THE GENETIC BASIS OF DEFENSIVE STRUCTURE VARIATION IN THREESPINE STICKLEBACK (*GASTEROSTEUS ACULEATUS*)

by

JACK H. PEPLINSKI

A THESIS

Presented to the Department of Biology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2019

An Abstract of the Thesis of

Jack H. Peplinski for the degree of Bachelor of Science
in the Department of Biology to be taken June 2019

Title: The Genetic Basis of Defensive Structure Variation in Threespine Stickleback (*Gasterosteus aculeatus*)

Approved: _____

Dr. William A. Cresko

A longstanding goal in evolutionary biology is to link differences in traits among organisms with genetic variants in their genomes. Hybrid populations are excellent models for studies that aim to associate such phenotypic variation with regions of the genome. The Riverbend section of the McKenzie river in Oregon is home to a hybrid population of freshwater-like and ocean-like threespine stickleback (*Gasterosteus aculeatus*). Stickleback are small fish that live in a variety of aquatic habitats and appear highly armored in oceanic and brackish environments and often exhibit a loss of armor in freshwater systems. Previous work in this population demonstrated that variation in a handful of bony traits encompassed the differences observed between oceanic and freshwater types. I hypothesized I would see similar variation in the pelvic defensive structure–a group of bones that surround the fish and protects it from predation–and that by association mapping, I would identify genetic variants contributing to the diversity in this trait. For this thesis, I measured 12 aspects of the defensive structure in 192 fish and used 19,540 genetic markers to perform a genome-wide association analysis. Here, I show that the defensive structure and its

components display abundant variation between individuals in this population. I describe the genetic architecture of this set of traits and report genetic regions of association, some of which overlap with previously discovered regions. In addition, I found novel regions of association for a subset of the traits and report candidate genes in these regions that may contribute to the phenotypic differences observed.

Acknowledgements

It's my pleasure to acknowledge and sincerely thank my mentor and advisor Mark Currey, whose time, mentorship, and expertise were invaluable to my growth as a scientist and a student. His work served as the inspiration for this thesis and his tireless enthusiasm for biology was inspirational on its own.

I would like to thank Dr. Bill Cresko for the opportunity to work in his lab and the doors and avenues that opportunity opened for me. His dedication to my project and the incredible wealth of knowledge and know-how he's shared with me has been exceptionally generous.

I would also like to thank the remaining members of the Cresko lab, who served as fantastic role models. It was a privilege to work alongside them. Much of the content of this thesis was inspired and directed by the work and notes of Dr. Kristin Alligood, a former Cresko lab member, to whom I'm very grateful.

Furthermore, I would like to acknowledge and thank Dr. Jeff McKnight in the biology department and Dr. Mark Carey in the honors college for hosting separate, rigorous, and helpful thesis prep classes that met throughout my final two years. Their guidance–and the help of my friends in those classes–helped me stay on track. In addition, I would like to thank my remaining committee members Dr. Nicole Dudukovic and Dr. Cristin Hulslander for their time and support.

I dedicate this project to my family, for supporting me through good and bad times, teaching me to work hard, and encouraging me to do my best.

Table of Contents

List of Figures

List of Tables

Chapter I

The Genetic Basis of Phenotypic Variation: How Do We Connect Genotypes to Phenotypes?

Biologists and other observers of the natural world have long been interested in the appearance of organisms and the diversity of forms organisms can take (Gessner 1551; Darwin 1859; Mendel 1865; Wallace 1871). *Phenotypes*–the appearances or behaviors of organisms (Roll-Hanson 2009)–can vary not just between different species, but also between individuals that belong to the same species (Liu et al. 2009; Branicki et al. 2011; Wood et al. 2014). Humans are a straightforward example of intraspecific variation. We have different skin tones, face and body shapes and hair color, among many other features. These variations are controlled by changes in our genome (Mendel 1865) and are called *genotypes* (Roll-Hanson 2009). Before the invention of recent technologies such as high-throughput sequencing tools, we did not have a clear understanding of the link between genotypes and phenotypes (Reuter et al. 2015). But with the discovery of DNA (Watson and Crick 1953) and how traits are passed from one generation to the next (Darwin 1859; Mendel 1865), we are beginning to understand how these differences in appearance arise.

A Brief History of Landmark Genetic Discoveries

In order to understand to explore the connection between phenotype and genotype, we must first understand the early discoveries of genetic drivers and inheritance that made studies like this thesis possible.

Czech scientist Gregor Mendel was the first to discover the particulate mechanism of inheritance with his famous pea experiments (Mendel 1865). Mendel cross-bred pea plants (*Pisum sativum*) over several generations and found that the seven traits: flower color, flower position, stem length, seed color, seed shape, pod color, pod shape were inherited without a blending of parental characteristics (Mendel 1865). This blending, called incomplete dominance, was the prevailing assumption at the time (Darwin 1868; Fisher 1931). The peas displayed complete dominance, where one allele completely masks the other. From these experiments, Mendel was able to conclude that each trait was inherited by factors we now call alleles. He also deduced that an individual inherits one allele from each parent for each trait and that the trait may not show up in an individual but can show up in future generations (Mendel 1865; Wood 1995). Mendel's research helped establish a basis for studies of genetic drivers of appearance (Bateson 1913; Parsons and Bodmer 1961; Dunn 1991).

Heritable traits are more likely to persist in a population–through a process called natural selection (Darwin 1859; Fisher 1930). English scientist Charles Darwin gathered significant evidence from the natural world–just before Mendel's work with peas–including on the now famous Galápagos finches. He concluded that the finches on the Galápagos islands were different but closely related species that, due to a process called adaptive radiation, evolved different beak shapes to become more effective foragers of a variety of foods (Darwin 1859). In natural selection, organisms better suited to their environment reproduce and pass on their genes, changing the phenotypic and genotypic makeup of a population (Fisher 1930; Williams 1966). Just like pea color, beak shape is a heritable trait (Darwin 1859; Grant and Grant 2002; Abzhanov et

al. 2004; Grant and Grant 2006). The genome 'instructs' the beak to look a certain way and that same script is more likely to be passed down to future generations if the beak shape allows the bird to acquire more resources (Darwin 1859; Grant and Grant 2002). Darwin's findings helped reveal the connection between environment and phenotypic variation and thus serves as a foundation for this thesis.

Technological Advancement Makes Improved Genetic Studies Possible

New sequencing technologies are producing genetic sequences from a vast array of organisms at an ever-improving rate (Reuter et al. 2015; Hohenlohe 2018). We now have the capacity to identify and quantify genetic variation of populations (Yu and Buckler 2006; Baird et al. 2008; Li 2011; Catchen et al. 2013; Alligood 2017). This allows us to potentially identify the genetic basis of phenotypic variation (Arya et al. 2011; Pallares et al. 2014; Kusakabe et al. 2016; Alligood 2017). One method to link genotype to phenotype is a technique called association mapping (Weigel and Nordborg 2005; Hoekstra et al. 2006; Yu and Buckler 2006; Li 2011; Reed et al. 2011; Pallares et al. 2014; Alligood 2017). Variation in the genome of an organism is correlated with differences in a quantifiable phenotype, such as height. When individuals within a population mate and reproduce, undergoing genetic recombination, the population evolves, through natural selection, to better adapt to its environment (Darwin 1859; Fisher 1930). By studying a single population, we can look at the small, individual changes in a near-identical genome called single-nucleotide polymorphisms (SNPs) and by controlling for location, we eliminate any kind of confounding environmental variables. Though we are beginning to understand the link between genetic and phenotypic variation, there's still much work to be done.

Genetic Architecture: Monogenic vs. Polygenic Bases for Phenotypes

Genetic architecture describes the way the genome codes for heritable phenotypic variation (Hansen 2006; Fuchsberger et al. 2016). One of the simplest ways of looking at the genetic architecture of a trait is to understand how many genes influence variation in a given phenotype. Phenotypes controlled by one or a few genes are called 'monogenic' while phenotypes controlled by many genes are called 'polygenic' (Cooke and Buckley 1987).

Monogenic traits are most often discrete in nature, such as color morphs or the presence/absence of a trait (Mendel 1865; Cooke and Buckley 1987; Colosimo et al. 2005; Hoekstra et al. 2006). The beach mouse (*Peromyscus polionotus*) is a good example of monogenic inheritance of a discrete phenotype. The mouse can display a dark or light color morph depending on two genes, the melanocortin-1 receptor (*Mc1r*) and *Agouti* (Hoekstra et al. 2006). On the other hand, polygenic inheritance is commonly found in complex, indiscrete, or quantitative traits such as human height (Cheverud 1984; Charlesworth 1990; Orr and Coyne 1992; Wood et al. 2014). About 80% of the human height is thought to be genetically controlled and about 50 genomic regions have been found to contribute to variation in human height (Yang et al. 2010). Being taller or shorter may not directly affect fitness but may contribute over generational and evolutionary time and may be a result of steady diet and lower rates of infectious disease which have been shown to negatively influence height (Mummert et al. 2011; Stulp and Barrett 2014).

The stickleback has also been analyzed for the purposes of unpacking the differences between mendelian and polygenic drivers of quantitative traits. For the most

part, the indiscrete and quantitative traits of the stickleback have been found to be polygenically inherited (Bell 1988; Blouw and Boyd 1992; Bell and Ortí 1994; Hatfield 1996; Benson et al. 2010; McGuigan et al. 2010; Li et al. 2017). However, this is not necessarily the case. Sometimes even discrete traits can be polygenically inherited and do not fit an 'additive model' (Banbura 1994; Cabot et al 1994; Davis et al. 1994; Hatfield 1996; Colosimo et al. 2004; Cresko et al. 2004; Glazer et al. 2014). This thesis aims–in part–to address the ongoing debate between the genetic architecture of certain traits. Based on previous findings, I expect to see polygenic inheritance for the linear traits that make up the defensive structure.

The Threespine Stickleback as a Model Organism

In order to study phenotypic variation, biologists use model organisms that possess characteristics that facilitate research. The stickleback is a model organism especially suited for studying the link between genotypes and phenotypes (**Figure 1**) (McPhail 1969; Moodie 1972; Bell and Foster 1994; Walker 1997; Cresko et al. 2007; Currey 2014; Alligood 2017). It has several characteristics that make it so:
Stickleback are relatively easy to trap and to maintain in the lab. They are tolerant to laboratory environments and have short generation times. Once euthanized, bleached, and stained, stickleback display bright bone structures, making for easy quantification of phenotypic traits (Cresko et al. 2004; Kimmel et al. 2005; Conte et al. 2015). Their small size makes euthanized specimens easy to image as well as to preserve and store. A major consideration for this thesis was that the stickleback genome had already been sequenced. This made it possible for genetic analysis to be performed.

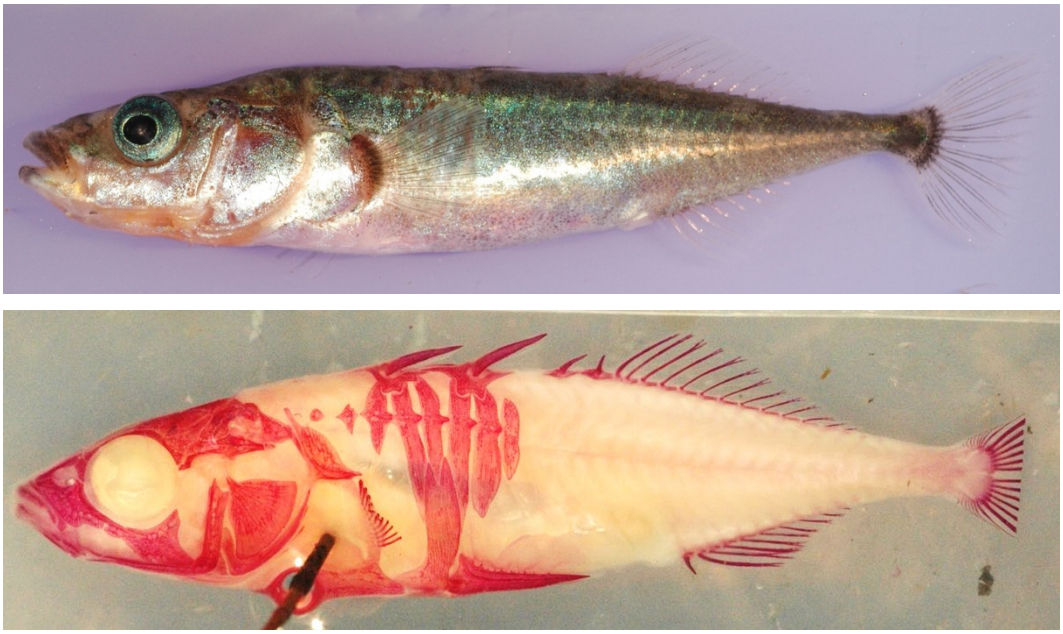

Figure 1. (Top) Live still of a threespine stickleback from Boot Lake, AK. (Bottom) Image of a euthanized stickleback, bleached and stained with Alizarin red from the Riverbend population in OR.

Besides lab handling, stickleback also have several natural characteristics that make them useful for studying the link between genotype and phenotype. Two of these characteristics make them especially suited for this thesis:

First, stickleback are found in a variety of aquatic habitats and life histories throughout the Holarctic region of the Northern hemisphere. These habitats include marine, estuarine, and freshwater environments (Bell and Foster 1994; Schluter 1995; Rundle et al. 2000; von Hippel 2010; Hendry et al. 2011). For example, anadromous stickleback develop and hatch in freshwater, migrate to the sea to grow, and then return to freshwater to spawn. Important for studies of adaptation, anadromous stickleback have repeatedly given rise to thousands of freshwater populations over the last millennia (Bell and Ortí 1994; Olafsdottir et al. 2007; Kitano et al. 2008; Gelmon et al. 2009; Reimchen et al. 2013). These different freshwater populations display a multiplicity of

adaptations necessary for survival in new habitats (Lavin and McPhail 1986; Bell and Foster 1994; Kalbe and Kurtz 2006; Hendry et al. 2011; Reimchen et al. 2013).

Second, stickleback display distinct phenotypes among populations, such as the number of lateral plates, which makes it relatively simple to identify noteworthy populations in order to further study their atypical appearances (Cresko et al. 2004; Kimmel et al. 2005; Conte et al. 2015). Lateral plates are plates of bone located along the sides of the fish and protect the fish from predators (Bell and Foster 1994; Reimchen et al. 1983). The number of lateral plates shows high variance among different populations (Colosimo et al. 2005; Alligood 2017). This is due to the range of environments that different populations can occupy, which leads to diversification as they rapidly adapt to the environment. More extensive armor is thought to protect the fish at the cost of slowing it down (Taylor and McPhail 1986; Bell and Foster 1994; Reimchen 1995; Walker 1997; Reimchen 2000). Marine and estuarine environments suit the greater armor mass whereas the freshwater systems promote the loss of armor to allow for faster swim speed. In addition, freshwater stickleback with reduced lateral plate armor have been shown to grow faster (Marchinko and Schluter 2007). Faster growth may increase the survival rate of juvenile stickleback against predating insects, as well as boosting nutritional reserves for overwintering and improving reproductive fitness (Barrett et al. 2008; Schluter et al. 2010).

Previous studies on threespine stickleback have identified the morphological variation in bone structure within natural populations of stickleback. This variation has been categorized into two distinct phenotypes: oceanic and freshwater (**Figure 2**). Other studies have shown that there are associations between separate bone structures

(Reimchen et al. 1983; Bell and Foster 1994). Furthermore, through QTL mapping studies, it has been shown that genomic regions are linked to these two phenotypes (Reimchen et al. 2013). Finally, a genome-wide association study (GWAS) has been performed on the skeletal variation on specific bone structures in Riverbend stickleback (Alligood 2017).

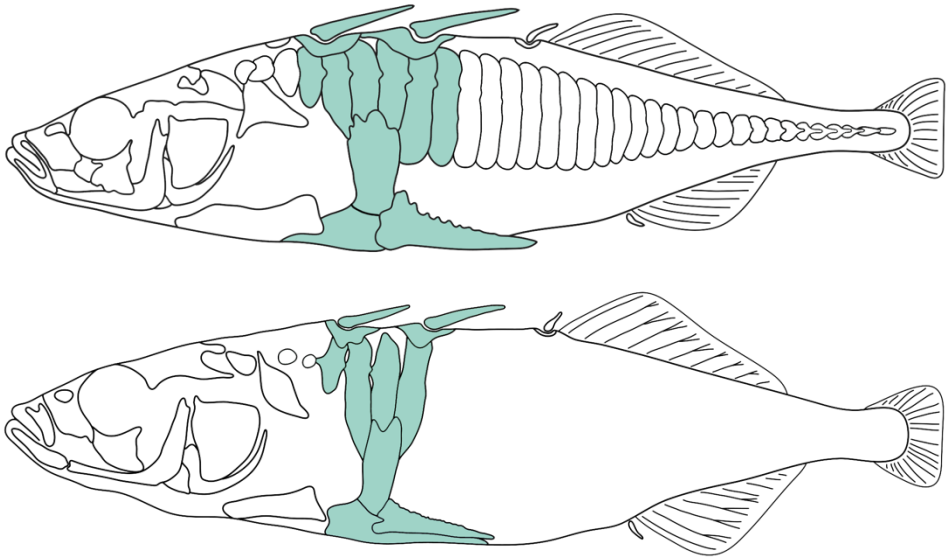

Figure 2. Ocean-like (top) vs. freshwater-like (bottom) stickleback, as seen in the Riverbend population. Posterior lateral plates are lost in freshwater-like stickleback, but parts of the defensive structure can also experience loss. Note: the smaller spines on the freshwater-like stickleback.

Association Mapping in a Hybrid Population

While quantitative trait loci (QTL) mapping studies are commonplace in addressing the link between phenotype and genotype, they require laboratory crosses that may not provide an accurate depiction of the mechanisms underlying natural populations. Laboratory crosses can also be difficult to execute if the wild species is

difficult to capture or if there are problems associated with the breeding and propagation of the laboratory populations (Rieseberg and Buerkle 2001). Thankfully, association mapping can be performed in natural populations by taking advantage of hybrid populations, where once separate genotypes and phenotypes have been allowed to undergo recombination (Hatfield et al. 1992; Sites et al. 1995; Kruuk et al. 1999; Currey 2014; Alligood 2017). Hybrid populations are extremely useful for studying the link between genotype and phenotype because recombination breaks down genomic linkage blocks and population structure that hinders the ability to associate variation with changes in the genome (Rieseberg et al. 1999, Rieseberg and Buerkle 2001; Currey 2014).

Genome-wide association studies (GWAS) are widely used methods of linking genes to phenotypes in natural populations (Visscher et al. 2012; Lee et al. 2014; Nadeau et al. 2014; Pallares et al. 2014; Turner and Harr 2014; Wang et al. 2014; Brelsford et al. 2017). GWAS identifies common variants in a set of sample individuals without loci or a candidate gene in mind (Cantor et al. 2010; Zhou and Stephens 2012). GWAS are exploratory in nature because of this, which suits the purposes of this thesis, and are usually followed up by more specific association studies that can narrow down a genomic locus of interest. GWAS has been used to great effect (Pasaniuc and Price 2016) in a variety of systems, including stickleback (Alligood 2017), though it has a major pitfall. In some cases, a GWAS will return numerous associations for genetic loci with common variants in the sample set (Ward and Kellis 2012). This is problematic in itself because it makes it difficult to parse the significance of the association but it also assumes that the marker is biologically relevant. Without knowing the exact

mechanisms that underlie the genome, it is difficult to tell whether that marker is actually coding for the phenotype of interest or merely mathematically associated (Smith and O'Brien 2005; Korte and Farlow 2013). Despite the drawbacks, GWAS still remains the best tool for exploratory studies of linking genotype to phenotype in natural populations (Pallares et al. 2014). Genomic associations can be cross-referenced with previously discovered genetic loci and analyzed for biological relevance–often based on orthologues. Then, as previously stated, candidate genes can be further analyzed for association.

Chapter II

Introduction

Variation in organism morphology has long interested biologists. Individuals within the same species can possess various phenotypes, often caused by environmental adaptation over evolutionary time. This phenotypic variation can be associated with genomic regions through analytical techniques such as genome-wide association studies (GWAS). However, such studies require both knowledge of the phenotype and the genotype of individuals within the population of interest. Fortunately, recent technological advances, such as high-throughput sequencing, have allowed researchers to sequence entire genomes, enabling the linking of genotype to phenotype.

While many association studies use laboratory crosses, GWAS can be utilized in natural populations. For this to work, individuals of different phenotypes and genotypes must interbreed and allow enough generational time for genomic recombination. Fortunately, the Riverbend section of the McKenzie river in Oregon contains a hybrid population of an evolutionary model organism, a small, bony fish called the threespine stickleback (*Gasterosteus aculeatus*). The stickleback inhabits a variety of environments, including marine, brackish, and freshwater systems and displays distinct skeletal phenotypes, thought to be adaptive in nature. In the Riverbend section, oceanic-like stickleback have been introduced to an endemic freshwater population and–as I show in this thesis–have had enough time to intermix.

Association studies in the Riverbend population have previously focused on cranial-facial morphology and lateral plate counts and have found that the Riverbend

population expresses intermediate phenotypes compared to other low- and high-plated Oregon populations. I am interested in the genomic basis of phenotypic variation of the mid-lateral region of the stickleback, responsible for the defense of the fish against natural predators.

The Riverbend Population: Ideal for Linking Genotype to Phenotype

Stickleback display clear phenotypes depending on their environment. Stickleback living in and around the coast, in salty or brackish water tend to display heavily-armored phenotypes most noticeably characterized by high numbers of lateral plates. These fish also tend to have cranial-facial morphologies that suit hunting instead of filter-feeding, so their heads are tall and rounded instead of long and thin. When two distinct phenotypes undergo recombination, they can produce intermediate phenotypes that fall between the two original ends of the spectrum. Depending on the nature of the system and the severity of the introduction process, hybrids can be selected against due to lowered fitness due to an inability to fit into an ecological niche (Schluter 1993). However, this is not always the case. Intermediate phenotypes can have no effect on a population or even a beneficial effect if the environment allows it. Stickleback hybrids have been shown to have not fitness disadvantages in the laboratory (McPhail 1984, 1992, 1994; Hatfield 1995) but have been shown to fare worse in the wild (Schluter 1995; Hatfield 1995) due to disruptive selection in natural environments.

There are several different hybrid zones, where high and low-plated individuals intermix. There are several causes for emergence of hybrid zones. A common cause is a natural environmental influence such as the formation of a temporary or permanent bridge between previously allopatric habitats. Flooding of low-elevation regions is a

common example of this. Another cause could be an environmental cline such as salt concentration as one moves up-river from the ocean. One could expect to see a mixed phenotype near the river delta and the extreme phenotypes in the ocean and up-river.

Stickleback are found throughout the Willamette Basin in Oregon. Low-plated stickleback are generally found in inland freshwater systems such as lakes and rivers and the high-plated stickleback are generally found in saltwater near the coast and in the ocean. An apparent hybrid population has formed in the McKenzie river near the Riverbend hospital located ~11.0 river miles from the confluence of the McKenzie and main stem of the Willamette river (**Figure 3**). This hybrid population is thought to be the result of human introduction of ocean-like stickleback upstream with subsequent admixture of the ocean-like fish with freshwater fish found on the valley floor (Currey 2019). The resulting population shows both high, intermediate, and low-plated individuals as well as much variation in other phenotypes. Previous work in the Riverbend population successfully used GWAS to identify genomic regions associated with lateral plate count and cranial-facial variation (Alligood 2017) suggesting that this population and this approach may prove useful for investigating the genetic basis of the defensive structure.

Aspects of the defensive structure also appear to vary in this population. This observation has led to the hypotheses for this thesis. The unique makeup of the Riverbend population makes it ideal for studying the genetic basis of defensive structure because it may allow us to associate apparent phenotypic variation in this structure with genetic variation and ultimately understand the genetic basis of this structure.

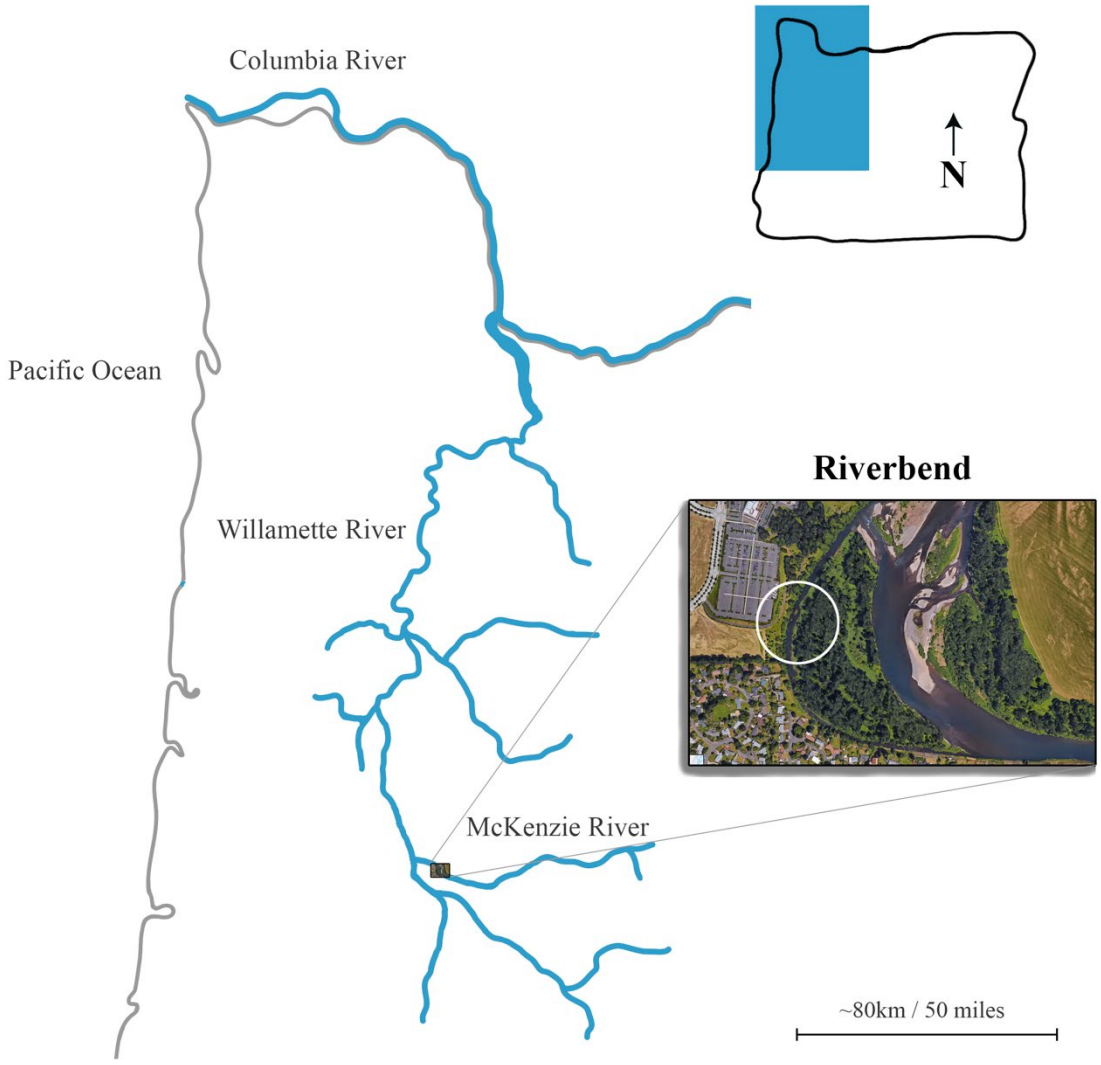

Figure 3. The location of Riverbend in the McKenzie river within the Willamette basin of Oregon.

When a divergent population is introduced to an existing population, there can be considerable differences–due to the natural processes of divergence–in the genome between the two groups. If there has not been enough time for recombination to disassociate this population structure, it can be an inhibitor of any GWAS study because any attempt at association will result in a reflection of the population structure. The

Riverbend population has had adequate time to hybridize between the oceanic-like and freshwater-like forms and thus lacks significant population structure (Alligood 2017). This makes the population suitable for association studies.

Addressing the Defensive Structure as a Sum of its Parts

The defensive structure, as defined by this thesis, is the entire mid-lateral region of the stickleback from the dorsal spines to the pelvic spines, and is composed of a number of overlapping and interacting protective overlying bones and cartilage that defend the fish against predation from avian and piscine predators such as pacific salmon and trout from the *Oncorhynchus* genus (**Figure 4**) (Bell and Foster 1994; Olsen et al. 2000; Reimchen 1992, 2000). The components of the structure work together to form a protective buttress around the midsection of the fish, attaching to and strengthening each other (Bruet et al. 2008; Song et al. 2010). The spines, when extended, create a wide and spiny bite radius for large piscine predators (Hoogland et al. 1956; Reimchen 2000, et al. 2013). Various components of the defensive structure have been studied, such as the pelvic structure and spines lengths, yet the structure has not been analyzed as a possibly-covarying unit. The literature has mostly ignored the basal plates and the ascending process, which work to hold the dorsal spines in place and connect both dorsal and ventral ends of the structure, respectively (Reimchen 1983).

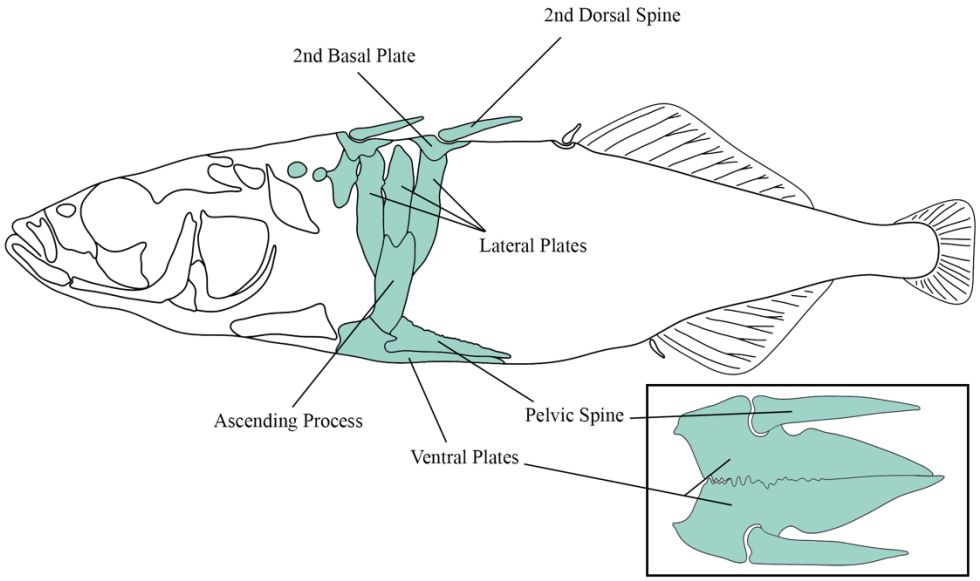

Figure 4. The Defensive Structure of a low-plated Threespine Stickleback (G. aculeatus). The ventral structure is shown in the box in the bottom-right of the figure.

In addition, the studied aspects of the structure have been linear or numerical in nature (Song et al. 2010; Reimchen et al. 2013), which may not provide a comprehensive view of the structure (Wiig et al. 2016).

Due to the importance of the structure in the overall fitness of the stickleback and the lack of literature on the entire structure, I set out to characterize the variation between individual traits within the structure and link that variation to regions on the genome. A previous study by Reimchen et al. (2013) identified eleven skeletal elements that appear to covary in the defensive structure. This finding serves as the starting point for this thesis, which aims to expand upon the traits in the previous study. The traits were all examples of linear measures; therefore, I expanded the scope of traits measured by including overall areas of different elements of the defensive structure. For example, as the ascending process is a major component that overlaps and integrates the dorsal

and ventral elements, the area of this trait may be more important than its height or length.

Research Aims and Hypothesis

Here I aim to take advantage of a hybrid population of stickleback to first quantify the phenotypic variation of the different elements of the defensive structure. I then aim to associate the phenotypic variation that I measure with genetic variation to understand the genomic architecture of these traits and to look for specific regions that associate with these traits or groups of traits.

Based on previous work and initial observations I expect to find significant variation within the different elements of the defensive structure and I also expect the individual elements to demonstrate a high degree of covariance with one another (Reimchen et al. 1983). Traits located closer to each other should show a stronger covariance than those which are farther away because they are forced to accommodate each other (Reimchen et al. 2013). The variation should be associated with regions on the genome that have already been located and also novel regions to unstudied traits in the defensive structure (Alligood 2017).

Methods

Collection and processing of stickleback samples

The first step in addressing the research questions was to collect stickleback from the Riverbend population. 600 individuals were all caught in one day. Non-baited 0.635 cm mesh minnow traps were placed in the water at the Riverbend location and left for 24 hours. 192 of the 600 were used for this analysis. The fish were caught and

euthanized using MS222 and fixed in 95% ETOH. Collections were made in accordance with IACUC protocols. The stickleback were then bleached and stained with Alizarin red solution in order to aid in visualization of the skeletal structures (Cresko et al. 2004; Kimmel et al. 2005; Conte et al. 2015).

Collection and scaling of phenotypic data

Instead of measuring the phenotypes of the stickleback using a physical measuring tool, each individual was imaged (with scale bars and labels) and then measured using the image analysis program ImageJ v. 1.x, for ease and accuracy (Schneider et al. 2012). In order to get a comprehensive view of the defensive structure as a whole, I collected numerical, linear, and area measurements from a variety of components of this trait. This required four different types of images for each individual: whole lateral, zoomed lateral, ventral, and anterior.

Whole lateral images of the stickleback were taken using a mounted Nikon D70 digital camera. In order to measure some of the more minute phenotypes, high quality imaging of the defensive structure, zoomed lateral, ventral, and anterior were taken using an Olympus SZX16 dissecting microscope equipped with an Olympus DP71 microscope digital camera. Olympus DP Controller version 3.3.1.292 was used to take and label images. In ImageJ, each individual was analyzed for 20 phenotypic traits, using the linear and freehand drawing functions to measure linear and area phenotypes respectively (**Figure 5**). Area measurements were performed by using the freehand draw tool in ImageJ v. 1.x to trace the outside of the bones and using the measure function to calculate the area of the inside.

I started with a broad approach to choosing which traits to measure. Previous research has measured a few linear aspects of the structure, including spines and the ventral plates (Reimchen 2013). I used those as a starting point, including other aspects of the structure such as the basal plates and ascending process, and including area measurements in addition to linear. The 20 traits were each thought to have some biological significance in terms of the fish's defense and were all observed to vary from individual to individual. Due to the amount of scholarship on lateral plate counts, they were included as a control to test the methodology of this thesis. Standard length was also included as a control to account for the varying sizes of the individuals.

Lateral Plates

The first of the 20 traits measured was a count of lateral plates (LPC). As previously stated, this measure is a control for the experimental design of this thesis. In addition to the count, I looked at the total area of all lateral plates (LPA), in case the size of the plates did not directly associate with the count (Wiig et al. 2016). A majority of association studies have focused on the lateral plates, as they are perhaps the most obvious source of variation (Bell 2001; Colosimo et al. 2004; Cresko et al. 2004; Shapiro et al. 2004; Colosimo et al. 2005; DeFaveri et al. 2011; Jones et al. 2012; Alligood 2017). They have also been heavily linked to a specific gene located in chromosome 4, the Eda gene (Colosimo et al. 2004). Although lateral plates attached to the basal plates and/or ascending process are part of the defensive structure, the rest are not (**Figure 2**)

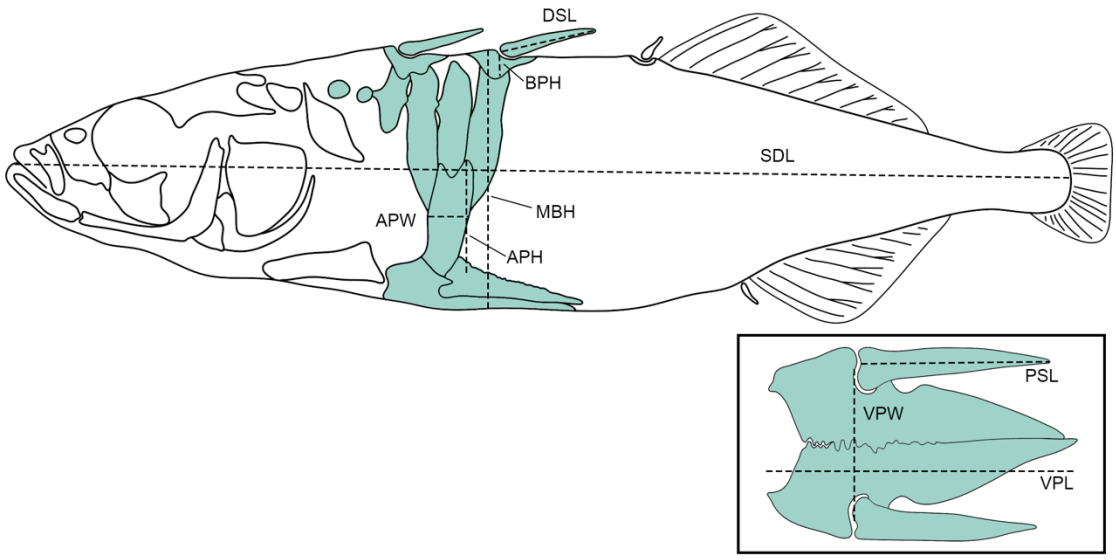

Figure 5. Linear traits measured on the stickleback (along the dashed lines). Area traits were traced along the outline of each of the outlined bones structures. APH – Ascending Process Height, APW – Ascending Process Width, BPH – Basal Plate Height, DSL – Dorsal Spine Length, MBH – Maximum Body Height, PSL – Pelvic Spine Length, SDL – Standard Length, VPL – Ventral Plate Length, VPW – Ventral Plate Width.

Size Axis

The standard length of the stickleback (SDL) was measured along with maximum body height (MBH). These size measures vary with the age and sex of the fish (Hunt et al. 2009; DeFaveri et al. 2013). SDL is a control for the size of the fish while the maximum body height can be considered part of the defensive structure, as it plays a role in increasing the bite-radius a predator must account for (Hoogland et al. 1956; Reimchen 2000, et al. 2013). SDL was measured by starting at the anterior edge of the lip bone and drawing to the posterior end of the fish, stopping where the tail fin begins. MBH was measured as the tallest point of the fish, from the dorsal end of the second basal plate to the ventral end of the pelvic structure, excluding spines (**Figure 5**).

Spine Axis

Much research has been done on the length of the spines in stickleback, including the second dorsal spine (DSL) and the pelvic spine (PSL) (Reimchen 1983; Peichel et al. 2001; Shapiro et al. 2004; Kitano et al. 2009, Reimchen 2013). In my initial sample, I analyzed the width of the second dorsal spine (DSW) and the cross-sectional diameter (CSD) which is the diameter of the outstretched dorsal and pelvic spines, mimicking the bite-radius of a potential predator (Reimchen 2013). DSL and PSL were measured linearly (as seen in **Figure 5**) and DSW and CSD were measured using anterior-facing images of the stickleback. DSW was measured at the base of the spine and CSD was measured from the anterior edge of the outstretched DSL in a straight line through the fish to the posterior edge of the outstretched PSL, forming a diameter.

Basal Plates

The basal plates play an important role in stabilizing the dorsal spines and connecting the spines to the rest of the defensive structure (Reimchen et al. 1983). Because of this, I analyzed the area of both plates (BPA1, BPA2) and the height of the second basal plate (BPH). I also looked at the margin of overlap between the lateral and basal plates as a measure of the strength of the attachment (LBO) (Reimchen 2013). BPH was drawn from the ventral edge of the second dorsal spine (DSL) to the ventral edge of BPA2. LBO was measured by tracing the length of the margin where BPA2 touches any lateral plates, including gaps between the touching plates (**Figure 5**).

Ascending Process

The ascending process is connected to the lateral plates and the ventral structure and shows a wide range of variance in appearance, most notably in the shape and number of forks (APC). I also measured the area (APA), width (APW), and the height of the ascending process (APH). Finally, I measured the margin of overlap between the ascending process and the lateral plates (LAO) to measure the strength of attachment. APW was measured at the widest part of the ascending process. APH was measured from the dorsal edge of the ascending process to the recession seen in the pelvic spine (PSL). LAO was measured by tracing the length of the margin where the ascending process touches any lateral plates, including gaps between the touching plates (**Figure 5**).

Ventral Structure

The ventral structure (or pelvic girdle), connects the ascending process to the pelvic spine and has been previously studied (Reimchen et al. 1985, Cresko et al. 2004; Shapiro et al. 2004; Chan et al. 2010). I measured the area of both left and right ventral plates (VPA) and the length and width of the plates (VPL and VPW). VPL was measured from the farthest anterior to the farthest posterior edge of the combined ventral plates. VPW was measured between the recessions in both pelvic spines (**Figure 5**).

Log-transformation and size-standardization

A replicate of each measurement was taken for every fish. The average between replicates was then taken and size-standardized. Size-standardization is necessary for

association mapping studies due to the high amount of variation of body size due to a variety of factors (age, sex, food availability, etc.) (Hunt et al. 2009; DeFaveri et al. 2013). In order to properly address the size of certain traits, I set the traits against the length of the individual. The traits were log-transformed to reduce the skewness of the data. Finally, the measurements were log-scaled and size-standardized. Using this standardization equation (Reimchen et al. 2013):

$$\log y'_{ij} = \log y_{ij} - \beta(\log x_i - \log \bar{x})$$

where y'_{ij} is the adjusted value of trait *j* for individual *i*, y_{ij} is the unadjusted value, β is the unstandardized regression slope, x_i is the standard length of individual *I*, and $\bar{x}$ is the overall mean standard length of all the individuals in the study (Reimchen et al. 1985; 2013). The unstandardized regression slope was calculated in *R* 3.4.0 using the package *smatr*, and the command *sma*.

Analysis of measurement accuracy and trait redundancy

A sample of 20 fish randomly selected from the larger sample of 192 were analyzed for replication accuracy and initial reading of variation. Regression analysis showed that the R^2 for all traits measured were all above 0.878 and a one-sample analysis of variance (ANOVA) found that the P values for all traits were above 0.05 (minimum 0.830). There was no significant difference between the two replicate measurements. This shows that there was high repeatability in the measuring technique.

The sample was then expanded from 20 to 50 randomly selected fish to increase the sample's power. Using principle component analysis, I found that several traits overlapped in their variance. Strongly covarying traits were eliminated from the final

analysis for the purpose of this thesis. 6 traits in total were eliminated: CSD, BP1A, DSW, LAO, LBO, LPA.

PC1 explained 45.7% of the covariance between traits and PC2 explained 9.7% (**Figure 6**). This indicates that this is a good model for studying covariance. The area of the first basal plate (BP1A) covaried strongly with that of the second basal plate (BP2A). I eliminated it due to redundancy.

Cross-sectional diameter (CSD) and dorsal spine width (DSW) are good measures of defensive traits, as they are thought to hinder the ability of a predator to bite down on the fish. However, due to difficulty of measurement, the measurements were not closely replicated. Time was also a factor in the reduction of the two traits, as they required a new set of images for each individual. The margin of overlap between the lateral plates and the basal plates (LBO) and the margin of overlap between the lateral plates and the ascending process (LBA) was initially included because the overlap was hypothesized to play a role in the overall strength of the defensive structure as a buttress against piercing predators and due to previous literature (Reimchen et al. 2013) that included a similar measure. Yet, the measures overlap in covariance with the height and area of the basal plates and were difficult to replicate. Individuals 8, 19, 20, and 24 were outliers along PC1. Initially this was thought to be due to a deformation of the preserved state of the individuals, but after close examination, they were preserved well and included in the final analysis.

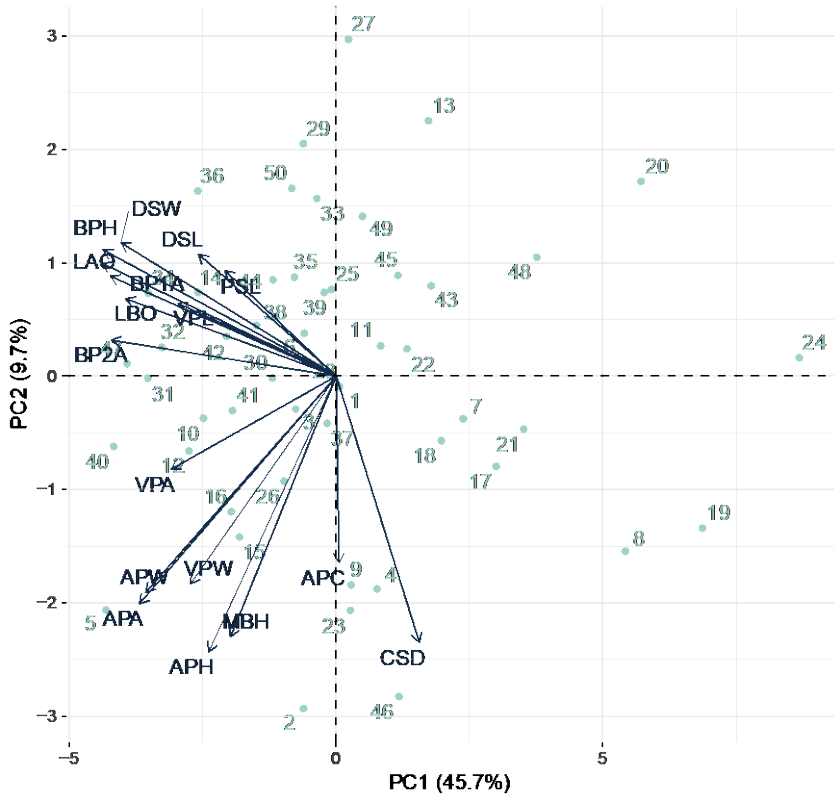

Figure 6. Principle component analysis for 20 traits. PC1 explains 45.1% of the variance and PC2 explains 9.5%.

This left 12 traits: Maximum body height (MBH), dorsal spine length (DSL), pelvic spine length (PSL), the area of the second basal plate (BP2A, re-abbreviated BPA), the height of the second basal plate (BPH), the number of forks on the ascending process (APC), the ascending process area (APA), the ascending process width (APW), the ascending process height (APH), the area of the ventral plates (VPA), the length of the ventral plates (VPL), and the width of the ventral plates (VPW). There was also the two controls: standard length (SDL) and the lateral plate count (LPC).

These traits were then grouped based on similar biological function to look for associations between the genome and highly covarying traits. There were eight groups

total: 1. The Ascending Process (included APA, APC, APH, APW), 2. The Basal Plate (BPA, BPH), 3. The Dorsal Complex (BPA, BPH, DSL), 4. The Size Axis (MBH, SDL), 5. The Spine Axis (DSL, PSL), 6. The Ventral Complex (PSL, VPA, VPL, VPW), 7. The Ventral Plates (VPA, VPL, VPW), and 8. The Defensive Structure (All traits). The size axis included the standard length (SDL) control as way to examine the relationship between height (MBH) and length (SDL).

The traits were then analyzed in all 192 individuals that had previously been RAD-sequenced. The procedure included the randomization of the order in which each individual was analyzed for each trait to reduce any bias with regards to improved accuracy of measurement over time. All counts and measurements were done twice for each individual. These replicates were not measured back-to-back per individual (ie. the APH was not taken twice on a single image before moving on), in order to reduce bias with regards to memorizing the previous measurement. The replications were then averaged. Principle component analyses were run for several combinations of phenotypes thought to contain biological significance in their groupings (**Table 2**).

Genetic Analysis

RAD library construction and SNP discovery

The genetic data used in this thesis was generated by Currey (2014) and Alligood (2017): DNA was collected through fin-clippings of each of the 192 stickleback individuals. This DNA was then digested with Sbfl-HF (NEB) restriction enzyme and RAD-seq libraries were prepared (Baird et al. 2008; Hohenlohe et al. 2012; Alligood 2017). This DNA was then run through the Illumina HiSeq 2500 platform, a

dye-sequencing tool. The barcode was demultiplexed and quality filtered through `process_radtag` in the *Stacks* software suite (Catchen et al. 2011). The reads were then aligned against the stickleback reference genome (version BROADSs1, Ensembl release 64) with GSnap (Wu and Watanabe 2005; Currey 2014; Alligood 2017).

Stacks programs `pstacks`, `cstacks`, and `sstacks` were used to identify SNPs across the genome. I used the populations program in *Stacks* to output filtered SNP data in PLINK in 21 linkage groups (Purcell et al. 2007). This resulted in 19,540 SNPs filtered and identified and included in subsequent association analysis.

In order to perform the association analysis I followed similar approaches as in Pallares and Alligood (Pallares et al. 2014; Alligood 2017). The associations were run using a univariate linear mixed model (LMM) in a genome-wide efficient mixed-model association (GEMMA v. 0.98.1) that controls for population structure and relatedness (Zhou and Stephens 2012). P-values were calculated through the likelihood-ratio test (LRT) in GEMMA and the likelihood of odds (LOD) scores were calculated using -log10 of the LRT p-values.

To correct for proximal contamination, a phenomenon that occurs when a mean-centered relatedness matrix is created where causative SNPs are fitted twice to the LMM, once in the matrix and once in the association, resulting in heightened background noise that makes it difficult to identify the causative SNP (Lippert et al. 2011; Listgarten et al. 2012). Using the method described in Pallares (2014) and Alligood (2017), associations were performed for individual linkage groups (LGs) using relatedness matrices that were created for all LGs except for the LG used in the association (Pallares et al. 2014).

In order to account for multiple testing, I used a Bonferroni correction with an alpha of 0.05. This resulted in correction was set at a LOD of 5.592. The Bonferroni correction is considered highly conservative for QTL mapping studies and thus, LOD scores lower than the Bonferroni correction may still be significant (Pallares et al. 2014).

Results

The Traits of the Defensive Structure Show Variation

All traits of the defensive structure show variation (**Figure 7**, **Table 1**). The average standard deviation between all unstandardized traits was 2.047, ranging between 0.208 for BPH and 6.246 for VPA (excluding controls). Lateral plate counts (LPC) and ascending process fork counts (APC) were not size-standardized because they are independent of size. The size-standardization showed reduced variation as a whole, however there is still a visible gap between sample set in each measurement, indicating there is a spectrum of phenotypes in the Riverbend population. The relatively low standard deviation is likely due to the presence of intermediate phenotypes in the Riverbend population (Alligood 2017).

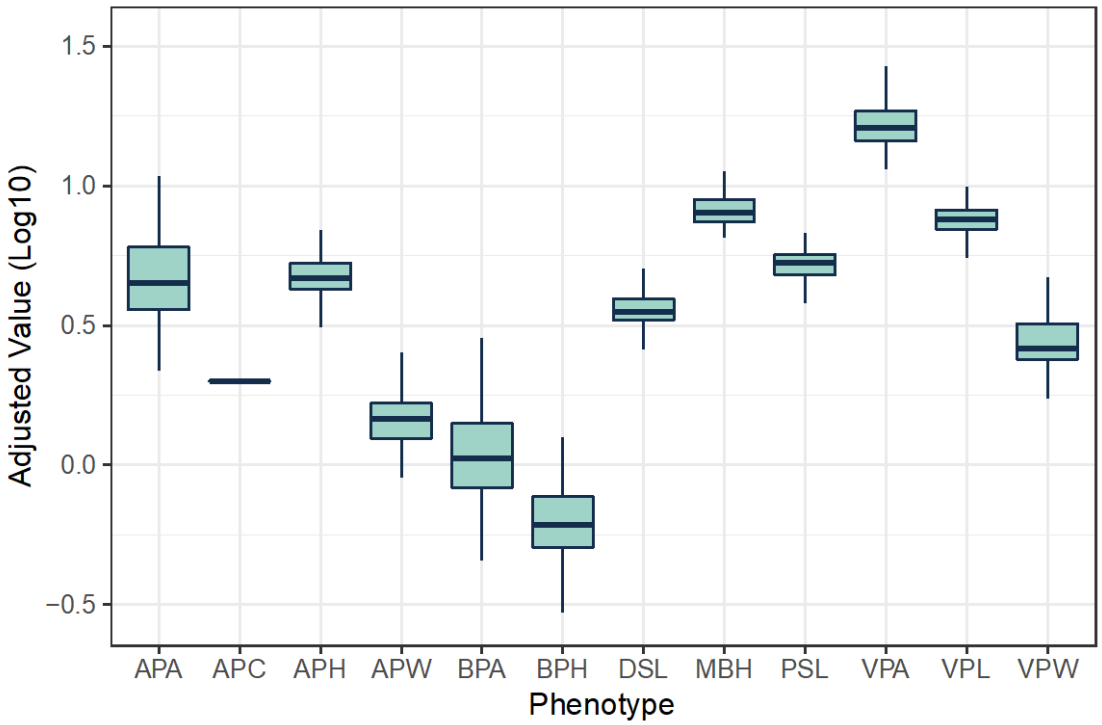

Figure 7. Adjusted values (Log10) for each of the measured phenotypic traits.

Table 1. Phenotypic variance between measured traits. Unstandardized and size-standardized mean and standard deviation. Measurements in millimeters or square millimeters.

		Unstandardized		Size-Standardized	
Trait	**Abbr.**	***M***	***SD***	***M***	***SD***
Ascending Process Area	**APA**	**5.027**	**1.904**	**0.674**	**0.152**
Ascending Process Fork Count	APC	1.948	0.594	N/A	N/A
Ascending Process Height	**APH**	**4.795**	**0.815**	**0.676**	**0.066**
Ascending Process Width	APW	1.495	0.331	0.164	0.094
Basal Plate Area	**BPA**	**1.175**	**0.533**	**0.035**	**0.170**
Basal Plate Height	BPH	0.653	0.208	-0.204	0.129
Dorsal Spine Length	**DSL**	**3.623**	**0.550**	**0.555**	**0.067**
Maximum Body Height	MBH	8.221	1.375	0.912	0.057
Pelvic Spine Length	**PSL**	**5.075**	**0.711**	**0.713**	**0.065**
Standard Length	SDL	35.326	5.273	1.544	0.061
Ventral Plate Area	**VPA**	**17.051**	**6.246**	**1.219**	**0.085**
Ventral Plate Length	VPL	7.585	1.104	0.877	0.055
Ventral Plate Width	**VPW**	**2.824**	**0.676**	**0.441**	**0.091**

Principle Component Analysis Shows Covariance in Defensive Structure Traits

We used principle component analysis (PCA) to group multiple traits into biologically relevant units to determine these grouped traits varied and covaried. The resulting principle components (PCs) were then used in subsequent association analysis (**Table 2**). All of the groupings except for the entire defensive structure (40.9%) had a PC1 greater than 0.5000 or 50% variance explained. By definition, there can only be as many principle components as there are traits, therefore groups with less than 4 traits did not have PC4s.

Table 2. Principle component scores for each grouping of traits for GWAS. Only the PC1 scores were used for GWAS.

		Principle Components			
Group	Traits	*PC1*	*PC2*	*PC3*	*PC4*
Ascending Process	APA, APC, APH, APW	0.5912	0.2395	0.1483	0.0210
Basal Plate	BPA, BPH	0.9022	0.0978		
Dorsal Complex	BPA, BPH, DSL	0.7111	0.2239	0.0650	
Size Axis	MBH, SDL	0.6214	0.3786		
Spine Axis	DSL, PSL	0.8269	0.1731		
Ventral Complex	PSL, VPA, VPL, VPW	0.5081	0.2841	0.1628	0.0450
Ventral Plates	VPA, VPL, VPW	0.6310	0.2951	0.0739	
Defensive Structure	All Traits	0.4092	0.1119	0.1102	0.0766

GWAS Indicates the Defensive Structure is a Polygenic System

19,540 SNPs were analyzed in 13 individual traits (including SDL) and 7 groupings of traits (**Figures 8-18**). I found an association on linkage group 4, with an LOD >40, with lateral plate counts similar to previous work done with this population (Alligood 2017), which provided confidence in our approach. I saw the same steep peak for lateral plate counts in the same region of the linkage group IV that the *Eda* gene is found, which gives me confidence in these findings (Colosimo et al. 2004). The baseline LOD for determining which SNPs are significant is LOD = 3, but this does not

optimally account for false discovery, so I also used a method of finding significance called the Bonferroni correction, which sets the number of trials against the number of traits. The line was set at 5.592, however, the Bonferroni correction is considered highly conservative for association studies. Because of this, I determined an LOD of 4 would be conservative enough to account for false discovery. Despite this, I found 2 peaks that had a LOD greater than the Bonferroni correction line, the dorsal spine length (DSL), and the spines PC1 (DSL and PSL).

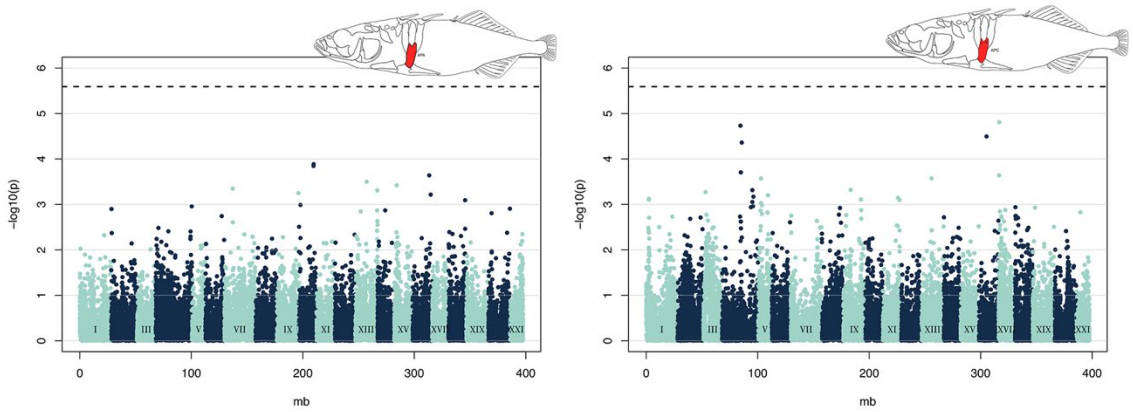

Figure 8. (Left) Ascending Process Area (APA) – Manhattan plot visualizing the association between genomic regions and the area of the ascending process. (Right) Ascending Process Fork Count (APC) – Manhattan plot visualizing the association between genomic regions and the number of forks on the ascending process.

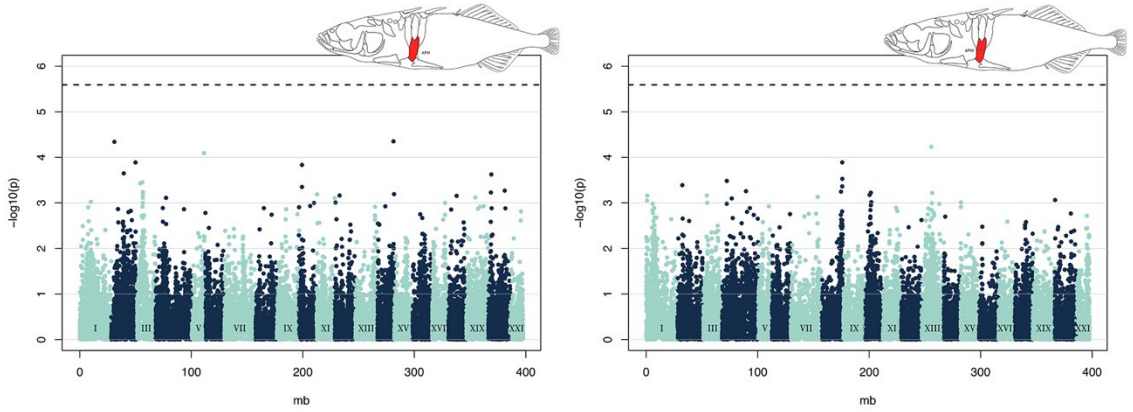

Figure 9. (Left) Ascending Process Height (APH) – Manhattan plot visualizing the association between genomic regions and the height of the ascending process. (Right) Ascending Process Width (APW) – Manhattan plot visualizing the association between genomic regions and the width of the ascending process.

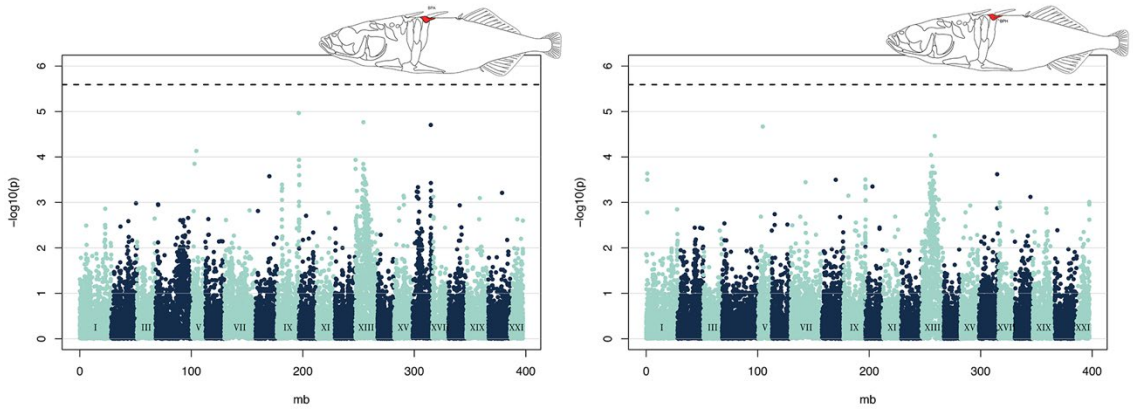

Figure 10. (Left) Basal Plate Area (BPA) – Manhattan plot visualizing the association between genomic regions and the area of the 2nd basal plate. (Right) Basal Plate Height (BPH) – Manhattan plot visualizing the association between genomic regions and the height of the 2nd basal plate.

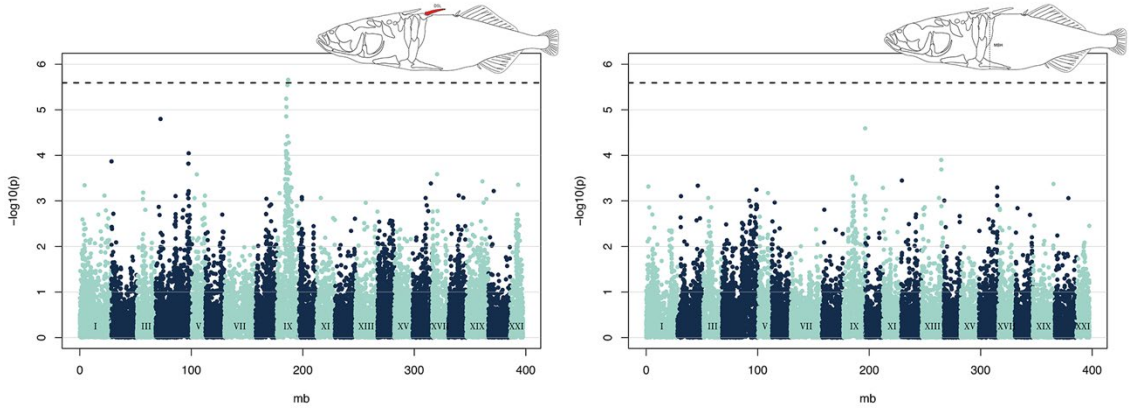

Figure 11. (Left) Dorsal Spine Length (DSL) – Manhattan plot visualizing the association between genomic regions and the length of the 2nd dorsal spine. (Right) Maximum Body Height (MBH) – Manhattan plot visualizing the association between genomic regions and the maximum body height.

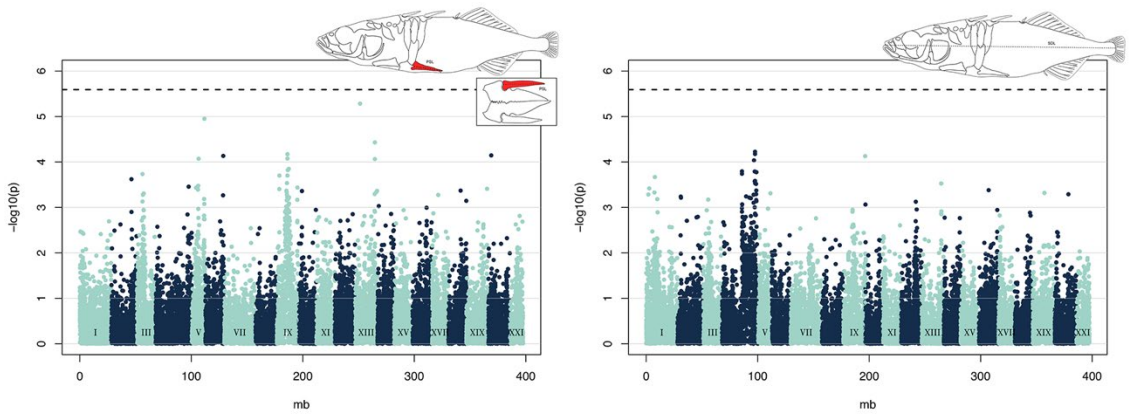

Figure 12. (Left) Pelvic Spine Length (PSL) – Manhattan plot visualizing the association between genomic regions and the length of the left pelvic spine. (Right) Standard Length (SDL) – Manhattan plot visualizing the association between genomic regions and the standard length of the fish.

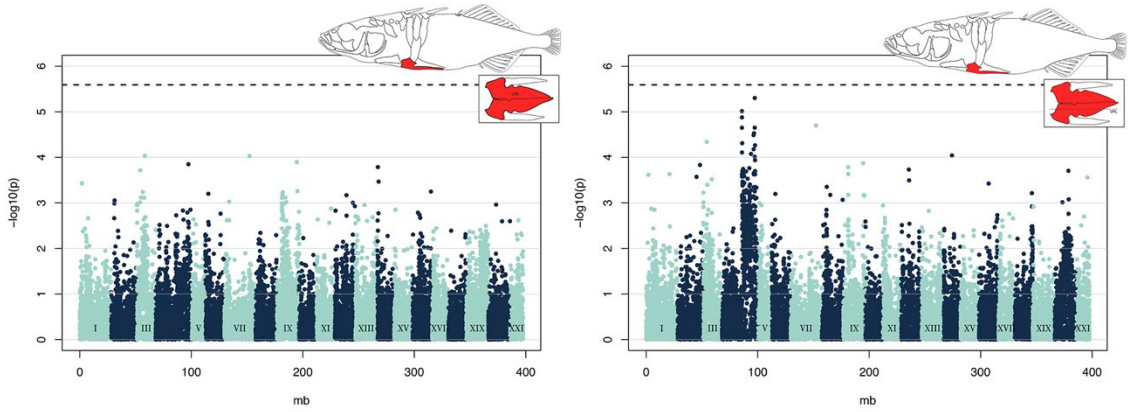

Figure 13. (Left) Ventral Plate Area (SDL) – Manhattan plot visualizing the association between genomic regions and the area of the ventral plates. (Right) Ventral Plate Length (VPA) – Manhattan plot visualizing the association between genomic regions and the length of the ventral plates.

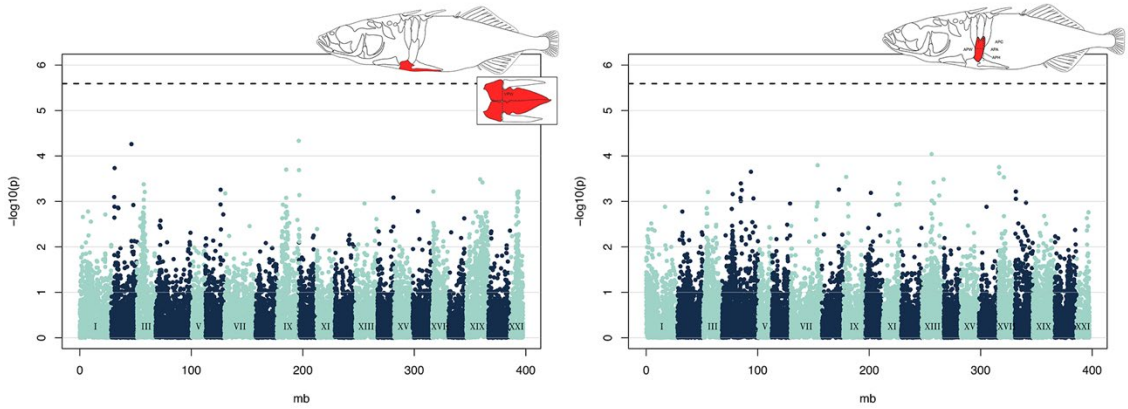

Figure 14. (Left) Ventral Plate Width (VPW) – Manhattan plot visualizing the association between genomic regions and the width of the ventral plates. (Right) Ascending Process PC1 – Manhattan plot visualizing the association between genomic regions and the PC1 score for APA, APC, APH, and APW.

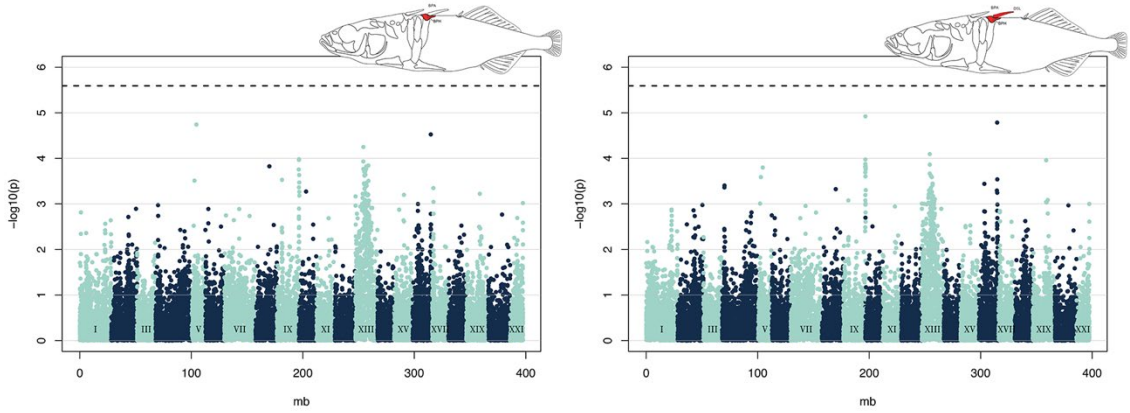

Figure 15. (Left) Basal Plate PC1 – Manhattan plot visualizing the association between genomic regions and the PC1 score for BPA and BPH. (Right) Dorsal Complex PC1 – Manhattan plot visualizing the association between genomic regions and the PC1 score for BPA, BPH, and DSL.

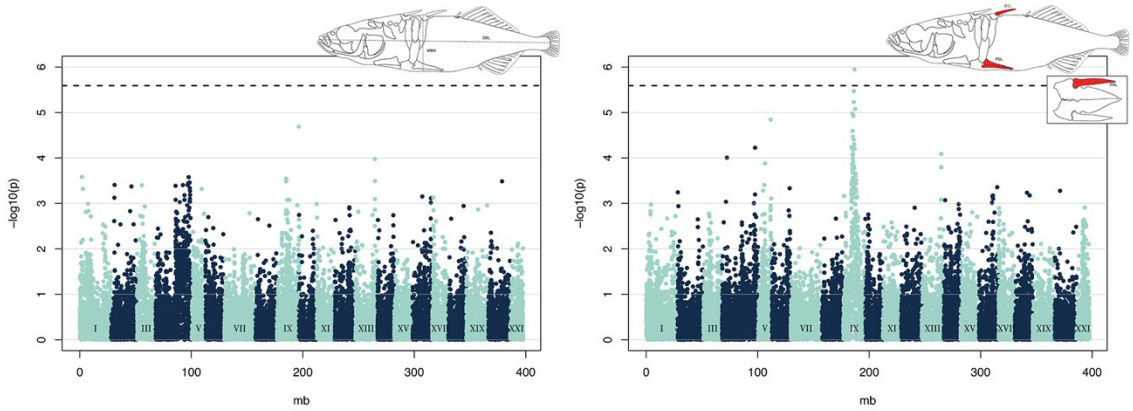

Figure 16. (Left) Size Axis PC1 – Manhattan plot visualizing the association between genomic regions and the PC1 score for MBH and SDL. (Right) Spine Axis PC1 – Manhattan plot visualizing the association between genomic regions and the PC1 score for DSL and PSL.

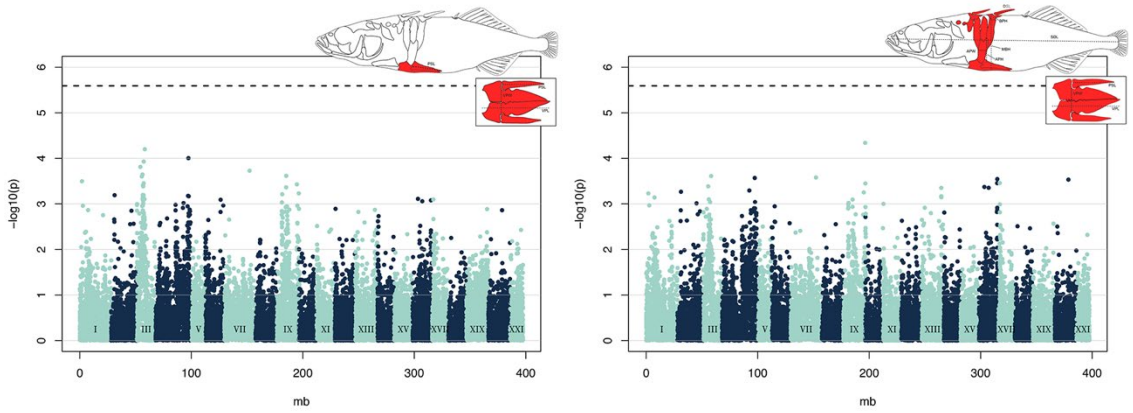

Figure 17. (Left) Ventral Complex PC1 – Manhattan plot visualizing the association between genomic regions and the PC1 score for PSL, VPA, VPL, VPW. (Right) Defensive Structure PC1 – Manhattan plot visualizing the association between genomic regions and the PC1 score for all defensive structure traits.

Peaks of Interest are Associated with the Basal Plate, Spines, Standard Length, and Ventral Plate Length

Significant peaks of association were found for the dorsal spine length (DSL) and the PC1 covariance for both the dorsal and pelvic spine lengths (DSL, PSL). I discovered 3 peaks above the Bonferroni correction line and several that were just below and might be associated with genetic regions that influence this trait due to the

conservative nature of the correction. The dorsal spine length (DSL) and spines PC1 (DSL & PSL) were highly significant. Other peaks of interest for us are the pelvic spine length (PSL), ventral plate length (VPL), the standard length (SDL), and the basal plate PC1 (**Table 3**).

Table 3. SNPs of interest for 4 different groups and traits. Bonferroni correction line is at a -log10 p-value of 5.592.

	SNPs of Interest		
Group/Trait	LG	*–log10 P-Value*	*Position on LG (mb)*
Basal Plate PC1 (BPA, BPH)	XIII	4.250	7.390
Spine Axis PC1 (DSL, PSL)	V	4.844	11.312
		3.878	6.158
	IX	5.950	10.122
		5.471	9.501
		5.230	9.512
SDL	IV	4.225	29.658
VPL	IV	5.302	29.424
		5.014	17.908

The basal plate PC1 combines both the basal plate area (BPA) and the basal plate height (BPH) and shows a strong association on LG XIII (**Figures 10, 15**). The highest SNP has a -log10 p-value of 4.250 at 7.390 mb (position on the linkage group) (**Figure 19**). The SNPs shows a large grouping of about 15 mb.

The spine axis PC1 combines both the dorsal spine length (DSL) and the pelvic spine length (PSL) and shows a strong association on LGs V and IX (**Figures 11, 12, 17**). The highest SNP on LG V has a -log10 p-value of 4.844 at 11.312 mb. The second highest SNP is located on a separate peak with a -log10 p-value of 3.878 at 6.158 mb (**Figure 20**). The highest SNP on LG IX has a -log10 p-value of 5.950 at 10.122 mb, the

second-highest has a -log10 p-value of 5.471 at 9.501 mb, and the third-highest has a -log10 p-value of 5.230 at 9.512 mb (**Figure 21**).

The standard length shows a strong association on LG IV (**Figure 13**). The highest SNP on LG IV has a -log10 p-value of 4.225 at 29.658 mb (**Figure 22**).

The ventral plate length shows a strong association on LG IV (**Figure 14**). The highest SNP on LG IV has a -log10 p-value of 5.302 at 29.424 mb and the second-highest SNP has a -log10 p-value of 5.014 at 17.908 mb (**Figure 23**).

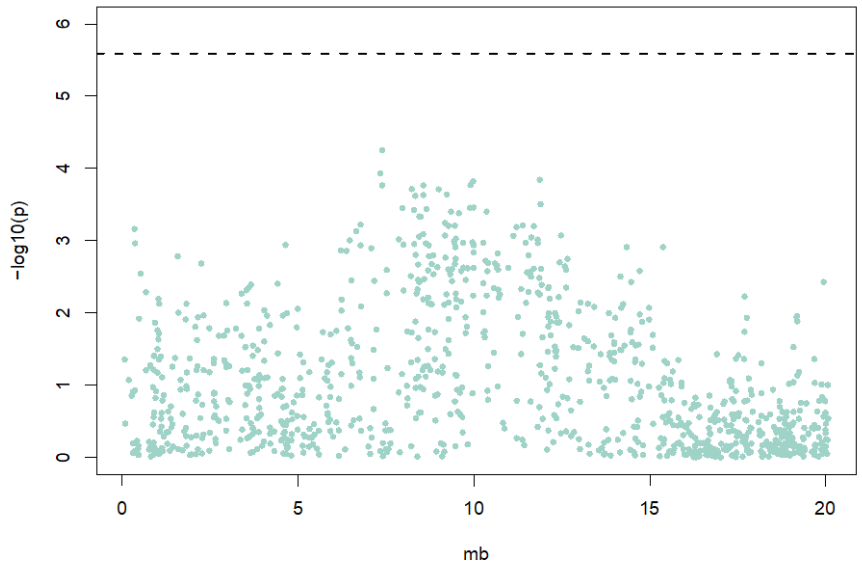

Figure 18. Basal Plate PC1, LG XIII – Manhattan plot visualizing the association between genomic regions and the PC1 score for BPA and BPH on LG XIII.

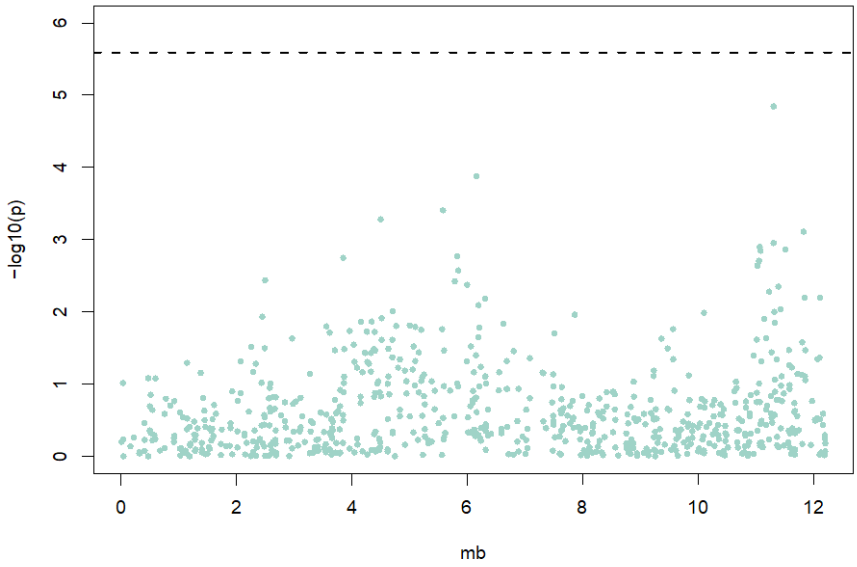

Figure 19. Spines PC1, LG V – Manhattan plot visualizing the association between genomic regions and the PC1 score for DSL and PSL on LG V.

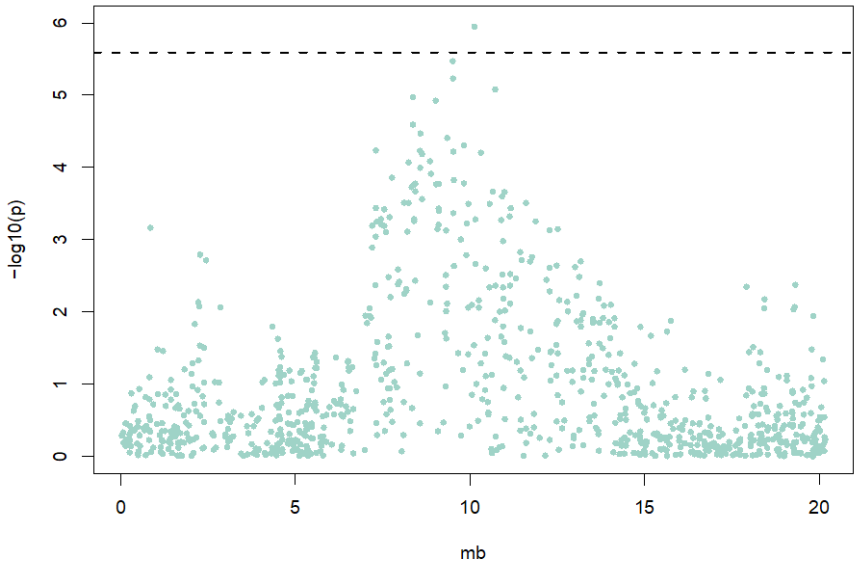

Figure 20. Spines PC1, LG IX – Manhattan plot visualizing the association between genomic regions and the PC1 score for DSL and PSL on LG IX.

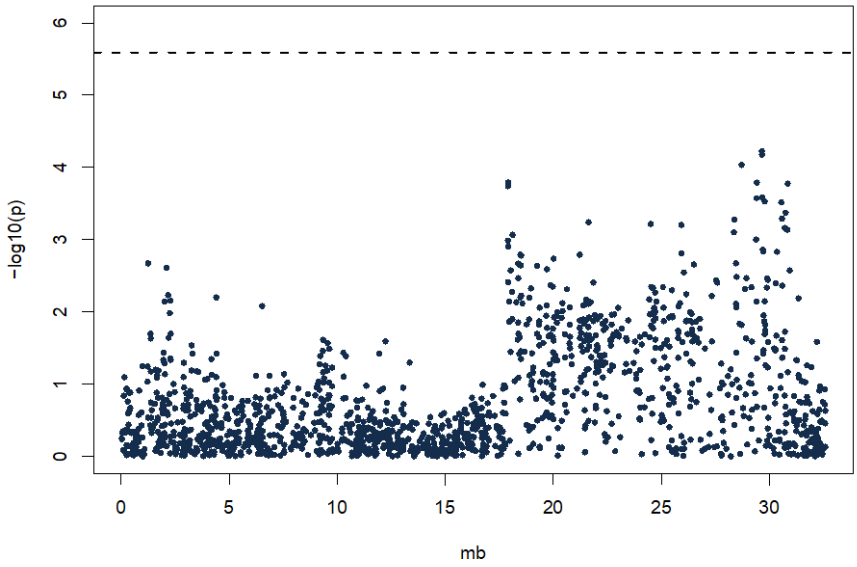

Figure 21. Standard Length (SDL), LG IV – Manhattan plot visualizing the association between genomic regions and the SDL on LG IV.

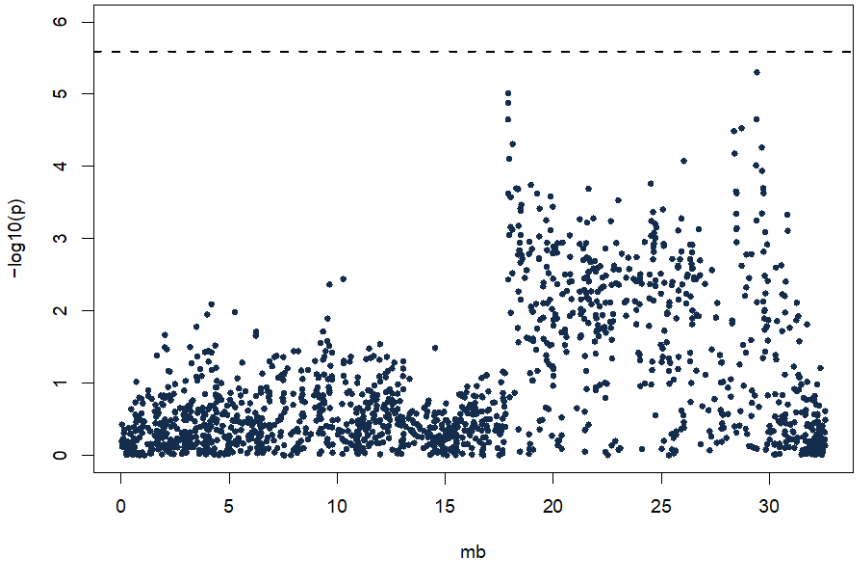

Figure 22. Ventral Plate Length (VPL), LG IV – Manhattan plot visualizing the association between genomic regions and the VPL on LG IV.

Discussion

The Riverbend Population of Stickleback: A Good System for Association Studies

Riverbend is a good model for studying overall variation between stickleback due to the intermixing of oceanic and freshwater forms (Currey 2014; Alligood 2017). Each of the defensive structure traits I examined showed high degrees of variance which is critical to the nature of the association study and may imply that the structure plays an important role in the fitness of the stickleback in different environments (Visscher et al. 2012; Lee et al. 2014; Nadeau et al. 2014; Pallares et al. 2014; Turner and Harr 2014; Wang et al. 2014; Brelsford et al. 2017). Counts and the more discrete linear measurements showed the greatest amount of variation, which could also be due to the effect these have on fitness (Hoekstra et al. 2006). A less discrete trait such as the width of the ascending process may not have a strong and direct fitness advantage or disadvantage as opposed to the body depth of the stickleback and thus, these smaller traits may not be selected for or against (Orr and Coyne 1992; Wood et al. 2014).

Alligood et al. (2017), who found that PC1 only explained ~5.624% of the covariation between opercle shape and lateral plate number in the Riverbend population. However, this makes sense because the opercle is not related to the lateral plate count in function or location and therefore likely evolves and varies separately.

In contrast, the PC1 for the entire defensive structure explained a high amount of the covariance (40.9%), which is positive for the purposes of this study because it means that the defensive structure is heavily influenced in the same direction: either a high amount of armor as in the marine phenotype or a low amount of armor in the freshwater phenotype. Reimchen (2013) saw similarly high covariance within the

defensive structure in hybrid Canadian populations of stickleback. This contributes to the evidence that the defensive structure is an important component to the fitness of individuals within a population (Bell and Foster 1994; Reimchen 2000; Marchinko and Schluter 2007; Barrett et al. 2008; Schluter et al. 2010).

The Defensive Structure is a Polygenic System

All of the analyzed traits showed multiple peaks of association and high background levels of SNP markers. The majority of the traits should be considered highly complex, in the sense that the width of a ventral plate may have something to do with defense but is probably not highly-specific to the defensive capabilities of an individual like lateral plate count or spine length (Davis et al. 1994; Ward and Kellis 2012). Instead of directly influencing defensive capabilities, the remaining traits likely play a synergistic role with each other, contributing as a whole instead of as an individual (Paaby and Rockman 2013). This is opposed to the Mendelian inheritance seen in several systems, including beach mice (Hoekstra et al. 2006) and the lateral plate counts seen in this thesis and in the stickleback literature (Reimchen et al. 1985; Cresko et al. 2004; Alligood 2017).

Unexplored Influences on Variation

Two factors were not included in this study, that have been reported to play a role in possible morphological differences between individuals: sexual dimorphism and bilateral asymmetry. Sexual dimorphism has been found to influence trait variation in stickleback (Leinonen et al. 2011). However, this should be at least partially controlled for by the size and randomness of our sample collection. Despite this, future work might

take sexual dimorphism into consideration. Bilateral asymmetry has been mostly observed in the ventral structure, between the two ventral plates, right and left (Trokovic et al. 2012). The pelvic spines have not been found to vary significantly between left and right sides of the stickleback. The ventral length, width, and area measurements performed in this thesis take into account both sides of the ventral structure and thus should not be influenced by this bilateral asymmetry.

Novel Genomic Regions Associate with Defensive Structure Traits in the Riverbend Population

Out of the multiple peaks of interest above LOD 4, I chose five that appeared to show the strongest association. These traits are the basal plate PC1, spine axis PC1, the standard length (SDL), and the ventral plate length (VPL). Excitingly, these traits all associate with genomic regions on the stickleback genome that have not been previously linked. In general, the literature has not addressed most of the indiscrete traits of the defensive structure, but for the more discrete traits, the literature has found different genomic regions associating with the traits.

The Basal Plate PC1 Associates with LG XIII

The basal plate PC1 incorporates both the basal plate area (BPA) and the basal plate height (BPH). Naturally, they covary with each other strongly (90.2% PC1), due to the association of linear height and area (**Table 2**). Association studies have neglected to analyze the properties of the basal plate. This is remiss as the basal plate plays a key role in stabilizing the dorsal spines and linking them to the rest of the defensive structure (Reimchen 1983). The area of the basal plates are also highly

variable, indicating a strong fitness benefit of having substantial or less substantial plates. The peak of interest was seen on LG XIII and is interesting not for the height of its peak but for the area the peak spans in genomic space (**Table 3, Figure 19**). Wide peaks can be attributed to several factors: the first, that there is not enough power in the sample size to make a strong association (Ward and Kellis 2012). The second, that there has been a recent recombination event that has affected the genome (Rieseberg et al. 1999; Rieseberg and Buerkle 2001). The third, is that there are inversions that subsequently affect more than one SNP (Jones et al. 2012; Kapun et al. 2016). In Jones et al. (2012), several stickleback genomes were sequenced, and chromosomal inversions were found on LGs I, XI, and XXI. This doesn't prove that such an inversion hasn't occurred on LG XIII in the Riverbend population but it may indicate the power of the study is not strong enough for the basal plates, as the recent recombination explanation is unlikely given what is known of the history of the Riverbend population (Currey 2014). Another explanation may be that the basal plates may be associated with a genomic island on LG XIII that includes genes that all have to be linked together in order for the individual to develop and function (Turner et al. 2005; Boyd et al. 2009; Feder and Nosil 2010).

The Standard Length and Ventral Plate Length Associate with LG IV

The standard length and ventral plate length have been understudied in the stickleback literature. Ventral plate length has been associated with LG VII and the *Pitx1* locus (Shapiro et al. 2004; Hohenlohe et al. 2010). The ventral plate length has been previously associated with LG IV in Shapiro et al. (2004) with the Gac4174 marker with a LOD of 4.7 at 32.4 cM. LG IV has been well-studied in the literature,

contributing to many of the genomic regions associated with discrete traits such as lateral plate counts. The association plots (**Figures 22, 23**) show a similarity between the standard length and the ventral plate length which may be due to an overlap in measuring, as the ventral plate length may vary with the standard length. However, the ventral plate length was size-standardized, which should remove any association with standard length. Also, in the PCA (**Figure 6**), the SDL and VPL did not show any covariance. The distal portion of LG IV may simply be a strong contributor to a variety of phenotypes in freshwater populations (Hohenlohe et al. 2010).

There are many more peaks of association that deserve analysis for previously-discovered and novel candidate genes. This thesis served as an exploratory analysis of the defensive structure as a system and is a starting point for future studies to further examine the novel genomic regions found in this study.

The Spine Axis PC1 Associates with LGs V and IX

Both the second dorsal spine length and the spine axis (the second dorsal spine length (DSL) and the pelvic spine length (PSL)) showed similar peaks of interest on LG V and IX, with the peak on LG IX rising above even the highly-conservative Bonferroni correction line. The peak on LG V is novel to what has been previously found with regards to both the dorsal and pelvic spine. The peak on LG IX is novel to the second dorsal spine, although the length of the first dorsal spine has been associated with LG IX in Japanese Sea stickleback (Kitano et al. 2009). The length of the second dorsal spine has been associated with several linkage groups including II, X (Miller et al. 2014), VIII, and XI (Peichel et al. 2001; Raeymaekers et al. 2007). The length of the pelvic spine has been associated with the Pitx1 gene on LG VII with minor peaks on LG

IV and II (Shapiro et al. 2004). It has also been associated with linkage group VIII (Peichel et al. 2001). The spines are of special interest, not only due to the strength of their association but also due to their discrete appearance and defensive function. Future studies should investigate the relationship between the first dorsal spine and the DSL and PSL because of the covariance between it and the DSL, to see if it associates with the same genome regions.

A Candidate Gene for the Spine Axis on LG IX

The peak on LG IX for the spine axis PC1 was the only peak above the Bonferroni correction line and thus demanded further examination. The highest four SNPs were located between 8.361 and 10.720 mb on LG IX, which became the region of interest for searching for a candidate gene. There are 94 named genes in that region, 11 of which were documented in the literature as being related to bone-growth or loss in orthologous regions in humans (**Table 4**) (Kiel et al. 2007; Xiao et al. 2010; Wang et al. 2014; Lee et al. 2015; Rauch et al. 2015; Watson et al. 2015; Abrahao-Machado et al. 2017; Feng et al. 2018). One in particular, prolyl 4-hydroxylase subunit beta (*P4HB*) stood out as the most likely candidate for further research. *P4HB* is a protein coding gene involved in metabolism and protein processing in the endoplasmic reticulum. In stickleback, it is located on LG IX between 10.226 and 10.236. In humans, it is located on LG XVII between 81.843 and 81.860 (Hunt et al. 2018).

Table 4. Annotated genes-of-interest within narrow region around SNP most highly associated with dorsal and pelvic spine length. Bold gene name indicates most-interesting candidate based on literature.

Start (bp)	Stop (bp)	Name	Ensembl ID	Supporting Literature
Literature or Human Patient-Supported Candidates				
8.355	8.365	*mtap*	ENSGACG00000017558	Abrahao-Machado et al. 2017
8.627	8.632	*ufsp2*	ENSGACG00000017761	Watson et al. 2015
8.825	8.827	*odam*	ENSGACG00000017898	Lee et al. 2015
9.406	9.408	*adra1d*	ENSGACG00000017994	Patient-Supported
9.498	9.505	*mettl14*	ENSGACG00000018009	Feng et al. 2018
9.646	9.651	*spock3*	ENSGACG00000018028	Kiel et al. 2007
9.72	9.734	*dlc1*	ENSGACG00000018034	Wang et al. 2014
10.07	10.08	*igfbp4*	ENSGACG00000018120	Xiao et al. 2010
10.123	10.13	*pip4k2b*	ENSGACG00000018135	Patient-Supported
10.131	10.134	*psmb3*	ENSGACG00000018151	Patient-Supported
10.226	**10.236**	***p4hb***	**ENSGACG00000018185**	**Rauch et al. 2015**

A heterozygous missense mutation in *P4HB* that impairs the disulfide isomerase activity of protein disulfide isomerase (PDI) is responsible for Cole-Carpenter syndrome: a bone fragility disorder distinguished by frequent fractures, craniosynostosis, ocular proptosis, hydrocephalus, and distinctive facial features (Cole and Carpenter 1987; Benham 2012; Rauch et al. 2015). Whether or not *P4HB* is controlling for the variation in spine lengths is yet to be seen and would require in-depth association studies that are hypothesis-focused. While I identify *P4HB* as the most likely candidate gene at this point in the study, the other 93 named genes in the chosen region, as well as the unnamed or unidentified genes in the region may also be controlling spine lengths. Finally, because there is a high level of background noise related to the spine axis, as well as the peak on LG V, this gene likely only plays a shared role in controlling spine lengths. Despite this knowledge, this finding represents an exciting association that could have positive consequences for human health as well as furthered understanding of the stickleback genome.

Conclusions

The defensive structure is composed of a complex set of bones that overlap and interact to protect the stickleback from predators. The structure varies widely between oceanic and freshwater populations of stickleback, implying that this structure is adaptive in nature. The Riverbend population has proven to be good for association studies of stickleback due to its hybrid population that has no population structure. Here, I find that for most of the individual traits and the complex as a whole (or in parts) may have a polygenic genetic basis. The basal plate (BPA, BPH) has been previously unstudied, but I found a strong association at linkage group XIII. I also

found associations between standard length (SDL) and the ventral plate length (VPL) on linkage group IV. The VPL has already been associated with LG IV but the SDL has not and contributes to the evidence that LG IV controls multiple phenotypes on the stickleback. Each of these traits are strong candidates for future, in-depth studies.

The genetic basis of the second dorsal and pelvic spine length (DSL, PSL) appear to be controlled by one or a few genetic loci and I have found evidence of never-before-seen genetic association on linkage group IX located between 8.361 and 10.720 mb on LG IX. This region contains 94 named genes, including a promising candidate, prolyl 4-hydroxylase subunit beta (*P4HB*) which has implications for Cole-Carpenter syndrome in humans. This candidate gene and the entire genomic region presents an exciting discovery for further research.

Bibliography

Abrahao-Machado, L. F., Antunes, B., Filippi, R. Z., Volc, S., Boldrini, E., Menezes, W. P., … de Camargo, O. P. (2017). Loss of MTAP expression is a negative prognostic marker in Ewing sarcoma family of tumors. *Biomarkers in Medicine*, *12*(1), 35–44. https://doi.org/10.2217/bmm-2017-0152

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*, *305*(5689), 1462–1465. https://doi.org/10.1126/science.1098095

Albertson, R.C. (2003). Genetic Basis of Adaptive Shape Differences in the Cichlid Head. 291-301. Journal of Heredity.

Alligood, K. S. (2017). Using Natural Populations of Threespine Stickleback to Identify the Genomic Basis of Skeletal Variation (Doctoral Thesis). University of Oregon, Eugene, OR.

Arya, G. H., Magwire, M. M., Huang, W., Serrano-Negron, Y. L., Mackay, T. F. C., & Anholt, R. R. H. (2015). The Genetic Basis for Variation in Olfactory Behavior in Drosophila melanogaster. *Chemical Senses*, *40*(4), 233–243. https://doi.org/10.1093/chemse/bjv001

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLOS ONE, 3(10), e3376.

Banbura, J. (1994). Lateral plate morph differentiation of freshwater and marine populations of the three-spined stickleback, *Gasterosteus aculeatus*, in Poland. *The Journal of Fish Biology*, 44:773-783.

Barrett, R. D. H., Rogers, S. M., & Schluter, D. (2008). Natural Selection on a Major Armor Gene in Threespine Stickleback. *Science*, *322*(5899), 255–257. https://doi.org/10.1126/science.1159978

Bateson, W., & Mendel, G. (1913). *Mendel's Principles of Heredity*. University Press.

Bell, M. A. (1998). Stickleback fishes: Bridging the gap between population biology and paleobiology. *Trends in Ecology & Evolution 3*(12), 320-324.

Bell, M. A. (2001). Lateral plate evolution in the threespine stickleback: getting nowhere fast. Genetica 112-113:445-461.

Bell, M. A., & Foster, Susan A. (1994). *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press.

Bell, M. A., & Ortí, G. (1994). Pelvic Reduction in Threespine Stickleback from Cook Inlet Lakes: Geographical Distribution and Intrapopulation Variation. *Copeia*, *1994*(2), 314–325. https://doi.org/10.2307/1446981

Benham, A. M. (2011). The Protein Disulfide Isomerase Family: Key Players in Health and Disease. *Antioxidants & Redox Signaling*, *16*(8), 781–789. https://doi.org/10.1089/ars.2011.4439

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., … Pomp, D. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences*, *107*(44), 18933–18938. https://doi.org/10.1073/pnas.1007028107

Blouw, D. M., & Boyd, G. J. (1992). Inheritance of reduction, loss, and asymmetry of the pelvis in Pungitius pungitius (ninespine stickleback). *Heredity*, *68*(1), 33–42. https://doi.org/10.1038/hdy.1992.4

Boyd, E. F., Almagro-Moreno, S., & Parent, M. A. (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends in Microbiology*, *17*(2), 47–53. https://doi.org/10.1016/j.tim.2008.11.003

Boyle Robert. (1667). An account of a very odd monstrous calf. *Philosophical Transactions of the Royal Society of London*, *1*(1), 10–10. https://doi.org/10.1098/rstl.1665.0007

Branicki, W., Liu, F., van Duijn, K., Draus-Barini, J., Pośpiech, E., Walsh, S., … Kayser, M. (2011). Model-based prediction of human hair color using DNA variants. *Human Genetics*, *129*(4), 443–454. https://doi.org/10.1007/s00439-010-0939-8

Brelsford Alan, Toews David P. L., & Irwin Darren E. (2017). Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1866), 20171106. https://doi.org/10.1098/rspb.2017.1106

Bruet, B. J. F., Song, J., Boyce, M. C., & Ortiz, C. (2008). Materials design principles of ancient fish armour. *Nature Materials*, *7*(9), 748–756. https://doi.org/10.1038/nmat2231

Buerkle, C. A., & Rieseberg, L. H. (2001). Low Intraspecific Variation for Genomic Isolation Between Hybridizing Sunflower Species. *Evolution*, *55*(4), 684–691. https://doi.org/10.1111/j.0014-3820.2001.tb00804.x

Cabot, E. L., Davis, A. W., Johnson, N. A., & Wu, C. I. (1994). Genetics of reproductive isolation in the Drosophila simulans clade: complex epistasis underlying hybrid male sterility. *Genetics*, *137*(1), 175–189.

Cantor, R. M., Lange, K., & Sinsheimer, J. S. (2010). Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application. American Journal of Human Genetics, 86(1), 6–22.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. (2011). Stacks: building and genotyping Loci de novo from short-read sequences. G3 (Bethesda) 1:171-182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W.A. Cresko. (2013). Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124-3140.

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., … Kingsley, D. M. (2010). Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science*, *327*(5963), 302–305. https://doi.org/10.1126/science.1182213

Charlesworth, B. (1990). Optimization Models, Quantitative Genetics, and Mutation. *Evolution*, *44*(3), 520–538. https://doi.org/10.1111/j.1558-5646.1990.tb05936.x

Cheverud, J. M. (1984). Quantitative genetics and developmental constraints on evolution by selection. Journal of Theoretical Biology *110*(2), 155-171.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, Jr., M. Dickson, J. Grimwood, J. Schmutz, R.M. Myers, D. Schluter, and D.M. Kingsley. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science 307:1928-1933.

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.

Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., … Peichel, C. L. (2015). Extent of QTL Reuse During Repeated Phenotypic Divergence of Sympatric Threespine Stickleback. *Genetics*, *201*(3), 1189–1200. https://doi.org/10.1534/genetics.115.182550

Cooke, F., & Buckley, P. A. (2013). *Avian Genetics: A Population and Ecological Approach*. Academic Press.

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., … Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proceedings of the National Academy of Sciences, 101(16), 6050–6055.

Cresko, W.A., K. L. McGuigan, P.C. Phillips, and J. H. Postlethwait. (2007). Studies of threespine stickleback developmental evolution: progress and promise. Genetica 129:105-126.

Currey Mark C., Bassham Susan, Perry Stephen, & Cresko William A. (2017). Developmental timing differences underlie armor loss across threespine stickleback populations. Evolution & Development, 19(6), 231–243.

Currey, M. C. (2014). The Phenotypic and Genetic Distribution of Threespine Stickleback that Inhabit the Willamette Basin, Oregon, USA (Masters Thesis). University of Oregon, Eugene, OR.

Currey, M. C., Bassham, S. L., & Cresko, W. A. (2019). Genetic divergence outpaces phenotypic divergence among threespine stickleback populations in old freshwater habitats. *BioRxiv*, 618017. https://doi.org/10.1101/618017

Darwin, Charles, 1809-1882. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London: John Murray.

David Cole, E. C., & Carpenter, T. O. (1987). Bone fragility, craniosynostosis, ocular proptosis, hydrocephalus, and distinctive facial features: A newly recognized type of osteogenesis imperfecta. *The Journal of Pediatrics*, *110*(1), 76–80. https://doi.org/10.1016/S0022-3476(87)80292-5

Davis, A. W., Noonburg, E. G., & Wu, C. I. (1994). Evidence for complex genic interactions between conspecific chromosomes underlying hybrid female sterility in the Drosophila simulans clade. *Genetics*, *137*(1), 191–199.

DeFaveri, J., & Merilä, J. (2013). Variation in Age and Size in Fennoscandian Three-Spined Sticklebacks (Gasterosteus aculeatus). PLoS ONE, 8(11).

DeFaveri, J., T. Shikano, Y. Shimada, A. Goto, and J. Merila. (2011). Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (Gasterosteus aculeatus). Evolution 65:1800-1807.

Dunn, L. C. (1991). *A Short History of Genetics: The Development of Some of the Main Lines of Thought, 1864-1939*. Iowa State University Press.

Feder, J. L., & Nosil, P. (2010). The Efficacy of Divergence Hitchhiking in Generating Genomic Islands During Ecological Speciation. *Evolution*, *64*(6), 1729–1747. https://doi.org/10.1111/j.1558-5646.2009.00943.x

Feng, Z., Li, Q., Meng, R., Yi, B., & Xu, Q. (2018). METTL3 regulates alternative splicing of MyD88 upon the lipopolysaccharide-induced inflammatory response in human dental pulp cells. *Journal of Cellular and Molecular Medicine*, *22*(5), 2558–2568. https://doi.org/10.1111/jcmm.13491

Fisher, R.A. (1930). *The Genetical Theory of Natural Selection*. The Clarendon Press.

Fisher, R.A. (1931). The Evolution of Dominance. *Biological Reviews*, *6*(4), 345-368. https://doil.org/10.1111/j.1469-185X.1931.tb01030.x

Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., … McCarthy, M. I. (2016). The genetic architecture of type 2 diabetes. *Nature*, *536*(7614), 41–47. https://doi.org/10.1038/nature18642

Gessner, Conrad. (1551). *Historiæ animalivm*. Tigvri. https://doi.org/10.5962/bhl.title.125499

Glazer, A. M., Cleves, P. A., Erickson, P. A., Lam, A. Y., & Miller, C. T. (2014). Parallel developmental genetic features underlie stickleback gill raker evolution. *EvoDevo*, *5*(1), 19. https://doi.org/10.1186/2041-9139-5-19

Grant, P. R., & Grant, B. R. (2002). Unpredictable Evolution in a 30-Year Study of Darwin's Finches. *Science*, *296*(5568), 707–711. https://doi.org/10.1126/science.1070315

Grant, P. R., & Grant, B. R. (2006). Evolution of Character Displacement in Darwin's Finches. *Science*, *313*(5784), 224–226. https://doi.org/10.1126/science.1128374

Hansen, T. F. (2006). The Evolution of Genetic Architecture. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 123–157. https://doi.org/10.1146/annurev.ecolsys.37.091305.110224

Hatfield, T. (1997). Genetic Divergence in Adaptive Characters between Sympatric Species of Stickleback. *The American Naturalist*, *149*(6), 1009–1029. https://doi.org/10.1086/286036

Hatfield, T., & Schluter, D. (1996). A Test for Sexual Selection on Hybrids of Two Sympatric Sticklebacks. *Evolution*, *50*(6), 2429–2434. https://doi.org/10.1111/j.1558-5646.1996.tb03629.x

Hatfield, T., Barton, N., & Searle, J. B. (1992). A Model of a Hybrid Zone Between Two Chromosomal Races of the Common Shrew (sorex Araneus). *Evolution*, *46*(4), 1129–1145. https://doi.org/10.1111/j.1558-5646.1992.tb00624.x

Hendry, A. P., Hudson, K., Walker, J. A., Räsänen, K., & Chapman, L. J. (2011). Genetic divergence in morphology–performance mapping between Misty Lake and inlet stickleback. *Journal of Evolutionary Biology*, *24*(1), 23–35. https://doi.org/10.1111/j.1420-9101.2010.02155.x

Hippel, F. von. (2010). *Tinbergen's Legacy in Behaviour: Sixty Years of Landmark Stickleback Papers*. BRILL.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. PLoS Genetics, 6(2), e1000862.

Hohenlohe, P. A., Hand, B. K., Andrews, K. R., & Luikart, G. (2018). Population Genomics Provides Key Insights in Ecology and Evolution. In O. P. Rajora (Ed.), Population Genomics: Concepts, Approaches and Applications (pp. 483–510).

Hohenlohe, P. A., J. Catchen, and W. A. Cresko. (2012). Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. Methods Mol Biol 888:235-260.

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science (New York, N.Y.)*, *313*(5783), 101–104. https://doi.org/10.1126/science.1126121

Hoogland, R., Morris, D., & Tinbergen, N. (1956). The Spines of Sticklebacks (Gasterosteus and Pygosteus) as Means of Defence against Predators (Perca and Esox). *Behaviour*, *10*(3/4), 205–236.

Hunt, J., Breuker, C. J., Sadowski, J. A., & Moore, A. J. (2009). Male–male competition, female mate choice and their interaction: determining total sexual selection. *Journal of Evolutionary Biology*, *22*(1), 13–26. https://doi.org/10.1111/j.1420-9101.2008.01633.x

Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution*, *39*(1), 57–106. https://doi.org/10.1006/jhev.2000.0403

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, P. Broad Institute Genome Sequencing, T. Whole Genome Assembly, J. Baldwin, T. Bloom, D. B. Jaffe, R. Nicol, J. Wilkinson, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55-61.

Kalbe, M., & Kurtz, J. (2006). Local differences in immunocompetence reflect resistance of sticklebacks against the eye fluke Diplostomum pseudospathaceum. Parasitology, 132(1), 105–116. https://doi.org/10.1017/S0031182005008681

Kapun, M., Fabian, D. K., Goudet, J., & Flatt, T. (2016). Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. Molecular Biology and Evolution, 33(5), 1317–1336. https://doi.org/10.1093/molbev/msw016

Kiel, D. P., Demissie, S., Dupuis, J., Lunetta, K. L., Murabito, J. M., & Karasik, D. (2007). Genome-wide association with bone mass and geometry in the Framingham Heart Study. BMC Medical Genetics, 8 Suppl 1, S14. https://doi.org/10.1186/1471-2350-8-S1-S14

Kimmel, C. B., Ullmann, B., Walker, C., Wilson, C., Currey, M., Phillips, P. C., … Cresko, W. A. (2005). Evolution and development of facial bone morphology in threespine sticklebacks. *Proceedings of the National Academy of Sciences*, *102*(16), 5791–5796. https://doi.org/10.1073/pnas.0408533102

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*, *18*(10), 769–774. https://doi.org/10.1016/j.cub.2008.04.027

Kitano, J., Ross, J. A., Mori, S., Kume, M., Jones, F. C., Chan, Y. F., … Peichel, C. L. (2009). A role for a neo-sex chromosome in stickleback speciation. Nature, 461(7267), 1079–1083.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*(1), 29. https://doi.org/10.1186/1746-4811-9-29

Kruuk, L. E. B., Baird, S. J. E., Gale, K. S., & Barton, N. H. (1999). A Comparison of Multilocus Clines Maintained by Environmental Adaptation or by Selection Against Hybrids. *Genetics*, *153*(4), 1959–1971.

Kusakabe, M., Ishikawa, A., Ravinet, M., Yoshida, K., Makino, T., Toyoda, A., … Kitano, J. (2017). Genetic basis for variation in salinity tolerance between stickleback ecotypes. *Molecular Ecology*, *26*(1), 304–319. https://doi.org/10.1111/mec.13875

Lavin, P. A., & McPhail, J. D. (1986). Adaptive Divergence of Trophic Phenotype among Freshwater Populations of the Threespine Stickleback (*Gasterosteus aculeatus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *43*(12), 2455–2463. https://doi.org/10.1139/f86-305

Lee, H.-K., Ji, S., Park, S.-J., Choung, H.-W., Choi, Y., Lee, H.-J., … Park, J.-C. (2015). Odontogenic Ameloblast-associated Protein (ODAM) Mediates Junctional Epithelium Attachment to Teeth via Integrin-ODAM-Rho Guanine Nucleotide Exchange Factor 5 (ARHGEF5)-RhoA Signaling. *The Journal of Biological Chemistry*, *290*(23), 14740–14753. https://doi.org/10.1074/jbc.M115.648022

Lees, D. R., & Creed, E. R. (1975). Industrial Melanism in Biston betularia: The Rôle of Selective Predation. Journal of Animal Ecology, 44(1), 67–83.

Leinonen, T., Cano, J. M., & Merilä, J. (2011). Genetic basis of sexual dimorphism in the threespine stickleback Gasterosteus aculeatus. Heredity, 106(2), 218–227.

Leonelli, S., & Ankeny, R. A. (2013). What makes a model organism? Endeavour, 37(4), 209–212.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, Z., Guo, B., Yang, J., Herczeg, G., Gonda, A., Balázs, G., … Merilä, J. (2017). Deciphering the genomic architecture of the stickleback brain with a novel multilocus gene-mapping approach. *Molecular Ecology*, *26*(6), 1557–1575. https://doi.org/10.1111/mec.14005

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nature Methods*, *8*(10), 833–835. https://doi.org/10.1038/nmeth.1681

Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., & Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. Nature Methods, 9(6), 525–526.

Liu, F., van Duijn, K., Vingerling, J. R., Hofman, A., Uitterlinden, A. G., Janssens, A. C. J. W., & Kayser, M. (2009). Eye color and the prediction of complex phenotypes from genotypes. *Current Biology*, *19*(5), R192–R193. https://doi.org/10.1016/j.cub.2009.01.027

Liu, J., Shikano, T., Leinonen, T., Cano, J. M., Li, M.-H., & Merilä, J. (2014). Identification of Major and Minor QTL for Ecologically Important Morphological Traits in Three-Spined Sticklebacks (Gasterosteus aculeatus). G3: Genes, Genomes, Genetics, 4(4), 595–604.

Marchinko, K. B. (2009). Predation's Role in Repeated Phenotypic and Genetic Divergence of Armor in Threespine Stickleback. Evolution, 63(1), 127–138.

Marchinko, K. B., & Schluter, D. (2007). Parallel Evolution by Correlated Response: Lateral Plate Reduction in Threespine Stickleback. *Evolution*, *61*(5), 1084–1090. https://doi.org/10.1111/j.1558-5646.2007.00103.x

McGuigan, K., Nishimura, N., Currey, M., Hurwit, D., & Cresko, W. A. (2011). Cryptic Genetic Variation and Body Size Evolution in Threespine Stickleback. *Evolution*, *65*(4), 1203–1211. https://doi.org/10.1111/j.1558-5646.2010.01195.x

McPhail, J. D. (1969). Predation and the Evolution of a Stickleback (*Gasterosteus*). *Journal of the Fisheries Research Board of Canada*, *26*(12), 3183–3208. https://doi.org/10.1139/f69-301

McPhail, J. D. (1984). Ecology and evolution of sympatric sticklebacks (Gasterosteus): morphological and genetic evidence for a species pair in Enos Lake, British Columbia. *Canadian Journal of Zoology*, *62*(7), 1402–1408. https://doi.org/10.1139/z84-201

McPhail, J. D. (1992). Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. *Canadian Journal of Zoology*, *70*(2), 361–369. https://doi.org/10.1139/z92-054

McPhail, J. D. (1993). Ecology and evolution of sympatric sticklebacks (Gasterosteus): origin of the species pairs. *Canadian Journal of Zoology*, *71*(3), 515–523. https://doi.org/10.1139/z93-072

Mendel, G. (1866). Versuche über Pflanzen-Hybriden. *Verhandlungen des Naturforschenden Vereines, Abhandlungern, Brünn*, **4**: 3– 47. Editions in different languages published by Matlová (1973).

Miller, C. T., Glazer, A. M., Summers, B. R., Blackman, B. K., Norman, A. R., Shapiro, M. D., … Kingsley, D. M. (2014). Modular Skeletal Evolution in Sticklebacks Is Controlled by Additive and Clustered Quantitative Trait Loci. Genetics, 197(1), 405–420.

Moodie, G. E. E. (1972). Predation, natural selection and adaptation in an unusual threespine stickleback. *Heredity*, *28*(2), 155–167. https://doi.org/10.1038/hdy.1972.21

Mouse Genome Sequencing Consortium, Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., … Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature, 420(6915), 520–562.

Mummert, A., Esche, E., Robinson, J., & Armelagos, G. J. (2011). Stature and robusticity during the agricultural transition: evidence from the bioarchaeological record. *Economics and Human Biology*, *9*(3), 284–301. https://doi.org/10.1016/j.ehb.2011.03.004

Nadeau, N. J., Ruiz, M., Salazar, P., Counterman, B., Medina, J. A., Ortiz-Zuazaga, H., … Papa, R. (2014). Population genomics of parallel hybrid zones in the mimetic butterflies, H. melpomene and H. erato. *Genome Research*, *24*(8), 1316–1333. https://doi.org/10.1101/gr.169292.113

Ólafsdóttir, G. Á., Snorrason, S. S., & Ritchie, M. G. (2007). Postglacial intra-lacustrine divergence of Icelandic threespine stickleback morphs in three neovolcanic lakes. *Journal of Evolutionary Biology*, *20*(5), 1870–1881. https://doi.org/10.1111/j.1420-9101.2007.01375.x

Olsen, B. R., A. M. Reginato, and W. F. Wang.(2000). Bone development. Annu Rev Cell Dev Bi 16:191-220.

Orr, H.A. & Coyne, J.A. (1990). The Genetics of Adaptation: A Reassessment. *The American Naturalist*. *140*(5), 725-742.

Paaby, A. B., & Rockman, M. V. (2013). The many faces of pleiotropy. Trends in Genetics : TIG, 29(2), 66–73.

Pallares, L. F., Harr, B., Turner, L. M., & Tautz, D. (2014). Use of a natural hybrid zone for genomewide association mapping of craniofacial traits in the house mouse. Molecular Ecology, 23(23), 5756–5770.

Parsons, P. A. & Bodmer, W. F. (1961). The Evolution of Overdominance: Natural Selection and Heterozygote Advantage. *Nature*, *190*(4770), 7-12.

Pasaniuc, B., & Price, A. L. (2017). Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics*, *18*(2), 117–127. https://doi.org/10.1038/nrg.2016.142

Peichel, C. L., K. S. Nereng, K. A. Ohgi, B. L. Cole, P. F. Colosimo, C. A. Buerkle, D. Schluter, and D. M. Kingsley. (2001). The genetic architecture of divergence between threespine stickleback species. Nature 414:901-905.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81:559-575.

Raeymaekers, J. a. M., Houdt, J. K. J. V., Larmuseau, M. H. D., Geldof, S., & Volckaert, F. a. M. (2007). Divergent selection as revealed by PST and QTL-based FST in three-spined stickleback (Gasterosteus aculeatus) populations along a coastal-inland gradient. Molecular Ecology, 16(4), 891–905.

Rauch, F., Fahiminiya, S., Majewski, J., Carrot-Zhang, J., Boudko, S., Glorieux, F., … Moffatt, P. (2015). Cole-Carpenter syndrome is caused by a heterozygous missense mutation in P4HB. *American Journal of Human Genetics*, *96*(3), 425–431. https://doi.org/10.1016/j.ajhg.2014.12.027

Reimchen, T. E. (1983). Structural Relationships Between Spines and Lateral Plates in Threespine Stickleback (Gasterosteus Aculeatus). Evolution, 37(5), 931–946.

Reimchen, T. E. (1992). Injuries on Stickleback from Attacks by a Toothed Predator (oncorhynchus) and Implications for the Evolution of Lateral Plates. *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., Stinson, E. M., & Nelson, J. S. (1985). Multivariate differentiation of parapatric and allopatric populations of threespine stickleback in the Sangan River watershed, Queen Charlotte Islands. *Canadian Journal of Zoology*, *63*(12), 2944–2951. https://doi.org/10.1139/z85-441

Reimchen, T.E., & Nosil, P. (2006). Replicated ecological landscapes and the evolution of morphological diversity among Gasterosteus populations from an archipelago on the west coast of Canada. Canadian Journal of Zoology, 84(5), 643–654.

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

Reimchen, Thomas E., & Bergstrom, C. A. (2009). The Ecology of Asymmetry in Stickleback Defense Structures. Evolution, 63(1), 115–126.

Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-Throughput Sequencing Technologies. *Molecular Cell*, *58*(4), 586–597. https://doi.org/10.1016/j.molcel.2015.05.004

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83*(4), 363–372. https://doi.org/10.1046/j.1365-2540.1999.00617.x

Roll-Hansen, N. (2009). Sources of Wilhelm Johannsen's genotype theory. Journal of the History of Biology, 42(3), 457–493.

Rundle, H. D., Nagel, L., Boughman, J. W., & Schluter, D. (2000). Natural Selection and Parallel Speciation in Sympatric Sticklebacks. *Science*, *287*(5451), 306–308. https://doi.org/10.1126/science.287.5451.306

Schluter, D. (1993). Adaptive Radiation in Sticklebacks: Size, Shape, and Habitat Use Efficiency. *Ecology*, *74*(3), 699–709. https://doi.org/10.2307/1940797

Schluter, D. (1995). Adaptive Radiation in Sticklebacks: Trade-Offs in Feeding Performance and Growth. *Ecology*, *76*(1), 82–90. https://doi.org/10.2307/1940633

Schluter, D. (1995). Adaptive Radiation in Sticklebacks: Trade-Offs in Feeding Performance and Growth. *Ecology*, *76*(1), 82–90. https://doi.org/10.2307/1940633

Schluter, D., Marchinko, K. B., Barrett, R. D. H., & Rogers, S. M. (2010). Natural selection and the genetics of adaptation in threespine stickleback. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2479–2486. https://doi.org/10.1098/rstb.2010.0036

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., … Kingsley, D. M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature, 428(6984), 717.

Sir, R. A. F., & Fisher, R. A. (1999). *The Genetical Theory of Natural Selection: A Complete Variorum Edition*. OUP Oxford.

Smith, M. W., & O'Brien, S. J. (2005). Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nature Reviews Genetics*, *6*(8), 623. https://doi.org/10.1038/nrg1657

Song, J., Reichert, S., Kallai, I., Gazit, D., Wund, M., Boyce, M. C., & Ortiz, C. (2010). Quantitative microstructural studies of the armor of the marine threespine stickleback (Gasterosteus aculeatus). *Journal of Structural Biology*, *171*(3), 318–331. https://doi.org/10.1016/j.jsb.2010.04.009

Stulp, G., & Barrett, L. (2016). Evolutionary perspectives on human height variation. *Biological Reviews*, *91*(1), 206–234. https://doi.org/10.1111/brv.12165

Taylor, E. B., & McPhail, J. D. (1986). Prolonged and burst swimming in anadromous and freshwater threespine stickleback, *Gasterosteus aculeatus*. *Canadian Journal of Zoology*, *64*(2), 416–420. https://doi.org/10.1139/z86-064

Trokovic, N., Herczeg, G., Ghani, N. I. A., Shikano, T., & Merilä, J. (2012). High levels of fluctuating asymmetry in isolated stickleback populations. BMC Evolutionary Biology, 12(1), 115.

Turner, L. M., & Harr, B. (n.d.). *Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions*. 25.

Turner, L. M., & Harr, B. (n.d.). *Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions*. 25.

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic Islands of Speciation in Anopheles gambiae. *PLOS Biology*, *3*(9), e285. https://doi.org/10.1371/journal.pbio.0030285

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five Years of GWAS Discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Walker, J. A. (1997). Ecological morphology of lacustrine threespine stickleback Gasterosteus aculeatus L. (Gasterosteidae) body shape. *Biological Journal of the Linnean Society*, *61*(1), 3–50. https://doi.org/10.1111/j.1095-8312.1997.tb01777.x

Wallace, RA. (1871). *Contributions to the Theory of Natural Selection*. Macmillan.

Wang, G., Yang, E., Smith, K. J., Zeng, Y., Ji, G., Connon, R., … Cai, J. J. (2014). Gene expression responses of threespine stickleback to salinity: implications for salt-sensitive hypertension. *Frontiers in Genetics*, *5*. https://doi.org/10.3389/fgene.2014.00312

Wang, Y., Lei, R., Zhuang, X., Zhang, N., Pan, H., Li, G., … Hu, G. (2014). DLC1-dependent parathyroid hormone-like hormone inhibition suppresses breast cancer bone metastasis. *The Journal of Clinical Investigation*, *124*(4), 1646–1659. https://doi.org/10.1172/JCI71812

Wang, Y., Lei, R., Zhuang, X., Zhang, N., Pan, H., Li, G., … Hu, G. (2014). DLC1-dependent parathyroid hormone-like hormone inhibition suppresses breast cancer bone metastasis. *The Journal of Clinical Investigation*, *124*(4), 1646–1659. https://doi.org/10.1172/JCI71812

Ward, L. D., & Kellis, M. (2012). Interpreting noncoding genetic variation in complex traits and human disease. *Nature Biotechnology*, *30*(11), 1095–1106. https://doi.org/10.1038/nbt.2422

Watson, C. M., Crinnion, L. A., Gleghorn, L., Newman, W. G., Ramesar, R., Beighton, P., & Wallis, G. A. (2015). Identification of a mutation in the ubiquitin-fold modifier 1-specific peptidase 2 gene, UFSP2, in an extended South African family with Beukes hip dysplasia. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*, *105*(7), 558–563. https://doi.org/10.7196/SAMJnew.7917

Watson, C. M., Crinnion, L. A., Gleghorn, L., Newman, W. G., Ramesar, R., Beighton, P., & Wallis, G. A. (2015). Identification of a mutation in the ubiquitin-fold modifier 1-specific peptidase 2 gene, UFSP2, in an extended South African family with Beukes hip dysplasia. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*, *105*(7), 558–563. https://doi.org/10.7196/SAMJnew.7917

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Weigel, D., & Nordborg, M. (2005). Natural Variation in Arabidopsis. How Do We Find the Causal Genes? *Plant Physiology*, *138*(2), 567–568. https://doi.org/10.1104/pp.104.900157

Wiig, E., Reseland, J. E., Østbye, K., Haugen, H. J., & Vøllestad, L. A. (2016). Variation in Lateral Plate Quality in Threespine Stickleback from Fresh, Brackish and Marine Water: A Micro-Computed Tomography Study. *PLOS ONE*, *11*(10), e0164578. https://doi.org/10.1371/journal.pone.0164578

Williams, G. C. (1966) *Adaptation and Natural Selection*. Princeton University Press.

Williams, G. C. (2018). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., … Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–1186. https://doi.org/10.1038/ng.3097

Wood, E. J. (1995). The encyclopedia of molecular biology: Editor in Chief, Sir John Kendrew, Executive Editor, Eleanor Lawrence. pp 1165. Blackwell Science, Oxford. 1994. £99.50. *Biochemical Education*, *23*(2), 105. https://doi.org/10.1016/0307-4412(95)90659-2

Wu, T. D. and C. K. Watanabe. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859-1875.

Xiao, Y., Cui, J., Li, Y.-X., Shi, Y.-H., & Le, G.-W. (2010). Expression of genes associated with bone resorption is increased and bone formation is decreased in mice fed a high-fat diet. *Lipids*, *45*(4), 345–355. https://doi.org/10.1007/s11745-010-3397-0

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., … Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569. https://doi.org/10.1038/ng.608

Yu, J., & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, *17*(2), 155–160. https://doi.org/10.1016/j.copbio.2006.02.003

Zhou, X., & Stephens, M. (2012). Genome-wide Efficient Mixed Model Analysis for Association Studies. *Nature Genetics*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310

Zou, W., & Zeng, Z.-B. (2008). Statistical Methods for Mapping Multiple QTL. International Journal of Plant Genomics, 2008.