of lateral plates might be outside of the 1.4 kb locus

Our results beg the question: if *LP3621* is not the causative mutation, then what is? There are 18 other polymorphisms (5 indels and 13 SNPs) besides *LP3621* within the 1.4 kb region that are differentiated between marine and freshwater sticklebacks. Of those, 12 are predicted to affect binding sites of transcription factors. Half of these (6) are predicted to be derived in freshwater, and of these, only three are moderately conserved in closely related species (analyses by Sophie Archambeault) (Supplementary Table S4). These three markers could potentially be the causative mutation in *Eda*. Two of them are close SNPs at positions 3749 (SNP3749) and 3757 (SNP3757) of the haplotype and together form a putative binding site for the transcription factor RHOXF11 (Rhox homeobox family member 11) in the marine allele (Supplementary Table S4). Homeobox family members tend to be involved in patterning during development (Gehring and Hiromi 1986). However, the Rhox family in mammals regulates germ cell development in the gonads (Maclean et al. 2005). The other marker is a set of two SNPs and a 1 bp insertion in freshwater, starting at position 3940 (SNP3940) (Supplementary Table S4). This SNP affects a putative binding domain for ESR2 (Estrogen Receptor 2). Estrogen is a major steroid hormone in all vertebrates (Amenyogbe et al. 2020), and while it is mostly

known for its effect in reproductive biology, it also affects immune responses and bone homeostasis in mammals and teleosts (Pinto et al. 2014; Burgos-Aceves et al. 2016), making the SNP3940 also an interesting candidate locus to mediate the effects of *Eda* on the lateral plates.

However, it is also possible that the causative mutation is not within the 1.4 kb locus. The previous association mapping first identified a 2.4 kb region within the 1st intron of *Eda* between *LP3621* and *Cnv770*, a 107 bp insertion in freshwater, as the region with the strongest association to the number of lateral plates and neuromast patterning. This was narrowed down to the 1.4 kb region based solely on a single low-plated recombinant individual that was LL at *LP3621* and CL at *Cnv770* (Archambeault, Bärtschi, et al. 2020). However, Archambeault et al,. also found six individuals that were low-plated despite having homozygous marine genotypes for both *LP3621* and *Cnv770*, hinting at the existence of other unliked modifier loci of the number of lateral plates beyond *Eda*. This is consistent with the fact that *Eda* explains only around 75% of the variation in the number of lateral plates in some populations, including the Lake Washington population used for association mapping to identify the region containing the causative mutation (Colosimo et al. 2004; Colosimo et al. 2005; Archambeault, Bärtschi, et al. 2020). In fact, there is also evidence for an effect of another gene of the *Eda* pathway, the *Ectodysplasin A receptor* (*Edar*), on the number of lateral plates (Knecht et al. 2007; Laurentino et al. 2022). Thus, it is possible that the one recombinant fish used to identify the 1.4 kb locus could have carried these unlinked modifiers of the number of lateral plates. If that is the case, it is possible that the causative mutation(s) of *Eda's* effect on the lateral plates might outside of the 1.4 kb region, within the 2.4 kb region first identified by Archambeault et al. This is supported by the fact that the LOD score in Archambeault et al., 2020 peaked at *Cnv770*, outside of the 1.4 kb region, suggesting that this is where the association with the number of lateral plates and the neuromast patterning was strongest within the *Eda* haplotype. This pattern however could also happen if completely-plated fish are recombinant between *Cnv770* and *LP3621* (e.g. individuals CC for *Cnv770* but CL for *LP3621*). Nonetheless, in the light of the results of this study, it is possible that the LOD peak at *Cnv770* might be true and that the one recombinant fish used to identify the 1.4 kb locus carried unlinked modifiers of the number of lateral plates. This means that *Cnv770* and all the polymorphisms within the 2.4 kb region should be considered as potential candidates to be the causative mutation mediating the phenotypic effects of *Eda*. *Cnv770* for example, is a 107 bp insertion in the freshwater allele. While no analysis of the transcription factor binding domains has been done yet for *Cnv770*, given its size it is not unlikely that some cis-regulatory elements might exist within it. However, before testing more individual mutations, future studies should first empirically test the causality of the 1.4 and 2.4 kb regions (e.g by swapping the marine or freshwater alleles by its counterpart in heterozygous fish) and looking at whether this is sufficient to change the lateral plate phenotype. This is important for two

reasons: 1) to confirm the causality of the 2.4 kb region on the phenotypic effects of *Eda*; and 2) to narrow down the future search for candidate causal mutations by determining whether the nested 1.4 kb region is truly causal, or if the causal region is in the other half of the 2.4 kb region.

5 Conclusion

Identifying the causal mutations underlying the phenotypic effects of genes continues to be a challenging task despite recent advances in sequencing and gene editing technologies. However, this is essential to answer many fundamental questions in the field, like the source of adaptive genetic variation, the role of pleiotropy in adaptation and how repeatable adaptive evolution is. Many studies use genotype-phenotype association analyses to identify promising candidate causal mutations, but then fail to test this connection empirically. The results of our study highlight the danger of such approach, by showing how even a promising candidate might turn out to not have any effect on the phenotype of interest. Thus, while challenging, we encourage researchers to empirically test their candidate adaptive mutations whenever possible. Only with accurate knowledge of the genetic basis of adaptation will we be able to answer the major questions in our field.

6 Acknowledgements

The authors would like to thank all Peichel Lab members for many fruitful discussions that greatly contributed to improve this study.

Supplementary Tables

Supplementary Tables for this chapter can be found in the following OneDrive Link:

https://unibe365-my.sharepoint.com/:f:/r/personal/carlos_rodriguezramirez_unibe_ch/Documents/DoctoralThesis_RodriguezRamirez_SupTables?csf=1&web=1&e=FOhHEN

Figures

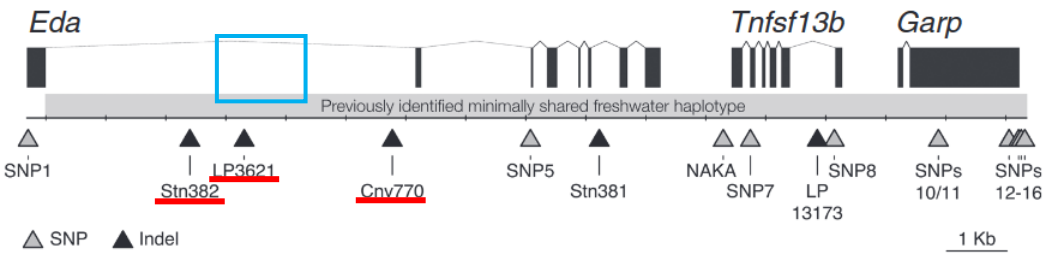


Figure 1 – Representation of the 16 kb *Eda* haplotype with the gene structure of the three genes that it contains: *Eda, Tnfsf13b* and *Garp*. Triangles represent the location of the markers used by Archambeault et al., 2020 to differentiate between the marine (C) and freshwater (L) alleles for fine mapping of the causative region of the phenotypic effects of *Eda* within the haplotype. Highlighted in red are two markers used in this study: *Stn382*, which we used to identify the marine or freshwater allele of *Eda,* and *LP3621,* our candidate causal mutation. Also highlighted in red is *Cnv770*, which together with *LP3621* makes up the boundaries of the 2.4 kb region with the strongest association to the number of lateral plates and patterning of the lateral line in Archambeault et al., 2020. Finally, the region highlighted by a blue box is the 1.4 kb region identified the causal for the phenotypic effects of *Eda* in the lateral plates by Archambeault et al., 2020. Figure is adapted from Archambeault, Bärtschi et al., 2020.

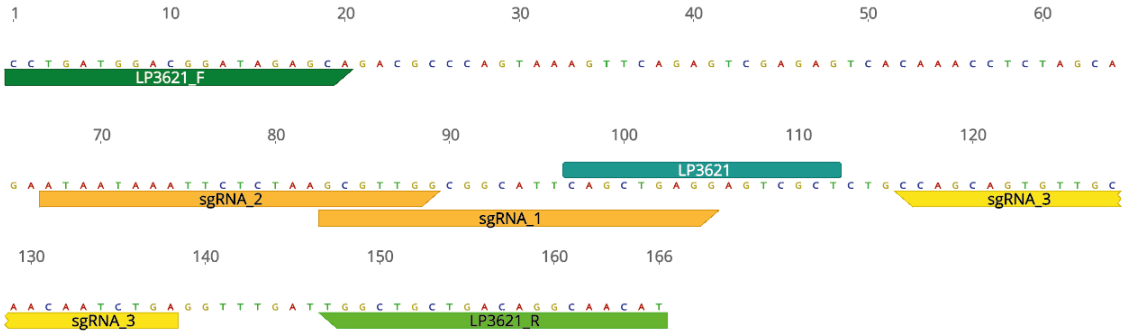


Figure 2 – C allele of *LP3621* as amplified by our PCR primers. In green are the binding sequences for the forward and reverse primers used to amplify and genotype *LP3621*. Cyan annotation starting at position 97 and finishing at position 112 highlights the 16 bp that make up *LP3621*. These 16 bp are deleted in the L allele. In yellow are indicated the target loci of the 3 sgRNAs that were used to guide the CRISPR-Cas9 protein.

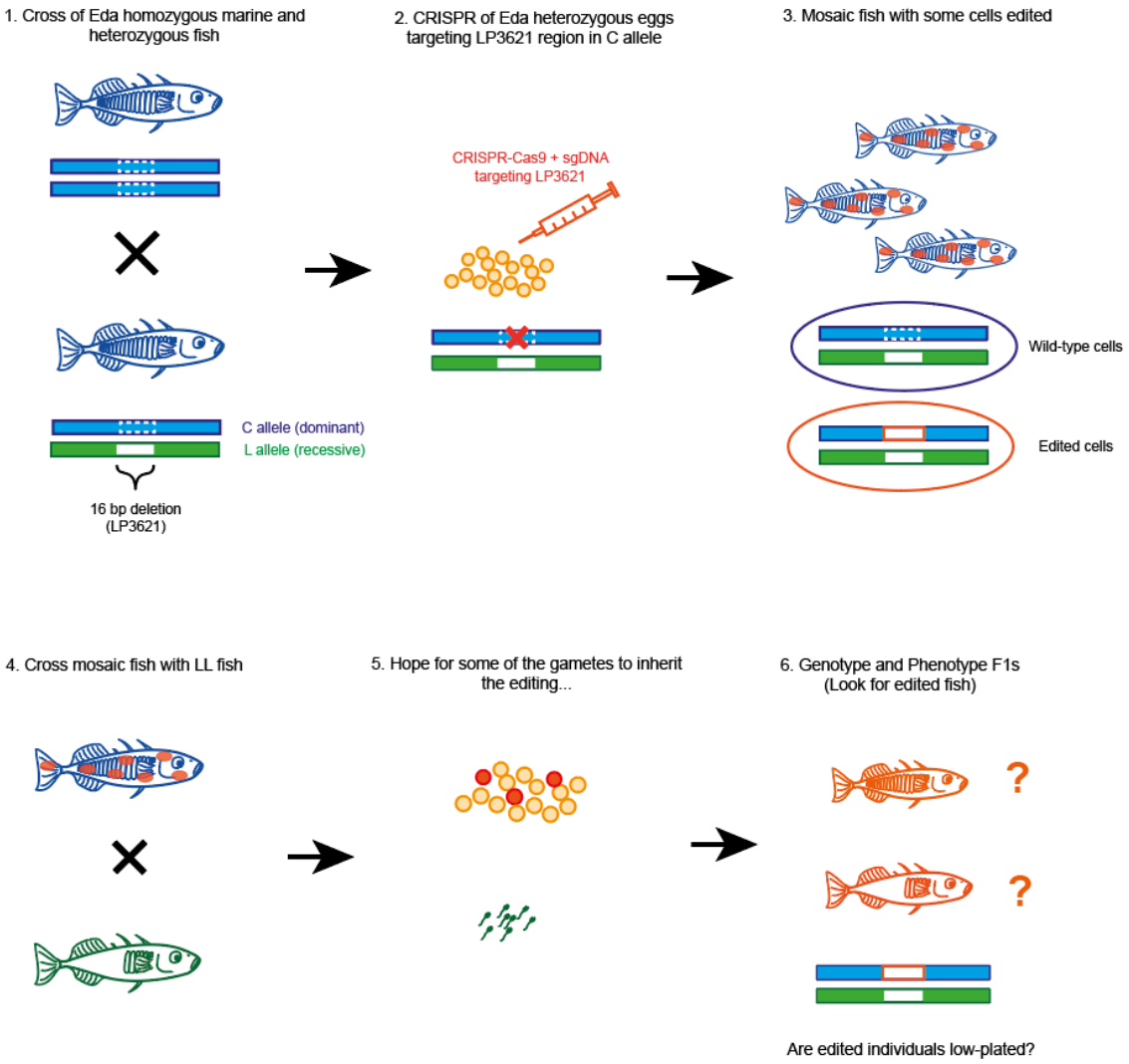
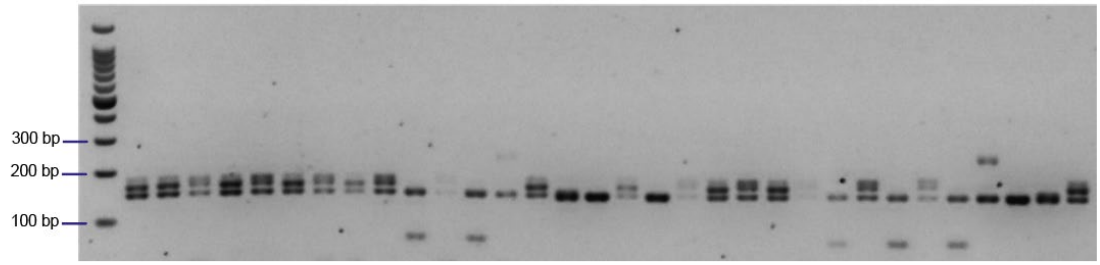


Figure 3 – Experimental design of the CRISPR-Cas9 experiment to delete *LP3621* from the marine allele of *Eda* heterozygous fish. The phenotypes of the resulting individuals reveal whether *LP3621* is sufficient to cause the loss of lateral plates. Note that while we prioritized heterozygous (CL) individuals, in a few cases, fish in steps 2 and 3 were homozygous marine (CC) (Supplementary Table S1).

a)



b)

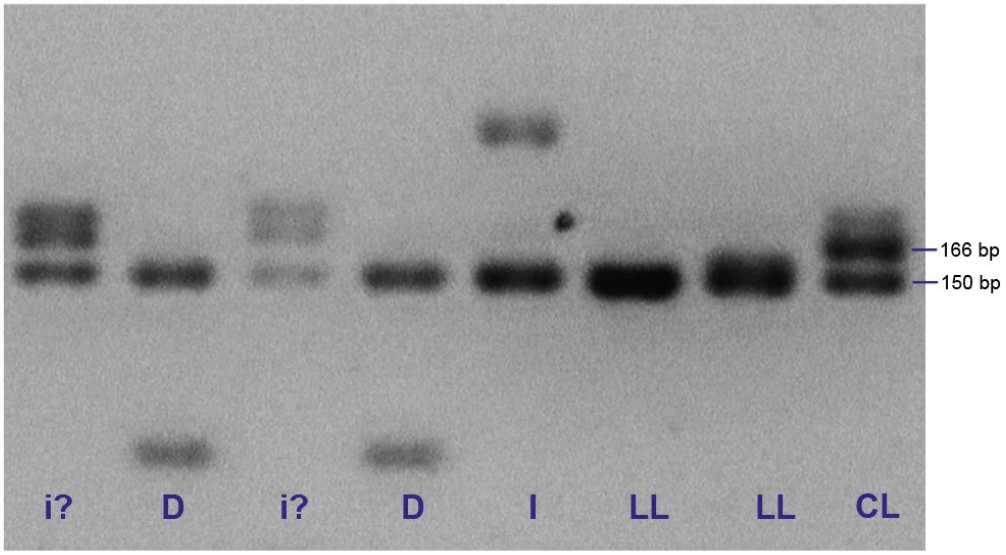


Figure 4 – A) Example of genotyping of *LP3621* in two F1 crosses, one with CL x LL parents and another with CC x LL parents. Edited individuals with large deletions and insertions are visible in these crosses. The 100 bp ladder is in the far-left lane for reference. LL individuals have a single 150 bp band. CL Individuals have two bands: the L band at 150bp and the C band at 166bp (the third top band in these individuals is probably a heterodimer of the two alleles). All individuals with other bands are edited at *LP3621* or in its close proximity. B) A zoom-in on the far-right side of the gel in a), with good examples of edits in the C allele. D – deletion; i? – potentially a small insertion; I – insertion; CL – CL individual; LL – LL individuals.

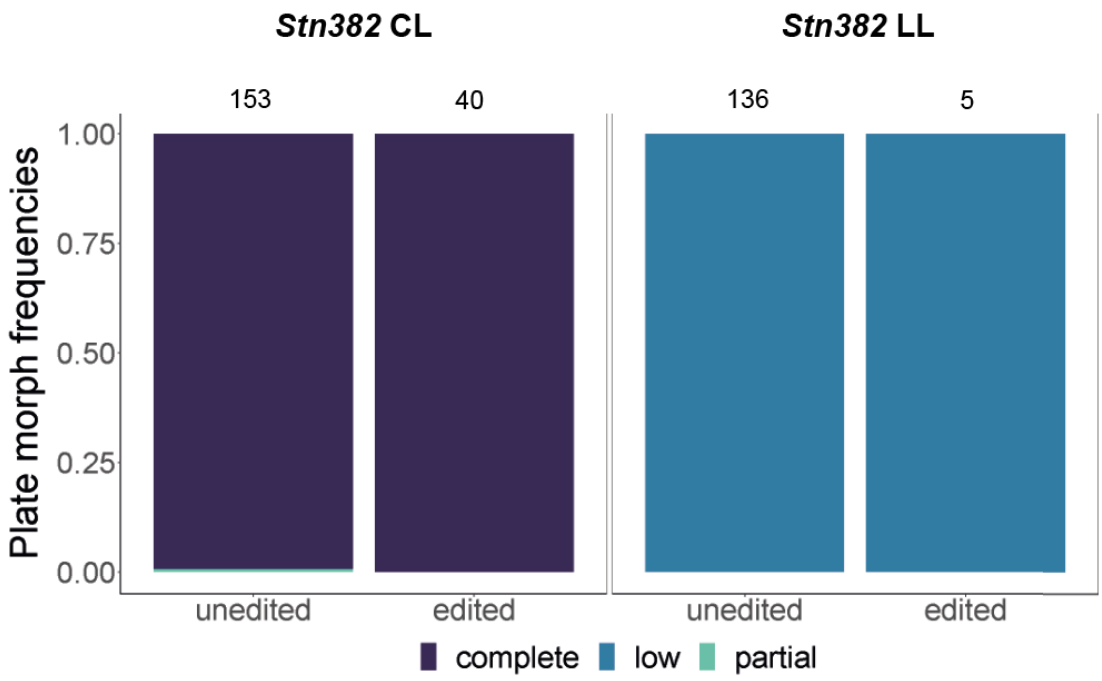


Figure 5 – Plate morph frequencies in F1 fish with unedited and edited genotypes at *LP3621*. Figure is separated by *Eda* (*Stn382*) CL fish (left panel) and *Eda* (*Stn382*) LL fish (right panel). There is one partial individual amongst the unedited CL individuals. Numbers of fish with each genotype are given.

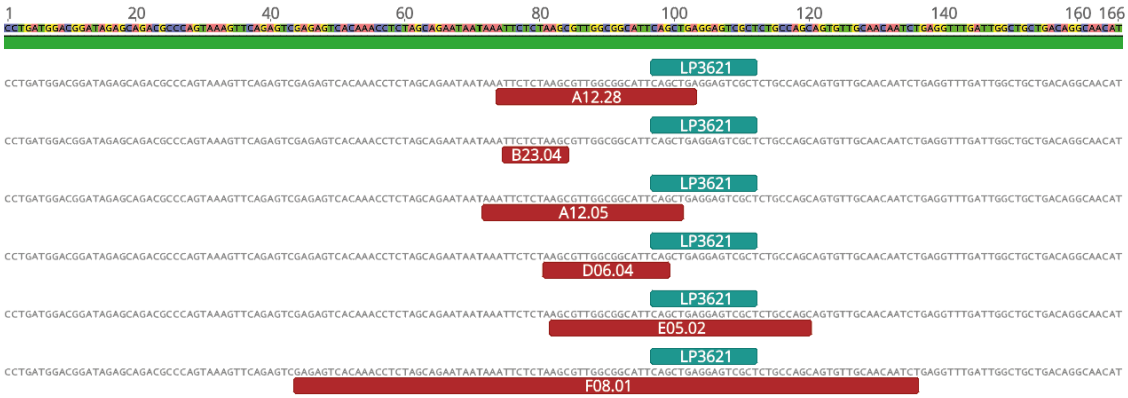


Figure 6 – Deletion edits found after Sanger sequencing of F1 transgenic fish. Annotations in red highlight the region around the marine allele of *LP3621* that was deleted. Names in the deletion annotations are fish IDs of the individuals where the deletion was identified. Two out of the six deletions are fully overlapping with *LP3621* (bottom two deletions).

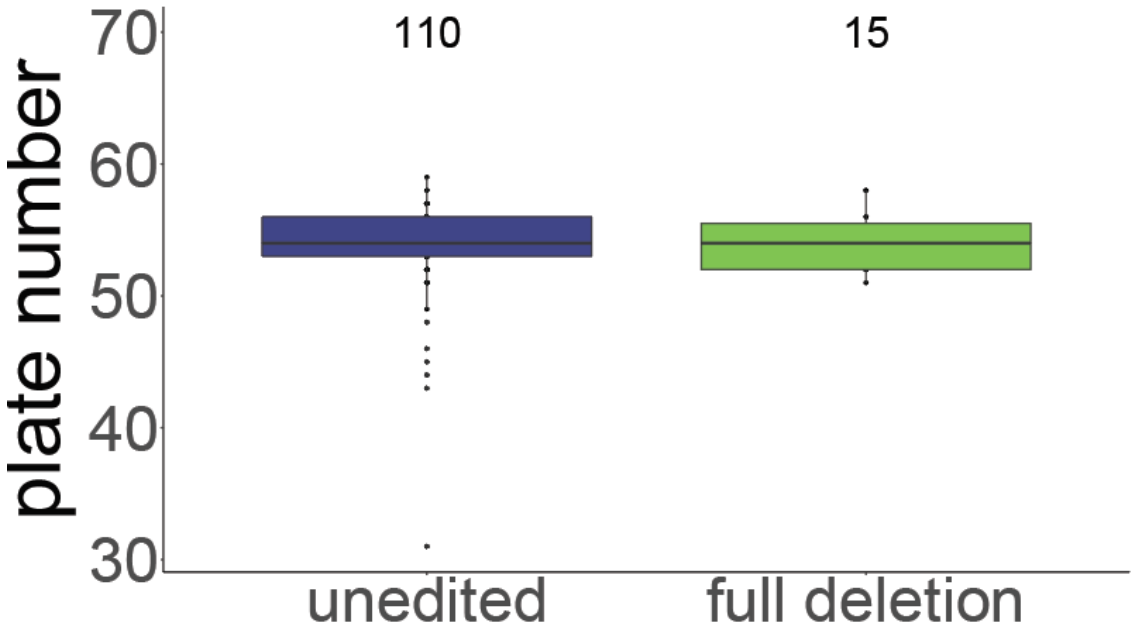
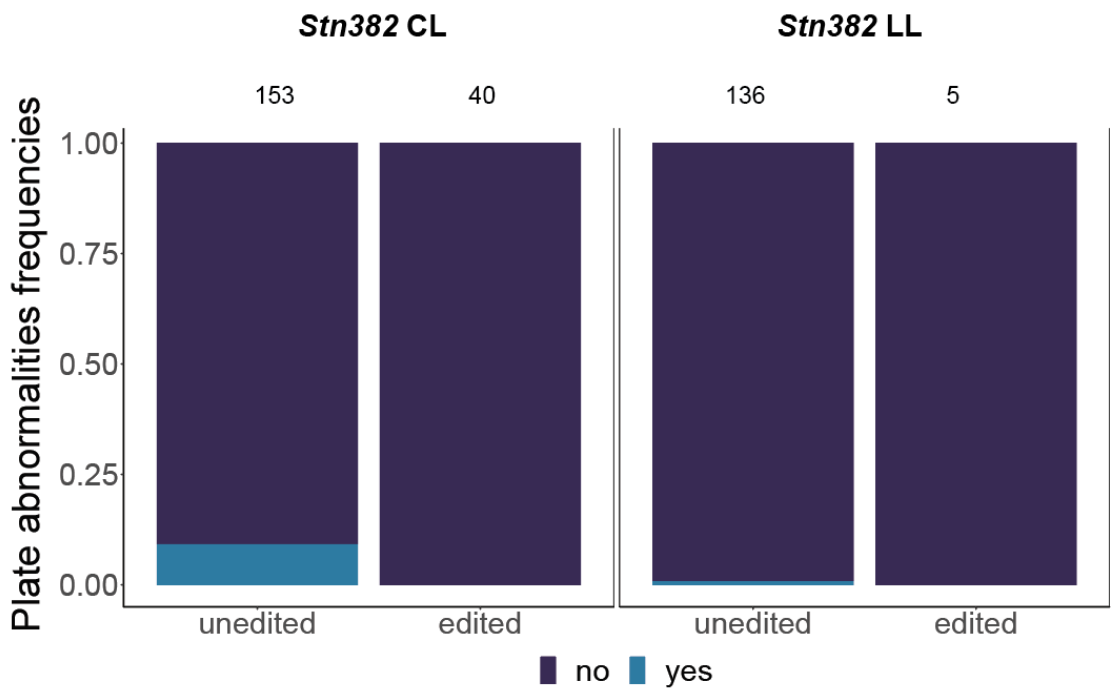


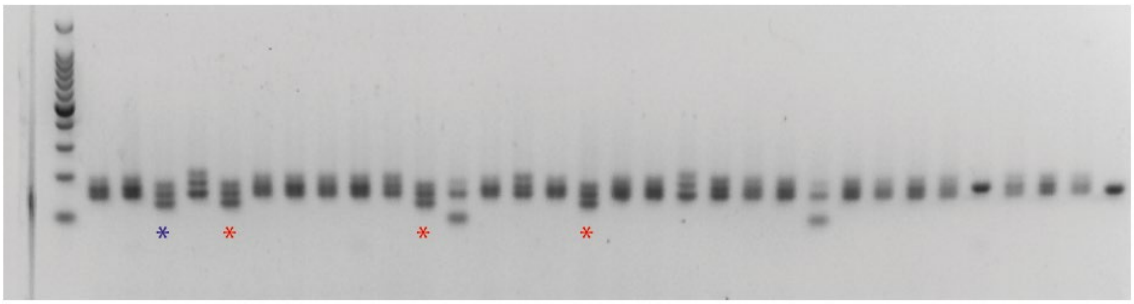
Figure 7 – Effect on the number of lateral plates of a full deletion of *LP3621* in the marine allele of heterozygous fish (CL at *Stn382*). Unedited individuals had a normal CL genotype at *LP3621*. The "full deletion" individuals to the right include the two sequenced *LP3621* full deletion fish and their 13 siblings with similar deletion bands in their *LP3621* genotype (Supplementary Figure S3).



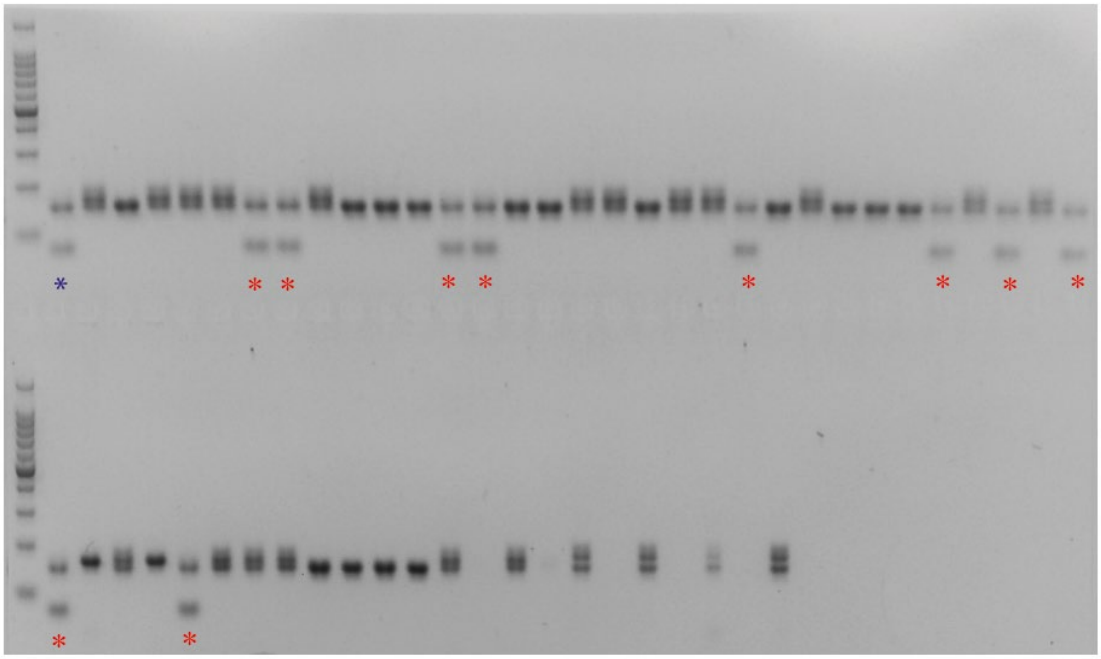
Supplementary Figure S1 – Frequencies of plate abnormalities between F1 fish with unedited and edited genotypes at *LP3621*. Figure is separated by *Eda* (*Stn382*) CL fish (left panel) and *Eda* (*Stn382*) LL fish (right panel). Numbers of fish with each genotype are given.

Supplementary Figure S2 – Insertion edits found after Sanger sequencing of a subset of the F1 fish with evidence of editing at *LP3621*. Sequence is the standard marine allele at *LP3621*, purple boxes highlight the two base pairs between which the insertions happened. Insertion IDs correspond to the fish IDs where they were identified and the size of the insertion in base pairs (bp).

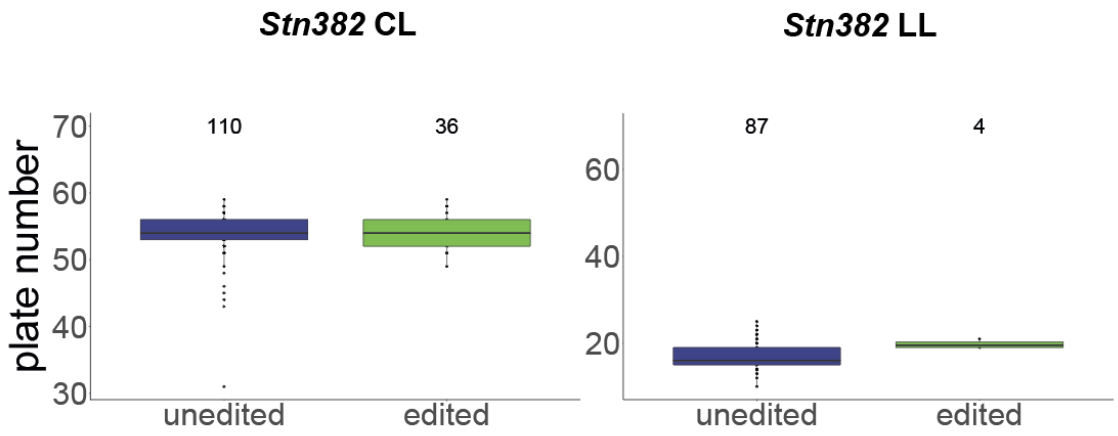
a)



b)



Supplementary Figure S3 – Genotyping results of the crosses where the a) 39 bp and the b) 93 bp deletions fully encompassing *LP3621* were found. Blue asterisk marks the sample that was Sanger sequenced, and red asterisk marks siblings whose edit genotype looked the same and therefore, where *LP3621* was also likely deleted.



Supplementary Figure S4 – Effect of editing (deletions and insertions) at *LP3621* on the number of lateral plates for *Eda* CL and LL F1 fish. This includes edits that are on-target, overlapping or off-target but in a vicinity of 150 bp from *LP3621* (Figure 6 and Supplementary Figure S2).

References

Amenyogbe E, Chen G, Wang Z, Lu X, Lin M, Lin AY. 2020. A review on sex steroid hormone estrogen receptors in mammals and fish. *Int J Endicrol* 2020:1–9.

Archambeault SL, Bärtschi LR, Merminod AD, Peichel CL. 2020. Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback Eda locus. *Evol Lett* 4:282–301.

Archambeault SL, Durston DJ, Wan A, El-Sabaawi RW, Matthews B, Peichel CL. 2020. Phosphorus limitation does not drive loss of bony lateral plates in freshwater stickleback (Gasterosteus aculeatus). *Evolution* 74:2088–2104.

Barrett RDH, Rogers SM, Schluter D. 2008. Natural selection on a major armor gene in threespine stickleback. *Science* 322:255–257.

Bell MA, Aguirre WE. 2013. Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evol Ecol Res* 15:377–411.

Bell MA, Aguirre WE, Buck NJ. 2004. Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution* 58:814–824.

Bell MA, Foster SA. 1994. The evolutionary biology of the threespine stickleback. Oxford; New York: Oxford University

Bomblies K, Peichel CL. 2022. Genetics of adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2122152119.

Burgos-Aceves MA, Cohen A, Smith Y, Faggio C. 2016. Estrogen regulation of gene expression in the teleost fish immune system. *Fish Shellfish Immunol* 58:42–49.

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327:302–305.

Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, Miller CT. 2014. Evolved tooth gain in sticklebacks is associated with a *cis* -regulatory allele of *Bmp6*. *Proc. Natl. Acad. Sci. U.S.A.* 111:13912–13917.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307:1928–1933.

Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, Kingsley DM. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol* 2:e109.

Courtier-Orgogozo V, Arnoult L, Prigent SR, Wiltgen S, Martin A. 2020. Gephebase, a database of genotype–phenotype relationships for natural and domesticated variation in Eukaryotes. *Nucleic Acids Res* 48:D696–D703.

Erickson PA, Baek J, Hart JC, Cleves PA, Miller CT. 2018. Genetic dissection of a supergene implicates *Tfap2a* in craniofacial evolution of threespine Sticklebacks. *Genetics* 209:591–605.

Gehring WJ, Hiromi Y. 1986. Homeotic genes and the homeobox. *Annu Rev Genet* 20:147–173.

Gelmond O, von Hippel FA, Christy MS. 2009. Rapid ecological speciation in three-spined stickleback Gasterosteus aculeatus from Middleton Island, Alaska: the roles of selection and geographic isolation. *J Fish Biol* 75:2037–2051.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.

Kitano J, Bolnick DI, Beauchamp DA, Mazur MM, Mori S, Nakano T, Peichel CL. 2008. Reverse evolution of armor plates in the threespine stickleback. *Curr Biol* 18:769–774.

Kitano J, Ishikawa A, Ravinet M, Courtier-Orgogozo V. 2022. Genetic basis of speciation and adaptation: from loci to causative mutations. *Phil. Trans. R. Soc. B* 377:20200503.

Knecht AK, Hosemann KE, Kingsley DM. 2007. Constraints on utilization of the EDA-signaling pathway in threespine stickleback evolution. *Evol Dev* 9:141–154.

Laurentino TG, Boileau N, Ronco F, Berner D. 2022. The ectodysplasin-A receptor is a candidate gene for lateral plate number variation in stickleback fish. *G3 (Bethesda)*:jkac077.

LE Rouzic A, Østbye K, Klepaker TO, Hansen TF, Bernatchez L, Schluter D, Vøllestad LA. 2011. Strong and consistent natural selection associated with armour reduction in sticklebacks. *Mol Ecol* 20:2483–2493.

Maclean JA, Chen MA, Wayne CM, Bruce SR, Rao M, Meistrich ML, Macleod C, Wilkinson MF. 2005. Rhox: a new homeobox gene cluster. *Cell* 120:369–382.

Mills MG, Greenwood AK, Peichel CL. 2014. Pleiotropic effects of a single gene on skeletal development and sensory system patterning in sticklebacks. *Evodevo* 5:5.

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:e05290.

Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Phil. Trans. R. Soc. B* 372:20150486.

Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. 2001. The genetic architecture of divergence between threespine stickleback species. *Nature* 414:901–905.

Pinto PIS, Estêvão MD, Power DM. 2014. Effects of estrogens and estrogenic disrupting compounds on fish mineralized tissues. *Mar Drugs* 12:4474–4494.

Reimchen TE. 1992. Injuries on stickleback from attacts by a toothed predator (Oncorhychus) and implications for the evolution of lateral plates. *Evolution* 46:1224–1230.

Reimchen TE. 2000. Predator handling failures of lateral plate morphs in *Gasterosteus* aculeatus: functional implications for the ancestral plate condition. *Behaviour* 137:1081–1096.

Reimchen TE, Bergstrom* C, Nosil P. 2013. Natural selection and the adaptive radiation of Haida Gwaii stickleback. *Evol Ecol Res* 15:241–269.

Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, et al. 2021. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv.* 7:eabg5285.

Rodríguez-Ramírez CE, Hiltbrunner M, Saladin V, Walker S, Urrutia A, Peichel CL. 2023. Molecular mechanisms of *Eda* -mediated adaptation to freshwater in threespine stickleback. *Mol Ecol*:mec.16989.

Schluter D, Marchinko KB, Arnegard ME, Zhang H, Brady SD, Jones FC, Bell MA, Kingsley DM. 2021. Fitness maps to a large-effect locus in introduced stickleback populations. *Proc. Natl. Acad. Sci. U.S.A.* 118:e1914889118.

Wong MM-K, Joyson SM, Hermeking H, Chiu SK. 2021. Transcription Factor AP4 Mediates Cell Fate Decisions: To Divide, Age, or Die. *Cancers (Basel)* 13:676.

Chapter 2

Molecular mechanisms of *Eda*-mediated adaptation to freshwater in threespine stickleback

Molecular mechanisms of *Eda*-mediated adaptation to freshwater in threespine stickleback

Carlos E. Rodríguez-Ramírez[1], Melanie Hiltbrunner[1], Verena Saladin[1], Stephanie Walker[1], Araxi Urrutia[2,3], Catherine L. Peichel[1]

[1]*Division of Evolutionary Ecology, Institute of Ecology and Evolution, University of Bern, Bern 3012, Switzerland*

[2]*Department of Biology and Biochemistry, and Milner Centre for Evolution, University of Bath, Bath, BA2 7AY, UK*

[3]*Institute of Ecology, UNAM, 04510, Mexico City, Mexico*

Abstract

A main goal of evolutionary biology is to understand the genetic basis of adaptive evolution. Although the genes that underlie some adaptive phenotypes are now known, the molecular pathways and regulatory mechanisms mediating the phenotypic effects of those genes often remain a black box. Unveiling this black box is necessary to fully understand the genetic basis of adaptive phenotypes, and to understand why particular genes might be used during phenotypic evolution. Here, we investigated which genes and regulatory mechanisms are mediating the phenotypic effects of the *Eda* haplotype, a locus responsible for the loss of lateral plates and changes in the sensory lateral line of freshwater threespine stickleback (*Gasterosteus aculeatus*) populations. Using a combination of RNAseq and a cross design that isolated the *Eda* haplotype on a fixed genomic background, we found that the *Eda* haplotype affects both gene expression and alternative splicing of genes related to bone development, neuronal development and immunity. These include genes in conserved pathways, like the BMP, netrin and bradykinin signalling pathways, known to play a role in these biological processes. Furthermore, we found that differentially expressed and differentially spliced genes had different levels of connectivity and expression, suggesting that these factors might influence which regulatory mechanisms are used during phenotypic evolution. Taken together, these results provide a better understanding of the mechanisms mediating the effects of an important adaptive locus in stickleback and suggest that alternative splicing could be an important regulatory mechanism mediating adaptive phenotypes.

Keywords: *Eda*, threespine stickleback, genetics of adaptation, alternative splicing, lateral plates, lateral line

1 Introduction

Understanding the connection between genetic variation and adaptive phenotypic variation is one of the main goals in evolutionary genetics. It is a challenging task, but in recent years the genes that underlie adaptive traits have been identified in some systems (Bomblies & Peichel, 2022). For example, a difference in coat color between deer mice (*Peromyscus maniculatus*) on different soils is controlled by *Agouti* (Linnen et al., 2009); loss of defensive lateral plates in freshwater threespine stickleback (*Gasterosteus aculeatus*) is controlled by *Eda* (Colosimo et al., 2005); industrial melanisation of the peppered moth (*Biston betularia*) was caused by the insertion of a transposable element in the first intron of *cortex* (Hof et al., 2016); and pollinator-specific flower colour in two sister species of monkeyflowers (*Mimulus lewisii* and *Mimulus cardinalis*) is controlled by *LAR1* (Yuan et al., 2016). However, even when a specific adaptive locus has been identified, the specific regulatory mechanisms and downstream molecular pathways mediating its effects on phenotypic variation often remain unknown (Bomblies & Peichel, 2022). A better understanding of how the genetic changes in adaptive loci impact the interactions of these genes in regulatory networks might explain why certain genes and molecular pathways tend to be re-used in the evolution of certain phenotypes instead of functionally similar alternatives (Stern, 2013).

While most studies have focused on changes in gene expression as a mechanism underlying phenotypic evolution, a growing body of evidence suggests that alternative splicing might also be important for adaptation and phenotypic evolution (Bush et al., 2017; L. Chen et al., 2012; Singh & Ahi, 2022; Verta & Jacobs, 2022; Wright et al., 2022). Alternative splicing (AS) regulates which exons and/or introns from a gene are retained in the mature messenger RNA (mRNA), allowing different mRNA isoforms and proteins to be coded from the same gene, thereby increasing proteomic diversity. AS has been found in animals, plants, and fungi (Bush et al., 2017; Chaudhary et al., 2019; Singh & Ahi, 2022; Wright et al., 2022). Between 92 to 95% of the genes in the human genome are estimated to undergo alternative splicing (Pan et al., 2008; E. T. Wang et al., 2008). Types of AS include exon skipping, exon shuffling, intron retention, and use of alternative 5’ and 3’ splice sites. Exon skipping is the most common type in animals, while intron retention is more common in plants (Kim et al., 2007; Marquez et al., 2012; Wang & Brendel, 2006). Several recent studies have found evidence for a role of AS in both phenotypic evolution and adaptation. For example, a mutation affecting splicing in *Msx2a* contributes to reduction in dorsal spine length, a trait involved in defense against predators, in freshwater threespine stickleback populations (Howes et al., 2017), and upregulation of an *Agouti* splice isoform is involved in the evolution of cryptic coat coloration in two species of deer mice (Mallarino et al., 2017). Genome-wide transcriptomic analyses have revealed changes in splicing between genetically-

similar but phenotypically-distinct head and body lice ecotypes (Tovar-Corona et al., 2015), between jaws from cichlid species occupying different trophic niches (Singh et al., 2017), and between benthic and pelagic ecotypes of Arctic charr (Jacobs & Elmer, 2021). These data point to the potential for AS to underlie adaptive phenotypic variation; however, the relative contribution of AS to adaptive phenotypic variation in comparison with differential gene expression is not well understood.

Threespine stickleback (*G. aculeatus*) are a great model to study the genetic and molecular mechanisms of adaptation. After the Last Glacial Maximum, approximately 15,000 years ago, individuals from marine populations in the Northern hemisphere independently colonized newly formed freshwater environments, resulting in the repeated evolution of phenotypic differences between marine and freshwater sticklebacks (Bell & Foster, 1994). This independent and replicated adaptation to freshwater makes threespine stickleback a very powerful system to study questions related to adaptation, phenotypic evolution, and the repeatability of evolution (Peichel & Marques, 2017). One well-studied trait is the repeated loss of bony lateral plates in most freshwater populations. These bony plates are known to provide protection against bird and fish predation in clear and open-water environments, such as the ocean or large lakes (Kitano et al., 2008; Leinonen et al., 2011; Reimchen, 1992, 2000; Reimchen et al., 2013). Several studies have documented rapid and strong selection for the loss of lateral plates in freshwater (Barrett et al., 2008; Bell et al., 2004; Bell & Aguirre, 2013; Gelmond et al., 2009; LE Rouzic et al., 2011; Rennison et al., 2015; Schluter et al., 2021) although the selective pressure driving this lateral plate reduction is still not clear (Archambeault, Durston, et al., 2020), Gene mapping and transgenic studies have shown that *Ectodysplasin A* (*Eda*) is the main gene controlling this phenotype (Colosimo et al., 2004, 2005). *Eda* signalling is known to affect the development of ectodermal appendages like hair, teeth, feathers and scales in vertebrates, (Cui & Schlessinger, 2006; Sadier et al., 2014). In threespine stickleback, *Eda* also has pleiotropic effects on the patterning of the sensory neuromasts that make up the lateral line (Archambeault, Bärtschi, et al., 2020; Mills et al., 2014; Wark et al., 2012) and in schooling behaviour (Greenwood et al., 2013, 2016). In the threespine stickleback genome, *Eda* is in a 16 kb haplotype on chromosome IV that contains fixed genetic differences between marine and freshwater populations (Archambeault, Bärtschi, et al., 2020; Colosimo et al., 2005; Jones et al., 2012; O'Brown et al., 2015). Individuals that have two marine alleles (hereafter called C) of this haplotype are completely-plated, while individuals that have two freshwater alleles (hereafter called L) are low-plated. In the Puget Sound population used for this study, fish that are heterozygous for *Eda* are completely-plated (Archambeault, Bärtschi, et al., 2020), but this is not the case in all populations (Colosimo et al., 2004; Laurentino et al., 2022). The haplotype also includes two other genes, *Tumor necrosis factor superfamily member 13b* (*Tnfsf13b*) and *Glycoprotein*

A rich protein (*Garp*). Both genes have immune functions in humans; *Tnfsf13b* codes for a cytokine (BAFF) that is important for B cell survival and homeostasis (Schweighoffer & Tybulewicz, 2018; Smulski & Eibel, 2018), while *Garp* codes for a transmembrane receptor protein that regulates the function of regulatory T-cells (Metelli et al., 2018). It is still unclear whether these two genes play a role in freshwater adaptation in threespine stickleback, by for example mediating immune differences between the ecotypes, or if they are just tightly linked with *Eda* in the haplotype. There is some evidence for an effect of the *Eda* haplotypes in the expression of target immune genes in F2 individuals derived from marine and freshwater crosses (Robertson et al., 2017), which raises the possibility of an adaptive role of these two gene. However, this study did not have the resolution to distinguish between the effects of the *Eda* haplotype and linked genes.

Despite our knowledge of the link between the *Eda* genotype and several phenotypes, we still have little knowledge of the downstream molecular mechanisms by which the *Eda* haplotype mediates its known phenotypic effects or whether there are other phenotypic effects of the haplotype. To address these questions, here we compare the transcriptomes of threespine stickleback siblings that possess the three different genotypes (CC, CL and LL) at the 16 kb *Eda* haplotype but otherwise share the same genomic background. We compared these individuals across two tissues: skin, where the lateral line and lateral plates develop; and head kidney which is a primary hemopoietic organ in bony fish similar to the bone marrow in mammals (Soulliere & Dixon, 2017). Specifically, we asked three main questions: 1) what is the effect of the *Eda* haplotype on differential gene expression and alternative splicing?; 2) can we identify candidate genes and pathways that mediate the known phenotypic effects of *Eda*?; and 3) does the *Eda* haplotype change the expression and/or splicing of other genes and pathways that might mediate other, previously unknown, phenotypic effects?

2 Materials and Methods

2.1 Ethics statement

Animal husbandry and experimental procedures were approved by the Veterinary Service of the Department of Agriculture and Nature of the Canton of Bern (VTHa# BE4/16 and BE82/17).

2.2 Fish cross design and care

To quantify the effects of the *Eda* haplotype on the transcriptome, we crossed marine threespine stickleback that were heterozygous for the *Eda* haplotype. This cross design provided fish with the same genomic background that varied only on their *Eda* genotype, thus disentangling the effect of the *Eda* haplotype from the rest of the genome (Figure 1). The individuals used in this study were F3 descendants of heterozygous wild fish collected in Puget Sound, WA, USA in the summers of 2015 and 2016 as previously described (Archambeault, Bärtschi, et al., 2020; Archambeault, Durston, et al., 2020). We generated these F3 individuals by making three independent crosses (families A, B and C) between F2 females and males that were heterozygous for the *Eda* haplotype. The resulting F3 fish were raised at approximately 15.0°C in near freshwater conditions of 3.5 parts per thousand (ppt) Instant Ocean salt (Aquarium Systems, Sarrebourg, France). Fish were fed brine shrimp nauplii twice a day, except for weekends when they were fed only once a day. They were exposed to a light cycle of 11h of daylight (3450 lumens), 1h of sunset, 11h of moonlight (600 lumens) and 1h of sunrise. When the F3 fish were between 129 day and 131 days post fertilization, two males and two females per *Eda* genotype (CC, CL or LL) from each of the three families (for a total of 36 individuals) were sacrificed in MS-222, skin and head kidney were dissected, and RNA was extracted for subsequent RNA sequencing (Figure 1).

2.3 DNA extractions and genotyping

DNA was extracted from fin tissue using a modified HotSHOT DNA extraction method as described (Archambeault, Bärtschi, et al., 2020; Archambeault, Durston, et al., 2020). Parents of the F3 crosses were genotyped at several markers in the *Eda* haplotype listed in Supplementary Table S1 to confirm they had the full 16 kb *Eda* haplotype. The F3 individuals were genotyped at *Stn382* to identify their *Eda* genotype and at *LRR* to identify their sex (Supplementary Table S1).

2.4 Dissections, RNA extraction, and sequencing

We dissected skin and head kidney from 36 individuals for RNA-sequencing. Skin was dissected from both sides of the posterior flank of the fish (starting at the level of the third spine, until the end of the dorsal fin), which is the region where LL sticklebacks do not have lateral plates and CL and CC sticklebacks do. RNA was extracted using an Invitrogen TRIzol kit (Invitrogen, Basel, Switzerland) according to the manufacturer's instructions. RNA concentration was measured for each sample using the Qubit RNA B Assay Kit (Invitrogen, Basel, Switzerland), and RNA quality was determined on a Fragment Analyzer CE12 (Advanced Analytics, Agilent, Santa Clara, CA, USA). The Next Generation Sequencing Platform of the University of Bern prepared the TruSeq Stranded mRNA library preparation for each of the 72 samples (36 skin, 36 head kidney) and performed the paired-end sequencing of the 72 libraries with 300 cycles on an Illumina NovaSeq 6000 S2 flow cell.

2.5 RNA-seq data pre-processing

The quality of the RNAseq reads was verified with *FastQC* v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We mapped the reads to the threespine stickleback reference genome v5 (Nath et al., 2021), using *STAR* v2.7.3a (Dobin et al., 2013) following the parameters previously used for threespine stickleback (Verta & Jones, 2019): *-outFilterIntronMotifs: RemoveNoncanonicalUnannotated*; *-chimSegmentMin* 50; *-alignSJDBoverhangMin* 1; *-alignIntronMin* 20; *-alignIntronMax* 200000; *--alignMatesGapMax* 200000 and *--limitSjdbInsertNsj* 2000000. However, we did not run STAR in the 2pass mode, because it increased our multimapping read rate by 5% and we did not benefit from the de novo splice junction identification since our downstream analysis focused on annotated genomic features. Next, we used *FeatureCounts* v2.0.1 (Liao et al., 2014) to count how many reads mapped to each genomic feature. We did this at two different levels: gene and exon. We ran featureCounts in paired-ended mode (-p), allowing only for reversely stranded alignments (-s 2), as per the characteristics of our read libraries and excluding read pairs where one of the mates did not map (-B) or if they mapped into a different strand or chromosome (-C). We used MultiQC v1.8 (Ewels et al., 2016) to summarize the quality reports for all samples from FastQC, STAR and featureCounts. One skin sample from a heterozygous (CL) female from family C was removed from all further analysis because it had high multimapping rates in STAR (35.7%). All computationally intensive calculations were performed on the University of Bern HPC cluster UBELIX (http://www.id.unibe.ch/hpc).

2.6 Identification of differentially expressed genes (DEGs)

For the differential expression analysis we used R v3.6.1 (R Core Team, 2019) and *edgeR* v3.26.8 (Robinson et al., 2010) available at the Bioconductor website (http://bioconductor.org). We used the gene-level read counts we obtained from *featureCounts* as input and started by filtering lowly-expressed genes, i.e. genes with fewer than 10 read counts in 11 or more of the 35 (skin) or 12 or more of the 36 (head kidney) samples analyzed for a given tissue. Next, we calculated library normalization factors for all samples and estimated gene expression dispersions using a weighted likelihood Empirical Bayes approach. Then, we used the *plotMDS()* function of the *limma* v3.40.6 R package (Ritchie et al., 2015) to run a modified multidimensional scaling (MDS) analysis which calculates the distance between each pair of samples based on the 500 top genes with the highest gene expression fold-changes between that pair of samples. Afterwards, we fitted all data to a negative binomial generalized linear model (GLM) model using genotype as the main explanatory variable and controlling for family and sex effects. Finally, we used a quasi-likelihood F-test to identify differentially expressed genes (DEGs) between genotypes. Instead of testing for fold-change differences from zero between our genotypes, we tested for differential expression relative to a minimum fold-change threshold using the *edgeR* implementation of the TREAT method (McCarthy & Smyth, 2009). We focused on genes with a significantly higher than 0.585 log2 fold-change (approximately a 1.5 fold-change in gene expression) between genotypes. We set the p-value cut-off to 0.05 and performed correction for multiple testing with the False Discovery Rate (FDR) method (Benjamini & Hochberg, 1995).

2.7 Identification of differentially spliced genes (DSGs)

One method to identify differential splicing is to test genes for differential exon usage. This is based on the principle that when the splicing pattern of a gene changes, the relative expression of the exons of that gene also change. Though it cannot identify all types of AS events, this method can identify exon skipping and exon swapping events, which comprise approximately half of the AS events in humans (Chaudhary et al., 2019). We used *edgeR's* implementation of the differential exon usage test to identify genes with evidence of differential splicing. We used the exon-level count data from *featureCounts* as the input, and applied the same filtering, variance estimation, and GLM fitting steps to the data as we did for the gene-level data for the differential expression analysis. We then used quasi-likelihood F-tests to identify differential exon usage using the two complementary methods in *edgeR*. The first method, called the "gene-level" method, uses the exon-level test statistics to obtain a gene-level p-value, while the second method, called the Simes' method (Simes, 1986), first calculates exon-level p-values and converts them into a single gene-level p-value . The "gene-level" method is better at detecting genes with several differentially spliced exons while the second method is better at

identifying genes with only a minority of differentially spliced exons (Y. Chen et al., 2008). Any gene found to have significant differential exon usage by either one or both methods was reported as a putatively differentially spliced gene (DSG).

2.8 Gene co-expression analysis

To identify putative gene interaction networks in the skin and head kidney transcriptomes, we did a weighted gene co-expression network analysis (Zhang & Horvath, 2005) using the R package *WGCNA v1.69* (Langfelder & Horvath, 2008). This analysis uses pairwise gene expression correlations across the transcriptome to infer how connected genes are to each other and to identify clusters of co-expressed genes (modules) whose gene expression is highly correlated and therefore expected to be working together in the same biological processes. We used *featureCounts* count data filtered by *edgeR*'s gene expression filter as input for the analysis. Following *WGCNA* recommendations (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/index.html),we normalized and applied a variance stabilizing transformation on the count data using the *vst()* function of the *DESeq2 v1.24.0* R package (Love et al., 2014) and adjusted for the family effect using the *ComBat()* function from the *sva v3.32.1* R package (Leek et al., 2012). Using the normalized and adjusted data, we created a gene similarity matrix using the absolute value of the pairwise biweight midcorrelation between all genes in our dataset. Next, we calculated a weighted adjacency matrix from the similarity matrix by rising the latter to a power of β. This power of β is referred to as the soft threshold of the analysis because it is used to emphasize strong correlations in the weighted adjacency matrix and de-emphasize weaker gene correlations. To calculate the appropriate value of the soft threshold for our data, we plotted the fit of our data to an approximate scale-free topology model (Zhang & Horvath, 2005) for different values of β using the *WGCNA* function *pickSoftThreshold()*. The plot revealed a saturation of the scale-free topology model fit for soft thresholds of 14 for the skin and 12 for the head kidney data (Supplementary Figure S1). To enable cross-tissue comparisons, we selected a conservative soft threshold of 14 for both tissues. Next, to further minimize the effect of noise and random correlations, we calculated a topological overlap matrix (TOM) from the adjacency matrix. The TOM was calculated by analysing not only the adjacency between a pair of genes, but also the overlap and similarity of their adjacency with other "third party" genes. Finally, a hierarchical clustering algorithm was used to define the gene co-expression modules. These steps were all performed by inputting the adjusted and normalized count data to the *blockwiseModules()* function of *WGCNA* with the following settings: *corType = "bicor", maxPOutliers = 0.10, maxBlockSize = 18000, TOMType = "signed", power = 14, randomSeed = 1234.*

We also used *WGCNA* to obtain measures of network total connectivity (kTotal) for every gene in the skin and head kidney transcriptomes. The total connectivity of a gene is a measure of how co-expressed that gene is with all other genes in the transcriptome, and it is calculated by summing the adjacency values of that gene with all other genes. Gene connectivity has previously been used as a proxy for pleiotropy (Featherstone & Broadie, 2002; Hämälä et al., 2020; Jacobs & Elmer, 2021; Rennison & Peichel, 2022; Wagner et al., 2007). We therefore used gene connectivity as a proxy to compare the levels of pleiotropy among three sets of genes: skin DEGs, skin DSGs, and the complete skin transcriptome. For this analysis we only included genes that were solely DEGs or DSGs, removing the six genes that were both DEGs and DSGs. We calculated the kTotal connectivity distribution of these sets of genes and then did pairwise comparisons of their medians. To test if the differences in the kTotal medians were significant, we used permutation tests. For each pairwise median kTotal comparison, we generated 10 000 random sets of genes with the same size as the sets of genes we were comparing and calculated the ratio of how many times the absolute difference in kTotal of the random permuted sets was the same or greater than the absolute differences in the real sets being compared. Using the same permutation approach, we similarly compared the medians of the distributions of average gene expression levels of skin DEGs-only, skin DSGs-only, and the complete skin transcriptome. It was not possible to do this for the head kidney data due to the lack of DEGs and low number of DSGs in this tissue.

2.9 Gene Ontology enrichment analysis

We did a Gene Ontology (GO) enrichment analysis to identify GO terms overrepresented in the DEGs, DSGs, and co-expressed gene modules in the g:GOst module of the g:Profiler webservice (Raudvere et al., 2019; Reimand et al., 2007) (https://biit.cs.ut.ee/gprofiler/gost). We selected the Ensembl stickleback annotation database for the analysis and used the list of genes that passed *edgeR's* gene expression filter as a background. All other settings were left on default. G:Profiler results also provided Human Phenotype (HP) annotations for stickleback, however these were not included in the analysis as they did not add more information than the GO Terms. To summarize these results, we followed a published protocol (Reimand et al., 2019) to build a network of enriched GO Terms by gene overlap using the Enrichmap v3.3.2 app of Cytoscape v3.8.2 (Merico et al., 2010; Shannon et al., 2003). The resulting GO Term networks were given representative names based on the terms present in the network using the default settings of the AutoAnnotate v1.3.4 app of Cytoscape (Kucera, 2017).

2.10 Identifying putative immune functions of DEGs and DSGs

To determine whether the *Eda* haplotype might have an influence on immunity, we also manually looked up the function of every LL vs CC DEG in skin and all DSGs found in both tissues in NCBI Gene (https://www.ncbi.nlm.nih.gov/gene) or in Zfin (https://zfin.org/) and GeneCards (https://www.genecards.org/). When genes were identified as having immune functions in these databases, we looked for supporting literature. Genes with clear evidence of having important immune roles in other organisms were considered putative immune genes.

3 Results

3.1 The *Eda* haplotype affects the expression and alternative splicing of hundreds of genes

Our results show an effect of the *Eda* haplotype in the skin and head kidney transcriptomes when controlling the genomic background, although the magnitude of this effect is quite different between the two tissues (Figure 2 and Figure 3). In both tissues, an MDS analysis of the pairwise expression changes of all genes between samples separates individuals only by family (Supplementary Figure S2). However, when focusing the analysis on the top 500 genes with the largest changes in gene expression between each pair of samples, the second dimension separated the LL *Eda* samples from the CC and CL *Eda* samples in skin (Figure 2a), mirroring the pattern of the plate phenotypes associated with these genotypes. This was not the case in head kidney, where the family effect was still the only factor driving the MDS (Figure 2b). We did not find any clustering of the samples by sex in the first two dimensions of either the skin or head kidney MDS (Supplementary Figure S2). Consistent with the MDS results, we found no differentially expressed genes (DEGs) in head kidney and hundreds of DEGs in skin (Figure 3 and Supplementary Table S2). There are dozens of differentially spliced genes (DSGs) between *Eda* genotypes in both skin and head kidney, but there are fewer DSGs in head kidney than in skin (Figure 3 and Supplementary Table S2). *Eda* itself was a DEG in the skin CC vs LL comparison, but the other two genes of the haplotype, *Tnfsf13b* and *Garp* were not. More than half of the skin DEGs and DSGs between LL vs CL are also present in the LL vs CC comparison. The LL vs CC comparison captures approximately 94.1% of the DEGs and 59.3% of the DSGs in skin (Supplementary Figure S3). Considering that CC and CL individuals have very similar lateral plate and lateral line phenotypes in this population (Archambeault, et al., 2020a), that the LL vs CC comparison allowed us to clearly distinguish the effect of the two *Eda* alleles, and that the effect of the *Eda* haplotype was stronger in the skin than in the head kidney transcriptome, we focused most of our downstream analysis on the LL vs CC comparison in the skin transcriptome.

3.2 Differentially expressed genes (DEGs) and differentially spliced genes (DSGs) in skin are mostly non-overlapping

Of the 241 DEGs and 34 DSGs between the *Eda* CC and LL skin samples, only six were both differentially expressed and differentially spliced. Although this overlap is significant ($p < 0.0001$, 10,000 permutations), the low overlap suggests that these two regulatory mechanisms are mostly independent from each other (Figure 3c). It has been suggested that differential splicing might avoid

constraints associated with differential expression of highly pleiotropic genes (Jacobs & Elmer, 2021; Rogers et al., 2021). To test whether pleiotropy could explain why some genes are regulated through gene expression and others through alternative splicing in our study, we compared gene co-expression connectivity of DEGs and DSGs as a proxy for pleiotropy (see Methods). We compared the connectivity distributions of the DEGs, DSGs, and the transcriptome-wide distribution and found that the DEGs have a higher total connectivity than the DSGs (DEGs median kTotal = 83.12, DSGs median kTotal = 46.22; $p < 0.0001$, 10,000 permutations), and that the DSGs had a total connectivity distribution not significantly different from the transcriptome-wide distribution (transcriptome median kTotal = 36.39, DSGs median kTotal = 46.22; $p = 0.2785$, 10,000 permutations) (Figure 4a). Interestingly, we found the opposite pattern when comparing gene expression levels between the DEGs and the DSGs, with the DSGs more highly expressed than the DEGs (DSGs median = 21.38 TPM, DEG median = 3.07 TPM; $p <0.0001$, 10,000 permutations) (Figure 4b). The DEGs also have a lower expression than the transcriptome-wide median (DEGs median = 3.07 TPM, transcriptome median = 12.84 TPM; $p <1e-4$, 10,000 permutations) while the DSGs have a higher expression (DSGs median = 21.38 TPM, transcriptome median = 12.84 TPM; $p = 0.019$, 10,000 permutations). These results suggest that factors like gene connectivity and expression level might be important in determining the type of regulatory mechanisms used to mediate phenotypic evolution.

To test whether DEGs and DSGs might be working in the same molecular pathways, we again used gene co-expression analysis to identify modules of strongly co-expressed genes. Of the 37 co-expression modules we identified in skin (Supplementary Figure S4), seven contain at least one DEG or DSG, and four of these contain both classes of genes (Table 1). Most of the DEGs (including *Eda*) are in a single co-expression module (module M5), and a smaller cluster of 12 DEGs is in another module (module M27). Only seven out of 241 DEGs were not present in any co-expressed module. By contrast, most (21 out of 34) of the DSGs were not in any of the co-expression modules (Table 1). Five out of the six genes that were both DSGs and DEGs were found with most of the DEGs in module M5. These results suggest that most of the DEGs we identified are strongly correlated in their gene expression in the skin transcriptome and thus might be working in the same or closely related molecular pathways. Genes that are both DSGs and DEGs are correlated with other DEGs and might be interacting with them; however, most of the DSGs have independent patterns of expression and might have more indirect interactions with the DEGs and each other (Table 1).

3.3 The *Eda* haplotype affects genes involved in bone development, neuronal development, and immune response

Gene Ontology enrichment analysis revealed that the DEGs in skin are enriched in general development and signalling and in more specific processes like bone development (i.e. GO Terms like "ossification", "odontogenesis" and "BMP signalling") and neuronal development ("netrin receptor activity" and "neuromuscular process controlling balance") (Figure 5 and Supplementary Table S3 for the full list of enriched GO Terms). There were no significantly enriched GO Terms for the DSGs, possibly because there were only 34 DSGs. However, more than half of the GO Terms present in DSGs are also present in the DEGs (51 out of 89) (Supplementary Figure S5). Inspection of the individual GO annotations present in the DSGs revealed the presence of two genes with annotations related to cartilage development (C*ol11a2* and *Runx2b*) and three genes with neuronal annotations (*Cln3, Zc4h2* and *Anks1b*) (Supplementary Table S4). Together, these results are consistent with the known effects of *Eda* on the lateral plates and sensory lateral line and suggest that the DEGs and DSGs underlying these GO terms are good candidates to be mediating these phenotypes.

The gene co-expression module M5 (where most of the DEGs are found) reveals similar GO enrichment results to the DEGs, except for the lack of the neuronal GO terms (Supplementary Table S3). However, module 27 has the second-most DEGs (Table 1) and has an enrichment of the "neuromuscular process controlling balance" GO term found in the DEGs (Supplementary Table S3). This module also has several genes annotated as being involved in lateral line development, vestibular reflex, and sound perception, which are all systems that rely on hair cells (Supplementary Table S4). Together, these results suggest that the effect of the *Eda* haplotype on the lateral plates seems to be represented mostly by co-expression module M5, while the effect of the *Eda* haplotype on the patterning of the lateral line is represented by module M27. Interestingly, the gene co-expression network of the ten genes most closely co-expressed with *Eda* (the *Eda* co-expression neighbourhood) plus the top ten connected genes (or hub genes) in modules M5 and M27 position M5 between *Eda* and module M27 (Figure 6). Since the distances in the network are based on how tightly genes are co-expressed, which should correlate with how closely genes interact, the topology of the network suggests that the effect of the *Eda* haplotype on module M5 could be mediated by genes in the *Eda* co-expression neighbourhood. These results further suggest that the effect of the *Eda* haplotype on module M27 could be mediated through the genes in module M5 (Figure 6). However, empirical studies manipulating the genes in these modules are necessary to test this hypothesis.

We also find evidence for a possible effect of the *Eda* haplotype on the immune response of skin and head kidney. The skin DEGs were enriched in GO terms for genes involved in the bradykinin pathway, which are strong inflammatory molecules, and in scavenger receptors which are involved in homeostasis and innate immunity (Alquraini & El Khoury, 2020; Canton et al., 2013) (Figure 5 and Supplemental Table S3). One of the DSGs (*ENSGACG00000014601*) has an Ig-like domain with putative immune function related GO term (Supplementary Table S4). Furthermore, a literature search of the skin DEGs revealed the presence of two genes with interesting links to immunity (Supplementary Table S5). The *Ets1* (ETS proto-oncogene 1) gene is involved in the development and/or function of T cells, B cells and natural killer cells (Dittmer, 2003; Garrett-Sinha, 2013). The *Laptm4b* (lysosomal protein transmembrane 4 beta) gene regulates the immunosuppressor activity of regulatory T cells and is known to interact with *Garp,* one of the genes in the *Eda* haplotype (Huygens et al., 2015) (Supplementary Table S5). Literature research also revealed an important immune gene, *Tbk1,* among the skin DSGs. *Tbk1* plays an important role mediating the interaction between multiple signaling pathways, many of which are related to immunity, namely pathogen detection, inflammation and immune response (Helgason et al., 2013).

In head kidney, the CC vs LL DSGs include genes annotated as involved in innate immune response, hemopoiesis, nervous system development, and epidermal development (Supplementary Table S4). Literature research revealed that the two genes with immune annotations, *itgb2* and *traf3,* do have important immune roles (Supplementary Table S5). *Itgb2* (integrin, beta 2)*,* also known as *lfa-1* and *cd177,* is important for the function and migration of T cells, neutrophils and killer cells (Bai et al., 2017; Barber et al., 2004; Kristóf et al., 2013; Ostermann et al., 2002; Walling & Kim, 2018). *traf3* (TNF receptor-associated factor 3) is a gene that plays an important role in anti-viral innate immune response (Gao et al., 2021; Oganesyan et al., 2006; Tseng et al., 2010) and the regulation of B and T cells (Lin et al., 2015; Yi et al., 2015). Taken together, these results suggest the potential for a pleiotropic role of the *Eda* haplotype on immune function in the skin and head kidney.

3.4 The *Rmnd5b* gene is consistently differentially spliced in both skin and head kidney

To look for general effects of the *Eda* haplotype, we looked for genes that are consistently differentially expressed and/or differentially spliced in both skin and head kidney. There are no consistent DEGs between skin and head kidney, but there is one DSG, *Rmnd5b* (*required for meiotic nuclear division 5 homolog B*) (Supplementary Table S2) which is a gene from chromosome IV located 500 kb downstream of *Eda. Rmnd5b* is a subunit of the GID/CTHL E3 ubiquitin ligase protein, which is involved

in regulating cell proliferation and glucose metabolism (Lampert et al., 2018; Maitland et al., 2022; Santt et al., 2008). In both skin and head kidney, there is differential usage of exon 1 between *Eda* LL and *Eda* CC individuals (Supplementary Figure S6). In head kidney, *Rmnd5b* is found in a co-expression module mostly related to transcriptional regulation but that also includes genes related to myeloid cell homeostasis and erythrocyte differentiation (Supplementary Tables S3 and S6). In skin, *Rmnd5b* is not part of any co-expression module (Supplementary Table S6). However, when we atomize the expression of *Rmnd5b* into its individual exons and analyse their co-expression with the rest of the genes, we find exon 1 of *Rmnd5b* in module M5 together with *Eda* and most DEGs (Supplementary Table S6). These results raise the possibility that differential splicing of *Rmnd5b* might mediate some of the *Eda* haplotype*'s* effects in skin and/or head kidney. The consistent effect on the splicing patterns of *Rmnd5b* suggests that there could be a cis-regulatory effect from the *Eda* haplotype driving these differences. However, we also find two SNPs within exon 11 of *Rmnd5b* (position 13320127 and 13320727) that are always homozygous for the reference allele in *Eda* CC individuals and heterozygous for the alternative allele in *Eda* LL individuals, such that one of the *Eda* haplotype L alleles is in linkage with the alternative allele at these SNPs (Supplementary Table S7). Nonetheless, it is not clear how SNPs in exon 11 of *Rmnd5b* or in the *Eda* haplotype could be acting in cis to drive the consistent change in splicing pattern of exon 1 between *Eda* CC and LL individuals because splicing regulatory regions tend to be found within the exon being regulated or in its adjacent introns (Lee & Rio, 2015; Lovci et al., 2013; Ule & Blencowe, 2019). Thus, further studies are necessary to verify whether and how the *Eda* haplotype affects the splicing of *Rmdn5b* and to test what role this might have in mediating its phenotypic effects.

4 Discussion

In this study we investigated the downstream effects of a 16 kb haplotype that has fixed differences between marine and freshwater threespine sticklebacks. This haplotype includes the gene *Eda* which is responsible for changes in lateral plates, lateral line and schooling behaviour between these ecotypes (Archambeault, Bärtschi, et al., 2020; Colosimo et al., 2004, 2005; Greenwood et al., 2016; Mills et al., 2014). We examined the effect of the *Eda* haplotype in the transcriptomes of skin and head kidney by isolating the three *Eda* genotypes (CC, CL and LL) on the same marine genomic background. There is a significant effect of the *Eda* haplotype on gene expression in skin, with hundreds of genes changing their expression by more than 1.5-fold between genotypes. We also found that the phenotypic effects of the *Eda* haplotype might not only be mediated through changes in the gene expression but also through changes in alternative splicing (Figure 3a). Although differentially spliced genes (DSGs) are mostly non-overlapping with differentially expressed genes (DEGs) (Figure 3b),

several DSGs are involved in the same biological processes as DEGs. This suggests that both processes might be important to mediate the phenotypic effects of the *Eda* haplotype. The DEGs and DSGs in skin are related to skeletal tissue development and neuronal processes, making them good candidates for mediating the known effects of *Eda* on the number of lateral plates and the patterning of the lateral line. Furthermore, we found some evidence for a pleiotropic effect of the *Eda* haplotype in immunity, with genes related to inflammation and leukocyte function differentially expressed or spliced in skin and dozens of genes differentially spliced in head kidney, the main immune organ in fish.

4.1 The *Eda* haplotype affects gene expression and splicing in mostly different genes

To identify the most important mediators of the phenotypic effects of the *Eda* haplotype, we only considered genes with a more than 1.5-fold difference in expression level between genotypes. That we found hundreds of DEGs in skin is a confirmation of the strong effect of this relatively small region of the genome (16 kb out of 450 Mb). Amongst these DEGs was *Eda* itself, which previously was found to have differences in expression levels between the C and L allele due to reduced responsiveness to *Wnt* signalling of the L allele (O'Brown et al., 2015). However, *Tnfsf13b* and *Garp,* the other two genes in the haplotype were not differentially expressed or spliced. The numbers of DEGs across the different genotype comparisons (CC vs LL, CC vs CL, and CL vs LL) mirror the phenotypic differences of these genotypes: there are no DEGs between the two completely-plated genotypes (CC and CL) but overlapping and similar number of DEGs in the comparisons of the low-plated genotype with the other two (CC vs LL and CL vs LL)(Figure 3a). This is consistent with the hypothesis that the DEGs we identified are involved in mediating the phenotypic effects of the *Eda* haplotype in skin.

To determine whether other regulatory mechanisms besides gene expression could be important in mediating the effects of *Eda*, we asked whether the *Eda* haplotype has an effect in alternative splicing, a regulatory mechanism that has recently been linked to phenotypic evolution and adaptation (Bush et al., 2017; L. Chen et al., 2012; Singh & Ahi, 2022; Verta & Jacobs, 2022; Wright et al., 2022). Though more limited than the effect on gene expression, we found that the *Eda* haplotype also affected alternative splicing of dozens of genes in skin. This is likely to be an underestimation of the number of DSGs, since the method that we used to identify DSGs, differential exon usage, is a conservative method that only detects changes in splicing involving complete exons. However differential exon usage still accounts for roughly half of the splicing events in human (Chaudhary et al., 2019) and has the greatest potential for mediating modular changes in the protein function. We found that differential expression and differential splicing affect mostly different sets of genes, with only six genes

being both differentially expressed and differentially spliced in the *Eda* CC vs LL comparison (Figure 3c). While a study comparing sympatric ecotypes of artic charr also found limited overlap between DEGs and DSGs (Jacobs & Elmer, 2021), a study comparing male and female transcriptomes of several bird species found almost half of the DSGs were also DEGs (Rogers et al., 2021) and a study in seasonal morphs of the *Bicyclus* butterfly found more than half of the DSGs were also DEGs (Steward et al., 2022).

In our study, DSGs and DEGs are also not found in the same gene co-expression networks. While most DEGs are found together on module M5, most DSGs are not part of any co-expression module (Table 1). This could be explained by the nature of the gene co-expression analysis, which clusters genes with similar expression profiles, something that DEGs will tend to share, and the fact that it is a gene-level analysis, so if a DSG has isoforms with different co-expression profiles, they will be missed in the gene co-expression analysis. This is supported by our results in the exon-level co-expression of *Rmnd5b,* where we find that the differentially spliced exon 1 is co-expressed with module M5 (Supplementary Table S6). However, despite these limitations of the gene co-expression analysis, 13 DSGs are found in a gene co-expression module, and 11 of those are found in modules that also include DEGs. These include five of the six DSGs that are also DEGs, which are found in module M5 together with most DEGs. So, while differential expression and differential splicing caused by the *Eda* haplotype tends to affect different groups of genes, some of the DEGs and DSGs are part of the same co-expression networks and so might be working together to mediate the phenotypic effects of the haplotype.

4.2 Downstream effects of the *Eda* haplotype in skin includes conserved and pleiotropic molecular pathways that are strong candidates to mediate the effects of *Eda* in skin

When comparing the transcriptomes of *Eda* CC and *Eda* LL individuals we find that genes with functions related to skeletal development (e.g skeletal system development, ossification, odontogenesis, calcium ion binding) are more often differentially expressed than we would expect by chance (Figure 5 and Supplementary Table S3). This result is consistent with the fact that these two genotypes underlie the two distinct lateral plate phenotypes in threespine sticklebacks (completely-plated versus low-plated) and makes these DEGs strong candidates to be mediators of the effects of the haplotype in the bony lateral plates. One excellent candidate is the bone morphogenetic protein (BMP) pathway, which is a conserved pathway in animals that was first discovered for its role in bone formation (Salazar et al., 2016; R. N. Wang et al., 2014). However, this pathway is now known to have pleiotropic effects

on tissue homeostasis, embryogenesis, and development, including the development of ectodermal appendages (Cui & Schlessinger, 2006; Sadier et al., 2014; R. N. Wang et al., 2014). The BMP pathway and the *Eda* pathway have been found to regulate each other in mice (Sadier et al., 2014). In stickleback, a cis-regulatory mutation in *Bmp6* is associated with increased number of pharyngeal teeth in benthic populations (Cleves et al., 2014), while loss-of-function mutations in *Eda* result in loss of pharyngeal teeth (Wucherpfennig et al., 2019). Five of the skin DEGs we found are members of the BMP family, namely *Bmp2a*, *Bmp4, Bmp5, Bmp7a and Bmp8a*. *Bmp4* is particularly interesting because it has previously been connected to adaptive phenotypic changes in beak size in Darwin's finches (Abzhanov et al., 2004). Taken together, these results suggest that the BMP pathway is a strong candidate to be a mediator of the effect of *Eda* on the bony lateral plates.

However, the BMP pathway was not the only pathway present in the skin DEGs. Genes from the Hedgehog pathway, which also plays a role in the development of ectodermal appendages (Sadier et al., 2014), were also present more often than expected by chance among the DEGs (Figure 5, Supplementary Table S3). This includes the Indian hedgehog molecule a (*Ihha*), which has been shown to regulate BMP expression and bone formation (Rahman et al., 2015). Furthermore, though not statistically overrepresented in GO terms, we also find DEGs from other important signalling pathways like Wnt (*Lef1* and *Dkk1a*), Fgf (*Fgf13b* and *Fgfr4*) and Notch (*Dld and Egfl6*) (Supplementary Table S3). Wnt and Fgf are known to also mediate the development of ectodermal appendages, including scale development in zebrafish (Aman et al., 2018; Cui & Schlessinger, 2006; Sadier et al., 2014). Furthermore, lower responsiveness to Wnt signalling has been connected to lower expression level of the freshwater *Eda* L allele (O'Brown et al., 2015). Among the DEGs from the Wnt pathway, we found the gene *Lef1*, a transcription factor which mediates Wnt activation of *Eda* expression in human cells (Durmowicz et al., 2002), and *Dkk1a* a Wnt antagonist (Supplementary Table S2). The presence of a Wnt activator of *Eda* and an antagonist of Wnt among the skin DEGs suggests that there might be a negative feedback interaction between the *Eda* pathway and the Wnt pathway during the development of lateral plates in threespine stickleback, as found in mouse hair and in zebrafish scales (Aman et al., 2018; Cui & Schlessinger, 2006).

These signaling pathways are highly pleiotropic and are involved in much more than just bone and ectodermal appendage development. For example, the Wnt, Fgf and the Notch pathways are also involved in the development and patterning of the lateral line in zebrafish (Dalle Nogare & Chitnis, 2017; Kniss et al., 2016). The two DEGs from the Wnt pathway mentioned previously, *Lef1* and *Dkk1a,* are also involved in neuromast development (Supplementary Table S4). Consistent with this, we found

an enrichment of DEGs related to neuronal processes, namely including netrin activity and genes related to “neuromuscular process controlling balance” (Figure 5). Netrins are a conserved family of diffusible proteins with chemotaxis characteristics, which are involved in axon guidance in the central nervous system (CNS). The enrichment of the “neuromuscular process controlling balance” annotation was driven by three genes: cadherin-related 23 (*cdh23*), otoferlin b (*otofb*), and calcium channel, voltage-dependent, L type, alpha 1D subunit a (*cacna1da*) (Supplementary Table S3). These three genes are also annotated as being involved in sound perception, and *cdh23* is also annotated as being involved in neuromast hair cell morphogenesis. These genes are interesting because the mammalian auditory and vestibular systems (the latter responsible for the sense of balance) and the fish lateral line system are all based on the use of hair cells to detect changes in balance, air, and water pressure, respectively (Mogdans, 2019; Roberts et al., 1988). As most functional annotations in stickleback are semi-automatically imported from model organisms (including mammals like mouse or human) (Gaudet et al., 2011), it is not surprising that some of the genes involved in lateral line development would be annotated as involved in balance and sound perception. Interestingly, gene co-expression module M27, which has 12 DEGs, is also enriched in genes with the “neuromuscular process controlling balance” GO Term. This includes one DEG, *otofb* which affects the development of neuromast hair cells in zebrafish (Manchanda et al., 2021) and regulates the release of neurotransmitters in hair cells in humans (Roux et al., 2006; Yasunaga et al., 1999). Besides *otofb*, module M27 also includes five other genes that are annotated as being involved in lateral line development, the auditory system, and/or the vestibular system (Supplementary Table S4). Taken together, these results suggest that M27 might represent a network of genes involved in mediating the effect of the *Eda* haplotype on lateral line patterning. In addition, some of the genes in module M5, such as the DEGs connected to the netrin pathway, auditory system, balance and neuromast development, probably also contribute to this phenotype. This module also contains the BMP pathway, which is also important for the development and patterning of the central and peripheral nervous system (Gámez et al., 2013). For example, *Bmp4* limited the number of sensory neurons and the extent of terminal peripheral nerve innervation in mouse skin (Guha et al., 2004). Interestingly, the topology of the gene co-expression network suggests that the influence of *Eda* on module M27 is mediated through module M5 (Figure 6). Taken together, these results suggest several DEGs in module M27 and M5 that are strong candidates for mediating the phenotypic effects of the *Eda* haplotype on the patterning of the lateral line.

Regarding the DSGs, we also found genes related to bone and neuronal development among the 34 DSGs between *Eda* LL and CC in skin. We did not find any significant enrichment of GO terms in the

DSGs, but the GO enrichment analysis does not have much power with the relatively small number of DSGs. Thus, we looked at the GO Terms present in the DSGs and found three with annotations related to neuronal development (*Zc4h2, Cln3*, and *Anks1b*), two genes related to cartilage development (*Runx2 and col11a2*), and an uncharacterized gene on chromosome IV (*ENSGACG00000017917*), predicted by Uniprot to have cadherin domains and be involved in cell adhesion and calcium binding, both processes connected to bone development. Furthermore, while most DSGs were not found in any co-expression module (Table 1), six out of 34 are found in co-expression module M5, together with most DEGs. This includes the two cartilage related genes, *Runx2* and *col11a2,* which are also among the six genes that are both differentially spliced and differentially expressed. Of these DSGs, *Runx2,* which codes a transcription factor protein, is of particular interest. Not only is it an essential gene for osteoblast differentiation, but it is also acts as an important integrator of the interaction between the BMP pathway and other major signalling pathways like Hedgehog and Wnt (Rahman et al., 2015). Considering that we find changes in expression in genes from these two pathways, it is possible that changes in the splicing and expression of *Runx2* could be partially mediating the changes in these pathways. Given that some DSGs also have functions consistent with the phenotypic effects of the *Eda* haplotype, changes in alternative splicing could be one of the mechanisms by which differences between the *Eda* haplotypes leads to changes in the lateral plate and lateral line phenotypes.

In summary our results suggest that several major developmental pathways that have been described in other systems to be involved in the development of ectodermal apendages and the lateral line, like Bmp, Wnt, Fgf and Notch, are probably also involved in mediating the phenotypic effects of the *Eda* haplotype in the lateral plates and lateral line. This suggests that that the effect of *Eda* on the lateral plate in threespine stickleback is at least in part mediated by conserved developmental pathways involved in the formation of homologous structures in other vertebrates. However, it is important to note that we only examined a single developmental timepoint after the plates had formed in CC and CL individuals, so there may be other genes or pathways that mediate the effects of the *Eda* haplotype at earlier stages of development. Elucidating the direct causal relationships between these pathways will require examination of expression at additional timepoints as well as manipulative experiments.

4.3 Possible effect of the *Eda* haplotype in immunity

The *Eda* haplotype includes two other genes, *Tnfsf13b* and *Garp*, both of which are predicted to have immune functions. However, it is not clear whether *Tnfsf13b* and *Garp* are important for freshwater adaptation or if they are simply tightly linked to the *Eda* haplotype. Although neither *Tnfsf13b* nor *Garp* is a DEG or DSG in the skin or the head kidney, these genes could still be differentially expressed or

spliced in tissues or timepoints not sampled in our study. Furthermore, there are coding changes between the C and L haplotype in both genes, which could also have phenotypic effects (Colosimo et al., 2005). Even if these two genes do not contribute to differences in immune function, *Eda* itself could still have an effect in immunity. Thus, we looked for an effect of the *Eda* haplotype on immune-related genes in two tissues important for immunity in teleost fish: the skin, one of the main physical barriers against pathogens, and the head kidney, one of the main leukocyte producing tissues (Smith et al., 2019). In skin, we found an enrichment of DEGs involved in bradykinin signalling (Figure 5), which are pro-inflammatory molecules (Kaplan et al., 2002; Marceau & Regoli, 2004), as well as scavenger receptors (Figure 5), which are a diverse family of receptors with roles in homeostasis and innate immunity, including identification and clearance of pathogens and inflammatory signalling (Alquraini & El Khoury, 2020; Canton et al., 2013). This suggests a potential for an effect of the *Eda* haplotype on innate immune responses in skin, in particular inflammation, which could be important to deal with the different pathogens in freshwater and marine environments. However, it is also important to note that the some inflammatory signalling proteins, including bradykinins, have been implicated in bone reabsorption (Epsley et al., 2021; Lerner et al., 1987). Thus, it is also possible that these inflammation-related genes are associated with homeostasis of the lateral plates rather than mediating inflammatory response differences. However, this still does not exclude the possibility than in an immune challenge scenario, the presence of different *Eda* genotypes could lead to differences in inflammatory response. Our literature research also revealed the presence of two DEGs involved in leukocyte function and/or development (*Ets1* and *Laptm4b)* (Dittmer, 2003; Garrett-Sinha, 2013), and one DSG (*Tbk1)* that is an important integrator of multiple signaling pathways related to immunity, namely pathogen detection, inflammation, and immune response (Helgason et al., 2013). Furthermore, *Laptm4b* was found to interact with *Garp,* one of the genes in the *Eda* haplotype, in mammalian cells (Huygens et al., 2015), raising the prospect that *Garp* could be mediating immune changes in skin between marine and freshwater threespine stickleback. Together, these data suggest a potential for an effect of the *Eda* haplotype on inflammation as well as other immune functions in skin.

We also found an effect, albeit small, of the *Eda* haplotype in the main immune tissue in fish, head kidney. In contrast to skin, the effect of the haplotype in head kidney was solely on splicing (Figure 3b). In the *Eda* CC vs LL comparison, we found 14 DSGs, two of which have important immune functions. The *Itgb2* (integrin beta 2) gene, also known as *Lfa-1* and *Cd177,* is important for the function and migration of T cells, neutrophils and killer cells (Bai et al., 2017; Barber et al., 2004; Kristóf et al., 2013; Ostermann et al., 2002; Walling & Kim, 2018). The *traf3* (TNF receptor associated factor 3) gene plays an important role in anti-viral innate immune response (Gao et al., 2021; Oganesyan et al., 2006; Tseng

et al., 2010) and in the regulation of B and T white cells (Lin et al., 2015; Yi et al., 2015). Interestingly, beyond immune functions, we also found two DSGs with neuronal development annotations (*Cables1* and *Nup98*)(Supplementary Table S4). It is possible that changes in the splicing of these genes could lead to changes in the innervation of head kidney between *Eda* genotypes, which could have an influence on how this organ reacts to external stimulus.

It is important to keep in mind that the individuals used in this study were healthy. Thus, it is possible that we are missing effects of the *Eda* haplotype that would only manifest during a situation of immune challenge. However, the results we find in healthy individuals already suggest that the *Eda* haplotype has the potential to influence immunity in two important immune organs, skin and head kidney. These results are consistent with a previous study that found evidence that the *Eda* haplotype affected parasite load and the expression of target immune genes in F2 individuals placed in enclosures in the wild (Robertson et al., 2017). However, due to the large blocks of linkage disequilibrium present in F2 crosses, the effect of mutations linked to the *Eda* haplotype could not be excluded in this study. Thus, although there is accumulating evidence that the *Eda* haplotype affects immunity, future follow-up work, using the crossing design like in our study, and immune challenge experiments like in Robertson et al. (2017) will be required to definitively establish whether the *Eda* haplotype affects immune phenotypes.

4.4 Differentially spliced genes are not more pleiotropic than differentially expressed genes in the skin of threespine stickleback

The effect of the *Eda* haplotype on both gene expression and alternative splicing of different genes putatively related to the same functions raises the question of why some genes are differentially expressed while others are differentially spliced. Even though this likely depends on the specifics of each individual gene, alternative splicing is a possible mechanism to avoid the functional constraints of pleiotropic genes by tinkering with the expression of different isoforms rather than expression of the entire gene. Two recent studies have provided some support for this hypothesis. Using tissue specificity as a proxy for pleiotropy, one study found DSGs to be more pleiotropic than DEGs or the rest of the transcriptome, with DEGs showing lower levels of pleiotropy than the rest of the transcriptome (Rogers et al., 2021). Using gene connectivity and number of associated GO terms as proxies for pleiotropy, the other study found that both DSGs and DEGs tend to be more pleiotropic than non-DSGs or non-DEGs, respectively (Jacobs & Elmer, 2021). In contrast to these two studies, we found that pleiotropy (measured as gene co-expression connectivity) in the stickleback skin DSGs does not differ from the rest of the transcriptome and that DEGs are more pleiotropic than both DSGs and

the rest of the transcriptome (Figure 4). These mixed results among the studies could result from the different biological contexts of the studies (genes affected by alleles of a single large effect haplotype in stickleback vs distinct freshwater ecotypes in Artic charr (Jacobs and Elmer 2021) and males and females in bird species (Rogers et al. 2021)) and/or from the use of different pleiotropy proxies (gene co-expression connectivity vs tissue specificity). In particular, gene co-expression connectivity, the proxy used in our study and by Jacobs and Elmer (2021), could be biased towards genes that are DEGs, since these genes will be highly co-expressed with each other, increasing their connectivity value. However, even if that is the case, the GO enrichment analysis suggests the DEGs are involved in pleiotropic developmental pathways, which is in line with a high connectivity value of these genes. Likewise, it is possible that gene co-expression connectivity is underestimated in DSGs since the analysis only assesses co-expression patterns at the gene level and not at the isoform level. Thus, genes with isoforms with different co-expression patterns would have a noisy co-expression signature at the gene-level. This idea is supported by the fact that when we atomized one of the skin DSGs that did not belong to any co-expression module (*Rmnd5b*) into its individual exons and input the exons as "genes" into the co-expression analysis, the first exon of Rmnd5b, which is differentially spliced between *Eda* genotypes, was part of the same co-expression network as most DEGs (Supplementary Table S6). However, despite these limitations of the connectivity proxy, Jacobs and Elmer (2021) did identify a higher gene co-expression connectivity of DSGs than non-DSGs in their study. Thus, it is possible that the differences in the results of the three studies might be related to their different biological contexts. This would suggest that connectivity by itself is not a determining factor for the use of differential splicing to mediate phenotypic differences.

Comparisons of the gene expression levels of DEGs and DSGs offer an alternative explanation. Although gene connectivity tends to increase in genes with higher expression levels (Supplementary Figure S7), DSGs tend to have higher expression levels than the average of the transcriptome or than DEGs, despite having lower connectivity. There is evidence that highly expressed genes evolve more slowly and are under stronger selective constraints, which has been suggested to be associated with the cost of transcription and/or translation (Drummond et al., 2005; Gout et al., 2010). Theoretical models also predict that highly expressed genes are more likely to be pleiotropic (Guillaume & Otto, 2012). Differential splicing could be a good mechanism to modulate the function of these highly expressed genes by changing the expression of alternative isoforms without affecting the expression level of the highly expressed isoform(s). In general, these results suggest that connectivity and expression level might be important factors in determining whether differential expression or alternative splicing is affected in genes mediating phenotypic effects. However, more studies are

needed for a more concrete understanding of whether these factors or others tend to determine the use of differential splicing and differential expression, or whether the use of these regulatory mechanisms is mainly context dependent.

5 Conclusions

Knowing the molecular mechanisms and pathways that connect adaptive genes to adaptive phenotypes is an important step towards understanding why particular genes and genetic changes might be used more often during phenotypic evolution (Stern, 2013). In this study we tackled this question by asking what genes and regulatory mechanisms are differentially affected by the marine and freshwater alleles of the *Eda* haplotype, a locus involved in lateral plate and lateral line differences between marine and freshwater sticklebacks. Our results show that the *Eda* haplotype affects hundreds of genes with different biological functions, like signalling, development, and immunity. These include conserved pathways and genes involved in bone formation and neuromast development, suggesting that the effects of the *Eda* haplotype on lateral plates and the patterning of the lateral line are mediated, at least in part, by conserved pathways. We also found that differential expression was not the only regulatory mechanisms at play, but that some genes were instead affected by changes in alternative splicing patterns. Furthermore, gene co-expression connectivity and expression levels were different between these two categories of genes, suggesting that these factors might influence the types of genetic changes that underlie adaptive phenotypic evolution.

Acknowledgements

We are grateful to Pooja Singh and Magdalena Bohutínská for very helpful discussions and feedback on the manuscript, Sophie Archambeault for setting up the *Eda* lineages used in this study, all Peichel Lab members for fruitful discussions throughout the study, and the University of Bern Next Generation Sequencing Center for library preparation and sequencing. This work was funded by the University of Bern.

Author Contributions

Carlos E. Rodríguez-Ramírez and Catherine L. Peichel conceived and designed the study; Melanie Hiltbrunner and Verena Saladin performed the experiments; Stephanie Walker contributed to the literature research of immune functions in differentially expressed genes; Carlos E. Rodríguez-Ramírez

analysed and interpreted the data with input from Araxi Urrutia and Catherine L. Peichel. Carlos E. Rodríguez-Ramírez wrote the paper with input from Catherine L. Peichel.

Supplementary Tables

Supplementary Tables for this chapter can be found in the following OneDrive Link:

https://unibe365-my.sharepoint.com/:f:/r/personal/carlos_rodriguezramirez_unibe_ch/Documents/DoctoralThesis_RodriguezRamirez_SupTables?csf=1&web=1&e=FOhHEN

Figures

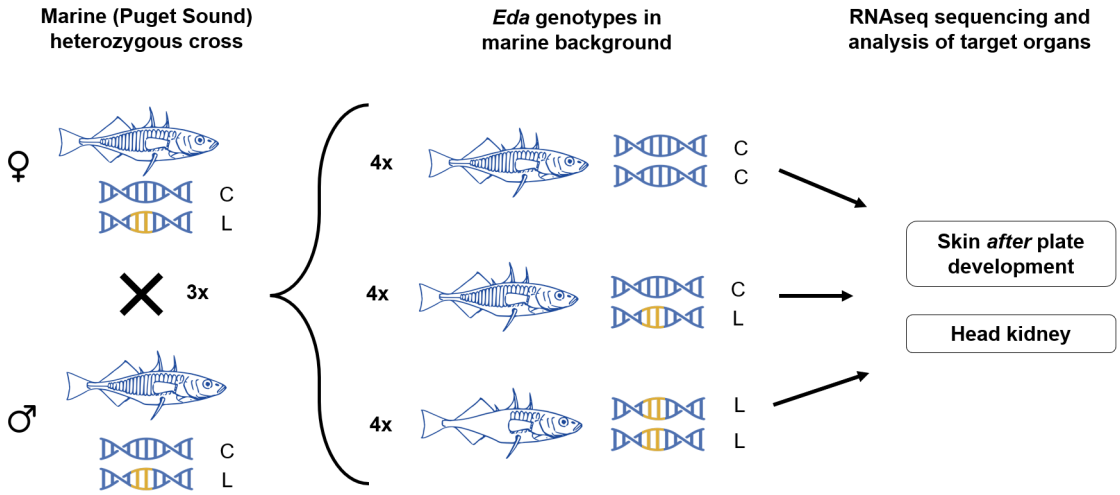


Figure 1. Experimental design. Marine sticklebacks that are heterozygous for the *Eda* haplotype have a completely plated phenotype and a marine genomic background (blue) but carry one copy of the completely plated C haplotype (blue) and one copy of the low-plated L haplotype (yellow). Crossing these individuals results in offspring with the three *Eda* genotypes (CC, CL, and LL) on the same marine genomic background. RNA from the skin and head kidney of these individuals were sequenced to test for the effect of the different *Eda* haplotypes on the transcriptome.

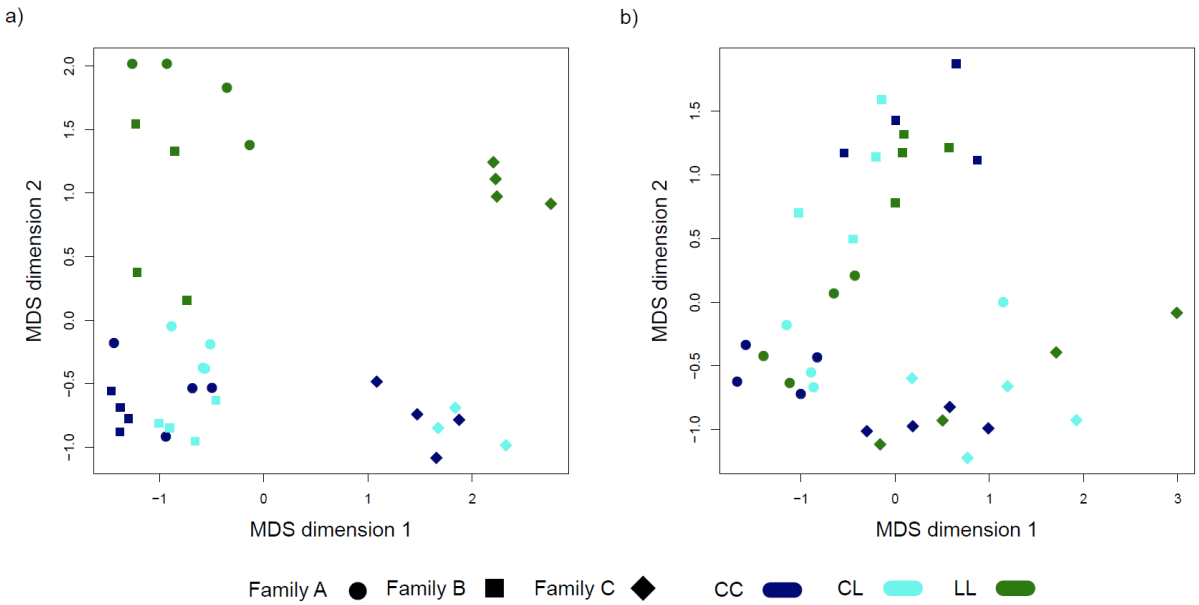


Figure 2. Samples cluster by *Eda* haplotype in skin but not head kidney. MDS plot of the pairwise distances between the gene expression profiles in a) skin samples and b) head kidney samples, based on the 500 genes with the largest pairwise changes in gene expression between each sample. Colour indicates the genotype of the samples: CC = dark blue; CL = light blue; and LL = green, and the different shapes indicate the different families. In skin the first MDS dimension separates one family (diamonds) from the other two (circles and squares). The second MDS dimension separates LL individuals from CC and CL individuals. In head kidney samples are clustered only by family.

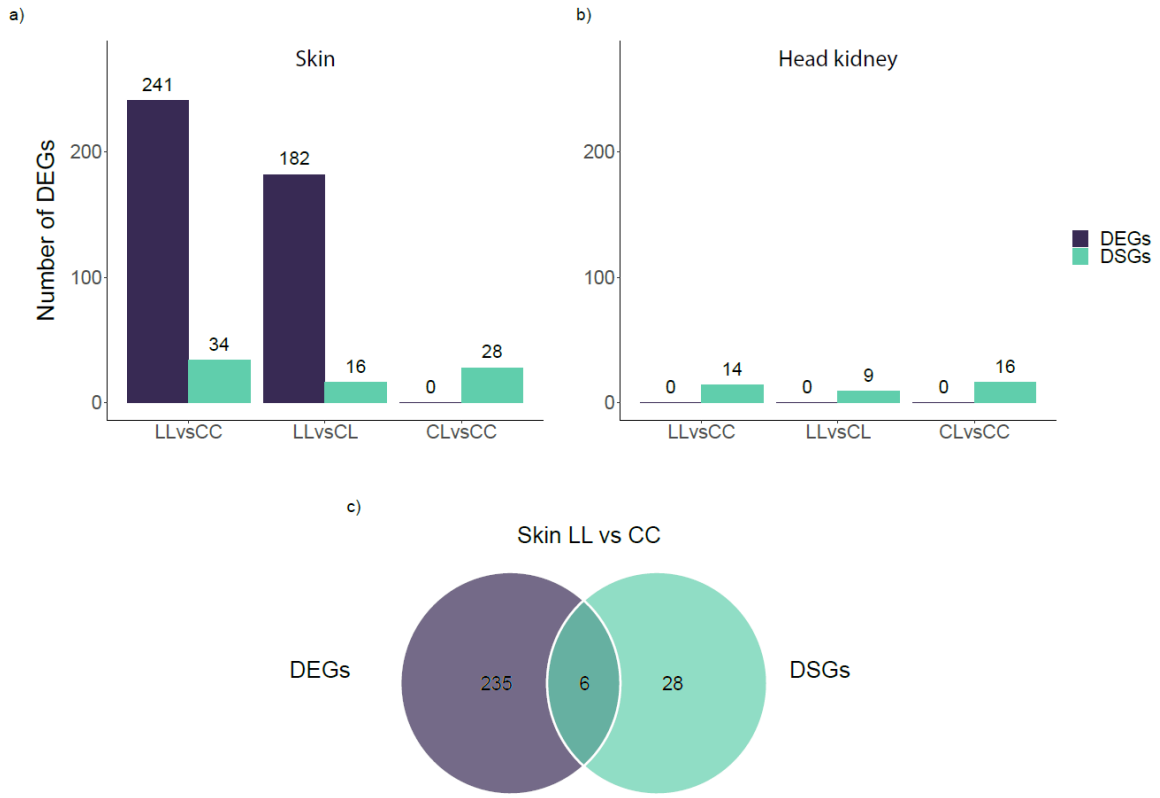


Figure 3. Differentially expressed genes (DEGs) and differentially spliced genes (DSGs) between *Eda* genotypes, in a) skin and b) head kidney. c) Venn diagram of the overlap between DEGs and DSGs in the skin CC vs LL comparison.

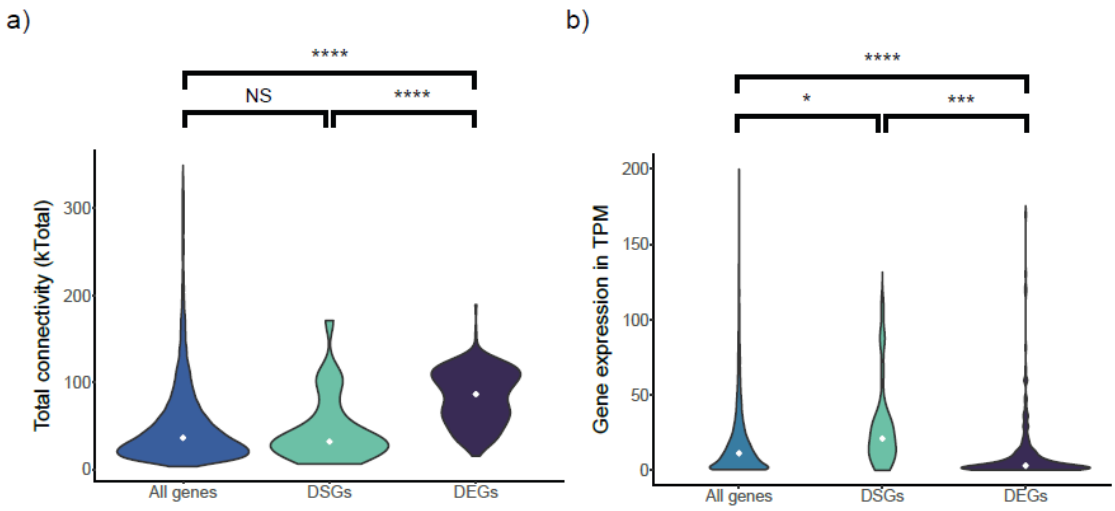


Figure 4. Skin DEGs have a greater gene co-expression connectivity than skin DSGs, but DSGs are more highly expressed. a) Violin plot showing the distribution of the values of total connectivity (kTotal) for all genes in the transcriptome, DSGs, and DEGs. b) Violin plot showing the distribution of the values of gene expression for all genes in the transcriptome, DSGs, and DEGs. Expression values are normalized in Transcripts per Million (TPM). For visual clarity, 635 outliers with an expression value over 200 TPM in the "All genes" category are not included in the plot. The six genes that were both DSGs and DEGs were not included in either analysis. In both plots the white diamond in the middle represents the median of the distribution, and the results of permutation tests for each pairwise comparison are shown with asterisks (*$p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$) or NS (non-significant).

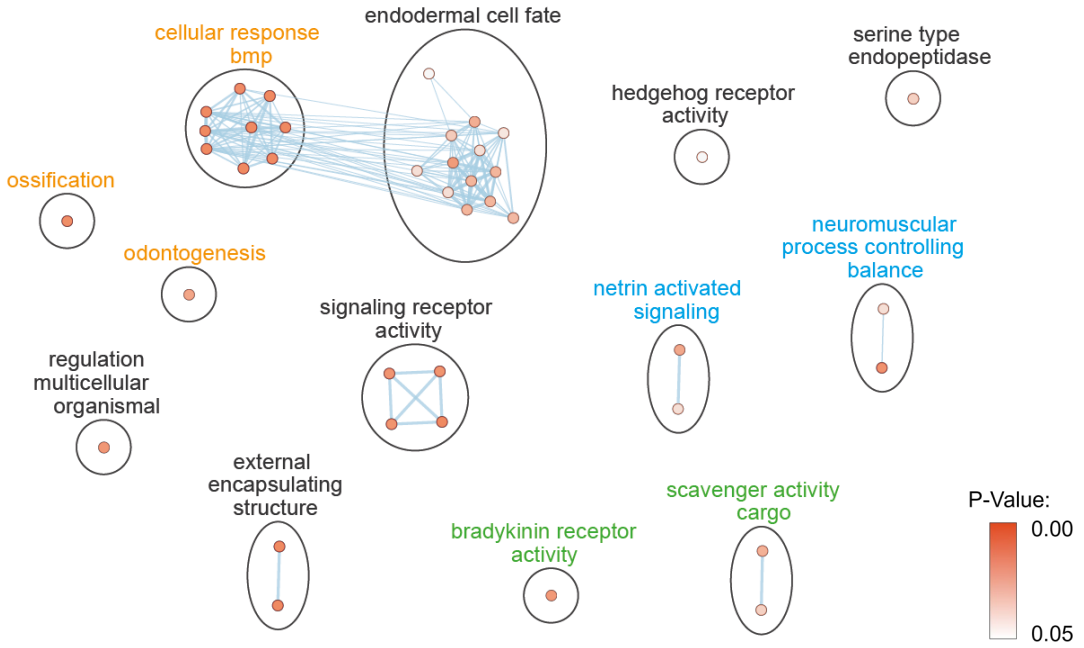


Figure 5. Summary networks of enriched GO terms in DEGs between *Eda* CC and LL individuals in skin. Nodes represent individual GO terms that were found to be significantly enriched in the 241 DEGs. Color of the nodes represents the P-value for the GO term. Blue lines represent gene similarity between GO terms. Black circles represent clusters of highly overlaping GO terms. Clusters labeled in orange have annotations related to bone development, clusters labeled in green have annotations related to immunity, clusters labeled in blue have annotations related to neuronal processes, and clusters labeled in black have general annotations. Labels of clusters of annotations are based on WordClouds of the GO terms present inside the clusters. For clarity, general GO Terms present in more than 150 genes in the genome were not included in the network. For the full list of enriched GO Terms see Supplementary Table S3.

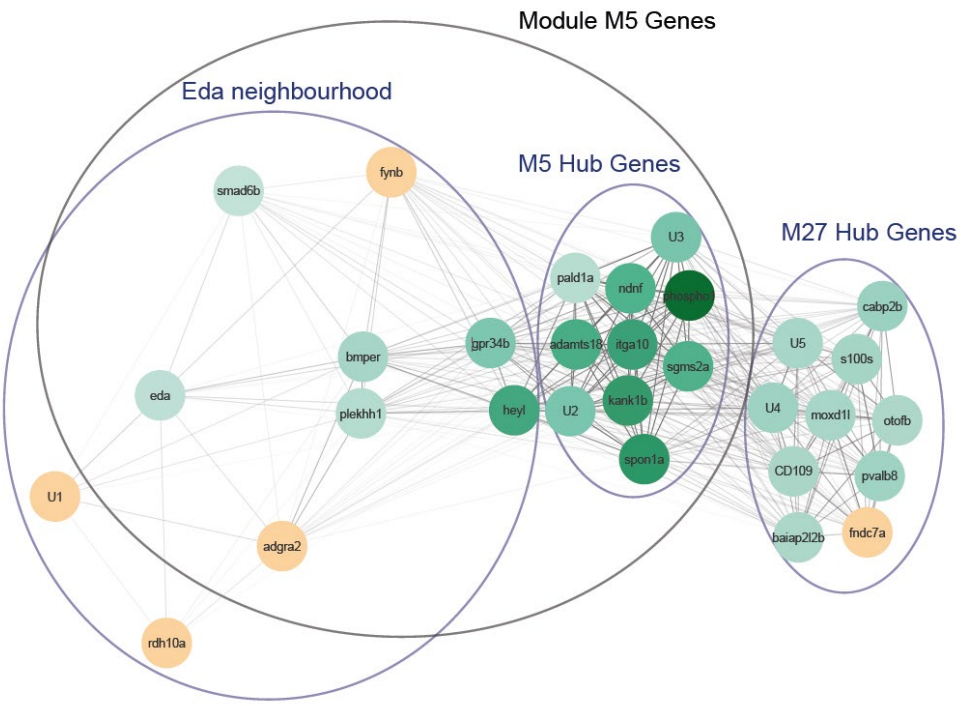


Figure 6. Gene co-expression network of the top ten genes co-expressed with *Eda* ("Eda neighbourhood") and the ten genes with the highest total connectivity in each of modules M5 and M27. Green circles indicate *Eda* LL vs CC DEGs, with darker shades indicating higher fold-change between LL and CC individuals. Yellow circles indicate genes that are not significantly differentially expressed. Lines indicate co-expression strength, with shorter and darker lines indicating stronger gene co-expression between two genes. Gene *Rdh10a*, on the bottom left of the figure was manually brought closer to the rest of the genes for visualization purposes. Nodes with U1 to U5 labels are uncharacterised genes. Their Ensembl IDs are: *ENSGACG00000017847* (U1), *ENSGACG00000016062* (U2), *ENSGACG00000008364* (U3), *ENSGACG00000015150* (U4) and *ENSGACG00000015153* (U5).

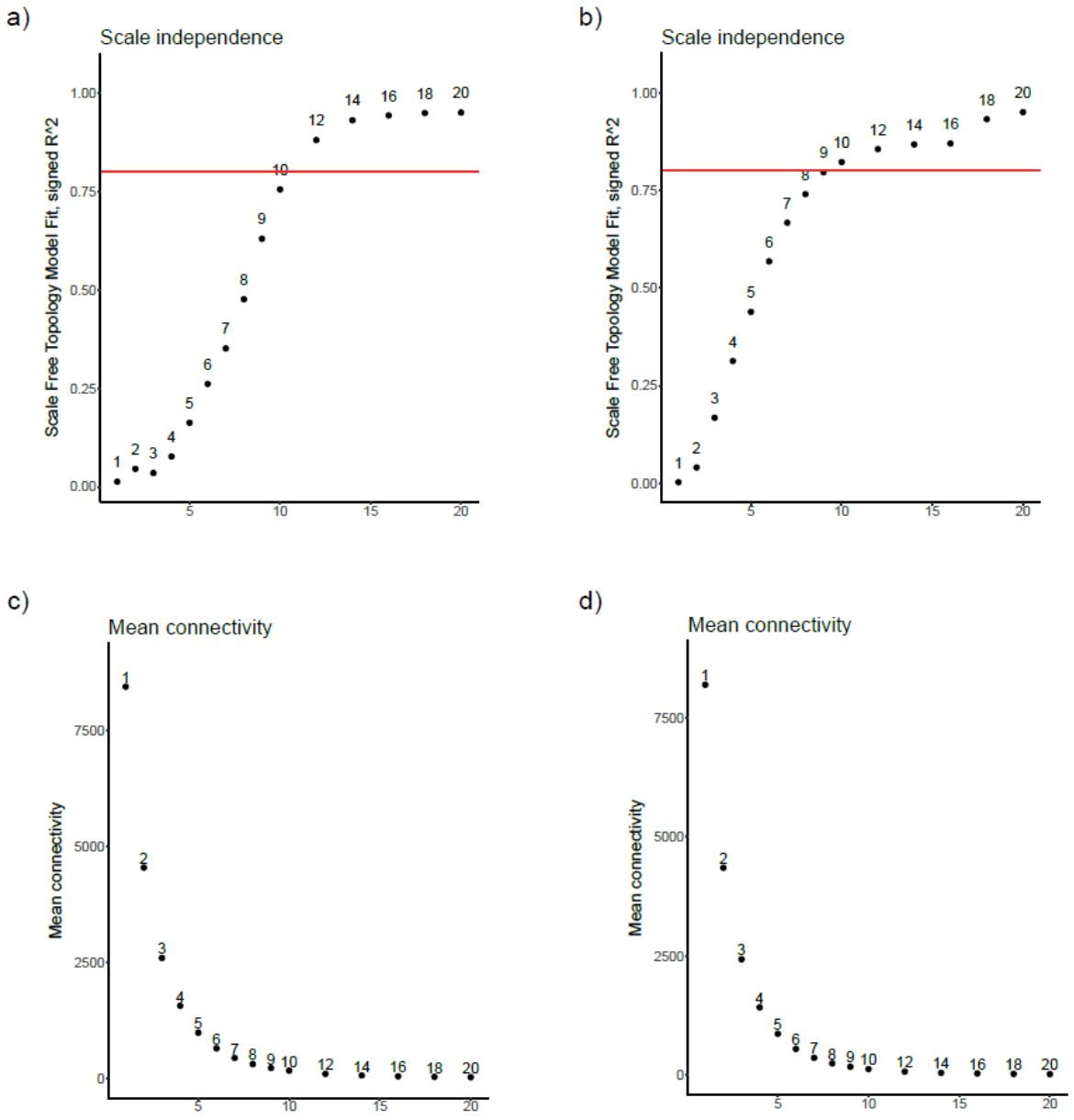


Figure S1. Scale-free topology model fits for **a)** skin and **b)** head kidney, and mean connectivity for **c)** skin and **d)** head kidney for different values of beta for co-expression analysis with WGCNA.

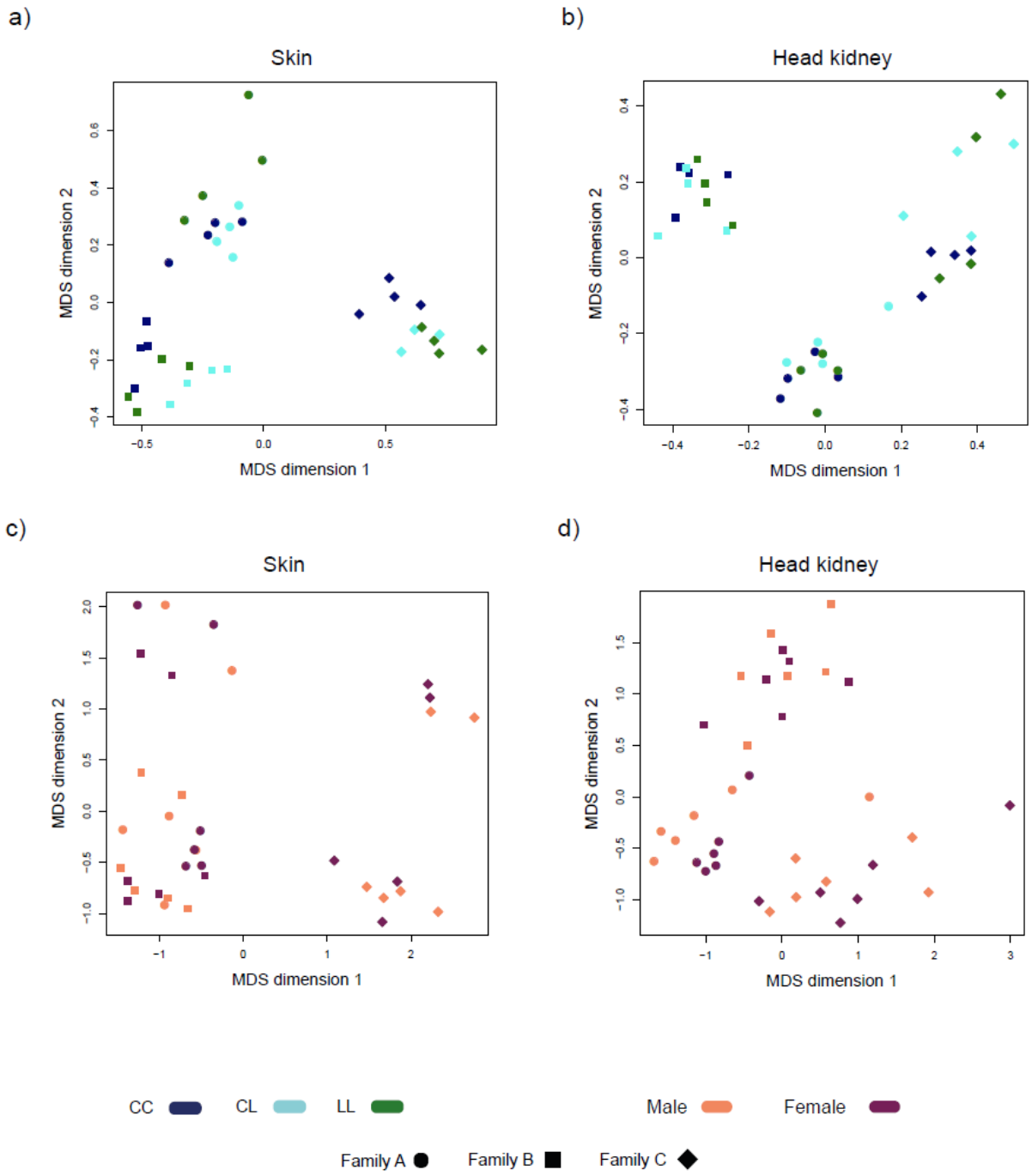


Figure S2. MDS plot of the pairwise distances between the gene expression profiles in skin and head kidney samples. MDS plots include all genes in the **a)** skin and **b)** head kidney transcriptomes or the top pairwise 500 genes with greatest expression fold-changes between each pair of samples in the **c)** skin and **d)** head kidney transcriptomes.

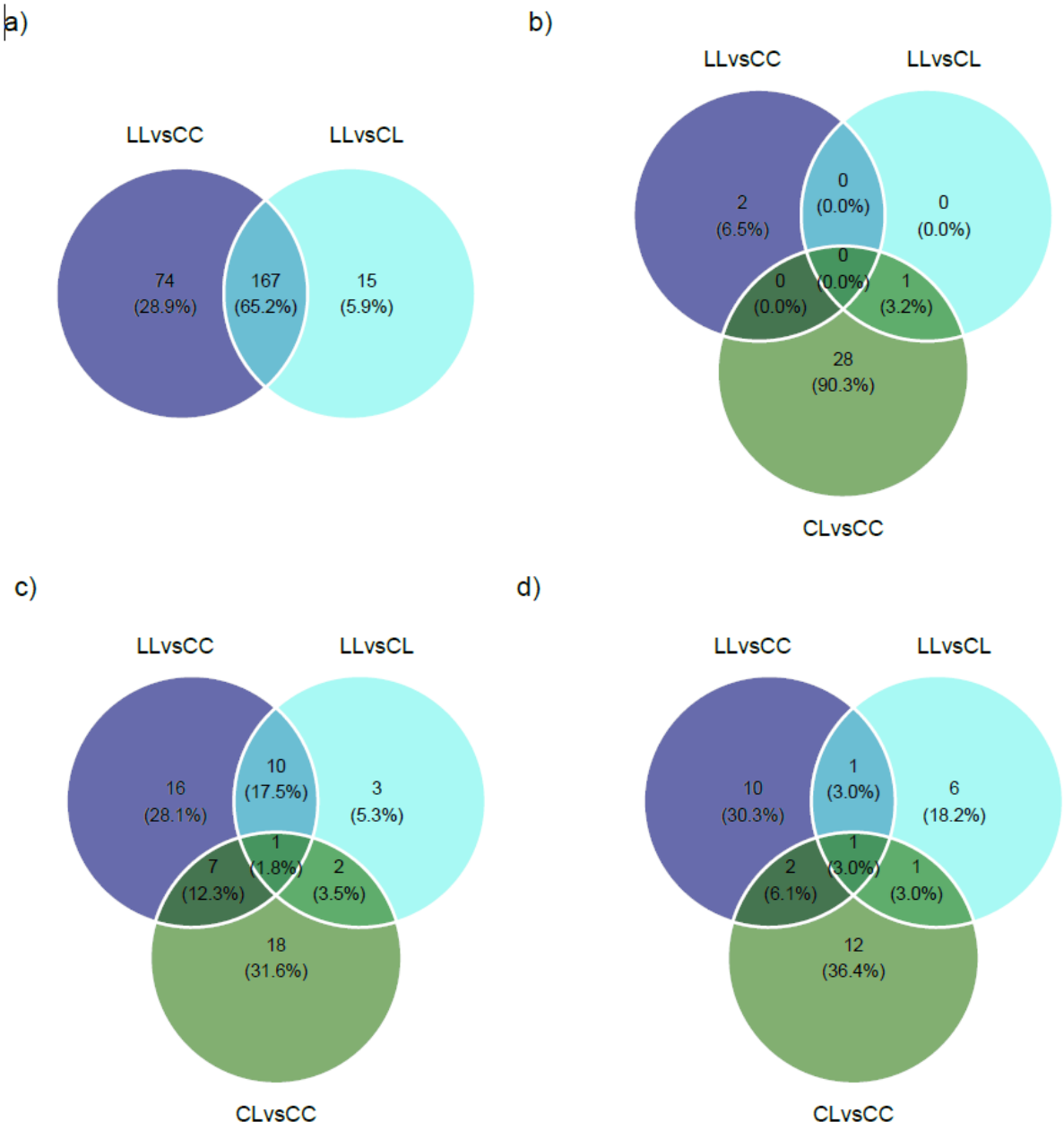
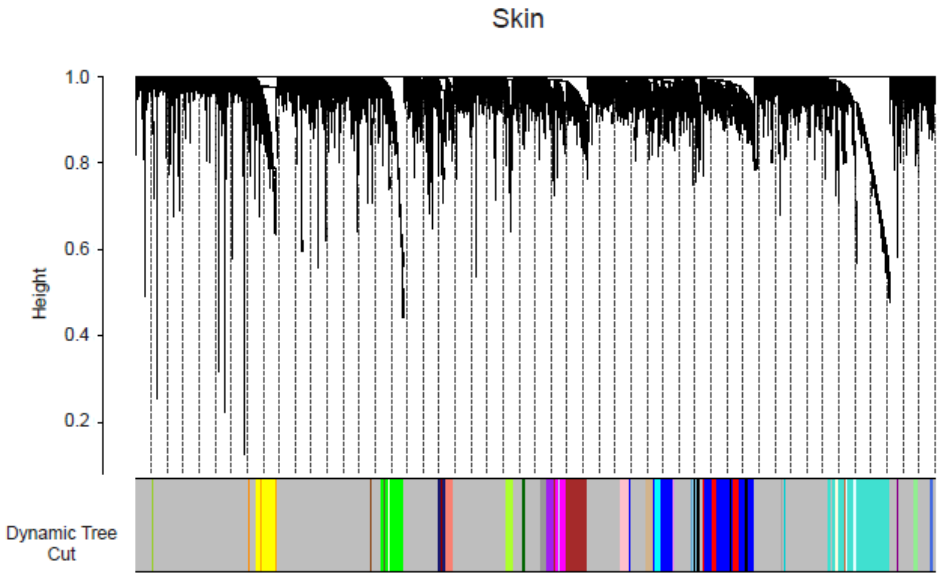


Figure S3. Overlap of genes identified in the different *Eda* genotype pairwise comparisons. **a)** DEGs in skin, **b)** DEGs in head kidney, **c)** DSGs in skin, and **d)** DSGs in head kidney.

a)



b)

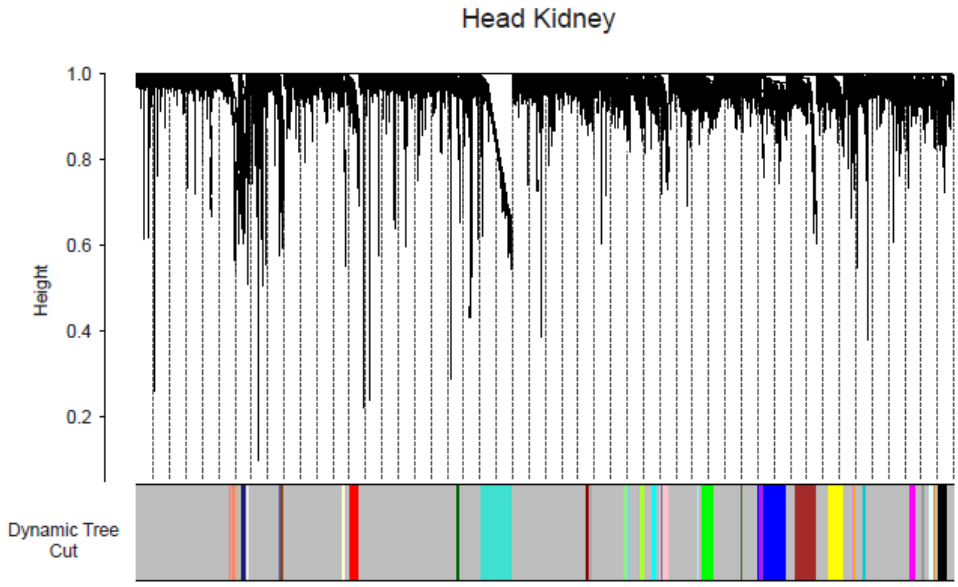


Figure S4. Topology Overlap Matrix (TOM) dissimilarity dendrograms calculated by WGCNA for beta = 14, for **a)** skin and **b)** head kidney. Each vertical line on the dendrogram represents a gene. Dynamic Tree Cut graph underneath represents identified modules. See Supplementary Table S8 for module ID and colour guide.

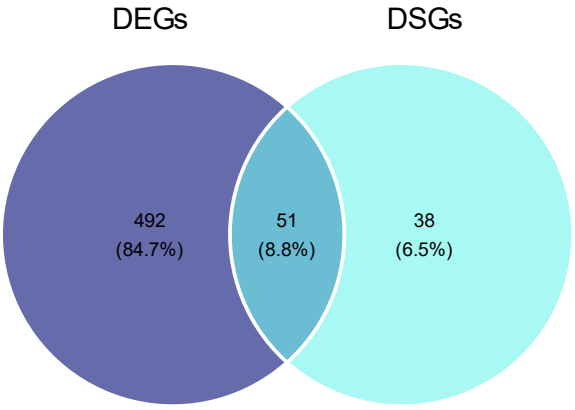


Figure S5. Overlap between the Gene Ontology (GO) Terms between DEGs and DSGs of the skin *Eda* LL vs CC comparison.

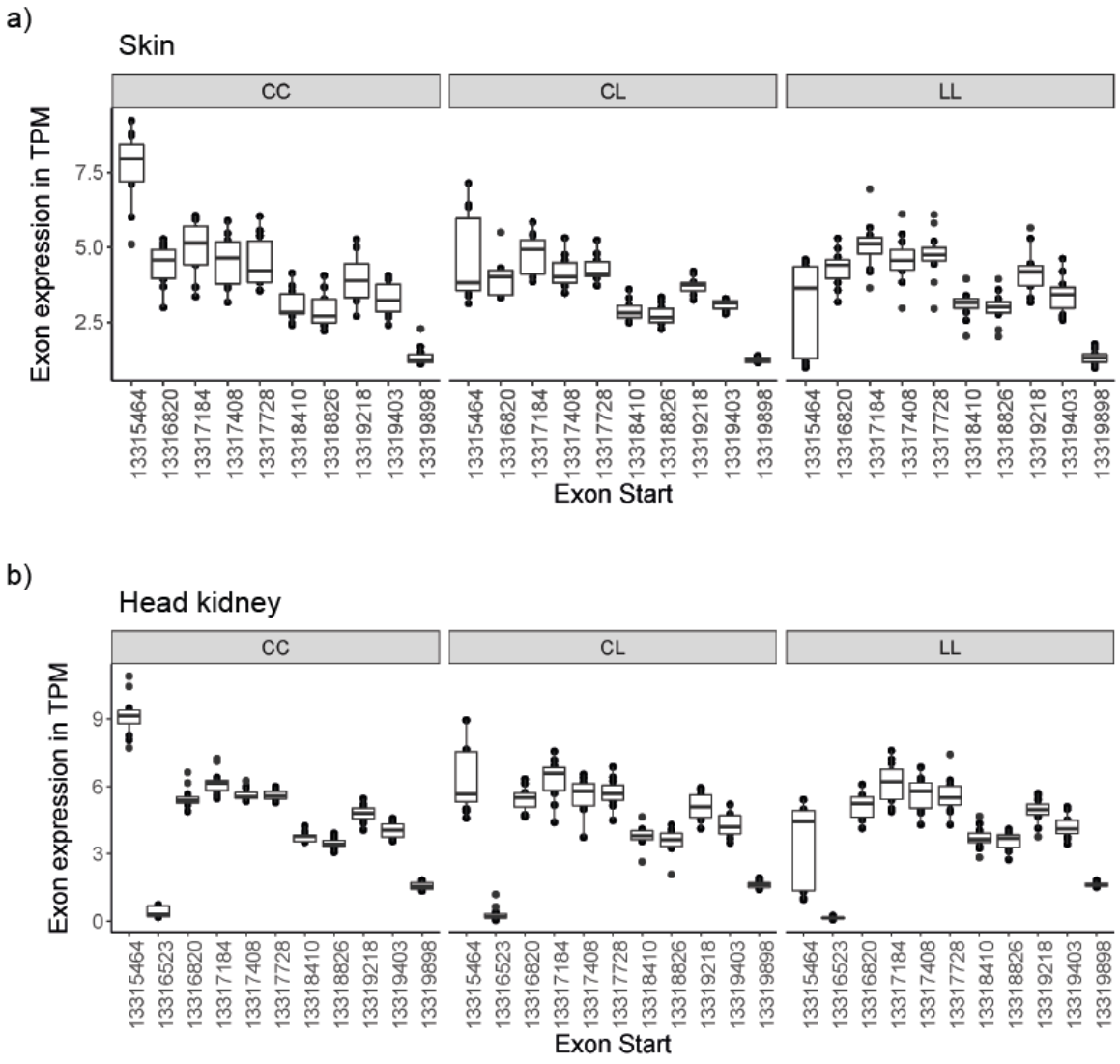


Figure S6. Exon expression plots for *Rmnd5b* in **a)** skin and **b)** head kidney of CC, CL and LL individuals. *Rmnd5b* first exon (starting at position 13315464 of chromosome IV), is differentially used between genotypes. The second exon (starting at position 13316523) is not represented in the skin exon plots because it was too lowly expressed to be included in the analysis.

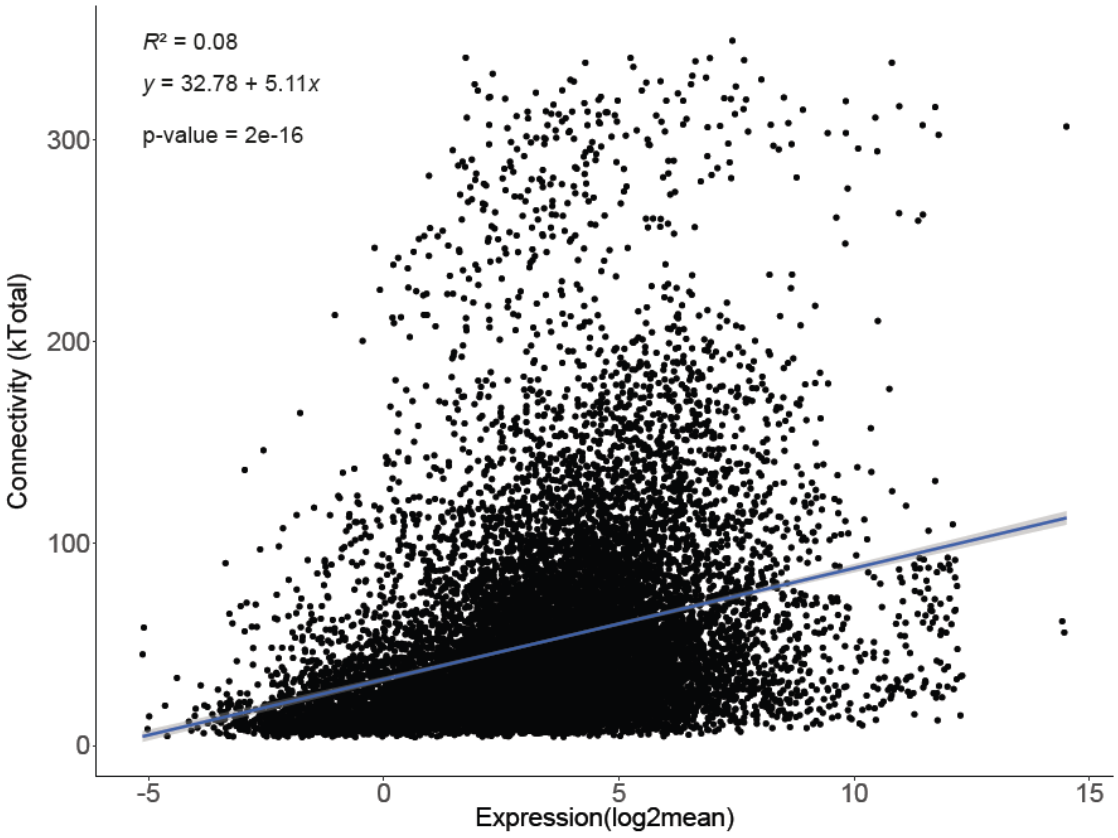


Figure S7. Scatterplot of the positive relationship between connectivity and gene expression across the skin transcriptome.

Tables

Table 1. Distribution of the skin DEGs and DSGs across M0, which represents genes not belonging to any co-expression module, and the seven out of 37 co-expression modules containing at least one of these categories of genes in the skin transcriptome. Details of all modules are provided Supplementary Figure S4 and Supplementary Table S8.

Module	M0	M2	M5	M6	M7	M10	M27	M32
Only DSGs	20	3	1	1	1	1	0	1
Only DEGs	6	2	210	2	3	0	12	0
DEG and DSG	1	0	5	0	0	0	0	0
Total genes	10791	1009	385	267	234	151	39	29

References

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., & Tabin, C. J. (2004). *Bmp4* and morphological variation of beaks in Darwin's finches. *Science (New York, N.Y.)*, *305*(5689), 1462–1465. https://doi.org/10.1126/science.1098095

Alquraini, A., & El Khoury, J. (2020). Scavenger receptors. *Current Biology*, *30*(14), R790–R795. https://doi.org/10.1016/j.cub.2020.05.051

Aman, A. J., Fulbright, A. N., & Parichy, D. M. (2018). Wnt/β-catenin regulates an ancient signaling network during zebrafish scale development. *ELife*, *7*, e37001. https://doi.org/10.7554/eLife.37001

Archambeault, S. L., Bärtschi, L. R., Merminod, A. D., & Peichel, C. L. (2020). Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback *Eda* locus. *Evolution Letters*, *4*(4), 282–301. https://doi.org/10.1002/evl3.175

Archambeault, S. L., Durston, D. J., Wan, A., El-Sabaawi, R. W., Matthews, B., & Peichel, C. L. (2020). Phosphorus limitation does not drive loss of bony lateral plates in freshwater stickleback (*Gasterosteus aculeatus*). *Evolution; International Journal of Organic Evolution*, *74*(9), 2088–2104. https://doi.org/10.1111/evo.14044

Bai, M., Grieshaber-Bouyer, R., Wang, J., Schmider, A. B., Wilson, Z. S., Zeng, L., Halyabar, O., Godin, M. D., Nguyen, H. N., Levescot, A., Cunin, P., Lefort, C. T., Soberman, R. J., & Nigrovic, P. A. (2017). CD177 modulates human neutrophil migration through activation-mediated integrin and chemoreceptor regulation. *Blood*, *130*(19), 2092–2100. https://doi.org/10.1182/blood-2017-03-768507

Barber, D. F., Faure, M., & Long, E. O. (2004). LFA-1 contributes an early signal for NK cell cytotoxicity. *Journal of Immunology (Baltimore, Md.: 1950)*, *173*(6), 3653–3659. https://doi.org/10.4049/jimmunol.173.6.3653

Barrett, R. D. H., Rogers, S. M., & Schluter, D. (2008). Natural selection on a major armor gene in threespine stickleback. *Science (New York, N.Y.)*, *322*(5899), 255–257. https://doi.org/10.1126/science.1159978

Bell, M. A., & Aguirre, W. E. (2013). Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evolutionary Ecology Research*, *15*(4), 377–411.

Bell, M. A., Aguirre, W. E., & Buck, N. J. (2004). Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution; International Journal of Organic Evolution*, *58*(4), 814–824. https://doi.org/10.1111/j.0014-3820.2004.tb00414.x

Bell, M. A., & Foster, S. A. (1994). *The evolutionary biology of the threespine stickleback*. Oxford University Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. JSTOR.

Bomblies, K., & Peichel, C. L. (2022). Genetics of adaptation. *Proceedings of the National Academy of Sciences*, *119*(30), e2122152119. https://doi.org/10.1073/pnas.2122152119

Bush, S. J., Chen, L., Tovar-Corona, J. M., & Urrutia, A. O. (2017). Alternative splicing and the evolution of phenotypic novelty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1713), 20150474. https://doi.org/10.1098/rstb.2015.0474

Canton, J., Neculai, D., & Grinstein, S. (2013). Scavenger receptors in homeostasis and immunity. *Nature Reviews. Immunology*, *13*(9), 621–634. https://doi.org/10.1038/nri3515

Chaudhary, S., Khokhar, W., Jabre, I., Reddy, A. S. N., Byrne, L. J., Wilson, C. M., & Syed, N. H. (2019). Alternative splicing and protein diversity: plants versus animals. *Frontiers in Plant Science*, *10*, 708. https://doi.org/10.3389/fpls.2019.00708

Chen, L., Tovar-Corona, J. M., & Urrutia, A. O. (2012). Alternative splicing: a potential source of functional innovation in the eukaryotic genome. *International Journal of Evolutionary Biology*, *2012*, 1–10. https://doi.org/10.1155/2012/596274

Chen, Y., McCarthy, D., Ritchie, M., Robinson, M., & Smyth, G. (2008). *EdgeR: differential analysis of sequence read count data User's Guide*.

Cleves, P. A., Ellis, N. A., Jimenez, M. T., Nunez, S. M., Schluter, D., Kingsley, D. M., & Miller, C. T. (2014). Evolved tooth gain in sticklebacks is associated with a *cis* -regulatory allele of *Bmp6*. *Proceedings of the National Academy of Sciences*, *111*(38), 13912–13917. https://doi.org/10.1073/pnas.1407567111

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*, *307*(5717), 1928–1933. https://doi.org/10.1126/science.1107239

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biology*, *2*(5), e109. https://doi.org/10.1371/journal.pbio.0020109

Cui, C.-Y., & Schlessinger, D. (2006). EDA signaling and skin appendage development. *Cell Cycle (Georgetown, Tex.)*, *5*(21), 2477–2483. https://doi.org/10.4161/cc.5.21.3403

Dalle Nogare, D., & Chitnis, A. B. (2017). A framework for understanding morphogenesis and migration of the zebrafish posterior lateral line primordium. *Mechanisms of Development*, *148*, 69–78. https://doi.org/10.1016/j.mod.2017.04.005

Dittmer, J. (2003). The biology of the *Ets1* proto-oncogene. *Molecular Cancer*, *2*, 29. https://doi.org/10.1186/1476-4598-2-29

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., & Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences*, *102*(40), 14338–14343. https://doi.org/10.1073/pnas.0504070102

Durmowicz, M. C., Cui, C. Y., & Schlessinger, D. (2002). The EDA gene is a target of, but does not regulate Wnt signaling. *Gene*, *285*(1–2), 203–211. https://doi.org/10.1016/s0378-1119(02)00407-9

Epsley, S., Tadros, S., Farid, A., Kargilis, D., Mehta, S., & Rajapakse, C. S. (2021). The effect of inflammation on bone. *Frontiers in Physiology*, *11*, 511799. https://doi.org/10.3389/fphys.2020.511799

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Featherstone, D. E., & Broadie, K. (2002). Wrestling with pleiotropy: Genomic and topological analysis of the yeast gene expression network. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *24*(3), 267–274. https://doi.org/10.1002/bies.10054

Gámez, B., Rodriguez-Carballo, E., & Ventura, F. (2013). BMP signaling in telencephalic neural cell specification and maturation. *Frontiers in Cellular Neuroscience*, *7*. https://doi.org/10.3389/fncel.2013.00087

Gao, P., Ma, X., Yuan, M., Yi, Y., Liu, G., Wen, M., Jiang, W., Ji, R., Zhu, L., Tang, Z., Yu, Q., Xu, J., Yang, R., Xia, S., Yang, M., Pan, J., Yuan, H., & An, H. (2021). E3 ligase Nedd4l promotes antiviral innate immunity by catalyzing K29-linked cysteine ubiquitination of TRAF3. *Nature Communications*, *12*(1), 1194. https://doi.org/10.1038/s41467-021-21456-1

Garrett-Sinha, L. A. (2013). Review of *Ets1* structure, function, and roles in immunity. *Cellular and Molecular Life Sciences: CMLS*, *70*(18), 3375–3390. https://doi.org/10.1007/s00018-012-1243-7

Gaudet, P., Livstone, M. S., Lewis, S. E., & Thomas, P. D. (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Briefings in Bioinformatics*, *12*(5), 449–462. https://doi.org/10.1093/bib/bbr042

Gelmond, O., von Hippel, F. A., & Christy, M. S. (2009). Rapid ecological speciation in three-spined stickleback *Gasterosteus aculeatus* from Middleton Island, Alaska: The roles of selection and geographic isolation. *Journal of Fish Biology*, *75*(8), 2037–2051. https://doi.org/10.1111/j.1095-8649.2009.02417.x

Gout, J.-F., Kahn, D., Duret, L., & Paramecium Post-Genomics Consortium. (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genetics*, *6*(5), e1000944. https://doi.org/10.1371/journal.pgen.1000944

Greenwood, A. K., Mills, M. G., Wark, A. R., Archambeault, S. L., & Peichel, C. L. (2016). Evolution of schooling behavior in threespine sticklebacks Is shaped by the *Eda* gene. *Genetics*, *203*(2), 677–681. https://doi.org/10.1534/genetics.116.188342

Greenwood, A. K., Wark, A. R., Yoshida, K., & Peichel, C. L. (2013). Genetic and neural modularity underlie the evolution of schooling behavior in threespine sticklebacks. *Current Biology: CB*, *23*(19), 1884–1888. https://doi.org/10.1016/j.cub.2013.07.058

Guha, U., Gomes, W. A., Samanta, J., Gupta, M., Rice, F. L., & Kessler, J. A. (2004). Target-derived BMP signaling limits sensory neuron number and the extent of peripheral innervation in vivo. *Development*, *131*(5), 1175–1186. https://doi.org/10.1242/dev.01013

Guillaume, F., & Otto, S. P. (2012). Gene functional trade-offs and the evolution of pleiotropy. *Genetics*, *192*(4), 1389–1409. https://doi.org/10.1534/genetics.112.143214

Hämälä, T., Gorton, A. J., Moeller, D. A., & Tiffin, P. (2020). Pleiotropy facilitates local adaptation to distant optima in common ragweed (*Ambrosia artemisiifolia*). *PLOS Genetics*, *16*(3), e1008707. https://doi.org/10.1371/journal.pgen.1008707

Helgason, E., Phung, Q. T., & Dueber, E. C. (2013). Recent insights into the complexity of Tank-binding kinase 1 signaling networks: The emerging role of cellular localization in the activation and substrate specificity of TBK1. *FEBS Letters*, *587*(8), 1230–1237. https://doi.org/10.1016/j.febslet.2013.01.059

Hof, A. E. van't, Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., Hall, N., Darby, A. C., & Saccheri, I. J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, *534*(7605), 102–105. https://doi.org/10.1038/nature17951

Howes, T. R., Summers, B. R., & Kingsley, D. M. (2017). Dorsal spine evolution in threespine sticklebacks via a splicing change in *MSX2A. BMC Biology*, *15*(1), 115. https://doi.org/10.1186/s12915-017-0456-5

Huygens, C., Liénart, S., Dedobbeleer, O., Stockis, J., Gauthy, E., Coulie, P. G., & Lucas, S. (2015). Lysosomal-associated Transmembrane Protein 4B (LAPTM4B) decreases Transforming Growth Factor β1 (TGF-β1) production in human regulatory T cells. *Journal of Biological Chemistry*, *290*(33), 20105–20116. https://doi.org/10.1074/jbc.M115.655340

Jacobs, A., & Elmer, K. R. (2021). Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Molecular Ecology*, *30*(20), 4955–4969. https://doi.org/10.1111/mec.15817

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Kaplan, A. P., Joseph, K., & Silverberg, M. (2002). Pathways for bradykinin formation and inflammatory disease. *Journal of Allergy and Clinical Immunology*, *109*(2), 195–209. https://doi.org/10.1067/mai.2002.121316

Kim, E., Magen, A., & Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Research*, *35*(1), 125–131. https://doi.org/10.1093/nar/gkl924

Kitano, J., Bolnick, D. I., Beauchamp, D. A., Mazur, M. M., Mori, S., Nakano, T., & Peichel, C. L. (2008). Reverse evolution of armor plates in the threespine stickleback. *Current Biology: CB*, *18*(10), 769–774. https://doi.org/10.1016/j.cub.2008.04.027

Kniss, J. S., Jiang, L., & Piotrowski, T. (2016). Insights into sensory hair cell regeneration from the zebrafish lateral line. *Current Opinion in Genetics & Development*, *40*, 32–40. https://doi.org/10.1016/j.gde.2016.05.012

Kristóf, E., Zahuczky, G., Katona, K., Doró, Z., Nagy, É., & Fésüs, L. (2013). Novel role of ICAM3 and LFA-1 in the clearance of apoptotic neutrophils by human macrophages. *Apoptosis: An International Journal on Programmed Cell Death*, *18*(10), 1235–1251. https://doi.org/10.1007/s10495-013-0873-z

Kucera, M. (2017). *Baderlab/Autoannotateapp: Autoannotate 1.2*. Zenodo. https://doi.org/10.5281/ZENODO.883467

Lampert, F., Stafa, D., Goga, A., Soste, M. V., Gilberto, S., Olieric, N., Picotti, P., Stoffel, M., & Peter, M. (2018). The multi-subunit GID/CTLH E3 ubiquitin ligase promotes cell proliferation and targets the transcription factor Hbp1 for degradation. *ELife*, *7*, e35528. https://doi.org/10.7554/eLife.35528

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Laurentino, T. G., Boileau, N., Ronco, F., & Berner, D. (2022). The ectodysplasin-A receptor is a candidate gene for lateral plate number variation in stickleback fish. *G3 Genes|Genomes|Genetics*, jkac077. https://doi.org/10.1093/g3journal/jkac077

LE Rouzic, A., Østbye, K., Klepaker, T. O., Hansen, T. F., Bernatchez, L., Schluter, D., & Vøllestad, L. A. (2011). Strong and consistent natural selection associated with armour reduction in sticklebacks. *Molecular Ecology*, *20*(12), 2483–2493. https://doi.org/10.1111/j.1365-294X.2011.05071.x

Lee, Y., & Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annual Review of Biochemistry*, *84*, 291–323. https://doi.org/10.1146/annurev-biochem-060614-034316

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, *28*(6), 882–883. https://doi.org/10.1093/bioinformatics/bts034

Leinonen, R., Sugawara, H., Shumway, M., & International Nucleotide Sequence Database Collaboration. (2011). The sequence read archive. *Nucleic Acids Research*, *39*(Database issue), D19-21. https://doi.org/10.1093/nar/gkq1019

Lerner, U. H., Jones, I. L., & Gustafson, G. T. (1987). Bradykinin, a new potential mediator of inflammation-induced bone resorption. Studies of the effects on mouse calvarial bones and articular cartilage in vitro. *Arthritis and Rheumatism*, *30*(5), 530–540. https://doi.org/10.1002/art.1780300507

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lin, W. W., Hostager, B. S., & Bishop, G. A. (2015). TRAF3, ubiquitination, and B-lymphocyte regulation. *Immunological Reviews*, *266*(1), 46–55. https://doi.org/10.1111/imr.12299

Linnen, C. R., Kingsley, E. P., Jensen, J. D., & Hoekstra, H. E. (2009). On the origin and spread of an adaptive allele in deer mice. *Science*, *325*(5944), 1095–1098. https://doi.org/10.1126/science.1175826

Lovci, M. T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T. Y., Stark, T. J., Gehman, L. T., Hoon, S., Massirer, K. B., Pratt, G. A., Black, D. L., Gray, J. W., Conboy, J. G., & Yeo, G. W. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nature Structural & Molecular Biology*, *20*(12), 1434–1442. https://doi.org/10.1038/nsmb.2699

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Maitland, M. E. R., Lajoie, G. A., Shaw, G. S., & Schild-Poulter, C. (2022). Structural and functional insights into GID/CTLH E3 ligase complexes. *International Journal of Molecular Sciences*, *23*(11), 5863. https://doi.org/10.3390/ijms23115863

Mallarino, R., Linden, T. A., Linnen, C. R., & Hoekstra, H. E. (2017). The role of isoforms in the evolution of cryptic coloration in *Peromyscus* mice. *Molecular Ecology*, *26*(1), 245–258. https://doi.org/10.1111/mec.13663

Manchanda, A., Bonventre, J. A., Bugel, S. M., Chatterjee, P., Tanguay, R., & Johnson, C. P. (2021). Truncation of the otoferlin transmembrane domain alters the development of hair cells and reduces membrane docking. *Molecular Biology of the Cell*, *32*(14), 1293–1305. https://doi.org/10.1091/mbc.E20-10-0657

Marceau, F., & Regoli, D. (2004). Bradykinin receptor ligands: Therapeutic perspectives. *Nature Reviews Drug Discovery*, *3*(10), 845–852. https://doi.org/10.1038/nrd1522

Marquez, Y., Brown, J. W. S., Simpson, C., Barta, A., & Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Research*, *22*(6), 1184–1195. https://doi.org/10.1101/gr.134106.111

McCarthy, D. J., & Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics (Oxford, England)*, *25*(6), 765–771. https://doi.org/10.1093/bioinformatics/btp053

Merico, D., Isserlin, R., Stueker, O., Emili, A., & Bader, G. D. (2010). Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PloS One*, *5*(11), e13984. https://doi.org/10.1371/journal.pone.0013984

Metelli, A., Salem, M., Wallace, C. H., Wu, B. X., Li, A., Li, X., & Li, Z. (2018). Immunoregulatory functions and the therapeutic implications of GARP-TGF-β in inflammation and cancer. *Journal of Hematology & Oncology*, *11*(1), 24. https://doi.org/10.1186/s13045-018-0570-z

Mills, M. G., Greenwood, A. K., & Peichel, C. L. (2014). Pleiotropic effects of a single gene on skeletal development and sensory system patterning in sticklebacks. *EvoDevo*, *5*(1), 5. https://doi.org/10.1186/2041-9139-5-5

Mogdans, J. (2019). Sensory ecology of the fish lateral-line system: Morphological and physiological adaptations for the perception of hydrodynamic stimuli. *Journal of Fish Biology*, jfb.13966. https://doi.org/10.1111/jfb.13966

Nath, S., Shaw, D. E., & White, M. A. (2021). Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 Genes|Genomes|Genetics*, *11*(2), jkab007. https://doi.org/10.1093/g3journal/jkab007

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene *EDA*. *ELife*, *4*, e05290. https://doi.org/10.7554/eLife.05290

Oganesyan, G., Saha, S. K., Guo, B., He, J. Q., Shahangian, A., Zarnegar, B., Perry, A., & Cheng, G. (2006). Critical role of TRAF3 in the Toll-like receptor-dependent and -independent antiviral response. *Nature*, *439*(7073), 208–211. https://doi.org/10.1038/nature04374

Ostermann, G., Weber, K. S. C., Zernecke, A., Schröder, A., & Weber, C. (2002). JAM-1 is a ligand of the beta(2) integrin LFA-1 involved in transendothelial migration of leukocytes. *Nature Immunology*, *3*(2), 151–158. https://doi.org/10.1038/ni755

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–1415. https://doi.org/10.1038/ng.259

Peichel, C. L., & Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1713), 20150486. https://doi.org/10.1098/rstb.2015.0486

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rahman, M. S., Akhtar, N., Jamil, H. M., Banik, R. S., & Asaduzzaman, S. M. (2015). TGF-β/BMP signaling and other molecular events: Regulation of osteoblastogenesis and bone formation. *Bone Research*, *3*(1), 15005. https://doi.org/10.1038/boneres.2015.5

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D., & Bader, G. D. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature Protocols*, *14*(2), 482–517. https://doi.org/10.1038/s41596-018-0103-9

Reimand, J., Kull, M., Peterson, H., Hansen, J., & Vilo, J. (2007). g:Profiler—A web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research*, *35*(suppl_2), W193–W200. https://doi.org/10.1093/nar/gkm226

Reimchen, T. E. (1992). Injuries on stickleback from attacks by a toothed predator (*Onchoryhnchus*) and implications for the evolution of lateral plates. *Evolution; International Journal of Organic Evolution*, *46*(4), 1224–1230. https://doi.org/10.1111/j.1558-5646.1992.tb00631.x

Reimchen, T. E. (2000). Predator handling failures of lateral plate morphs in *Gasterosteus aculeatus*: functional implications for the ancestral plate condition. *Behaviour*, *137*(7–8), 1081–1096. https://doi.org/10.1163/156853900502448

Reimchen, T. E., Bergstrom, C., & Nosil, P. (2013). Natural selection and the adaptive radiation of Haida Gwaii stickleback. *Evolutionary Ecology Research*, *15*(3), 241–269.

Rennison, D. J., Heilbron, K., Barrett, R. D. H., & Schluter, D. (2015). Discriminating selection on lateral plate phenotype and its underlying gene, *Ectodysplasin*, in threespine stickleback. *The American Naturalist*, *185*(1), 150–156. https://doi.org/10.1086/679280

Rennison, D. J., & Peichel, C. L. (2022). Pleiotropy facilitates parallel adaptation in sticklebacks. *Molecular Ecology*, *31*(5), 1476–1486. https://doi.org/10.1111/mec.16335

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Roberts, W. M., Howard, J., & Hudspeth, A. J. (1988). Hair cells: Transduction, tuning, and transmission in the inner ear. *Annual Review of Cell Biology*, *4*, 63–92. https://doi.org/10.1146/annurev.cb.04.110188.000431

Robertson, S., Bradley, J. E., & MacColl, A. D. C. (2017). *Eda* haplotypes in three-spined stickleback are associated with variation in immune gene expression. *Scientific Reports*, *7*(1), 42677. https://doi.org/10.1038/srep42677

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rogers, T. F., Palmer, D. H., & Wright, A. E. (2021). Sex-specific selection drives the evolution of alternative splicing in birds. *Molecular Biology and Evolution*, *38*(2), 519–530. https://doi.org/10.1093/molbev/msaa242

Roux, I., Safieddine, S., Nouvian, R., Grati, M., Simmler, M.-C., Bahloul, A., Perfettini, I., Le Gall, M., Rostaing, P., Hamard, G., Triller, A., Avan, P., Moser, T., & Petit, C. (2006). Otoferlin, defective in a human deafness form, is essential for exocytosis at the auditory ribbon synapse. *Cell*, *127*(2), 277–289. https://doi.org/10.1016/j.cell.2006.08.040

Sadier, A., Viriot, L., Pantalacci, S., & Laudet, V. (2014). The ectodysplasin pathway: From diseases to adaptations. *Trends in Genetics: TIG*, *30*(1), 24–31. https://doi.org/10.1016/j.tig.2013.08.006

Salazar, V. S., Gamer, L. W., & Rosen, V. (2016). BMP signalling in skeletal development, disease and repair. *Nature Reviews Endocrinology*, *12*(4), 203–221. https://doi.org/10.1038/nrendo.2016.12

Santt, O., Pfirrmann, T., Braun, B., Juretschke, J., Kimmig, P., Scheel, H., Hofmann, K., Thumm, M., & Wolf, D. H. (2008). The yeast GID complex, a novel ubiquitin ligase (E3) involved in the regulation of carbohydrate metabolism. *Molecular Biology of the Cell*, *19*(8), 3323–3333. https://doi.org/10.1091/mbc.e08-03-0328

Schluter, D., Marchinko, K. B., Arnegard, M. E., Zhang, H., Brady, S. D., Jones, F. C., Bell, M. A., & Kingsley, D. M. (2021). Fitness maps to a large-effect locus in introduced stickleback populations. *Proceedings of the National Academy of Sciences*, *118*(3), e1914889118. https://doi.org/10.1073/pnas.1914889118

Schweighoffer, E., & Tybulewicz, V. L. (2018). Signalling for B cell survival. *Current Opinion in Cell Biology*, *51*, 8–14. https://doi.org/10.1016/j.ceb.2017.10.002

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, *73*(3), 751–754. https://doi.org/10.1093/biomet/73.3.751

Singh, P., & Ahi, E. P. (2022). The importance of alternative splicing in adaptive evolution. *Molecular Ecology*, *31*(7), 1928–1938. https://doi.org/10.1111/mec.16377

Singh, P., Börger, C., More, H., & Sturmbauer, C. (2017). The role of alternative splicing and differential gene expression in cichlid adaptive radiation. *Genome Biology and Evolution*, *9*(10), 2764–2781. https://doi.org/10.1093/gbe/evx204

Smith, N. C., Rise, M. L., & Christian, S. L. (2019). A comparison of the innate and adaptive immune systems in cartilaginous fish, ray-finned fish, and lobe-finned fish. *Frontiers in Immunology*, *10*, 2292. https://doi.org/10.3389/fimmu.2019.02292

Smulski, C. R., & Eibel, H. (2018). BAFF and BAFF-Receptor in B cell selection and survival. *Frontiers in Immunology*, *9*, 2285. https://doi.org/10.3389/fimmu.2018.02285

Soulliere, C., & Dixon, B. (2017). Immune system organs of bony fishes. In *Reference Module in Life Sciences* (p. B978012809633812179X). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.12179-X

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, *14*(11), 751–764. https://doi.org/10.1038/nrg3483

Steward, R. A., de Jong, M. A., Oostra, V., & Wheat, C. W. (2022). Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nature Communications*, *13*(1), 755. https://doi.org/10.1038/s41467-022-28306-8

Tovar-Corona, J. M., Castillo-Morales, A., Chen, L., Olds, B. P., Clark, J. M., Reynolds, S. E., Pittendrigh, B. R., Feil, E. J., & Urrutia, A. O. (2015). Alternative splice in alternative lice. *Molecular Biology and Evolution*, *32*(10), 2749–2759. https://doi.org/10.1093/molbev/msv151

Tseng, P.-H., Matsuzawa, A., Zhang, W., Mino, T., Vignali, D. A. A., & Karin, M. (2010). Different modes of ubiquitination of the adaptor TRAF3 selectively activate the expression of type I interferons and proinflammatory cytokines. *Nature Immunology*, *11*(1), 70–75. https://doi.org/10.1038/ni.1819

Ule, J., & Blencowe, B. J. (2019). Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Molecular Cell*, *76*(2), 329–345. https://doi.org/10.1016/j.molcel.2019.09.017

Verta, J.-P., & Jacobs, A. (2022). The role of alternative splicing in adaptation and evolution. *Trends in Ecology & Evolution*, *37*(4), 299–308. https://doi.org/10.1016/j.tree.2021.11.010

Verta, J.-P., & Jones, F. C. (2019). Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *ELife*, *8*, e43785. https://doi.org/10.7554/eLife.43785

Wagner, G. P., Pavlicev, M., & Cheverud, J. M. (2007). The road to modularity. *Nature Reviews Genetics*, *8*(12), 921–931. https://doi.org/10.1038/nrg2267

Walling, B. L., & Kim, M. (2018). LFA-1 in T cell migration and differentiation. *Frontiers in Immunology*, *9*, 952. https://doi.org/10.3389/fimmu.2018.00952

Wang, B.-B., & Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(18), 7175–7180. https://doi.org/10.1073/pnas.0602039103

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470–476. https://doi.org/10.1038/nature07509

Wang, R. N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., Idowu, O., Li, M., Shen, C., Hu, A., Haydon, R. C., Kang, R., Mok, J., Lee, M. J., Luu, H. L., & Shi, L. L. (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes & Diseases*, *1*(1), 87–105. https://doi.org/10.1016/j.gendis.2014.07.005

Wark, A. R., Mills, M. G., Dang, L.-H., Chan, Y. F., Jones, F. C., Brady, S. D., Absher, D. M., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., & Peichel, C. L. (2012). Genetic architecture of variation in the lateral line sensory system of threespine sticklebacks. *G3 (Bethesda, Md.)*, *2*(9), 1047–1056. https://doi.org/10.1534/g3.112.003079

Wright, C. J., Smith, C. W. J., & Jiggins, C. D. (2022). Alternative splicing as a source of phenotypic diversity. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-022-00514-4

Wucherpfennig, J. I., Miller, C. T., & Kingsley, D. M. (2019). Efficient CRISPR-Cas9 editing of major evolutionary loci in sticklebacks. *Evolutionary Ecology Research*, *20*(1), 107–132.

Yasunaga, S., Grati, M., Cohen-Salmon, M., El-Amraoui, A., Mustapha, M., Salem, N., El-Zir, E., Loiselet, J., & Petit, C. (1999). A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes DFNB9, a nonsyndromic form of deafness. *Nature Genetics*, *21*(4), 363–369. https://doi.org/10.1038/7693

Yi, Z., Wallis, A. M., & Bishop, G. A. (2015). Roles of TRAF3 in T cells: Many surprises. *Cell Cycle (Georgetown, Tex.)*, *14*(8), 1156–1163. https://doi.org/10.1080/15384101.2015.1021524

Yuan, Y.-W., Rebocho, A. B., Sagawa, J. M., Stanley, L. E., & Bradshaw, H. D. (2016). Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. *Proceedings of the National Academy of Sciences*, *113*(9), 2448–2453. https://doi.org/10.1073/pnas.1515294113

Zhang, B., & Horvath, S. (2005). A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1). https://doi.org/10.2202/1544-6115.1128

Chapter 3

The role of alternative splicing in marine-freshwater divergence in threespine stickleback

The role of alternative splicing in marine-freshwater divergence in threespine stickleback

Carlos E. Rodríguez-Ramírez[1] and Catherine L. Peichel[1]

[1]*Division of Evolutionary Ecology, Institute of Ecology and Evolution, University of Bern, Bern 3012, Switzerland*

A preprint of this chapter is available at bioRxiv:

Rodríguez-Ramírez CE, Peichel CL. 2024. The role of alternative splicing in marine-freshwater divergence in threespine stickleback. *bioRxiv*

http://biorxiv.org/lookup/doi/10.1101/2024.06.14.598968

The preprint is also in revision in Genome Biology and Evolution

Abstract

Alternative splicing (AS) regulates which parts of a gene are kept in the messenger RNA and has long been appreciated as a mechanism to increase the diversity of the proteome within eukaryotic species. There is a growing body of evidence that AS might also play an important role in adaptive evolution. However, the overall contribution of AS to phenotypic evolution and adaptation is still unknown. In this study we asked whether AS played a role in adaptation to divergent marine and freshwater habitats in threespine stickleback (*Gasterosteus aculeatus*). Using two published gill RNAseq datasets, we identified differentially expressed and differentially spliced genes (DEGs and DSGs) between population pairs of marine-freshwater stickleback in the Northeast Pacific and tested whether they are preferentially found in regions of the genome involved in freshwater-marine divergence. We found over one hundred DSGs, and they were found more often than expected by chance in peaks of genetic divergence and quantitative trait loci (QTL) that underlie phenotypic divergence between ecotypes. The enrichment of DSGs in these regions is similar to the enrichment of DEGs. Furthermore, we find that among the different types of AS, mutually exclusive exon splicing is the most strongly correlated with genetic divergence between ecotypes. Taken together, our results suggests that AS might have played a role in the adaptive divergence of marine and freshwater sticklebacks and that some types of AS might contribute more than others to adaptation.

Key words: adaptation, alternative splicing, threespine stickleback, regulatory evolution, gene expression, gill

1 Introduction

Gene regulatory evolution has long been proposed to be an important driver of phenotypic evolution (King and Wilson 1975). In particular, cis-regulatory evolution has been argued to be an important driver of phenotypic evolution due to its potential to reduce antagonistic pleiotropy compared to protein evolution (Carroll 2005; Hoekstra and Coyne 2007; Stern and Orgogozo 2008; Wittkopp and Kalay 2012; Bomblies and Peichel 2022). Supporting these ideas, changes in gene expression mediated by cis-regulatory mutations have been shown to underlie phenotypic evolution and adaptation to different environments (Stern and Orgogozo 2008; Rebeiz et al. 2009; Chan et al. 2010; Wittkopp and Kalay 2012; Indjeian et al. 2016; Hill et al. 2021; Wooldridge et al. 2022). However, the literature on the role of regulatory evolution in phenotypic evolution and adaptation has for the most part focused solely on changes in gene expression and until recently has ignored other mechanisms of gene regulation, such as alternative splicing, which is common in many eukaryotic lineages (Chen et al. 2014). When alternative splicing occurs, different combinations of exons, and sometimes introns are included in the mature mRNA, leading to alternative mRNA isoforms that might result in the translation of different proteins from the same gene (Bush et al. 2017; Singh and Ahi 2022; Verta and Jacobs 2022; Wright et al. 2022). There are five types of alternative splicing (AS) events: 1) exon skipping (ES), when an alternative exon is not included in the mRNA; 2) mutually exclusive exons (MXE), when one exon out of a group is always included in the mRNA, but never more than one of the exons at the same time; 3) intron retention (IR), when an intron is kept in the mRNA instead of being spliced out as usual; 4) alternative 3' splice sites (A3SS); and 5) alternative 5' splice sites (A5SS), when part of the 3' or 5' end of an exon is spliced out of the mRNA (Figure 1). One gene can undergo a combination of these types of alternative splicing events, allowing for multiple different mRNA isoforms of a gene and leading to an overall increase in the diversity of the proteome. This might be the reason why AS has been found to be correlated with complexity (as measured by cell-type diversity) across eukaryotic lineages (Chen et al. 2014).

The potential of AS to create different proteins has led to the question of whether it might play a role in phenotypic evolution and adaptation (Bush et al. 2017; Singh and Ahi 2022; Verta and Jacobs 2022; Wright et al. 2022). Indeed, recent studies have found that changes in AS can lead to phenotypic differences between ecotypes or species. For example, dorsal spine reduction in freshwater populations of threespine stickleback (*Gasterosteus aculeatus)* is mediated by the use of an A5SS in the first exon of *Msx2a,* a gene involved in osteoblast differentiation (Howes et al. 2017). Infrared sensation in vampire bats is linked to changes in the ratio of ES in the temperature receptor gene *Trpv1,* making it sensitive to temperatures around 30 ^{o}C instead of 40 ^{o}C (Gracheva et al. 2011). Increased

lipid accumulation in two populations of the cavefish *Astyanax mexicanus* is caused by ES, resulting in a premature stop codon in the gene *per2,* a suppressor of lipid metabolism (Xiong et al. 2022). Eighteen other examples of intra or interspecific phenotypic variation mediated by alternative splicing are found at the genotype-phenotype database GePheBase (Martin and Orgogozo 2013; Courtier-Orgogozo et al. 2020), highlighting the potential of alternative splicing to cause phenotypic evolution. Supporting a role for AS in adaptation, many differentially spliced genes (DSGs) have been found between species or divergent ecotypes in human lice, cichlids, charr, sunflower and *Arabidopsis* (Tovar-Corona et al. 2015; Singh et al. 2017; Smith et al. 2018; Wang et al. 2019; Jacobs and Elmer 2021), and across environmental clines in wild house mice (Manahan and Nachman 2024). However, to our knowledge, to date only one study in benthic and pelagic ecotypes of Artic charr (Jacobs and Elmer 2021) has looked for evidence of divergent selection in DSGs between ecotypes of a species.

The threespine stickleback (*G. aculeatus*) is a good model to study questions related to the genetic basis of phenotypic evolution and adaptation. This small teleost fish is distributed across the Northern Hemisphere and the ancestral marine form independently colonised and adapted to many freshwater habitats approximately 12 000 years ago after the Last Glacial Maximum. Marine and freshwater stickleback diverge in ecology, physiology, and morphology, with the repeated evolution of many phenotypes in freshwater (Bell and Foster 1994). Genetic studies have found several of these parallel phenotypes are due to mutations in the same gene (Colosimo et al. 2005; Miller et al. 2007; Chan et al. 2010; Ishikawa et al. 2019). Many quantitative trait loci (QTL) mapping studies have identified regions of the genome strongly associated with phenotypic divergence (Peichel and Marques 2017). In addition, global genomic studies incorporating marine and freshwater population pairs from across the Northern Hemisphere have identified parallel peaks of genetic divergence across marine-freshwater population pairs that are putatively under divergent selection. Interestingly, most of these peaks of divergence are found in non-coding regions, hinting an important role of cis-regulatory evolution in the divergence between marine and freshwater sticklebacks. (Jones et al. 2012; Roberts Kingman et al. 2021).

Here, we use this system to ask whether alternative splicing could be a regulatory mechanism that contributes to adaptation to divergent environments. More precisely, we ask whether alternative splicing is important for marine-freshwater divergence in threespine stickleback. Using publicly-available RNA-seq data from marine and freshwater populations from the Northeast Pacific, we ask the following questions: 1) are there differentially spliced genes (DSGs) between the two ecotypes?; 2) is there evidence that these DSGs might mediate phenotypic divergence between the ecotypes?; and 3) is there any evidence that natural selection acted on these DSGs?

2 Results

2.1 Over one hundred DSGs between marine and freshwater stickleback

Using an RNA-seq dataset from gill tissue of marine and freshwater sticklebacks from Canada (Supplementary Table S1), we found 16 999 expressed genes, of which 1882 are differentially expressed genes (DEGs) between marine and freshwater samples. We detected alternative splicing events in 1345 genes, of which 139 are differentially spliced genes (DSGs) between ecotypes (Figure 2a, Supplementary Table S2). Thirty genes are simultaneously DEGs and DSGs (differentially expressed and spliced genes, or DESGs). We found all five types of differential splicing events amongst the DSGs (Figure 2b). Differential MXE is the most common and is present in 71 DSGs; this is followed by 45 DSGs with differential ES, 22 DSGs with IR, 18 DSGs with A3SS, and 8 DSGs with differential A5SS (Figure 2b, Supplementary Table S2). There is almost no enrichment of GO Terms in DEGs or DSGs. The only exception is an enrichment for the terms “phosphatidylinositol monophosphate phosphatase activity” and “phosphatidylinositol-3-phosphate phosphatase activity” in the DEGs (Supplementary Table S3).

2.2 Pacific DSGs are significantly enriched in some categories of QTL

To test whether DEGs and DSGs might be mediating phenotypic divergence between marine and freshwater sticklebacks, we tested their enrichment in 316 QTL that underlie traits that diverge between Pacific marine and freshwater populations (Supplementary Table S4) (Peichel and Marques 2017; Liu et al. 2022; Rennison and Peichel 2022). The QTL span most of the gill transcriptome: out of the 16 999 genes in this dataset, 12 129 genes (71.4%) are inside at least one QTL. We found that DEGs are enriched in the overall set of QTL, as well as in most phenotypic sub-categories (Figure 3 and Supplementary Table S5). Meanwhile, the DSGs are not enriched in the overall set of QTL, but they are enriched in QTL sub-categories associated with body shape, defence and feeding (p-value = 0.001, 0.013 and 0.033, respectively; permutation tests, 1000 permutations). DESGs are also enriched in QTL associated with body shape and swimming (p-value = 0.01 and 0.009, respectively; permutation tests, 1000 permutations) (Figure 3 and Supplementary Table S5).

2.3 DEGs and DSGs are significantly enriched in EcoPeaks

To determine whether DEGs and DSGs might be under divergent selection, we tested whether they are significantly enriched in EcoPeaks, which are regions of the genome with peaks of genetic divergence between multiple marine and freshwater populations from either the Northeast Pacific ("Pacific EcoPeaks") or from the Northeast Pacific, California, and Europe ("Global EcoPeaks") (Roberts Kingman et al. 2021). The Pacific EcoPeaks include 22.7% of the genes in our transcriptome (3854 out of 16 999) (Supplementary Table S2). We found 43.4% of DEGs (803 out of 1852), 40.4% of the DSGs (44 out of 109) and 56.7% of the DESGs (17 out of 30) are inside the Pacific EcoPeaks. This is significantly more than the 22.7% to 23.7% of background genes that are in these regions of the genome (Figure 4a, Supplementary Table S6). DEGs and DSGs are similarly enriched in Global EcoPeaks (Figure 4b, Supplementary Table S6).

DEGs and DSGs overlapping both EcoPeaks and QTL are the strongest candidates for genes with an adaptive role in the marine-freshwater divergence in threespine stickleback. Thus, we asked whether DEGs and DSGs overlapping Pacific EcoPeaks and QTL are enriched in any particular biological functions. However, we found that these genes are not significantly enriched in any GO Term category (Supplementary Table S3). Thus, we used the GeneCards database to investigate the function of the 15 DSGs found within both Global EcoPeaks and QTL. We found multiple genes involved in essential amino acid metabolism, chromatin remodelling, immunity, vesicle transport and muscle function (Supplementary Table S7).

Examining the types of differential splicing events underlying the DSGs, we found that DSGs with significant MXE, ES and IR events are enriched in Pacific EcoPeaks (54.9%, 37.8% and 31.8% respectively; p-values < 0.001, 0.021 and 0.034 respectively; permutation test, 1000 permutations) (Supplementary Figure S1a, Supplementary Table S6). Though not significant, DSGs with A3SS show a similar trend, with 7 out of 18 (38%) found in EcoPeaks (p-value = 0.172; permutation test, 1000 permutations). DSGs with MXE and IR are also significantly enriched in the Global EcoPeaks (Supplementary Figure S1b, Supplementary Table S6).

Finally, we also examined whether there is a difference in the overall fold-change in expression (DEGs) or the inclusion isoform change ("isoform difference" in DSGs) between genes inside versus outside of EcoPeaks. We found that DSGs in both Pacific and Global EcoPeaks have a significantly higher isoform difference than those outside, particularly DSGs within Global EcoPeaks (Figure 5, Table S8). Interestingly, DEGs do not differ in their fold-change inside and outside of EcoPeaks (Figure 5, Table S8).

2.4 No strong correlation between transcriptome-wide genetic divergence and differential expression and splicing.

To determine to what extent marine-freshwater splicing divergence is correlated with marine-freshwater genetic divergence, we compared the isoform difference of all 1345 genes used in the differential splicing analysis with their average SNP p-value from the Pacific EcoPeak data (Roberts Kingman et al. 2021). We found a very weak positive correlation between isoform difference and genetic distance (R2 = 0.015, p-value = 1.3e-06) (Supplementary Figure S2). We found a similarly weak positive correlation between genetic distance and the log of expression fold-change between ecotypes (R2 = 0.010, p-value = 2.2e-16) (Supplementary Figure S2). These weak correlations are likely driven by the EcoPeaks. When we separated the data inside and outside of the EcoPeaks, the correlations between genetic distance and isoform difference/fold-change are weaker, and in most cases no longer significant (Supplementary Figure S2).

2.5 Differences in effect size and genetic divergence between types of AS

To gain insights into whether certain types of splicing could be more important to adaptation than others, we tested whether genetic divergence and strength of DS changes between the five types of AS events. For this analysis, we classified DSGs by the type of AS event that was the strongest DS event in that gene. When comparing the different types of AS with each other, there is a tendency for genes with MXE and IR to have a higher isoform difference between ecotypes, though this is only significant in the comparison between MXE and ES (Supplementary Figure S3a and Table S9a). Regarding genetic divergence, there is a tendency for genes with MXE to be more divergent than other types of DSGs, but it is only significant when compared to IR (Supplementary Figure S3b and Table S9b). However, when we compare the DSGs to their non-DSG counterparts for each type of AS, we find that DSGs with differential MXE, ES and IR have a significantly higher genetic divergence than genes that have MXE, ES and IR, respectively, but that are not differentially spliced between ecotypes (Figure 6 and Table S10). This suggests that DSGs with these types of splicing are more likely to have been targeted by selection between marine and freshwater sticklebacks.

3 Discussion

The magnitude of the role of alternative splicing in adaptation is still unknown. We sought to tackle this question by assessing the role of alternative splicing in the marine-freshwater divergence of threespine stickleback (*G. aculeatus*). We show that there are more than one hundred DSGs in the gill transcriptome between marine and freshwater stickleback populations from the Northeast Pacific. DSGs are enriched not only in regions under putative divergent selection in the genome, but also within QTL underlying phenotypic divergence between ecotypes in the Pacific. Furthermore, the enrichment of DSGs in these regions of the genome was similar to the enrichment found for DEGs, suggesting that DSGs might be as important for adaptation as DEGs. Among the five types of AS, MXE splicing events are the most commonly divergent between ecotypes and are in the most genetically divergent genes. Thus, mutually exclusive exon use could be particularly important for adaptation to different environments. Taken together our results suggest that alternative splicing might play an important role in marine-freshwater divergence in threespine stickleback.

3.1 Over one hundred genes have changes in alternative splicing between marine and freshwater sticklebacks in gill

To determine the role of alternative splicing in the freshwater-marine divergence in threespine stickleback, we first assessed the extent of differential expression and differential splicing between ecotypes. We found 1852 DEGs, 109 DSGs and 30 DESGs (Figure 2a). This disparity in abundance of DEGs and DSGs is a pattern commonly found in other studies (Grantham and Brisson 2018; Jacobs and Elmer 2021; Steward et al. 2022; Rodríguez-Ramírez et al. 2023), though not always (Singh et al. 2017). This might suggest that differential splicing is less commonly used than differential expression for adaptation. Since AS directly affects the protein sequence it is possibly under stronger purifying selection than differential expression. However, it is important to note that methods to detect AS in short-read data have greater limitations in their ability to identify and quantify changes in splicing than methods for identification of DEGs in short-read data. For example, methods based on splice junctions like the one used in this study and others (e.g. Steward et al., 2022) work mainly with reads overlapping these junctions. Because most reads cannot be used for analyses, the power of the DS analyses is greatly reduced compared to the DE analyses, which can use most reads in the transcriptome. Methods based on differential exon usage have been used in other studies (e.g. Singh et al., 2017; Jacobs & Elmer 2021; Rodríguez-Ramírez et al., 2023) and can use most of the same reads as DE analyses but cannot detect certain types of splicing (i.e. A3SS and A5SS). Thus, all AS studies based on

short-read RNA-seq data are likely to underestimate the extent of DS. In contrast, long-read RNA sequencing methods like Iso-seq, give us information on the full mRNA sequence and have revealed many previously unknown isoforms both in animals and plants (Singh and Ahi 2022). Future studies using long-read RNA sequencing technologies are essential to properly assess the relative role of DEGs and DSGs in adaptation.

3.2 DSGs are associated with regions underlying phenotypic divergence in body shape, defence and feeding between marine and freshwater stickleback.

Consistent with a role for DEGs and DSGs in mediating phenotypic divergence between marine and freshwater stickleback, we found that DEGs and DSGs were enriched in QTL associated with a variety of phenotypes. While DEGs were enriched in more QTL, DSGs were still associated with QTL underlying phenotypes like body shape, defence traits and feeding traits. However, an important caveat is that these QTL are mostly associated with phenotypes completely unrelated to gill tissue, except for some of the feeding trait QTL (Supplementary Table S4). The expression of genes associated with these QTL in gill tissue could be due to several non-mutually exclusive reasons. One possibility is that some of these genes have pleiotropic effects in both gill and the tissues associated with the QTL. For example, the gene *stat3* is amongst the most divergent DSGs in the dataset (Supplementary Table S7) and is a pleiotropic transcription factor with described roles in multiple biological processes including skeletal development, hair maintenance, immunity and cellular respiration (Levy and Lee 2002; O'Shea et al. 2002; Wegrzyn et al. 2009; Hillmer et al. 2016; Zhou et al. 2021). Furthermore, *stat3* in mice is known to produce two alternative splicing isoforms: a full-length isoform, *stat3α;* and a truncated isoform, *stat3β.* The isoforms have different phenotypic effects and are thought to partially explain the high pleiotropy of this gene (Maritano et al. 2004). A second possibility is that most of the QTL have very low resolution and span large regions of the genome, so most of the genes found within them are unlikely to be related to the focal QTL phenotype. An enrichment of DEGs and DSGs in these QTL could occur if these genes affect other unmapped phenotypes that are linked to these QTL, consistent with the QTL clustering observed in stickleback (Peichel and Marques 2017). A final possibility is that strong selection for DE and DS in a gene within a specific tissue could allow the fixation of regulatory variants that cause leaky expression and splicing in unrelated tissues like gill. Mutations in cis-regulatory elements are known to affect gene expression noise (Richard and Yvert 2014) and the strength of selection against noise depends on the function of the gene and its position in the gene pathway (Barroso et al. 2018). Thus, it is possible that when rapid adaptation to different environments occurs, regulatory variants that are favourable in one tissue might get fixed despite increasing the

transcriptional noise of that gene in other tissues. While the DEGs and DSGs that we found are enriched in QTL and EcoPeaks between marine and freshwater sticklebacks, many of the DEGs and DSGs have relatively small effect sizes (Supplementary Table S2), which is what we would expect if some of these genes represented some sort of transcriptional leakage resulting from divergent selection between marine and freshwater stickleback in other tissues.

3.3 DSGs are associated with regions under divergent selection between ectoypes

Having determined that DSGs are enriched in regions of the genome associated to phenotypic divergence between marine and freshwater sticklebacks, next we asked whether the same was true for the smaller subset of the genome with putative signatures of divergent selection between ecotypes. We used the EcoPeaks database (Roberts Kingman et al. 2021) and found that DEGs and DSGs were enriched in both Pacific and Global EcoPeaks (Figure 4). This suggests that DSGs are as likely as DEGs to be under divergent selection between threespine stickleback ecotypes in Northeast Pacific populations. It is possible that some of these genes are DE and DS simply because their causative cis-regulatory variants merely hitchhiked with the actual targets of selection in the EcoPeaks. However, we find only a very weak linear correlation between genetic divergence in the EcoPeaks data and the strength of DS and DE across the gill transcriptome (Supplementary Figure S2), suggesting that the DEGs and DSGs inside of the Pacific EcoPeaks are not just a side-effect of local genetic divergence.

There is not a strong linear correlation between genetic distance in the Pacific EcoPeak data and the strength of DE or DS in the gill transcriptomes. However, it is important to note that we do not have genomic sequencing data for the individuals used to identify DEGs and DSGs, so we cannot directly look at the correlation between genetic distance and DE and DS. Nonetheless, we do find that DSGs in Global EcoPeaks have stronger DS than those outside of Global EcoPeaks (Figure 5). We find a similar trend for DSGs in the Pacific EcoPeaks (p-value = 0.059, Figure 5, Supplementary Table S8). Interestingly, we do not find an increase in the fold-change of DEGs inside and outside of EcoPeaks (Figure 5, Supplementary Table S8). If the strength of DE and DS is correlated with their adaptive importance, this result suggests selection for stronger splicing divergence in the EcoPeaks but not for stronger expression divergence. It is also possible that divergent selection is equally strong in DE and DS, but splicing-mediated phenotypic effects require greater splicing divergence than expression-mediated phenotypic effects. Finally, it is possible that DS is less constrained by purifying selection and thus diverges more quickly inside the EcoPeaks, potentially through hitchhiking of splicing regulatory variants with other targets of selection. However, if this hypothesis were true, we would expect to find

a correlation between genetic divergence and DS inside and outside of the Pacific EcoPeaks, which we do not (Supplementary Figure S2). Taken together, our results suggest that there is selection for greater splicing divergence than expression divergence between marine and freshwater sticklebacks inside the EcoPeaks.

3.4 MXE, IR and ES may play an important role in marine-freshwater divergence

Alternative splicing can generate different types of splicing events (Figure 1). To test whether they might play different roles in adaptation, we assessed the presence of five types of AS in our data. We found all five types of AS in the DSGs of our dataset. MXE, ES and IR events are not only more common, but they are also the only types of AS enriched in EcoPeaks (Supplementary Figure S1). Furthermore, we also find that only DSGs with either of these three types of differential AS events have a higher genetic divergence their non-DSGs counterparts. Taken together these results suggest that MXE, ES and IR are more likely than A3SS and A5SS to be under selection between the marine-freshwater sticklebacks and play a role in their adaptive divergence.

It is not clear why A3SS and A5SS would be less used in adaptation. Just like ES and IR, A3SS and A5SS remove and add mRNA sequence and thus have the potential to increase proteomic diversity. However, most of the time this likely leads to aberrant proteins, truncated proteins or degradation of the mRNA by nonsense-mediated mRNA decay (NMD). Nonetheless, all of these can still be a mechanism for indirect down-regulation of genes, which could be adaptive. This is the case for gene *Msx2a*, which underlies dorsal spine differences between marine and freshwater threespine sticklebacks. The freshwater allele of *Msx2a* leads to shorter dorsal spines and has increased A5SS splicing of the first exon of this gene, which introduces an early stop codon and leads to a truncated non-functional version of the protein (Howes et al. 2017). Similarly, an ES event in the gene *TBXT* removes part of the transcriptional regulation domain of the protein and leads to the loss of tail in apes (Xia et al. 2024). Overall, ES and IR probably tend to add or remove larger sequences to the mRNA than A3SS and A5SS, which could increase the probability of new protein domains to emerge. This might also make down-regulation of genes by NMD more reliable through ES and IR than A3SS and A5SS. Indeed, IR has been found to both regulate gene expression by NMD and to contribute to functional isoforms (Wong et al. 2016). ES has similarly been found to regulate the incorporation of “poison” exons in mammalian splicing regulator (SR family) genes that lead to NDM but are important for cell fitness (Wright et al. 2022).

Furthermore, ES can also regulate new exons that emerge in intronic sequences. This can occur for example, through exonization of intronic transposable elements (TEs) (Lev-Maor et al. 2007; Sorek 2007; Wright et al. 2022). Studies in human and mice suggest that most new exons are alternatively spliced and expressed at low frequencies, making them nearly neutral (Xing and Lee 2006; Sorek 2007). This can allow new exons to evolve and potentially acquire a new function over time. For example, primates possess many lineage-specific exons that have originated from a class of TEs known as *Alu* elements. While their function is still not clear, *Alu*-containing exons can be expressed. Although they tend to be alternatively skipped (Sorek et al. 2002), some have been found to contribute to the sequence of proteins (Lin et al. 2016; Martinez-Gomez et al. 2020; Wright et al. 2022). While over 79% of new exons in humans are predicted to be deleterious (Sorek 2007), some are known to have acquired new functions. For example, the human ADAR2 gene is an RNA editing enzyme with a primate-specific exon 8 derived from an *Alu* element. This exon can be included in the catalytic region of the protein, which still works with the same substrate but with an altered catalytic activity (Sorek 2007).

Despite the potential of ES and IR to contribute to adaptation, MXE is the type of splicing that stood out the most in our data. This is the most common differential AS event (Figure 2), the most enriched in EcoPeaks (Supplementary Figure S1), and the AS event found in the DSGs with the greatest isoform differences and genetic divergence (Figure 6; Supplementary Figure S3). MXE is a type of splicing that switches one alternative exon for another in the mRNA (Figure 1). This allows for modular changes in the protein structure where one protein domain can be switched for a different one without affecting the integrity of the protein. One extreme example of this is the *Dscam* gene in *Drosophila melanogaster,* which encodes for a transmembrane protein that is important for neuronal connection and self-avoidance. The gene has 115 exons, of which 95 belong to one of four clusters of mutually exclusive exons and has been found to produce up to 18 496 isoforms (Sun et al. 2013), more than the number of genes in the genome of *D. melanogaster.* Furthermore, MXE is thought to be associated with exon duplications (Kondrashov 2001; Letunic 2002; Wright et al. 2022). Similar to gene duplications and exonization of intronic sequences, MXE can allow for the evolution of new functions in one of the paralogous exons while maintaining its ancestral function in the other paralog. Evolving new exons through duplications rather than exonization of TEs could be advantageous in that duplicated exons are immediately able to code for a functioning part of the existing protein. Thus, exon duplication and combined with MXE could promote the evolution of alternative protein domains that might become advantageous when the species adapts to a different environment. Taken together, our results suggest that MXE could be a powerful mechanism to maintain standing genetic variation in

protein isoforms and could play an important role in the divergence of marine and freshwater stickleback in the Pacific.

4 Conclusion

The role of alternative splicing in adaptation is still poorly understood. However, alternative splicing can be quite a versatile regulatory mechanism that can both indirectly down-regulate the expression of a gene by creating non-functional isoforms and mediate modular changes in the protein by permuting different exons from the gene into the mRNA. Consistent with a growing body of evidence in other systems, we find evidence for a role of alternative splicing in the marine-freshwater adaptive divergence of threespine stickleback. DSGs between marine and freshwater populations are enriched in QTL underlying phenotypic divergence between ecotypes, and DSGs are as enriched as DEGs in regions of the genome putatively under divergent selection between marine and freshwater populations. We also find evidence that different types of alternative splicing might contribute differently to adaptation, with MXE standing out in our data and suggesting that the modular change of exons could be particularly important for adaptation. Finally, our results are quite likely an underestimate of the true extent of differential splicing between marine and freshwater sticklebacks. Limitations of geographical representation, number of tissues, developmental time points and short-read data mean that we are probably detecting only a small part of the alternative splicing events in the threespine stickleback transcriptome. Future studies with more populations and tissues, long-read sequencing, and functional analyses will be essential to have a more precise picture of the role of alternative splicing in marine-freshwater divergence in threespine stickleback.

5 Materials and Methods

5.1 RNA-seq data

We searched the NCBI Sequence Read Archive (SRA) for publicly available RNA-seq data that met three criteria: 1) the data included samples from marine and freshwater population pairs of threespine stickleback (*G. aculeatus*); 2) the data came from the same tissue, so it could be merged and compared; 3) the data came from Pacific populations, since these are the ones with the most complete QTL and genetic divergence data (Peichel and Marques 2017; Roberts Kingman et al. 2021). Following these criteria, we found data from two RNA-seq studies in stickleback gill tissue (Gibbons et al. 2017; Verta and Jones 2019) (Supplementary Table S1). The Verta and Jones (2019) data is from first-generation descendants of wild-caught individuals grown in a common garden in the lab at 3.5 parts per thousand (ppt) salinity. From this dataset we used four marine and four freshwater individuals from the Little Campbell River (British Columbia, Canada). The Gibbons et al. (2017) data included ten marine wild-caught individuals from Oyster Lagoon (British Columbia, Canada) that spent four weeks at 20 ppt in the laboratory before being gradually moved to either 0.0 ppt (five individuals) or 30 ppt (five individuals) for three months, and ten freshwater wild-caught individuals from Trout Lake (British Columbia, Canada) that spent four weeks at 2.0 ppt in the laboratory before gradually being exposed to either 0.0 ppt (five individuals) or 30 ppt (five individuals) for three months.

5.2 RNA-seq data pre-processing

Quality control of the RNAseq read libraries was done with FastQC v0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were quality trimmed using Trimmomatic v0.36 (Bolger et al. 2014). Reads where both paired ends passed quality filtering were then mapped against version 5 of the *G. aculeatus* genome (Nath et al. 2021) using STAR v2.7.10b (settings: --twopassMode --chimSegmentMin [1/3 of read length] --alignSJDBoverhangMin 3 --alignIntronMin 70 --alignIntronMax 562000 --alignMatesGapMax 562000 --limitSjdbInsertNsj 2000000). Library quality metrics from FastQC and alignment quality metrics from STAR and featureCounts were summarized and visualized with MultiQC v1.14 (Ewels et al. 2016) to assess sample quality (Supplementary Table S1).

5.3 Identification of differentially expressed genes (DEGs)

We obtained count tables of the RNA-seq reads mapped to each gene in the genome using the gene annotations from NCBI (build 100) for version 5 of the *G. aculeatus* genome (https://www.ncbi.nlm.nih.gov/datasets/taxonomy/481459/) and *featureCounts* from the *Subread* v2.0.3 (Liao et al. 2014) package. We excluded reads where one of the pairs was unaligned, or when the two pairs mapped to different chromosomes or different strands. Next, we used the *edgeR* v3.28.1 package (Chen et al. 2008) in *R* v3.6.1 (R Core Team 2019) for the differential expression analysis. First, we filtered lowly-expressed genes using the *filterByExpr()* function, which removes genes with less than 10 counts in a minimum number of samples based on sample size (for further details consult the *filterByExpr()* documentation). For our data, this meant that genes with less than ten counts in nine or more samples were removed. Second, we fit the count data in *edgeR* to a negative binomial general linear model to control for batch effects present in the datasets. We used a ~(study+ecotype) model to control for the effect of having data from two different studies. Third, we ran a quasi-likelihood F test on the fitted data to test for differential expression between marine and freshwater samples.

5.4 Identification of differentially spliced genes (DSGs)

To identify DSGs we used a method based on the identification of splicing events in RNA-seq data implemented in the program *rMATs* v4.1.2 (Shen et al. 2014). First, rMATs uses reads that STAR maps to exon boundaries in the mRNA or read pairs that map to different exons in a gene, to identify five types of alternative splicing events: Exon skipping (ES), alternative 5' and 3' start sites (A5SS and A3SS), intron retention (IR), and mutually exclusive exons (MXE). For each splicing event, rMATs defines an inclusion isoform and a skipping isoform (Figure 1), counts how many reads map to each isoform, and calculates an Isoform Inclusion Difference metric (Shen et al. 2014), also known as Percent Spliced-In, or PSI (Grantham and Brisson 2018; Rogers et al. 2021). This metric, which we will refer to as isoform difference, measures how much the ratio of the inclusion isoform changes between treatments (Shen et al. 2014). Since rMATs does not include batch-correction or minimum expression filters, we took the raw counts of the splicing events identified by *rMATs* and used the *edgeR* function *filterByExp(),* to filter out lowly expressed splicing events. Then, we corrected for the study effect in the data just as in the differential expression analysis using the function *ComBat-seq(),* from the package *sva* v3.35.2, which outputs batch-corrected counts (Zhang et al. 2020). Finally, using the "*—task stat*" mode in *rMATs,* we re-calculated the isoform differences using the batch-corrected counts from the splicing events and did the statistical test for differential splicing. We used the default rMATs settings for the analysis.

5.5 EcoPeak and QTL enrichment analyses

To test whether DEGs and DSGs might be under divergent selection, we asked whether they were enriched in "EcoPeaks", regions of the genome with peaks of genetic divergence between multiple marine and freshwater populations from across the Northern hemisphere (Roberts Kingman et al. 2021). The dataset is divided into Northeast Pacific (which we will refer to as Pacific for simplicity) and Global EcoPeaks depending on the samples used. The Pacific EcoPeaks were identified by comparing 12 marine and 57 freshwater populations from Alaska (US), Haida Gwaii (Canada), British Columbia (Canada), and Washington State (US). The Global EcoPeaks were identified by comparing 28 marine and 56 freshwater populations from the Northeast Pacific, California, and Europe. EcoPeaks were identified using two approaches: 1) a window-based genetic distance approach; and 2) a SNP-level statistical test for imbalance of genetic variants between marine and freshwater populations. Sensitive EcoPeaks were defined when an FDR of 5% was obtained in either of the two analyses, while Specific EcoPeaks were defined when an FDR of 1% was obtained in both analyses (for more details consult Roberts Kingman et al., 2021). Since the number of DSGs in our dataset was not very high, we used the Sensitive EcoPeaks to increase the power of our analyses. We tested the enrichment of our DEGs and DSGS in both Pacific and Global EcoPeaks. Since the EcoPeaks were originally identified in the v4.1 (*gasAcu1-4*) genome assembly (Roberts Kingman et al. 2021), we lifted over the coordinates of the EcoPeaks to the coordinates of the version 5 genome (Supplementary Table S11) using the *liftOver()* function of the *R* package *rtracklayer* v1.46.0 (Lawrence et al. 2009) and the chain file "v4.1_to_v5.chain" available at the Stickleback Genome Browser (https://stickleback.genetics.uga.edu/).

Briefly, the enrichment analysis involved: 1) identifying the proportion of DEGs and DSGs that overlapped with EcoPeaks; 2) identifying the proportion of background genes that overlapped with EcoPeaks; and 3) testing if the proportion of DEGs and DSGs in EcoPeaks is significantly higher than the proportion of background genes using a permutation test (1000 permutations) in R. The background genes included in this analysis corresponded to the genes that were tested for differential expression or differential splicing, respectively, in each dataset. For the differential expression analysis, the background genes are all that passed the minimum gene expression filter. For the differential splicing analysis, the background genes are limited to those genes for which rMATs found evidence of alternative splicing and were therefore tested for differential splicing between the marine and freshwater ecotypes. For the enrichment analysis of genes that were both differentially expressed and differentially spliced (DESGs), we used the intersection of the DEG and DSG background genes. To

compare whether genes in EcoPeaks had stronger differential expression or differential splicing than genes outside EcoPeaks, we used a permutation test (10 000 permutations) to compare whether the median of the distribution of log2 fold-change of DEGs or the distribution of isoform difference of DSGs was different for genes inside and outside of EcoPeaks.

To test whether the DEGs and DSGs we identified might be involved in phenotypic divergence between the ecotypes, we did a similar enrichment analysis for quantitative trait loci (QTL) that underlie traits that differ between Pacific marine and freshwater populations (Peichel and Marques 2017; Rennison and Peichel 2022). We used the dataset of Liu et al. (2022), which excluded QTL that did not go in the expected direction (i.e QTL with marine alleles that lead to more freshwater-like phenotypes and vice-versa) (Liu et al. 2022). We lifted the coordinates of the QTL windows from v1 to v5 of the stickleback genome using the chain file "v1_withChrUn_to_v5.chain.txt" available at the Stickleback Genome Browser (https://stickleback.genetics.uga.edu/). Similar to the EcoPeak enrichment analysis, we: 1) identified the proportion of DEGs and DSGs that overlapped with QTL; 2) identified the proportion of background genes that overlapped with QTL; and 3) tested if the proportion of DEGs and DSGs in QTL is significantly higher than the proportion of background genes using a permutation test (1000 permutations) in R. We did this for the complete set of QTL and also for the different phenotypic categories of QTL defined by Peichel and Marques (2017): defence, behaviour and sensory system, body shape, body size, swimming, feeding, pigmentation, and respiration (Supplementary Table S4). We did not test for QTL category enrichment for the different types of DSGs because we did not have enough genes for the analysis to have statistical power.

5.6 Effect of genetic distance in expression and splicing

To complement the EcoPeak enrichment analysis, we also looked for evidence of transcriptome-wide selection for stronger differential expression and splicing by looking at the correlation of splicing isoform difference and gene expression fold-change with genetic divergence between ecotypes. A significant positive correlation could mean transcriptome-wide selection for stronger differential expression (DE) and differential splicing (DS). To obtain a measure of genetic distance in both coding and non-coding regions across the stickleback genome, we used the SNP-level p-value data used to identify Pacific EcoPeaks by Kingman et al., 2021 (data kindly provided by the authors). These p-values result from a Fisher Exact Test for the probability of an imbalance in allele counts between ecotypes at each SNP (for more details consult Supplementary Section 9 in Kingman et al., 2021). A p-value < 0.05 means the SNP differs in allele frequencies between ecotypes and suggests divergent selection. As for the EcoPeak analysis, we translated all SNP coordinates from the v4.1 (*gasAcu1-4*) genome to

the v5 coordinates. Then, for each gene in the transcriptome, we calculated a gene-level genetic divergence based on the average p-value of the SNPs overlapping each gene. With this data we tested if there was a correlation between genetic distance and strength of differential expression (using the log2fold-change metric from EdgeR) or splicing (measured as isoform difference, as described above) using all the genes tested in the differential expression and differential splicing analyses, separately. In the case of genes with more than one alternative splicing event, the isoform difference of the most differentially spliced event between ecotypes was selected. To test the significance of the correlations we used a non-parametric linear model based on ranks implemented in the *R* package *Rfit* v0.24.2 (Kloke and McKean 2012). We also did this analysis separately for genes inside and outside of EcoPeaks.

5.7 Comparisons between types of AS

To see whether selection could be acting differently on particular types of AS, we compared both the genetic distance distributions and isoform difference distributions of the five types of AS events. DSGs were categorized into the five types of AS based on their most differential AS event. Then, we calculated the median of the isoform difference and genetic distance (average SNP p-value per gene) distributions for the five categories of DSGs, and we did pairwise tests on the difference of the medians using a Mann Whitney U Test, as implemented in the function wilcox.test() from the *R* package *stats* v3.6.1. In addition, we asked whether DSGs had a higher genetic distance than genes with alternative splicing but no significant differential splicing between ecotypes. For all genes with each type of AS event, we compared the median of the genetic distance distributions of the DSGs and the non-DSGs. We tested if this difference was significant with a permutation test (10 000 permutations).

5.8 Gene Ontology enrichment analysis

We tested whether DSGs and DEGs were enriched in specific biological processes, through a Gene Ontology Enrichment analysis. We used the g:Profiler (Reimand et al. 2007) as implemented in the *R* package *gprofiler2* v0.2.2. We used the *G. aculeatus* functional annotations from Ensembl implemented in *gprofiler* and used custom backgrounds for the statistical analysis. As for the EcoPeak and QTL enrichment analysis, the background for the DEGs was the set of genes that passed the minimum expression, and the background for the DSGs was the set of genes with evidence of AS that passed the minimum expression filter junction filtering process (i.e the set of genes that were tested for DE by EdgeR and DS by rMATs). In addition, we also tested the enrichment of the subset of DEGs and DSGs that overlapped both Pacific EcoPeaks and QTL. Finally, we manually looked at the gene

function of DSGs that overlapped both Global EcoPeaks and QTL in the GeneCards database (https://www.genecards.org/), which integrates information from several databases including NCBI Gene (https://www.ncbi.nlm.nih.gov/gene) and UniProt (https://www.uniprot.org/).

Data availability

All RNAseq data used in this study was already publicly available at NCBI's Sequence Read Archive (SRA) under BioProjects PRJNA371616 and PRJNA530695. Accession numbers for all samples used can be found in Supplementary Table S1. The EcoPeak data can be obtained from the UCSC Genome Browser (http://genome.ucsc.edu/) (instructions for downloading the data are in Kingman et al., 2021, under the "Data and materials availability" section). QTL data can be found in the supplementary material of Rennison and Peichel, 2022 and Liu et al., 2022. Scripts used for the analysis will be submitted to GitHub before publication.

Acknowledgements

We are grateful to: Daniel Jeffries for very helpful discussion and feedback on the manuscript; Mike White for providing valuable chain files to lift the EcoPeaks from their v4.1 (*gasAcu*1-4) genomic coordinates to version 5 coordinates; Garrett Roberts Kingman, Krishna Veeramah and David Kingsley for kindly providing the SNP P-value data for our genetic distance analyses; Virginie Courtier-Orgogozo for kindly suggesting and helping with the search of phenotypes caused by splicing changes in GePheBase; and all Peichel Lab member for multiple helpful discussions during the realization of this study. This work was funded by the University of Bern.

Supplementary Tables

Supplementary Tables for this chapter can be found in the following OneDrive Link:

https://unibe365-my.sharepoint.com/:f:/r/personal/carlos_rodriguezramirez_unibe_ch/Documents/DoctoralThesis_RodriguezRamirez_SupTables?csf=1&web=1&e=FOhHEN

Figures

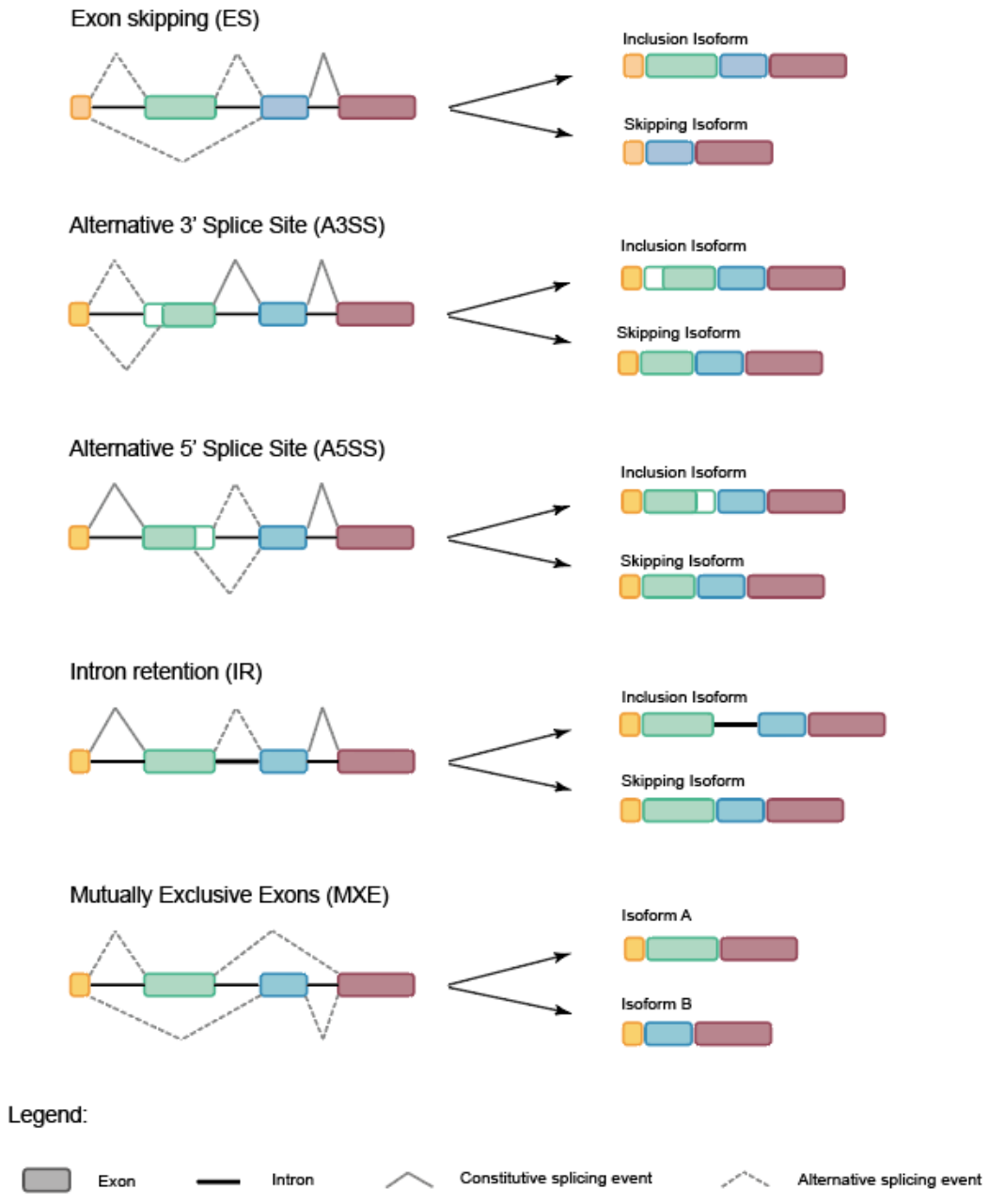


Figure 1. Different types of alternative splicing events. The DNA sequences are indicated to the left of the arrows, and the resulting mRNA sequences are indicated to the right. rMATs classifies one isoform as the "inclusion isoform" and another as the "skipping isoform", then it counts all unambiguous reads mapping to each isoform to calculate the isoform inclusion difference (i.e change in the proportion of inclusion isoforms relative to skipping isoforms between ecotypes). For MXE splicing, the "inclusion" label is given to the isoform with the 5'-most exon.

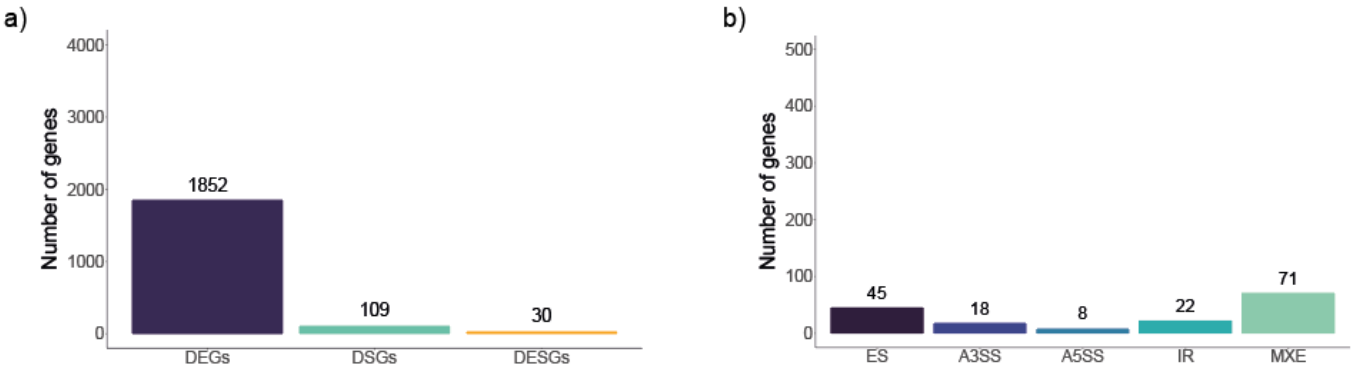


Figure 2. a) Number of differentially expressed genes (DEGs), differentially spliced genes (DSGs), and differentially expressed and spliced genes (DESGs) in the gill transcriptome. DEGs and DSGs counts in the figure do not include DESGs. b) Number of differentially spliced events of each AS type found within the 139 DSGs (including DESGs) in the gill transcriptome.

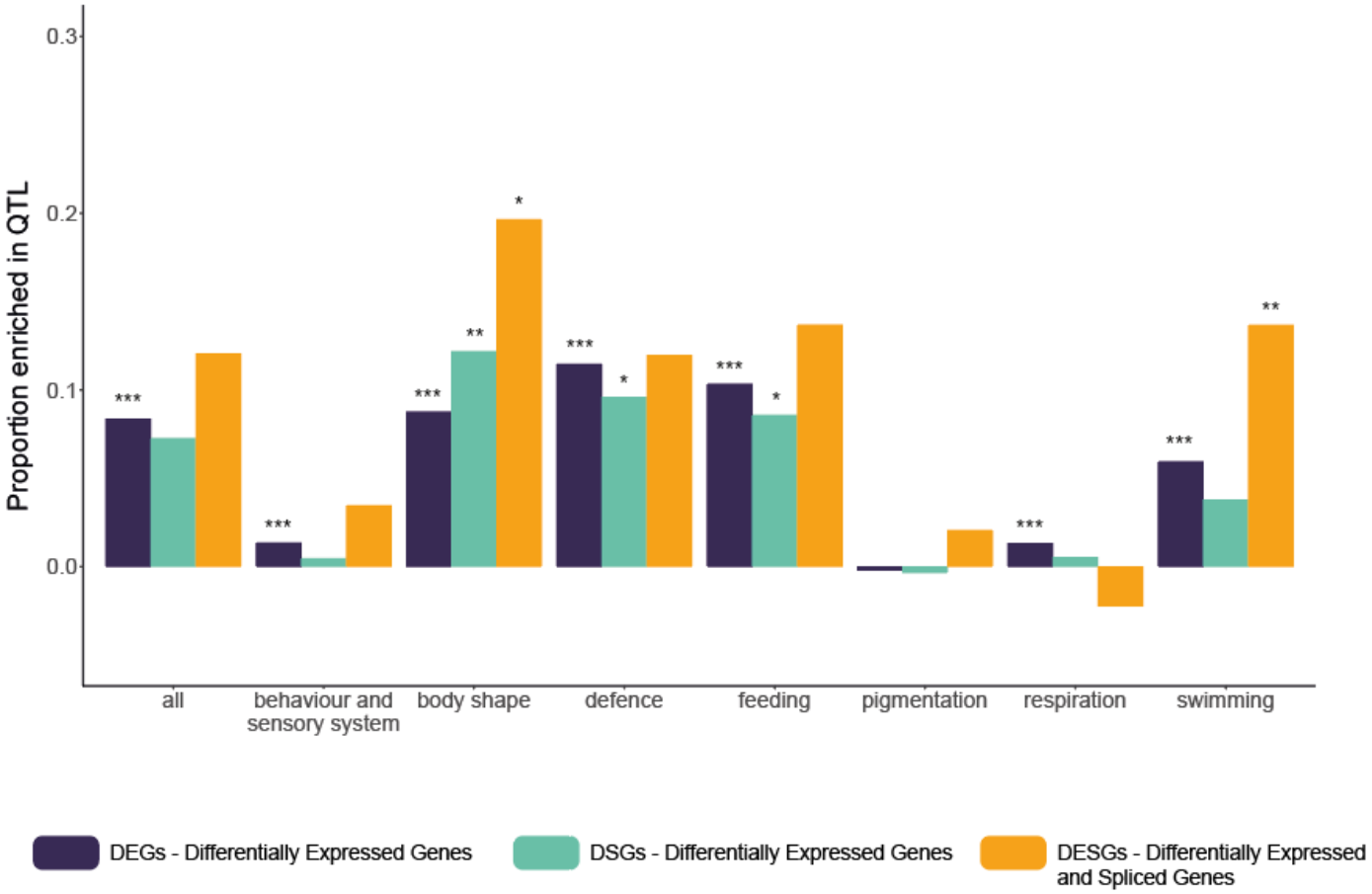


Figure 3. QTL enrichment analysis results for DEGs (dark blue), DSGs (aqua), and DESGs (orange). Bars represent the proportion of DEGs, DSGs, or DESGs in a QTL for a given category minus the proportion of background genes in a QTL for that category. Asterisks represent significance levels for the QTL enrichment test (permutation test, 1000 permutations): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

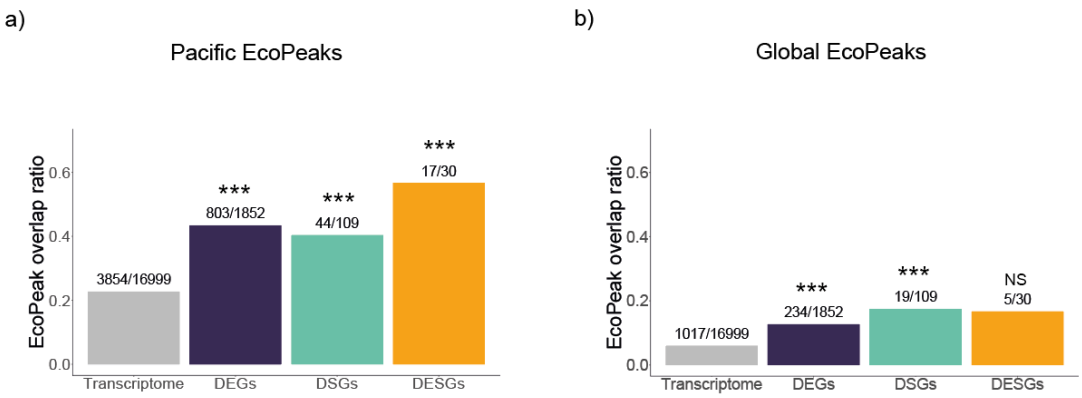


Figure 4. Proportion of DEGs, DSGs, DESGs and transcriptome genes overlapping Pacific (a) and Global (b) EcoPeaks. Transcriptome genes are all genes that passed the minimum expression filter (i.e the background of the DEGs). DSGs and DESGs were compared to their respective backgrounds, but for simplicity they are not represented in the figure (Supplementary Table S6). The number of genes of each category inside the EcoPeaks relative to the total number of genes in that category are shown. Asterisks represent significance levels for the EcoPeak enrichment test (permutation test, 1000 permutations): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; NS – not significant, p-value > 0.05.

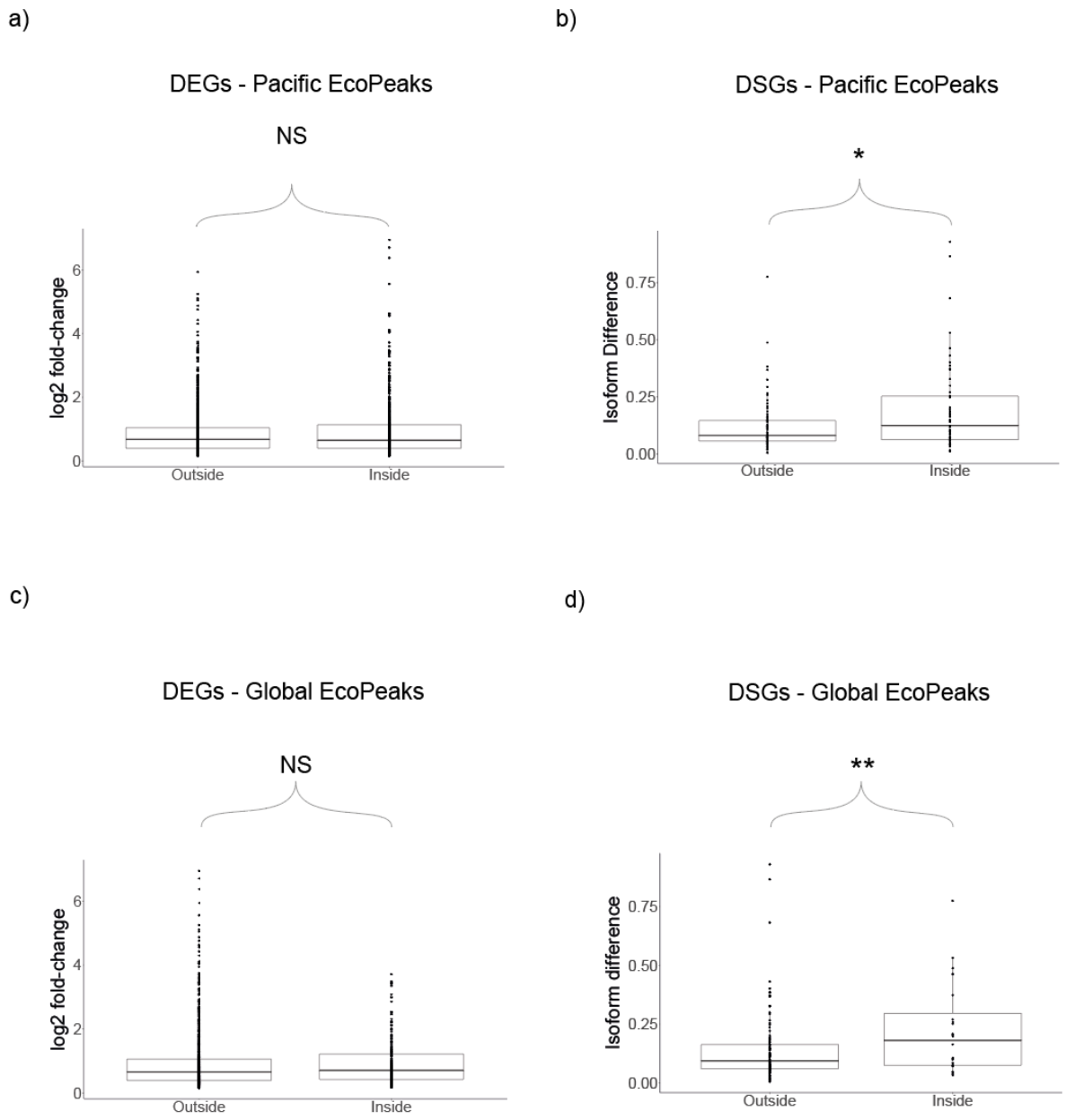


Figure 5. Difference in fold-change (a and c) and isoform difference (b and d) of DEGs and DSGs inside and outside of Pacific (a and b) or Global (c and d) EcoPeaks. Asterisks represent significance levels for the difference of the medians (permutation test, 1000 permutations): * p-value < 0.05; ** p-value < 0.01; NS – not significant, p-value > 0.05.

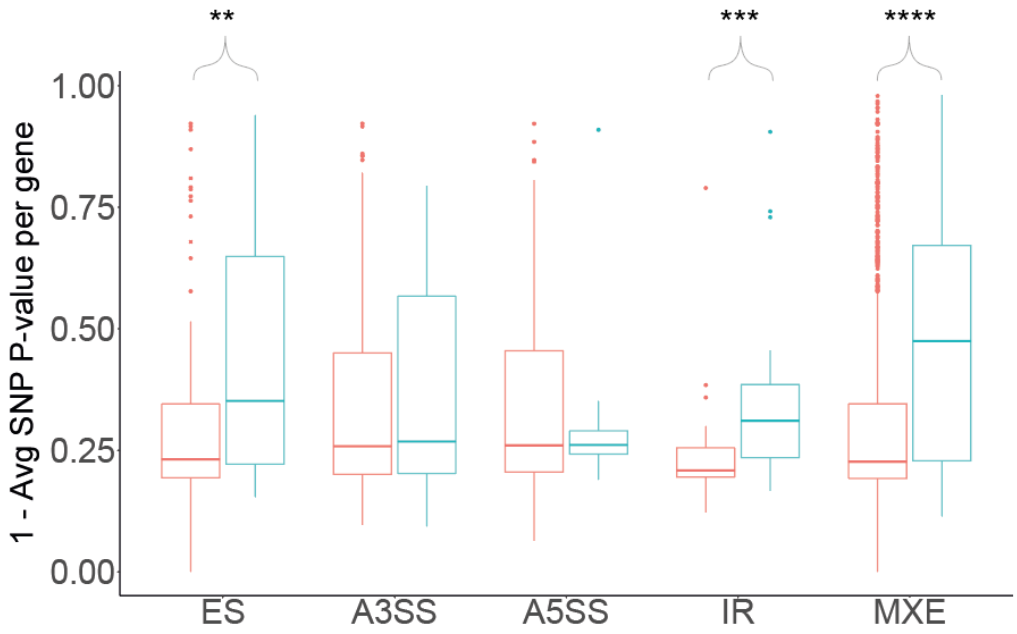
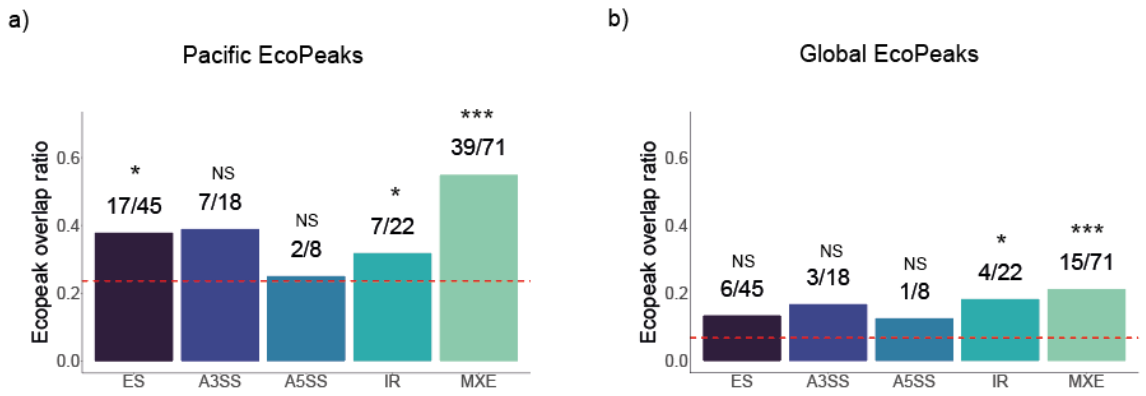
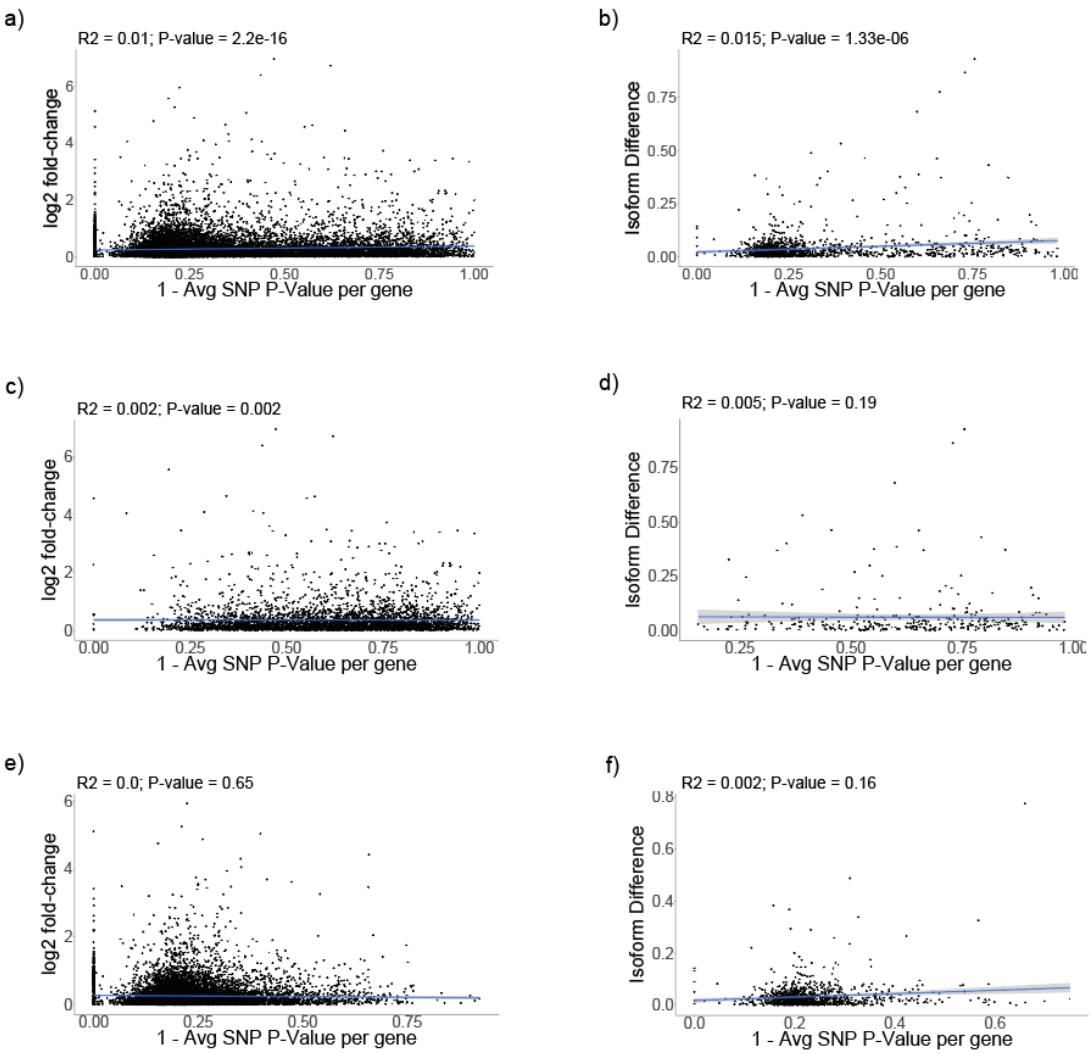


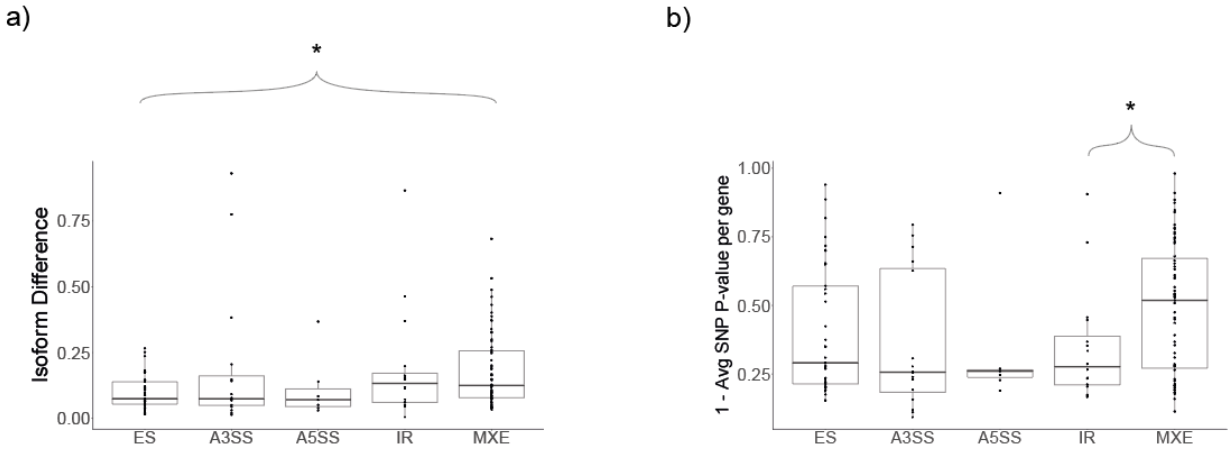
Figure 6. Distributions of average SNP p-value for significant DSGs (blue) and non-DSGs (red) per splicing type. Non-DSGs are genes that are alternatively spliced but not significantly differentially spliced between marine and freshwater samples. To make the Y axis more intuitive, the average SNP p-value is subtracted from 1 so that larger values represent greater genetic divergence. Asterisks indicate whether DSGs have a significantly higher genetic divergence median than non-DSGs (permutation test, 10,000 permutations): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; **** p-value < 0.0001.



Supplementary Figure S1. Proportion of differential splicing events of each type found in genes within Pacific EcoPeaks (a) or Global EcoPeaks (b). Red dashed line represents the proportion of genes in the transcriptome that are found inside EcoPeaks. On top of each bar are shown the proportions of significantly divergent splicing events inside the EcoPeaks versus the total number of divergent splicing events of that type. Asterisks represent significance levels for the EcoPeak enrichment test (permutation test, 1000 permutations): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; NS – not significant, p-value > 0.05.



Supplementary Figure S2. Correlation between SNP average p-value and log2 fold-change (a, c, e) or isoform difference (b, d, f) for all background genes of the Pacific DE and DS analysis (a and b), only the background genes inside Pacific EcoPeaks (c and d), or only background genes outside of Pacific EcoPeaks (e and f). We subtracted the average SNP p-value to 1 so that higher values on the X axis represent higher genetic divergence of the SNPs in the genes.



Supplementary Figure S3. Comparisons of the distribution of isoform differences (a) and SNP average p-value (b) between DSGs with different types of alternative splicing. Only the strongest splicing event per DSG was considered. In panel b) we subtracted the average SNP p-value from 1 so that larger values on the y-axis represent greater genetic divergence. Asterisks represent significant differences in the medians of the distributions according to a Mann Whitney U Test (p-value < 0.05). All pairwise combinations were tested, and only significant differences are shown.

References

Barroso GV, Puzovic N, Dutheil JY. 2018. The evolution of gene-specific transcriptional noise is driven by selection at the pathway level. *Genetics* 208:173–189.

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

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Bomblies K, Peichel CL. 2022. Genetics of adaptation. *Proc Natl Acad Sci U S A.* 119:e2122152119.

Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO. 2017. Alternative splicing and the evolution of phenotypic novelty. *Phil Trans R Soc B.* 372:20150474.

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol.* 3:e245.

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327:302–305.

Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. 2014. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol Biol Evol.* 31:1402–1413.

Chen Y, McCarthy D, Ritchie M, Robinson M, Smyth G. 2008. edgeR: differential analysis of sequence read count data - User's Guide.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307:1928–1933.

Courtier-Orgogozo V, Arnoult L, Prigent SR, Wiltgen S, Martin A. 2020. Gephebase, a database of genotype–phenotype relationships for natural and domesticated variation in Eukaryotes. *Nucleic Acids Res.* 48:D696–D703.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.

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.* 26:2711–2725.

Gracheva EO, Cordero-Morales JF, González-Carcacía JA, Ingolia NT, Manno C, Aranguren CI, Weissman JS, Julius D. 2011. Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* 476:88–91.

Grantham ME, Brisson JA. 2018. Extensive differential splicing underlies phenotypically plastic aphid morphs. *Mol Biol Evol.* 35:1934–1946.

Hill MS, Vande Zande P, Wittkopp PJ. 2021. Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet.* 22:203–215.

Hillmer EJ, Zhang H, Li HS, Watowich SS. 2016. STAT3 signaling in immunity. *Cytokine Growth Factor Rev.* 31:1–15.

Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995–1016.

Howes TR, Summers BR, Kingsley DM. 2017. Dorsal spine evolution in threespine sticklebacks via a splicing change in MSX2A. *BMC Biol*. 15:115.

Indjeian VB, Kingman GA, Jones FC, Guenther CA, Grimwood J, Schmutz J, Myers RM, Kingsley DM. 2016. Evolving new skeletal traits by cis -regulatory changes in bone morphogenetic proteins. *Cell* 164:45–56.

Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A, et al. 2019. A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* 364:886–889.

Jacobs A, Elmer KR. 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol Ecol.* 30:4955–4969.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.

Kloke JD, McKean JW. 2012. Rfit: Rank-based estimation for linear models. *The R Journal* 4:57–64.

Kondrashov FA. 2001. Origin of alternative splicing by tandem exon duplication. *Hum Mol Genet.* 10:2661–2669.

Lawrence M, Gentleman R, Carey V. 2009. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 25:1841–1842.

Letunic I. 2002. Common exon duplication in animals and its role in alternative splicing. *Hum Mol Genet.* 11:1561–1567.

Lev-Maor G, Goren A, Sela N, Kim E, Keren H, Doron-Faigenboim A, Leibman-Barak S, Pupko T, Ast G. 2007. The "alternative" choice of constitutive exons throughout evolution. *PLoS Genet.* 3:e203.

Levy DE, Lee C. 2002. What does Stat3 do? *J Clin Invest.* 109:1143–1148.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.

Lin L, Jiang P, Park JW, Wang J, Lu Z, Lam MPY, Ping P, Xing Y. 2016. The contribution of Alu exons to the human proteome. *Genome Biol.* 17:15.

Liu Z, Roesti M, Marques D, Hiltbrunner M, Saladin V, Peichel CL. 2022. Chromosomal fusions facilitate adaptation to divergent environments in threespine stickleback. *Mol Biol Evol.* 39:msab358.

Manahan DN, Nachman MW. 2024. Alternative splicing and environmental adaptation in wild house mice. *Heredity* 132:133–141.

Maritano D, Sugrue ML, Tininini S, Dewilde S, Strobl B, Fu X, Murray-Tait V, Chiarle R, Poli V. 2004. The STAT3 isoforms α and β have unique and specific functions. *Nat Immunol.* 5:401–409.

Martin A, Orgogozo V. 2013. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* 67:1235–1250.

Martinez-Gomez L, Abascal F, Jungreis I, Pozo F, Kellis M, Mudge JM, Tress ML. 2020. Few SINEs of life: Alu elements have little evidence for biological relevance despite elevated translation. *NAR Genom Bioinform.* 2:lqz023.

Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, Kingsley DM. 2007. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* 131:1179–1189.

Nath S, Shaw DE, White MA. 2021. Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 (Bethesda)* 11:jkab007.

O'Shea JJ, Gadina M, Schreiber RD. 2002. Cytokine signaling in 2002. *Cell* 109:S121–S131.

Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Phil Trans R Soc B.* 372:20150486.

R Core Team. 2019. R: A language and environment for statistical computing. Available from: https://www.R-project.org/

Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB. 2009. Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* 326:1663–1667.

Reimand J, Kull M, Peterson H, Hansen J, Vilo J. 2007. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* 35:W193–W200.

Rennison DJ, Peichel CL. 2022. Pleiotropy facilitates parallel adaptation in sticklebacks. *Mol Ecol.* 31:1476–1486.

Richard M, Yvert G. 2014. How does evolution tune biological noise? *Front Genet.* 5:374.

Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, et al. 2021. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci Adv.* 7:eabg5285.

Rodríguez-Ramírez CE, Hiltbrunner M, Saladin V, Walker S, Urrutia A, Peichel CL. 2023. Molecular mechanisms of *Eda* -mediated adaptation to freshwater in threespine stickleback. *Mol Ecol.* doi: 10.1111/mec.16989.

Rogers TF, Palmer DH, Wright AE. 2021. Sex-specific selection drives the evolution of alternative splicing in birds. *Mol Biol Evol*. 38:519–530.

Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A.* 111.

Singh P, Ahi EP. 2022. The importance of alternative splicing in adaptive evolution. *Mol Ecol.* 31:1928–1938.

Singh P, Börger C, More H, Sturmbauer C. 2017. The role of alternative splicing and differential gene expression in cichlid adaptive radiation. *Genome Biol Evol.* 9:2764–2781.

Smith CCR, Tittes S, Mendieta JP, Collier-zans E, Rowe HC, Rieseberg LH, Kane NC. 2018. Genetics of alternative splicing evolution during sunflower domestication. *Proc Natl Acad Sci U S A.* 115:6768–6773.

Sorek R. 2007. The birth of new exons: Mechanisms and evolutionary consequences. *RNA* 13:1603–1608.

Sorek R, Ast G, Graur D. 2002. *Alu* -containing exons are alternatively spliced. *Genome Res.* 12:1060–1067.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62:2155–2177.

Steward RA, de Jong MA, Oostra V, Wheat CW. 2022. Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nat Commun.* 13:755.

Sun W, You X, Gogol-Döring A, He H, Kise Y, Sohn M, Chen T, Klebes A, Schmucker D, Chen W. 2013. Ultra-deep profiling of alternatively spliced *Drosophila* Dscam isoforms by circularization-assisted multi-segment sequencing. *EMBO J.* 32:2029–2038.

Tovar-Corona JM, Castillo-Morales A, Chen L, Olds BP, Clark JM, Reynolds SE, Pittendrigh BR, Feil EJ, Urrutia AO. 2015. Alternative splice in alternative lice. *Mol Biol Evol.* 32:2749–2759.

Verta J-P, Jacobs A. 2022. The role of alternative splicing in adaptation and evolution. *Trends Ecol Evol.* 37:299–308.

Verta J-P, Jones FC. 2019. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *Elife* 8:e43785.

Wang X, Yang M, Ren D, Terzaghi W, Deng X, He G. 2019. *Cis* -regulated alternative splicing divergence and its potential contribution to environmental responses in Arabidopsis. *Plant J.* 97:555–570.

Wegrzyn J, Potla R, Chwae Y-J, Sepuri NBV, Zhang Q, Koeck T, Derecka M, Szczepanek K, Szelag M, Gornicka A, et al. 2009. Function of mitochondrial Stat3 in cellular respiration. *Science* 323:793–797.

Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet.* 13:59–69.

Wong JJ -L., Au AYM, Ritchie W, Rasko JEJ. 2016. Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. *BioEssays* 38:41–49.

Wooldridge TB, Kautt AF, Lassance J-M, McFadden S, Domingues VS, Mallarino R, Hoekstra HE. 2022. An enhancer of *Agouti* contributes to parallel evolution of cryptically colored beach mice. *Proc Natl Acad Sci U S A.* 119:e2202862119.

Wright CJ, Smith CWJ, Jiggins CD. 2022. Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet*. 23:697-710.

Xia B, Zhang W, Zhao G, Zhang X, Bai J, Brosh R, Wudzinska A, Huang E, Ashe H, Ellis G, et al. 2024. On the genetic basis of tail-loss evolution in humans and apes. *Nature* 626:1042–1048.

Xing Y, Lee C. 2006. Alternative splicing and RNA selection pressure — evolutionary consequences for eukaryotic genomes. *Nat Rev Genet.* 7:499–509.

Xiong S, Wang W, Kenzior A, Olsen L, Krishnan J, Persons J, Medley K, Peuß R, Wang Y, Chen S, et al. 2022. Enhanced lipogenesis through Pparγ helps cavefish adapt to food scarcity. *Curr Biol.* 32:2272-2280.e6.

Zhang Y, Parmigiani G, Johnson WE. 2020. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* 2:lqaa078.

Zhou S, Dai Q, Huang X, Jin A, Yang Y, Gong X, Xu H, Gao X, Jiang L. 2021. STAT3 is critical for skeletal development and bone homeostasis by regulating osteogenesis. *Nat Commun.* 12:6891.

General discussion

When I started studying evolutionary biology, I learned about many beautiful examples of adaptive phenotypes and their genetic basis in systems like stickleback, deer mice, *Drosophila* or peppered moth. At that point it was easy to believe that we had already "figured out" adaptation and we understood its genetic basis. This idea was reinforced by the increasing number of cases where candidate loci underlying phenotypic change were being identified thanks to advances in sequencing technologies. However, when I started working in this fascinating field of biology, I quickly learned that there are still many "black boxes" in our knowledge about the genetics of adaptation. While we have increased substantially our knowledge of the genetic architecture of phenotypic variation in the wild, in most cases we still do not know nearly as much about the mechanisms through which these loci affect the phenotype, particularly in non-model species (Bomblies and Peichel 2022; Kitano et al. 2022). Not to mention that for most phenotypes, their adaptive status is putative since their fitness effects have not been tested empirically, making it unclear whether they are the true target of natural selection or whether another genetically linked phenotype is (Barrett and Hoekstra 2011; Bomblies and Peichel 2022). Proper understanding the molecular mechanism of adaptation is essential to understand why particular genes are used in adaptation (Bomblies and Peichel 2022). Was it just chance, and other genes in the same gene pathway could have played the same role? Or are there constraints or benefits that favoured the use of a specific gene? While the answer to these questions likely depends on a case-by-case basis, understanding under which circumstances each scenario is more likely to happen is essential to further our understanding of the genetic basis of adaptation and to consider ambitious enterprises like being able to predict adaptation at the genetic level.

Threespine sticklebacks are a great example of the progress but also the limitations in our knowledge of the genetics of adaptation. Using forward genetic approaches, we have identified many quantitative trait loci (QTL) in the genome underlying phenotypic variation between marine and freshwater populations (Peichel and Marques 2017). Using reverse genetic approaches, we have also managed to identify many regions in the genome with signatures of divergent selection (Jones et al. 2012; Roberts Kingman et al. 2021). Using genetic manipulation experiments we have managed to identify the genes underlying some of these phenotypic changes (Colosimo et al. 2005; Miller et al. 2007; Chan et al. 2010; Cleves et al. 2014; Indjeian et al. 2016; Howes et al. 2017). However, there are still many gaps on our knowledge of the marine-freshwater adaptive divergence in threespine stickleback. For example, most of the phenotypes that have been studied in threespine stickleback are external morphological phenotypes. We know very little about the genetic basis of adaptation for physiological or behavioral traits (but see Greenwood et al. 2016 and Ishikawa et al. 2019). Even for the morphological phenotypes, we have identified many QTL controlling their variation but are still far

from identifying candidate genes (Peichel and Marques 2017). Furthermore, while these phenotypes show consistent changes between marine and freshwater populations, actual tests of the fitness effects of these phenotypes are missing for the most part.

However, threespine stickleback is a powerful system to address these questions. It has replicated instances of freshwater adaptation; rich ecological knowledge of this system; well-developed genetic and genomic tools; and amenability to lab conditions. This made threespine stickleback a really exciting and fun system to work with during my PhD, allowing me to explore different questions on the genetic basis of adaptation. Here now I discuss the main findings of my PhD, what they teach us about the genetic basis of the marine-freshwater divergence in threespine stickleback, as well as what they teach us about genetic basis of adaptation in general. I also discuss what I think are interesting venues for future research based on these results.

The molecular mechanisms of the phenotypic effects of *Eda*

Eda is a classic example in the field of the genetic basis of adaptation. It is a large effect gene that controls most of the genetic variation (75-94%) in the number of anti-predator lateral plates (Colosimo et al. 2005; Archambeault, Durston, et al. 2020), and that it is strongly under divergent selection between marine and freshwater habitats (Barrett et al. 2008). We also know that it is a pleiotropic gene which also affects the patterning of the lateral line (Mills et al. 2014) and schooling behaviour (Greenwood et al. 2016). However, despite almost 20 years having passed since the discovery that *Eda* controls the loss of lateral plates in freshwater threespine (Colosimo et al. 2005), we know strikingly little about the molecular mechanisms that mediate the phenotypic effects of *Eda.* We know that: 1) the causative mutation should be cis-regulatory and located in a 1.4 kb region in the first intron of *Eda* (Colosimo et al. 2005; O'Brown et al. 2015; Archambeault, Durston, et al. 2020)*,* but the exact causal mutation is still a mystery*;* 2) *Eda* is a pleiotropic gene, with effects also in the patterning of the lateral line and schooling behaviour (Mills et al. 2014; Greenwood et al. 2016), but whether these pleiotropic effects are controlled independently or a direct result from each other is unclear; 3) *Eda* is found in linkage with two putative immune genes *Tnfsf13b* and *Garp*, but whether they play a role in marine-freshwater adaptation or not is unclear (Colosimo et al. 2005; Archambeault, Bärtschi, et al. 2020); and 4) the freshwater allele of *Eda* is down-regulated compared to the marine allele, and this seems to be caused by a lower response to *Wnt* signaling (Colosimo et al. 2005; O'Brown et al. 2015), but what gene pathways downstream of *Eda* mediate its phenotypic effects is unknown. In the studies described in Chapter 1 and 2 of this thesis I sought to address several of these questions.

In Chapter 1, I identified and tested a candidate causative mutation for the phenotypic effects of *Eda* within the 1.4 kb region in the first intron previously identified. This candidate, *LP3621* is a derived 16 bp deletion in the freshwater allele of *Eda* that overlaps the binding site of TFAP4, a transcription factor that regulates cell fate determination and other cellular processes in mammals (Wong et al. 2021) and whose family member has been linked to ecotypic divergence in craniofacial skeleton development (Erickson et al. 2018). However, despite having all the characteristics of a promising candidate gene, our CRISPR-Cas9 experiment found no effect of this region in the number of lateral plates. This came as quite a surprise to me at first, but then I realized that there are 18 other polymorphisms in the 1.4 kb region, many with overlapping binding sites for transcription factors. *LP3621* was simply the strongest candidate, but not the only one by far. Considering that after the NAKA SNP this is the second candidate mutation for the effects of *Eda* to be disproved (O'Brown et al. 2015; Archambeault et al. 2020), I believe that narrowing the causative region down further before testing more individual candidate mutations is essential. Particularly because there are two important possibilities that need to be considered: 1) that two or more closely linked mutations are behind the effects of *Eda;* and 2) that the 1.4 kb region identified by Archambeault et al. 2020 might not be causative at all. I believe the latter in particular to be an important possibility. As already mentioned in the discussion of Chapter 1, the hypothesis of the causality of the 1.4 kb region is based on a single recombinant fish found by Archambeault et al. (2020), and there is the risk that unlinked modifier loci could have been present in this individual. A more conservative causative region is the larger 2.4 kb region that had the strongest association to variation on the number of lateral plates and neuromast patterning in this study.

However, expanding the causative region inevitably increases even more the number of candidate causative mutations, highlighting the need to narrow down the causative region in order to find the causative mutation. This could be done by dividing the 2.4 kb region into smaller sub-regions and testing their causality with CRISPR-Cas9 genetic manipulation. For example, by swapping the freshwater allele of each region into the marine allele of *Eda* heterozygous fish and testing if this causes any phenotypic changes. Once we identify the sub-region of the 2.4 kb region that is causal, this procedure can be repeated until we have few enough mutations to test individually, with hopefully some stronger candidates than others. Furthermore, if multiple mutations are responsible for the phenotypic effects of *Eda,* it might be easier to detect with this approach than if we were to test the mutations one by one.

Chapter 2 of the thesis addressed a different but complementary question about *Eda*: regardless of which is the causative mutation, what are the molecular pathways that it affects to cause the

phenotypic effects of *Eda*? To answer this question, we performed a RNAseq experiment where we sequenced individuals with the three *Eda* genotypes (CC, CL and LL) in an otherwise marine genomic background. The only exception were the two other genes in the *Eda* haplotype, *Tnfsf13b and Garp,* which we did not unlink from *Eda,* preserving the haplotype. We did this to test whether there was any effect of the haplotype in immunity that could be linked these genes and explain why they tend to be linked to *Eda*.

The results of this study were far more enlightening than I had ever hoped. While at first, the list of over two hundred genes looked messy, with the help of the GO enrichment analysis, the gene co-expression analysis, and a good dose of literature research, a clearer picture of what we were looking at started to emerge. I found that the haplotype was affecting several members of the BMP pathway, a conserved pathway involved in bone development and the development of ectodermal appendages (Cui and Schlessinger 2006; Sadier et al. 2014; Wang et al. 2014), which is a strong candidate to mediate the effects of *Eda* on the number of lateral plates. I also found strong candidates for the effect of *Eda* in the patterning of the lateral line. I found an effect of the haplotype in genes associated to neuronal development (netrins) and also in genes specifically involved in neuromast or cell hair development. Interestingly, several of these genes formed their own co-expression module, hinting at a potential regulatory network for neuromast development affected by the haplotype. Furthermore, I also found an effect of the haplotype in immune genes in skin, and in a few genes in head kidney, an important immune organ in fish that plays a role akin to the bone marrow in mammals (Soulliere and Dixon 2017). These results support the possibility that *Tnfsf13* or *Garp* might be kept on the *Eda* haplotype because they mediate important immune effects between ecotypes. One of the immune genes that were differentially expressed in skin (*Laptm4b*) even codes for a protein that is known to interact with the GARP protein in mammals (Huygens et al. 2015). However, more studies are necessary to know which one of the three genes in the haplotype are affecting these immune genes, and whether their effect has any adaptive value in the marine-freshwater divergence of threespine stickleback.

Finally, I find it very interesting that the *Eda* haplotype affects genes in highly pleiotropic pathways like Fgf, Wnt and Notch. As mentioned in the discussion of this chapter, these three pathways have been described to have effects in both in scale development and lateral line development in zebrafish. The same is true for bradykinins, which are important inflammatory molecules (Kaplan et al. 2002; Marceau and Regoli 2004), but have also been implicated in bone homeostasis (Lerner et al. 1987; Epsley et al. 2021). These results suggest a potential molecular mechanism to explain the pleiotropic effects of *Eda.* Instead of *Eda* directly regulating different pathways involved in different phenotypes,

it might potentially be affecting always the same pleiotropic pathways, which then, depending on the tissue and cellular environment, might have different phenotypic effects.

I believe that the results of these studies also have interesting implications for our understanding of the genetics of adaptation in general. The results from Chapter 1 on our candidate causative mutation highlight how seemingly strong candidate mutations might turn out to have no effect on the phenotype under study. Considering that most genes and mutations underlying phenotypic variation in the wild have been identified through association studies but not tested empirically, it is important to keep in mind that a lot of our knowledge on the genetic basis of adaptation remains empirically untested, and there could be many wrong candidates. This has the potential to be a bias on our understanding of the genetics of adaptation (Barrett and Hoekstra 2011).

Likewise, the results of Chapter 2 highlight how conserved and pleiotropic developmental pathways can be used to drive phenotypic change, even when they are not directly under selection. A better understanding of the constraints between selection acting directly in genes within these conserved pathways (like the role of *Bmp6* role in pharyngeal tooth variation in stickleback (Cleves et al. 2014)) or acting in genes in adjacent but interacting pathways (like *Eda*'s effect on the lateral plates) is essential to properly understand the genetic basis of adaptation.

The role of alternative splicing and gene expression in adaptive evolution

Alternative splicing is a fascinating molecular mechanism. It takes genes, which otherwise would be static and inflexible blocs of information and makes them into interactive and modular sources of information that can change this information based on changes in their environment. Alternative splicing is also interesting because despite being regulated by cis-regulatory elements, it directly impacts the coding sequence, often probably more than most non-synonymous mutations. This puts alternative splicing mutations in an interesting intermediate position between cis-regulatory mutations affecting gene expression and coding mutations. This is why I was surprised when I started to work in transcriptomics and found that, for the most part, this really cool mechanism has been largely ignored in evolutionary genetics. The reasons for this seem to be partly because of lack of awareness about alternative splicing in the field and partly because of a perception that most alternative splicing isoforms are not functional. This views have changed in recent years with increasingly more studies showing evidence for a role of alternative splicing in adaptive evolution (Bush et al. 2017; Singh and Ahi 2022; Verta and Jacobs 2022; Wright et al. 2022). However, despite this progress, we still know relatively little about the role of alternative splicing in adaptive evolution.

Like how often does natural selection act directly on alternative splicing? And when it does, how is alternative splicing used? Is it mostly used to promote nonsense-mediated decay (NMD) of the transcripts of a gene and thus to indirectly down-regulate it? Or does selection often use the potential of alternative splicing to increase protein diversity? What kind of constraints favour, or not, the use of alternative splicing as mechanism for adaptive evolution?

I sought to address some of these questions in Chapters 2 and 3 of this thesis. In Chapter 2, I studied alternative splicing as a regulatory mechanism that mediates the phenotypic effects of an adaptive gene, while in Chapter 3, I explored whether alternative splicing itself can be the target and driver of divergent adaptive evolution. While much work remains to be done, I believe these two studies gave us some very interesting insights into these questions. I found differentially spliced genes in both studies, supporting the hypothesis that it is an ubiquitous mechanism in adaptive evolution. While the number of differentially spliced genes (DSGs) that I identified was in general always much less than the number of differentially expressed genes (DEGs) (with the notable exception of head kidney in Chapter 2), this unlikely to be the true relative frequency of DSGs to DEGs. While different methods were used in the two studies to infer changes in alternative splicing, both methods had limitations in their power to detect DSGs. An interesting pattern found in both studies is that DEGs and DSGs are for the most part non-overlapping. This is consistent with previous studies in artic charr and house mouse (Jacobs and Elmer 2021; Manahan and Nachman 2024), though studies in birds and butterflies have found a greater overlap of DEGs and DSGs (Rogers et al., 2021; Steward et al., 2020). My results suggest that in threespine stickleback, these molecular mechanisms are also for the most part independent of each other and used independently to regulate a gene's function.

The results of the DSGs downstream of *Eda* in Chapter 2 suggest that characteristics of the genes like their level of pleiotropy and average expression level could be correlated with the gene being differentially expressed or spliced. Surprisingly, I found no evidence for a higher than average pleiotropy of DSGs, contrary to previous studies (Jacobs and Elmer 2021; Rogers et al. 2021). Our results therefore did not support the hypothesis that alternative splicing might be a mechanism to regulate pleiotropic genes while minimizing pleiotropic effects (Singh and Ahi 2022; Verta and Jacobs 2022; Wright et al. 2022). I did, however, find evidence that DEGs had a higher pleiotropy than average. As discussed in Chapter 2, it is unclear whether our result is due to limitations on the proxy of pleiotropy used (gene connectivity) or the context of the genes being studied (DSGs downstream of an adaptive haplotype, rather than DSGs diverging between ecotypes), or if they indeed reflect lineage-specific differences on the use of alternative splicing. Taken at face value, our results would suggest that differential expression is favored as a regulatory mechanism of highly pleiotropic genes, while alternative splicing is favored to regulate highly expressed genes. It is possible that this is true

and that for example, even though alternative splicing could preserve the original protein of a gene, this is not enough to mitigate pleiotropic effects of alternative isoforms in highly pleiotropic genes. Likewise, it is possible that highly expressed genes are more sensitive to changes in their gene expression levels, and thus other mechanisms of regulation like alternative splicing are favored.

However, as discussed in Chapter 2, it is possible that gene co-expression connectivity is not a good proxy for pleiotropy in alternatively spliced genes. If the isoforms produced by a pleiotropic gene have very different functions, it is possible that their co-expression patterns are very different. A gene-level co-expression analysis would lose this signal by merging together the expression patterns of the different isoforms. This is supported by our results on the co-expression patterns of individual exons on the DSG *Rmnd5b.* Here I found evidence that that exon 1, which is differentially spliced between different *Eda* haplotypes, has a different co-expression pattern to the rest of the gene, which is only revealed when looking at this exon individually. However, co-expression analysis is a very computationally demanding analysis even at the gene-level, since it involves calculating the pairwise correlations of all genes in the transcriptome. Thus, an exon-level analysis likely increases the computational resources necessary in several orders of magnitude. While I experimented with this for a little bit, unfortunately I was not able to figure out the resources necessary for this kind of analysis in time to add it to the studies in Chapter 2 and 3.

Even without a pleiotropy analysis, Chapter 3 revealed many interesting insights on the role of alternative splicing in the marine-freshwater divergence of threespine stickleback. As mentioned previously, there are several great examples of adaptive genes whose effects are mediated by changes in their alternative splicing, however it is still unclear whether these are isolated instances or part of a wider pattern of a recurrent contribution of alternative splicing to adaptive evolution. In this chapter I explored this question in the marine-freshwater adaptive divergence of threespine stickleback through three sub-questions: 1) are there differences in alternative splicing between marine and freshwater threespine stickleback populations?; 2) do these changes in alternative splicing underlie phenotypic divergence between the ecotypes?; and 3) are these changes in alternative splicing the result of natural selection? Of course, answering these questions completely requires more than a single study, however we sought to start exploring these questions combining publicly available RNAseq data on marine and freshwater populations with the extensive resources on the genetic basis of this divergence that already exist for threespine stickleback. Using this approach, I found evidence not only of many genes with diverging patterns of alternative splicing between ecotypes, but also that these DSGs were enriched in regions of the genome underlying phenotypic divergence between ecotypes as well as regions putatively under divergent selection. While more work is needed, our

results are consistent with the hypothesis that alternative splicing might be an important mechanism in the marine-freshwater divergence of threespine stickleback.

Furthermore, I found evidence that not all types of alternative splicing probably contribute equally to adaptation. Mutually exclusive exons (MXE) are the most common type of alternative splicing that diverged between ecotypes and is the type with the largest genetic divergence between DSGs, suggesting that DSGs with mutually exclusive exons result from divergent selection. This is a particularly exciting result in my opinion because I believe that MXE is the type of splicing that is better poised to mediate modular tinkering of the protein sequence through alternative splicing. Furthermore, MXE is thought to be associated with exon duplications (Kondrashov 2001; Letunic 2002; Wright et al. 2022). This makes MXE a great mechanism to first hide and later reveal genetic variation accumulating in duplicated exons, similarly to how gene duplications can promote new genetic variation to accumulate in one duplicate while another maintains the ancestral function. The fact that MXE turned out to be the most common and divergent type of splicing in our dataset adds support for it to play an important role in adaptive evolution. An interesting venue for future studies would be to explore whether MXE is happening in duplicated exons and whether these show evidence of relaxation of natural selection. Likewise, it would be interesting to use long-read RNA sequencing to compare isoforms resulting from differential MXE between ecotypes and have a better idea of how different the proteins translated from different isoforms truly are.

In conclusion, the results presented in this thesis suggest that alternative splicing is likely an important mechanism on the adaptive divergence of marine and freshwater stickleback. More studies are necessary to better elucidated the relationship of alternative splicing with pleiotropy and the contribution of different types of alternative splicing to the marine-freshwater divergence in threespine stickleback evolution. However, I believe the results presented in this thesis are promising enough to encourage such future studies both in stickleback and in other systems. Beyond the ideas already discussed, I believe there are two particularly interesting venues for future research on the role of alternative splicing in evolution. One of them is studies of alternative splicing in systems of experimental evolution, where we can track the frequency changes of alleles affecting alternative splicing to test empirically whether natural selection acts directly on alternative splicing. Another one is linked to the fact that alternative splicing can be regulated not only by intronic elements, but also by exonic enhancers and silencers (Lee and Rio 2015). This raises the interesting possibility that some coding mutations with apparently minimal effect on the coding sequence, might be having a greater impact by affecting the alternative splicing of genes. It would be very interesting to run mutational scan studies on alternatively spliced genes to try to assess the percentage of coding mutations that

might also be having an effect on the alternative splicing pattern of genes, effectively acting both as coding and cis-regulatory mutations.

Final thoughts

I believe that the results presented across the three chapters of this thesis have increased, even if by a little bit, our understanding of the genetic basis of adaptation in threespine stickleback. They suggest that conserved gene pathways BMP and regulatory mechanisms like alternative splicing might be important molecular mechanism mediating adaptive phenotypic change. Likewise, our results highlight the importance of testing candidate causal mutations and how strong candidates might turn out to not have any of the hypothesized phenotypic effects. I think these studies also highlight the strength of using other approaches beyond genomics to study the genetic basis of adaptation. While genomic studies can be an extremely powerful tool, by themselves are not enough to fully understand the molecular mechanisms underlying adaptive evolution. Genetic manipulation studies and transcriptomic studies can greatly help in this regard. While the usefulness of genetic manipulation studies is widely acknowledged, this is not always the case for transcriptomic studies. Concerns about plasticity, transcriptional noise and the volatility of the transcriptome can raise questions about its reliability and usefulness.

However, I think people often forget that the transcriptome is also a phenotype, just like the number of plates in a stickleback, or its behavioral patterns during mating season. In fact, the transcriptome is arguably the first phenotypic effect of the genome, and the bridge connecting the genome to all its other phenotypic effects. Its dynamic nature reflects the molecular and physiological nature of biological systems, and thus a strength rather than a downside on its usefulness as a tool to understand biological phenomena. At least, as long as we can design robust experiments and analyses that take in consideration the dynamic nature of the transcriptome. If we take this approach, I believe much can be learned of the genome-phenotype link by studying this first phenotypic effect of the genome, as attested by the many excellent transcriptomic studies on the genetic basis of adaptation cited across this thesis and, hopefully, also by the two transcriptomic studies in it.

References

Archambeault SL, Bärtschi LR, Merminod AD, Peichel CL. 2020. Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback Eda locus. Evol Lett 4:282–301.

Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO. 2017. Alternative splicing and the evolution of phenotypic novelty. Phil. Trans. R. Soc. B 372:20150474.

Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, Miller CT. 2014. Evolved tooth gain in sticklebacks is associated with a cis -regulatory allele of Bmp6. Proc. Natl. Acad. Sci. U.S.A. 111:13912–13917.

Cui C-Y, Schlessinger D. 2006. EDA signaling and skin appendage development. Cell Cycle 5:2477–2483.

Epsley S, Tadros S, Farid A, Kargilis D, Mehta S, Rajapakse CS. 2021. The effect of inflammation on bone. Front. Physiol. 11:511799.

Erickson PA, Baek J, Hart JC, Cleves PA, Miller CT. 2018. Genetic dissection of a supergene implicates Tfap2a in craniofacial evolution of threespine sticklebacks. Genetics 209:591–605.

Huygens C, Liénart S, Dedobbeleer O, Stockis J, Gauthy E, Coulie PG, Lucas S. 2015. Lysosomal-associated transmembrane protein 4B (LAPTM4B) decreases transforming growth factor β1 (TGF-β1) production in human regulatory T cells. J Biol Chem 290:20105–20116.

Jacobs A, Elmer KR. 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. Mol Ecol 30:4955–4969.

Kaplan AP, Joseph K, Silverberg M. 2002. Pathways for bradykinin formation and inflammatory disease. J Allergy Clin Immunol 109:195–209.

Kondrashov FA. 2001. Origin of alternative splicing by tandem exon duplication. Hum Mol Genet 10:2661–2669.

Lee Y, Rio DC. 2015. Mechanisms and regulation of alternative pre-mRNA splicing. Annu Rev Biochem 84:291–323.

Lerner UH, Jones IL, Gustafson GT. 1987. Bradykinin, a new potential mediator of inflammation-induced bone resorption. Studies of the effects on mouse calvarial bones and articular cartilage in vitro. Arthritis Rheum 30:530–540.

Letunic I. 2002. Common exon duplication in animals and its role in alternative splicing. Hum Mol Genet 11:1561–1567.

Manahan DN, Nachman MW. 2024. Alternative splicing and environmental adaptation in wild house mice. *Heredity* 132:133–141.

Marceau F, Regoli D. 2004. Bradykinin receptor ligands: therapeutic perspectives. Nat Rev Drug Discov 3:845–852.

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:e05290.

Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. Phil. Trans. R. Soc. B 372:20150486.

Rogers TF, Palmer DH, Wright AE. 2021. Sex-specific selection drives the evolution of alternative splicing in birds. Mol Biol Evol 38:519–530.

Sadier A, Viriot L, Pantalacci S, Laudet V. 2014. The ectodysplasin pathway: from diseases to adaptations. Trends Genet 30:24–31.

Steward, R. A., de Jong, M. A., Oostra, V., & Wheat, C. W. (2022). Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nature Communications*, *13*(1), 755. https://doi.org/10.1038/s41467-022-28306-8

Singh P, Ahi EP. 2022. The importance of alternative splicing in adaptive evolution. Mol Ecol 31:1928–1938.

Soulliere, C., & Dixon, B. (2017). Immune system organs of bony fishes. In *Reference Module in Life Sciences* (p. B978012809633812179X). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.12179-X

Verta J-P, Jacobs A. 2022. The role of alternative splicing in adaptation and evolution. Trends Ecol. Evol 37:299–308.

Wang RN, Green J, Wang Z, Deng Y, Qiao M, Peabody M, Zhang Q, Ye J, Yan Z, Denduluri S, et al. 2014. Bone Morphogenetic Protein (BMP) signaling in development and human diseases. Genes & Diseases 1:87–105.

Wright CJ, Smith CWJ, Jiggins CD. 2022. Alternative splicing as a source of phenotypic diversity. Nat Rev Genet.

Acknowledgements

"Home is not where you live, but who cares when you are gone"

– *In the Blood,* song by Darren Korb

I often feel that while life does not always go the way one expects, I have always been extremely lucky in the things that really matter. In particular, the fact that, wherever I go, I always find really great and kind people. Thus, I was lucky to be surrounded by a large group of amazing people that supported me so much throughout my PhD. Some were geographically closer than others, but every single one of these people gave me something valuable in every little interaction that we had together. I do not think I can come up with words that will ever make justice to the deep gratitude and appreciation I have for all these people. So, I will keep this very simple, but know that, to all of you, I am deeply grateful.

I would like to start by expressing my gratitude to Katie, for being an extremely kind and supportive mentor who helped me grow immensely as a scientist and who always encouraged me to follow the questions that I am most passionate about. Directly or by example, you taught me both how to be a better scientist, but also how to be a kind and good person in science. I know that I will remember those lessons for the rest of my life.

I would like to thank Vítor Sousa, who was also an amazing mentor during master's studies and who encouraged me to contact Katie for a PhD. Thanks to you, I ended up here and had some of the best years of my life.

I would like to thank to to all the amazing members of the Peichel lab. All of you make the lab a great (and really fun!) place to work. I could not have asked for better people to spend most of my days around.

I would also like to thank the amazing group of friends that made my life in Bern so fun in these last couple of years. While there are too many people for me to mention individually, I would like in particular to express my deepest gratitude and appreciation to (in no specific order): Ludo, Marion, Dave, Hiranya, Juli Maria, Ben, Åsa, Qindong, Tamara, Pooja, Itzel, Dan and Majda. We have been through so many fun and amazing times together. From delicious dinners and lunches at each other's places, to sunny days by the Aaare; from intense töggeli matches to intense board game evenings.

From fun evenings drinking beers together and laughing at ridiculous conversation to some truly amazing trips abroad. I have done so many new things thanks to you and learned so much. You are all truly dear to me and while we do not know where our lives will take us, it is my deepest wish that we find a way to keep meeting and having fun and amazing times together for many, many, years to come.

I know this is possible thanks to my great friends back in Portugal. Who despite the distance that separate us, have never given up on reaching out to me and are always there for me when I go back to my Lusitana home. I would like to express my gratitude in particular to Inês Ventura, Gonçalo Miguel, Patrícia Santos, Catarina Mouta, Bruno Carvalho, Cláudia Belém, Miguel Almeida, Antonio Brito, Diogo and Pedro Reis. You are all truly amazing friends, and I am really lucky that, despite not managing to stay in contact nearly as much as I wish I would, all of you are still there for me whenever I reach out. Thank you so much.

Finally, I would like to express my deep gratitude and love to my family, who helped me become the person that I am today and that have given me so much love and support throughout these years that we have been half-a-world of distance from each other. It has been hard to miss so much of your lives, and I miss you all a lot. I would not be here in Bern finishing a doctoral thesis if it was not for all your support. So, with all my love I dedicate this thesis to you.

Finally, I would like to express my deep love and gratitude to Cátia, with whom I have been sharing the ups and downs of life for nearly a decade now. Again, I find myself at loss for words because no words I can think of can convey all the love and gratitude I feel for the support that you have given me on a daily basis. You have seen all my struggles and all my victories, and you have always been there to celebrate with me or give me strength. You give me every day a loving home to comeback to. Thank you so much, I love you.