Exploring the Evolutionary Significance of Chromosomal Fusions and Inversions: Implications for Adaptive Evolution and Sex Chromosome Evolution

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Zuyao Liu

von China

Leiterin der Arbeit:
Prof. Dr. Catherine. L. Peichel
Institut für Ökologie und Evolution, Universität Bern

Creative Commons License Deed

Attribution 4.0 International (CC BY 4.0)

This is a human-readable summary of (and not a substitute for) the license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

Bedeutung Chromosomaler Fusionen und Inversionen für die Evolutive Anpassung und Evolution von Sexchromosomen

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Zuyao Liu

von China

Leiterin der Arbeit:
Prof. Dr. Catherine. L. Peichel
Institut für Ökologie und Evolution, Universität Bern

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, 22.6.2023

Der Dekan
Prof. Dr. Marco Herwegh

Public Defense

Institute of Ecology and Evolution,

University of Bern,

22. June 2023

Composition of the jury:

Chair: Prof. Dr. Claudia Bank

Supervisor: Prof. Dr. Catherine. L. Peichel

Examiner: Prof. Dr. Qi Zhou, Zhejiang University

Table of Contents

Summary

For several decades, identifying the genetic basis of evolutionary changes has been a primary focus of evolutionary biology. While empirical studies have largely concentrated on single nucleotide polymorphisms (SNPs), chromosomal rearrangements have garnered less attention due to the challenges associated with accurate detection via traditional sequencing approaches. Nevertheless, it is predicted that chromosomal rearrangements can have a greater impact on evolutionary processes since they affect more genomic regions and elements. Further, due to the ability of chromosomal rearrangements to modify the recombination landscape, they have been hypothesized to play a significant role in various evolutionary processes, such as adaptation to distinct environments and the evolution of sex chromosomes.

This thesis first utilized stickleback species (Gasterosteidae) as a model system to explore the role of chromosomal fusion in adaptive evolution. Sticklebacks, particularly threespine sticklebacks (*Gasterosteus aculeatus*), have gained attention for their ability to colonize and adapt to freshwater environments from ancestral marine habitats over the past several million years. Notably, repeated patterns of phenotypic variation between the marine and freshwater ecotypes have been observed, making sticklebacks a unique and valuable system for identifying the genetic changes underlying adaptive evolution. In Chapter 1, I generated a fourspine stickleback (*Apeltes quadracus*) genome assembly and identified two fusion events in *G. aculeatus* by comparative genomics. On the two fused chromosomes, I also found an enrichment of adaptive quantitative trait loci (QTL) and population genomic signals of selection between marine and freshwater ecotypes of *G. aculeatus*. My research suggests that adaptive clusters on the fused chromosomes in *G. aculeatus* have more likely arisen from new mutations that occurred after the fusion rather than the linking of pre-existing adaptive alleles.

Aside from their recognized role as a model system to study adaptive evolution, sticklebacks have remarkable variation in sex chromosome composition, even among closely-related species. Furthermore, chromosomal inversions have been observed on the sex chromosomes of *G. aculeatus* and are believed to have contributed to the formation of distinct evolutionary strata. In Chapter 2, I investigated the sex chromosome system of *A.*

quadracus and identified a recent sex chromosome turnover in the identity of the sex chromosome and sex determination gene. Utilizing linked-read sequencing data, I found two polymorphic inversions on the X and Y chromosomes across different populations. In Chapter 3, I developed a novel pipeline to assemble the Y chromosome of blackspotted stickleback (*G. wheatlandi*) using long-read and whole-genome resequencing data. The evolutionary strata were defined at a high resolution, and several chromosomal inversions were identified between the X and Y chromosomes of *G. wheatlandi*. By comparing Y assemblies between *G. aculeatus* and *G. wheatlandi*, I discovered that *G. wheatlandi* exhibits a faster rate of gene loss and higher levels of deleterious mutation accumulation, even in the homologous region on the Y chromosome. My findings in these two chapters highlight the role of evolutionary forces, such as drift and sexually-antagonistic selection, in driving sex chromosome evolution and turnover.

Through an examination of the observed patterns of chromosomal rearrangements in stickleback species, my thesis work has expanded our understanding of how large structural variation contributes to genomic evolution and adaptation.

General Introduction

The genetic changes that underlie adaptive evolution

Adaptive evolution plays a pivotal role in ensuring the survival and prosperity of all living organisms, as it facilitates the emergence of mechanisms that enhance their ability to cope with environmental pressures, such as the acquisition of disease resistance and improved foraging capabilities. Gaining insights into the underlying mechanisms that propel adaptive evolution is of utmost importance for comprehending the remarkable diversity of life on our planet and elucidating how organisms have evolved traits that confer fitness advantages, enabling them to flourish in diverse ecological niches (Barrett and Hoekstra 2011; Blount et al. 2018). Genetic changes play a crucial role in adaptive evolution by providing the raw material for natural selection to act upon (Charlesworth et al. 2017).

In the past decades, most studies of genetic variation have focused on single nucleotide polymorphisms (SNPs), which can arise due to various mechanisms, including errors during DNA replication, exposure to mutagens, or genetic recombination, which might impact phenotypes (Brown 2002). SNPs frequently serve as genetic markers to track evolutionary relationships due to their prevalence and easy detectability across the genome (Morin et al. 2004). Despite the extensive scrutiny directed towards SNPs in genetic research, another crucial form of genetic variation, chromosomal rearrangements, has frequently been neglected. This is partly due to the fact that identifying and characterizing chromosomal rearrangements can be more challenging (Wellenreuther et al. 2019). Traditional cytogenetic techniques, such as karyotyping, are capable of detecting major chromosomal rearrangements, but their resolution is limited when it comes to identifying more subtle genomic changes. Next-generation sequencing technologies offer a promising avenue for characterizing chromosomal rearrangements, but their efficacy is constrained by the quality of the reference genome assembly. A continuous and complete genome assembly is essential for studying chromosomal evolution, as short-read sequencing technologies often yield reads that are only a few hundred base pairs in length, impeding the precise reconstruction of large-scale genomic rearrangements. Recent advances in long-read sequencing technology have revolutionized the study of genomic rearrangements, which allow researchers to assemble an entire chromosome, and to detect and analyze chromosomal rearrangements (Amarasinghe

et al. 2020), leading to a renewed interest in understanding their roles in adaptation and speciation.

Types and effects of chromosomal rearrangements

Chromosomal rearrangements, which refer to alterations in the structure or number of chromosomes, play a crucial role in the evolution of genomes (Lander et al. 2001; Jaillon et al. 2004; Putnam et al. 2008). The various types of chromosomal rearrangements, which include insertions, deletions, fusions, inversions, translocations, duplications, as well as whole genome duplication, contribute to the generation of genetic diversity, which is essential for populations to adapt to varying environmental conditions (Rieseberg 2001; Mérot et al. 2020). These changes in chromosome structure can be brought about by several mechanisms, such as DNA recombination, repair, and replication processes, and the activity of transposable elements (Carvalho and Lupski 2016).

Chromosomal rearrangements can have multiple effects on an organism, including direct effects in meiosis, changes in gene regulation, and changes in copy number, potentially resulting in significant evolutionary effects. While chromosomal rearrangements can have various effects, fusions and inversions are particularly important in modifying the genetic landscape, potentially resulting in significant evolutionary effects. In particular, fusions and inversions have a significant impact on recombination patterns, which can lead to reproductive isolation between populations and the evolution of novel traits with adaptive advantages. One of the predicted effects of chromosomal inversion and fusion is their ability to create reproductive isolation between populations (Hou et al. 2014; Fuller et al. 2019). This occurs when chromosomal rearrangements prevent individuals from interbreeding or producing viable offspring due direct meiotic effects and/or to the accumulation of fixed differences in the regions of low or no recombination. Over time, this can lead to the divergence of the two populations and the formation of new species, a process known as speciation. Another predicted effect of chromosomal inversion and fusion is the facilitation of adaptive evolution. The importance of chromosomal inversions and fusions in facilitating adaptative evolution is highlighted by numerous examples across different taxa. For example, chromosomal inversions have been found to be related to adaptive traits in plants (Barb et al. 2014; Aguirre-Liguori et al. 2019), insects (Corbett-Detig and Hartl 2012; Lindtke et al. 2017),

birds (Hooper and Price 2017; Lundberg et al. 2023), fish (Barth et al. 2017; Leitwein et al. 2017) and mammals (Stefansson et al. 2005; Puig et al. 2015).

Likewise, chromosomal rearrangements have influenced the evolution of sex chromosomes in numerous organisms. In many species, sex is determined by the presence of sex chromosomes (Bergero and Charlesworth 2009; Bachtrog et al. 2014; Ponnikas et al. 2018). Over time, these sex chromosomes can undergo rearrangements, leading to the evolution of new sex determination systems. Here, I mainly focused on two types of chromosomal rearrangements, fusions and inversions, as they have the ability to modify the recombination landscape on a substantial scale, potentially resulting in significant and immediate evolutionary effects (Cicconardi et al. 2021).

The effects of chromosomal fusion and inversion in adaptive evolution

Many studies have revealed that the distribution of adaptive alleles, which are genetic variants that provide selective advantages to an organism in specific environments, is not random across the genome (Turner et al. 2005; Nosil et al. 2009; Nadeau et al. 2012; Duranton et al. 2018; Irwin et al. 2018). This non-random distribution suggests that certain regions of the genome may be more prone to accumulate beneficial mutations or be more sensitive to selection pressures than others. Chromosomal fusion and inversion are two types of chromosomal rearrangements that are predicted to have a significant impact on shaping the genomic distribution of adaptive alleles as both can alter the recombination landscape, which is essential for avoiding recombination between adaptive alleles (Rieseberg 2001; Kirkpatrick and Barton 2006; Guerrero and Kirkpatrick 2014).

Chromosomal inversions, the reversal of the orientation of a segment of a chromosome relative to its homologous counterpart, are prevalent in natural populations and were first detected in *Drosophila* (Sturtevant 1921). One of the primary ways that inversions can play an important role in promoting local adaptation is by reducing recombination between loci that are beneficial in specific environments. This is because inversions effectively suppress recombination within the inverted region, which can facilitate the formation of co-adapted gene complexes (Dobzhansky 1947; Dobzhansky 1970) or capture locally adaptive alleles (Kirkpatrick and Barton 2006). Inversions can thus allow for the maintenance of locally adapted alleles by preventing them from being broken up by recombination with other alleles

that may be maladaptive in the same environment (Kirkpatrick and Barton 2006; Kirkpatrick 2010). In addition to promoting local adaptation, inversions can also function as a barrier to gene flow between populations, thereby promoting divergence and potentially driving the evolution of new species (Fuller et al. 2018; Huang and Rieseberg 2020).

There are numerous examples of inversions promoting local adaptation in various organisms (Wellenreuther and Bernatchez 2018). For example, in monkey flowers (*Mimulus guttatus*), a chromosomal inversion has been found to be associated with adaptation to different elevations. The inversion is associated with the regulation of genes involved in the response to abiotic stresses, including drought and cold, suggesting that the inversion is involved in local adaptation to different elevations with different climatic conditions (Twyford and Friedman 2015). Similarly, in the butterfly species *Heliconius numat*a, a chromosomal inversion has been found to be associated with mimicry adaptation. The inversion is associated with the regulation of genes involved in wing coloration in adaptation to mimic the wing patterns of other butterflies in the local environment (Jay et al. 2018). Another example of the role of inversion in adaptation comes from the study of the apple maggot fly, *Rhagoletis pomonella* (Feder et al. 2003). In the northeastern United States, the apple maggot flies feed on hawthorn fruits, while in the Midwest, the fly feeds on apples. The shift in host plants is thought to have been facilitated by the evolution of an inversion. The inversion traps a set of genes associated with apple feeding, reducing gene flow between the two populations and promoting the evolution of host plant specialization.

In the genetic study of adaptation, chromosomal fusions have received relatively scant attention when compared to chromosomal inversions. Despite this, it is worth noting that chromosomal fusion constitutes another type of chromosomal rearrangement capable of facilitating adaptation via alteration of recombination landscapes. Notably, unlike chromosomal inversion, chromosomal fusions can additionally bring together previously unlinked adaptive loci (Guerrero and Kirkpatrick 2014). Furthermore, it should be noted that chromosomal fusion tends to result in a longer chromosome, which typically results in a lower average recombination rate across the chromosome (Roesti et al. 2013; Haenel et al. 2018; Cicconardi et al. 2021). A notable example of the effect of chromosomal fusion can be seen in *Heliconius* butterflies. These butterflies inherited fused chromosomes from a common ancestor, which has resulted in decreased genetic diversity due to a decrease in

recombination and lower effective population size. Moreover, there is increased selection against introgression among diverging populations on the fused chromosomes(Cicconardi et al. 2021; Yoshida et al. 2023).

Despite the fact that chromosomal fusions and inversions have been observed to facilitate adaptive evolution either directly or indirectly, several aspects of these processes remain shrouded in mystery. Although it is widely recognized that such chromosomal rearrangements have the capacity to modify patterns of linkage disequilibrium and recombination rates, as well as generate novel gene combinations or interrupt existing ones, the precise mechanisms by which these effects occur have yet to be fully elucidated. On the one hand, recombination would be reduced or completely ceased when chromosomal fusions and inversions happen, resulting in the easier fixation of newly emerged adaptive alleles (Feder et al. 2012; Via 2012). On the other hand, these chromosomal rearrangements could link or trap previously existing adaptive loci, forming a cluster (Yeaman and Whitlock 2011). Further research is needed to distinguish between whether inversions and fusions capture existing adaptive loci or gain adaptive loci after their formation, and to provide a more comprehensive understanding of the role of chromosomal rearrangements in adaptive evolution.

Diversity and convergence of sex chromosomes

Sex chromosomes play a critical role in determining the sex of some species, but there is a great deal of diversity in the size, shape, and composition of these chromosomes (The Tree of Sex Consortium 2014). There are many types of sex determination systems, the most familiar of which are the genetic sex determination systems, which can include the presence of sex chromosomes. The XX/XY system is a commonly known sex chromosome system found in mammals, including humans, where females possess two X chromosomes and males have one X and one Y chromosome. Other species, such as birds, have evolved the ZW/ZZ system, in which females possess a pair of Z chromosomes, while males have one Z and one W chromosome. In mammals, sex chromosomes are heteromorphic, in which the X and Y differ in size and gene content, and contain many genes involved in sexual development and reproduction (Bull 1983; Hughes et al. 2005; Bellott et al. 2014; Cortez et al. 2014). However, in some species, such as ratite birds (Mank and Ellegren 2007; Vicoso, Kaiser, et al. 2013) , pythons (Vicoso, Emerson, et al. 2013) and turtles (Valenzuela and Adams 2011) the sex

chromosomes are homomorphic and have fewer sex-linked genes. In some cases, the identity of the sex chromosome as well as whether the sex chromosomes are heteromorphic or homomorphic is highly variable among closely-related species, as seen in some species of fish (El Taher et al. 2021) and frogs (Jeffries et al. 2018).

Despite the astonishing diversity of sex chromosomes among species, sex chromosomes are also highly convergent in terms of their evolutionary history and patterns. Similar patterns have been detected in many different organisms, such as mammals, birds, and fish (Bachtrog et al. 2014; Furman et al. 2020) Sex chromosomes typically originate from a pair of homologous autosomes that gradually diverge in both function and structure over time. Sex chromosomes are thought to evolve when a mutation arises on one of the chromosomes, leading it to become a sex-determining locus and initiate the differentiation of sexes. If there is suppression of recombination between the two homologous chromosomes, the sex-determining chromosome accumulates further mutations, resulting in further divergence in structure and function from its homologous counterpart on the other chromosome. These mutations can include changes in gene expression and even the addition or deletion of genes. However, the lack of recombination in the sex chromosomes usually leads to the convergent pattern of degeneration of the sex chromosome, with loss of functional genes and accumulation of repetitive DNA. This process of degeneration can ultimately lead to differences in morphological traits between males and females (Vicoso 2019).

Hence, the diversity and convergence of sex chromosomes constitute two essential characteristics that make them a captivating and intricate system for exploring the following questions: 1) why do sex chromosomes differ between closely-related species? 2) which evolutionary forces drive the evolution of sex chromosomes from autosomes?

The effects of chromosomal fusion and inversion in sex chromosome evolution

Chromosomal fusion and inversion are also critical chromosomal rearrangements that have had a profound impact on the evolution of sex chromosomes. These events lead to a modification of the recombination landscape, which constitutes a fundamental step in the process of sex chromosome evolution.

Chromosomal fusion can result in the formation of a new pair of sex chromosomes by fusing a chromosome containing a sex-determining region with another autosome, thereby

converting it into a neo-sex chromosome. Such fusions have been observed in various organisms, such as fruit flies (Zhou and Bachtrog 2012; Zhou and Bachtrog 2015; Bracewell and Bachtrog 2020) and fish (Kitano and Peichel 2012; Pennell et al. 2015). Asymmetric sexually antagonistic selection between the two sexes is predicted to result in a higher frequency of fusion events between a sex chromosome and an autosome, compared to fusion events between two autosomes (Charlesworth and Charlesworth 1980; Matsumoto and Kitano 2016). The linkage of sexually antagonistic alleles on an autosome to the ancestral sex chromosome is expected to increase the average fitness of both sexes. This is because such a fusion event would ensure that female-beneficial and male-beneficial alleles remain associated with the appropriate respective sex chromosome. Some evidence in support of this hypothesis has been found the jumping spider genus *Habronattus*, which has repeatedly evolved X-autosome fusions that could link male-favored alleles on the neo-Y chromosome (Maddison and Leduc-Robert 2013). Another possible force is meiotic drive. During female meiosis in animals, one of the meiotic products goes into the egg while the others are discarded in the polar bodies. Female meiotic drive can preferentially transmit either fused or unfused chromosomes in different species, resulting in a higher likelihood of X-autosome fusions or opposing their formation, respectively (de Villena and Sapienza 2001a; de Villena and Sapienza 2001b; Yoshida and Kitano 2012).

Chromosomal inversion is another type of chromosomal rearrangement that has played an important role in sex chromosome evolution. Inversions can suppress recombination between the X and Y chromosomes, which is thought to be an important factor in the evolution of sex chromosome differentiation (Charlesworth 2023). The prevention of recombination between the X and Y chromosomes via chromosomal inversion can lead to the accumulation of genetic mutations and structural alterations, which play a significant role in the evolution of sex chromosomes. Consequently, chromosomal inversion is regarded as a principal factor contributing to the formation of regions where different levels of divergence between sex chromosomes have evolved, referred to as evolutionary strata (Lahn and Page 1999; Wang et al. 2012; Jay et al. 2022; Olito et al. 2022; Olito and Abbott 2023). In addition to suppressing recombination, chromosomal inversion can also promote the emergence of novel sex-determining regions. In the case of a stickleback species (*Pungitius pungitius*), a recently evolved XY sex chromosome formed in a large inversion that is associated with hybrid

male sterility between two divergent lineages, which might have facilitated the emergence of a new male-determining gene to overcome the sterility in hybrid populations (Natri et al. 2019).

Despite the frequent observation of chromosomal rearrangements in the evolution of sex chromosomes, the mechanisms by which they occur and their specific roles in the formation and degeneration of sex chromosomes remain unclear. Consequently, further comprehensive studies are required to elucidate the evolutionary processes involved in the formation and degeneration of sex chromosomes. Such studies may help to uncover key insights into the molecular and genetic mechanisms of sex chromosome evolution and contribute to a more thorough understanding of the diversity of sex determination systems across different species.

Sticklebacks: an excellent system to study the effects of chromosomal rearrangements on genome evolution

Sticklebacks are a group of fish species belonging to the family Gasterosteidae, and they are an excellent model to explore the effects of chromosomal rearrangements in adaptive evolution and sex chromosome evolution (Reid et al. 2021). There are around 20 species of sticklebacks, which are found in freshwater and saltwater environments throughout the world (Houde and Zastrow 1993). The most common and well-known species of stickleback is the threepine stickleback (*Gasterosteus aculeatus*), which is found in both freshwater and marine environments in the Northern Hemisphere (Wootton 1976; Bell and Foster 1994; Jones et al. 2012; Reid et al. 2021). *G. aculeatus* have undergone multiple cycles of colonization from marine habitats to freshwater habitats after the retreat of Pleistocene glaciers. This repeated process has led to significant changes in the physical and behavioral characteristics of the species, including body shape, skeletal armor, trophic specializations, pigmentation, salt handling, life history, and mating preferences (Bell and Foster 1994; McKinnon and Rundle 2002; Reid et al. 2021). For example, in freshwater environments, *G. aculeatus* typically have a reduced number of plates and spines compared to their marine counterparts. Previous studies have identified some of the genetic changes that underlie the repeated evolution of reduced plates and spines when sticklebacks colonize freshwater habitats, such as *Eda* (Colosimo et al. 2005; Barrett et al. 2009; Mills et al. 2014; Archambeault et al. 2020) and *Pitx1* (Chan et al. 2010; Thompson et al. 2018; Xie et al. 2019). Thus, the

threespine stickleback is a good model system to study the genetic and genomic changes that underlie repeated adaptation to divergent environments (Peichel and Marques 2017; Reid et al. 2021). Despite the repeated cycles of colonization from marine to freshwater habitats in *G. aceulatus*, not all stickleback species can successfully colonize freshwater environments. This provides a unique opportunity to compare the genomic architecture between stickleback species that can live in freshwater and those that do not.

In particular, sticklebacks provides an opportunity to understand the role of chromosomal fusions and inversions in adaptive evolution. Genomic studies have found three chromosomal inversions that are related to marine-freshwater adaptation in *G. aculeatus* (Jones et al. 2012). Many adaptive quantitative trait loci (QTL) have been located around these inversions (Peichel and Marques 2017). Apart from chromosomal inversions, a change of chromosome numbers among species also indicates the potential possibility of a role for chromosome fusions in adaptation to freshwater habitats. There are 21 pairs of chromosomes (2n=42) in *G. aculeatus*, and 23 pairs (2n=46) in fourspine stickleback (*Apeltes quadracus*), which is only found in marine and brackish water (Fig 1). Cytogenetic studies have shown that the difference in chromosome numbers between the two species came from either chromosomal fusion or fission on Chromosome (Chr) 4 and Chr 7 (Urton et al. 2011). In addition, an enrichment of adaptive QTL has been detected on these two chromosomes (Peichel and Marques 2017). These findings suggest that chromosomal fusion or fission might have potentially facilitated freshwater adaptation in *G. aculeatus*.

Stickleback species also exhibit great potential in the study of sex chromosome evolution since sex chromosomes among species show astonishing diversity (Fig 1) (Jeffries et al. 2022). Sex chromosome turnovers, which involve the swapping of sex chromosomes between species, have been observed multiple times (Ross et al. 2009; Dixon et al. 2019; Jeffries et al. 2022). Moreover, various degrees of differentiation and degeneration can be seen on sex chromosomes across different species (Dixon et al. 2019; Natri et al. 2019; Sardell et al. 2021; Dagilis et al. 2022). In addition, chromosomal rearrangements are quite common on the sex chromosomes of sticklebacks. In Japan Sea stickleback (*Gasterosteus nipponicus*) and blackspotted stickleback (*Gasterosteus wheatlandi*), the shared ancestral Y chromosome has independently fused with an autosome, resulting in a neo-sex chromosome (Kitano et al. 2009; Ross et al. 2009). The Y-autosome fusion in *G. nipponicus* is considered to facilitate the

reproductive isolation between species with fused and unfused chromosomes, showing the importance of sex chromosomes during speciation (Kitano et al. 2009). Apart from chromosomal fusion, chromosomal inversions have been identified as highly correlated with the evolution of strata on the Y chromosome in *G. aculeatus* (Peichel et al. 2020). Because all *Gasterosteus* species have a shared ancestral sex chromosome, Chr19, and the sex chromosomes of the three different species possess an independent evolutionary history after speciation happened, this system provides a rare case to study how sex chromosomes evolved, and how chromosomal fusion and inversion affect the evolution of sex chromosomes (Peichel et al. 2020; Sardell et al. 2021).

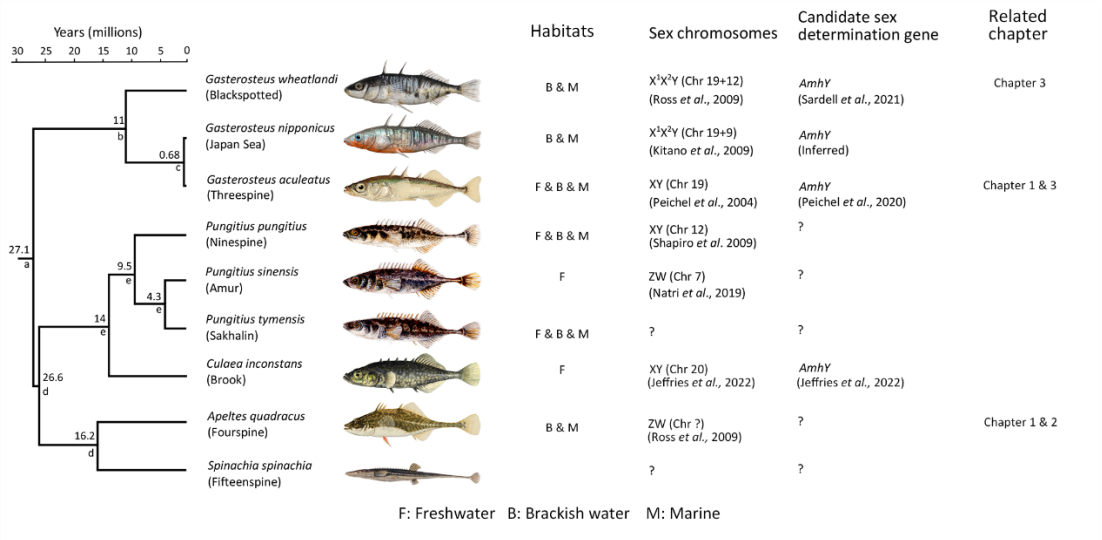


Fig 1. Phylogeny, habitats and sex chromosomes in Gasterosteidae modified from Jeffries *et al*. (2022). "+" indicates a Y- autosome fusion. The tree topology and node dates were obtained from: (a) Varadharajan *et al*. (2019), (b) Betancur-R *et al*. (2015), Friedman *et al*. (2013), Near *et al*. (2013), Sanciangco *et al*. (2016), (c) Ravinet *et al*. (2018), (d) timetree.org, and (e) Guo *et al*. (2019).

Outline of this thesis

In summary, stickleback is an exceptional system for investigating the effects of chromosomal rearrangements on genome evolution and evolutionary processes. With the help of long-read sequencing and whole-genome sequencing, I addressed fundamental

questions in evolutionary biology, including the effect of chromosomal fusion in adaptive evolution, the diversity of sex chromosomes in sticklebacks, and the role of chromosomal inversion in sex chromosome degeneration.

In **Chapter 1**, I assembled the genome of *A. quadracus*. By comparing it with the latest genome assembly of *G. aculeatus*, and the outgroup species, tubesnout (*Aulorhynchus flavidus*), I was able to detect chromosomal fusion events in *G. aculeatus*. Through mapping the non-redundant QTL dataset to each chromosome and conducting population genomic analyses, I found signatures of enrichment of adaptive QTL and genetic differentiation on the fused chromosomes, and provided first evidence that chromosome fusions can facilitate the formation of genetically adaptive clusters. I also inferred the most likely effect of how chromosomal fusion facilitates adaptive evolution.

In **Chapter 2,** I utilized the genome assembly from the previous chapter to identify the sex chromosome in *A. quadracus*. With help of linked-read sequencing of wild samples and pool-sequencing of genetic crosses, I detected a novel sex chromosome in *A. quadracus*, which is completely different from previous predictions. The sex-determining region was located, and it contains two novel candidate sex-determining genes. Most interestingly, chromosome-specific inversions have been identified in different populations, one of which is on the X chromosome. This rare X-specific inversion might include genes under sexually antagonistic selection.

In **Chapter 3**, I developed a novel pipeline, incorporating PacBio and HiC data to completely phase and assembly the X and Y chromosomes of *G. wheatlandi*. With the new assembly, I detected many chromosomal inversions between X and Y chromosomes and re-defined the evolutionary strata in *G. wheatlandi*. By comparing to the Y assembly of *G. wheatlandi* to the existing Y assembly of *G. aculeatus*, I found the relationship between genes lost and the age of each stratum. Also, I investigated patterns of neutral evolution and the accumulation of deleterious mutation in each stratum. I revealed a positive but nonlinear correlation between the age and the level of degeneration of sex chromosomes and inferred potential factors. This study represents the first comparison of two homologous and complete assemblies of Y chromosomes, providing valuable insights into our understanding of sex chromosomes.

References

Aguirre-Liguori JA, Gaut BS, Jaramillo-Correa JP, Tenaillon MI, Montes-Hernández S, García-Oliva F, Hearne SJ, Eguiarte LE. 2019. Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (*Zea mays parviglumis* and *Zea mays mexicana*). *Mol. Ecol.*:mec.15098.

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21:30.

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

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex determination: why so many ways of doing it? *PLoS Biol.* 12:e1001899.

Barb JG, Bowers JE, Renaut S, Rey JI, Knapp SJ, Rieseberg LH, Burke JM. 2014. Chromosomal Evolution and Patterns of Introgression in *Helianthus*. *Genetics* 197:969–979.

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

Barrett RDH, Vines TH, Bystriansky JS, Schulte PM. 2009. Should I stay or should I go? The *Ectodysplasin* locus is associated with behavioural differences in threespine stickleback. *Biol. Lett.* 5:788–791.

Barth JMI, Berg PR, Jonsson PR, Bonanomi S, Corell H, Hemmer-Hansen J, Jakobsen KS, Johannesson K, Jorde PE, Knutsen H, et al. 2017. Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Mol. Ecol.* 26:4452–4466.

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

Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghlul S, Graves T, Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* 508:494–499.

Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* 24:94–102.

Betancur-R R, Ortí G, Pyron RA. 2015. Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes. *Ecol. Lett.* 18:441–450.

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* 362:eaam5979.

Bracewell R, Bachtrog D. 2020. Complex Evolutionary History of the Y Chromosome in Flies of the Drosophila obscura Species Group. *Genome Biol. Evol.* 12:494–505.

Brown TA. 2002. Mutation, repair and recombination. In: Genomes. 2nd edition. Wiley-Liss.

Bull JJ. 1983. Evolution of sex determining mechanisms. The Benjamin/Cummings Publishing Company, Inc.

Carvalho CMB, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17:224–238.

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a *Pitx1* Enhancer. *Science* 327:302–305.

Charlesworth D. 2023. Why and how do Y chromosome stop recombining? *J. Evol. Biol.* 36:632–636.

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

Charlesworth D, Charlesworth B. 1980. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet. Res.* 35:205–214.

Cicconardi F, Lewis JJ, Martin SH, Reed RD, Danko CG, Montgomery SH. 2021. Chromosome Fusion Affects Genetic Diversity and Evolutionary Turnover of Functional Loci but Consistently Depends on Chromosome Size. *Mol. Biol. Evol.* 38:4449–4462.

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

Corbett-Detig RB, Hartl DL. 2012. Population Genomics of Inversion Polymorphisms in Drosophila

melanogaster. *PLoS Genet.* 8:e1003056.
Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grützner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508:488–493.
Dagilis AJ, Sardell JM, Josephson MP, Su Y, Kirkpatrick M, Peichel CL. 2022. Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. *Philos. Trans. R. Soc. B Biol. Sci.* 377:20210205.
Dixon G, Kitano J, Kirkpatrick M. 2019. The Origin of a New Sex Chromosome by Introgression between Two Stickleback Fishes. *Mol. Biol. Evol.* 36:28–38.
Dobzhansky T. 1947. Genetics of natural populations. XIV. A response of certain gene arrangements in the third chromosome of *Drosophila pseudoobscura* to natural selection. Genetics 32(2):142–160.
Dobzhansky T. 1970. Genetics of the evolutionary process. Columbia University Press
Duranton M, Allal F, Fraïsse C, Bierne N, Bonhomme F, Gagnaire P-A. 2018. The origin and remolding of genomic islands of differentiation in the European sea bass. *Nat. Commun.* 9:2518.
El Taher A, Ronco F, Matschiner M, Salzburger W, Böhne A. 2021. Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. *Sci. Adv.* 7:eabe8215.
Feder JL, Gejji R, Yeaman S, Nosil P. 2012. Establishment of new mutations under divergence and genome hitchhiking. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367:461–474.
Feder JL, Roethele JB, Filchak K, Niedbalski J, Romero-Severson J. 2003. Evidence for Inversion Polymorphism Related to Sympatric Host Race Formation in the Apple Maggot Fly, *Rhagoletis pomonella*. *Genetics* 163:939–953.
Friedman M, Keck BP, Dornburg A, Eytan RI, Martin CH, Hulsey CD, Wainwright PC, Near TJ. 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. *Proc. R. Soc. B Biol. Sci.* 280:20131733.
Fuller ZL, Koury SA, Phadnis N, Schaeffer SW. 2019. How chromosomal rearrangements shape adaptation and speciation: Case studies in *Drosophila pseudoobscura* and its sibling species *Drosophila persimilis*. *Mol. Ecol.* 28:1283–1301.
Fuller ZL, Leonard CJ, Young RE, Schaeffer SW, Phadnis N. 2018. Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLOS Genet.* 14:e1007526.
Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank JE. 2020. Sex Chromosome Evolution: So Many Exceptions to the Rules. *Genome Biol. Evol.* 12:750–763.
Guerrero RF, Kirkpatrick M. 2014. Local adaptation and the evolution of chromosome fusions. *Evolution.* 68:2747–2756.
Guo B, Fang B, Shikano T, Momigliano P, Wang C, Kravchenko A, Merilä J. 2019. A phylogenomic perspective on diversity, hybridization and evolutionary affinities in the stickleback genus *Pungitius*. *Mol. Ecol.* 28:4046–4064.
Haenel Q, Laurentino TG, Roesti M, Berner D. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol. Ecol.* 27:2477–2497.
Hooper DM, Price TD. 2017. Chromosomal inversion differences correlate with range overlap in passerine birds. *Nat. Ecol. Evol.* 1:1526–1534.
Hou J, Friedrich A, de Montigny J, Schacherer J. 2014. Chromosomal Rearrangements as a Major Mechanism in the Onset of Reproductive Isolation in *Saccharomyces cerevisiae*. *Curr. Biol.* 24:1153–1159.
Houde ED, Zastrow CE. 1993. Ecosystem-and taxon-specific dynamic and energetics properties of larval fish assemblages. *Bull. Mar. Sci.* 53:290–335.
Huang K, Rieseberg LH. 2020. Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants. *Front. Plant Sci.* 11:296.
Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, Wilson RK, Page DC. 2005. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* 437:100–103.
Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN, Grossen C, Delmore KE, Alcaide M, Irwin JH. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. *Mol. Ecol.* 27:4839–4855.
Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz

C, Bernot A, et al. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* 431:946–957.
Jay P, Tezenas E, Véber A, Giraud T. 2022. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLOS Biol.* 20:e3001698.
Jay P, Whibley A, Frézal L, Rodríguez de Cara MÁ, Nowell RW, Mallet J, Dasmahapatra KK, Joron M. 2018. Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Curr. Biol.* 28:1839-1845.e3.
Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, Borzée A, Barrow LN, Canestrelli D, Crochet P-A, Dufresnes C, et al. 2018. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nat. Commun.* 9:4088.
Jeffries DL, Mee JA, Peichel CL. 2022. Identification of a candidate sex determination gene in *Culaea inconstans* suggests convergent recruitment of an *Amh* duplicate in two lineages of stickleback. *J. Evol. Biol.*:jeb.14034.
Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
Kirkpatrick M. 2010. How and Why Chromosome Inversions Evolve. *PLoS Biol.* 8:e1000501.
Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* 173:419–434.
Kitano J, Peichel CL. 2012. Turnover of sex chromosomes and speciation in fishes. *Environ. Biol. Fishes* 94:549–558.
Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz J, Myers RM, et al. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* 461:1079–1083.
Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
Lahn BT., and Page DC. 1999. Four evolutionary strata on the human X chromosome. Science, 286(5441), 964-967.
Leitwein M, Garza JC, Pearse DE. 2017. Ancestry and adaptive evolution of anadromous, resident, and adfluvial rainbow trout (*Oncorhynchus mykiss*) in the San Francisco bay area: application of adaptive genomic variation to conservation in a highly impacted landscape. *Evol. Appl.* 10:56–67.
Lindtke D, Lucek K, Soria-Carrasco V, Villoutreix R, Farkas TE, Riesch R, Dennis SR, Gompert Z, Nosil P. 2017. Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Mol. Ecol.* 26:6189–6205.
Lundberg M, Mackintosh A, Petri A, Bensch S. 2023. Inversions maintain differences between migratory phenotypes of a songbird. *Nat. Commun.* 14:452.
Maddison WP, Leduc-Robert G. 2013. Multiple origins of sex chromosome fusions correlated with chiasma localization in Habronattus jumping spiders (Araneae: Salticidae). *Evolution* 67:2258–2272.
Mank JE, Ellegren H. 2007. Parallel divergence and degradation of the avian W sex chromosome. *Trends Ecol. Evol.* 22:389–391.
Matsumoto T, Kitano J. 2016. The intricate relationship between sexually antagonistic selection and the evolution of sex chromosome fusions. *J. Theor. Biol.* 404:97–108.
McKinnon JS, Rundle HD. 2002. Speciation in nature: the threespine stickleback model systems. *Trends Ecol. Evol.* 17:480–488.
Mérot C, Oomen RA, Tigano A, Wellenreuther M. 2020. A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol.* 35:561–572.
Mills MG, Greenwood AK, Peichel CL. 2014. Pleiotropic effects of a single gene on skeletal development and sensory system patterning in sticklebacks. *EvoDevo* 5:5.
Morin PA, Luikart G, Wayne RK, the SNP workshop group. 2004. SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19:208–216.
Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, Quail MA, Joron M, ffrench-Constant RH, Blaxter ML, et al. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. B Biol. Sci.*

367:343–353.
Natri HM, Merilä J, Shikano T. 2019. The evolution of sex determination associated with a chromosomal inversion. *Nat. Commun.* 10:145.
Near TJ, Dornburg A, Eytan RI, Keck BP, Smith WL, Kuhn KL, Moore JA, Price SA, Burbrink FT, Friedman M, et al. 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl. Acad. Sci.* 110:12738–12743.
Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18:375–402.
Olito C, Abbott JK. 2023. The evolution of suppressed recombination between sex chromosomes and the lengths of evolutionary strata. *Evolution* 77:1077–1090.
Olito C, Ponnikas S, Hansson B, Abbott JK. 2022. Consequences of partially recessive deleterious genetic variation for the evolution of inversions suppressing recombination between sex chromosomes. *Evolution* 76:1320–1330.
Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 372.
Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM, Kingsley DM, et al. 2020. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* 21:177.
Pennell MW, Kirkpatrick M, Otto SP, Vamosi JC, Peichel CL, Valenzuela N, Kitano J. 2015. Y Fuse? Sex Chromosome Fusions in Fishes and Reptiles. *PLOS Genet.* 11:e1005237.
Ponnikas S, Sigeman H, Abbott JK, Hansson B. 2018. Why Do Sex Chromosomes Stop Recombining? *Trends Genet.* 34:492–503.
Puig M, Casillas S, Villatoro S, Cáceres M. 2015. Human inversions and their functional consequences. *Brief. Funct. Genomics* 14:369–379.
Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.
Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. 2018. The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. *PLOS Genet.* 14:e1007358.
Reid K, Bell MA, Veeramah KR. 2021. Threespine Stickleback: A Model System For Evolutionary Genomics. *Annu. Rev. Genomics Hum. Genet.* 22:357–383.
Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16:351–358.
Roesti M, Moser D, Berner D. 2013. Recombination in the threespine stickleback genome-patterns and consequences. *Mol. Ecol.* 22:3014–3027.
Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). *PLoS Genet.* 5:e1000391.
Sanciangco MD, Carpenter KE, Betancur-R R. 2016. Phylogenetic placement of enigmatic percomorph families (Teleostei: Percomorphaceae). *Mol. Phylogenet. Evol.* 94:565–576.
Sardell JM, Josephson MP, Dalziel AC, Peichel CL, Kirkpatrick M. 2021. Heterogeneous Histories of Recombination Suppression on Stickleback Sex Chromosomes. *Mol. Biol. Evol.* 38:4403–4418.
Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, et al. 2005. A common inversion under selection in Europeans. *Nat. Genet.* 37:129–137.
Sturtevant AH. 1921. A Case of Rearrangement of Genes in Drosophila. *Proc. Natl. Acad. Sci.* 7:235–237.
The Tree of Sex Consortium. 2014. Tree of Sex: A database of sexual systems. *Sci. Data* 1:140015.
Thompson AC, Capellini TD, Guenther CA, Chan YF, Infante CR, Menke DB, Kingsley DM. 2018. A novel enhancer near the Pitx1 gene influences development and evolution of pelvic appendages in vertebrates. *eLife* 7:e38555.
Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic Islands of Speciation in Anopheles gambiae. *PLoS Biol.* 3:e285.
Twyford AD, Friedman J. 2015. Adaptive divergence in the monkey flower Mimulus guttatus is maintained by a chromosomal inversion. *Evolution* 69:1476–1486.
Urton JR, McCann SR, Peichel CL. 2011. Karyotype Differentiation between Two Stickleback Species

(Gasterosteidae). *Cytogenet. Genome Res.* 135:150–159.
Valenzuela N, Adams DC. 2011. Chromosome number and sex determination coevolve in turtles. *Evol. Int. J. Org. Evol.* 65:1808–1813.
Varadharajan S, Rastas P, Löytynoja A, Matschiner M, Calboli FCF, Guo B, Nederbragt AJ, Jakobsen KS, Merilä J. 2019. A high-quality assembly of the nine-spined stickleback (Pungitius pungitius) genome. *Genome Biol. Evol.*:evz240.
Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos. Trans. R. Soc. B Biol. Sci.* 367:451–460.
Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nat. Ecol. Evol.* 3:1632–1641.
Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. *PLoS Biol.* 11:e1001643.
Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl. Acad. Sci.* 110:6453–6458.
de Villena FP-M, Sapienza C. 2001a. Nonrandom segregation during meiosis: the unfairness of females. *Mamm. Genome* 12:331–339.
de Villena FP-M, Sapienza C. 2001b. Female meiosis drives karyotypic evolution in mammals. *Genetics* 159:1179–1189.
Wang J, Na J-K, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W, et al. 2012. Sequencing papaya X and Y ^{h} chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl. Acad. Sci.* 109:13710–13715.
Wellenreuther M, Bernatchez L. 2018. Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends Ecol. Evol.* 33:427–440.
Wellenreuther M, Mérot C, Berdan E, Bernatchez L. 2019. Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Mol. Ecol.* 28:1203–1209.
Wootton RJ. 1976. The Biology of the Sticklebacks. Lodon: Academic Press
Xie KT, Wang G, Thompson AC, Wucherpfennig JI, Reimchen TE, MacColl ADC, Schluter D, Bell MA, Vasquez KM, Kingsley DM. 2019. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* 363:81–84.
Yeaman S, Whitlock MC. 2011. THE GENETIC ARCHITECTURE OF ADAPTATION UNDER MIGRATION-SELECTION BALANCE. *Evolution* 65:1897–1911.
Yoshida K, Kitano J. 2012. THE CONTRIBUTION OF FEMALE MEIOTIC DRIVE TO THE EVOLUTION OF NEO-SEX CHROMOSOMES. *Evolution* 66:3198–3208.
Yoshida, K, Rödelsperger C, Röseler W, Riebesell M, Sun S, Kikuchi T, Sommer RJ. 2023. Chromosome fusions repatterned recombination rate and facilitated reproductive isolation during *Pristionchus* nematode speciation. *Nat Ecol Evol* **7**, 424–439
Zhou Q, Bachtrog D. 2012. Sex-Specific Adaptation Drives Early Sex Chromosome Evolution in *Drosophila*. *Science* 337:341–345.
Zhou Q, Bachtrog D. 2015. Ancestral Chromatin Configuration Constrains Chromatin Evolution on Differentiating Sex Chromosomes in Drosophila. *PLOS Genet.* 11:e1005331.

Chapter 1

Chromosomal fusions facilitate adaptation to divergent environments in threespine stickleback

Zuyao Liu[1], Marius Roesti[1], David Marques[2,3,4], Melanie Hiltbrunner[1], Verena Saladin[1], and Catherine L. Peichel[1,*]

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

[2]Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

[3]Department of Fish Ecology and Evolution, Centre for Ecology, Evolution, and Biogeochemistry, Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Kastanienbaum, Switzerland.

[4]Natural History Museum Basel, Basel, Switzerland.

***Corresponding author:** E-mail: catherine.peichel@iee.unibe.ch

This chapter has been published in Molecular Biology and Evolution:

Liu Z, Roesti M, Marques D, Hiltbrunner M, Saladin V, Peichel CL. 2022. Chromosomal Fusions Facilitate Adaptation to Divergent Environments in Threespine Stickleback. *Mol. Biol. Evol.* 39:msab358. DOI: https://doi.org/10.1093/molbev/msab358

Abstract

Chromosomal fusions are hypothesized to facilitate adaptation to divergent environments, both by bringing together previously unlinked adaptive alleles and by creating regions of low recombination that facilitate the linkage of adaptive alleles. But, there is little empirical evidence to support this hypothesis. Here, we address this knowledge gap by studying threespine stickleback (*Gasterosteus aculeatus*), in which ancestral marine fish have repeatedly adapted to freshwater across the northern hemisphere. By comparing the threespine and ninespine stickleback (*Pungitius pungitius*) genomes to a *de novo* assembly of the fourspine stickleback (*Apeltes quadracus*) and an outgroup species, we find two chromosomal fusion events involving the same chromosomes have occurred independently in the threespine and ninespine stickleback lineages. On the fused chromosomes in threespine stickleback, we find an enrichment of quantitative trait loci (QTL) underlying traits that contribute to marine versus freshwater adaptation. By comparing whole genome sequences of freshwater and marine threespine stickleback populations, we also find an enrichment of regions under divergent selection on these two fused chromosomes. There is elevated genetic diversity within regions under selection in the freshwater population, consistent with a simulation study showing that gene flow can increase diversity in genomic regions associated with local adaptation and our demographic models showing gene flow between the marine and freshwater populations. Integrating our results with previous studies, we propose that these fusions created regions of low recombination that enabled the formation of adaptative clusters, thereby facilitating freshwater adaptation in the face of recurrent gene flow between marine and freshwater threespine sticklebacks.

Keywords

Adaptation; chromosomal fusion; natural selection; genome assembly; threespine stickleback; fourspine stickleback; Gasterosteidae

Introduction

Understanding what facilitates rapid adaptation to new environments is of fundamental interest in evolutionary biology. A key question is whether adaptive loci are linked together in particular regions of the genome (Yeaman 2013; Schwander et al. 2014; Thompson and Jiggins 2014). Theoretical work has predicted that tight physical linkage between adaptive alleles would facilitate adaptation to divergent environments, particularly when there is gene flow, by preventing the production of unfit combinations of phenotypes through recombination (Charlesworth and Charlesworth 1979; Lenormand and Otto 2000; Hoffmann and Rieseberg 2008). In support of these theoretical predictions, empirical work from many systems shows that the distribution of adaptive loci across the genome is not random. For example, population genomic studies in many systems that show divergence despite the presence of gene flow have found that adaptive loci tend to be clustered in the genome, forming highly differentiated regions called "genomic islands" (Turner et al. 2005; Nadeau et al. 2012; Duranton et al. 2018; Irwin et al. 2018). Similarly, genetic linkage mapping studies have revealed evidence for the clustering of quantitative trait loci (QTL) underlying putatively adaptive phenotypes (e.g. Protas et al. 2008; Friedman et al. 2015; Peichel and Marques 2017).

Although these empirical findings support the theoretical predictions, it is still unclear how such QTL clusters and/or genomic islands form. Genomic clusters could evolve because of the higher probability of an adaptive mutation to fix near another locally adapted mutation since such architectures are seldom disrupted by recombination (the divergence hitchhiking hypothesis) (Feder et al. 2012; Via 2012). Genomic clusters could also be formed by genomic rearrangements that bring adaptive loci together (the genomic architecture change hypothesis) (Yeaman and Whitlock 2011). A study incorporating both analytical models and individual-based simulations suggested that genomic clusters are more likely to form through genomic rearrangements that bring together adaptive loci than through the establishment of an adaptive mutation near another locally adapted mutation (Yeaman 2013). Consistent with this finding, empirical studies have often found that such genomic clusters are often associated with chromosomal rearrangements, such as inversions (Kirkpatrick and Barton 2006; Schwander et al. 2014; Thompson and Jiggins 2014; Wellenreuther and Bernatchez 2018). However, there are not many studies focusing on other kinds of chromosomal rearrangements, such as chromosomal fusions.

Unlike chromosome inversions, which can only create clusters by reducing

recombination between loci that are already physically linked, chromosomal fusions have been predicted to facilitate adaption both by bringing together previously unlinked loci and by changing the recombination landscape to create a new region of reduced recombination (Guerrero and Kirkpatrick 2014). Chromosomal fusions (and fissions) are common, as evidenced by the dramatic differences in chromosome number among species. Across multicellular eukaryotes, diploid chromosome number ranges from 2 to 1260 (Sinha et al. 1979; Crosland and Crozier 1986). Chromosome numbers can even vary between closely related species (Wang and Lan 2000; Lysak et al. 2006; Ross et al. 2009; Urton et al. 2011; Valenzuela and Adams 2011) or be polymorphic within species (Dobigny et al. 2017; Wellband et al. 2019). Robertsonian fusions (i.e. fusions between two acrocentric chromosomes at their centromeres) are the most common type of chromosomal rearrangement in plants and animals (Robinson and King 1995). These Robertsonian fusions can have profound impacts on the recombination landscape across the entire genome (Vara et al. 2021). These effects are most obvious on the Robertsonian chromosomes, where recombination is restricted to the distal ends of the chromosome in fusion heterozygotes as well as in fusion homozygotes (Bidau et al. 2001; Castiglia and Capanna 2002; David and Janice 2002; Franchini et al. 2016; Franchini et al. 2020; Vara et al. 2021). More generally, chromosomal fusions create larger chromosomes, which have a lower average recombination rate (Roesti et al. 2013; Haenel et al. 2018; Cicconardi et al. 2021). Despite this clear impact of chromosomal fusions on recombination, there is little empirical evidence supporting the hypothesis that chromosomal fusions play a role in adaptation (but see Kitano et al. 2009; Bidau et al. 2012; Wellband et al. 2019).

In this study, we used stickleback fish species in the family Gasterosteidae to examine whether chromosomal fusions have contributed to the formation of adaptive genomic clusters. This system provides an excellent opportunity to address the role of chromosome fusion in adaptation as closely related stickleback species differ in chromosome number (Fig. 1). In particular, we focused on the fourspine stickleback (*Apeltes quadracus*), which has 23 pairs of chromosomes (2n=46) and is primarily found in marine and brackish habitats, and the threespine stickleback (*Gasterosteus aculeatus*), which has only 21 pairs of chromosomes (2n=42) and can live in freshwater as well as marine and brackish habitats (Chen and Reisman 1970; Wootton 1976; Ross and Peichel 2008; Ross et al. 2009; Fig. 1). Previous studies have shown that the difference in chromosome numbers between *A. quadracus* and *G. aculeatu*s

involves the large metacentric chromosomes 4 and 7 in *G. aculeatus*, which each represent two pairs of acrocentric chromosomes in *A. quadracus* (Urton et al. 2011). However, without data from a closely-related outgroup species, it was impossible to determine whether there had been chromosomal fissions in *A. quadracus* or chromosomal fusions in *G. aculeatus.* However, it was intriguing to note that both chromosomes 4 and 7 have frequently been associated with QTL and genomic islands of divergence between marine and freshwater *G. aculeatus* (Hohenlohe et al. 2010; Jones et al. 2012; Roesti et al. 2014; Peichel and Marques 2017; Nelson and Cresko 2018; Fang et al. 2020; Magalhaes et al. 2021; Roberts Kingman et al. 2021), suggesting the possibility that chromosomal fusions might have facilitated adaptation to divergent habitats in this species. However, previous population genomic studies had not directly tested whether these chromosomes were specifically enriched for genomic clusters of adaptive loci.

Here, we generated a high-quality de novo assembly for *A. quadracus*, and then integrated comparative genomics and population genomics to address the following questions: (1) is the difference in chromosome number between threespine stickleback (*G. aculeatus*) and fourspine stickleback (*A. quadracus*) due to chromosomal fusion in *G. aculeatus* or chromosomal fission in *A. quadracus*?; (2) is there an enrichment of QTL contributing to adaptive divergence in traits on chromosomes 4 and 7 in *G. aculeatus*?; (3) is there an enrichment of molecular signatures of divergent adaptation on chromosomes 4 and 7 in *G. aculeatus*?; and (4) how did chromosomal fusions facilitate adaptation to divergent habitats in *G. aculeatus*?

Results and Discussion

Phylogenetic relationship and chromosome numbers of stickleback species

We generated phylogenetic trees for seven species of the Gasterosteidae family plus the outgroup species (*Aulorhynchus flavidus*) using 1734 single-copy, orthologous coding gene sequences obtained from whole genome sequencing data (*G. aculeatus*, *Pungitius pungitius*, *A. quadracus*, *A. flavidus*) and RNA-seq data (*G. nipponicus*, *G. wheatlandi*, *Culaea inconstans*, *Spinachia spinachia*) (Supplementary Table S1). The phylogeny generated by concatenated sequences is highly supported with all bootstrap values equal to 100 (Fig. 1A). It is consistent with a previous phylogeny generated from 11 nuclear genes and mitochondrial genomes (Kawahara et al. 2009). To account for incomplete lineage sorting, we also built a species tree.

First, gene trees were reconstructed for each ortholog. Then, these trees were combined to find a topology that agrees with the largest number of quartet trees. The species tree is the same as the concatenated tree with high support values (Fig. 1B).

Based on this phylogeny, it is likely that the ancestor of the Gasterosteidae family inhabited marine and brackish water. The brook stickleback (*C. inconstans*) is the only species that lives primarily in freshwater, while the threespine stickleback (*G. aculeatus*) and the ninespine stickleback (*P. pungitius*) are able to inhabit both marine and freshwater habitats, with the opportunity for gene flow between the marine and freshwater populations. Interestingly, these two species also have a diploid chromosome number of 42 (2n=42), which is reduced relative to the diploid chromosome number (2n=46) in the fourspine stickleback (*A. quadracus*), the brook stickleback (*C. inconstans*), and the outgroup *A. flavidus* (Li et al. submitted). We also found that the fifteenspine stickleback (*S. spinachia*) has a lower diploid chromosome number (2n=40) by counting metaphase chromosomes from three independent males (41 metaphases counted, mode 2n=40, range 2n=38-42) and three independent females (9 metaphases counted, mode 2n=40, range 2n=38-41; Supplementary Fig. S1). Given that most teleosts have a diploid chromosome number of 48 or 50 (Naruse et al. 2004; Amores et al. 2014), it is likely that lower chromosome number in species within the stickleback family results from chromosomal fusions. However, it is also possible that the fusions were ancestral and that the greater number of chromosomes in some species results from chromosomal fission. To distinguish between these possibilities, we used the newly available whole-genome assemblies of the outgroup *A. flavidus* (Li et al. submitted), *P. pungitius* (Varadharajan et al. 2019), and *G. aculeatus* (Nath et al. 2021), as well as the high-quality assembly of *A. quadracus* generated in this study. We then focused on the whole-chromosome rearrangements that have occurred in *G. aculeatus* to determine whether these rearrangements are associated with genetic loci that underlie adaptation to divergent marine and freshwater habitats in this species.

De novo assembly and annotation of the *A. quadracus* genome

To generate a high-quality assembly of the *A. quadracus* genome, we used high-coverage PacBio long-read sequencing to assemble the genome of a female fish derived from a laboratory cross between two populations from Nova Scotia, Canada. Raw read coverage was 91.58x (39.2 Gbp in total). 10X Genomics linked reads and HiC reads from the same individual

were used for scaffolding the assembly separately. The final assembly is 428.91 Mbp, and it contains 890 scaffolds, including 21 chromosome-level scaffolds. The N50 length is 18.10 Mbp, and the assembly quality assessed by BUSCO was relatively high with 96.9% completeness. *A. quadracus* has a smaller genome than the other existing stickleback genome assemblies (~449 Mbp for *G. aculeatus* (Nath et al. 2020) and ~521 Mbp for *P. pungitius* (Varadharajan et al. 2019)). We constructed a repeat library for *A. quadracus* using de novo and homology-based approaches (See Materials and Methods). After masking the repetitive regions, the rest of the genome was annotated with the evidence from RNA-seq data, homologous protein databases, and ab initio annotation. We filtered out annotated genes with poor quality (typically AED > 0.5), leading to 21,955 genes in the final version of the annotation. The accession numbers for the *A. quadracus* assembly and annotation are available in Supplementary Table S1.

Independent fusions of the same chromosomes in *G. aculeatus* and *P. pungitius*

The difference in chromosome number between *G. aculeatus* (2n=42) and *A. quadracus* (2n=46) found in previous cytogenetic studies could either result from fission events in *A. quadracus* or fusion events in *G. aculeatus* (Ross et al. 2009; Urton et al. 2011). By comparing the genome assemblies of *G. aculeatus* and *A. quadracus*, as well as *P. pungitius*, to the outgroup species (*A. flavidus*), we conclude that two fusions occurred in *G. aculeatus* (Fig. 2). The synteny map reveals that chromosomes 4 and 7 in *G. aculeatus* are likely the result of end-to-end fusions between chromosomes 4 and 22, and 7 and 23, respectively in *A. quadracus* (Supplementary Figs. S2-S4). These four chromosomes are also unfused in the outgroup *A. flavidus*, which also has 23 chromosome pairs. Zooming into the detailed synteny map, we also find evidence for inversion and gene transposition between *A. quadracus* and *G. aculeatus* (Supplementary Figs. S2-4). On *G. aculeatus* chromosome 4, two large inversions have occurred near the fusion point. In contrast, the inversions on *G. aculeatus* chromosome 7 have occurred towards the chromosome ends. However, based on the order of the genes in the outgroup, these inversions have likely occurred in *A. quadracus*, not *G. aculeatus*.

Interestingly, chromosome 4 in *P. pungitius* is also the result of a fusion between *A. quadracus* chromosomes 4 and 22. However, taking the phylogeny (Fig. 1) as well as a closer analysis of the fusion breakpoints into account (Supplementary Fig. S3), the fusion events involving *A. quadracus* chromosomes 4 and 22 in both *G. aculeatus* and *P. pungitius* are likely to have occurred independently. Further, chromosome 12 in *P. pungitius*, which is the sex

chromosome (Shapiro et al. 2009; Rastas et al. 2016; Natri et al. 2019) is the result of a fusion between *A. quadracus* chromosomes 7 and 12 (Fig. 2). Although *A. quadracus* chromosome 7 is involved in fusion events in both *G. aculeatus* and *P. pungitius*, it has fused to different chromosomes in these species (Fig. 2 and Supplementary Fig. S4), again suggesting independent fusions have occurred in the two lineages. Together, these data demonstrate that chromosomal fusions have occurred in the two stickleback lineages that include species (*G. aculeatus* and *P. pungitius*) able to inhabit both marine and freshwater habitats, raising the possibility that such fusions have contributed to the ability of these species to adapt to divergent habitats in the face of gene flow.

Enrichment of marine-freshwater QTL on chromosomes 4 and 7 in *G. aculeatus*

If fusions facilitate adaptation by linking adaptive alleles, we would predict that an increased number of QTL underlying adaptive traits would map to the fused chromosomes, and that the these QTL would have congruent effects in the expected direction (i.e. a marine allele confers a marine phenotype and vice versa) on multiple traits. Thus, we tested whether there was an enrichment of QTL with effects in the expected direction on *G. aculeatus* chromosomes 4 and 7 using a database of QTL identified in crosses between marine and freshwater populations (Peichel and Marques 2017). Indeed, we found that chromosomes 4 and 7, as well as chromosomes 16, 20, and 21, have significantly more QTL with effects in the expected direction than other chromosomes, accounting for variation in either the length of chromosomes or the number of genes on the chromosomes (Fig. 3 and Supplementary Table S2). Chromosome 21 has an inversion that is polymorphic within *G. aculeatus*, which is one of the strongest signals of divergence between worldwide marine and freshwater populations (Jones et al. 2012; Roesti et al. 2015; Fang et al. 2020; Magalhaes et al. 2021; Roberts Kingman et al. 2021). Although there are no apparent large-scale chromosomal rearrangements between marine and freshwater populations associated with chromosomes 16 or 20, the adaptive clusters on chromosomes 4, 7 and 21 are associated with chromosomal rearrangements that might facilitate linkage of adaptive traits.

No enrichment of gene transpositions or gene duplications on chromosomes 4 and 7

It has also been proposed that such adaptive clusters could form via small-scale genomic rearrangements, such as transposition of single genes and/or gene duplications (Yeaman

2013). We therefore examined the distribution of gene duplication and gene transposition events in *G. aculeatus* relative to *P. pungitius*, *A. quadracus*, and *A. flavidus*. There were too few gene transposition events to determine whether the distribution of these genes varied among chromosomes. There are more gene duplications than expected on chromosomes 10, 11, 16 and 21, given either the length of the chromosome or the number of genes on the chromosome (Supplementary Table S3). A comparison of the *G. aculeatus* and *A. flavidus* genomes also revealed no evidence for an enrichment of micro-rearrangements, lineage-specific genes, or gene duplications on *G. aculeatus* chromosomes 4 or 7, although gene duplications are enriched specifically within one region on chromosome 4 (Li et al. submitted). It is therefore possible that gene duplication might also play a role in the formation of the QTL clusters on chromosomes 16 and 21, but not on the fusion chromosomes 4 and 7.

Enrichment of genomic signatures of selection on chromosomes 4 and 7 in *G. aculeatus*

The clustering of adaptive QTL on chromosomes 4 and 7 suggests that these chromosome fusions link adaptive loci together. However, from the QTL analysis, we can only observe this at the phenotypic level. To further explore whether chromosome fusions show signatures of selection at the sequence level, we examined different signatures of selection using whole genome sequencing data. Using Hidden Markov Models (HMM), we identified genomic islands of differentiation between a marine (Puget Sound) and freshwater (Lake Washington) population of *G. aculeatus*. The distribution of genomic islands is uneven across the genome, and chromosomes 4, 7, 9, 11, and 20 have a significantly higher number of windows with outlier SNPs in genomic islands than expected, given either the length of the chromosome or the number of genes on the chromosome (For details of all enrichment analyses in this section, see Methods, Supplementary Fig. S5, and Supplementary Table S4). Next, we used a window-based method to calculate F_{ST} across the genome. Fst within genomic islands is elevated, and peaks are enriched on chromosomes 4 and 7 (Fig. 4 and Supplementary Fig. S5). For these two chromosomes, regions with elevated F_{ST} are found in the middle of the chromosomes. A similar pattern is also revealed by a topology weighting analysis (Supplementary Fig. S6), in which regions in the middle of chromosomes 4 and 7 show a higher proportion of topology 1, indicating adaptation of freshwater populations.

We also calculated window-based nucleotide diversity (Pi) across the genome to trace the signature that selection left within each population. Overall, the nucleotide diversity of

the Lake Washington freshwater population is higher than in the Puget Sound marine population, with delta Pi ($Pi_{Lake\ Washington} - Pi_{Puget\ Sound}$) always greater than 0. The greatest differences in nucleotide diversity between the populations are found on chromosomes 1, 4, 7, 20 and 21, with more diversity in the freshwater Lake Washington population (Fig. 4 and Supplementary Fig. S5). Within Lake Washington, there are more top 5% outlier windows for Pi than expected on chromosomes 4 and 7 (as well as on chromosomes 8, 20 and 21), particularly in the middle of the chromosomes (Fig. 4 and Supplementary Fig S5 and Supplementary Table S4). Interestingly, genetic diversity in the regions under selection is lower in the Puget Sound marine population and elevated in the Lake Washington freshwater population (Fig. 4 and Supplementary Fig. S5).

The nucleotide diversity results are surprising. Most current-day freshwater populations of *G. aculeatus,* such as the Washington Lake population, were founded by marine stickleback after the end of the last ice age, approximately 12,000 years ago (Bell and Foster 1994). Thus, selection towards a novel environment is mainly thought to occur in the freshwater environment, leading to a reduction in genetic diversity near selected sites. Furthermore, freshwater populations are expected to have a smaller population size, where genetic drift would have a more powerful influence, leading to a faster loss of genetic diversity in the freshwater population. However, a recent simulation study has pointed out that gene flow can not only homogenize the genome but also increase diversity near regions under selection (Jasper and Yeaman 2020). To determine whether gene flow can explain the distribution of nucleotide diversity in our data, we built several demographic models (Supplementary Fig. S7) to explore the most plausible evolutionary history of the Puget Sound marine and Lake a h n ton re h ater pop at on a e on C a e the e t o e ha a ott enec event in the ancestral population, followed by two reciprocal migration regimes (Fig. 5 and Supplementary Table S5). The effective population size in Puget Sound is 33,111, which is larger than the effective population size of 3,775 in Lake Washington, consistent with the expectation that the marine population has a larger population size. The inferred bottleneck is consistent with a previous Pairwise Sequentially Markovian Coalescent (PSMC) inference of the demographic histories of these two populations (Shanfelter et al. 2019). Two migration regimes are inferred with an increase in migration at 111 years ago, which is roughly consistent with when the Lake Washington Ship Canal, which connects Lake Washington and Puget Sound, was built in 1917 (Edmondson 1991). During both periods of migration, the

actual number of migrants from Puget Sound to Lake Washington is lower than the reverse, suggesting that more fish migrate from the freshwater environment to the marine environment. Overall, our demographic model suggests that migration between marine and freshwater populations is common, especially after the build-up of the Lake Washington Ship Canal. This is consistent with a scenario of gene flow increasing diversity near regions under selection (Jasper and Yeaman 2020) and our result that regions with high genetic diversity are associated with regions under selection. Similar results have been observed in Alaskan populations of *G. aculeatus*, with low genetic diversity in marine populations and high genetic diversity in freshwater populations in regions of the genome under divergent selection (Nelson et al. 2019). Their simulations suggest that this pattern results from asymmetries in population structure between the habitats, especially near locally adapted sites, and that this effect on diversity is strongest in regions of low recombination, such as we find on chromosomes 4 and 7.

Lastly, we used two haplotype-based methods to detect footprints of recent or ongoing selection. iHS is a statistic for detecting incomplete selective sweeps across the genome within a population (Voight et al. 2006), while XPEHH is a statistic for detecting (nearly) complete selective sweeps in one of two populations (Sabeti et al. 2007). We calculated the proportion of extreme values (w-iHS and w-XPEHH) in 20kb windows with a step size of 10kb. Signatures of recent selection exist across the whole genome in both populations, with more windows containing signatures of divergent selection (XPEHH) than expected between the populations on chromosomes 5, 9 and 17 (Fig. 4, Supplementary Fig. S5 and Supplementary Table S4). Chromosomes 8 and 10 exhibit more windows of elevated iHS in Lake Washington, and chromosomes 4, 17, 18 and 21 exhibit more windows of elevated iHS in Puget Sound (Supplementary Fig. S5 and Supplementary Table S4). Thus, these patterns of recent selection differ from the patterns nucleotide diversity and F_{ST}, particularly on chromosomes 4 and 7 (Fig. 4 and Supplementary Fig. S5), consistent with previous results suggesting that most regions of strong divergence between marine and freshwater ecotypes are on the order of millions of years old (Nelson and Cresko 2018; Roberts Kingman et al. 2021).

How might chromosomal fusions facilitate the formation of adaptive clusters?

Overall, we find that signatures of divergent selection between marine and freshwater are distributed across the *G. aculeatus* genome, but that some regions of the genome show

evidence for clustering of adaptive loci. The patterns we find in our population genomic analyses using whole genome sequencing of a single marine-freshwater pair from the Eastern Pacific are consistent with the results of many population genomic studies, mostly using RAD-seq, in global marine-freshwater pairs (Hohenlohe et al. 2010; Jones et al. 2012; Roesti et al. 2014; Peichel and Marques 2017; Haenel et al. 2018; Nelson and Cresko 2018; Fang et al. 2020; Magalhaes et al. 2021; Roberts Kingman et al. 2021). In contrast to previous studies, we explicitly tested whether particular chromosomes are enriched for different signatures of selection. We found that chromosomes 4 and 7 have significantly more QTL associated with traits that diverge between marine and freshwater populations, more outlier SNPs in genomic islands of divergence, and higher levels of diversity in freshwater. By contrast these chromosomes do not have an excess of gene transposition or duplication events, or signatures of recent selection. These strong signals on chromosomes 4 and 7 have been previously observed, and they have been attributed to the fact that these are regions of low recombination (Roesti et al. 2014; Nelson et al. 2019; Roberts Kingman et al. 2021). Indeed, using genetic diversity as a proxy for recombination rate (Cicconardi et al. 2021), we find that chromosomes 4 and 7 have lower recombination rates than the unfused chromosomes in the *G. aculeatus* genome and that recombination rates on these chromosomes are lower than on their unfused homologues in *A. quadracus* (Supplementary Fig. S8). Interestingly, the patterns on these two chromosomes are different than those on chromosome 1, which is also a large metacentric chromosome with similar patterns of reduced recombination across the middle of the chromosome (Roesti et al. 2013; Glazer et al. 2015; Shanfelter et al. 2019). This suggests that the reduction of recombination observed on chromosomes 4 and 7 is greater than we would predict for metacentric chromosomes of similar size. Furthermore, chromosome 1 does not show chromosome-wide enrichment for any signatures of selection or for QTL (Supplementary Fig. S5 and Supplementary Table S2 and Supplementary Table S4). Thus, we hypothesize that the clustering of adaptive loci on chromosomes 4 and 7 is associated with the reduced recombination created by the chromosomal fusions.

There are two non-mutually exclusive hypotheses for how chromosomal fusions might facilitate adaptation (Guerrero and Kirkpatrick 2014). The first is that the fusion brings together pre-existing locally adapted alleles. The second is that the fusion creates a region of low recombination, which then enables the formation of adaptive clusters, as has been seen in the case of a chromosomal inversion in *Mimulus guttatus* (Coughlan and Willis 2019). In

the case of the fusions found in *G. aculeatus*, it is difficult to determine whether one of these explanations may be most important, or whether both are playing a role. This is because the two sister species of *G. aculeatus* (*G. wheatlandi* and *G. nipponicus*) also have 21 pairs of chromosomes (Fig. 1), and our preliminary assembly of a *G. wheatlandi* genome suggests that chromosomes 4 and 7 show the same arrangement as in *G. aculeatus*. Thus, the fusions were likely present in the common ancestor of the three *Gasterosteus* species. However, both *G. wheatlandi* and *G. nipponicus* can only live in marine or brackish habitats (Fig. 1). Thus, the presence of the fusion itself was not enough to enable adaptation to freshwater. Previous work has suggested that duplications of the *Fads2* gene occurred in *G. aculeatus*, but not in *G. wheatlandi* or *G. nipponicus*, and that these duplications enabled *G. aculeatus* to take advantage of nutritionally depauperate freshwater habitats (Ishikawa et al. 2019). Interestingly, there is also a duplication of *Fads2* in *P. pungitius*, which can also live in freshwater. We speculate that once *G. aculeatus* (and perhaps *P. pungitius*) was able to invade freshwater, the region of low recombination created by the fusions provided a genomic region that could allow the buildup of adaptive alleles that were resistant to gene flow between marine and freshwater populations. Nonetheless, it is possible that the fusions we find in these species were fixed due to selection for linkage between alleles that provided an advantage in the ancestral habitat. A role for selection is suggested by convergent involvement of the same chromosomes in fusions in *Gasterosteus* and *Pungitius*. However, with our current data, we are unable to determine whether selection, drift, and/or another force like meiotic drive was responsible for the fixation of chromosomal fusions in sticklebacks (Dobigny et al. 2017).

Regardless of the mechanism of initial fixation, once fixed, we hypothesize that these fusions provided a unique genomic substrate for the formation of adaptive clusters in *G. aculeatus* as it was moving between marine and freshwater habitats during repeated bouts of glaciation and deglaciation during its evolutionary history over the past several million years. It does not appear that new genes were moving into these regions (Li et al. submitted), and therefore they must have been built by what has been called “allele-only clustering”, which is when selection builds clusters of locally adapted alleles at loci already co-localized in the genome (Roesti 2018). The patterns of divergence we see indeed suggest that multiple adaptive clusters are embedded in the larger regions of particularly low recombination across chromosomes 4 and 7 (Fig. 4 and Supplementary Fig. S8). As many of these adaptive clusters

in *G. aculeatus* (including those on chromosome 4 and 7) are at least a million years old (Nelson and Cresko 2018; Roberts Kingman et al. 2021) , there has been much time for the buildup of these adaptive alleles. Interestingly, older adaptive regions seem to be larger, suggesting that adaptive alleles are accumulating in these regions over time (Roberts Kingman et al. 2021). The accumulation of many adaptive alleles within these adaptive clusters is also consistent with a detailed study of the *Eda* region on chromosome 4, which showed evidence that multiple mutations within a 16kb region of high divergence between marine and freshwater populations contribute to lateral plate and sensory lateral line phenotypes, and that linked mutations outside the *Eda* region are responsible for the QTL cluster observed on chromosome 4 (Archambeault et al. 2020). Taken together, these data are more consistent with the divergence hitchhiking hypothesis (Feder et al. 2012; Via 2012) than the genomic architecture change hypothesis (Yeaman 2013). Our data suggest that even if the fusions themselves were not initially selected to link adaptive alleles, they have provided a genomic substrate that facilitates the process of divergence hitchhiking.

Conclusion

While the role of chromosomal rearrangements, such as inversions, in adaptation have been well-studied, the contribution of chromosomal fusions to adaptation is still unclear. By comparing genome assemblies, we found that two chromosomal fusions have occurred in *G. aculeatus*, and further demonstrate that these fused chromosomes are enriched in adaptive QTL and signatures of selection between marine and freshwater populations. We propose that these chromosomal fusions facilitated adaptation by altering the recombination landscape to create regions of low recombination that enabled the formation of adaptive clusters that can persist in the face of gene flow.

Materials and methods

Ethics statement

All experiments involving animals were approved by the Veterinary Service of the Department of Agriculture and Nature of the Canton of Bern (VTHa# BE4/16, BE17/17 and BE127/17).

Sample collections

In 2017, *A. quadracus* were collected from Rainbow Haven Beach (44.654857, -63.42113) and Canal Lake (44.498298, -63.90205) in Nova Scotia, Canada by Anne Dalziel. In 2018, *G. wheatlandi* were collected from Rainbow Haven Beach (44.654857, -63.42113) in Nova Scotia, Canada by Anne Dalziel. In 2017, *C. inconstans* were collected from the Sass River (60.073328, -113.312240) in the Northwest Territories, Canada by Julia Wucherpfennig; brains were dissected by Ian Heller and placed into RNAlater (Life Technologies, Carlsbad, California, USA). In 2018, *S. spinachia* were collected from the Baltic Sea (54.387423, 10.494736) near Hohenfelde, Germany by Arne Nolte. All samples were shipped to the University of Bern for further processing.

DNA and RNA extraction and sequencing

For assembly of the *A. quadracus* genome, DNA from a single laboratory-reared female resulting from a cross between a Rainbow Haven Beach female and a Canal Lake male (both from Nova Scotia, Canada) was used. High molecular weight DNA was extracted from the blood following (Peichel et al. 2020) and used to prepare a SMRTbell Express library for PacBio sequencing and a 10X Genomics library for Linked-Reads sequencing. The liver of the same individual was used to prepare a Hi-C sequencing library using the Phase Genomics Proximo Hi-C animal kit (Phase Genomics, Seattle, Washington, USA). Four SMRT cells were sequenced on a PacBio Sequel Platform, and the 10X Genomics and Hi-C libraries were sequenced for 300 cycles on an Illumina NovaSeq SP flow cell. To polish the PacBio reads, DNA from wild-caught individuals from Canal Lake (4 females, 4 males) was extracted using phenol-chloroform and used to prepare Illumina DNA TruSeq libraries, which were sequenced for 300 cycles on an Illumina NovaSeq SP flow cell. All library preparation and sequencing were performed by the University of Bern Next Generation Sequencing Platform.

Total RNA was extracted from whole brains of wild-caught adult *G. wheatlandi* (4 females, 4 males), *C. inconstans* (5 females, 5 males), *A. quadracus* from Canal Lake (4 females, 4 males), and *S. spinachia* (4 females and 4 males) using Trizol (Life Technologies, Carlsbad, California, USA) following the manufacturer's instructions. Illumina mRNA TruSeq libraries were prepared and either subject to 150bp paired-end sequencing on an Illumina HiSeq3000 (*G. wheatlandi*, *C. inconstans*, *A. quadracus*) or 150bp paired-end sequencing on an Illumina NovaSeq SP flow cell (*S. spinachia*) at the University of Bern Next Generation Sequencing Platform.

For this study, we also used the available genome assemblies for *G. aculeatus* (Nath et al. 2021)), *P. pungitius* (Varadharajan et al. 2019), and the outgroup *A. flavidus* (Li et al. submitted). We also used available RNA-seq data from *G. nipponicus* (Ishikawa et al. 2019). Supplementary Table S1 summarizes all samples and sequencing data used for this study and provides all relevant accession numbers.

Reconstruction of the stickleback phylogeny

To determine if the phylogenetic relationships among the species in the Gasterosteidae family are consistent with previous studies using 11 nuclear genes and mitochondrial genomes (Kawahara et al. 2009), we built a phylogenetic tree using seven species in the family (*A. quadracus*, *C. inconstans*, *G. aculeatus*, *G. nipponicus*, *G. wheatlandi*, *P. pungitius*, *S. spinachia*) and an outgroup *A. flavidus*. For species with a reference genome (*A. quadracus*, *G. aculeatus*, *P. pungitius*, and *A. flavidus*), nucleotide and amino acid sequences of the coding regions were extracted. For species without a reference genome, we used RNA-seq data to build transcriptome assemblies.

RNA-seq reads were trimmed using Trimmomatic (v 0.36), and the reads were de novo assembled by the Trinity assembler (v 2.10.0). The open reading frames (ORF) were predicted by Transdecoder (accessed on 02/10/2020) (Haas et al. 2013). Redundancy at the amino acid level was removed by cd-hit (v 4.8.1) (Li and Godzik 2006) with a threshold of 95% identity. Next, amino acid sequences of the eight species were compared to search for orthologs by OrthoFinder (v 2.3.12) (Emms and Kelly 2019), and only single-copy orthologs were kept for the downstream analysis. Then, we aligned amino acid sequences using muscle (v 3.8.1511) to guide the alignment of the corresponding nucleotides sequences. Sites with gaps or missing data were removed entirely, resulting in 1734 alignments of single-copy orthologs. Phylogenies were built in two ways: 1) we concatenated alignments of 1734 orthologs to build a supermatrix and reconstructed a phylogeny using RaxML (v8) (Stamatakis 2006); 2) for each alignment, we first built gene trees in RaxML (v8) and then estimated the species tree using ASTRAL-III (V 5.7.4) (Zhang et al. 2018).

Identification of chromosome number in *S. spinachia*

For the phylogenies shown in Fig. 1, we also added information on the known habitats of each species (Wootton 1976; Guo et al. 2019) and the diploid chromosome number (Chen and

Reisman 1970; Ocalewicz et al. 2008; Ross and Peichel 2008; Kitano et al. 2009; Ross et al. 2009; Ocalewicz et al. 2011). However, there was no prior information on the diploid chromosome number for *S. spinachia*. We therefore used the protocol of Ross and Peichel (2008) to generate metaphase spreads from 3 of the *S. spinachia* females and 3 of the *S. spinachia* males used for the RNA-sequencing data (Supplementary Table S1). Sex was determined by inspection of the gonads. The fish were euthanized in 0.2% tricaine methanesulfonate (MS-222), and the spleen was used for the metaphase spreads. Metaphase spreads from each individual were stained with DAPI and photographed on a Nikon Eclipse 80i microscope using a Photometrics CoolSNAP ES2 camera (Photometrics, USA) and NIS-Elements BR 3.22.15 imaging software (Nikon, Japan). Chromosomes were counted from photos of individual metaphase spreads.

A. quadracus de novo genome assembly

The PacBio assembly was generated using Flye 2.6 with default parameters (Kolmogorov et al. 2019), followed by the polishing step using Arrow (v 3.0) and Pilon (Walker et al. 2014) separately with default parameters in both cases. For polishing, whole-genome resequencing data described above from eight *A. quadracus* individuals (four males, four females) from Canal Lake, Nova Scotia, Canada (Supplementary Table S1) were used. Raw reads were trimmed by Trimmomatic (v 0.36) (Bolger et al. 2014) with a sliding window of 4 bp. The first 13 bp of reads were dropped, and windows of the remaining reads were also dropped with an average quality score below 15. Genome size estimation was run by GenomeScope 2.0 (Ranallo-Benavidez et al. 2020) with trimmed data.

Contig scaffolding was conducted using the 10x Genomics linked reads and Hi-C proximity guided assembly separately. Contigs were linked by linked reads using ARCS (v 1.1.1) and LINKS (Warren et al. 2015; Yeo et al. 2018). Raw Hi-C reads were first processed with HiCUP (Wingett et al. 2015) and then assembled by Juicer (v. 1.5) (Durand et al. 2016) and 3D-DNA (v. 180922) (Dudchenko et al. 2017). After the first round of Hi-C scaffolding, the assembly was revised manually based on the contact map and then scaffolded again. The final step, gap-closing, was run by LR_Gapcloser (Xu et al. 2019). Assembly quality was evaluated by BUSCO v3 (Simão et al. 2015; Waterhouse et al. 2018).

A. quadracus genome annotation

The genome assembly was annotated in a two-step pipeline. The first step was the annotation of repeat elements. MITE-Tracker (Crescente et al. 2018) was used to detect miniature inverted-repeat transposable elements (MITE). Full-length long terminal repeat (LTR) sequences were identified using LTR_finder (Xu and Wang 2007) and LTRharvest (Ellinghaus et al. 2008), and were further combined by LTR_retriever (Ou and Jiang 2018). Subsequently, RepeatMolder (v. 2.0) (Flynn et al. 2020) was used to identify novel repeat sequences. Libraries from MITE, LTR, and RepeatMolder were merged into a non-redundant library and passed to the final annotation of repetitive sequences with RepeatMasker (v. 4.0.9) (Smit et al. 2013).

The RNA-sequencing data generated from eight *A. quadracus* individuals (four males, four females) from Canal Lake, Nova Scotia, Canada (Supplementary Table S1) and described above was used to aid in genome annotation. The raw reads were trimmed by Trimmomatic (v. 0.36) and then used as the input for Trinity assembler with default parameters (v. 2.10.0) (Grabherr et al. 2011).

The prediction and annotation of genes were conducted on the repeat-masked genome assembly with the Maker2 (v. 2.31.10) pipeline (Holt and Yandell 2011), including four rounds of annotation. In the first round, the transcriptome assembly generated by Trinity and protein data from *Danio rerio*, *G. aculeatus*, *P. pungitius*, *Takifugu flavidus,* and the Uniprot database (UniProt Consortium 2015) were used as evidence for the program. The second round of annotation included two training and prediction steps by AUGUSTUS (v. 3.2.3) (Stanke et al. 2008) and SNAP (Korf 2004). The results were then passed to MAKER2. For the third round annotation, GeneMARK-ES (Ter-Hovhannisyan et al. 2008) was combined with MAKER2. Finally, the second round annotation was repeated with the resulting files from the third round. The final annotation was checked based on annotation edit distance (AED), and only annotations with AED score 0.5 or less were retained for downstream analysis. Functional annotation was conducted by eggnog-mapper (v2) (Huerta-Cepas et al. 2017).

Genomic synteny analyses and detection of rearrangements between species

Synteny analyses were conducted in two ways. First, Mummer4 and nucmer (Marçais et al. 2018) were used to compare the order of genes between *G. aculeatus* and *A. quadracus* on *G. aculeatus* chromosomes 4 and 7. Alignments shorter than 2000bp with an identity less than 85% were removed. Second, non-redundant coding sequence sets from four species (*G.*

aculeatus, *A. quadracus*, *P. pungitius* and *A. flavidus*) were used for cross synteny analysis. We used MCScan (Tang et al. 2008) in JCVI package (Tang et al. 2015) to compare synteny on the chromosome level as well as the gene level. *A. flavidus* was chosen as the outgroup based on the phylogeny to examine whether the reduction of chromosome number in *G. aculeatus* and *P. pungitius* relative to *A. quadracus* is due to fission or fusion.

Identification of gene transposition and duplication events

To detect gene duplication and transposition events, we first extracted single-copy orthologues from four species (*G. aculeatus*, *P. pungitius*, *A. quadracus*, *A. flavidus*) using OrthoFinder (v 2.3.12) (Emms and Kelly 2019). For gene duplication events, we used the duplication summary from OrthoFinder and focused on genes only duplicated in *G. aculeatus*; we included both intra- and inter-chromosomal duplications in the analyses. For gene transposition events, we focused on inter-chromosomal gene transpositions, in which a gene had moved to the focal chromosome in *G. aculeatus* from another chromosome in the other species. The homology of chromosomes from different species is based on our synteny map (Fig. 2). If a gene is only present on a focal chromosome in *G. aculeatus* but is not present on the homologous chromosomes in other species, we considered it as a valid transposition event. The sex chromosome was excluded from these analyses.

To test whether any chromosomes had an excess of duplicated genes, the expected distribution of duplicated genes on each chromosome was calculated based on both the chromosome length in base pairs and the number of genes on the chromosome. The expected and observed distributions were compared in R through a goodness-of-fit test (chisq.test). Chromosomes with significantly higher values than expected were identified by standardized residuals with a value larger than 3 in both comparisons (Supplementary Table S3). There were too few gene transposition events to analyze.

Genomic distribution of marine-freshwater QTL in *G. aculeatus*

To test if the fusion events in *G. aculeatus* are associated with clustering of adaptive traits, we used a modified version of a QTL database (Peichel and Marques 2017). The QTL data were filtered to remove redundant QTL following Rennison and Peichel (in review), and only the 655 QTL found in crosses between marine and freshwater populations were retained for the downstream analysis (Supplementary Table S2). We first mapped all the retained QTL with

confidence intervals to the *G. aculeatus* v.5 genome (Nath et al. 2021) in 50kb windows, following Peichel and Marques (2017). Next, we used the data from the original QTL papers to determine whether the marine allele at these QTL confers a marine phenotype and vice versa, which would suggest that these QTL contribute to adaptation to the divergent marine and freshwater habitats. A chi-square test following (Peichel and Marques 2017) was used to test if the number of QTL with effects in the expected direction on a given chromosome is significantly different from the expected number of QTL with effects in the expected direction on that chromosome, given either the length of the chromosome or the number of genes on the chromosome. To identify significant deviations from the expectation on a particular chromosome, the standardized residuals for each chromosome were examined, with a value of 3 indicating the observed data is significantly larger than expected and a value of -3 indicated the observed data is significantly lower than expected (Supplementary Table S2).

Identifying genomic islands of differentiation

Previous population genomic studies of marine-freshwater divergence were either based on very low coverage (2-5X) whole genome sequence or RAD-seq data (Hohenlohe et al. 2010; Jones et al. 2012; Roesti et al. 2014; Nelson and Cresko 2018; Fang et al. 2020; Magalhaes et al. 2021; Roberts Kingman et al. 2021). To identify genomic islands of differentiation and signatures of selection between *G. aculeatus* marine and freshwater fish, we therefore used the only high-coverage (17-22X), whole-genome sequencing data available at the time of our analyses, which was from 25 freshwater individuals from Lake Washington and 24 marine individuals from Puget Sound (Supplementary Table S1; Shanfelter et al. 2019). Trimmed reads (methods described as above) were mapped to the *G. aculeatus* v.5 genome assembly (Nath et al. 2021) by BWA (v 0.7.11) (Li 2013). Bam files were sorted with duplicates marked by Samtools (v 1.9) (Li et al. 2009) and MarkDuplicates in GATK4 (Van der Auwera and O'Connor 2020) separately. Variants were called using HaplotypeCaller, and joint genotyping was conducted by combining all individuals for the population with GATK4 (Van and O'Connor 2020). For SNP filtration, we used Vcftools (0.1.16) and kept sites with minimum genotype qualities greater than 30, fewer than 20% missing genotypes, and a minor allele frequency greater than 0.05. To prevent bias caused by too high or too low sequencing depth, we also filtered out sites if the population mean depth coverages were less than half or greater than twice the average value for each population. Finally, sites that were not in Hardy-Weinburg

equilibrium in each population were removed.

Using this dataset, we followed the approach of (Hofer et al. 2012; Marques et al. 2016) to identify genomic islands of differentiation between the Puget Sound marine and Lake Washington freshwater populations of *G. aculeatus*. A Hidden Markov model (HMM) was used to find regions with exceptionally low and high divergence compared to the background divergence (assumed to be neutral). Only SNPs with minor allele frequencies > 0.25 were used for this analysis because low-frequency allele SNPs tend to disrupt the detection of high differentiation regions which will never reach a high level of differentiation (Roesti et al. 2012). Locus level F_{ST} was estimated in Arlequin (v 3.5.2.2) (Excoffier and Lischer 2010), and outliers were identified assuming an infinite island model. An HMM method was run to model every chromosome separately based on the probability of an SNP being an outlier from the Fst analysis. Scripts can be found at https://github.com/marqueda/HMM-detection-of-genomic-islands (Marques et al. 2016). Only regions passing the multiple-testing correction with a false discovery rate of 0.001 were recognized as "genomic islands". We excluded chromosome 19, which is the *G. aculeatus* sex chromosome (Peichel et al. 2004) from the analysis.

Detecting signatures of selection across the genome

Scans for signatures of selection were performed between the Puget Sound marine and Lake Washington freshwater populations in various ways using the dataset described above. A window-based F_{ST} distribution and nucleotide diversity were calculated with Vcftools (v 0.1.16) with a window size of 20kb and a window step of 10kb. To further identify selected regions, we also adopted haplotype-based statistics. We first extracted mapped reads with mapping quality larger than 20 and inferred haplotypes using WhatsHap (v1.0) (Martin et al. 2016) and shapeit4 (v 4.1.3) (Delaneau et al. 2019) with default parameters. Then, the output file was imported into the R package rehh (Gautier et al. 2017) to detect soft and incomplete sweeps within populations (iHS) and to detect complete sweeps that occurred in one population and not the other (XPEHH). We followed (Voight et al. 2006) to calculate the proportion of extreme iHS and XPEHH values (w-iHS and w-XPEHH, the proportion of |iHS| and |XPEHH| > 2) in the same 20kb overlapping windows. The sex chromosome, chromosome 19 , was also excluded from this analysis.

To examine whether particular chromosomes were enriched for these signatures of selection, we compared the observed number of: 1) SNPs within genomic islands; 2) top 5%

Pi outliers within each population; 3) top 5% |iHS| regions of outliers within each population; and 4) top 5% XPEHH regions of outliers on each chromosome to the expected numbers, given either the length of the chromosome or the number of genes on the chromosome in R through a goodness-of-fit test (chisq.test). Chromosomes with significantly higher values than expected were identified by standardized residuals with a value larger than 3 in both comparisons (Supplementary Table S4).

Topology weighting analyses

To explore the evolutionary histories of marine and freshwater alleles on the fusion chromosomes, we used a topology weighting approach. We built phylogenetic trees with the SNP dataset for the genome scan in non-overlapping windows for every 50 SNPs by RaxML (v8) (Stamatakis 2006) and conducted tree weighting in Twisst (Martin and Van Belleghem 2017). The analysis was performed on the two fused chromosomes, chromosomes 4 and 7, separately. For comparison, we performed the analysis on chromosome 1 because it is a large submetacentric chromosome with a similar length and recombination patterns as on chromosomes 4 and 7 (Urton et al. 2011; Roesti et al. 2013; Glazer et al. 2015; Shanfelter et al. 2019). However, it has not experienced inter-chromosomal fusion between the *G. aculeatus* and *A. quadracus* lineages.

Inferring demographic history

The SNP dataset used for demographic simulations was the same as the one for detecting genomic islands with two differences. First, all rare alleles (i.e. a minor allele frequency less than 0.05) were kept. Second, we removed sites located in the genomic islands of differentiation. To account for linkage disequilibrium (LD), we used PLINK (v 1.9) to calculate and prune the SNP matrix to those with LD < 0.1. To prevent bias from SNPs in repeated regions, we checked the distance between consecutive SNPs and discarded those where the distance was less than five base pairs.

To explore the evolutionary history of these two *G. aculeatus* populations and explain the patterns of genomic diversity, we reconstructed their demographic history with fastsimcoal2 (v 2.6) (Excoffier et al. 2013). The one-dimensional folded observed site frequency spectrum (SFS) was calculated with easySFS (https://github.com/isaacovercast/easySFS) for each population. To maximize the number of segregating sites, 22 and 18 individuals of Lake

Washington and Puget Sound were kept for downstream analyses respectively. We fixed the split time of freshwater and marine population to 12,000 years ago, assuming a generation time of 1 year (Bell and Foster 1994). Thirteen models were built to identify the best scenario (Supplementary Fig. S7): 1) constant population size; 2) two bottlenecks while splitting; 3) two bottlenecks after splitting; 4) one bottleneck before splitting; 5) one bottleneck and splitting; 6) one bottleneck and splitting followed by a constant and reciprocal migration; 7) one bottleneck and splitting followed by an early reciprocal migration; 8) one bottleneck and splitting followed by a recent reciprocal migration; 9) one bottleneck and splitting followed by two reciprocal migration regimes; 10) one bottleneck and splitting followed by introgression from Lake Washington to Puget Sound; 11) one bottleneck and splitting followed by introgression from Puget Sound to Lake Washington; 12) one bottleneck and splitting followed by introgression from Lake Washington to Puget Sound and two reciprocal migration regimes; 13) one bottleneck and splitting followed by introgression from Puget Sound to Lake Washington and two reciprocal migration regimes. To maximize the likelihood of each model, we randomly started from 100 parameter combinations in 50 Expectation-Conditional Maximization (*ECM*) *cycles* with a total of 200,000 coalescent simulations. A mutation rate of 7.9×10^{-9} was used, following (Guo et al. 2013). For each model, we obtained the best likelihood values and estimated parameters from 100 optimizations. The best model was selected based on the a e t C pp e entar a e

Genetic diversity analysis of each chromsome in fused and unfused taxa

To explore whether fused chromosomes have a lower recombination rate, we compared genetic diversity of each chromsome in *G. aculeatus* and *A. quadracus*. Genetic diversity can be used as a proxy for recombination rate because a decrease in recombination rate should lead to an increase in levels of background selection and therefore decrease in genetic diversity. Such a relationship between genetic diversity and recombination rate has been observed in *Heliconius* butterflies (Cicconardi et al. 2021). To obtain diversity data in *A. quadracus*, the whole-genome resequencing data described above from eight individuals from Canal Lake, Nova Scotia, Canada (Supplementary Table S1) were mapped by BWA (v 0.7.11) (Li 2013) to the *A. quadracus* reference genome generated in this study. Bam files were sorted with duplicates marked by Samtools (v 1.9) (Li et al. 2009) and MarkDuplicates in GATK4 (Van der Auwera and O'Connor 2020) separately. Variants were called using

HaplotypeCaller, and joint genotyping was conducted by combining all individuals with GATK4 (Van and O'Connor 2020). For SNP filtration, we used Vcftools (0.1.16) and kept sites with minimum genotype qualities greater than 30, fewer than 20% missing genotypes, and a minor allele count greater than 2. For *G. aculeatus*, the same SNP dataset for identifying genomic islands was used, except that we only used data from the marine population (Puget Sound) to prevent potential bias due to linkage to adaptive sites in the freshwater population. For both species, we extracted four-fold degenerate sites with the script codingSiteTypes.py available at (https://github.com/simonhmartin/genomics_general). Genetic diversity was calculated in windows of 50 SNPs with the script popgenWindows.py (https://github.com/simonhmartin/genomics_general). The average value of each chromosome was calculated by hand, and genetic diversity on each chromosomes was normalized relative to the average diversity of unfused chromosomes within a species.

Acknowledgments

This work was supported by grants from the Swiss National Science Foundation (grant number 31003A_176130 to C.L.P.) and the US National Institutes of Health (grant number R01 GM116853 to C.L.P.). We thank Sam Yeaman for discussions and for sharing the *A. flavidus* genome assembly before publication, Anne Dalziel, Julia Wucherpfennig, Ian Heller, David Kingsley, and Arne Nolte for providing stickleback samples, the University of Bern Next Generation Sequencing Platform for sequencing support, and the members of the Peichel lab for discussions.

Data availability

All data used in this study were already publicly available or are available at the NCBI Sequence Read Archive under project number PRJNA746773. The *A. quadracus* genome annotations are available on Dryad: doi:10.5061/dryad.wh70rxwpf. All accession numbers are listed in Supplementary Table S1.

References

Amores A, Catchen J, Nanda I, Warren W, Walter R, Schartl M, Postlethwait JH. 2014. A RAD-Tag genetic map for the platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish. *Genetics* 197:625–641.

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

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

Bidau CJ, Giménez MD, Palmer CL, Searle JB. 2001. The effects of Robertsonian fusions on chiasma frequency and distribution in the house mouse (*Mus musculus domesticus*) from a hybrid zone in northern Scotland. *Heredity* 87:305–313.

Bidau CJ, Miño CI, Castillo ER, Martí DA. 2012. Effects of abiotic factors on the geographic distribution of body size variation and chromosomal polymorphisms in two neotropical grasshopper species (*Dichroplus* : Melanoplinae: Acrididae). *Psyche J. Entomol.* 2012:1–11.

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

Castiglia R, Capanna E. 2002. Chiasma repatterning across a chromosomal hybrid zone between chromosomal races of Mus musculus domesticus. Genetica 114:35–40.

Charlesworth D, Charlesworth B. 1979. Selection on recombination in clines. Genetics 91:581–589.

Chen T-R., Reisman HM. 1970. A comparative chromosome study of the North American species of sticklebacks (Teleostei: Gasterosteidae). Cytogenet. Genome Res. 9:321–332.

Cicconardi F, Lewis JJ, Martin SH, Reed RD, Danko CG, Montgomery SH. 2021. Chromosome fusion affects genetic diversity and evolutionary turnover of functional loci but consistently depends on chromosome size. *Mol. Biol. Evol.* 38:4449-4462.

Coughlan JM, Willis JH. 2019. Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. *Mol. Ecol.* 28:1343-1357.

Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC Bioinformatics* 19:348.

Crosland MWJ, Crozier RH. 1986. *Myrmecia pilosula*, an ant with only one pair of chromosomes. *Science* 231:1278–1278.

David D, Janice B-D. 2002. Chromosomal rearrangements and evolution of recombination: comparison of chiasma distribution patterns in standard and Robertsonian populations of the house mouse. *Genetics* 162:1355–1366.

Delaneau O, Zagury J-F, Robinson MR, Marchini JL, Dermitzakis ET. 2019. Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* 10:5436.

Dobigny G, Britton-Davidian J, Robinson TJ. 2017. Chromosomal polymorphism in mammals: an evolutionary perspective. *Biol. Rev.* 92:1–21.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92–95.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3:95–98.

Duranton M, Allal F, Fraïsse C, Bierne N, Bonhomme F, Gagnaire P-A. 2018. The origin and remolding of genomic islands of differentiation in the European sea bass. *Nat. Commun.* 9:2518.

Edmondson WT. 1991. The uses of ecology: Lake Washington and beyond. Seattle: University of Washington Press.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9:e1003905.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population

genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10:564–567.
Fang B, Kemppainen P, Momigliano P, Feng X, Merilä J. 2020. On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat. Ecol. Evol.* 4:1105–1115.
Feder JL, Gejji R, Yeaman S, Nosil P. 2012. Establishment of new mutations under divergence and genome hitchhiking. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367:461–474.
Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* 117:9451–9457.
Franchini P, Colangelo P, Meyer A, Fruciano C. 2016. Chromosomal rearrangements, phenotypic variation and modularity: a case study from a contact zone between house mouse Robertsonian races in Central Italy. *Ecol. Evol.* 6:1353–1362.
Franchini P, Kautt AF, Nater A, Antonini G, Castiglia R, Meyer A, Solano E. 2020. Reconstructing the evolutionary history of chromosomal races on islands: a genome-wide analysis of natural house mouse populations. *Mol. Biol. Evol.* 37:2825–2837.
Friedman J, Twyford AD, Willis JH, Blackman BK. 2015. The extent and genetic basis of phenotypic divergence in life history traits in *Mimulus guttatus*. *Mol. Ecol.* 24:111–122.
Gautier M, Klassmann A, Vitalis R. 2017. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* 17:78–90.
Glazer AM, Killingbeck EE, Mitros T, Rokhsar DS, Miller CT. 2015. Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3 (Bethesda)* 5:1463–1472.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–652.
Guerrero RF, Kirkpatrick M. 2014. Local adaptation and the evolution of chromosome fusions. *Evolution.* 68:2747–2756.
Guo B, Chain FJ, Bornberg-Bauer E, Leder EH, Merilä J. 2013. Genomic divergence between nine- and three-spined sticklebacks. *BMC Genomics* 14:756.
Guo B, Fang B, Shikano T, Momigliano P, Wang C, Kravchenko A, Merilä J. 2019. A phylogenomic perspective on diversity, hybridization and evolutionary affinities in the stickleback genus *Pungitius*. *Mol. Ecol.* 28:4046–4064.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–1512.
Haenel Q, Laurentino TG, Roesti M, Berner D. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol. Ecol.* 27:2477–2497.
Hofer T, Foll M, Excoffier L. 2012. Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics* 13:107.
Hoffmann AA, Rieseberg LH. 2008. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annu. Rev. Ecol. Evol. Syst.* 39:21–42.
Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet.* 6:e1000862.
Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34:2115–2122.
Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN, Grossen C, Delmore KE, Alcaide M, Irwin JH. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. *Mol. Ecol.* 27:4839–4855.
Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A, et al. 2019. A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* 364:886–889.
Jasper RJ, Yeaman S. 2020. Local adaptation can cause both peaks and troughs in nucleotide diversity

within populations. http://biorxiv.org/lookup/doi/10.1101/2020.06.03.132662
Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz J, Myers RM, et al. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* 461:1079–1083.
Koch EL, Morales HE, Larsson J, Westram AM, Faria R, Lemmon AR, Lemmon EM, Johannesson K, Butlin RK. 2021. Genetic variation for adaptive traits is associated with polymorphic inversions in *Littorina saxatilis*. *Evol. Lett.* 5:196–213.
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37:540–546.
Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
Lenormand T, Otto SP. 2000. The Evolution of Recombination in a Heterogeneous Environment. *Genetics* 156:423–438.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 25:2078–2079.
Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.
Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci.* 103:5224–5229.
Magalhaes IS, Whiting JR, D'Agostino D, Hohenlohe PA, Mahmud M, Bell MA, Skúlason S, MacColl ADC. 2021. Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nat. Ecol. Evol.* 5:251–261.
Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* 14:e1005944.
Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLOS Genet.* 12:e1005887.
Martin M, Patterson M, Garg S, O Fischer S, Pisanti N, Klau GW, Schöenhuth A, Marschall T. 2016. WhatsHap: fast and accurate read-based phasing. http://biorxiv.org/lookup/doi/10.1101/085050
Martin SH, Van Belleghem SM. 2017. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* 206:429–438.
Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, Quail MA, Joron M, ffrench-Constant RH, Blaxter ML, et al. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. B Biol. Sci.* 367:343–353.
Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani. 2004. A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. *Genome Res.* 14:820–828.
Nath S, Shaw DE, White MA. 2021. Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 (Bethesda)* 11:jkab007.
Nelson TC, Crandall JG, Ituarte CM, Catchen JM, Cresko WA. 2019. Selection, linkage, and population structure interact to shape genetic variation among threespine stickleback genomes. *Genetics* 212:1367–1382.
Nelson TC, Cresko WA. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.* 2:9–21.
Ocalewicz K, Fopp-Bayat D, Woznicki P, Jankun M. 2008. Heteromorphic sex chromosomes in the ninespine stickleback Pungitius pungitius. *J. Fish Biol.* 73:456–462.
Ocalewicz K, Woznicki P, Furgala-Selezniow G, Jankun M. 2011. Chromosomal location of Ag/CMA $_3$ -NORs, 5S rDNA and telomeric repeats in two stickleback species. *Ital. J. Zool.* 78:12–19.
Ou S, Jiang N. 2018. LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176:1410–1422.
Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J, Myers RM, Mori S, Schluter D,

Kingsley DM. 2004. The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr. Biol.* 14:1416-1424.
Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Phil Trans R Soc B*. 372(1713):20150486.
Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM, Kingsley DM, et al. 2020. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* 21:177.
Protas M, Tabansky I, Conrad M, Gross JB, Vidal O, Tabin CJ, Borowsky R. 2008. Multi-trait evolution in a cave fish, *Astyanax mexicanus. Evol. Dev.* 10:196–209.
Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11:1432.
Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, et al. 2021. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv.* 7:eabg5285.
Robinson TJ, King M. 1995. Species evolution: the role of chromosome change. *Syst. Biol.* 44:578.
Roesti M. 2018. Varied genomic responses to maladaptive gene flow and their evidence. *Genes* 9:298.
Roesti M, Gavrilets S, Hendry AP, Salzburger W, Berner D. 2014. The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* 23:3944–3956.
Roesti M, Kueng B, Moser D, Berner D. 2015. The genomics of ecological vicariance in threespine stickleback fish. *Nat. Commun.* 6:8767.
Roesti M, Moser D, Berner D. 2013. Recombination in the threespine stickleback genome-patterns and consequences. *Mol. Ecol.* 22:3014–3027.
Roesti M, Salzburger W, Berner D. 2012. Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evol. Biol.* 12:94.
Ross JA, Peichel CL. 2008. Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. *Genetics* 179:2173–2182.
Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genet.* 5:e1000391.
Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.
Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr. Biol.* 24:R288–R294.
Shanfelter AF, Archambeault SL, White MA. 2019. Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biol. Evol.* 11:1552–1572.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
Sinha BMB, Srivastava DP, Jha J. 1979. Occurrence of various cytotypes of *Ophioglossum Reticulatum* L. in a population from N. E. India. *Caryologia* 32:135–146.
Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0. Available from: <http://www.repeatmasker.org>.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 24:637–644.
Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. *Science* 320:486–488.
Tang H, Krishnakumar V, Li J. 2015. jcvi: JCVI utility libraries. Zenodo Available from: https://zenodo.org/record/31631
Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18:1979–1990.
Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity* 113:1–8.
Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic Islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3:e285.

UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43:D204-212.
Urton JR, McCann SR, Peichel CL. 2011. Karyotype differentiation between two stickleback species (Gasterosteidae). *Cytogenet. Genome Res.* 135:150–159.
Valenzuela N, Adams DC. 2011. Chromosome number and sex determination coevolve in turtles. *Evolution.* 65:1808–1813.
Van der Auwera G, O'Connor B. 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media.
Vara C, Paytuví-Gallart A, Cuartero Y, Álvarez-González L, Marín-Gual L, Garcia F, Florit-Sabater B, Capilla L, Sanchéz-Guillén RA, Sarrate Z, et al. 2021. The impact of chromosomal fusions on 3D genome folding and recombination in the germ line. *Nat. Commun.* 12:2981.
Varadharajan S, Rastas P, Löytynoja A, Matschiner M, Calboli FCF, Guo B, Nederbragt AJ, Jakobsen KS, Merilä J. 2019. A high-quality assembly of the nine-spined stickleback (*Pungitius pungitius*) genome. *Genome Biol. Evol* 11:3291-3308.
Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos. Trans. R. Soc. B Biol. Sci.* 367:451–460.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963.
Wang W, Lan H. 2000. Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. *Mol. Biol. Evol.* 17:1326–1333.
Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJM, Birol I. 2015. LINKS: scalable, alignment-free scaffolding of draft genomes with long reads. *GigaScience* 4:35.
Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35:543–548.
Wellband K, Mérot C, Linnansaari T, Elliott JAK, Curry RA, Bernatchez L. 2019. Chromosomal fusion and life history-associated genomic variation contribute to within-river local adaptation of Atlantic salmon. *Mol. Ecol.* 28:1439–1459.
Wellenreuther M, Bernatchez L. 2018. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* 33:427–440.
Wingett SW, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* 4:1310.
Wootton RJ. 1976. The biology of the sticklebacks. London: Academic Press
Xu G-C, Xu T-J, Zhu R, Zhang Y, Li S-Q, Wang H-W, Li J-T. 2019. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience* 8(1), giy157
Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35:W265–W268.
Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci.* 110:E1743–E1751.
Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* 65:1897–1911.
Yeo S, Coombe L, Warren RL, Chu J, Birol I. 2018. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics* 34:725–731.
Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153.

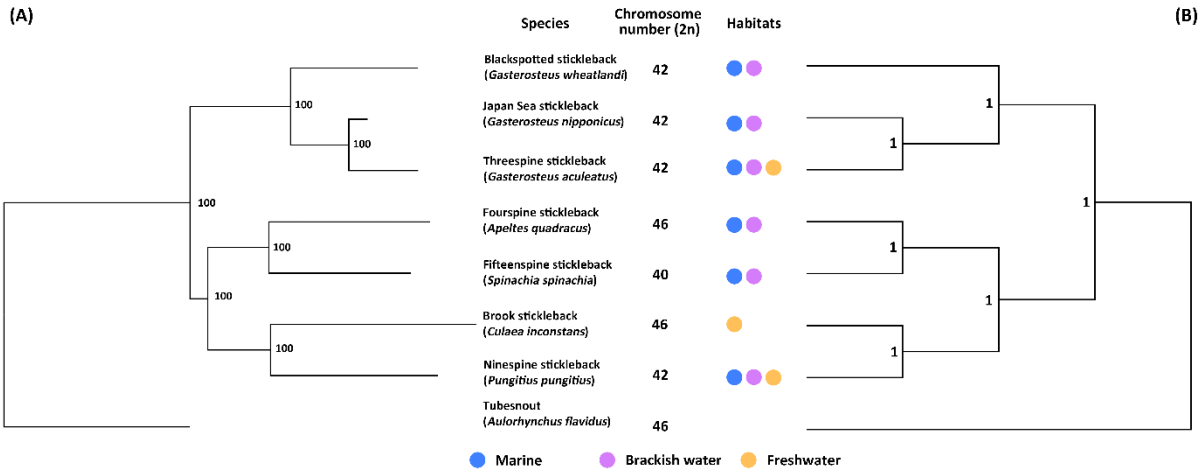


Fig. 1. Phylogeny of stickleback species and the *A. flavidus* outgroup. (A) Phylogenetic relationship among species was reconstructed in RaxML using a concatenated supermatrix of 1734 single-copy, orthologous genes. Numbers near nodes are bootstrap values. (B) Species tree was reconstructed in ASTRAL-III based on individual gene trees. Numbers near nodes are support values from ASTRAL-III. Data on diploid chromosome number are from (Chen and Reisman 1970; Ocalewicz et al. 2008; Ross and Peichel 2008; Kitano et al. 2009; Ross et al. 2009; Ocalewicz et al. 2011) and this study for *S. spinachia*, and data on habitats are from (Wootton 1976; Guo et al. 2019).

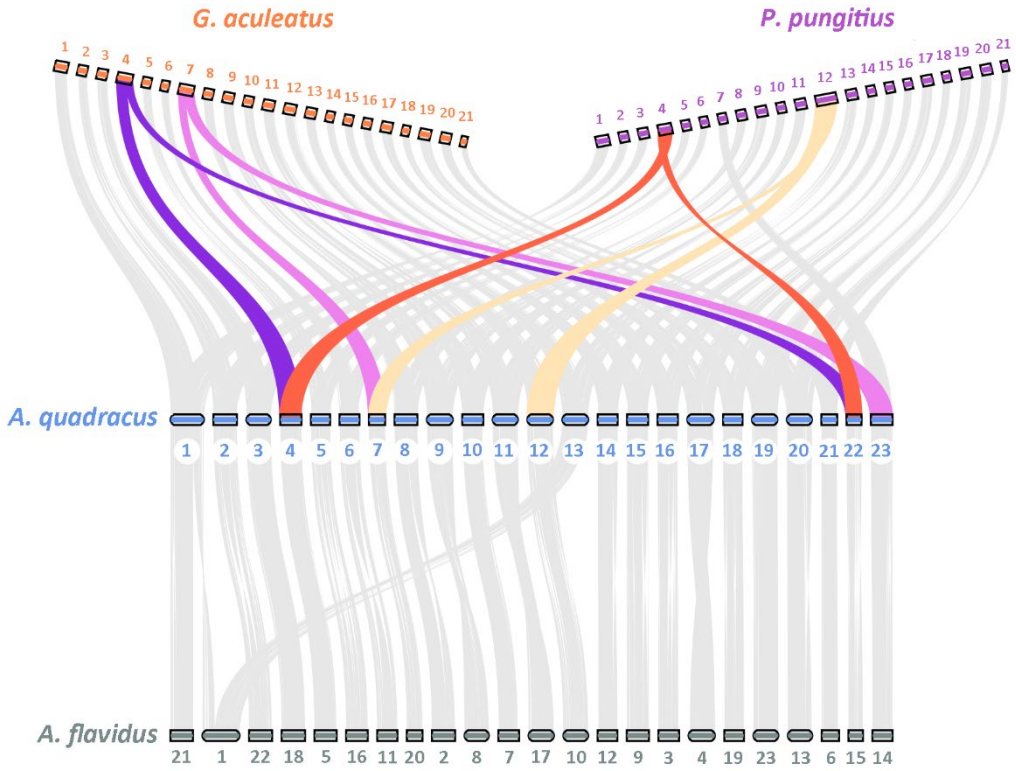


Fig. 2. Synteny map of the *A. flavidus*, *A. quadracus*, *G. aculeatus* and *P. pungitius* genomes. The comparison is based on homologous coding region sequences. Colored rectangles are chromosomes and numbers indicate the corresponding chromosomes. Colored lines represent the fusion events in *G. aculeatus* and *P. pungitius*.

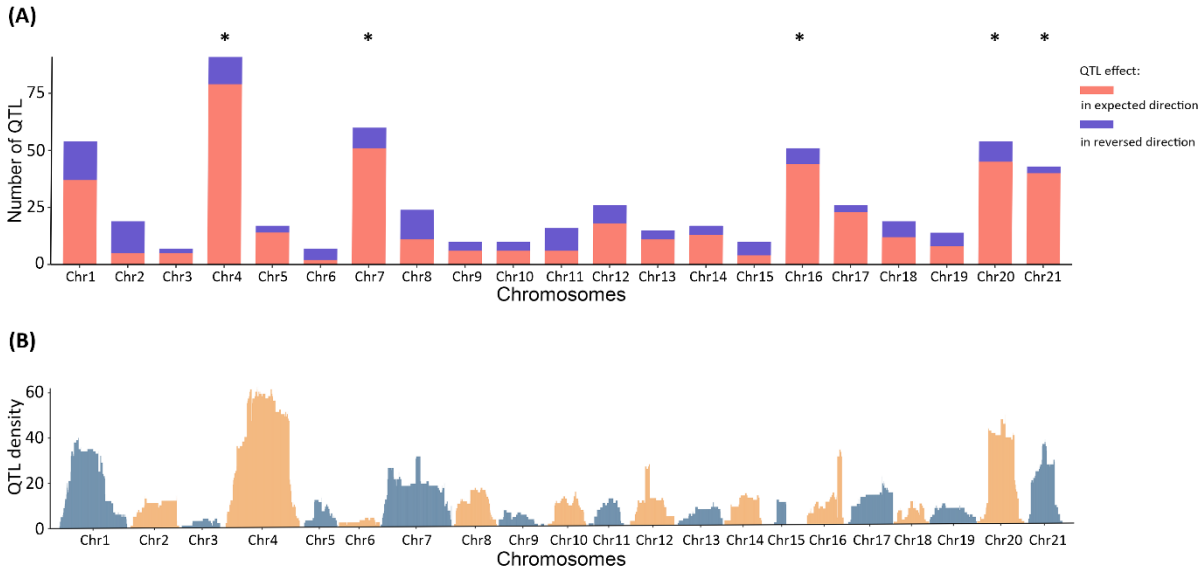


Fig. 3. (A) Counts of QTL underlying traits that differ between marine and freshwater populations with QTL conferring an effect in the expected direction in red, and QTL conferring an effect in the reversed direction in purple. (B) Density of QTL confidence intervals mapped to the *G. aculeatus* genome in 50kb windows. QTL data are collected from previous studies (Supplementary Table S2). Chromosomes with asterisks have significantly more QTL with effects in the expected direction than expected given either the number of genes on the chromosome or the chromosome length (Supplementary Table S2).

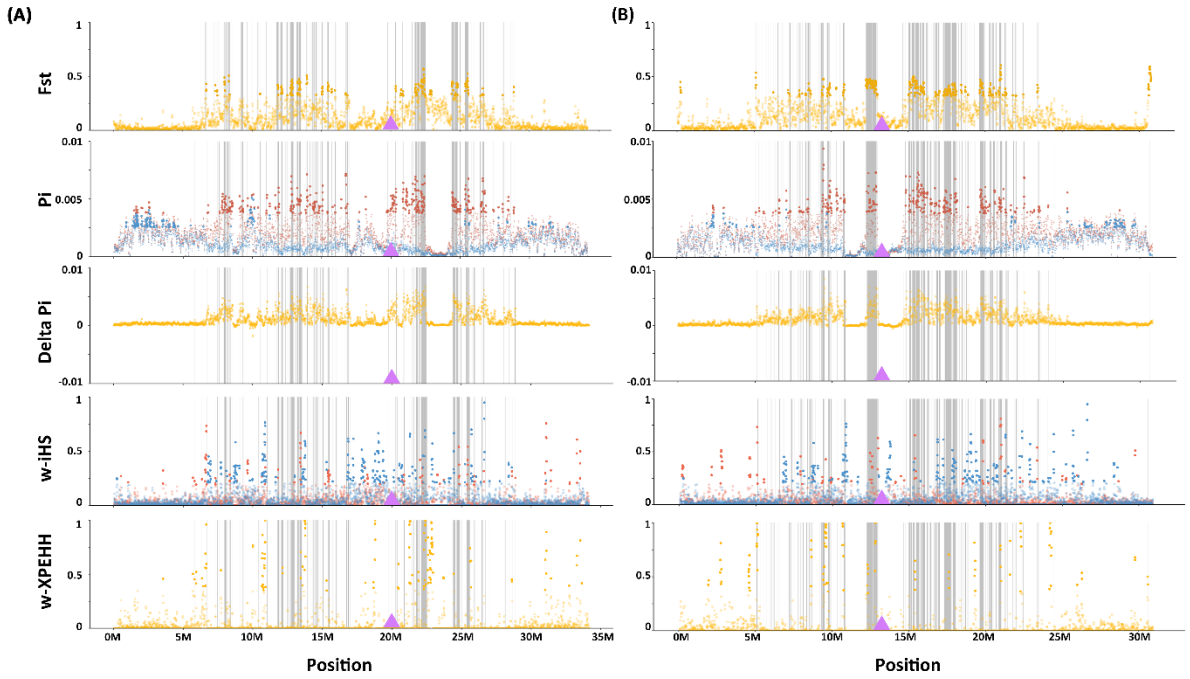


Fig. 4. Signatures of selection in the Lake Washington freshwater and Puget Sound marine populations of *G. aculeatus*. Statistics are shown here for chromosomes 4 (A) and 7 (B), with all chromosomes shown in Supplementary Fig. S5. All statistics were calculated in 20kb sliding windows with a step size of 10 kb. Dark grey bars indicate the genomic islands and the purple triangle indicates the fusion points. From top to bottom: Fst across the whole chromosome, with solid dots highlighting SNPs in the top 5% of genome-wide Fst; nucleotide diversity (Pi) of Lake Washington (red) and Puget Sound (blue) populations, with solid dots highlighting SNPs with the top 5% highest values of Pi in each population; differences of nucleotide diversity between the two populations. (Delta Pi = $Pi_{Lake\ Washington} - Pi_{Puget\ Sound}$); haplotype-based selection statistic iHS, with solid dots indicating the top 5% genome-wide outliers for Lake Washington (red) and Puget Sound (blue); and haplotype-based selection statistic XPEHH, with top 5% genome-wide outliers labeled in solid yellow dots.

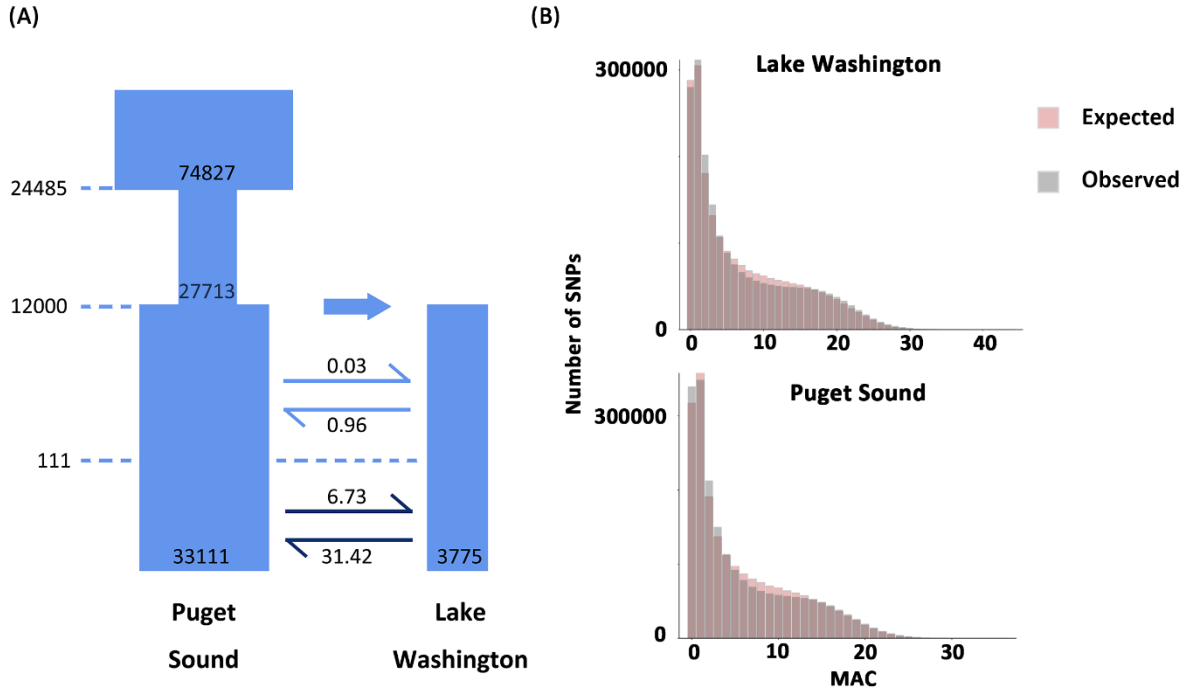
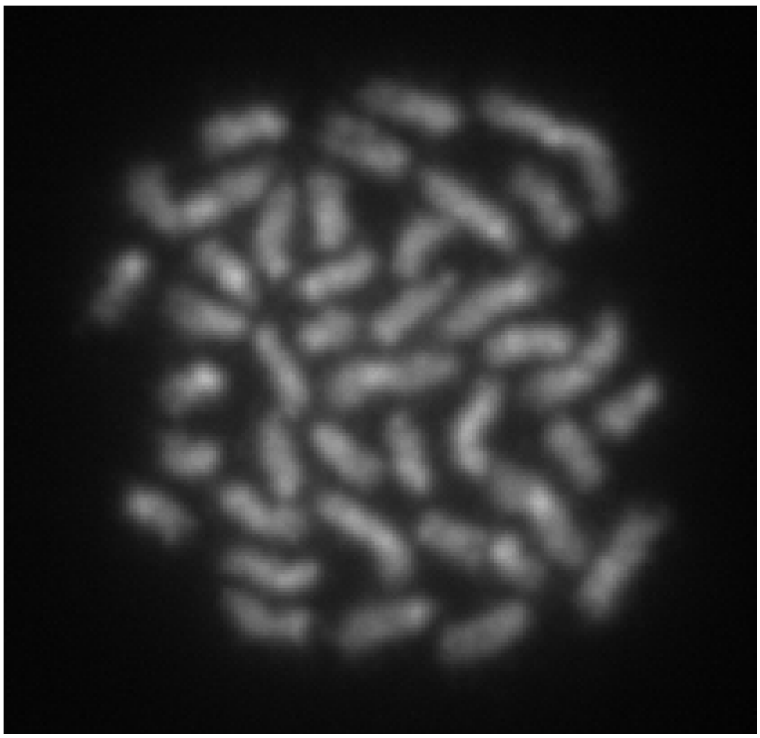
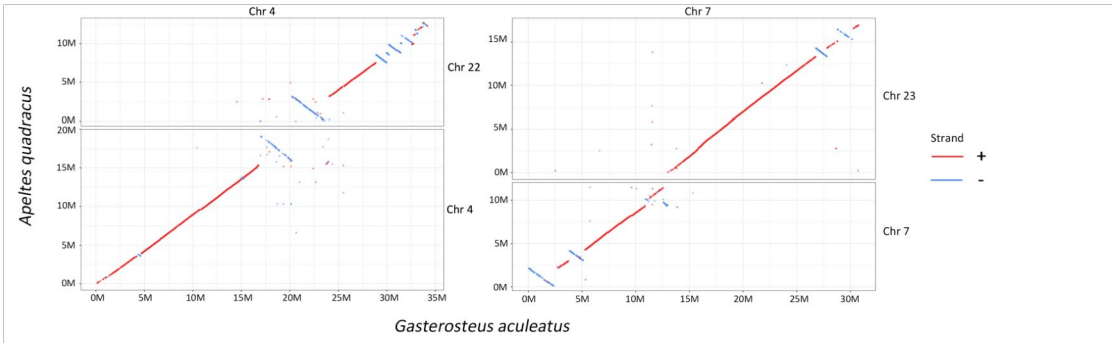


Fig. 5. Demographic model of Lake Washington and Puget Sound populations. (A) Best demographic model inferred by fastsimcoal2. Dashed lines represent the time of the events. (B) Comparison of the observed minor allele count (MAC) spectrum (grey bars) and the simulated minor allele count spectrum (red bars).

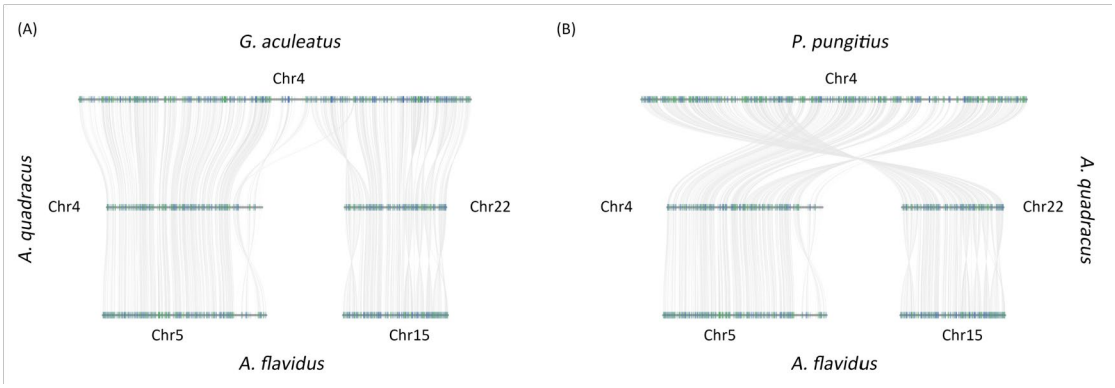
Supplementary Information



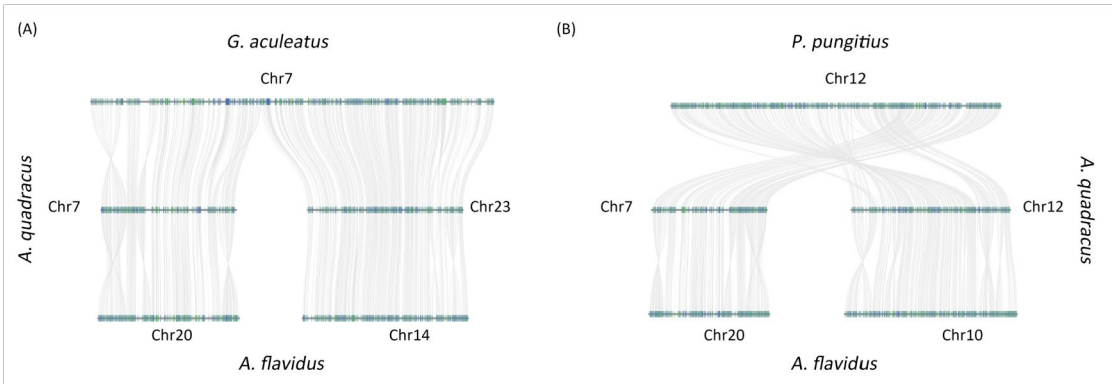
Supplementary Fig. S1. Metaphase spread from a *S. spinachia* male, showing the diploid chromosome number of 40.



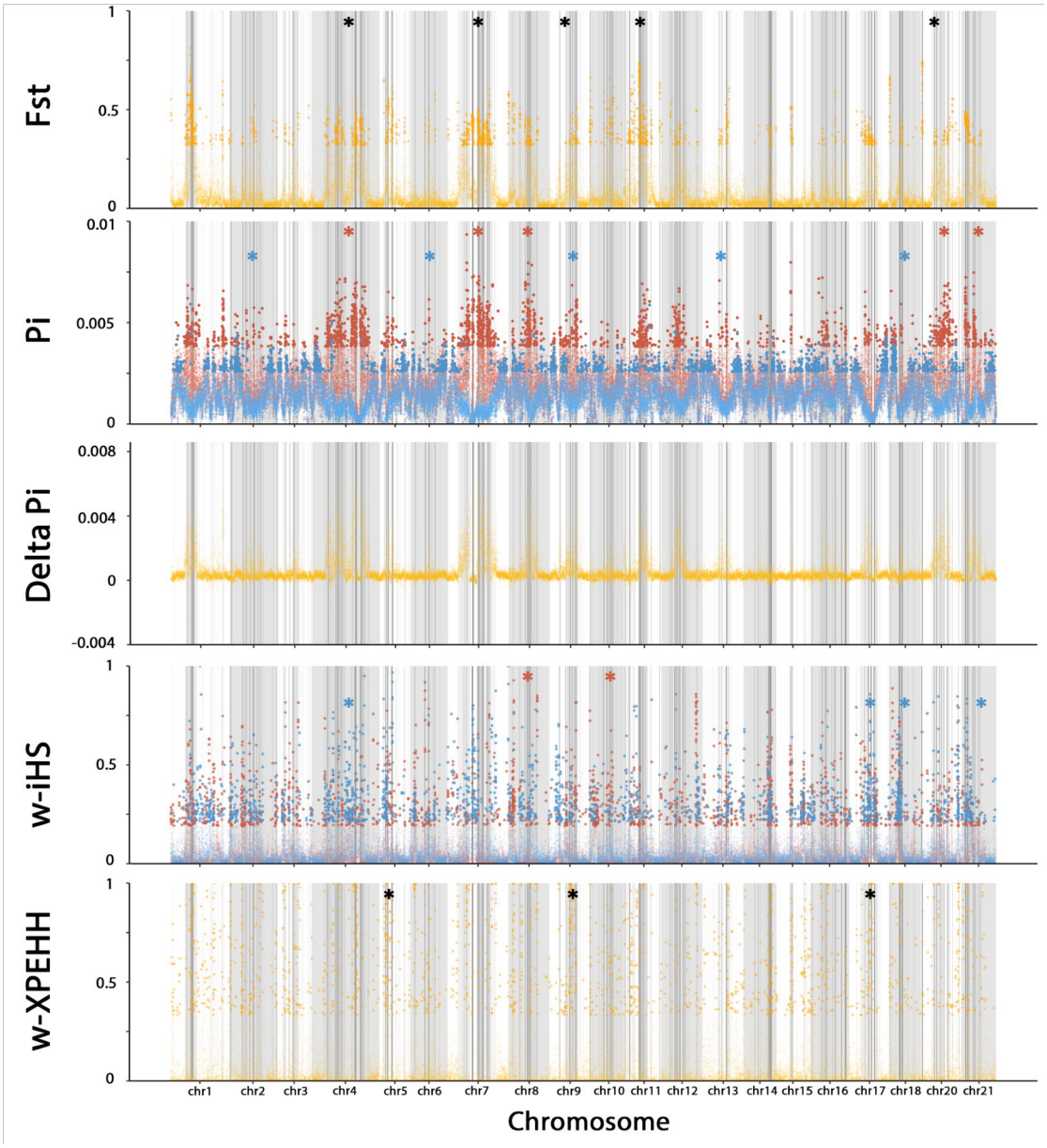
Supplementary Fig. S2. Synteny map of *G. aculeatus* chromosomes 4 and 7 compared with *A. quadracus*, based on coding region sequences using Mummer4 and nucmer.



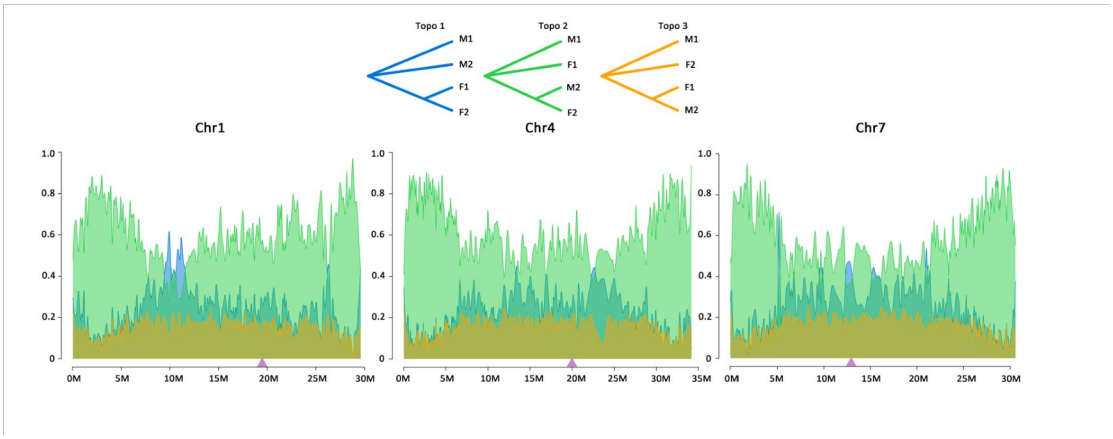
Supplementary Fig. S3. (A) Gene map of *G. aculeatus* chromosome 4 compared with *A. quadracus* and *A. flavidus*. (B) Gene map of *P. pungitius* chromosome 4 compared with *A. quadracus* and *A. flavidus*.



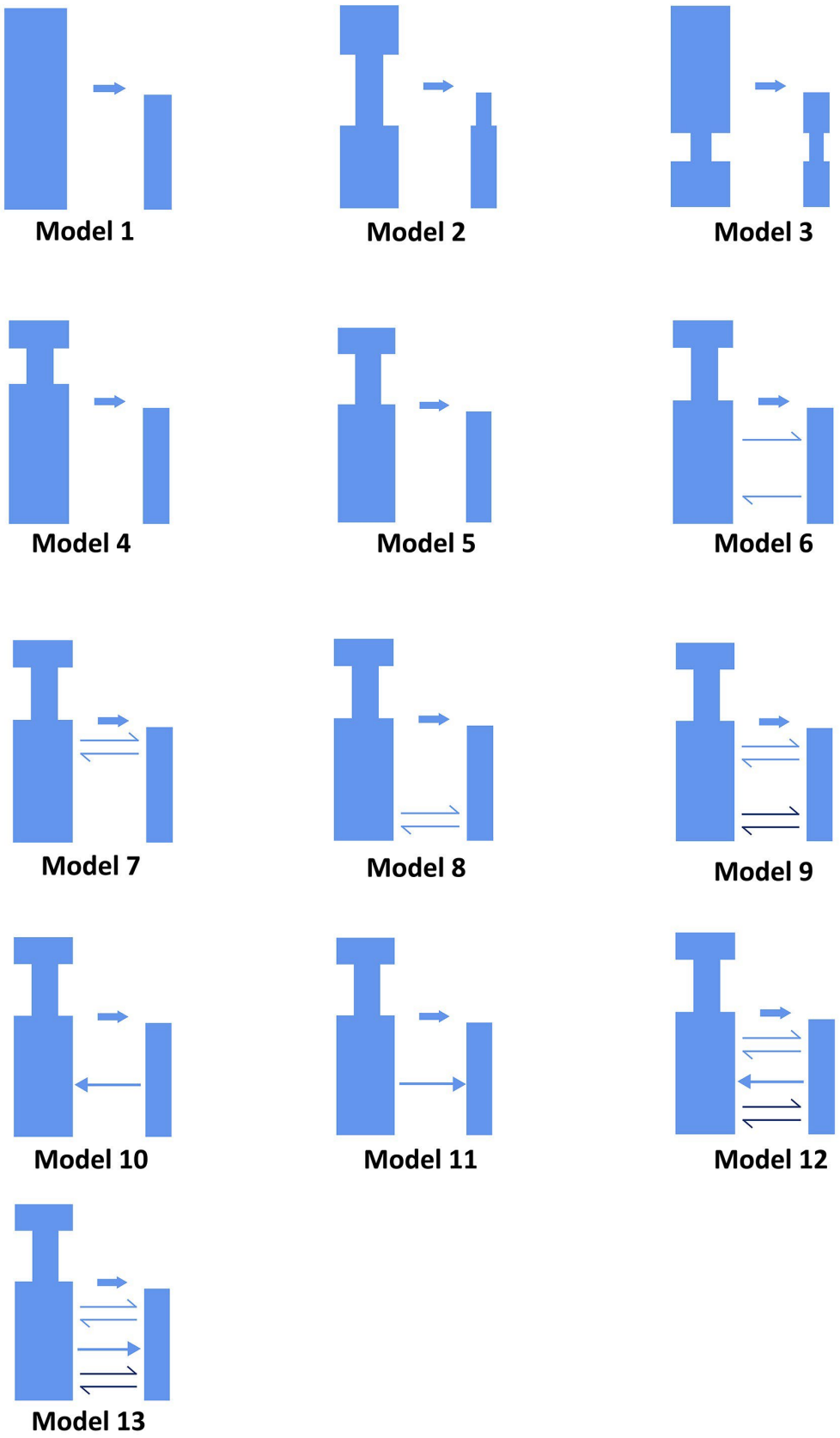
Supplementary Fig. S4. (A) Gene map of *G. aculeatus* chromosome 7 compared with *A. quadracus* and *A. flavidus*. (B) Gene map of *P. pungitius* chromosome 12 compared with *A. quadracus* and *A. flavidus*.



Supplementary Fig. S5. Signatures of selection in the Lake Washington freshwater and Puget Sound marine populations across the whole genome. All statistics were calculated in 20kb sliding windows with a step size of 10 kb. Dark grey bars indicate the genomic islands. From top to bottom: Fst distribution across the genome, with solid dots highlighting SNPs in the top 5% of genome-wide Fst; nucleotide diversity (Pi) of Lake Washington (red) and Puget Sound (blue) populations, with solid dots highlighting SNPs with the top 5% highest values of Pi in each population; differences of nucleotide diversity between the two populations. (Delta Pi = $Pi_{Lake\ Washington} - Pi_{Puget\ Sound}$); haplotype-based selection statistic iHS, with solid dots indicating the top 5% genome-wide outliers for Lake Washington (red) and Puget Sound (blue); and haplotype-based selection statistic XPEHH, with top 5% genome-wide outliers labeled in solid yellow dots. Asterisks represent chromosomes that show significantly greater evidence for selection in Lake Washington (red), Puget Sound (blue) or between the populations (black) than expected, given both the length of the chromosome and the number of genes on the chromosome, based on the standardized residuals from a chi-squared test (Supplementary Table S4).

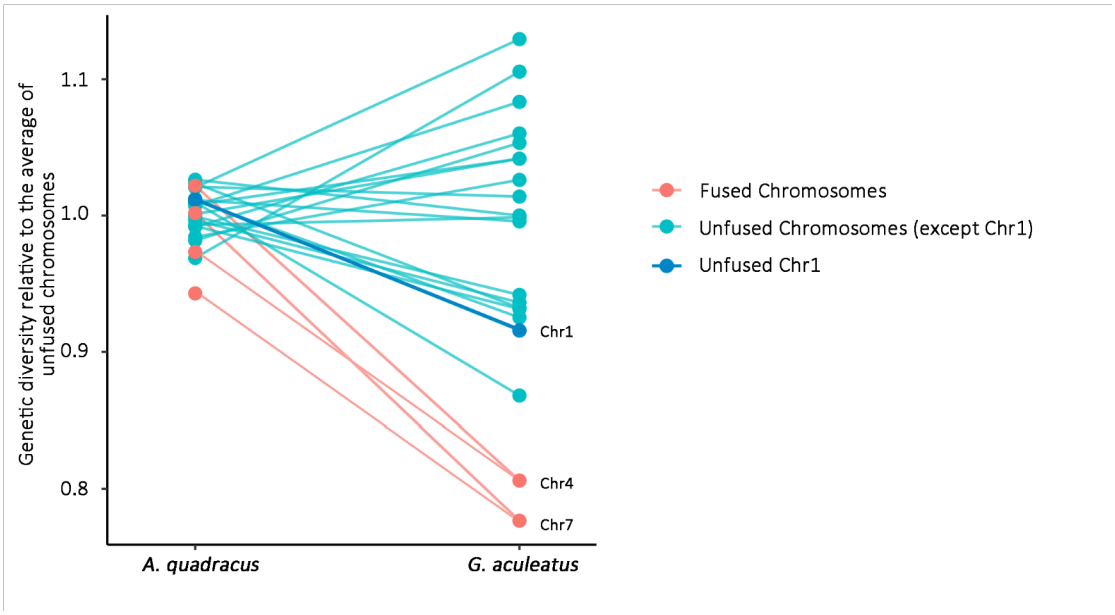


Supplementary Fig. S6. Topology weightings of marine and freshwater haplotypes in *G. aculeatus*. M represents marine haplotypes, while F represents freshwater haplotypes. M1 and F1 represent the major alleles in the respective populations, while M2 and F2 represent the minor alleles. Topo 1 represents the topology in which marine and freshwater ecotypes consistently diverge. Topo 2 and 3 represent topologies in marine and freshwater haplotypes that are not divergent. Purple triangles represent centromeres as well as the fusion points on chromosomes 4 and 7.



Supplementary Fig. S7. Models used in demographic modeling, with the Puget Sound population indicated on the left and the Lake Washington population indicated on the right: 1) constant population size; 2) two bottlenecks while splitting; 3) two bottlenecks after

splitting; 4) one bottleneck before splitting; 5) one bottleneck and splitting; 6) one bottleneck and splitting followed by a constant and reciprocal migration; 7) one bottleneck and splitting followed by an early reciprocal migration; 8) one bottleneck and splitting followed by a recent reciprocal migration; 9) one bottleneck and splitting followed by two reciprocal migration regimes; 10) one bottleneck and splitting followed by introgression from Lake Washington to Puget Sound; 11) one bottleneck and splitting followed by introgression from Puget Sound to Lake Washington; 12) one bottleneck and splitting followed by introgression from Lake Washington to Puget Sound and two reciprocal migration regimes; 13) one bottleneck and splitting followed by introgression from Puget Sound to Lake Washington and two reciprocal migration regimes.



Supplementary Fig. S8 Comparison of genetic diversity between fused and unfused chromosomes in *A. quadracus* and *G. aculeatus*. Genetic diversity is a proxy for measuring the recombination rate on each chromosome and was calculated based on four-fold degenerate sites and normalized relative to the average diversity of unfused chromosomes. The two fused chromosomes (red) have lower genetic diversity relative to the average of unfused chromsomes (dark and light blue) in *G. aculeatus* as well as relative to their unfused homologues in *A. quadracus*, suggesting lower recombination rates on the fused chromosomes.

Supplementary Table S1. Sample information and accession numbers for sequencing data in this study.

Supplementary Table S2. QTL database used in this study and results for the chi-square test of QTL distribution on *G. aculeatus* chromosomes. All QTL are related to traits that differ between marine and freshwater ecotypes, with redundancies removed. Each QTL was examined to determine whether the phenotypic effect of the QTL was in the expected direction, based on the direction of divergence between the parental populations. The expected number of QTL with effects in the expected direction on each chromosome was calculated both by length in base pairs and number of genes on the chromosome and compared to the observed distributions in R using a goodness-of-fit test (chisq.test). Following Peichel and Marques (2017), chromosomes with significantly more QTL in the expected direction were identified by standardized residuals with a value larger than 3 in both comparisons and are highlighted in bold.

Supplementary Table S3. Distribution of gene transposition and gene duplication events on *G. aculeatus* chromosomes. The expected distribution of duplicated genes on each chromosome was calculated both by chromosome length in base pairs and the number of genes on the chromosome and compared to the observed distribution in R using a goodnessof-fit test (chisq.test). Chromosomes with significantly higher values than expected were identified by standardized residuals with a value larger than 3 in both comparisons and are highlighted in bold. There were too few gene transposition events to perform a similar analysis. Chromosome 19 (the sex chromosome) is omitted from these analyses.

Supplementary Table S4. Results of the chi-square test of distribution of signatures of selection on *G. aculeatus* chromosomes, including SNP numbers in genomic islands, nucleotide diversity (Pi), the proportion of iHS, and proportion of XPEHH. The expected distribution on each chromosome was calculated both by chromosome length in base pairs and number of genes on the chromosome and compared to the observed distribution in R through a goodness-of-fit test (chisq.test). Chromosomes with significantly higher values than expected were identified by standardized residuals with a value larger than 3 in both comparisons and are highlighted in bold. Chromosome 19 (the sex chromosome) is omitted

from these analyses.

Supplementary Table S5. Results of the comparisons among the 13 demographic models, andthe medians and 95% confidence interval of the parameters in the best-fitting model (model9: one bottleneck and splitting followed by two reciprocal migration regimes).

See https://academic.oup.com/mbe/article/39/2/msab358/6462204#supplementary-data

Reference

Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Phil Trans R Soc B.* 372(1713):20150486.

Chapter 2

The fourspine stickleback (*Apeltes quadracus*) has an XY sex chromosome with polymorphic inversions on both X and Y chromosomes

Zuyao Liu[1], Yingguang Frank Chan[2] are a[2], and Catherine L. Peichel[1, *]

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

[2]Friedrich Miescher Laboratory of the Max Planck Society, 72076 Tübingen, Germany

***Corresponding author:** E-mail: catherine.peichel@unibe.ch

This chapter will be submitted to PLoS Genetics.

Abstract

Teleost fish are well-known for possessing a diversity of sex chromosomes and for undergoing frequent turnovers of these sex chromosomes. However, previous studies have mainly focused on species with heteromorphic sex chromosomes, while comparatively little attention has been given to species with homomorphic sex chromosomes, which may capture early stages of sex chromosome turnovers. To better understand the evolution of sex chromosomes, we used the fourspine stickleback (*Apeltes quadracus*) as a model organism. Previously, it was believed that females of this species possessed a ZW heteromorphic sex chromosome system. However, our whole-genome sequencing of wild populations and genetic crosses revealed that *A. quadracus* actually has a homomorphic XY sex chromosome on Chr23. This chromosome has not previously been identified as a sex chromosome in other stickleback species, indicating a recent sex chromosome turnover. We also identified two genes - *rxfp2a* and *zar1l* - as novel candidate sex determination genes. Notably, we observed inversions on both the X and Y chromosomes in different populations, which have shaped distinct strata among populations. We propose that the inversion on the X chromosome may have been favored by beneficial selection in females, in contrast to the Y-specific inversion. The new sex chromosome and polymorphic inversions observed in *A. quadracus* provide an excellent system for studying the evolution of sex chromosomes and their turnovers.

Keywords

Fourspine stickleback; sex chromosome; sex determination; chromosomal inversion; recombination; Gasterosteidae

Author Summary

As compared to mammals and birds, teleost fish exhibit a very high level of diversity in their sex chromosomes, even among closely-related species. Thus far, little attention has been paid to variation within species, particularly those with homomorphic sex chromosomes, although they offer a valuable opportunity to advance our understanding of the mechanisms underlying the formation and turnover of sex chromosomes. Through the analysis of sequences obtained from diverse populations, we determined that instead of the previously reported ZW system, *A. quadracus* has an XY sex determination system on Chr23. Within the sex determining region, we identified *rxfp2a* and *zar1l* as putative sex determining genes. Notably, we also observed polymorphic inversions present on both the X and Y chromosomes that are still segregating within each population. Based on our findings, we hypothesize that the X-linked inversions are favored by sexually antagonistic selection. These observations represent a rare condition in which sex chromosomes are still polymorphic for sex-linked inversions, which offers important insights into the early stages of sex chromosome evolution.

Introduction

Although sex determination systems are diverse across species, genetic sex determination mechanisms associated with the presence of heteromorphic sex chromosomes have independently evolved many times across the tree of life [1]. There are two main types of sex chromosomes. When males are the heterogametic sex, as in mammals, females have two X chromosomes and males have an X chromosome and a Y chromosome. When females are the heterogametic sex, as in birds, females carry a Z and a W chromosome, and males have two Z chromosomes. Although some groups like mammals and birds have very stable sex chromosome systems, in other groups like frogs [2,3], lizards [4], and fishes [5–7], even closely related species have different sex chromosome systems [8].

This diversity of sex chromosome systems is due to sex chromosome turnover, which occurs either when an existing sex determination gene moves to a new chromosome or when a novel sex determination gene arises on a chromosome [9,10]. According to the classical model of sex chromosome evolution, the acquisition of a new sex determination gene on an autosome can lead to the loss of recombination between the X and Y (or Z and W) chromosomes, resulting in the accumulation of deleterious mutations on the sex-specific and therefore non-recombining chromosome (Y or W) and the eventual formation of the heteromorphic sex chromosome pair [11,12]. This process can be interrupted by sex chromosome turnover, which resets the cycle and initiates the process again [9,10]. Although the evolutionary forces driving these turnovers are still unknown, sex chromosome turnovers have been hypothesized to occur due to selection for linkage of sexually antagonistic alleles with a sex-determination locus [13], selection to purge deleterious mutations that have accumulated on sex chromosomes [14,15], selection to maintain unbiased sex ratios [16,17] , or random genetic drift [16,18,19]. However, testing these hypotheses remains challenging because it is difficult to catch such turnovers when they occur [10].

The presence of polymorphic sex chromosomes within species might provide an opportunity to catch turnovers at an early stage. Intraspecies variation has been found in different groups, including frogs [2,20–22] and fishes [23–25]. Studies of these polymorphic systems have provided some insights into the evolutionary forces driving turnovers. For example, invasion of a new sex chromosome in cichlids is associated with linkage to a trait under sexually antagonistic selection [26]. Population-specific variation in the presence of a

sex chromosome in clawed frog is consistent with selection to purge deleterious mutations on the sex-specific chromosome [20]. Despite these examples, the evolutionary drivers and genetic mechanisms that underlie intraspecies variation in sex chromosomes is still mostly unknown.

A variety of sex chromosome systems have been identified in the species of the stickleback family (Gasterosteidae) that have diverged within the past 30 million years [27], suggesting that there have been recent sex chromosome turnovers. The three species in the genus *Gasterosteus* possesses a conserved heteromorphic XY sex chromosome on chromosome 19, with the *amhy* gene as the candidate sex determination gene [28,29]. The ancestral Y chromosome has independently fused to different autosomes in *G. nipponicus* and *G. wheatlandi* [7,30,31]. An independent duplication of *amh* has also been found as a candidate sex determination genes on chromosome 20 of *Culaea inconstans* [27]. Chromosome 12 is involved in an XY sex determination system in some *Pungitius* species, and another ZW sex determination system on chromosome 7 has been detected in *P. sinensis* based on genetic mapping [32,33]. Even so, sex chromosomes have not yet been fully identified in other species in this family and, therefore, additional work is needed to explore the origins and evolution of sex chromosome evolution and turnovers in this family.

Fourspine sticklebacks (*Apeltes quadracus*) are of interest, as previous studies suggest they might possess a different sex chromosome than other stickleback species, but the sex chromosome has still not been identified. Initially, cytogenetic analysis of a population from Maine showed that *A. quadracus* has a heteromorphic ZW sex chromosome [34]. In the years following the initial study, conflicting evidence from new cytogenetic analyses were reported. Females from a Massachusetts (MA) population were found to have a heteromorphic sex chromosome, while no heteromorphic sex chromosome was identified in females (or males) from a Connecticut (CT) population [7,35]. These data suggested that the sex chromosome system might be polymorphic within *A. quadracus*, making this species an attractive target for further study of the evolution of sex chromosome turnover.

To identify the sex chromosome in fourspine stickleback, we collected samples from three different populations, two of which (MA and CT) were used for the previous cytogenetic studies, and the other population from Nova Scotia (NS) was used for the recently published female genome assembly [36]. For each population, we collected wild samples and created crosses from a single mother and father per population. We generated haplotagging linked-

read sequencing data from individuals of the wild populations and pooled sequencing data from the crosses to identify the sex chromosome and sex determination region in fourspine stickleback. We further used these data to explore the variation on the sex chromosome among populations and identify candidate sex-determination genes.

Results

***A. quadracus* has an unexpected XY sex determination system on Chr23**

We first utilized pool-seq data from crosses of each of three populations (CT, MA, and NS in Fig 1) to determine the location of the sex chromosomes in *A. quadracus*. Analysis of the sequencing depth ratio between males and females from the three crosses did not reveal any clear reductions in either sex, suggesting the presence of homomorphic sex chromosomes (Supplementary Fig S1). However, if mutations specific to either a female-specific or male-specific chromosome have accumulated, increased genetic differentiation between the two sexes as well as increased diversity within the heterogametic sex on the sex chromosome would be expected. Consistent with these predictions, fixation index (Fst) between males and females shows elevated differentiation on Chr23 in all three crosses (Supplementary Fig S2). The extent of genetic differentiation on Chr23 varies among crosses. Differentiation is elevated from 4.38 to 9.40 Mb in the CT cross, from 8.58 to 9.40 Mb in the MA cross, and from 0 to 15.00 Mb in the NS cross (Fig 2A). Also, there is a region with extremely low genetic differentiation between two sexes in the MA cross (Fig 2A). Distributions of genetic diversity (Pi) on Chr23 within the two sexes also vary among the three crosses (Supplementary Fig S3). In the CT cross, males have higher diversity than females between 4.38 and 9.40 Mb, while in the MA cross, both sexes have high levels of diversity in the same region. In the NS cross, only males have higher diversity between 8.58-9.40 Mb (Fig 2B). We also counted the number of sex-specific SNPs in each cross. In the CT cross, male-specific SNPs are enriched between 4.38-9.40 Mb, while the enrichment of male-specific SNPs in both the MA and NS crosses is between 8.58-9.40 Mb (Fig 3).

Because these data are from relatively small genetic crosses, the number of recombination events limits our ability to narrow down the location of the sex determination region. Hence, we conducted linked-read sequencing of wild samples from the same three populations (20 females and 20 males for CT populations, 13 females and 11 males from MA populations, and 15 females and 14 males from NS population). Consistent with the cross data,

the ratio of sequencing depth between males and females shows no clear difference among three populations or between the sexes (Fig 4A), but there is high genetic differentiation between the sexes on Chr23 (Fig 4B). Also consistent with the data from the crosses, the distributions of Fst and Pi on Chr23 show different patterns among the three populations (Fig 5; Supplementary Fig S4). Differentiation is elevated from 4.38 to 9.40 Mb in the CT population and from 8.58 to 9.40 Mb in the MA population. In the NS population, there is a moderate elevation between 4.11 and 12.20 Mb with an extremely high elevation between 8.58 and 9.40 Mb (Fig 5A). The CT population has higher genetic diversity in both females and males between 4.38 and 9.40 Mb compared to the genomic background, with males exhibiting even higher levels of diversity compared to females. The MA and the NS populations only show elevated male diversity between 8.58 and 9.40 Mb (Fig 5B).

In summary, the genetic differentiation between males and females is prominent on Chr23 and varies among populations. Males exhibit higher levels of diversity than females, as seen in both crosses and wild samples. And, in all three crosses, there are more male-specific SNPs than female-specific SNPs on Chr23. Further, our analysis of RNA-seq data from the NS cross by SEX-DETector [37] shows more sex-linked transcripts with evidence of male heterogamety than female heterogamety (Supplementary Table S1). Taking these results together, we conclude that *A. quadracus* has an XY sex determination system and that Chr23 is the sex chromosome. Using data from the wild populations, we identify a shared sex determination region between 8.58 Mb and 9.40 Mb.

Different populations have different X- and Y-linked inversions

The different patterns of differentiation and diversity in the three populations led us to hypothesize that there might be population-specific inversions on the sex chromosomes. To further investigate this possibility, we took advantage of the linked-read sequencing that we performed on wild fish from each population. To confirm the presence of inversions and identify the location of breakpoints, we combined three lines of evidence (See Material and Methods and Supplementary Table S2 for details).

Analysis of the linked reads in the CT population reveals that there is an inversion relative to the reference genome [36] between 4.38 and 9.117 Mb, with polymorphism in both sexes. Specifically, 17 of 20 females are homozygous for the inverted orientation, while 17 of 20 males are heterozygous with only one copy of the inversion (Supplementary Table S2). Hence,

it can be inferred that the inversion is on the X chromosome, which explains its existence in both sexes, and that it is almost fixed based on the SNP density plots at the individual level (Supplementary Appendix 1). The presence of this X-linked chromosome inversion explains the high genetic differentiation between males and females in both the cross (Fig 2A) and population data (Fig 5A). The elevated diversity in CT wild females (Fig 5B) is also consistent with the fact that a few females are heterozygous for the inversion. The lack of elevated diversity in females from the CT cross (Fig 2B) suggests that the mother of this cross was homozygous for the X-linked inversion and that the father also had an X chromosome with the inversion, such that all daughters were homozygous for the inversion on the X chromosome (Supplementary Fig S5).

In the MA population, genetic differentiation and the difference in genetic diversity between sexes is only found in the sex determination region (Fig 4B and Fig 5A). These data suggest that the X-linked inversion is not found at a high frequency in this population. However, we were unable to assess this directly because the average sequencing depth of the MA population was not high enough to confidently genotype the inversions. It is noted, however, that in the region of the X-linked inversion, there is low divergence between males and females and high genetic diversity within males and female in the cross data (Fig 2). This could be explained if the mother of the cross was heterozygous for the inversion and the father did not carry the inversion. In this case, daughters and sons would have equal frequencies of the inversion, resulting in no differentiation between males and females (Supplementary Fig S5) and similar levels of diversity within males and females, as we observe (Fig 1). Together, these data suggest that the X-linked inversion might be present at low frequency in the MA population.

For the NS population, there is an inversion between 4.47 and 12.15 Mb that is polymorphic in males only, indicating a Y-specific inversion. Consistent with this, there is a region of slightly elevated genetic differentiation between 4.11 and 12.20 Mb in the NS population (Fig 5A). However, in both the cross and population data, genetic differentiation between males and females as well as elevated diversity within males is strongest within the shared sex-determination region, suggesting that there has not been much divergence between the X and the Y within this inverted region.

In addition, there is a potentially smaller inversion in both sexes and all populations between 8.60 and 10.27 Mb, which is supported by excessive shared barcodes between

windows near breakpoints, split reads, and discordantly mapped read pairs. However, due to the enrichment of repeats in the entire sex determination region, the genotype of individuals cannot be determined for this inversion (Supplementary Table S2).

Summarizing the above evidence, a model is proposed for visualizing the pattern of inversion in different populations (Fig 6). The CT population has an X-specific inversion partially covering the sex determination region, whereas the NS population has a Y-specific inversion covering the entire sex determination region and most of the Y chromosome as well. The two identified inversions in the CT and NS populations are derived, as they are inverted relative to the female genome assembly from the NS population, whose orientation appears to be ancestral by comparison to the genome assemblies of other stickleback species [36].

Population-specific evolutionary trajectories of sex chromosomes

A population-specific inversion would result in very different evolutionary trajectories of the sex chromosomes in the different populations. To further investigate this, we first calculated male-specific SNPs in each population with linked-read data (Supplementary Table S3). Note that none of the following analyses involved the MA population because of its poor sequencing quality. In the CT population, all fixed male-specific SNPs are distributed within the shared sex-determination region, and no male-specific SNP is detected within the X-linked inversion except for the region that overlaps the sex determination region. This is not surprising given that males in this population also carry X chromosomes with the inversion. In the NS population, most male-specific SNPs are detected in the shared sex determination region. However, several male-specific SNPs are found outside this region, but within the Y-specific inversion, providing further evidence for its existence (Supplementary Table S3). We then compared the distribution of dS values between X and Y chromosomes across the entire Chr23 (Fig 6). In both the CT and NS populations, genes within the shared sex-determination region have relatively high dS values, while regions that are involved in the inversions have intermediate dS values. A notable exception is a region between 0 and 4.475 Mb in the NS population, which is likely due to the small number of genes and a single outlier value in this region. Nonetheless, these lower levels of dS in the inverted regions suggest that the two population-specific inversions developed after the initial divergence of the sex determination region.

No evidence of extensive degeneration within the inversions

Inversions can lead to a suppression of recombination in heterozygotes, causing an accumulation of deleterious mutations. Thus, inversions have been suggested to be associated with loss of recombination and subsequent degeneration on sex chromosomes [38,39]. Although the sex chromosomes of *A. quadracus* have homomorphic sex chromosome with no large region of depth reduction, we explored whether there has been degeneration at a fine scale. One method to identify degenerated genes on the sex chromosomes involves comparing read depth of genes between males and females in wild populations using linked-read data. If the ratio of male to female depth is less than 0.75, the gene is considered to be degenerate, indicating a loss of its content [31]. In the CT population, there are four genes on Chr23 that are degenerate based on this criteria (Supplementary Table S4). Three of them are located within the sex determination region, and one is located outside of the sex determination region but within the X inversion. In the NS population, two genes are identified as degenerate and both are located in the sex determination region (Supplementary Table S4). We also looked for the presence of premature stop codons as evidence for degeneration. There are 8 premature stop codons in males and 2 premature stop codons in females in the CT population, and no premature stop codons in either males or females in the NS population (Supplementary Table S4). Considering the evidence, the X inversion may contribute to the degeneration of sex chromosomes.

Rxfp2a* and *zar1l* are candidate sex-determination genes in *A. quadracus

Following the identification of the sex determination system and sex chromosome in *A. quadracus*, we were interested in identifying the key gene responsible for sex determination. Since the *A. quadracus* genome assembly was from a female individual, it is missing any genes that are specific to males. Using a kmer-based approach, we reconstructed male-specific fragments separately from the pool-seq data for each of the three populations and searched the NCBI ref-seq database to identify genes. For each gene present in all populations, we counted the number of SNPs located in the coding region and the non-coding region, separately for both sexes. We focused on genes that had male-specific SNPs, but not female-specific SNPs, due to the XY sex determination system. In total, there are 17 such candidate sex-determination genes located on Chr23 within the sex-determination region (Supplementary Table S5). Among these genes, there are two genes of interest, *rxfp2a* and

zar1l, which are related to the development of reproductive system (see Discussion for details). There are 4 male-specific SNPs in the *zar1l* coding region, and one is predicted to cause a deleterious mutation, according to PROVEAN analysis. The *rxfp2*a gene has three SNPs that are specific to males. Two of these SNPs are located in the regulatory region of the gene, and one is a nonsynonymous mutation located in the coding region. However, the SNP in the coding region is predicted to have no significant effect on the function of the gene.

Discussion

Variation and turnover of sex chromosomes in *A. quadracus*

Previous evidence from cytogenetic studies suggested that populations of *A. quadracus* from Maine and Massachusetts have a heteromorphic ZW sex chromosome in females [7,34]. However, no heteromorphic sex chromosome was detected in metaphase spreads of males or females from Connnecticut [35]. Using sequencing data from genetic crosses and wild fish from three populations, we determined that *A. quadracus* has an XY sex determination system on Chr23. Our analyses included samples from both the Massachusetts and Connecticut populations used in the previous cytogenetic studies. Thus, the discovery that *A. quadracus* has an XY sex determination system is surprising, as the morphology of chromosomes in the MA population clearly indicated the presence of a heteromorphic pair in females [7]. One possible explanation for this result is if the MA individuals used for cytogenetics were heterozygous for the X-linked inversion that we identified in the CT population. If the inversion caused a change in chromosome morphology at the cytogenetic level, we might see what appears to be a heteromorphic chromosome pair. Although we performed linked-read sequencing of some of the MA females used for the cytogenetic study [7], we did not obtain good enough sequence to confidently assess inversion genotypes in these individuals. However, the MA cross data suggest that the X-linked inversion is present in the MA population (Fig 2). As we do not have samples from the Maine population used in the older cytogenetic study [34], we could not assess whether the X-linked inversion is present in this population. If heteromorphic chromosomes in females are indeed due to heterozygosity for the X-linked inversion, it is not surprising that the CT females were homomorphic in the previous cytogenetic study since these females are mostly fixed for the inversion (Supplementary Table S2). However, to fully resolve this mystery, a more detailed molecular cytogenetic analyses of these different populations is needed, which will be facilitated by our

identification of the sex-determination region on Chr23 in *A. quadracus*.

Chr23 has not previously been identified as a sex chromosome in sticklebacks, suggesting that there has been a sex chromosome turnover in *A. quadracus*. However, it is interesting to note that *A. quadracus* Chr23 is homologous to part of chromosome 7 in both *Gasterosteus aculeatus* and *Pungitius pungitius* [36]. The non-homologous part of chromosome 7 has fused to chromosome 12 in the *Pungitius* lineage, and there is evidence that chromosome 7 carries a female heterogametic (ZW) sex determination locus in *P. sinensis* and that chromosome 12 carries a male heterogametic sex (XY) determination locus in *P. pungitius* [32]. Given that the sex-determination region in these two species is not homologous to that in *A. quadracus*, it is unlikely that they have the same sex determination gene. However, testing this hypothesis requires identifying the sex-determination gene in all three species. It is clear that *A. quadracus* has a different sex chromosome and sex determination gene from the *Gasterosteus* species, in which the master sex determination gene *amhy* is found on chromosome 19 [28,29,31,40], or in *C. inconstans*, in which there has been an independent duplication of *amhy* on chromosome 20 [27]. No duplicated copy of *amh* has been found in *Pungitius* species or in *A. quadracus* [27]. Further supporting a sex chromosome turnover in *A. quadracus* is the lack of extensive differentiation between the X and the Y or degeneration on the Y chromosome. Similar patterns on sex chromosomes in *P. pungitius*, *P. sinensis*, and *C. inconstans* hint that these turnovers also occurred quite recently [27,32,41]. The sex chromosomes in these species are in contrast to the Y chromosome in the *Gasterosteus* lineage, which evolved approximately 22 million years ago and has experienced extensive degeneration, albeit at different rates in the three species in this genus [28,29,31]. This variation in turnover among different stickleback lineages provides an opportunity to further investigate the factors that lead to sex chromosome stability in some lineages and turnover in others.

Two novel candidate sex determination genes

We identified two novel candidate sex determination genes in the shared sex determination region on Chr23. The genes *zar1l* and *rxfp2a* are the only two male-specific genes (characterized by male-specific SNPs within the sex determination region across all populations studied), which are known to play roles in the development of the reproductive system (Supplementary Table S5).

The *rxfp2a* (relaxin/insulin-like family peptide receptor 2) gene encodes a receptor that plays a crucial role in the development of placental mammals by binding with high affinity to the peptide *INSL3* (insulin-like 3). This *INSL3/RXFP2* pairing is essential for the proper descent of the testicles during development in mammals [42,43]. Loss of *rxfp2a* results in cryptorchidism in mice [44–47]. Phylogenetic analysis of 71 mammalian genomes has revealed that the *rxfp2a* gene is lost or non-functional in four afrotherian species that lack testicular descent [48]. Studies in zebrafish have shown that *INSL3* is involved in regulating spermatogonial stem cell differentiation from mitosis to meiosis [49]. Since *rxfp2a* is a receptor for *INSL3*, mutations in this gene have the potential to disrupt the entire *INSL3/RXFP2* signaling pathway, ultimately affecting spermatogenesis. In *A. quadracus*, *rxfp2a* exhibits male-specific mutations in both the coding and regulatory regions. Given the conserved role of this gene in testes development and spermatogenesis, we consider *rxfp2a* a potential sex determination gene that warrants further investigation.

As a maternal effect gene conserved across vertebrates*, zar1* plays an important role in oocyte-embryo transition and impacts female fertility in mice [50]. In *Xenopus laevis*, the *zar1* gene controls the translation of Wee1 and Mos mRNAs in immature oocytes [51]. Additional evidence about *zar1* impacting the sex ratio was found in *Danio rerio*, where a complete male-biased sex ratio was observed in *zar1* knock-out mutants [52]. In addition, it was reported to have an effect on a number of known translation factors, such as CEPB, ePAB, and 4E-T [52,53], among which CPEB and ePAB are known for controlling the process of oogenesis [54] and 4E-T is associated with human primary ovarian insufficiency [44]. In our study, we found that there are two *zar1* genes in the *A. quadracus* genome: the ancestral copy is on Chr8, and the duplicated copy (*zar1l*) is found on both the X and Y copies of Chr23, with one amino acid change in males that is predicted to be deleterious. Considering this gene is quite conserved across species, it is likely that the amino acid change disrupts the function. Hence, having one functional copy of *zar1* could leads to male development, which would be consistent with the zebrafish data. Therefore, we conclude that *zar1l* is another appropriate candidate gene. However, further experiments, such as gene knock-outs and/or SNP editing by CRISPR-Cas9 are be necessary to determine whether *rxfp2a* or *zar1l* is the master sex determination gene in *A. quadracus*.

While numerous sex determination genes have been identified and studied in fish, the two genes mentioned above, *rxfp2a* and *zar1l*, have not been previously identified as sex

determination genes. In contrast, other genes, such as *amh*, *amhr2*, *dmrt1*, and *gdf6*, have been repeatedly identified as master sex determination genes in various fish species [55]. In stickleback species, one key sex determination gene is the independent duplication of the *amh* gene on the Y chromosome of *Gasterosteus* species [29] and *C. inconstans* [27]. However, there is no evidence for an additional copy of the *amh* gene on Chr23 (this study) or elsewhere in the *A. quadracus* genome [27], suggesting that *A. quadracus* has probably undergone a turnover in the sex determination gene. However, it is possible that a duplication of the *amh* gene or other genes unique to the Y chromosome may have been missed in our analysis due to the limited resolution of short-read data. Therefore, assembling a complete Y chromosome is necessary to confirm the absence or presence of a duplication of *amh* or other putative sex determination genes in *A. quadracus*.

Polymorphic X- and Y-linked inversions on the sex chromosomes of *A. quadracus*

We have also identified polymorphic and derived inversions on both the X and Y chromosomes in *A. quadracus* populations. There was a high frequency of an X-linked inversion in both males and females in the CT population, partially covering the sex determination region (Fig 5, Supplementary Table S2 and Appendix 1). Evidence for a similar X-linked inversion was also found with the Pool-seq data from the MA cross, indicating that it might be present at a low frequency in this population (Fig 2B). Although this X-linked inversion does not seem to be present in the NS population, discordantly mapped reads and shared barcodes point to a Y-specific inversion in this population (Supplementary Table S2).

Inversions have been proposed as a mechanism to suppress recombination between X and Y chromosomes [39]. A number of hypotheses have been proposed to explain the suppression of recombination on sex chromosomes, including sexual antagonism [12,56–59], meiotic drive [60], dosage compensation [61], sheltering of recessive deleterious mutations in heterozygotes [62,63], neutral processes [64], and genetic drift [65,66]. Our finding of a Y-linked inversion in the NS population is consistent with all of these models for suppression of recombination between the X and the Y (except for [64] which models the suppression of recombination in the absence of inversions). Indeed, several studies have now found evidence for Y-linked inversions associated with suppression of recombination on Y chromosomes [29,67].

However, the fixation of the inversion on X chromosomes, as we observe in the CT

population, is not predicted by all models. Neither the dosage compensation model nor the sheltering of recessive deleterious mutations in heterozygotes should select for inversions on the X chromosome. This is because only Y-linked inversions are always heterozygous in males and therefore lead directly to suppression of recombination in males and/or to the sheltering of recessive deleterious alleles. Although inversions on the X or the Y could theoretically contribute to suppression of recombination between meiotic drivers and a sex determination locus, this has not been explicitly modeled. Of the remaining models (sexually antagonistic selection, genetic drift), we favor the hypothesis that sexually antagonistic selection might be involved. Recessive beneficial mutations, including inversions, do have a higher probability of fixation via genetic drift on X chromosomes than on autosomes, but this is only the case when X-linked alleles are hemizygous in males as in highly degenerate sex chromosomes [65]. As the *A. quadracus* Y chromosome has not experienced much degeneration, drift alone is not likely to be the explanation for the fixation of the X-linked inversion. Rather, selection for linkage between the sex-determination locus and a locus with beneficial fitness effects in one sex and detrimental effects in the other (sexual antagonism) could select for inversions on the X chromosome [65]. Direct comparisons between the rate of fixation of X and Y-linked inversions under the sexually antagonistic selection hypothesis have not been done, but X-autosome fusions (which also could suppress recombination between the sex determination locus and a sexually antagonistic allele) can spread under sexually antagonistic selection, albeit more slowly than a Y-autosome fusion [57]. Consistent with the predictions of the sexual antagonism hypothesis, the inversion on the X chromosome does contain the shared sex determination region. However, additional work is necessary to test this hypothesis, including performing long-read sequencing in order to fully assemble the sequence of the X-linked inversion and the Y chromosome, assessing the frequencies of the X-linked inversion across many *A.quadracus* populations, and determining whether phenotypes under sexually antagonistic selection are associated with the inversion .

Conclusions

Although variation in sex chromosomes systems among closely-related species is now well-documented, the mechanisms behind sex chromosome turnover remain unclear. By examining population data from wild-caught samples and genetic crosses, we find evidence of a recent turnover in both the sex determination gene and the sex chromosome in *A.*

quadracus. Furthermore, there are polymorphic inversions on the X and Y chromosomes, with relatively little degeneration on the Y chromosomes. This within-species variation on the *A. quadracus* sex chromosomes provides an opportunity for further studies to determine the role of evolutionary forces such as drift and sexually-antagonistic selection in driving sex chromosome evolution and turnover.

Material and Methods

Ethics statement

All experiments involving animals were approved by the Veterinary Service of the Department of Agriculture and Nature of the Canton of Bern (VTHa# BE4/16, BE17/17 and BE127/17).

Sample collections and genetic crosses

For wild populations, *A. quadracus* were collected from the following localities: Canal Lake (44.49830, -63.90205) in Nova Scotia (NS), Canada in 2021 by Anne Dalziel; Demarest Lloyd State Park (41.5289936, -70.9833719) in Massachusetts (MA), USA in 2007 by Catherine Peichel; and West River Memorial Park (41.314148, -72.956544) in Connecticut (CT), USA in 2021 by Natalie Steinel and Daniel Bolnick (Fig 1). The sex of each individual was identified by dissection of the gonads, and a fin clip was sampled and preserved in 95% ethanol for DNA extraction and sequencing.

Genetic crosses were also made from the same populations, with the wild parents of the crosses collected from: Canal Lake (44.49830, -63.90205) in Nova Scotia (NS), Canada in 2019 by Anna Dalziel; Demarest Lloyd State Park (41.5289936, -70.9833719) in Massachusetts (MA), USA in 2007 by Catherine Peichel; West River Memorial Park (41.314148, -72.956544) in Connecticut (CT), USA in 2009 by Thomas Near. For each population, a single cross was generated using a single female and a single male. The sex of each F1 offspring was identified by dissection of the gonads, and a fin clip was sampled and preserved in ethanol for DNA extraction and sequencing. For the NS cross, brains were also dissected from 12 males and 12 females from the F1 offspring as well as from the male and female F0 parents used for crossing for further RNA-seq analysis. Total numbers of individuals sequenced for each population and cross are provided in Supplementary Table S6.

Note that a previous cytogenetic study of the same MA population used here suggested

it had a ZW sex chromosome [7], while a previous cytogenetic study of the same CT population used here did not identify any heteromorphic sex chromosome pair [35].

DNA and RNA extraction and sequencing libraries

For all samples, DNA was extracted by phenol-chloroform extraction, followed by ethanol precipitation. Total brain RNA from the F0 parents and F1 offspring of the NS cross was extracted using Trizol (Life Technologies, Carlsbad, California, USA) following the manufacturer's instructions.

For wild-caught populations, genomic DNA was extracted from fin clips as described. Multiplexed haplotagging libraries were prepared as described in [68] with the following modifications in WASH buffer volumes, Tn5 stripping, subsampling and exonuclease reaction. Briefly, DNA were processed in batches of 96 samples. For each sample, 0.75 ng input DNA at 0.15 ng/µl concentration were mixed with 2.5µl haplotagging beads resuspended in 20µl of WASH buffer (20 mM Tris pH8, 50 mM NaCl, 0.1% Triton X-100). We reduced the volume of the tagmentation reaction by using only 5µl of 5x tagmentation buffer (50 mM TAPS pH 8.5 with NaOH, 25 mM MgCl2, 50% N,N-dimethylformamide) and 15 µl of 0.6% SDS for Tn5 stripping following tagmentation. Next, the samples were pooled with 1/3 bead subsampling. This corresponds to a final input DNA of 0.25 ng per sample.

With only 8 pooled samples on the magnetic stand, the buffer was removed, and 20 µl of 1x Lambda Exonuclease buffer, supplemented with 10 units of Exonuclease I (M0293L, New England BioLabs), was added to each sample. Samples were incubated at 48 °C for 20 minutes, and then washed twice for 5 minutes with 150 µl of WASH buffer. DNA library was then amplified using NEBNext® High-Fidelity 2X PCR Master Mix (M0541L, New England BioLabs) in eight 50 µl PCR reaction according to manufacturer's instructions, using 3 µl of 10 µM TruSeq-F AATGATACGGCGACCACCGAGATCTACAC and TruSeq-R CAAGCAGAAGACGGCATACGAGAT primers, with the following cycling conditions: 10 min at 72°C followed by 30 sec 98°C and 10 cycles of: 98°C for 15 sec, 65°C for 30 sec and 72°C for 60 sec. Libraries were pooled after PCR into a single library pool, size selected using 0.9x volume of Ampure magnetic beads (Beckman Coulter), Qubit quantified, followed by a second size selection with 0.45x and 0.85x volume of Ampure magnetic beads, to remove library longer than 800 bp and smaller than 300 bp, respectively. Pooled libraries were sequenced on a whole S4 lane of Novaseq 6000 (Illumina) instrument with a 151+13+13+151 cycle run setting,

such that the run produced 13 and 13 nt in the i7 and i5 index reads, respectively. Sequence data were first converted into fastq format using --create-fastq-for-index-reads using the bcl2fastq program (Illumina). Then we performed beadTag demultiplexing to generate the modified fastq files using a custom demult_fastq program, resulting in a fastq file supplemented with molecular and sample barcode in the header of each read (e.g. BX:Z:A01C02B03D04). This program is available at https://github.com/evolgenomics/haplotagging.

For pool-sequencing of F1 offspring of genetic crosses and DNA-sequencing of F0 parents, sequencing libraries were created by standard Illumina DNA TruSeq kits. For the RNA-sequencing of F0 parents and F1 offspring from the NS cross, libraries were prepared with the Illumina mRNA TruSeq kit. All libraries were subject to 150bp paired-end sequencing on Illumina NovaSeq SP flow cells by the University of Bern Next Generation Sequencing Platform.

Short read data processing and SNP calling

All raw reads were trimmed by Trimmomatic (v 0.36) [69] with a sliding window of 4 bp. The first 13 bp of all reads were dropped, and windows with an average quality score below 15 were also dropped.

For DNA linked-reads sequencing from wild populations, trimmed reads were first mapped to the latest *A. quadracus* female assembly [36] by EMA [70], and remaining unmapped reads were further mapped by BWA (v 0.7.11) [71]. Bam files were sorted and duplicates were removed by Picard 2.0.1 (http://broadinstitute.github.io/picard). SNP calling was done by ANGSD 0.9.7 [72] with the GATK algorithm. Vcftools 0.1.16 [73] was used to further filter the SNP matrix with the following criteria: (1) individuals witha mean coverage lower than 6; (2) the population mean depth coverage at the SNP was less than 4x or greater than 40x; (3) the proportion of missing data at the SNP was greater than 0.2 in either the CT population or the NS population; (4) the minor allele frequency of the SNP was less than 0.05. The MA population had poor sequencing quality (likely due to the age of the samples) and was therefore not used for SNP filtering, in order to rescue as much information as possible from this population.

For Pool-seq reads from genetic crosses, trimmed reads were first mapped to the latest assembly by BWA (v 0.7.11), and sorted with duplicates removed by Picard 2.0.1. Pooplation2 [74] was used to create a sync file containing all the variants for each cross separately.

For RNA-seq reads from the NS genetic cross, trimmed reads were mapped to the assembly by STAR 2.7.10 [75], and merged by Samtools v1.15 [76].

Identification of the sex determination system and sex chromosome in *A. quadracus*

The sex determination system and sex chromosome were identified in *A. quadracus* using multiple lines of evidence. Using mosdepth 0.3.3 [77] with a sliding window of 20kb and a step size of 10kb, sequencing depth was calculated for both linked read data from wild populations and Pool-Seq data from genetic crosses. For the wild populations, Fst, Pi, and SNP density were calculated by VCFtools 0.1.16 [73] with a sliding window of the same size. For the genetic crosses, PoPoolation1 [78] was used for calculating Pi, and PoPoolation2 [74] was used for calculating Fst. PSASS 3.0.1 [79] was used to calculate sex-specific SNPs in each genetic cross. To further confirm the sex determination pattern, RNA-seq data from the parents and offspring of the NS cross was fed into read2snp 2.0 [80] to obtain SNP data, and also fed into Trinity 2.11.2 [81] to obtain a de novo transcriptome assembly. The output SNP array and assembly were then processed by SEX-DETector [37].

Identification of population-specific inversions

To determine whether there are inversions on the sex chromosome, three methods were used. First, the linked-read sequences were used to identify shared barcodes among any pairs of windows of 10kb on each chromosome by LRez v2.2.3 [82]. Windows with shared barcodes were divided into two categories: windows that are adjacent, and windows that are 500kb apart on the same chromosome. Putative inversions were identified based on the number of shared barcodes between 500 kb apart non-adjacent window pairs. Second, inversions on the sex chromosomes were identified by LEVIATHAN V1.0.2 [83]. Third, screening of bam files for split and discordantly mapped read pairs near the breakpoints of inversion was done by IGV 2.14.1 [84]. Genotypes of inversions were determined by the divergence and diversity pattern between sexes within the inversion as well as the SNP density plot generated by VCFtools 0.1.16 at the individual level. The above analyses were not conducted in the MA population due to the poor sequencing quality.

Pattern of molecular evolution within inversions

Due to the poor sequencing quality, the MA population was not used in the following

analyses of the linked-read data from the wild populations. First, fixed male-specific SNPs were identified in each population separately following the rules: (1) SNPs only exist in males; and (2) the allele frequency is higher than 0.45. Second, dS values between X and Y-linked alleles for genes were calculated. To phase the X and Y-linked alleles, females with homozygous inversions and males with heterozygous inversions were selected for the CT population. The NS population was composed of females without inversions and males with heterozygous inversions. Phasing of SNPs was first done by Hapcut2 [85], then by WhatsHap v1.1 [86,87]. Genes with more than three SNPs between X and Y copies were retained. dS values between X and Y-linked alleles for genes were compared using KaKs Calculator 3 [88].

Estimating degeneration on sex chromosome

Degeneration of the Y chromosome usually appears in two forms: (1) the loss of genes either due to complete deletion or degeneration such that the Y allele can no longer be aligned to the X allele; or (2) accumulation of premature stop codons. To identify the genes that degenerated on the *A. quadracus* sex chromosomes, we calculated male to female read-depth ratios for each wild population by mosdepth. Further, mutations that cause premature stop codons were identified by snpEff v5.1 [89] separately for each sex and each population. SNPs with accumulated stop codon were identified with an allele frequeny greater than 0.9 in females or 0.45 in males. The above analyses were not conducted on the MA population.

Identification of potential sex determination genes

Because the fourspine stickleback genome assembly was obtained from a female individual [36], a kmer-based approach was employed to extract male-specific sequences. KmerGo [90] was first run on the Pool-seq data from each cross separately. Shared male-specific kmers were identified, and corresponding reads were extracted from the raw dataset by BBmap [91]. Then raw male-specific reads were assembled into short contigs by ABYSS 2.0 [92], and genes shared by all three crosses were identified by Blast 2.11.0 [93]. The prediction of the effect of SNPs in coding regions that were differentiated between males and females were done by PROVEAN [94] after target genes were identified. Male-specific SNPs of the candidate genes were identified separately in each population.

Data availability

All sequences are uploaded and available on the NCBI Sequence Read Archive (Supplementary Table S6).

Acknowledgments

We thank Daniel Bolnick, Anne Dalziel, Thomas Near, and Natalie Steinel for collecting wild fish, Melanie Hiltbrunner, Shaugnessy McCann, and Verena Saladin for making crosses, Melanie Hiltbrunner for extractions, and the University of Bern Next Generation Sequencing Platform for library preparation and sequencing. This research was funded by Swiss National Science Foundation grant 31003A_176130 to C.L.P.

References

1. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, et al. Sex Determination: Why So Many Ways of Doing It? PLoS Biol. 2014;12: e1001899. doi:10.1371/journal.pbio.1001899

2. Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, Borzée A, et al. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nat Commun. 2018;9: 4088. doi:10.1038/s41467-018-06517-2

3. Ma WJ, Veltsos P. The Diversity and Evolution of Sex Chromosomes in Frogs. Genes. 2021;12: 483. doi:10.3390/genes12040483

4. Gamble T, Coryell J, Ezaz T, Lynch J, Scantlebury DP, Zarkower D. Restriction Site-Associated DNA Sequencing (RAD-seq) Reveals an Extraordinary Number of Transitions among Gecko Sex-Determining Systems. Mol Biol Evol. 2015;32: 1296–1309. doi:10.1093/molbev/msv023

5. Darolti I, Wright AE, Sandkam BA, Morris J, Bloch NI, Farré M, et al. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. Proc Natl Acad Sci. 2019;116: 19031–19036. doi:10.1073/pnas.1905298116

6. El Taher A, Ronco F, Matschiner M, Salzburger W, Böhne A. Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. Sci Adv. 2021;7: eabe8215. doi:10.1126/sciadv.abe8215

7. Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). PLoS Genet. 2009;5: e1000391. doi:10.1371/journal.pgen.1000391

8. The Tree of Sex Consortium, Ashman T-L, Bachtrog D, Blackmon H, Goldberg EE, Hahn MW, et al. Tree of Sex: A database of sexual systems. Sci Data. 2014;1: 140015. doi:10.1038/sdata.2014.15

9. Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, et al. Sex Chromosome Evolution: So Many Exceptions to the Rules. Genome Biol Evol. 2020;12: 750–763. doi:10.1093/gbe/evaa081

10. Vicoso B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat Ecol Evol. 2019;3: 1632–1641. doi:10.1038/s41559-019-1050-8

11. Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. 2013;14: 113–124. doi:10.1038/nrg3366

12. Charlesworth B. The evolution of sex chromosomes. Science. 1991;251: 1030–1033. doi:10.1126/science.1998119

13. van Doorn GS, Kirkpatrick M. Turnover of sex chromosomes induced by sexual conflict. Nature. 2007;449: 909–912. doi:10.1038/nature06178

14. Blaser O, Grossen C, Neuenschwander S, Perrin N. Sex-chromosome turnovers induced by deleterious mutation load. Evolution. 2013;67: 635–645. doi:10.1111/j.1558-5646.2012.01810.x

15. Blaser O, Neuenschwander S, Perrin N. Sex-chromosome turnovers: the hot-potato model. Am Nat. 2014;183: 140–146. doi:10.1086/674026

16. Bull JJ, Charnov EL. Changes in the heterogametic mechanism of sex determination. Heredity. 1977;39: 1–14. doi:10.1038/hdy.1977.38

17. Werren JH, Beukeboom LW. Sex determination, sex ratios, and genetic conflict. Annu Rev Ecol Syst.

1998;29: 233–261. doi:10.1146/annurev.ecolsys.29.1.233

18. Saunders PA, Neuenschwander S, Perrin N. Sex chromosome turnovers and genetic drift: a simulation study. J Evol Biol. 2018;31: 1413–1419. doi:10.1111/jeb.13336

19. Veller C, Muralidhar P, Constable GWA, Nowak MA. Drift-Induced Selection Between Male and Female Heterogamety. Genetics. 2017;207: 711–727. doi:10.1534/genetics.117.300151

20. Evans BJ, Mudd AB, Bredeson JV, Furman BLS, Wasonga DV, Lyons JB, et al. New insights into *Xenopus* sex chromosome genomics from the Marsabit clawed frog *X. borealis*. J Evol Biol. 2022;35: 1777–1790. doi:10.1111/jeb.14078

21. Furman BLS, Cauret CMS, Knytl M, Song XY, Premachandra T, Ofori-Boateng C, et al. A frog with three sex chromosomes that co-mingle together in nature: Xenopus tropicalis has a degenerate W and a Y that evolved from a Z chromosome. PLoS Genet. 2020;16: e1009121. doi:10.1371/journal.pgen.1009121

22. Rodrigues N, Vuille Y, Loman J, Perrin N. Sex-chromosome differentiation and 'sex races' in the common frog (*Rana temporaria*). Proc R Soc B Biol Sci. 2015;282: 20142726. doi:10.1098/rspb.2014.2726

23. Feller AF, Ogi V, Seehausen O, Meier JI. Identification of a novel sex determining chromosome in cichlid fishes that acts as XY or ZW in different lineages. Hydrobiologia. 2021;848: 3727–3745. doi:10.1007/s10750-021-04560-7

24. Kocher TD, Behrens KA, Conte MA, Aibara M, Mrosso HDJ, Green ECJ, et al. New Sex Chromosomes in Lake Victoria Cichlid Fishes (Cichlidae: Haplochromini). Genes. 2022;13: 804. doi:10.3390/genes13050804

25. Schultheis C, Böhne A, Schartl M, Volff JN, Galiana-Arnoux D. Sex Determination Diversity and Sex Chromosome Evolution in Poeciliid Fish. Sex Dev. 2009;3: 68–77. doi:10.1159/000223072

26. Roberts RB, Ser JR, Kocher TD. Sexual Conflict Resolved by Invasion of a Novel Sex Determiner in Lake Malawi Cichlid Fishes. Science. 2009;326: 998–1001. doi:10.1126/science.1174705

27. Jeffries DL, Mee JA, Peichel CL. Identification of a candidate sex determination gene in *Culaea inconstans* suggests convergent recruitment of an *Amh* duplicate in two lineages of stickleback. J Evol Biol. 2022; jeb.14034. doi:10.1111/jeb.14034

28. Dagilis AJ, Sardell JM, Josephson MP, Su Y, Kirkpatrick M, Peichel CL. Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. Philos Trans R Soc B Biol Sci. 2022;377: 20210205. doi:10.1098/rstb.2021.0205

29. Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, et al. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol. 2020;21: 177. doi:10.1186/s13059-020-02097-x

30. Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, et al. A role for a neo-sex chromosome in stickleback speciation. Nature. 2009;461: 1079–1083. doi:10.1038/nature08441

31. Sardell JM, Josephson MP, Dalziel AC, Peichel CL, Kirkpatrick M. Heterogeneous Histories of Recombination Suppression on Stickleback Sex Chromosomes. Mol Biol Evol. 2021;38: 4403–4418. doi:10.1093/molbev/msab179

32. Natri HM, Merilä J, Shikano T. The evolution of sex determination associated with a chromosomal

inversion. Nat Commun. 2019;10: 145. doi:10.1038/s41467-018-08014-y

33. Rastas P, Calboli FCF, Guo B, Shikano T, Merilä J. Construction of Ultradense Linkage Maps with Lep-MAP2: Stickleback F $_2$ Recombinant Crosses as an Example. Genome Biol Evol. 2016;8: 78–93. doi:10.1093/gbe/evv250

34. Chen TR, Reisman HM. A comparative chromosome study of the North American species of sticklebacks (Teleostei: Gasterosteidae). Cytogenet Genome Res. 1970;9: 321–332. doi:10.1159/000130102

35. Urton JR, McCann SR, Peichel CL. Karyotype Differentiation between Two Stickleback Species (Gasterosteidae). Cytogenet Genome Res. 2011;135: 150–159. doi:10.1159/000331232

36. Liu Z, Roesti M, Marques D, Hiltbrunner M, Saladin V, Peichel CL. Chromosomal Fusions Facilitate Adaptation to Divergent Environments in Threespine Stickleback. Mol Biol Evol. 2022;39: msab358. doi:10.1093/molbev/msab358

37. Muyle A, Käfer J, Zemp N, Mousset S, Picard F, Marais GA. SEX-DETector: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. Genome Biol Evol. 2016;8: 2530–2543. doi:10.1093/gbe/evw172

38. Bergero R, Charlesworth D. The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol. 2009;24: 94–102. doi:10.1016/j.tree.2008.09.010

39. Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. Heredity. 2005;95: 118–128. doi:10.1038/sj.hdy.6800697

40. Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J, et al. The Master Sex-Determination Locus in Threespine Sticklebacks Is on a Nascent Y Chromosome. Curr Biol. 2004;14: 1416–1424. doi:10.1016/j.cub.2004.08.030

41. Dixon G, Kitano J, Kirkpatrick M. The Origin of a New Sex Chromosome by Introgression between Two Stickleback Fishes. Mol Biol Evol. 2019;36: 28–38. doi:10.1093/molbev/msy181

42. Arai T, Ueno D, Kitamura T, Goto A. Habitat preference and diverse migration in threespine sticklebacks, *Gasterosteus aculeatus* and *G. nipponicus*. Sci Rep. 2020;10: 14311. doi:10.1038/s41598-020-71400-4

43. Overbeek PA, Gorlov IP, Sutherland RW, Houston JB, Harrison WR, Boettger-Tong HL, et al. A transgenic insertion causing cryptorchidism in mice. genesis. 2001;30: 26–35. doi:10.1002/gene.1029

44. Kasippillai T, MacArthur DG, Kirby A, Thomas B, Lambalk CB, Daly MJ, et al. Mutations in *eIF4ENIF1* Are Associated With Primary Ovarian Insufficiency. J Clin Endocrinol Metab. 2013;98: E1534–E1539. doi:10.1210/jc.2013-1102

45. Kumagai J, Hsu SY, Matsumi H, Roh JS, Fu P, Wade JD, et al. INSL3/Leydig Insulin-like Peptide Activates the LGR8 Receptor Important in Testis Descent. J Biol Chem. 2002;277: 31283–31286. doi:10.1074/jbc.C200398200

46. Timms KL. Rodent Models for Translational Research in Endometriosis. Biol Reprod. 2012;87: 143–143. doi:10.1093/biolreprod/87.s1.143

47. Zimmermann S, Steding G, Emmen JMA, Brinkmann AO, Nayernia K, Holstein AF, et al. Targeted Disruption of the Insl3 Gene Causes Bilateral Cryptorchidism. Mol Endocrinol. 1999;13: 681–691. doi:10.1210/mend.13.5.0272

48. Sharma V, Lehmann T, Stuckas H, Funke L, Hiller M. Loss of RXFP2 and INSL3 genes in Afrotheria shows that testicular descent is the ancestral condition in placental mammals. PLoS Biol. 2018;16: e2005293. doi:10.1371/journal.pbio.2005293

49. Assis LHC, Crespo D, Morais RDVS, França LR, Bogerd J, Schulz RW. INSL3 stimulates spermatogonial differentiation in testis of adult zebrafish (*Danio rerio*). Cell Tissue Res. 2016;363: 579–588. doi:10.1007/s00441-015-2213-9

50. Wu X, Viveiros MM, Eppig JJ, Bai Y, Fitzpatrick SL, Matzuk MM. Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. Nat Genet. 2003;33: 187–191. doi:10.1038/ng1079

51. Yamamoto TM, Cook JM, Kotter CV, Khat T, Silva KD, Ferreyros M, et al. Zar1 represses translation in Xenopus oocytes and binds to the TCS in maternal mRNAs with different characteristics than Zar2. Biochim Biophys Acta BBA - Gene Regul Mech. 2013;1829: 1034–1046. doi:10.1016/j.bbagrm.2013.06.001

52. Miao L, Yuan Y, Cheng F, Fang J, Zhou F, Ma W, et al. Translation repression by maternal RNA binding protein zar1 is essential for early oogenesis in zebrafish. Development. 2016; dev.144642. doi:10.1242/dev.144642

53. Cook J, Charlesworth A. The Developmentally Important RNA-binding Protein, Zygote arrest (Zar), Regulates mRNA Translation. FASEB J. 2015;29. doi:10.1096/fasebj.29.1_supplement.711.18

54. Guzeloglu-Kayisli O, Lalioti MD, Aydiner F, Sasson I, Ilbay O, Sakkas D, et al. Embryonic poly(A)-binding protein (EPAB) is required for oocyte maturation and female fertility in mice. Biochem J. 2012;446: 47–58. doi:10.1042/BJ20120467

55. Pan Q, Kay T, Depincé A, Adolfi M, Schartl M, Guiguen Y, et al. Evolution of master sex determiners: TGF- signalling pathways at regulatory crossroads. Philos Trans R Soc B Biol Sci. 2021;376: 20200091. doi:10.1098/rstb.2020.0091

56. Charlesworth B, Charlesworth D. A Model for the Evolution of Dioecy and Gynodioecy. Am Nat. 1978;112: 975–997.

57. Charlesworth D, Charlesworth B. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genet Res. 1980;35: 205–214. doi:10.1017/S0016672300014051

58. Fisher RA. The evolution of dominance. Biol Rev. 1931; 345–368.

59. Rice WR. The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. Evolution. 1987;41: 911. doi:10.2307/2408899

60. Úbeda F, Patten MM, Wild G. On the origin of sex chromosomes from meiotic drive. Proc R Soc B Biol Sci. 2015;282: 20141932. doi:10.1098/rspb.2014.1932

61. Lenormand T, Roze D. Y recombination arrest and degeneration in the absence of sexual dimorphism. Science. 2022;375: 663–666. doi:10.1126/science.abj1813

62. Charlesworth B, Wall JD. Inbreeding, heterozygote advantage and the evolution of neo–X and neo–Y sex chromosomes. Proc. Royal Soc. B. 1999;266: 51–56. doi:10.1098/rspb.1999.0603

63. Jay P, Tezenas E, Véber A, Giraud T. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. PLoS Biol.

2022;20: e3001698. doi:10.1371/journal.pbio.3001698

64. Jeffries DL, Gerchen JF, Scharmann M, Pannell JR. A neutral model for the loss of recombination on sex chromosomes. Philos Trans R Soc B Biol Sci. 2021;376: 20200096. doi:10.1098/rstb.2020.0096

65. Charlesworth B, Coyne JA, Barton NH. The Relative Rates of Evolution of Sex Chromosomes and Autosomes. Am Nat. 1987;130: 113–146.

66. Ponnikas S, Sigeman H, Abbott JK, Hansson B. Why Do Sex Chromosomes Stop Recombining? Trends Genet. 2018;34: 492–503. doi:10.1016/j.tig.2018.04.001

67. Lahn BT, Page DC. Four Evolutionary Strata on the Human X Chromosome. Science. 1999;286: 964–967. doi:10.1126/science.286.5441.964

68. e er a a ar a a e r a et a Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. Proc Natl Acad Sci. 2021;118: e2015005118. doi:10.1073/pnas.2015005118

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

70. ha ana I, Berger B. Latent variable model for aligning barcoded short-reads improves downstream analyses. Bioinformatics; 2017 Nov. doi:10.1101/220236

71. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997. 2013 Mar 16.

72. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014;15: 356. doi:10.1186/s12859-014-0356-4

73. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330

74. Kofler R, Pandey RV, Schlotterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011;27: 3435–3436. doi:10.1093/bioinformatics/btr589

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

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

77. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 2018;34: 867–868. doi:10.1093/bioinformatics/btx699

78. Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, et al. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. PLoS ONE. 2011;6: e15925. doi:10.1371/journal.pone.0015925

79. Feron R. RomainFeron/PSASS: Psass 3.0.1. Zenodo; 2020. doi:10.5281/ZENODO.3702337

80. Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, et al. Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. PLoS Genet. 2013;9: e1003457. doi:10.1371/journal.pgen.1003457

81. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8: 1494–1512. doi:10.1038/nprot.2013.084

82. Morisse P, Lemaitre C, Legeai F. LRez: a C++ API and toolkit for analyzing and managing Linked-Reads data. Bioinforma Adv. 2021;1: vbab022. doi:10.1093/bioadv/vbab022

83. Morisse P, Legeai F, Lemaitre C. LEVIATHAN: efficient discovery of large structural variants by leveraging long-range information from Linked-Reads data. Bioinformatics; 2021 Mar. doi:10.1101/2021.03.25.437002

84. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14: 178–192. doi:10.1093/bib/bbs017

85. Edge P, Bafna V, Bansal V. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res. 2017;27: 801–812. doi:10.1101/gr.213462.116

86. Martin M, Patterson M, Garg S, O Fischer S, Pisanti N, Klau GW, et al. WhatsHap: fast and accurate read-based phasing. Bioinformatics; 2016 Nov. doi:10.1101/085050

87. Patterson M, Marschall T, Pisanti N, van Iersel L, Stougie L, Klau GW, et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. J Comput Biol. 2015;22: 498–509. doi:10.1089/cmb.2014.0157

88. Zhang Z. KaKs_calculator 3.0: Calculating selective pressure on coding and non-coding sequences. Genomics Proteomics Bioinformatics. 2022; S167202292100259X. doi:10.1016/j.gpb.2021.12.002

89. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6: 80–92. doi:10.4161/fly.19695

90. Wang Y, Chen Q, Deng C, Zheng Y, Sun F. KmerGO: A Tool to Identify Group-Specific Sequences With k-mers. Front Microbiol. 2020;11: 2067. doi:10.3389/fmicb.2020.02067

91. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014. Available: https://www.osti.gov/biblio/1241166

92. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 2017;27: 768–777. doi:10.1101/gr.214346.116

93. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421

94. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31: 2745–2747. doi:10.1093/bioinformatics/btv195

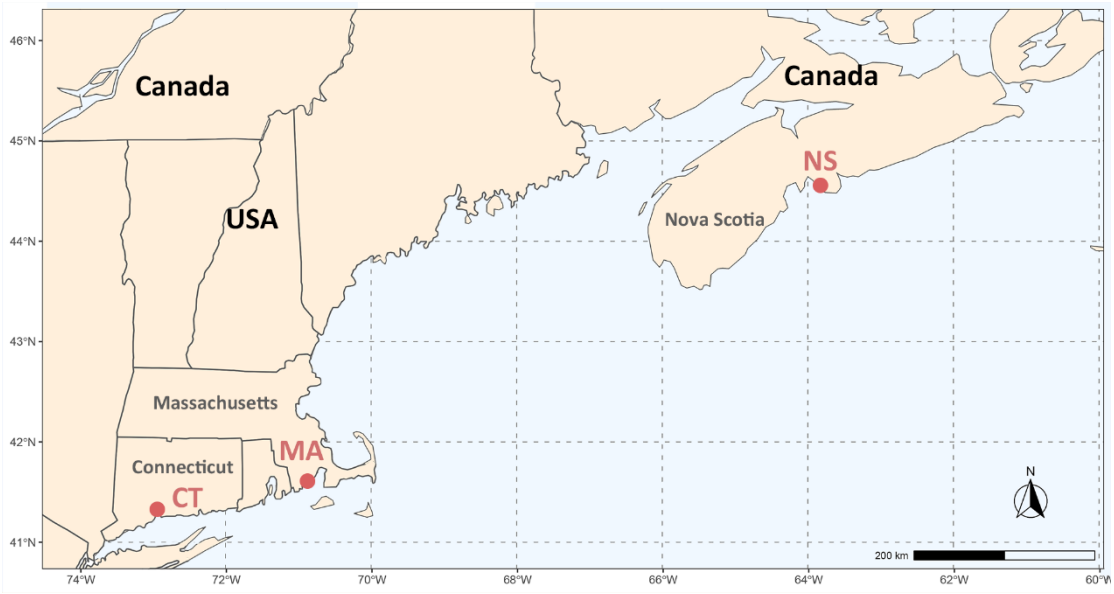


Fig 1. Fourspine stickleback sampling locations in this study. Red dots represent sampled populations.

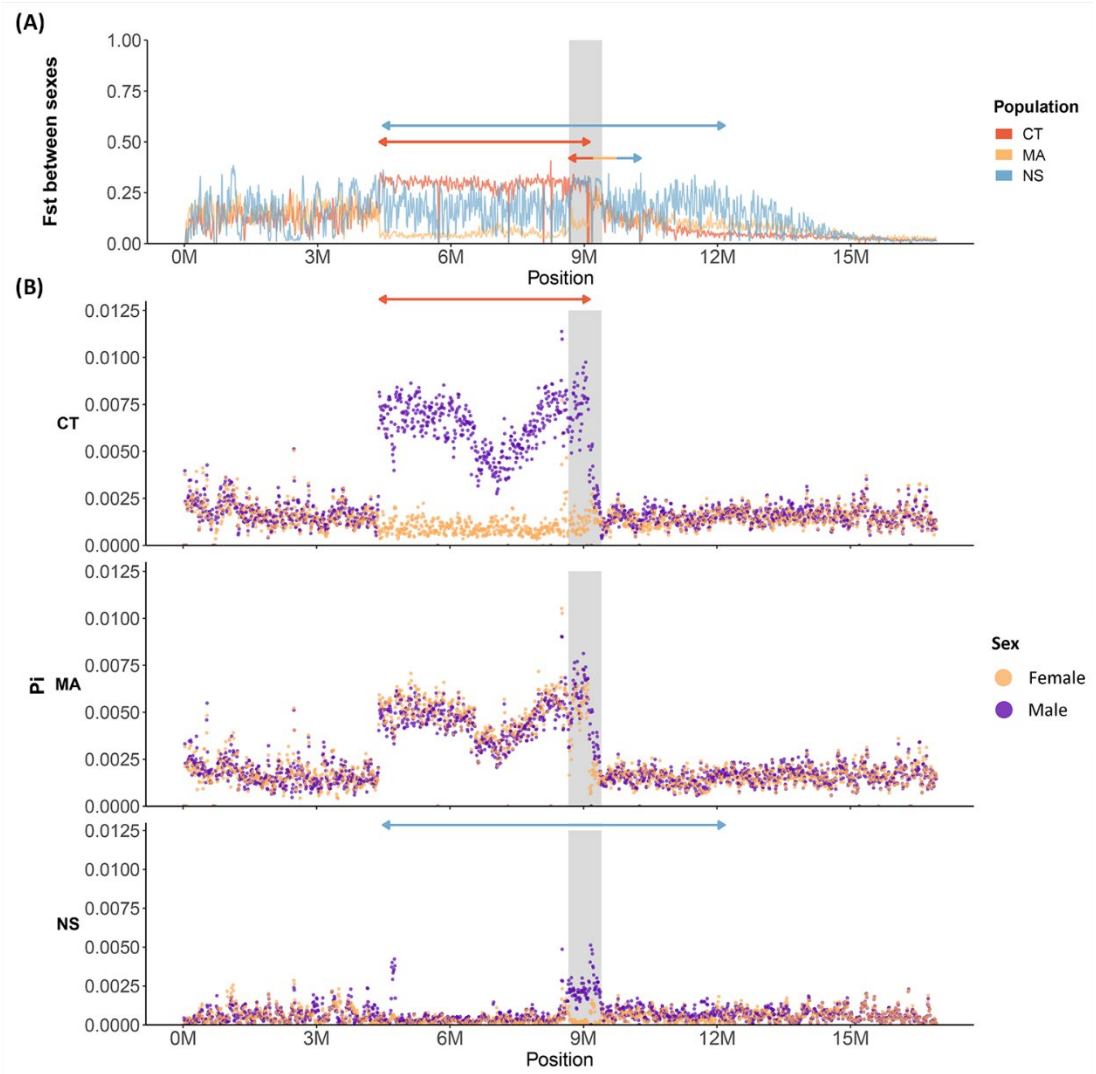


Fig 2. Patterns of genetic differentiation and diversity on Chr23 in genetic crosses. A) Genetic differentiation (Fst) between males and females on Chr23 calculated from pool-seq data from three crosses, with the Connecticut (CT) cross in coral, the Massachusetts (MA) cross in yellow, and the Nova Scotia (NS) cross in light blue. The coral and light blue arrows represent the corresponding inversions in the CT and NS populations. The arrow with three colors represents a shared inversion among all populations. B) Distribution of genetic diversity (Pi) within males (purple dots) and females (yellow dots) in 20kb sliding windows on Chr23 calculated from pool-seq data from the three crosses. The grey region represents the shared sex determination region, which was identified in the sequencing data from wild individuals from the same three populations used for these crosses. The coral and light blue arrows represent the corresponding inversions in the CT and NS populations.

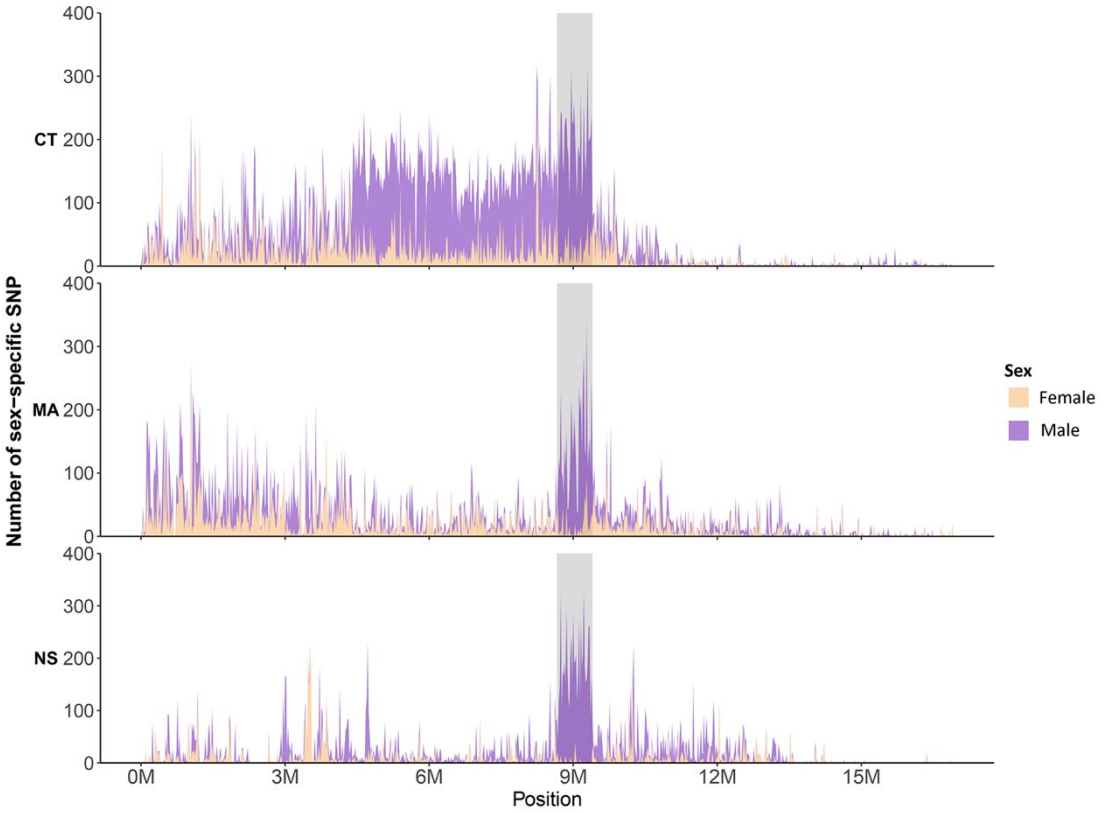


Fig 3. Distribution of sex-specific SNPs on Chr23 in genetic crosses derived from the Connecticut (CT), Massachusetts (MA), and Nova Scotia (NS) populations. A sliding window of 20kb and a step size of 10kb was used. Purple bars represent the number of male-specific SNPs, and yellow bars represent the number of female-specific SNPs in each population. A shared region around 9Mb in the sex determination regions has an increase in male-specific SNPs in all crosses (the grey region), consistent with the shared sex-determination region identified in the sequencing data from wild individuals from the same three populations used for these crosses.

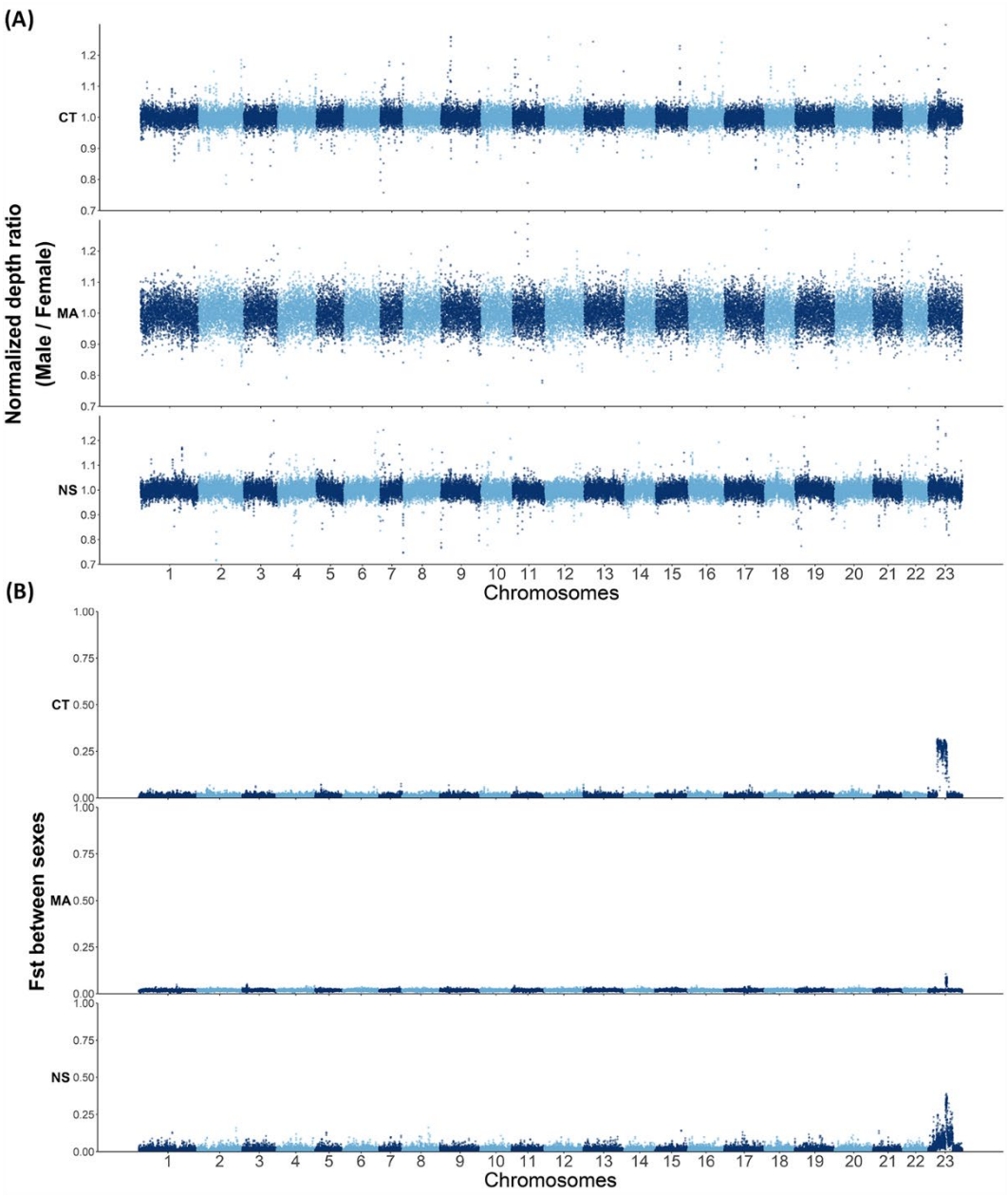


Fig 4. Genomic distributions of normalized male-to-female depth ratio (A) and genetic differentiation (Fst) between males and females (B) were calculated using linked-read sequences from three wild populations from Connecticut (CT), Massachusetts (MA), and Nova Scotia (NS). A sliding window of 20kb and a step size of 10kb was used. Alternating colors in each panel are used to highlight the different chromosomes.

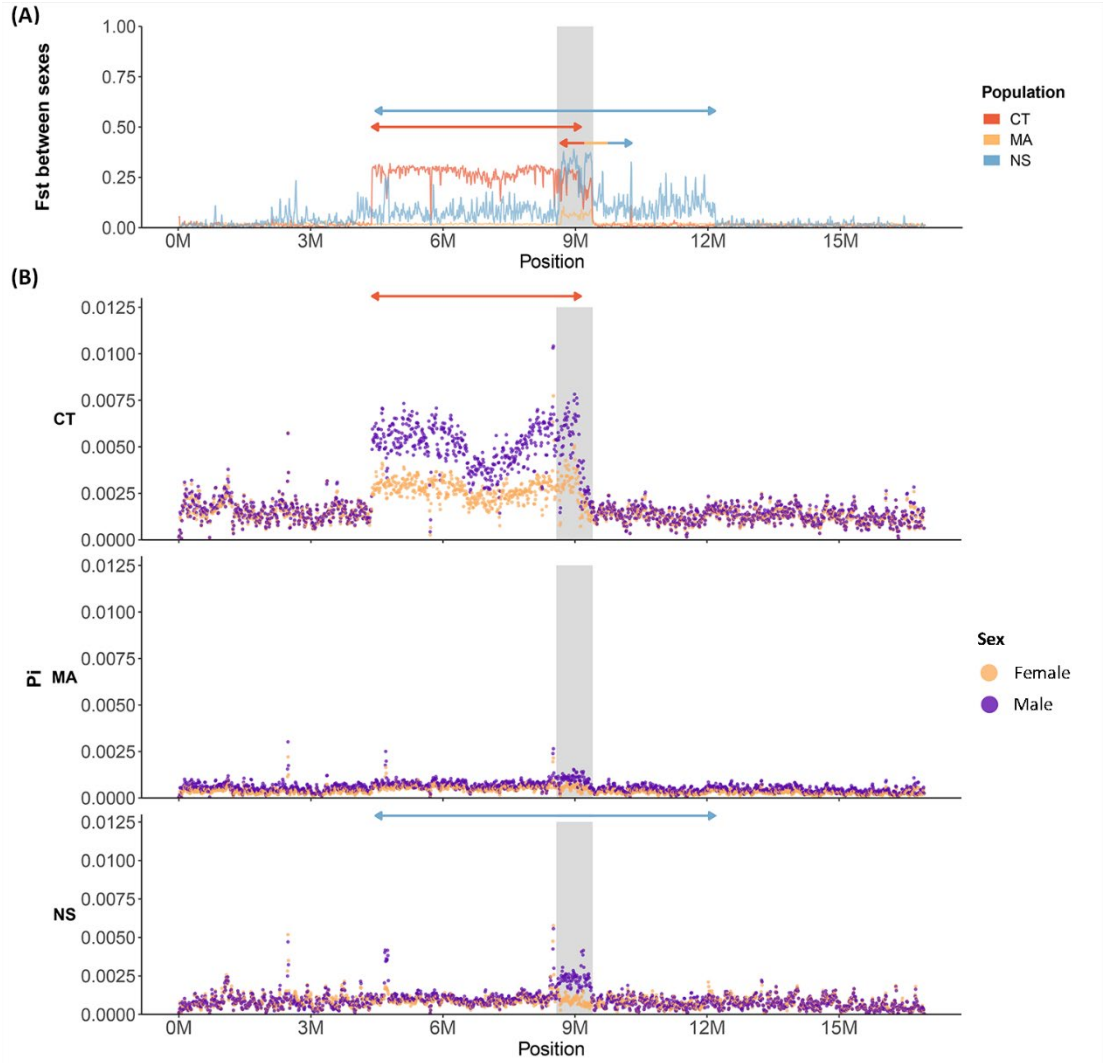


Fig 5. Patterns of genetic differentiation and diversity on Chr23 in wild populations. A) Genetic differentiation (Fst) between males and females on Chr23 was calculated with linked-read sequencing data from three wild populations, with the Connecticut (CT) population in coral, the Massachusetts (MA) population in yellow, and the Nova Scotia (NS) population in light blue. The coral and light blue arrows represent the corresponding inversions in CT and NS populations. The arrow with the three colors represents a shared inversion among all populations. B) Distribution of genetic diversity (Pi) with males (purple dots) and females (yellow dots) on Chr23 calculated using linked-read sequencing data from the three wild populations. The grey region represents the shared sex-determination region across the three populations. The coral and light blue arrows represent the corresponding inversions in the CT and NS populations.

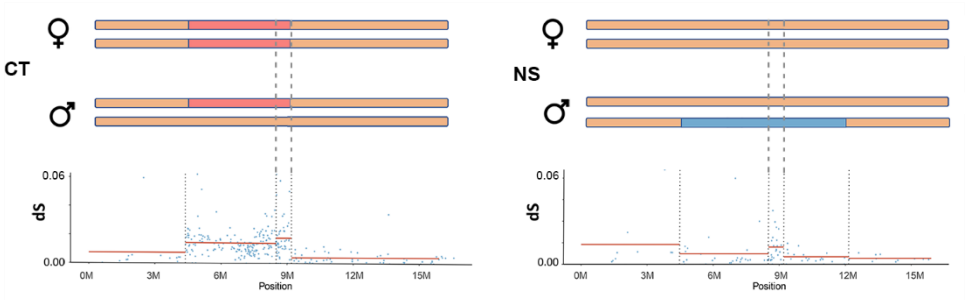


Fig 6. Model for population-specific inversions on the *A. quadracus* sex chromosomes. Orange bars represent the sex chromosome pair on Chr23. Coral bars show positions of the X-specific inversion in both sexes in the CT population, and the light blue bar shows the position of the Y-specific inversion in males in the NS population. Grey dashed lines indicated the sex determination region. The distributions of dS values between X and Y on Chr23 in each population. Blue dots are dS values of corresponding genes, and red lines represent the average values. Dotted lines are borders of regions where dS values change dramatically.

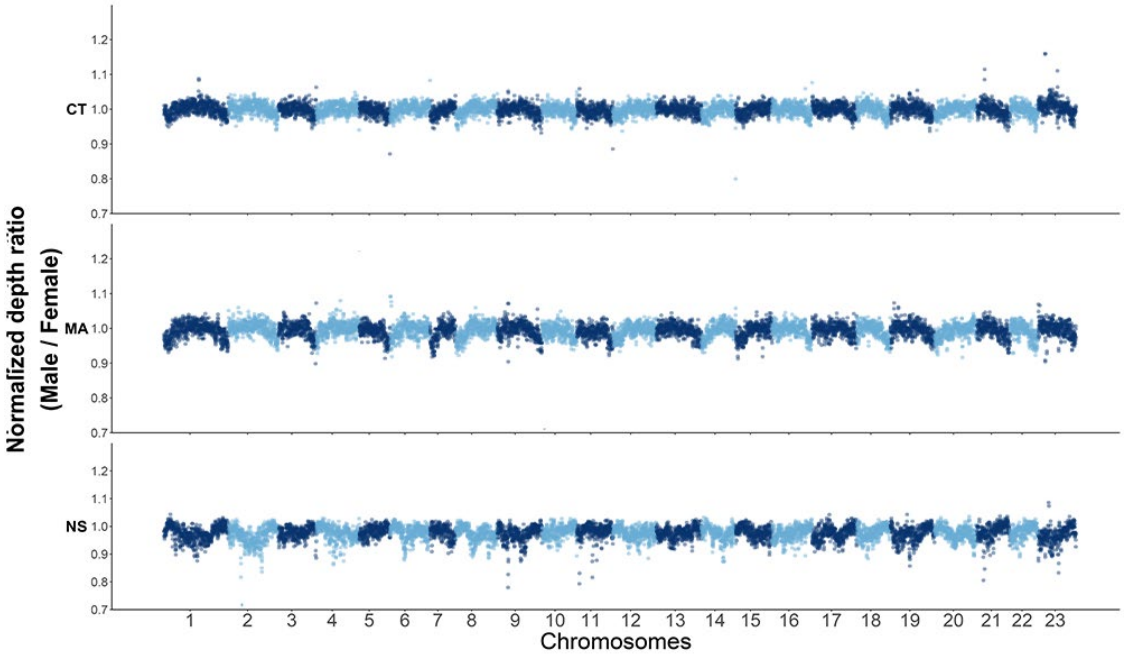


Fig S1. Male-to-female depth ratio across the genome with pool-seq data in genetic crosses from three populations (CT, MA, and NS). Raw depth values were normalized to eliminate the difference between two sexes. The size of the sliding window is 20 kb and the step size is 10kb. Chromosomes are indicated on the X-axis, and the normalized depth ratio is shown on the Y-axis. Dark and light blue regions indicate the different chromosomes.

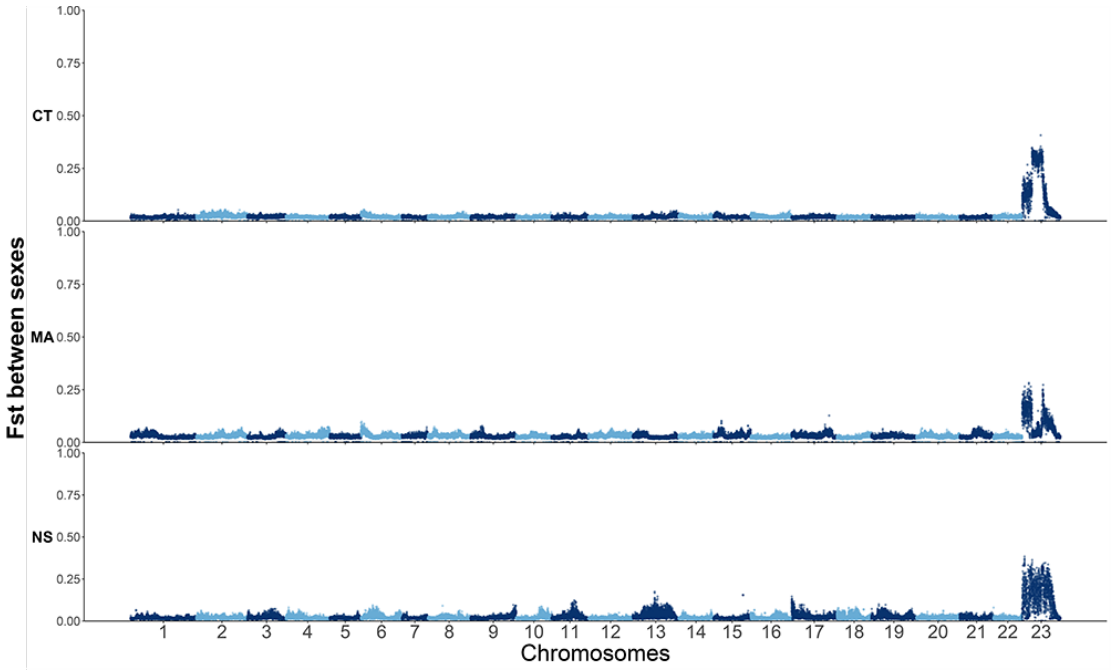


Fig S2. Genomic distribution of fixation index (Fst) between males and females in genetic crosses from three populations (CT, MA, and NS). The size of the sliding window is 20 kb and the step size is 10kb. Chromosomes are indicated on the X-axis, and the Fst values are shown on the Y-axis. Dark and light blue regions indicate the different chromosomes.

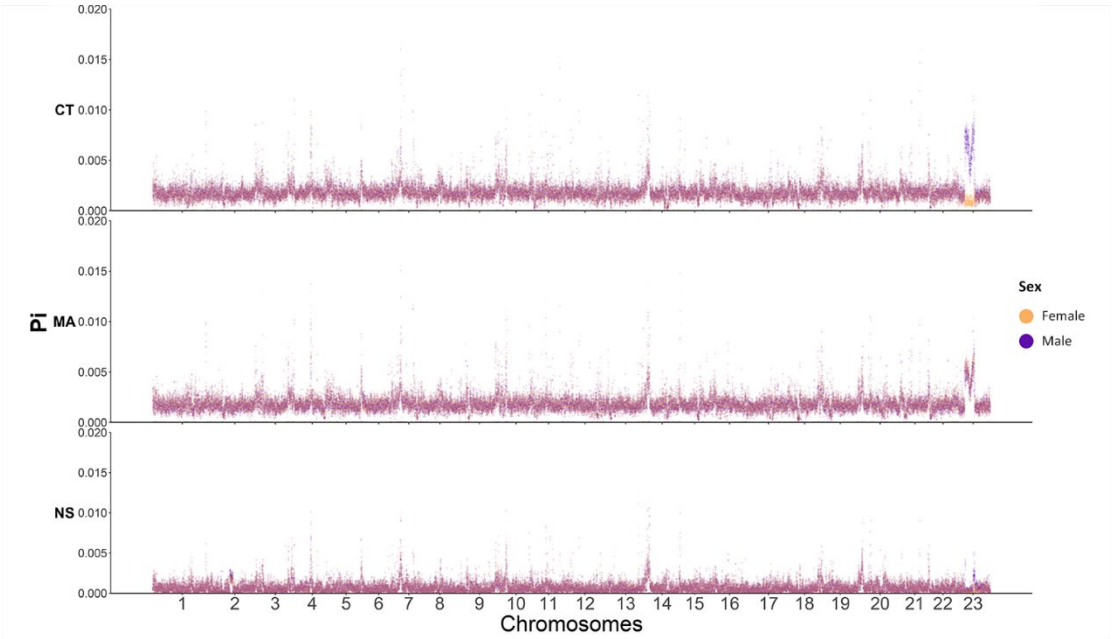


Fig S3. Genomic distribution of genetic diversity (Pi) in 20kb sliding windows within males and females in genetic crosses from three populations (CT, MA, and NS). Chromosomes are indicated on the X-axis, and the values of genetic diversity (Pi) are shown on the Y-axis. Purple dots represent males and yellow dots represent females.

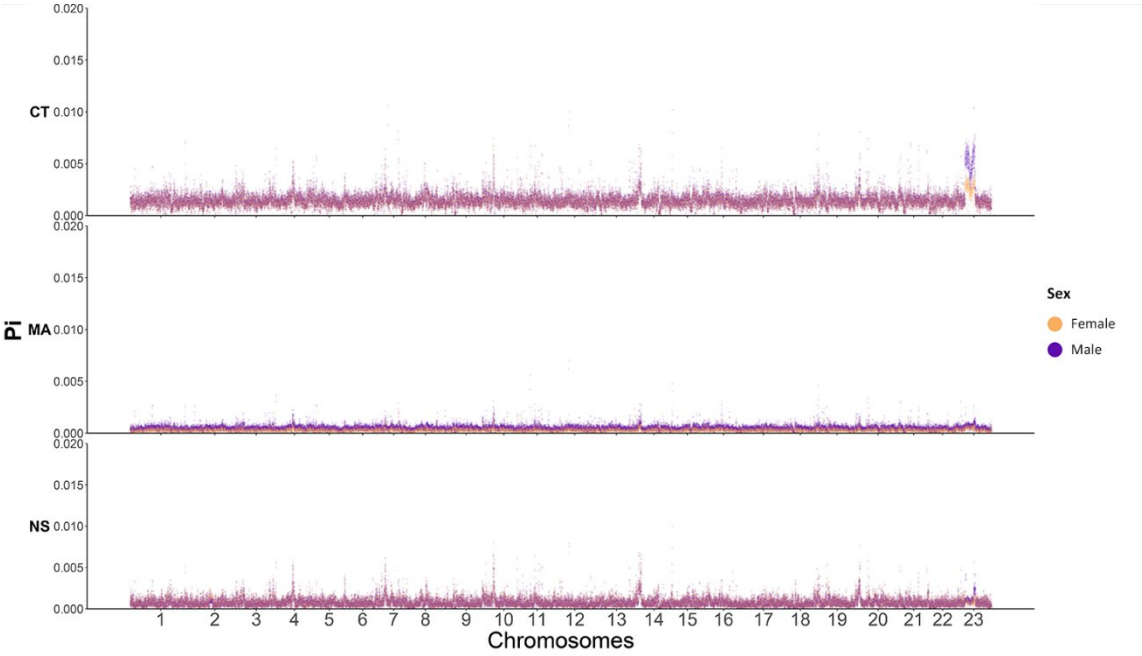


Fig S4. Genomic distribution of genetic diversity (Pi) in 20kb sliding windows within males and females calculated from linked-read data from three populations (CT, MA, and NS). Chromosomes are indicated on the X-axis, and the values of genetic diversity (Pi) are shown on the Y-axis. Purple dots represent males and yellow dots represent females.

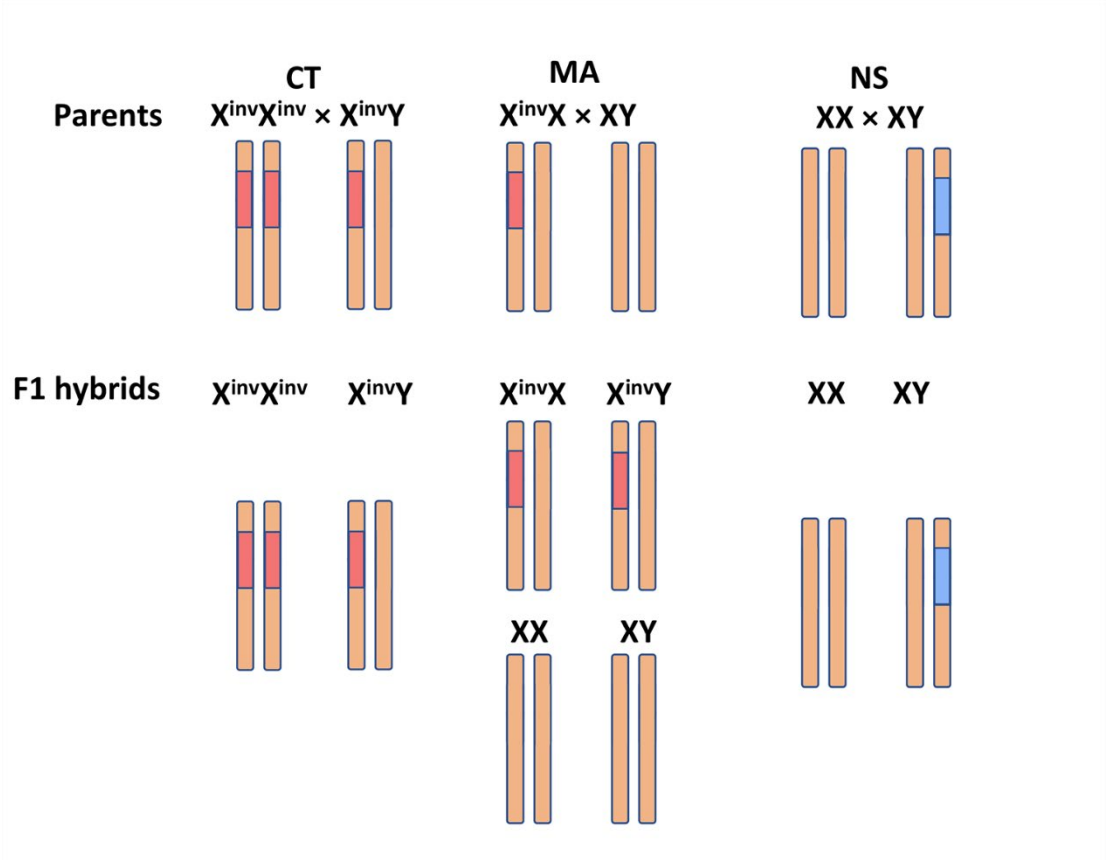


Fig S5. Segregation patterns of inversions in genetic crosses inferred from pool-seq data. Red bars represent the X-specific inversion, and blue bars represent the Y-specific inversion. For the MA cross, the number of individuals with each genotype is assumed to be equal within a sex.

Table S1. Result of SEX-Detector. The first column represents the model used in each run. The second column shows the number of sex-linked transcripts detected.

Model	Number of sex-linked transcripts
XY	86
ZW	0

Table S2. Identification of inversions and inferred genotypes in the CT and NS populations. Number of shared barcodes are calculated in a sliding window with a size of 50kb. The genotypes of the inversion between 8.60M-10.27M were not inferred due to the poor alignments of sequence in this region.

Sample	Shared barcodes between adjacent windows	X-specific inverion (4.38M-9.12M)		
		Shared barcodes between windows of inversion	Evidence from insertion size or clipped reads at breakpoints	Inferred Genotype
CT_fem_1	146.784	158	YES	1/1
CT_fem_10	44.225	49	YES	1/1
CT_fem_11	65.4448	64	YES	1/1
CT_fem_12	35.2056	40	YES	1/1
CT_fem_13	57.2558	49	YES	1/1
CT_fem_14	47.3166	54	YES	1/1
CT_fem_15	87.8311	70	YES	1/1
CT_fem_16	61.0821	56	YES	1/1
CT_fem_17	129.146	113	YES	1/1
CT_fem_18	75.0549	81	YES	1/1
CT_fem_19	72.808	48	YES	0/1
CT_fem_2	147.93	139	YES	1/1
CT_fem_20	59.4601	55	YES	1/1
CT_fem_3	80.3556	70	YES	1/1
CT_fem_4	141.901	179	YES	1/1
CT_fem_5	73.5794	82	YES	1/1
CT_fem_6	45.6108	39	YES	1/1
CT_fem_7	77.0803	36	YES	0/1
CT_fem_8	88.0201	32	YES	0/1
CT_fem_9	65.6976	64	YES	1/1
CT_male_1	104.123	52	YES	0/1
CT_male_10	107.733	60	YES	0/1
CT_male_11	82.4849	30	YES	0/1
CT_male_12	100.3	51	YES	0/1
CT_male_13	87.9238	48	YES	0/1
CT_male_14	29.5812	0	NO	0/0
CT_male_15	63.0449	30	YES	0/1
CT_male_16	80.4188	38	YES	0/1
CT_male_17	99.1335	45	YES	0/1
CT_male_18	106.076	50	YES	0/1
CT_male_19	119.126	51	YES	0/1
CT_male_2	93.1234	0	NO	0/0
CT_male_20	74.2546	32	YES	0/1
CT_male_3	101.255	71	YES	0/1
CT_male_4	101.571	63	YES	0/1
CT_male_5	133.293	87	YES	0/1
CT_male_6	87.3751	39	YES	0/1
CT_male_7	57.0254	0	NO	0/0
CT_male_8	105.576	62	YES	0/1
CT_male_9	65.8175	27	YES	0/1
NS_fem_1	77.5009	0	NO	0/0
NS_fem_10	111.621	0	NO	0/0
NS_fem_11	78.7797	0	NO	0/0
NS_fem_12	131.76	0	NO	0/0

NS_fem_13	129.083	0	NO	0/0
NS_fem_14	71.3881	0	NO	0/0
NS_fem_15	78.1536	0	NO	0/0
NS_fem_16	65.3349	0	NO	0/0
NS_fem_2	97.2144	0	NO	0/0
NS_fem_4	83.2056	0	NO	0/0
NS_fem_5	122.728	0	NO	0/0
NS_fem_6	101.697	0	NO	0/0
NS_fem_7	123.125	0	NO	0/0
NS_fem_8	99.0573	0	NO	0/0
NS_fem_9	183.191	0	NO	0/0
NS_male_1	112.904	0	NO	0/0
NS_male_10	58.5842	0	NO	0/0
NS_male_11	61.6031	0	NO	0/0
NS_male_12	76.4997	0	NO	0/0
NS_male_13	92.3385	0	NO	0/0
NS_male_14	54.2871	0	NO	0/0
NS_male_15	65.1559	0	NO	0/0
NS_male_16	85.0697	0	NO	0/0
NS_male_2	156.896	0	NO	0/0
NS_male_3	110.404	0	NO	0/0
NS_male_4	122.949	0	NO	0/0
NS_male_5	78.4436	0	NO	0/0
NS_male_7	108.999	0	NO	0/0
NS_male_9	102.568	0	NO	0/0

Sample	Shared inversion (8.60m-10.27m)		
	Shared barcodes between windows of inversion	Evidence from clipped reads or insertion size	Inferred Genotype
CT_fem_1	55	YES	NA
CT_fem_10	36	YES	NA
CT_fem_11	41	YES	NA
CT_fem_12	21	YES	NA
CT_fem_13	43	YES	NA
CT_fem_14	35	YES	NA
CT_fem_15	46	YES	NA
CT_fem_16	43	YES	NA
CT_fem_17	91	YES	NA
CT_fem_18	41	YES	NA
CT_fem_19	21	YES	NA
CT_fem_2	70	YES	NA
CT_fem_20	42	YES	NA
CT_fem_3	49	YES	NA
CT_fem_4	71	YES	NA
CT_fem_5	59	YES	NA
CT_fem_6	34	YES	NA
CT_fem_7	29	YES	NA
CT_fem_8	36	YES	NA
CT_fem_9	36	YES	NA
CT_male_1	59	YES	NA
CT_male_10	36	YES	NA
CT_male_11	35	YES	NA
CT_male_12	38	YES	NA
CT_male_13	55	YES	NA
CT_male_14	17	YES	NA
CT_male_15	35	YES	NA
CT_male_16	35	YES	NA
CT_male_17	65	YES	NA
CT_male_18	62	YES	NA
CT_male_19	42	YES	NA
CT_male_2	33	YES	NA
CT_male_20	35	YES	NA
CT_male_3	44	YES	NA
CT_male_4	76	YES	NA
CT_male_5	57	YES	NA
CT_male_6	54	YES	NA
CT_male_7	24	YES	NA
CT_male_8	69	YES	NA
CT_male_9	46	YES	NA
NS_fem_1	26	YES	NA
NS_fem_10	22	YES	NA
NS_fem_11	0	YES	NA

NS_fem_12	37	YES	NA
NS_fem_13	36	YES	NA
NS_fem_14	38	YES	NA
NS_fem_15	38	YES	NA
NS_fem_16	30	YES	NA
NS_fem_2	25	YES	NA
NS_fem_4	26	YES	NA
NS_fem_5	38	YES	NA
NS_fem_6	11	YES	NA
NS_fem_7	20	YES	NA
NS_fem_8	33	YES	NA
NS_fem_9	36	YES	NA
NS_male_1	71	YES	NA
NS_male_10	59	YES	NA
NS_male_11	48	YES	NA
NS_male_12	50	YES	NA
NS_male_13	59	YES	NA
NS_male_14	77	YES	NA
NS_male_15	51	YES	NA
NS_male_16	57	YES	NA
NS_male_2	90	YES	NA
NS_male_3	91	YES	NA
NS_male_4	59	YES	NA
NS_male_5	57	YES	NA
NS_male_7	90	YES	NA
NS_male_9	79	YES	NA

Sample	Y-specific inversion (4.47m-12.15m)		
	Shared barcodes between windows of inversion	Evidence from insertion size or clipped reads at breakpoints	Inferred Genotype
CT_fem_1	0	NO	0/0
CT_fem_10	0	NO	0/0
CT_fem_11	0	NO	0/0
CT_fem_12	0	NO	0/0
CT_fem_13	0	NO	0/0
CT_fem_14	0	NO	0/0
CT_fem_15	0	NO	0/0
CT_fem_16	0	NO	0/0
CT_fem_17	0	NO	0/0
CT_fem_18	0	NO	0/0
CT_fem_19	0	NO	0/0
CT_fem_2	0	NO	0/0
CT_fem_20	0	NO	0/0
CT_fem_3	0	NO	0/0
CT_fem_4	0	NO	0/0
CT_fem_5	0	NO	0/0
CT_fem_6	0	NO	0/0
CT_fem_7	0	NO	0/0
CT_fem_8	0	NO	0/0
CT_fem_9	0	NO	0/0
CT_male_1	0	NO	0/0
CT_male_10	0	NO	0/0
CT_male_11	0	NO	0/0
CT_male_12	0	NO	0/0
CT_male_13	0	NO	0/0
CT_male_14	0	NO	0/0
CT_male_15	0	NO	0/0
CT_male_16	0	NO	0/0
CT_male_17	0	NO	0/0
CT_male_18	0	NO	0/0
CT_male_19	0	NO	0/0
CT_male_2	0	NO	0/0
CT_male_20	0	NO	0/0
CT_male_3	0	NO	0/0
CT_male_4	0	NO	0/0
CT_male_5	0	NO	0/0
CT_male_6	0	NO	0/0
CT_male_7	0	NO	0/0
CT_male_8	0	NO	0/0
CT_male_9	0	NO	0/0
NS_fem_1	0	NO	0/0
NS_fem_10	0	NO	0/0
NS_fem_11	0	NO	0/0
NS_fem_12	0	NO	0/0
NS_fem_13	0	NO	0/0

NS_fem_14	0	NO	0/0
NS_fem_15	0	NO	0/0
NS_fem_16	0	NO	0/0
NS_fem_2	0	NO	0/0
NS_fem_4	0	NO	0/0
NS_fem_5	0	NO	0/0
NS_fem_6	0	NO	0/0
NS_fem_7	0	NO	0/0
NS_fem_8	0	NO	0/0
NS_fem_9	0	NO	0/0
NS_male_1	86	YES	0/1
NS_male_10	42	YES	0/1
NS_male_11	26	YES	0/1
NS_male_12	22	YES	0/1
NS_male_13	35	YES	0/1
NS_male_14	0	NO	0/0
NS_male_15	39	YES	0/1
NS_male_16	61	YES	0/1
NS_male_2	89	YES	0/1
NS_male_3	60	YES	0/1
NS_male_4	54	YES	0/1
NS_male_5	37	YES	0/1
NS_male_7	39	YES	0/1
NS_male_9	40	YES	0/1

Table S3. Distributions of male-specific SNPs on the sex chromosome in the CT and NS populations. Only SNPs with allele frequency higher than 0.45 are included. Proportion refers to the ratio of the number of male-specific SNPs in a specific region to the total number of male-specific SNPs on Chr23.

Distribution of male-specific snps				
	Chr	Region	Number	Proportion
Population CT	Chr23	0-4.37 Mb	0	0%
	Chr23	4.37-8.58 Mb	0	0%
	Chr23	8.58-9.40 Mb	31	100%
	Chr23	9.40-15.9 Mb	0	0%
Population NS	Chr23	0-4.40 Mb	0	0%
	Chr23	4.40-8.58 Mb	8	38%
	Chr23	8.58-9.40 Mb	9	43%
	Chr23	9.40-12.15 Mb	4	19%
	Chr23	12.15-15.9Mb	0	0%

Table S4. Genes with reduced coverage in males or stop codons in the CT and NS populations. Normalized sequencing depth was calculated for each sex. Only genes with a male:female depth ratio lower than 0.75 are listed.

Population	Chr	Start	End	Depth ratio (male vs female)	Gene	Blast results
CT	Chr23	8519701	8523240	0.659718092	ApeQuad13845	C1QL4
CT	Chr23	9108579	9109544	0.545791631	ApeQuad13892	NA
CT	Chr23	9113579	9114544	0.664014034	ApeQuad13894	OR2AT4
CT	Chr23	10259651	10265787	0.726332324	ApeQuad13960	SEPT4
NS	Chr23	9106872	9107129	0.577207278	ApeQuad13891	COA4
NS	Chr23	9108579	9109544	0.706774799	ApeQuad13892	NA

CT female				**CT male**			
Gene	**Chr**	**Start**	**End**	**Gene**	**Chr**	**Start**	**End**
Apequad13671	Chr23	6051869	6074700	ApeQuad13520	Chr23	2486399	2489497
Apequad13733	Chr23	7139796	7148738	ApeQuad13626	Chr23	5016305	5019935
				ApeQuad13652	Chr23	5486707	5487769
				ApeQuad13657	Chr23	5676005	5693345
				ApeQuad13665	Chr23	5935001	5967739
				ApeQuad13671	Chr23	6051869	6074700
				ApeQuad13713	Chr23	6926771	6932354
				ApeQuad13864	Chr23	8686046	8688041
NS female				**NS male**			
Gene	Chr	Start	End	Gene	Chr	Start	End
NA	NA	NA	NA	NA	NA	NA	NA

Table S5. Male-specific genes assembled in the sex-determination region on chromosome 23. The first column shows the gene name in *A. quadracus*. The second and third column show the blast hits to the NCBI database. The last column shows the types of mutations. Genes in bold are candidate sex-determination genes.

Genes	**Blast results**	**Abbreviation**	**Mutation type**
Apequad13864-RA	LIM domain kinase 1-like	LIMK1	Stop codon gained
Apequad13868-RA	NA	NA	Nonsynonymous mutation
Apequad13869-RA	NA	NA	Nonsynonymous mutation
Apequad13871-RA	NA	NA	Regulatory change
Apequad13875-RA	Methyltransferase Like 27	METTL27	Regulatory change
Apequad13877-RA	transcription elongation factor SPT6	SUPT6H	Nonsynonymous mutation
Apequad13878-RA	RAB34, member RAS oncogene family	RAB34	Regulatory change
Apequad13879-RA	nuclear speckle splicing regulatory protein 1	NSRP1	Nonsynonymous mutation
Apequad13880-RA	slingshot protein phosphatase 2	SSH2	Nonsynonymous mutation
Apequad13884-RA	telomerase-binding protein EST1A	SMG6	Nonsynonymous mutation & Regulatory change
Apequad13887-RA	reticulon 4 receptor like 2	RTN4RL2	Regulatory change
Apequad13889-RA	SET and MYND domain-containing protein 4	SMYD4	Nonsynonymous mutation & Regulatory change
Apequad13894-RA	olfactory receptor 24	OR2AT4	Nonsynonymous mutation
Apequad13895-RA	NXPE family member 3	NXPE3	Nonsynonymous mutation
Apequad13902-RA	**relaxin family peptide receptor 2**	**RXFP2**	**Nonsynonymous mutation & Regulatory change**
Apequad13904-RA	FRY microtubule binding protein	FRY	Nonsynonymous mutation & Regulatory change
Apequad13905-RA	**zygote arrest 1-**	**ZAR1L**	**Nonsynonymous mutation**

Table S6. Sample information and accession numbers for sequencing data in this study.

Population	Females	Males	Locality	Latitude	Longitude	Source	Supplier	Tissue
CT	20	20	West River Memorial Park, Connecticut, USA	41.314148	-72.95654	Wild-caught	Natalie Steinel, Daniel Bolnick	fin clips
MA	13	11	Demarest Lloyd State Park, Massachusetts, USA	41.528994	-70.98337	Wild-caught	Catherine Peichel	fin clips
NS	15	14	Canal Lake, Nova Scotia, Canada	44.498300	-63.90205	Wild-caught	Anne Dalziel	fin clips
CT cross F0 parents	1	1	West River Memorial Park, Connecticut, USA	41.314148	-72.95654	Wild-caught	Thomas Near	fin clips
CT cross F1 offspring	53	55	West River Memorial Park, Connecticut, USA	41.314148	-72.95654	Lab-reared	Peichel lab	fin clips
MA cross F0 parents	1	1	Demarest Lloyd State Park, Massachusetts, USA	41.528994	-70.98337	Wild-caught	Catherine Peichel	fin clips
MA cross F1 offspring	18	22	Demarest Lloyd State Park, Massachusetts, USA	41.528994	-70.98337	Lab-reared	Peichel lab	fin clips
NS cross F0 parents	1	1	Canal Lake , Nova Scotia, Canada	44.498300	-63.90205	Wild-caught	Anne Dalziel	fin clips
NS cross F1 offspring	19	16	Canal Lake , Nova Scotia, Canada	44.498300	-63.90205	Lab-reared	Peichel lab	fin clips
NS cross F1 offspring	12	12	Canal Lake , Nova Scotia, Canada	44.498300	-63.90205	Lab-reared	Peichel lab	brain
Assembly of *Apeltes quadracus*	1	0	Rainbow Haven Beach x Canal Lake, Nova Scotia, Canada	NA	NA	Lab-reared	Peichel lab	NA

Data type	**Library prep**	**Sequence method**	**Sequence location**	**Publication**	**Funder**
DNA-SEQ	Haplotagging	150 bp paired end Illumina NovaSeq S4	Genewiz, Germany	this study	SNF 31003A_176130
DNA-SEQ	Haplotagging	150 bp paired end Illumina NovaSeq S4	Genewiz, Germany	this study	SNF 31003A_176130
DNA-SEQ	Haplotagging	150 bp paired end Illumina NovaSeq S4	Genewiz, Germany	this study	SNF 31003A_176130
DNA-SEQ	Illumina DNA TruSeq	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	SNF 31003A_176130
POOL-SEQ	Illumina DNA TruSeq	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	SNF 31003A_176130
DNA-SEQ	Illumina DNA TruSeq	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	SNF 31003A_176130
POOL-SEQ	Illumina DNA TruSeq	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	SNF 31003A_176130
DNA-SEQ	Illumina DNA TruSeq	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	SNF 31003A_176130
POOL-SEQ	Illumina DNA TruSeq	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	SNF 31003A_176130
RNA-SEQ	Illumina mRNA TruSeq	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	SNF 31003A_176130
GENOME ASSEMBLY	NA	NA	NA	Liu et al. 2022	SNF 31003A_176130

Supplementary Appendix 1. Heterozygosity of Chr23 of each individual of CT and MA populations. See https://doi.org/10.6084/m9.figshare.22774919.v1 for details.

Chapter 3

Unveiling the convergent and divergent evolution of sex chromosomes from comparative genomics of complete Y assemblies in stickleback fish

Zuyao Liu[1], Catherine Peichel[1] and Daniel Jeffries[1, *]

1. Division of Evolutionary Ecology, Institute of Ecology and Evolution, University of Bern, 3012,Bern, Switzerland

***Corresponding author:** E-mail: daniel.jeffries@unibe.ch

This chapter is a manuscript in preparation.

Abstract

The emergence of sex chromosomes is a remarkable instance of convergent evolution in the natural world. The phenomenon of recombination suppression between sex chromosomes and subsequent degeneration of the non-recombining and heteromorphic sex chromosome is prevalent across eukaryotes. However, a comprehensive understanding of the underlying mechanisms remains elusive due to the lack of complete heteromorphic sex chromosome assemblies, which limits the ability to compare the sequence content of sex chromosomes across different species. In this study, we have developed a novel assembly pipeline utilizing long-read PacBio sequencing data and HiC data to fully phase the sex chromosomes. As a result, a complete assembly of the Y chromosome measuring 35.75 Mb was generated for the blackspotted stickleback (*Gasterosteus wheatlandi*). Our findings provide confirmation of an end-to-end fusion between one copy of Chr12 and the ancestral Y chromosome, Chr19, and the identification of five evolutionary strata on this neo-Y chromosome. Additionally, we discovered multiple chromosomal inversions associated with strata on the Y chromosome. Upon comparing the Y assembly of *G. wheatlandi* with that of *G. aculeatus*, we observed a faster rate of degeneration and the accumulation of deleterious mutations on the Y chromosome of *G. wheatlandi*, despite the Y chromosomes being partially homologous. Furthermore, the rate of degeneration on the Y chromosome was not constant, exhibiting rapid initial degeneration followed by gradual deceleration. In conclusion, our findings demonstrate that while the evolution of sex chromosomes can exhibit convergent patterns, closely-related sex chromosomes can also rapidly diverge over a short period of time.

Keywords: degeneration ,sex chromosome, stickleback, strata ,Y chromosome

Introduction

Heteromorphic sex chromosomes, including the commonly found XY and ZW systems, are a striking example of convergent evolution in nature. Similar patterns have been repeatedly observed across various organisms, including mammals, birds, fish, and flowering plants, with one sex chromosome being morphologically distinct from the other (Bachtrog et al. 2014; Pennell et al. 2018; Kratochvíl et al. 2021). Such heteromorphic sex chromosomes are thought to evolve after a pair of autosomes acquires a new master sex determining gene, , and recombination is reduced or ceased entirely around that gene (Charlesworth et al. 2005; Bachtrog 2006; Vicoso 2019). Several hypotheses have been proposed to explain the suppression of recombination on sex chromosomes: 1) linkage between sex determining genes and sexually antagonistic loci is favored by selection (Rice 1987; Wright et al. 2016); 2) meiotic drive (Úbeda et al. 2015); 3) sheltering of deleterious mutations (Charlesworth and Wall 1999); and 4) the accumulation of neutral divergence (Jeffries et al. 2021). No matter the cause, the suppression of recombination creates an opportunity for deleterious mutations and repeats to accumulate over time due to Hill-Robertson interference, which reduces the effectiveness of selection (Hill and Robertson 1966), and Muller's rachet (Muller 1964; Felsenstein 1974), whereby deleterious mutations accumulate due to the absence of purifying selection. Both of these mechanisms will lead to the subsequent degeneration and loss of content on the non-recombining member of the sex chromosome pair.

Although sex chromosome degeneration is quite common, the extent of degeneration can vary, even between closely related species (Ma and Rovatsos 2022). At one extreme, some species possess a homomorphic and undifferentiated sex chromosome pair, with little degeneration. Such homomorphic sex chromosomes have been found in frogs (Ma and Veltsos 2021), non-avian reptiles (Vicoso, Emerson, et al. 2013), ratite birds (Vicoso, Kaiser, et al. 2013; L. Xu et al. 2019), and teleost fish (El Taher et al. 2021; Xue et al. 2021; Long et al. 2023). At the other extreme, some species have sex chromosomes where the vast majority of the Y chromosome gene content has been lost, as in most mammals (Bull 1983; Hughes et al. 2005; Bellott et al. 2014; Cortez et al. 2014) , or where the Y chromosome has been lost entirely, as in Amami spiny rats (Terao et al. 2022) and creeping voles (Couger et al. 2021). Nevertheless, sex chromosomes do not start losing content as a whole. In contrast, many studies have found that recombination suppression has occurred at different times in the evolutionary history of a sex chromosome, which results in regions with different degrees of

divergence between the X and the Y, referred as 'evolutionary strata' (Lahn and Page 1999; Handley et al. 2004). Some studies have tried to tackle the question of why sex chromosomes evolve at different rates across species. Yet, most of them focus on species that possess homologous but highly degenerated sex chromosomes (Hughes et al. 2005; Valenzuela and Adams 2011; L. Xu et al. 2019), or on homomorphic sex chromosomes but with ambiguous homology (Vicoso and Bachtrog 2013; Papadopulos et al. 2014; Jeffries et al. 2018). Therefore, there is still a gap in understanding how and why strata form and evolve at different rates.

Stickleback fishes (Gasterosteidae) are an excellent model system to investigate the evolution of recombination suppression and its consequences as different species in the family have evolved different sex chromosomes that show differences in their extent of differentiation (Jeffries et al., 2022, also see Chapter 2). Of these species, the *Gasterosteus* clade is of interest as they possess an XY sex chromosome system that is shared across the three *Gasterosteus* species but not with other species in the Gasterosteidae family (Sardell et al. 2021; Dagilis et al. 2022). The XY sex chromosome pair of threespine stickleback (*G. aculeatus*) is chromosome (Chr) 19, and it has a recombining pseudo-autosomal region (PAR) of 2.5 Mb and three different strata in the 16 Mb non-recombining region (Peichel et al. 2004; Ross and Peichel 2008; Peichel et al. 2020). These strata correspond to three independent inversions between X and Y chromosomes, suggesting step-wise suppressions of recombination. Of these strata, S1 is the oldest and most degenerated, which was estimated to have originated around 22 Mya. The other two strata, S2 and S3, were estimated to have stopped recombining more recently (4.7-5.9 Mya) and are less degenerated (Peichel et al. 2020). The sister species, the Japan Sea stickleback (*G. nipponicus*), shares the same strata on Chr19 with no obvious divergence found between the species (Dagilis et al. 2022). However, around 1-2 Mya, one copy of Chr9 was fused with the Y chromosome of Chr19, leading to a neo-Y chromosome with a 6.9 Mb stratum that has undergone little degeneration or divergence between the X and the Y (Kitano et al. 2009; Yoshida et al. 2014; Dagilis et al. 2022). Similar to the Japan Sea stickleback, the blackspotted stickleback (*G. wheatlandi*), has experienced a fusion between the ancestral Y chromosome and an autosome. However, this is an independent fusion involving Chr12 (Ross et al. 2009). Although Chr19 was identified as the ancestral sex chromosome in blackspotted stickleback, only the oldest stratum S1 is shared with the other two species (*G. aculeatus* and *G. nipponicus*) (Sardell et al. 2021). However, significant differences exist on stratum S1, with blackspotted stickleback evolving

faster and losing more genes than threespine stickleback (Sardell et al. 2021). While this study explored the sex chromosomes of blackspotted stickleback, it used the threespine stickleback X chromosome as a reference, potentially introducing bias due to species divergence and limitations in short-read sequencing (Sardell et al. 2021). However, these results suggest that a comparison of the blackspotted and threespine sex chromosomes presents an opportunity to study the convergence and repeatability of sex chromosomes over time. Additionally, the young neo-sex chromosome of blackspotted stickleback provides insight into the step-by-step loss of recombination after fusion.

In this study, we generated high-quality assemblies of the autosomes, X chromosomes, and Y chromosome of blackspotted stickleback (*G. wheatlandi*) using PacBio and HiC sequencing data. We used these assemblies, combined with population level whole genome resequencing data, as well as the available assembly of the threespine stickleback (*G. aculeatus*) Y chromosome (Peichel et al. 2020) to identify structural rearrangements on the *G. wheatlandi* Y chromosome, estimate the times of recombination suppression on this chromosome, and compare the rate of degeneration between different evolutionary strata on the Y chromosome both within and between species.

Results

De novo assemblies and annotations of the blackspotted stickleback genome

Using high-coverage PacBio and Hi-C data, we successfully assembled the genome of the blackspotted stickleback (*G. wheatlandi*). The raw PacBio read coverage and HiC read coverage was approximately 56x (~28 Gb total sequence) and 109x (~109 Gb total sequence), respectively, across the entire genome (See Supplementary Table S1 for sequencing details). After extracting Y-specific reads from both datasets (see Methods and Fig. 1), the final assembly of autosomes and X chromosomes is 520.48 Mb with 490 scaffolds, of which the longest 21 scaffolds correspond to the number of chromosomes and are at the chromosome level. The N50 length is 23.51 Mb, and the assembly quality, evaluated by BUSCO, exhibits relatively high completeness (~95%) (Supplementary Fig S1). Compared to other sticklebacks, the blackspotted stickleback has a larger genome size (~429 Mb for *A. quadracus* [Liu et al., 2022], ~449 Mb for *G. aculeatus* [Nath et al., 2021], and ~521 Mb for *P. pungitius* [Varadharajan et al., 2019]). Our assembly aligns very well with the chromosome-level assembly of threespine stickleback (Supplementary Fig S2).

To assemble the Y chromosome of the blackspotted stickleback, we performed phasing of the sex chromosomes with PacBio and HiC reads to extract PacBio reads that contain Y-specific variants (see Methods and Fig 1). Integrated with the isolated Y-specific PacBio reads, we were able to successfully assemble the entire Y chromosome, which has a length of 35.75 Mb (Fig 2A). Validation of our method for isolating Y reads indicated that only putative Y reads were used for the assembly (Fig 2B).

Following the construction of a repeat library, we performed repeat masking of the genome assemblies. For the assembly of autosomes and X chromosomes, 27.80% of the total content was masked, whereas, for the assembly of the Y chromosome, 47.05% of the total content was masked. The remaining genome was annotated using evidence from RNA-seq data, homologous proteins, and ab initio annotation. Any annotated genes that exhibited poor quality, typically characterized by an annotation edit distance (AED) of > 0.5, were filtered out. This process led to the final version of the annotation containing 25,806 genes for the assembly of autosomes and X chromosomes and 1,478 genes for the assembly of the Y chromosome.

Centromeric repeats were identified and extracted from raw PacBio reads, and higher order structure, a block of centromere repeats forming a larger unit of tandem repeats, was further assessed to locate the position of the centromere. The positions of centromeres on Chr12 and the Y assembly are around 24.9Mb and 26.4 Mb respectively (Fig 3A). The position of the centromere on Chr19 was not found; this sequence might have been deleted due to the uneven coverage of repeats during the HiC scaffolding step (Dudchenko et al. 2017). We infer that the centromere of Chr12 was kept as the centromere of the Y chromosome in blackspotted stickleback. This is because the centromere of the threespine stickleback Y is highly degenerated (Cech and Peichel 2015; Sardell et al. 2021), and the the centromere of the fused autosome (Chr9) was retained on the neo-Y chromosome in the Japan Sea stickleback (Cech and Peichel 2016).

Multiple chromosomal inversions are found between the X and Y chromosomes of blackspotted stickleback

We compared the assemblies of their X and Y chromosomes and found four chromosomal inversions on Chr12 (Fig 3A and Supplementary Fig S2). We also determined that a 2.3 Mb segment of Chr12 was translocated to the other side of the centromere on the

Y chromosome, and one of the four chromosomal inversions (black inversion in Fig. 3A) on Chr12 was nested in a larger inversion. Our analysis revealed that three inversion events have occurred, with the first between 12.05Mb and 21.59Mb (purple and green blocks in Fig 3) and the second between 6.35Mb and 12.05Mb (blue and green blocks in Fig 3). Interestingly, a small block (the green block in Fig 3) was involved in both of these inversion events. There is third nested inversion (the black block in Fig 3) that could have occurred in either of the two major inversion events mentioned above. The alignments of blocks of Chr19 are chaotic, and no clear and large homologous fragments are identified due to the presence of a large number of repeats. Despite this, we do observe homologies between genes on X and Y chromosomes, which are dispersed across the entire Y chromosome.

The Y chromosome of blackspotted stickleback has five evolutionary strata

Chromosomal inversions have been considered as one of the major factors that lead to suppression of recombination between X and Y chromosomes. Once recombination stops, divergence starts to accumulate between X and Y chromosomes, and the extent of divergence should reflect the time since recombination was suppressed. Regions in which recombination has been suppressed at different time correspond to evolutionary strata. To ask whether the inversions and translocations we identified contribute to recombination suppression and thus the formation of strata, we distinguished four regions on Chr12: 1) a pseudo-autosomal region (PAR) between 0 Mb to 6.35Mb; 2) putative stratum SW5 (corresponding to the blue inversion) between 6.35 Mb and 12.05Mb; 3) putative stratum SW4 (corresponding to the purple and green inversions) between 12.05 Mb and 21.59 Mb; and 4) putatitve stratum SW3 (corresponding to the translocation in orange) between 21.59 Mb to the end of the chromosome (Fig 3B). For Chr19, despite that there is little homology between the X and the Y chromosomes, we still partitioned it into two regions based on the fact that stratum S1 (between 14.69M and the end of the chromosome) is shared across all species of *Gasterosteus* (Sardell et al. 2021), and the rest of Chr19 is taken as a single stratum, stratum SW2. We then calculated several statistics across these putative strata on both Chr12 and Chr19 (Fig 3B).

First, the ratio of normalized depth between the two sexes was calculated across the X chromosome assemblies (Fig 3B). As the ancestral sex chromosome, Chr19 shows an obvious reduction in male read depth across the entire chromosome, suggesting a highly degenerated

Y chromosome. On the contrary, there is no reduction in male read depth on Chr 12, except for SW3 where the read depth ratios are more dispersed (Fig 3B).

Second, Fst was calculated using phased X and Y-sequences to determine the extent of accumulated divergence between X and Y chromosomes (Fig 3B). The distribution of divergence across the entire Chr19 was found to be dispersed, with Fst values ranging from 0 to 1. In contrast, Chr12 exhibited four distinct regions of divergence, including: 1) a region with low Fst in the pseudoautosomal region (PAR); 2) a region with elevated Fst in SW5 and SW4; 3) a region nested within SW4 with lower Fst; and 4) a region with the highest Fst in SW3. These observed regions of X-Y divergence were found to align reasonably well with the evolutionary strata defined by the inversions in the middle of SW4 on Chr12.

Third, to calculate the dS values, we identified all one-to-one gene pairs between the X and Y chromosomes. Although there is considerable variation in dS values across Chr19, which may be due to insufficient data points, we still observed a discernable difference between S1 and SW2. On Chr12, SW3 has the highest dS value, while the other two strata, SW4 and SW5, possess similar dS values.

Finally, the detection of strata can be enhanced by increasing the number of species analyzed (Zhang et al. 2022). To date the divergence time between sex chromosomes, we employed single-copy genes obtained from Orthofinder. Our analysis revealed that the average divergence time of stratum S1 (~ 21.2 mya) is older than the SW2 region (~ 13.9 mya) on Chr19. Regions SW5, SW4, and SW3 have estimated divergence times of 2.42 mya, 2.21 mya, and 3.59 mya, respectively (Fig. 3B).

Taken together, these results suggest that there are five distinct evolutionary strata on the blackspotted Y chromosome, which likely reflect a step-wise suppression of recombination through chromosomal rearrangements events at different points in the evolutionary past (Peichel et al. 2020).

The progressive loss of genes on the Y Chromosome over time

As recombination suppression occurs, the accumulation of deleterious mutations on the chromosome leads to the eventual loss or non-functionalization of Y chromosome genes. Accordingly, we investigated the number of genes lost on the Y chromosome in each stratum of both species. Our analysis revealed that in *G. wheatlandi*, the oldest stratum, S1, exhibited the highest number of lost genes, and SW2, the second oldest stratum, lost fewer genes. On

the neo-Y chromosome, more genes are lost in older strata. The numbers of genes lost in S4 and S5 is about the same, which is consistent with the fact that these two strata have similar ages (Fig 3B and Fig 4). Similarly, in *G. aculeatus* the oldest strata S1 has lost more Y chromosome genes than the younger strata. Surprisingly, we found that stratum S2 lost more genes than stratum S3, despite both strata having similar times of recombination suppression (Fig 4 and Supplementary Table S2). Additionally, it is noteworthy that the rate of gene loss appears to be non-linear in both species. In G. *wheatlandi*, genes have appeared to be lost at a faster rate earlier, with a leveling off in older strata (Fig 4). A similar observation can be made in *G. aculeatus* (Fig 4), although the pattern is not as obvious due to fewer strata and the similar ages of S2 and S3 (Peichel et al. 2020).

Deleterious mutations have accumulated in the older strata

It is expected that deleterious mutations will accumulate on the Y chromosome, due to its reduced ability to undergo recombination compared to the X chromosome. To investigate the evolution of the Y chromosome after recombination suppression, we specifically analyzed variants that exist only in males. By examining the distribution of alternative allele frequencies, we categorized male-specific SNPs into distinct groups (Supplementary Fig S4): sites that are fixed on the Y chromosome represent differences between the X and Y chromosomes, while sites that are polymorphic within males should have accumulated after recombination between the X and Y was suppressed.

Our analysis reveals a positive correlation between the age of evolutionary strata and the accumulation of genes carrying loss of function (LOF) mutations. Specifically, we observe a greater number of genes with LOF mutations in older strata compared to younger ones, suggesting that the Y chromosome has undergone a progressive loss of functional genes over evolutionary time. Notably, we observe a higher frequency of genes carrying LOF mutations in stratum S1 of *G. wheatlandi* compared to its counterpart in *G. aculeatus*, despite both strata originating from a common ancestor (Sardell et al. 2021), suggesting a more rapid loss of genes in *G. wheatlandi* (Fig 5A).

The dN/dS analysis yields conflicting results. In both *G. aculeatus* and *G. wheatlandi*, the oldest stratum, S1, on the ancestral sex chromosome, Chr19, exhibits the lowest dN/dS values compared to the other strata, as these regions are probably under purifying selection. However, on the neo sex chromosome of *G. wheatlandi*, the oldest stratum, SW3, has a

significantly higher dN/dS value than the other two strata, suggesting possible positive selection (Fig 5B).

Last, the unbiased genetic diversity of 4-fold degenerate sites was calculated to measure the neutral divergence in sites that are still polymorphic across autosomes, X chromosomes, and Y chromosomes. In both species, the diversities of X chromosomes are slightly lower than that of autosomes, which is close to 75%, the predicted relative effective population size of X chromosome (Table 1). As expected, the diversity on the Y chromosomes is much lower than on the X chromosomes in both species, although the diversity on the *G. aculeatus* Y is much greater than that of *G. wheatlandi*. In *G. aculeatus*, the oldest stratum S1 has the highest Y diversity, which is approximately 25% of the autosomal diversity, while the diversity of strata S2 and S3 is lower. In *G. wheatlandi*, the oldest strata S1 has the highest diversity. The diversity on the ancestral Y (Chr19) is higher than that of the neo-Y (Chr12), corresponding to a longer time of recombination cessation. However, on the neo-Y chromosome, the oldest stratum, SW3, has the lowest diversity.

Discussion

The rate of Y chromosome degeneration is dynamic: rapid initial decline followed by gradual deceleration

According to Bachtrog (2008), the loss of functional genes on the Y chromosome initially occurs at fast rate following recombination suppression. This rate gradually slows down and eventually stops. This is because as degeneration continues, the number of functional genes that can tolerate the accumulation of deleterious mutations decreases, and the remaining genes are more likely to be beneficial or crucial for males. As a result, deleterious mutations are more likely to be purged due to strong purifying selection (Bachtrog 2008).

In both species studied here, the proportion of gene loss on the Y chromosome positively correlates with the dS value, indicating an ongoing process of gene non-functionalization on the Y chromosome. Remarkably, the rate of gene loss appears to be non-linear, especially in *G. wheatlandi*, following the theoretical model mentioned above. Specifically, the loss of genes is fast in younger strata and gradually slows down in older strata after reaching a certain threshold, as expected (Fig 4). The results of analyzing single-copy gene pairs between the X and Y chromosomes in both species show that the number of genes with LOF mutations is higher in older strata compared to younger strata. Although there were no genes with LOF

mutations in stratum S3, this result could be attributed to the low number of single-copy genes in this stratum (only 54 genes). Moreover, the ancestral Y chromosome (Chr19) had a much higher proportion of genes with LOF mutations compared to the neo-Y chromosome (Chr12). These results provide evidence that supports the idea of a continuous process of non-functionalization of genes on the Y chromosome and the gradual accumulation of deleterious mutations.

It is worth noting that the dN/dS ratio does not show a positive correlation with the age of each stratum (Fig 5B). In both species, on the ancestral sex chromosome, the dN/dS ratio is significantly lower in the oldest stratum, S1, compared to the other strata. Peichel et al. (2020) demonstrated in threespine stickleback that genes retained in stratum S1 on the Y chromosome exhibit a higher degree of testis-biased expression, indicating their potential involvement in male development. Additionally, White et al. (2015) showed that genes retained in S1 in threespine stickleback are predicted to be dosage sensitive, suggesting that genes retained in S1 are under purifying selection. These results are further consistent with the model proposed by Bachtrog (2008) that genes under strong purifying selection tend to be maintained on Y chromosomes. In contrast, an accelerated gene evolution was observed on the neo-sex chromosome, with the younger strata exhibiting significantly greater values of dN/dS, particularly in *G. wheatlandi*. The higher dN/dS ratio suggests that there could positive selection in genes on the young Y chromosome. These observations suggests that the evolution of the Y chromosome is a complex and dynamic process, and the selective pressures acting on Y chromosome genes is not solely determined by their age.

Faster degeneration of the Y chromosome in *G. wheatlandi*

Despite the shared ancestry of S1 on the ancestral Y chromosome (Chr19) of *G. aculeatus* and *G. wheatlandi*, the rate of degeneration differs between the two species. Notably, for single-copy genes presented on both X and Y chromosomes, *G. wheatlandi* has fewer genes (73 out of 732) retained on its ancestral Y chromosome (Chr19) compared to *G. aculeatus* (307 out of 651). While the ages of strata could explain the difference to some degree, this explanation does not apply to stratum S1 (16.1% genes retained in *G. aculeatus* and 5.6 % genes reatained in *G. wheatlandi*, see Supplementary Table S2) where most genes come from a shared ancestor. Additionally, we observed a higher proportion of genes with LOF mutations on Chr19 in *G. wheatlandi*, particularly in stratum S1. These results collectively suggest a

faster rate of degeneration of the Y chromosome in *G. wheatlandi*.

The underlying mechanisms responsible for the differences in the evolutionary rates of homologous regions on sex chromosomes remain unclear. One potential hypothesis is that the effect of genetic drift is stronger in *G. wheatlandi*. This species is found within a narrower geographic range compared to *G. aculeatus* (Wootton 1976) and exhibits about half the amount of genetic diversity across the genome, as revealed by the unbiased genetic diversity of 4-fold degenerate sites (Table 1). Furthermore, the difference between *G. wheatlandi* and *G. aculeatus* may also be attributed to differences in their historical demographic patterns. *G. aculeatus* has undergone multiple colonization events from marine habitats to freshwater habitats, resulting in numerous isolated populations serving as reservoir of genetic variation (Nelson et al. 2019). Gene flow between marine and freshwater ecotypes has been observed frequently in the past million years (Jones et al. 2012; Østbye et al. 2018; Liu et al. 2022), which may have helped to connect isolated populations and maintain genetic diversity in *G. aculeatus*. By contrast, *G. wheatlandi*, as a near-coastal species, may have experienced multiple strong bottleneck events during the glacial cycles with limited genetic input from refugial populations. The findings suggest that *G. wheatlandi* may have a smaller effective population size, leading to a stronger impact of genetic drift. This could result in the random fixation of more deleterious mutations and repeats in the non-recombining region, ultimately leading to a faster loss of genes. However, the reduction in effective size alone cannot fully explain the extremely low diversity on the Y chromosome of *G. wheatlandi*, as the reduction of genetic diversity on the Y is not proportional to the reduction of genetic diversity on autosomes between *G. wheatlandi* and *G. aculeatus* (Table 1). An additional hypothesis is that the strength of sexually antagonistic selection between the two species could also contribute to the observed differences in evolving rates of Y chromosomes. Stronger sexually antagonistic selection would facilitate the fixation of sexually antagonistic alleles as well as of the inversions which hinder recombination. Further investigation of the loci that are subject to sexually antagonistic selection is required to address this question. There are some other forces that accelerate the rate of degeneration of sex chromosomes, such as faster mutation rates (Graves 2006), and fewer beneficial and dosage-sensitive genes on the Y chromosome . However, further detailed investigations are necessary to determine whether they contribute to differential rates of Y chromosome degeneration between *G. aculeatus* and *G. wheatlandi*.

Short-read sequencing biases the detection of evolutionary strata

By mapping short-read sequences to the X chromosome assembly of *G. aculeatus*, Sardell et al. (2021) previously identified six strata on the sex chromosomes of *G. wheatlandi*. By comparing assemblies of the *G. wheatlandi* X and Y chromosomes, our study identified only five strata with distinct boundaries (Fig. 3). Both studies identified the ancestral sex-determining region of all *Gasterosteus* species, stratum S1. However, unlike the multiple strata identified in Sardell et al. (2021) based on short reads mapped to *G. aculeatus*, no clear evidence of subsequent separation of extra strata on Chr19 is observed in our study, and thus the rest of the chromosome was recognized as a single stratum, SW2 (Fig 3B). No obvious signal of the PAR was observed on Chr19 in our study; however, our previous assembly of the *G. aculeatus* Y chromosome also did not include the PAR (Peichel et al. 2020), likely because our assembly strategy focused on identifying reads that were divergent from the X. On the neo-Y (Chr 12), we identified three strata, which is more than were previously found (Fig 3B). Stratum SW3 is consistent with the results from the previous study, while SW4 and SW5 were previously recognized as a single stratum. Although our data did not find a large difference in Fst or dS values between SW4 and SW5, we did estimate that they occurred at slightly different times (2.21 and 2.42 mya, respectively) and they are associated with two distinct inversion events. Thus, we still consider them as two different strata, although the fact that these two inversion may have occurred close in time has resulted in similar level of divergence.

Accurately defining evolutionary strata is an essential first step towards understanding the evolution of sex chromosomes. However, many studies on sex chromosomes rely solely on short read whole-genome sequencing data, which can limit the ability to detect regions that are evolving independently (Wright et al. 2017; Rifkin et al. 2021; Hearn et al. 2022). One limitation of using whole-genome sequencing data to detect strata is the difficulty in detecting events that occurred within a short time frame. However, in this study, we were able to identify an additional stratum on Chr12 by first identifying two inversions using long-read sequencing data (Fig 3A). Although further analyses of these two inversions using read depth, Fst and dS suggest that they occurred within a similar timeframe, the long-read sequencing data show that there have been two independent evolutionary events. The second limitation pertains to potential biases in mapping reads to sex chromosomes. Due to the cessation of recombination, repeat elements on sex chromosomes may have undergone amplification and accumulation, which can result in the underestimation or overestimation

of the number of SNPs. This phenomenon was observed in Peichel et al. (2020), wherein the estimation of synonymous mutations using PacBio assembly was higher than that using short reads, such that Strata S2 and S3 were considered as a single stratum until the generation of the Y chromosome assembly (Peichel et al. 2020). In addition to promoting the use of long-read data, it is important to use accurate reference assemblies to avoid bias when studying sex chromosome evolution. The faster rate of evolution of sex chromosomes often results in significant differences in genetic structure and architecture, even among closely related species. Thus, it is crucial to be mindful of potential errors associated with whole-genome sequencing data when characterizing sex chromosomes. And the use of phased assemblies, as we have done here, is fundamental as the accurate calling of haplotypes and structural variants between sister chromosomes are needed to precisely characterizing the evolutionary history of sex chromosomes.

Material and Methods

Ethics Statement

All experiments involving animals were approved by the Veterinary Service of the Department of Agriculture and Nature of the Canton of Bern (VTHa# BE4/16, BE17/17, and BE127/17).

Sample collection

For the genome assembly of *G. wheatlandi*, one lab-reared male resulting from a cross between wild-caught fish collected by Anne Dalziel from Canal Lake, Nova Scotia, Canada (44.498298, --63.90205) was used for PacBio sequencing, and two of his brothers were used for Hi-C sequencing. For whole-genome sequencing and RNA-sequencing of wild-caught fish, four males and four females of *G. wheatlandi* were collected from Rainbow Haven Beach in o a cot a Cana a 11 nne a e

DNA extraction and sequencing

High molecular weight DNA was extracted from the blood following the procedure in Peichel et al. (2020) of a single male and used to prepare a SMRTbell Express Library for PacBio sequencing. The liver of two brothers was used to prepare a Hi-C sequencing library using the

Phase Genomics Proximo Hi-C animal kit (Phase Genomics, Seattle, WA). Four SMRT cells were sequenced on the PacBio Sequel Platform, and the Hi-C libraries were sequenced for 300 cycles on an Illumina NovaSeq SP flow cell. DNA from eight wild-caught individuals (four females, four males) was extracted using phenol–chloroform and used to prepare Illumina DNA TruSeq libraries, which were sequenced for 300 cycles on an Illumina NovaSeq S1 flow cell. Brain RNA from the same individuals was used for RNA-sequencing, as previously described (Liu et al. 2022). All library preparation and sequencing was performed by the University of Bern Next Generation Sequencing Platform.

In this study, we also used available genome assemblies for *G. aculeatus* (Nath et al., 2021), *A. quadracus* (Liu et al., 2022), and *A. flavidus* (Li et al. 2022), plus the Y chromosome assembly for *G. aculeatus* (Peichel et al. 2020). We also used available whole-genome sequencing data from 17 male and 27 female wild-caught *G. aculeatus* (Shanfelter et al. 2019), 15 male wild-caught *G. wheatlandi* and four female wild-caught *G. aculeatus* (Sardell et al. 2021), 30 F1 interspecies hybrids (15 males and 15 females) resulting from the crosses of the wild-caught *G. aculeatus* females by *G. wheatlandi* males (Sardell et al. 2021) and brain RNA-seq data of four male and four female wild-caught *G. wheatlandi* (Liu et al. 2022). Detailed information for all samples and accession numbers are provided in Supplementary Table S1.

Identifying and isolating highly divergent Y-specific reads in *G. wheatlandi*

Before assembling the genome, we removed reads that are Y-specific in both the PacBio reads and the Hi-C reads to reduce potential assembly errors. The whole-genome sequencing data of four male and four female *G. wheatlandi* and the raw PacBio reads were used as input to isolate long reads that are specific to the Y chromosome by SRY (X.-B. Wang et al. 2020) with an estimated chromosome size of 28Mb (Sardell et al., 2021). These extracted PacBio reads were considered Y-specific reads that are highly divergent from ther X counterparts and excluded when assembling the autosomes and X chromosomes. For the Hi-C sequencing data, we first used KmerGO (Y. Wang et al. 2020) to isolate male-specific kmers with a length of 21 bp using the whole genome data from the same four males and four females. Next, we used the BBDuk.sh module in the BBMap program (Bushnell, 2014) to extract reads with at least one male-specific kmer covered. These extracted Hi-C reads were considered Y-specific reads and excluded in the scaffolding of the autosomes and X chromosomes.

Genome assembly of the autosomes and X chromosomes of *G. wheatlandi*

The PacBio assembly of *G. wheatlandi* was generated by flye 2.9.1 (Kolmogorov et al. 2019) with default parameters and then was polished by Racon 1.5.0 (Vaser et al. 2017) for two rounds using Hi-C reads from the brothers. We applied Purge_Dups (Guan et al. 2020) to reduce the impact of duplicated haplotigs in the assembly.

Scaffolding contigs was conducted using Hi-C proximity-guided assembly. Raw Hi-C reads were first trimmed with the same pipeline as described above. Trimmed Hi-C reads were further processed with HiCUP (Wingett et al. 2015) and Juicer (Durand et al. 2016) with default parameters and then used for scaffolding the assembled contigs using 3D-DNA (v. 180922)(Dudchenko et al. 2017). After the first round, we manually revised the assembly based on the HiC contact map and then scaffolded it again. Finally, LR_Gapcloser (G.-C. Xu et al. 2019) was used to close any gaps in the assembly. The quality of the assembly was validated by BUSCO v5 (Simão et al. 2015).

To validate our approach of removing Y-specific reads, we also generated another assembly with the Y-specific reads included. Then, we used bwa (v 0.7.11) (Li 2013) to align the whole-genome sequencing data of four females and four males to the two assemblies separately. We then calculated normalized sequencing depths in 1 kb sliding windows using reads per kilobase per million mapped reads (RPKM) method for each sex separately.

Alignment of short reads and SNP calling

Whole-genome DNA sequencing reads from each *G. aculeatus* and *G. wheatlandi* individual were first trimmed by Trimmomatic (v 0.36) (Bolger et al. 2014) with a sliding window of 4 bp. The first 15 bp of reads were dropped, and windows with an average quality score below 15 in the remaining reads were also dropped. Trimmed reads were mapped to the autosome and X chromosome assembly of the corresponding species using bwa (v 0.7.11) (Li 2013). Then, Samtools(Li et al. 2009) and GATK4 (Van and O'Connor 2020) were used to sort alignments and remove PCR duplicates from the bam files. SNP calling was conducted using Haplotype Caller, and joint genotyping was run on all individuals for each species separately in GATK4.

For SNP filtration, we first applied the hard filtration following the GATK best practices. Then, we used Vcftools (0.1.16) (Danecek et al. 2011) to keep sites with minimum genotype qualities greater than 20, more than five genotyped individuals and the count of the

alternative allele greater than 1. To prevent biases caused by paralogs, repeats and low sequencing depth, we also filtered out sites with a population mean coverage that was less than half or greater than twice the average value for each sex of each species.

Identifying male-specific SNPs

For *G. aculeatus*, we identified male-specific SNPs based on their occurrence only in males and not in females. Since we had whole-genome sequencing data from 23 males and 21 females, our approach should filter out the vast majority of SNPs that are shared between two sexes.

For *G. wheatlandi*, traditional identification of male-specific SNPs could not be applied to our whole-genome sequencing datatset with only four females and four males. Therefore, we used pedigrees to phase SNPs in 15 interspecies F1 hybrid crosses between *G. aculeatus* females and *G. wheatlandi* males (Sardell et al. 2021). For each heterozygous SNP present in the progeny, we leveraged the genotypic information of the parental individuals to ascertain the source of the paternally and maternally inherited alleles. Specifically, in the case of male and female offspring, paternally transmitted alleles were derived from sperm bearing a Y chromosome and an X chromosome, respectively. This provided us with 15 phased X chromosomes, 15 phased Y chromosomes and 30 phased autosomes.

Assembly of the Y chromosome in *G. wheatlandi*

On a Y chromosome, there can be three types of regions: (1) the PAR, where recombination happens and thus there is no divergence between X and Y chromosomes; (2) slightly divergent regions, where recombination has stopped and divergence has started to accumulate between X and Y chromosomes; and (3) highly divergent regions, where the divergence between X and Y chromosomes is quite high (Fig 1). Our approach to isolating Y-specific reads mainly focused on those that are highly divergent. To phase the regions that are slightly divergent between X and Y chromosomes, the putatively Y-specific PacBio reads were first mapped to the assembly of autosomes and X chromosomes using minimap2 (Li 2018). Then, we used Longshot (Edge and Bansal 2019) to call heterozygous sites in this individual and mapped all Hi-C reads to the same assembly using bwa (v0.7.11) (Li 2013) with default parameters. The SNP matrix from SNP calling with the PacBio reads and alignments from the Hi-C reads were fed to Hapcut2 (Edge et al. 2017) and Whatshap (Martin et al. 2016)

to conduct chromosome-level phasing. Finally, we extracted mapped PacBio reads of each haplotype from chromosomes 12 and 19, which are the sex chromosomes of *G. wheatlandi* (Ross et al. 2009; Sardell et al. 2021). Considering that Y chromosomes have degenerated, the dataset with a smaller size was considered to contain reads from the PAR and slightly divergent regions of the Y.

Thus, with data from PAR, slighly divergent reads and highly divergent reads, we generated a putative assembly of the Y chromosome (Fig 1). For Hi-C scaffolding, only reads with male-specific kmers were used. Fig 1 shows our complete pipeline to assemble the Y chromosome.

Genome annotation of *G. wheatlandi*

The genome assemblies were annotated in a two-step pipeline. In the first step, repeat elements were identified and annotated with a combined assembly including the autosomes, X chromosomes, and Y chromosome. Miniature inverted-repeat transposable elements (MITE)-Tracker (Crescente et al. 2018) was used to detect MITEs. Then, EDTA 2.0.1(Ou et al. 2019) was applied to annotate long-terminal repeats and novel repeats. Repeats libraries of MITE and EDTA were then merged into a non-redundant library and passed to RepeatMasker (v. 4.1.2) (Smit et al. 2013) for a final round of annotation.

In the second step, gene structures were predicted in the assemblies by four rounds of Maker3 (Holt and Yandell 2011) runs with a repeat-masked assembly. For the assembly of autosomes and X chromosomes, in the first round, we used the RNA-seq data the same four males and four females sequenced in this study (Supplementary Table S1) to create an assembly using Trinity 2.14.0(Haas et al. 2013). Protein data from *Danio rerio*, *G. aculeatus*, the Uniprot database as well as the RNA assembly were used as evidence for the program. The second round of annotation included two training and prediction steps by AUGUSTUS (v. 3.3.2) (Stanke et al. 2008) and SNAP (Korf 2004). Then, these results were passed to MAKER3. In the third round, GeneMARK-ES (Ter-Hovhannisyan et al. 2008) was used to train models and combined with MAKER3. Finally, the second-round annotation was repeated with the outputs from the third round to further polish gene structures. The final annotation was checked based on AED values, and only annotations with an AED score of 0.5 or less were retained for downstream analysis.

For the assembly of the Y chromosome, similar procedures were implemented, with the

distinction that only data from the four male individuals were utilized for RNA assembly.

Identifying the centromere position of *G. wheatlandi*

Identification of centromere repeats of *G. wheatlandi* was conducted following Melters et al. (2013). The most frequent repeat was identified as the candidate for the centromeric repeat. We used Perl scripts from Melters et al. (2013)to identify the higher order structure of the centromere, and then blasted the sequences of the higher order structure against each chromosome. Regions where most hits were located were considered as the position of the centromere on each chromosome.

Genomic synteny analyses to identify structural rearrangements between the X and Y chromosomes

First, we used minimap2 (Li 2018) to align chromosomes between *G. wheastlandi* and *G. aculeatus* and generated synteny plots using D-Genies(Cabanettes and Klopp 2018). Using the same approach, we generated synteny plots between the X and Y chromosomes within each species. Second, we extracted coding sequences from the X chromosomes and Y chromosome of *G. wheatlandi*, kept the longest transcript for each gene, and used JCVI (Haibao Tang et al. 2015) to compare synteny at the gene level. To investigate the order of occurrence of inversions between X and Y chromosomes, we used GRSR (Wang and Wang 2018) to infer the most likely order of events.

Identifying orthologs among stickleback species

Sequences from coding regions from *G. wheatlandi*, *G. aculeatus*, *A. quadracus* and *A. flavidus* were extracted or downloaded. The longest transcript of each gene from each species was extracted to identify orthologs using Orthofinder 2 (Emms and Kelly 2019). We identified one-to-one orthologs between X and Y chromosomes for *G. wheatlandi* and *G. aculeatus* separately. Gene rearrangements were identified based on the results of phylogenetic hierarchical orthogroups.

Molecular evolution of the Y chromosome of *G. wheatlandi*

We used our newly generated Y assembly to compare the features between the two X chromosomes and the Y chromosome. First, we extracted SNPs from the four males and four

females sequenced in this study and kept sites with minor allele counts greater than 1. The fixation index (Fst) between males and females was calculated with sliding windows of 1kb using Vcftools 0.1.16 (Danecek et al. 2011). Next, the ratio of normalized depths between the two sexes was calculated using mosdepth (Pedersen and Quinlan 2018) with sliding windows of the same size.Then, we used KaKsCalculator 3.0 (Zhang 2022) to estimate dS for each one-to-one gene pair between the two X chromosomes and the Y chromosome, which should reflect the neutral divergence. Finally, we extracted single-copy orthologs among all species from orthofinder analysis and aligned the corresponding protein sequences using PRANK(Löytynoja 2014). Pal2nal (Suyama et al. 2006) was used to guide the alignment of coding region sequences with protein data, and alignments longer than 200bp were kept for the following analysis. BEAST2 (Bouckaert et al. 2014) was used to estimate the divergence time of each gene between X and Y chromosomes. Since no calibration time was applied, we only calculated the relative divergence time compared to the root node. We defined strata on X chromosomes based on results from the above analyses and the chromosome rearrangements we identified.

Estimating deleterious mutation load on the Y chromosome

Mutations are expected to accumulate on the Y chromosome but are sheltered by their counterparts on the X chromosome as predicted by theory (Charlesworth and Wall 1999; Antonovics and Abrams 2004; Jay et al. 2022). To estimate mutation load on the Y chromosome for each species, we extracted one-to-one gene pairs present on both the X and Y chromosomes in the non-recombining regions within each species and grouped them into different categories by the defined strata. Next, we extracted SNPs that have at least one alternative allele, are male-specific, and are only present within the coding regions of one-to-one gene pairs between X and Y chromosomes, and annotated them using snpEff (Cingolani et al. 2012). We then counted the frequency of alternative alleles for each site and separated them into two groups: sites that are fixed with a frequency larger than 0.45 and sites that are are polymorphic with a frequency equal to or smaller than 0.45 on the Y chromosome. To compare with autosomes and X chromosomes, we repeated the same pipeline for SNPs on the autosomes and X chromosomes. However, we used female individuals when extracting SNPs on X chromosomes, and a frequency of 0.9 was used as the cut-off for identifying fixed SNPs on the X chromosome.

Once a gene gains a LOF mutation, the function of the gene is disrupted completely. Hence, the subsequent accumulation of LOF mutations would have no extra effect on the same gene. Thus, the proportions of genes with fixed LOF mutations were calculated for autosomes, X chromosomes, and the separate Y chromosome strata for each species. For the remaining genes with fixed non-synonymous mutations but no LOF mutation on the Y chromosome, haplotypes of each gene were reconstructed with bcftools (Li et al. 2009), and the dN/dS ratio was calculated for each gene. Also, we extracted SNPs that are still polymorphic on Y chromosomes and calculated the unbiased genetic diversity at 4-fold sites of each gene in each stratum using pixy 1.2.7 (Korunes and Samuk 2021).

References

Antonovics J, Abrams JY. 2004. Intratetrad mating and the evolution of linkage relationships. *Evolution* 58:702–709.

Bachtrog D. 2006. A dynamic view of sex chromosome evolution. *Curr. Opin. Genet. Dev.* 16:578–585.

Bachtrog D. 2008. The Temporal Dynamics of Processes Underlying Y Chromosome Degeneration. *Genetics* 179:1513–1525.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex determination: why so many ways of doing it? *PLoS Biol.* 12:e1001899.

Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghlul S, Graves T, Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* 508:494–499.

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

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* 10:e1003537.

Bull JJ. 1983. Evolution of sex determining mechanisms. The Benjamin/Cummings Publishing Company, Inc.

Bushnell B. 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner. Available from: https://www.osti.gov/biblio/1241166

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958.

Cech JN, Peichel CL. 2015. Identification of the centromeric repeat in the threespine stickleback fish (*Gasterosteus aculeatus*). *Chromosome Res.* 23:767–779.

Cech JN, Peichel CL. 2016. Centromere inactivation on a neo-Y fusion chromosome in threespine stickleback fish. *Chromosome Res.* 24:437–450.

Charlesworth B, Wall JD. 1999. Inbreeding, heterozygote advantage and the evolution of neo–X and neo–Y sex chromosomes. *Proc. R. Soc. Lond. B Biol. Sci.* 266:51–56.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6:80–92.

Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grützner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508:488–493.

Couger MB, Roy SW, Anderson N, Gozashti L, Pirro S, Millward LS, Kim M, Kilburn D, Liu KJ, Wilson TM, et al. 2021. Sex chromosome transformation and the origin of a male-specific X chromosome in the creeping vole. *Science* 372:592–600.

Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC Bioinformatics* 19:348.

Dagilis AJ, Sardell JM, Josephson MP, Su Y, Kirkpatrick M, Peichel CL. 2022. Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. *Philos. Trans. R. Soc. B Biol. Sci.* 377:20210205.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92–95.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* 3:95–98.

Edge P, Bafna V, Bansal V. 2017. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res.* 27:801–812.

Edge P, Bansal V. 2019. Longshot enables accurate variant calling in diploid genomes from single-

molecule long read sequencing. *Nat. Commun.* 10:4660.
El Taher A, Ronco F, Matschiner M, Salzburger W, Böhne A. 2021. Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. *Sci. Adv.* 7:eabe8215.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.
Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* 78:737–756.
Graves JAM. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124:901–914.
Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36:2896–2898.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–1512.
Haibao Tang, Krishnakumar V, Jingping Li. 2015. Jcvi: Jcvi Utility Libraries. Available from: https://zenodo.org/record/31631
Handley L-JL, Ceplitis H, Ellegren H. 2004. Evolutionary Strata on the Chicken Z Chromosome: Implications for Sex Chromosome Evolution. *Genetics* 167:367–376.
Hearn KE, Koch EL, Stankowski S, Butlin RK, Faria R, Johannesson K, Westram AM. 2022. Differing associations between sex determination and sex-linked inversions in two ecotypes of *Littorina saxatilis*. *Evol. Lett.* 6:358–374.
Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* 8:269–294.
Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, Wilson RK, Page DC. 2005. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* 437:100–103.
Jay P, Tezenas E, Véber A, Giraud T. 2022. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLOS Biol.* 20:e3001698.
Jeffries DL, Gerchen JF, Scharmann M, Pannell JR. 2021. A neutral model for the loss of recombination on sex chromosomes. *Philos. Trans. R. Soc. B Biol. Sci.* 376:20200096.
Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, Borzée A, Barrow LN, Canestrelli D, Crochet P-A, Dufresnes C, et al. 2018. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nat. Commun.* 9:4088.
Jeffries DL, Mee JA, Peichel CL. 2022. Identification of a candidate sex determination gene in *Culaea inconstans* suggests convergent recruitment of an *Amh* duplicate in two lineages of stickleback. *J. Evol. Biol.*:jeb.14034.
Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz J, Myers RM, et al. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* 461:1079–1083.
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37:540–546.
Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
Korunes KL, Samuk K. 2021. PIXY : Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* 21:1359–1368.
Kratochvíl L, Stöck M, Rovatsos M, Bullejos M, Herpin A, Jeffries DL, Peichel CL, Perrin N, Valenzuela N, Pokorná MJ. 2021. Expanding the classical paradigm: what we have learnt from vertebrates about sex chromosome evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 376:20200097.
Lahn BT, Page DC. 1999. Four Evolutionary Strata on the Human X Chromosome. *Science* 286:964–967.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* [Internet]. Available from: http://arxiv.org/abs/1303.3997
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome

Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* 25:2078–2079.
Li Q, Lindtke D, Rodríguez-Ramírez C, Kakioka R, Takahashi H, Toyoda A, Kitano J, Ehrlich RL, Chang Mell J, Yeaman S. 2022. Local Adaptation and the Evolution of Genome Architecture in Threespine Stickleback. *Genome Biol. Evol.* 14:evac075.
Liu Z, Roesti M, Marques D, Hiltbrunner M, Saladin V, Peichel CL. 2022. Chromosomal Fusions Facilitate Adaptation to Divergent Environments in Threespine Stickleback. *Mol. Biol. Evol.* 39:msab358.
Long X, Charlesworth D, Qi J, Wu R, Chen M, Wang Z, Xu L, Fu H, Zhang X, Chen X, et al. 2023. Independent Evolution of Sex Chromosomes and Male Pregnancy–Related Genes in Two Seahorse Species. *Mol. Biol. Evol.* 40:msac279.
Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. Mult. Seq. Alignment Methods:155–170.

Ma W, Rovatsos M. 2022. Sex chromosome evolution: The remarkable diversity in the evolutionary rates and mechanisms. *J. Evol. Biol.* 35:1581–1588.
Ma W-J, Veltsos P. 2021. The Diversity and Evolution of Sex Chromosomes in Frogs. *Genes* 12:483.
Martin M, Patterson M, Garg S, O Fischer S, Pisanti N, Klau GW, Schöenhuth A, Marschall T. 2016. WhatsHap: fast and accurate read-based phasing. Bioinformatics Available from: http://biorxiv.org/lookup/doi/10.1101/085050
Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby J, Sebra R, Peluso P, Eid J, Rank D, et al. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* 14:R10.
Muller HJ. 1964. The relation of recombination to mutational advance. *Mutat. Res.* 106:2–9.
Nath S, Shaw DE, White MA. 2021. Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 GenesGenomesGenetics* 11:jkab007.
Nelson TC, Crandall JG, Ituarte CM, Catchen JM, Cresko WA. 2019. Selection, Linkage, and Population Structure Interact To Shape Genetic Variation Among Threespine Stickleback Genomes. *Genetics* 212:1367–1382.
Østbye K, Taugbøl A, Ravinet M, Harrod C, Pettersen RA, Bernatchez L, Vøllestad LA. 2018. Ongoing niche differentiation under high gene flow in a polymorphic brackish water threespine stickleback (*Gasterosteus aculeatus*) population. *BMC Evol. Biol.* 18:14.
Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20:275.
Papadopulos AST, Kaye M, Devaux C, Hipperson H, Lighten J, Dunning LT, Hutton I, Baker WJ, Butlin RK, Savolainen V. 2014. Evaluation of genetic isolation within an island flora reveals unusually widespread local adaptation and supports sympatric speciation. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130342.
Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34:867–868.
Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM, Kingsley DM, et al. 2020. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* 21:177.
Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J, Myers RM, Mori S, Schluter D, Kingsley DM. 2004. The Master Sex-Determination Locus in Threespine Sticklebacks Is on a Nascent Y Chromosome. *Curr. Biol.* 14:1416–1424.
Pennell MW, Mank JE, Peichel CL. 2018. Transitions in sex determination and sex chromosomes across vertebrate species. *Mol. Ecol.* 27:3950–3963.
Rice WR. 1987. The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. *Evolution* 41:911.
Rifkin JL, Beaudry FEG, Humphries Z, Choudhury BI, Barrett SCH, Wright SI. 2021. Widespread Recombination Suppression Facilitates Plant Sex Chromosome Evolution. *Mol. Biol. Evol.* 38:1018–1030.
Ross JA, Peichel CL. 2008. Molecular Cytogenetic Evidence of Rearrangements on the Y Chromosome of the Threespine Stickleback Fish. *Genetics* 179:2173–2182.

Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). *PLoS Genet.* 5:e1000391.
Sardell JM, Josephson MP, Dalziel AC, Peichel CL, Kirkpatrick M. 2021. Heterogeneous Histories of Recombination Suppression on Stickleback Sex Chromosomes. *Mol. Biol. Evol.* 38:4403–4418.
Shanfelter AF, Archambeault SL, White MA. 2019. Divergent Fine-Scale Recombination Landscapes between a Freshwater and Marine Population of Threespine Stickleback Fish. *Genome Biol. Evol.* 11:1552–1572.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
Smit A, R H, P G. 2013. RepeatMasker Open-4.0. Available from: <http://www.repeatmasker.org>.
Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinforma. Oxf. Engl.* 24:637–644.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.
Terao M, Ogawa Y, Takada S, Kajitani R, Okuno M, Mochimaru Y, Matsuoka K, Itoh T, Toyoda A, Kono T, et al. 2022. Turnover of mammal sex chromosomes in the *Sry* -deficient Amami spiny rat is due to male-specific upregulation of *Sox9*. *Proc. Natl. Acad. Sci.* 119:e2211574119.
Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18:1979–1990.
Úbeda F, Patten MM, Wild G. 2015. On the origin of sex chromosomes from meiotic drive. *Proc. R. Soc. B Biol. Sci.* 282:20141932.
Valenzuela N, Adams DC. 2011. Chromosome number and sex determination coevolve in turtles. *Evol. Int. J. Org. Evol.* 65:1808–1813.
Van der AG, O'Connor B. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media.
Varadharajan S, Rastas P, Löytynoja A, Matschiner M, Calboli FCF, Guo B, Nederbragt AJ, Jakobsen KS, Merilä J. 2019. A high-quality assembly of the nine-spined stickleback (Pungitius pungitius) genome. *Genome Biol. Evol.*:evz240.
a er o a ara an 1 a t an acc rate e no o eno e a e ro on uncorrected reads. *Genome Res.* 27:737–746.
Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nat. Ecol. Evol.* 3:1632–1641.
Vicoso B, Bachtrog D. 2013. Reversal of an ancient sex chromosome to an autosome in Drosophila. *Nature* 499:332–335.
Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. *PLoS Biol.* 11:e1001643.
Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl. Acad. Sci.* 110:6453–6458.
Wang D, Wang L. 2018. GRSR: a tool for deriving genome rearrangement scenarios from multiple unichromosomal genome sequences. *BMC Bioinformatics* 19:291.
Wang X-B, Liu Q-Y, Li A-L, Ruan J. 2020. SRY: An Effective Method for Sorting Long Reads of Sex-limited Chromosome. Bioinformatics Available from: http://biorxiv.org/lookup/doi/10.1101/2020.05.25.115592
Wang Y, Chen Q, Deng C, Zheng Y, Sun F. 2020. KmerGO: A Tool to Identify Group-Specific Sequences With k-mers. *Front. Microbiol.* 11:2067.
White MA, Kitano J, Peichel CL. 2015. Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. *Mol. Biol. Evol.* 32:1981–1995.
Wingett SW, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* 4:1310.
Wootton RJ. 1976. The Biology of the Sticklebacks. Lodon: Academic Press
Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam B, Buechel SD, Kolm N, Breden F, Vicoso B, Mank JE. 2017. Convergent recombination suppression suggests role of sexual selection in guppy sex

chromosome formation. *Nat. Commun.* 8:14251.
Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. *Nat. Commun.* 7:12087.
Xu G-C, Xu T-J, Zhu R, Zhang Y, Li S-Q, Wang H-W, Li J-T. 2019. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience* [Internet] 8.
Xu L, Wa Sin SY, Grayson P, Edwards SV, Sackton TB. 2019. Evolutionary Dynamics of Sex Chromosomes of Paleognathous Birds. *Genome Biol. Evol.* 11:2376–2390.
Xue L, Gao Y, Wu M, Tian T, Fan H, Huang Y, Huang Z, Li D, Xu L. 2021. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. *Genome Biol.* 22:203.
Yoshida K, Makino T, Yamaguchi K, Shigenobu S, Hasebe M, Kawata M, Kume M, Mori S, Peichel CL, Toyoda A, et al. 2014. Sex Chromosome Turnover Contributes to Genomic Divergence between Incipient Stickleback Species. *PLoS Genet.* 10:e1004223.
Zhang H, Sigeman H, Hansson B. 2022. Assessment of phylogenetic approaches to study the timing of recombination cessation on sex chromosomes. *J. Evol. Biol.* 35:1721–1733.
Zhang Z. 2022. KaKs_calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinformatics*:S167202292100259X.

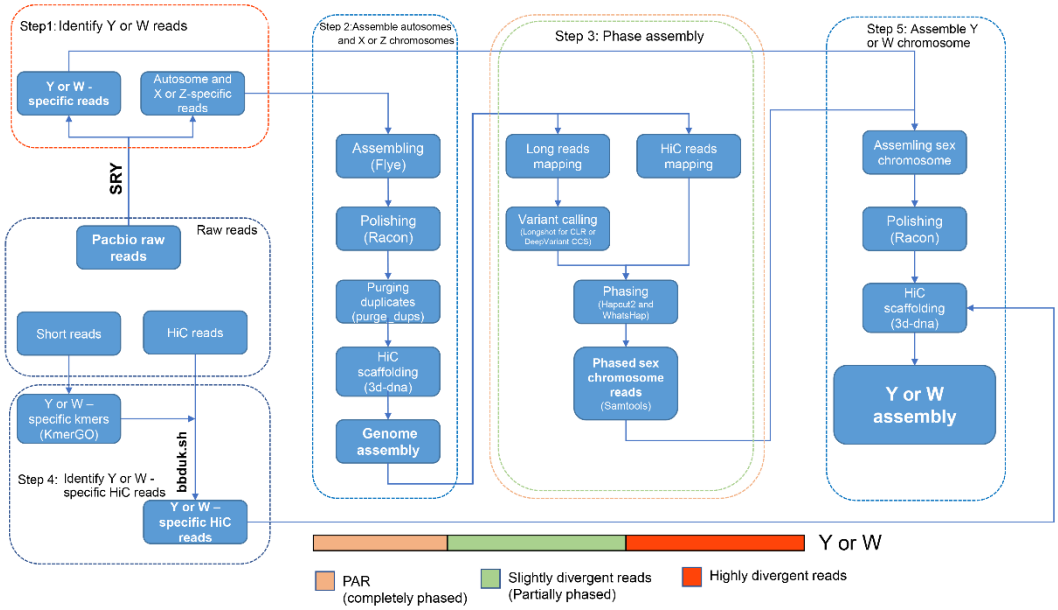


Fig 1. Pipeline for assembling the Y or W chromosome using reads from the PAR (light salmon), slightly divergent (green), and highly divergent (red) regions on the sex chromosome. Blue dotted boxes show normal assembly steps, while other colored dotted boxes show additional steps for incorporating corresponding reads from Y or W chromosomes.

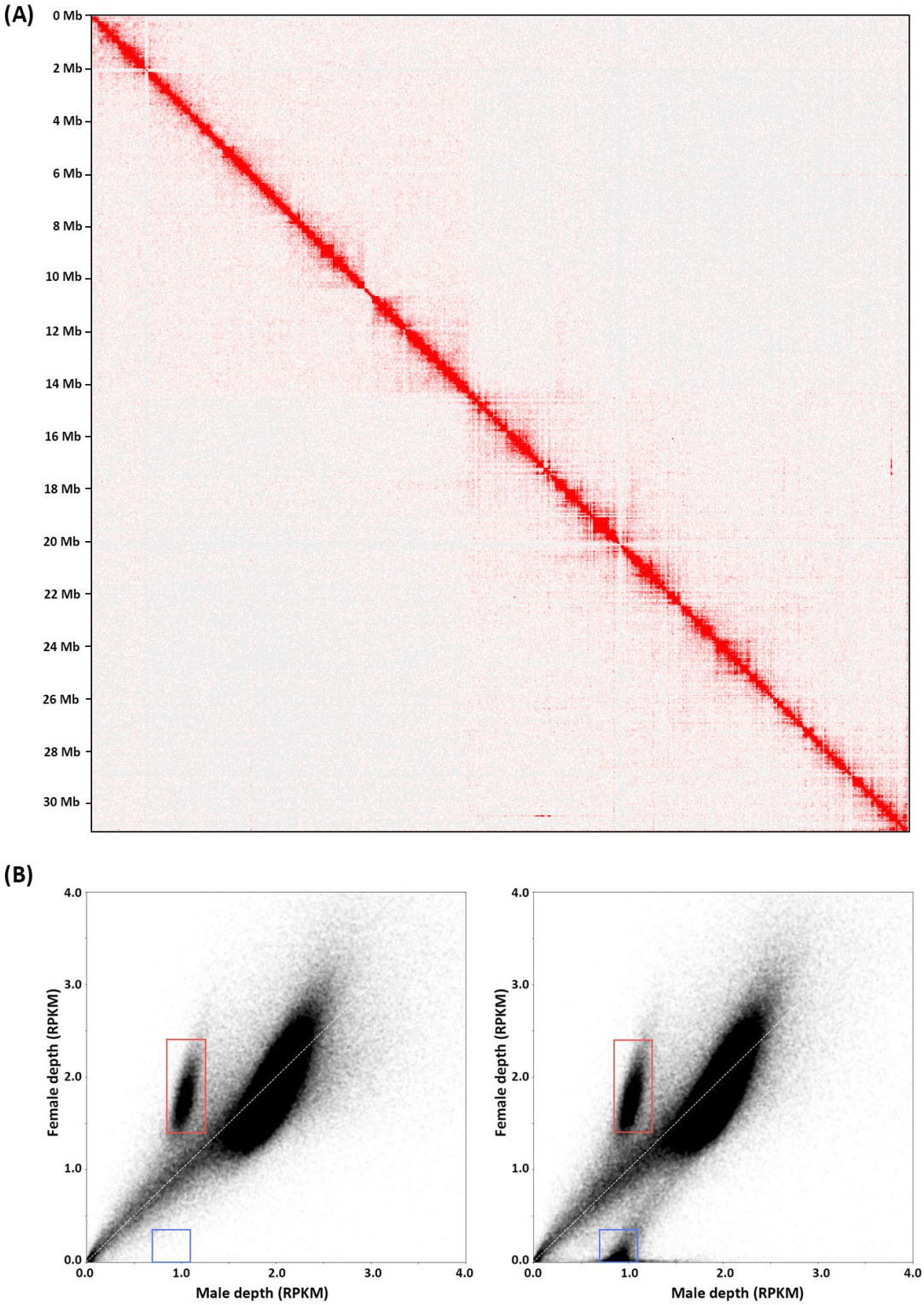


Fig 2. (A) HiC contact map of the neo-Y assembly. (B) Validation of Y-specific read removal. To validate the accuracy of the process of removing Y-specific reads, sequencing depth was calculated in 1kb windows for both sexes and normalized using the RPKM method. The left panel shows the distribution of sequencing depths with Y-specific reads removed, while the right panel shows the distribution of sequencing depths with Y-specific reads included.

Hemizygous regions on the X chromosome are indicated in the red box, and Y-specific regions on the Y chromosome are indicated in the blue box.

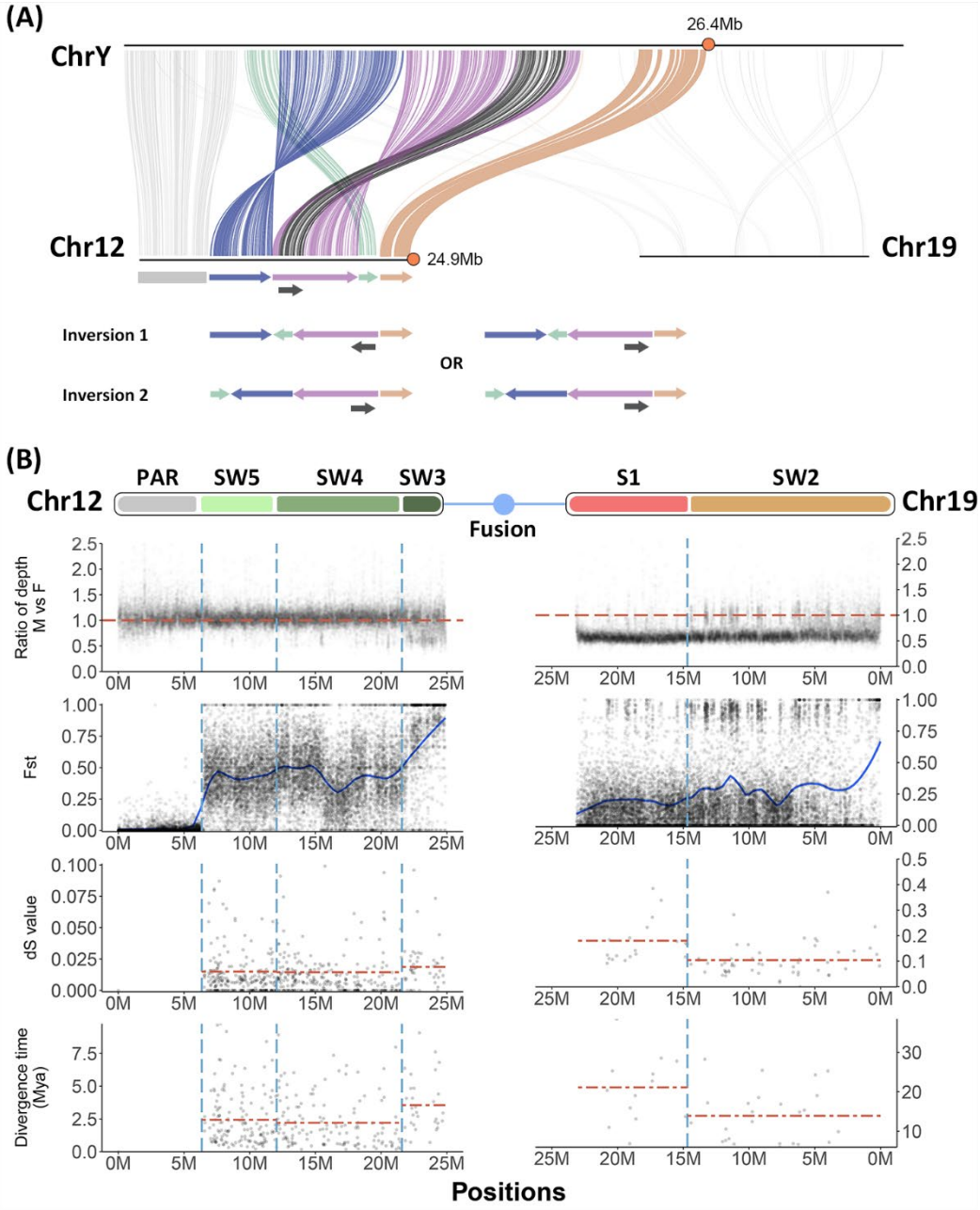


Fig 3. (A) Synteny map of X and Y chromosomes in *G. wheatlandi*. Homologous coding region sequences were used to generate this comparison, with colored lines indicating chromosomal rearrangements between the two X (Chrs 12 and 19) and Y chromosomes. Below Chr12, the grey rectangle represents the PAR, with colored arrows showing the orientation of the corresponding segments on the X (Chr 12). The predicted order of chromosomal inversions is shown below the synteny plot. The first inversion includes the purple and green regions, and the second inversion includes the blue and green regions. Notably, the black inversion could occur at any stage of either the first inversion or the second inversion. (B) Molecular evolution of the sex chromosomes in *G. wheatlandi.* The colored bars represent the defined

evolutionary strata on the X chromosomes, with blue circles indicating the fusion event. Arrows below the bars represent the putative inversions. Light blue dotted lines indicate the boundaries between the strata. From top to bottom, the figure displays the male:female read depth ratio (the red dotted line indicates the autosomal mean), F_{ST} between 15 phased X and 15 phased Y chromosomes (blue curved lines show the smoothed pattern), distribution of dS values between phased X and Y chromosomes (red dotted lines indicate the mean), and divergence time between phased X and Y chromosomes (red dotted lines indicate the mean; note the difference in scale between Chr12 and Chr19). Due to lack of differentiation between X and Y, the PAR is omitted from the distribution of dS values and divergence time.

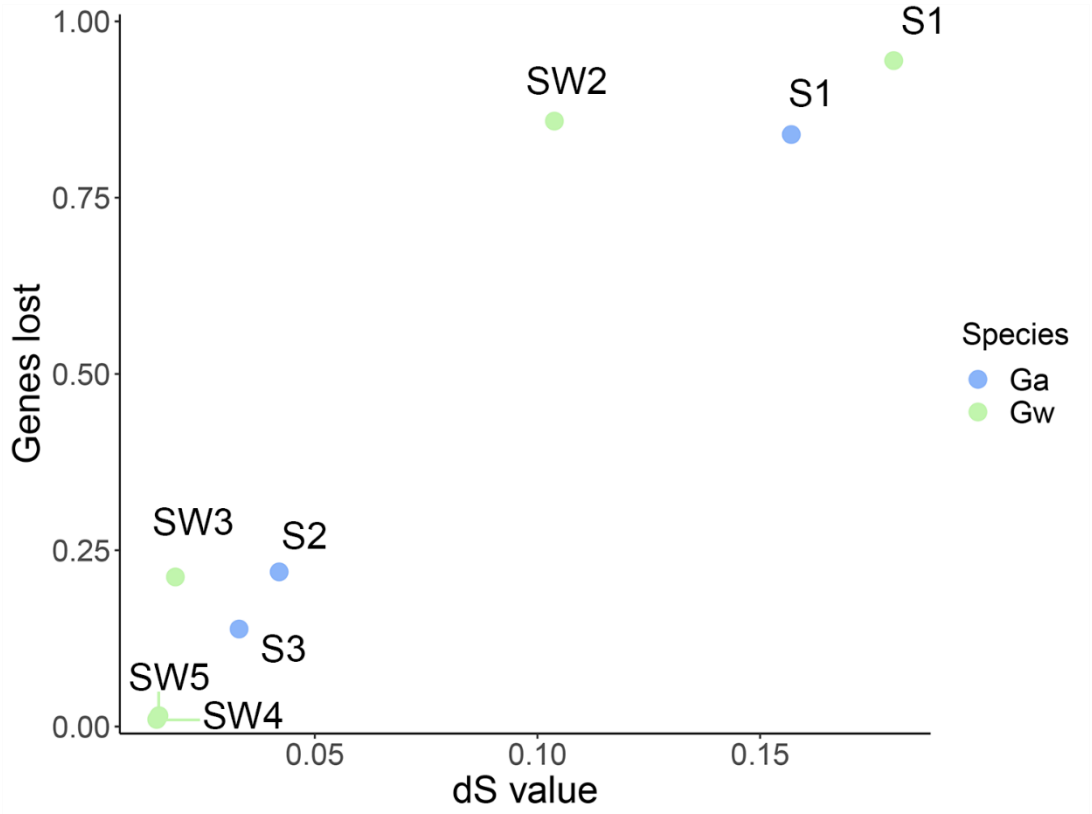


Fig 4. Proportions of gene loss in each strata. The blue dots represent the three strata in *G. aculeatus*, while the green dots represent the five strata in *G. wheatlandi*. The x-axis shows the dS value, while the y-axis shows the proportion of genes lost on the Y chromosome in the corresponding species.

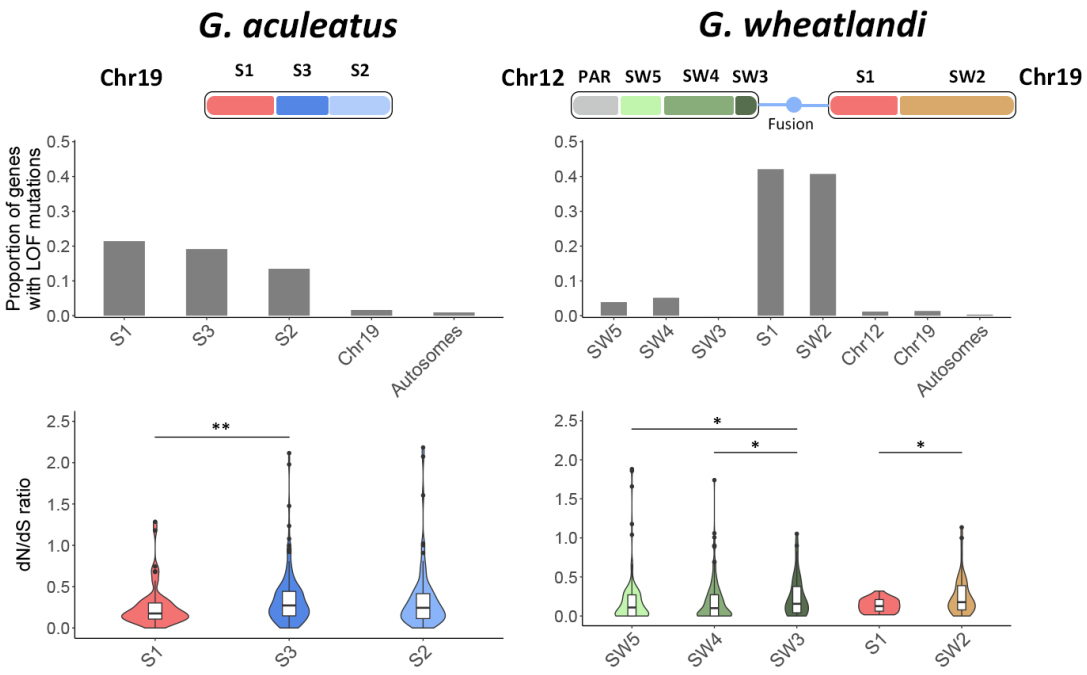


Fig 5. Molecular evolution of single-copy genes on Y chromosomes. The top panel shows the proportion of genes on Y chromosomes with LOF mutations, while the bottom panel shows the distribution of dN/dS ratios between X and Y copies of genes with no LOF mutation in each strata. Comparisons with asterisks indicate statistical significance (*p < 0.01; **p < 0.001).

Table 1. Genetic diversity in autosomes, X chromosomes, and each stratum on the Y chromosomes in *G. aculeatus* and *G. wheatlandi*. Unbiased genetic diversity was calculated with four-fold degenerate sites only.

Stratum	Genetic diversity	Proportion relative to the average of autosomes
	G. aculeatus	
Autosomes	4.13E-03	NA
Chr19	2.57E-03	62.16%
S1	1.03E-03	24.98%
S2	4.16E-04	10.07%
S3	4.75E-04	11.50%
	G. wheatlandi	
Autosomes	2.37E-03	NA
Chr12	2.14E-03	90.31%
Chr19	1.80E-03	75.86%
S1	1.10E-04	4.63%
SW2	8.63E-05	3.63%
SW3	2.83E-06	0.12%
SW4	3.26E-05	1.37%
SW5	3.49E-05	1.47%

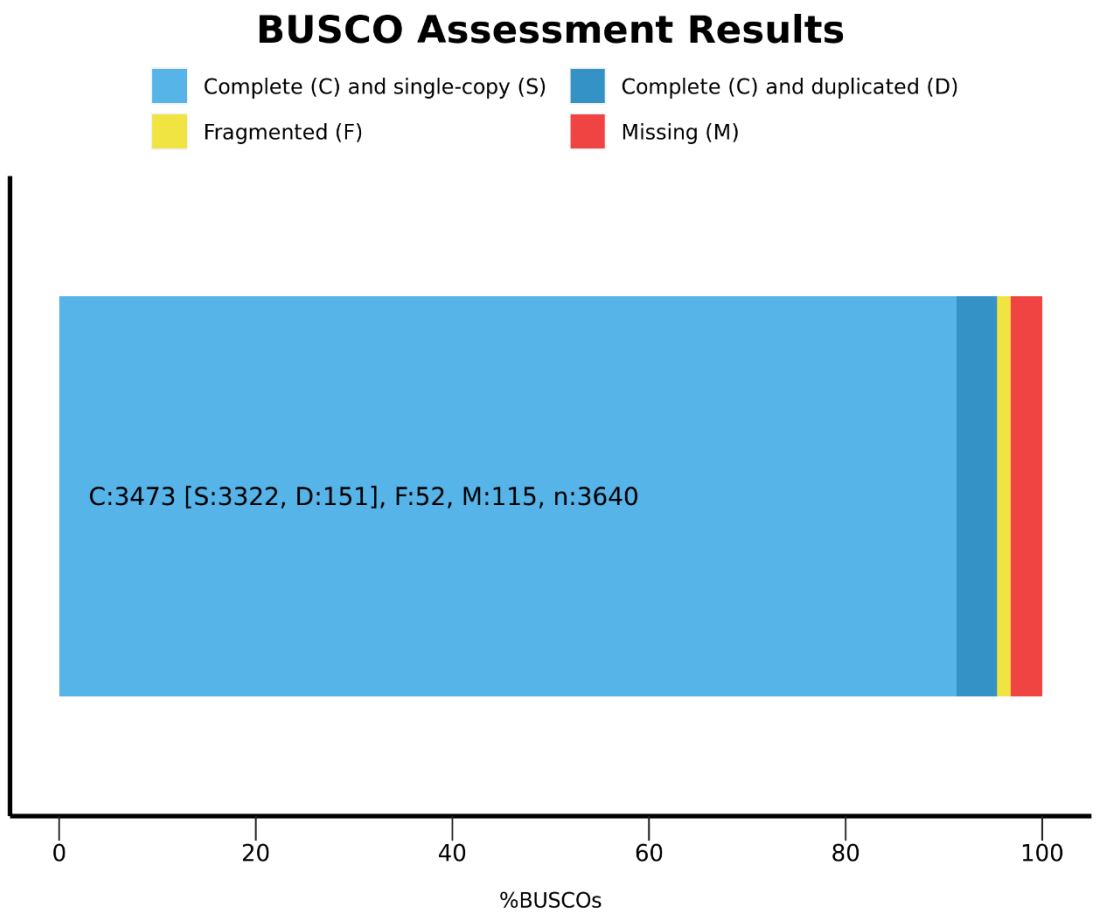


Fig S1. BUSCO score of *G. wheatlandi*. The assembly includes autosomes, X chromosomes, Y chromosome and unplaced scaffolds.

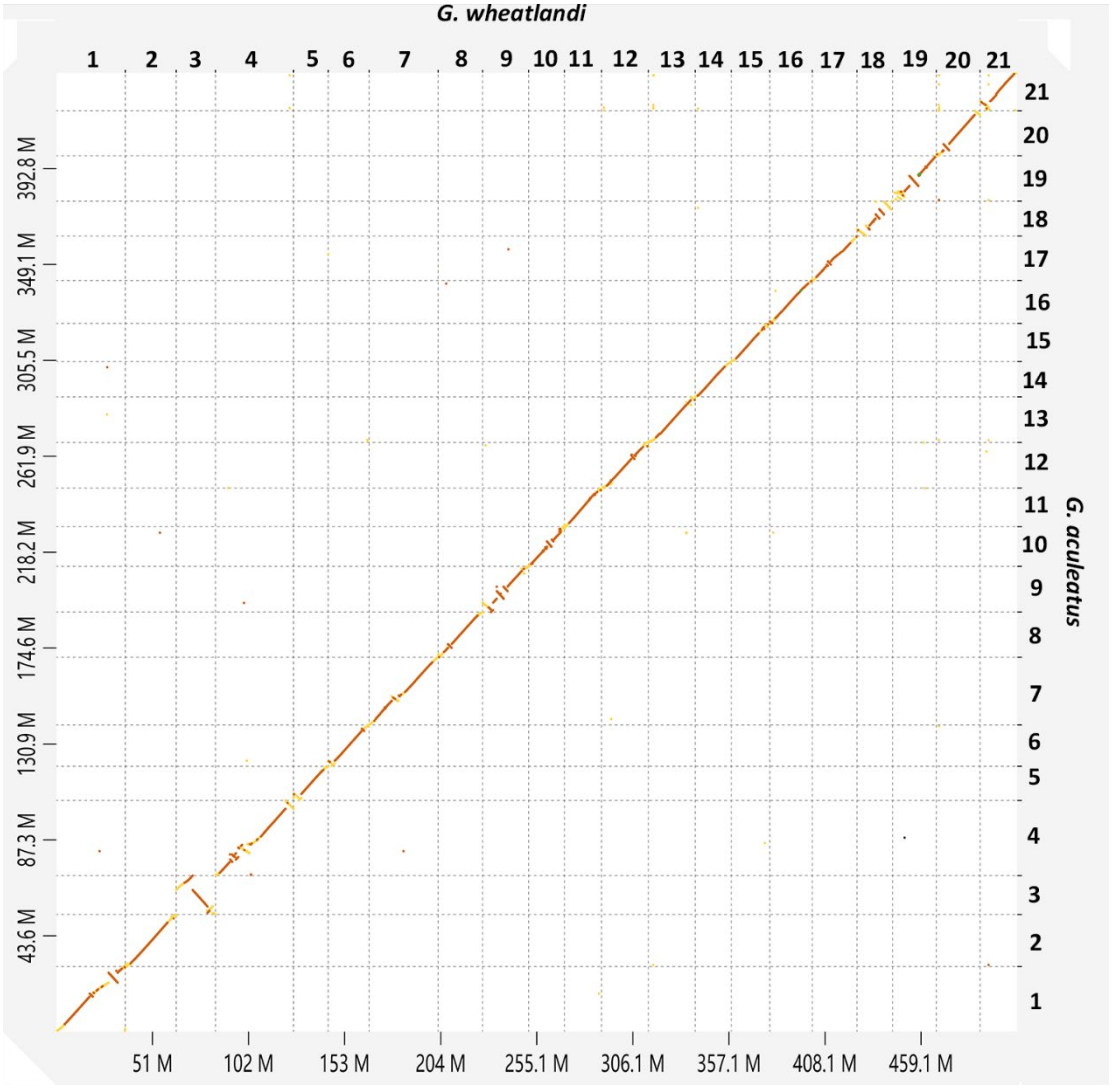


Fig S2. Synteny map of autosomal genome assemblies between *G. aculeatus* and *G. wheatlandi*.

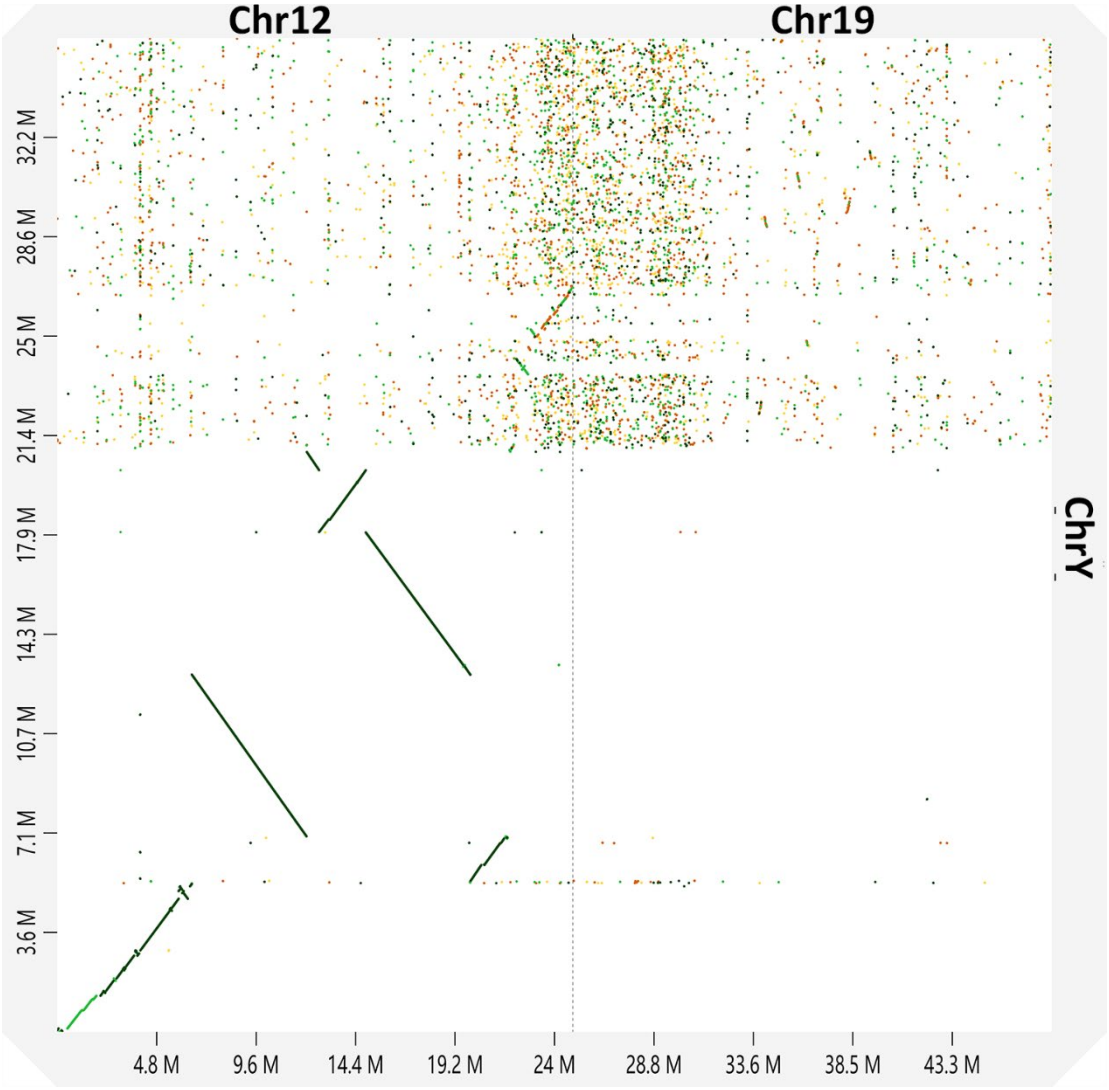


Fig S3. Synteny map of genome assemblies between X and Y chromosomes in *G. wheatlandi*.

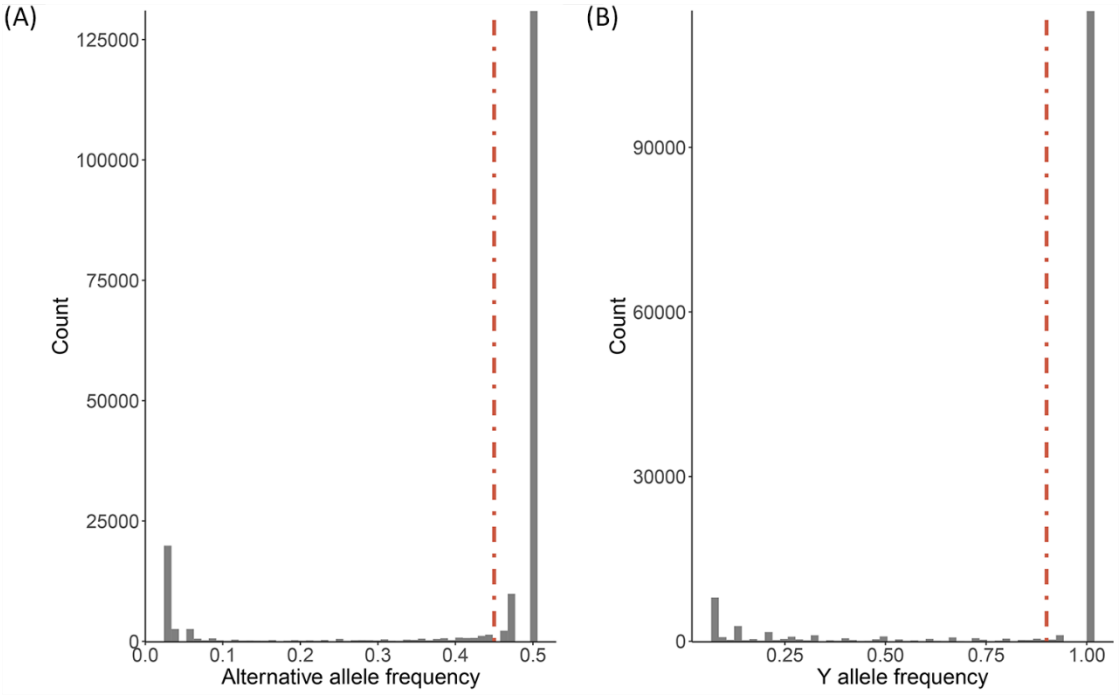


Fig S4. (A) Distribution of the allele frequency of male-specific SNPs. Red dotted line represents the cut-off for identifying fixed SNPs in males. (B) Distribution of the allele frequency of SNPs on phased Y chromosome. Red dotted line represents the cut-off for identifying fixed SNPs on Y chromosome.

Table S1. Sample information and accession numbers for sequencing data in this study.

Species	Population	Females	Males	Latitude	Longitude	Source	Supplier	Tissue	Data type
Gasterosteus aculeatus	Lake Washington	13	11	NA	NA	Wild-caught	Peichel lab	fin clips	DNA-seq
Gasterosteus aculeatus	Puget Sound	14	6	NA	NA	Wild-caught	Peichel lab	fin clips	DNA-seq
Gasterosteus aculeatus	Paxton Lake	0	1	NA	NA	Lab-reared	Peichel lab	blood	Y chromosome assembly
Gasterosteus aculeatus	Canal Lake, Nova Scotia, Canada and Humber Arm, Newfoundland, Canada	4	0	NA	NA	Wild-caught	Dalziel lab	fin clips	DNA-seq
Gasterosteus wheatlandi	Canal Lake, Nova Scotia, Canada and York Harbour, Newfoundland, Canada	0	15	NA	NA	Wild-caught	Dalziel lab	fin clips	DNA-seq
Gasterosteus wheatlandi	F1 interspecies hybrids	15	15	NA	NA	Lab-reared	Peichel lab	embryos	DNA-seq
Gasterosteus wheatlandi	Rainbow Haven Beach, Nova Scotia, Canada	4	4	44.654857	-63.421126	Wild-caught	Dalziel lab	fin clips	DNA-seq
Gasterosteus wheatlandi	Rainbow Haven Beach, Nova Scotia, Canada	4	4	44.654857	-63.421126	Wild-caught	Dalziel lab	brain	RNA-seq
Gasterosteus wheatlandi	Canal Lake, Nova Scotia, Canada	0	1	44.498298	-63.902046	Lab-reared	Peichel lab	blood	PacBio
Gasterosteus wheatlandi	Canal Lake, Nova Scotia, Canada	0	1	44.498298	-63.902046	Lab-reared	Peichel lab	liver	DNA-seq
Gasterosteus wheatlandi	Canal Lake, Nova Scotia, Canada	0	1	44.498298	-63.902046	Lab-reared	Peichel lab	liver	DNA-seq
Gasterosteus wheatlandi	Rainbow Haven Beach, Nova Scotia, Canada	0	1	NA	NA	NA	Peichel lab	NA	Genome assembly

Sequence method	Sequence platform	Sequence location	Publication	Accession number
Illumina DNA Truseq	300 cycles Illumina NextSeq	Georgia Genomics and Bioinformatics Core	Shanfelter et al. 2019	NCBI Sequence Read Archive SRP137809
Illumina DNA Truseq	300 cycles Illumina NextSeq	Georgia Genomics and Bioinformatics Core	Shanfelter et al. 2019	NCBI Sequence Read Archive SRP137809
NA	NA	NA	Peichel et al. 2020	https://stickleback.genetics.uga.edu/downloadData/
Illumina DNA Truseq	150bp paired end Illumina NovaSeq S2	University of Bern Next Generation Sequencing Platform	Sardell et al. 2021	NCBI Sequence Read Archive PRJNA742065
Illumina DNA Truseq	150bp paired end Illumina NovaSeq S2	University of Bern Next Generation Sequencing Platform	Sardell et al. 2021	NCBI Sequence Read Archive PRJNA742065
Illumina DNA Truseq	150bp paired end Illumina NovaSeq S2	University of Bern Next Generation Sequencing Platform	Sardell et al. 2021	NCBI Sequence Read Archive PRJNA742065
Illumina DNA Truseq	150bp paired end Illumina NovaSeq S1	University of Bern Next Generation Sequencing Platform	this study	
Illumina mRNA Truseq	150bp paired end Illumina HiSeq3000	University of Bern Next Generation Sequencing Platform	Liu et al. 2022	NCBI Sequence Read Archive PRJNA746773
SMRTbell express	PacBio Sequel 4 SMRT flow cells	University of Bern Next Generation Sequencing Platform	this study	
PHASE genomics proximo Hi-C	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	
PHASE genomics proximo Hi-C	150bp paired end Illumina NovaSeq SP	University of Bern Next Generation Sequencing Platform	this study	
NA	NA	NA	this study	

Notes
Same individuals used for dna and rna-seq
Same individuals used for dna and rna-seq
Brother of male used for pacbio
Brother of male used for pacbio

Table S2. Proportion of single-copy genes lost in each strata.

Stratum	Number of single-copy genes retained	Number of single-copy genes	Proportion of genes lost
		G. aculeatus	
S1	56	349	0.839541547
S2	89	114	0.219298246
S3	162	188	0.138297872
		G. wheatlandi	
S1	19	342	0.944444444
SW2	54	382	0.858638743
SW3	52	66	0.212121212
SW4	293	296	0.010135135
SW5	257	261	0.01532567

General Discussion

Examining the role of chromosomal fusions and inversions in adaptation: separating natural selection from other evolutionary processes

Many studies have shown that adaptive loci or genomic islands are prevalent throughout the genome, and chromosomal rearrangements are thought to play a significant role in their formation (Yeaman 2013). Among different types of chromosomal rearrangements, inversions have been extensively studied, but much less is known about the role of chromosomal fusions. In Chapter 1, I identified two chromosomes that have undergone chromosomal fusion in the *Gasterosteus* lineage and have a significant abundance of adaptive loci. However, certain questions remain unanswered, particularly regarding the formation of adaptive clusters, which may arise from either de novo mutations after the fusions occur or by bringing together pre-existing loci due to the locally and dramatically altered recombination landscape created by the fusions. Based on my comparative analysis of *G. aculeatus* with outgroup species, I concluded that de novo mutations are more likely to become fixed within the fusion in sticklebacks. Nevertheless, the possibility of enrichment through physical linkage of pre-existing adaptive loci cannot be entirely ruled out.

Accurate identification of sites under selection is essential to address questions related to the role of chromosomal rearrangements in the formation of adaptive clusters. However, the challenge lies in identifying these sites due to their genetic linkage with other regions where recombination is suppressed. With recent advancements in long-read sequencing technology, the task has become feasible through the use of complete haploid assembly as a foundation. The breakthrough in linked-read sequencing allows for the phasing of entire chromosomes, enabling comparisons between multiple populations at a reasonable cost (Meier et al. 2021). Cross-population comparisons can target specific regions within chromosomal rearrangements, as the fixation of neutral sites in the surrounding areas is likely to be random. Although sample requirements are strict and may not be applicable to all species, our *G. aculeatus* study species, with its hundreds of instances of isolation and colonization from marine to freshwater habitats, is ideal for this purpose (Wootton 1976; Bell and Foster 1994; Jones et al. 2012; Roberts Kingman et al. 2021).

In addition to the bioinformatics aspect, it is essential to evaluate how chromosomal rearrangements contribute to adaptation in nature. Recent advances in CRISPR technology

have facilitated the manipulation of genomes, providing an opportunity to investigate the role of chromosomal inversions and fusions in adaptation. A potential experiment involves "flipping" an inversion, resulting in a region that still has a heterozygous allelic content but with restored recombination. Similarly, for chromosomal fusion, breaking the fused chromosome in a species with adaptive traits and observing the outcome after recombination is recovered is a feasible strategy. Comparable experiments have been successfully conducted in maize and Arabidopsis (Schmidt et al. 2020; Schwartz et al. 2020; Angelopoulou et al. 2022; Rönspies et al. 2022). The ultimate objective is to recover recombination in regions where it is suppressed by chromosomal rearrangements through genetic engineering and to assess the effects on phenotypic evolution and adaptation. The development of new methods for both sequencing and genetic engineering means that it is an opportune time to investigate the evolutionary implications of these chromosomal rearrangements in nature.

Revisiting the relationship between chromosomal rearrangements and recombination suppression on sex chromosomes: could rearrangements be a consequence, rather than a cause?

Chromosomal rearrangements, particularly fusions and inversions, have long been recognized as drivers of recombination suppression (Charlesworth et al. 2005). While my thesis focused on these two types of rearrangements, in Chapter 1, I specifically investigated the effect of fusion on recombination reduction by comparing genetic diversity between fused and unfused chromosomes as well as between taxa with fused and unfused chromosomes. While other studies have demonstrated or discussed that heterozygous chromosomal inversions can also lead to recombination suppression in autosomes (Coyne et al. 1991; Kirkpatrick 2010; Farré et al. 2013; Fishman et al. 2013; Lundberg et al. 2017), there has been less empirical research on the impact of such inversions on sex chromosomes.

Nonetheless, chromosomal inversions have been considered play a significant role in the formation of evolutionary strata on sex chromosomes (Wang et al. 2012; Jay et al. 2022; Olito et al. 2022; Olito and Abbott 2023). While chromosomal inversions have been identified as a major factor, it is important to note that not all identified strata are associated with the presence of a chromosomal inversion between two sex chromosomes (Furman et al. 2020). Other factors, such as transposable elements (Ponnikas et al. 2018) and epigenetic changes (Zhang et al. 2008; Metzger and Schulte 2018), are also predicted to contribute to the

formation of evolutionary strata. It is also possible that sex chromosomes can evolve in regions of low recombination, in the absence of inversions. For example, in seahorses, chromosomal inversions were not detected in the low-recombination regions of sex chromosomes. Instead, the low recombination rate in this region may represent the ancestral state, as homologous regions in other species span centromeres (Long et al. 2023). Similarly, recent studies of fungi have shown that the suppression of recombination on sex chromosomes is an ancestral state, preceding the divergence of species. Yet, no large inversion has been found related to the sex-linked regions (Branco et al. 2017; Carpentier et al. 2019). Despite the fact that inversions are not the only mechanism that can cause recombination suppression on sex chromosomes, in Chapters 2 and 3, I identified sex-linked chromosomal inversions in multiple species, which are associated with evolutionary strata, suggesting they played a role in recombination suppression between the X and Y chromosomes of these species.

However, it is also possible that these inversions are not the driver but a consequence of recombination suppression. Inversions can be accumulated in regions of low recombination for several reasons: 1) the fixation rate of a structural variation, such as inversion, is higher in regions of low recombination for the same reasons that mutations accumulate on sex chromosomes; 2) the accumulation of TE in regions of low recombination potentially facilitates the formation of inversions; and 3) chromosomal inversions may help to shelter deleterious mutations that are accumulated in regions of relatively low recombination rate from purifying selection by further reducing the recombination rate. A recent theoretical study also predicted that the sheltering of mutation load can explain the extension of evolutionary strata on Y chromosomes (Jay et al. 2022). The establishment of a chromosomal inversion can further strengthen the existing divergent selection, protecting them from being reversed by any rare recombining events. These alternative hypotheses proposing that inversions result from recombination suppression provide a more straightforward explanation for the prevalence of inversions on sex chromosomes compared to the previously held notion that inversions are necessary for recombination suppression. These alternative hypotheses do not rely on the low probability of capturing a permanent heterozygous site for a chromosomal inversion. Instead, the inversion can reinforce the existing positive feedback loop in the evolution of sex chromosomes, requiring fewer assumptions and coincidences.

In order to evaluate the proposed hypotheses, further chromosome-level assemblies of sex chromosomes are required. If an inversion event has captured a sex determination gene, it is possible to estimate the divergence time between the sex determination gene and the remaining part of the inversion. This can be done by comparing the homologous regions from closely-related taxa. Additionally, by comparing with a homologous region from a closely-related taxa, one could infer the ancestral recombination landscape, which can provide information on when the suppression of recombination occurred. By doing this, it is possible to infer the evolutionary path of the formation of a new sex chromosome.

Recent advancements in sequencing technology and computational tools have enabled the investigation of large structural variations in genomes

Over the past few decades, research on genome evolution has primarily centered around SNPs due to their ease of quantification with high-throughput sequencing data (Wellenreuther et al. 2019). In contrast to SNPs, structural variations have received less attention in the study of genome evolution due to the limited detection power of short-read sequencing. Nonetheless, structural variations are considered to have more significant effects on genomic evolution, as they involve a larger number of affected base pairs. Recent research has highlighted the crucial role of structural variations in adaptive evolution and sex chromosome evolution (Wellenreuther and Bernatchez 2018; Wellenreuther et al. 2019; Huang and Rieseberg 2020; Orteu and Jiggins 2020). This thesis presents evidence that chromosomal fusions can have a global impact on the recombination landscape, leading to the formation of adaptive clusters (Chapter 1) and that chromosomal inversions play a critical role in the step-wise evolution of sex chromosomes (Chapter 2 and 3).

The field of sequencing technologies has recently undergone a period of explosive growth, with the emergence of new technologies enabling researchers to address questions related to structural variations. For instance, third-generation sequencing technologies such as PacBio and Nanopore have made it feasible to generate chromosome-level assemblies, facilitating the detection of chromosomal rearrangements at various scales. Additionally, the development of linked-read sequencing has overcome the limitations of traditional short-read sequencing (Fang et al. 2019), enabling the detection of structural variation at a population level at a low cost. Chapter 1 details the successful generation of a chromosome-level assembly and identification of two chromosomal fusions in sticklebacks, which

constitutes critical evidence of adaptation through large structural variation. Chapter 2 employs the latest haplotagging technology to detect two inversions on X and Y chromosomes in various *A. quadracus* populations, providing new insights into the evolution of sex chromosomes.

The development of new sequencing technologies has created opportunities to investigate the effects of chromosomal rearrangements in genomic evolution. However, to fully exploit the potential of these technologies, it is crucial to assemble genomes correctly. While genome assembly pipelines have become more mature, assembling sex chromosomes completely remains challenging. Previous studies have been limited to fragmented pieces of sex chromosomes due to the inability to recover the entire chromosome using short reads. Although PacBio and Nanopore sequencing offer potential solutions, generating complete sex chromosome assemblies remains difficult because current approaches aim to generate merged haploid assemblies. Alternative approaches have been developed, such as sequencing YY homozygous individuals or sperm cells, but these are species-specific and time-consuming. In Chapter 3, I developed a new pipeline that combines population data, PacBio sequencing, and HiC reads to generate a complete assembly of the Y chromosome of *G. wheatlandi*. This pipeline does not require complicated experimental designs and instead fully utilizes the variance in existing data and phasing methods. Thus, it can be applied to both model and non-model species, expanding the ability to study sex chromosome evolution across different taxa. In summary, with the maturation of sequencing technologies, now is an opportune time to investigate the effects of chromosomal rearrangements on genomic evolution.

References

Angelopoulou A, Papaspyropoulos A, Papantonis A, Gorgoulis VG. 2022. CRISPR-Cas9-mediated induction of large chromosomal inversions in human bronchial epithelial cells. *STAR Protoc.* 3:101257.

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

Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, Siguenza S, Brandenburg J-T, Coelho MA, Hood ME, et al. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc. Natl. Acad. Sci.* 114:7067–7072.

Carpentier F, Rodríguez de la Vega RC, Branco S, Snirc A, Coelho MA, Hood ME, Giraud T. 2019. Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi. *Genome Res.* 29:944–953.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128.

Coyne JA, Aulard S, Berry A. 1991. Lack of underdominance in a naturally occurring pericentric inversion in Drosophila melanogaster and its implications for chromosome evolution. *Genetics* 129:791–802.

Fang L, Kao C, Gonzalez MV, Mafra FA, Pellegrino Da Silva R, Li M, Wenzel S-S, Wimmer K, Hakonarson H, Wang K. 2019. LinkedSV for detection of mosaic structural variants from linked-read exome and genome sequencing data. *Nat. Commun.* 10:5585.

Farré M, Micheletti D, Ruiz-Herrera A. 2013. Recombination rates and genomic shuffling in human and chimpanzee—a new twist in the chromosomal speciation theory. *Mol. Biol. Evol.* 30:853–864.

Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP. 2013. Chromosomal rearrangements and the genetics of reproductive barriers in Mimulus (monkey flowers). *Evolution* 67:2547–2560.

Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank JE. 2020. Sex Chromosome Evolution: So Many Exceptions to the Rules. *Genome Biol. Evol.* 12:750–763.

Huang K, Rieseberg LH. 2020. Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants. *Front. Plant Sci.* 11:296.

Jay P, Tezenas E, Véber A, Giraud T. 2022. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLOS Biol.* 20:e3001698.

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

Kirkpatrick M. 2010. How and Why Chromosome Inversions Evolve. *PLoS Biol.* 8:e1000501.

Long X, Charlesworth D, Qi J, Wu R, Chen M, Wang Z, Xu L, Fu H, Zhang X, Chen X, et al. 2023. Independent Evolution of Sex Chromosomes and Male Pregnancy–Related Genes in Two Seahorse Species. *Mol. Biol. Evol.* 40:msac279.

Lundberg M, Liedvogel M, Larson K, Sigeman H, Grahn M, Wright A, Åkesson S, Bensch S. 2017. Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evol. Lett.* 1:155–168.

e er a a ar a a e réau A, Aldás I, Box Power O, Nadeau NJ, Bridle JR, Rolian C, et al. 2021. Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proc. Natl. Acad. Sci.* 118:e2015005118.

Metzger DCH, Schulte PM. 2018. The DNA Methylation Landscape of Stickleback Reveals Patterns of Sex Chromosome Evolution and Effects of Environmental Salinity. *Genome Biol. Evol.* 10:775–785.

Olito C, Abbott JK. 2023. The evolution of suppressed recombination between sex chromosomes and the lengths of evolutionary strata. *Evolution* 77:1077–1090.

Olito C, Ponnikas S, Hansson B, Abbott JK. 2022. Consequences of partially recessive deleterious genetic variation for the evolution of inversions suppressing recombination between sex chromosomes. *Evolution* 76:1320–1330.

Orteu A, Jiggins CD. 2020. The genomics of coloration provides insights into adaptive evolution. *Nat. Rev. Genet.* 21:461–475.

Ponnikas S, Sigeman H, Abbott JK, Hansson B. 2018. Why Do Sex Chromosomes Stop Recombining? *Trends Genet.* 34:492–503.

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

Rönspies M, Schmidt C, Schindele P, Lieberman-Lazarovich M, Houben A, Puchta H. 2022. Massive crossover suppression by CRISPR–Cas-mediated plant chromosome engineering. *Nat. Plants* 8:1153–1159.

Schmidt C, Fransz P, Rönspies M, Dreissig S, Fuchs J, Heckmann S, Houben A, Puchta H. 2020. Changing local recombination patterns in Arabidopsis by CRISPR/Cas mediated chromosome engineering. *Nat. Commun.* 11:4418.

Schwartz C, Lenderts B, Feigenbutz L, Barone P, Llaca V, Fengler K, Svitashev S. 2020. CRISPR–Cas9-mediated 75.5-Mb inversion in maize. *Nat. Plants* 6:1427–1431.

Wang J, Na J-K, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W, et al. 2012. Sequencing papaya X and Y ^{h} chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl. Acad. Sci.* 109:13710–13715.

Wellenreuther M, Bernatchez L. 2018. Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends Ecol. Evol.* 33:427–440.

Wellenreuther M, Mérot C, Berdan E, Bernatchez L. 2019. Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Mol. Ecol.* 28:1203–1209.

Wootton RJ. 1976. The Biology of the Sticklebacks. Lodon: Academic Press

Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci.* 110:E1743–E1751.

Zhang W, Wang X, Yu Q, Ming R, Jiang J. 2008. DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Res.* 18:1938–1943.

Acknowledgements

I would like to express my sincere gratitude to everyone who has supported me throughout my academic journey and has contributed to the completion of this thesis. Firstly, I am grateful to UniBern for providing me with the opportunity to pursue my research interests. I am also thankful to the Institute of Ecology and Evolution and the scientific communities that have provided me with guidance and helped me to grow both professionally and personally. I am especially grateful to my PhD students and Postdoc colleagues, who have been my constant support system through daily communication and discussions.

I would like to extend my gratitude to Professor Qi Zhou, my external examiner, for his valuable comments and for attending my defense. It is amazing how our paths crossed at ESEB, and I am honored that you agreed to be my external examiner. Also, I would like to thank Claudia Bank for serving as the chair of my PhD defense.

I extend my heartfelt appreciation to the Peichel Lab for their unwavering support throughout my research journey. I am particularly grateful to Nicole Nesvadba, Melanie Hilterbrunner, and Verena Saladin for their immense help in the lab and fish room. Special thanks to Melanie Hilterbrunner for her time-consuming efforts in making all the libraries. I would also like to express my gratitude to Jun Kitano's Lab for showing interest in my project and providing valuable comments, and for the enjoyable time we had in Prague. My sincere thanks to David Marques for his insightful feedback on my paper and his assistance with the o e n a a o than to ran n an Chan an are a or the r collaboration on the Haplotagging technology and their valuable inputs to my draft.

I wish to express my heartfelt gratitude to some individuals who have played an integral role in the success of my PhD journey. Firstly, I extend my deepest appreciation to my advisor and mentor, Catherine Peichel, for her invaluable guidance, unyielding support, and dedication to the field of biology. She is a remarkable scientist who has inspired me with her brilliance, passion, and unwavering commitment to excellence. Her unwavering encouragement and steadfast support throughout my doctoral studies have been instrumental in building my confidence and shaping my research interests. Moreover, her significant contributions to the scientific community serve as a testament to her exceptional leadership and dedication to science. Also, I am thankful to Sam Yeaman from University of

Calgory to be my external supervisor and for the comments on my paper and the encouragement about my future career.

Particularly, I am indebted to my colleagues and friends, Marius Roesti and Dan Jeffries, for their unwavering faith in my abilities and constant support during my times of stress and anxiety. Their scientific insights and approach to life have been an inspiration to me. I am also grateful to Matt Zuellig for sharing his knowledge on CRISPR and engaging in insightful discussions with me. Additionally, I would like to express my appreciation to Mike White for his invaluable contribution to my sex chromosome project. Lastly, I am thankful to Bo and Devi for their support and welcoming me as a dog person in our lab parties, meetings, and social events hosted by Katie.

Finally, I would like to express my gratitude to the friends and family members who have been instrumental in keeping me motivated and grounded throughout this journey. I am particularly thankful to Xuejing Wang, whose constant support and insightful discussions have been invaluable throughout my PhD career. I am also grateful to Xiangyi Li Richter for her guidance on pursuing a career in science and her warm hospitality. Furthermore, I want to thank Yifan Yao, Yang Zhou, Allen Xingyong Zhang, and Mengyou Zhang for always welcoming me whenever and wherever we meet. Additionally, I would like to thank Samuel Tonini and Andrea von Gunten for offering me a free dancing studio where I could relieve my stress and anxiety. Finally, I want to express my heartfelt appreciation to my parents, whose unwavering faith in me and support for my decisions have been a source of strength throughout my life.

Life is like a journey, and during my PhD journey, I've learned how to identify and solve problems, as well as how to better understand and relate to both myself and those around me. It's been both an outward exploration and an inward journey of the soul. Completing my PhD is not an ending, but rather a revelation of a whole new beginning. I want to express my heartfelt gratitude to all the amazing people in my life who have supported me along the way.

Declaration of consent

on the basis of Article 18 of the PromR Phil.-nat. 19

Name/First Name:

Registration Number:

Study program:

Bachelor ☐ Master ☐ Dissertation ☐

Title of the thesis:

Supervisor:

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of September 5th, 1996 and Article 69 of the University Statute of June 7th, 2011 is authorized to revoke the doctoral degree awarded on the basis of this thesis.
For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with theses submitted by others.

Place/Date

Signature