Article: Discoveries

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

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. aculeatus* 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.*

193 *pungitius*, it has fused to different chromosomes in these species (Fig. 2 and Supplementary Fig. S4),
194 again suggesting independent fusions have occurred in the two lineages. Together, these data
195 demonstrate that chromosomal fusions have occurred in the two stickleback lineages that include
196 species (*G. aculeatus* and *P. pungitius*) able to inhabit both marine and freshwater habitats, raising the
197 possibility that such fusions have contributed to the ability of these species to adapt to divergent
198 habitats in the face of gene flow.
199
200 **Enrichment of marine-freshwater QTL on chromosomes 4 and 7 in *G. aculeatus***
201 If fusions facilitate adaptation by linking adaptive alleles, we would predict that an increased number
202 of QTL underlying adaptive traits would map to the fused chromosomes, and that the these QTL would
203 have congruent effects in the expected direction (i.e. a marine allele confers a marine phenotype and
204 vice versa) on multiple traits. Thus, we tested whether there was an enrichment of QTL with effects in
205 the expected direction on *G. aculeatus* chromosomes 4 and 7 using a database of QTL identified in
206 crosses between marine and freshwater populations (Peichel and Marques 2017). Indeed, we found
207 that chromosomes 4 and 7, as well as chromosomes 16, 20, and 21, have significantly more QTL with
208 effects in the expected direction than other chromosomes, accounting for variation in either the length
209 of chromosomes or the number of genes on the chromosomes (Fig. 3 and Supplementary Table S2).
210 Chromosome 21 has an inversion that is polymorphic within *G. aculeatus*, which is one of the strongest
211 signals of divergence between worldwide marine and freshwater populations (Jones et al. 2012; Roesti
212 et al. 2015; Fang et al. 2020; Magalhaes et al. 2021; Roberts Kingman et al. 2021). Although there are
213 no apparent large-scale chromosomal rearrangements between marine and freshwater populations
214 associated with chromosomes 16 or 20, the adaptive clusters on chromosomes 4, 7 and 21 are
215 associated with chromosomal rearrangements that might facilitate linkage of adaptive traits.
216
217 **No enrichment of gene transpositions or gene duplications on chromosomes 4 and 7**
218 It has also been proposed that such adaptive clusters could form via small-scale genomic
219 rearrangements, such as transposition of single genes and/or gene duplications (Yeaman 2013). We
220 therefore examined the distribution of gene duplication and gene transposition events in *G. aculeatus*
221 relative to *P. pungitius*, *A. quadracus*, and *A. flavidus*. There were too few gene transposition events to
222 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 Washington freshwater populations. Based on ΔAIC values, the best model has a bottleneck 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, there is an overall reduction in recombination on these two chromosomes relative to 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). Thus, 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.

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).

493 The prediction and annotation of genes were conducted on the repeat-masked genome assembly
494 with the Maker2 (v. 2.31.10) pipeline (Holt and Yandell 2011), including four rounds of annotation. In
495 the first round, the transcriptome assembly generated by Trinity and protein data from *Danio rerio*, *G.*
496 *aculeatus*, *P. pungitius*, *Takifugu flavidus,* and the Uniprot database (UniProt Consortium 2015) were
497 used as evidence for the program. The second round of annotation included two training and
498 prediction steps by AUGUSTUS (v. 3.2.3) (Stanke et al. 2008) and SNAP (Korf 2004). The results were
499 then passed to MAKER2. For the third round annotation, GeneMARK-ES (Ter-Hovhannisyan et al. 2008)
500 was combined with MAKER2. Finally, the second round annotation was repeated with the resulting
501 files from the third round. The final annotation was checked based on annotation edit distance (AED),
502 and only annotations with AED score 0.5 or less were retained for downstream analysis. Functional
503 annotation was conducted by eggnog-mapper (v2) (Huerta-Cepas et al. 2017).
504
505 **Genomic synteny analyses and detection of rearrangements between species**
506 Synteny analyses were conducted in two ways. First, Mummer4 and nucmer (Marçais et al. 2018) were
507 used to compare the order of genes between *G. aculeatus* and *A. quadracus* on *G. aculeatus*
508 chromosomes 4 and 7. Alignments shorter than 2000bp with an identity less than 85% were removed.
509 Second, non-redundant coding sequence sets from four species (*G. aculeatus*, *A. quadracus*, *P.*
510 *pungitius* and *A. flavidus*) were used for cross synteny analysis. We used MCScan (Tang et al. 2008) in
511 JCVI package (Tang et al. 2015) to compare synteny on the chromosome level as well as the gene level.
512 *A. flavidus* was chosen as the outgroup based on the phylogeny to examine whether the reduction of
513 chromosome number in *G. aculeatus* and *P. pungitius* relative to *A. quadracus* is due to fission or fusion.
514
515 **Identification of gene transposition and duplication events**
516 To detect gene duplication and transposition events, we first extracted single-copy orthologues from
517 four species (*G. aculeatus*, *P. pungitius*, *A. quadracus*, *A. flavidus*) using OrthoFinder (v 2.3.12) (Emms
518 and Kelly 2019). For gene duplication events, we used the duplication summary from OrthoFinder and
519 focused on genes only duplicated in *G. aculeatus*; we included both intra- and inter-chromosomal
520 duplications in the analyses. For gene transposition events, we focused on inter-chromosomal gene
521 transpositions, in which a gene had moved to the focal chromosome in *G. aculeatus* from another
522 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.

583

584 **Detecting signatures of selection across the genome**

585 Scans for signatures of selection were performed between the Puget Sound marine and Lake
586 Washington freshwater populations in various ways using the dataset described above. A window-
587 based F_{ST} distribution and nucleotide diversity were calculated with Vcftools (v 0.1.16) with a window
588 size of 20kb and a window step of 10kb. To further identify selected regions, we also adopted
589 haplotype-based statistics. We first extracted mapped reads with mapping quality larger than 20 and
590 inferred haplotypes using WhatsHap (v1.0) (Martin et al. 2016) and shapeit4 (v 4.1.3) (Delaneau et al.
591 2019) with default parameters. Then, the output file was imported into the R package rehh (Gautier
592 et al. 2017) to detect soft and incomplete sweeps within populations (iHS) and to detect complete
593 sweeps that occurred in one population and not the other (XPEHH). We followed (Voight et al. 2006)
594 to calculate the proportion of extreme iHS and XPEHH values (w-iHS and w-XPEHH, the proportion of
595 |iHS| and |XPEHH| > 2) in the same 20kb overlapping windows. The sex chromosome, chromosome
596 19 , was also excluded from this analysis.

597 To examine whether particular chromosomes were enriched for these signatures of selection, we
598 compared the observed number of: 1) SNPs within genomic islands; 2) top 5% Pi outliers within each
599 population; 3) top 5% |iHS| regions of outliers within each population; and 4) top 5% XPEHH regions
600 of outliers on each chromosome to the expected numbers, given either the length of the chromosome
601 or the number of genes on the chromosome in R through a goodness-of-fit test (chisq.test).
602 Chromosomes with significantly higher values than expected were identified by standardized residuals
603 with a value larger than 3 in both comparisons (Supplementary Table S4).

604

605 **Topology weighting analyses**

606 To explore the evolutionary histories of marine and freshwater alleles on the fusion chromosomes, we
607 used a topology weighting approach. We built phylogenetic trees with the SNP dataset for the genome
608 scan in non-overlapping windows for every 50 SNPs by RaxML (v8) (Stamatakis 2006) and conducted
609 tree weighting in Twisst (Martin and Van Belleghem 2017). The analysis was performed on the two
610 fused chromosomes, chromosomes 4 and 7, separately. For comparison, we performed the analysis
611 on chromosome 1 because it is a large submetacentric chromosome with a similar length and
612 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 smallest ΔAIC (Supplementary Table S5).

Genetic diversity analysis of each chromosome in fused and unfused taxa

To explore whether fused chromosomes have a lower recombination rate, we compared genetic diversity of each chromosome 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 chromosome 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

673 publication, Anne Dalziel, Julia Wucherpfennig, Ian Heller, David Kingsley, and Arne Nolte for providing
674 stickleback samples, the University of Bern Next Generation Sequencing Platform for sequencing
675 support, and the members of the Peichel lab for discussions.
676

677 **Data availability**

678 All data used in this study were already publicly available or are available at the NCBI Sequence Read
679 Archive under project number PRJNA746773. The *A. quadracus* genome annotations are available on
680 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

769 humans. *BMC Genomics* 13:107.
770 Hoffmann AA, Rieseberg LH. 2008. Revisiting the Impact of Inversions in Evolution: From Population Genetic
771 Markers to Drivers of Adaptive Shifts and Speciation? *Annu. Rev. Ecol. Evol. Syst.* 39:21–42.
772 Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population Genomics of Parallel
773 Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet.* 6:e1000862.
774 Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-
775 generation genome projects. *BMC Bioinformatics* 12:491.
776 Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide
777 functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34:2115–2122.
778 Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN, Grossen C, Delmore KE, Alcaide M, Irwin JH. 2018.
779 A comparison of genomic islands of differentiation across three young avian species pairs. *Mol. Ecol.* 27:4839–
780 4855.
781 Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A, et al. 2019. A
782 key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* 364:886–889.
783 Jasper RJ, Yeaman S. 2020. Local adaptation can cause both peaks and troughs in nucleotide diversity within
784 populations. http://biorxiv.org/lookup/doi/10.1101/2020.06.03.132662
785 Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al.
786 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
787 Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz J, Myers RM, et al. 2009.
788 A role for a neo-sex chromosome in stickleback speciation. *Nature* 461:1079–1083.
789 Koch EL, Morales HE, Larsson J, Westram AM, Faria R, Lemmon AR, Lemmon EM, Johannesson K, Butlin RK. 2021.
790 Genetic variation for adaptive traits is associated with polymorphic inversions in *Littorina saxatilis*. *Evol. Lett.*
791 5:196–213.
792 Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat.*
793 *Biotechnol.* 37:540–546.
794 Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
795 Lenormand T, Otto SP. 2000. The Evolution of Recombination in a Heterogeneous Environment. *Genetics*
796 156:423–438.
797 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.
798 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project
799 Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.*
800 25:2078–2079.
801 Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide
802 sequences. *Bioinformatics* 22:1658–1659.
803 Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I. 2006. Mechanisms of chromosome number
804 reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci.* 103:5224–5229.
805 Magalhaes IS, Whiting JR, D'Agostino D, Hohenlohe PA, Mahmud M, Bell MA, Skúlason S, MacColl ADC. 2021.
806 Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nat. Ecol. Evol.*
807 5:251–261.
808 Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile
809 genome alignment system. *PLOS Comput. Biol.* 14:e1005944.
810 Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2016. Genomics of rapid incipient
811 speciation in sympatric threespine stickleback. *PLOS Genet.* 12:e1005887.
812 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. *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.
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.

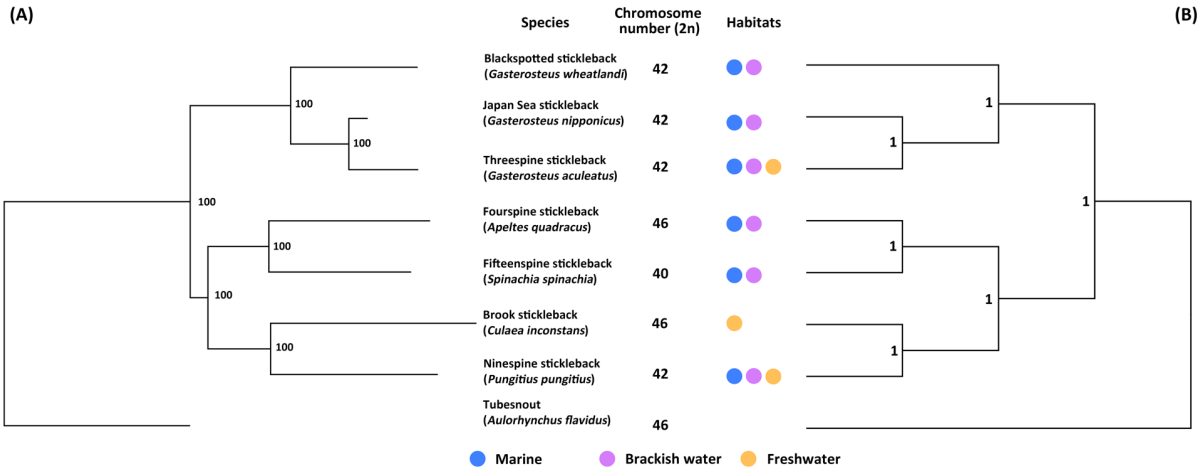
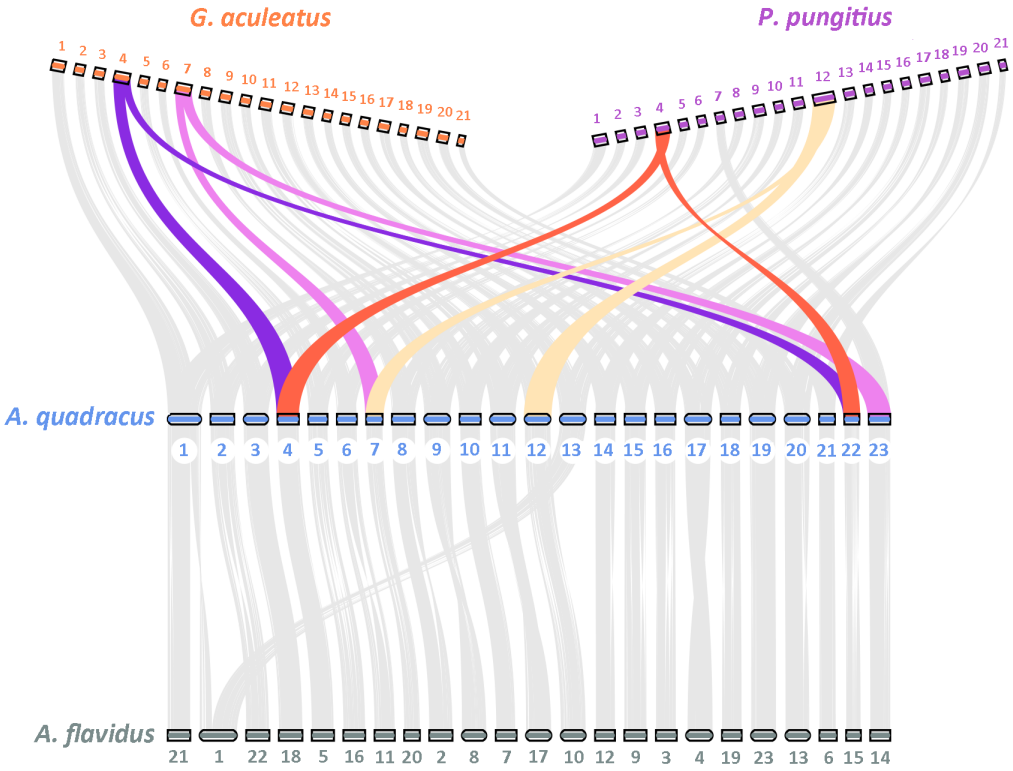


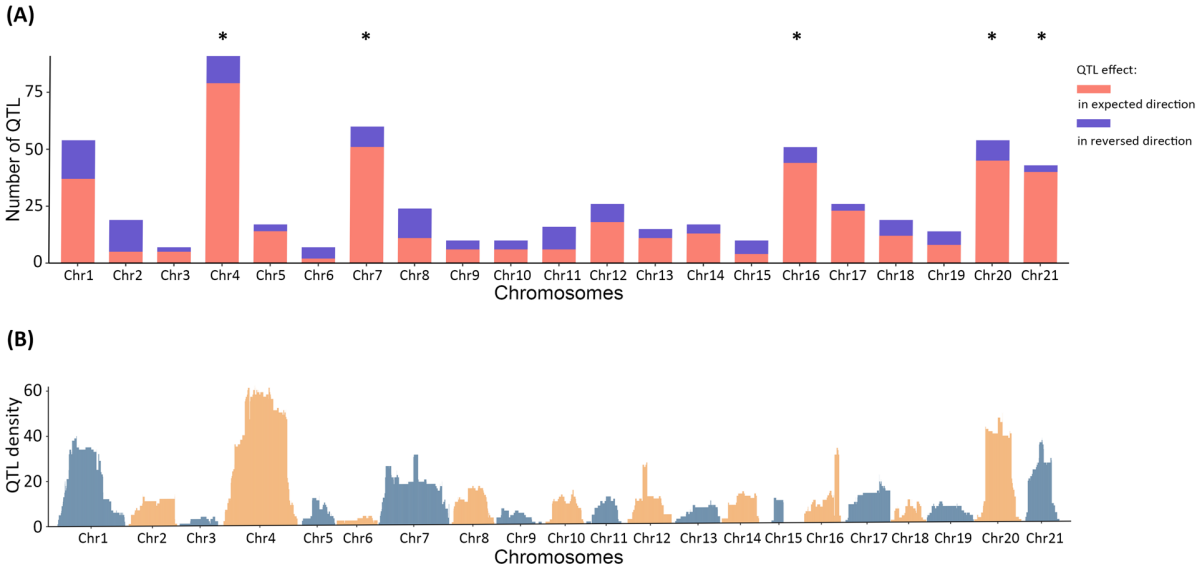
Fig. 1. Phylogeny of stickleback species and the *A. flavidus* outgroup. (A) Phylogenetic relationship among species was reconstructed in RaxML using a concatenated supermatrix of 1734 single-copy, orthologous genes. Numbers near nodes are bootstrap values. (B) Species tree was reconstructed in ASTRAL-III 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).



943

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

948



949

950 **Fig. 3.** (A) Counts of QTL underlying traits that differ between marine and freshwater populations with
951 QTL conferring an effect in the expected direction in red, and QTL conferring an effect in the reversed
952 direction in purple. (B) Density of QTL confidence intervals mapped to the *G. aculeatus* genome in
953 50kb windows. QTL data are collected from previous studies (Supplementary Table S2). Chromosomes
954 with asterisks have significantly more QTL with effects in the expected direction than expected given
955 either the number of genes on the chromosome or the chromosome length (Supplementary Table S2).
956

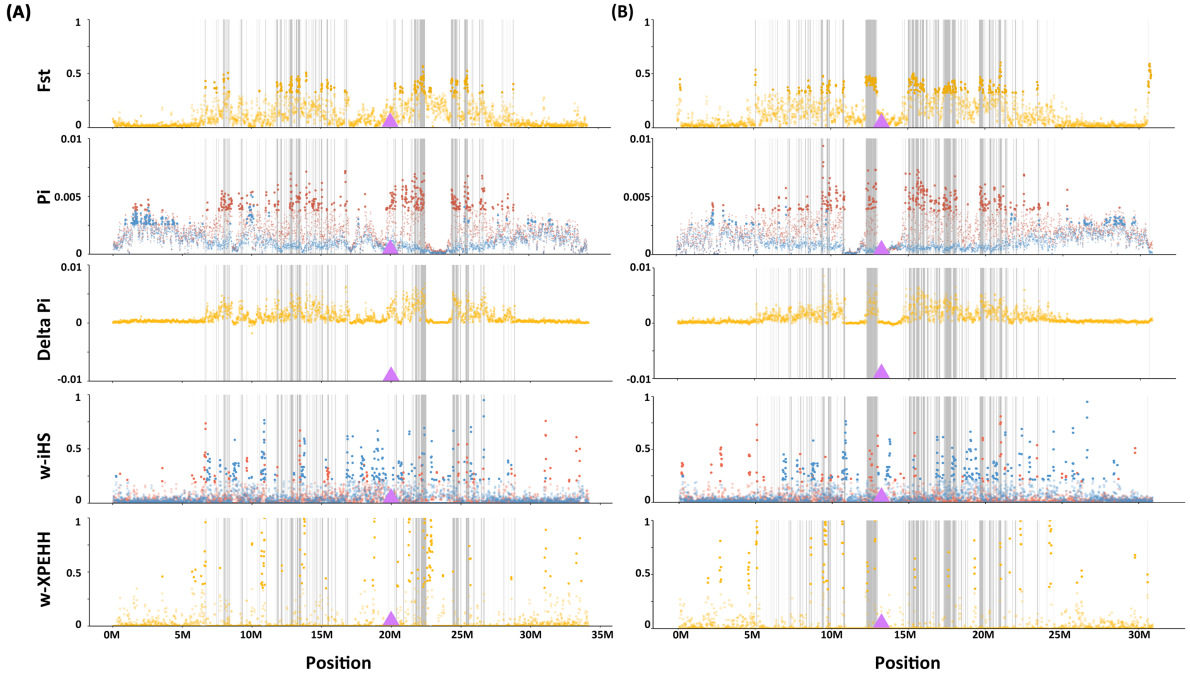
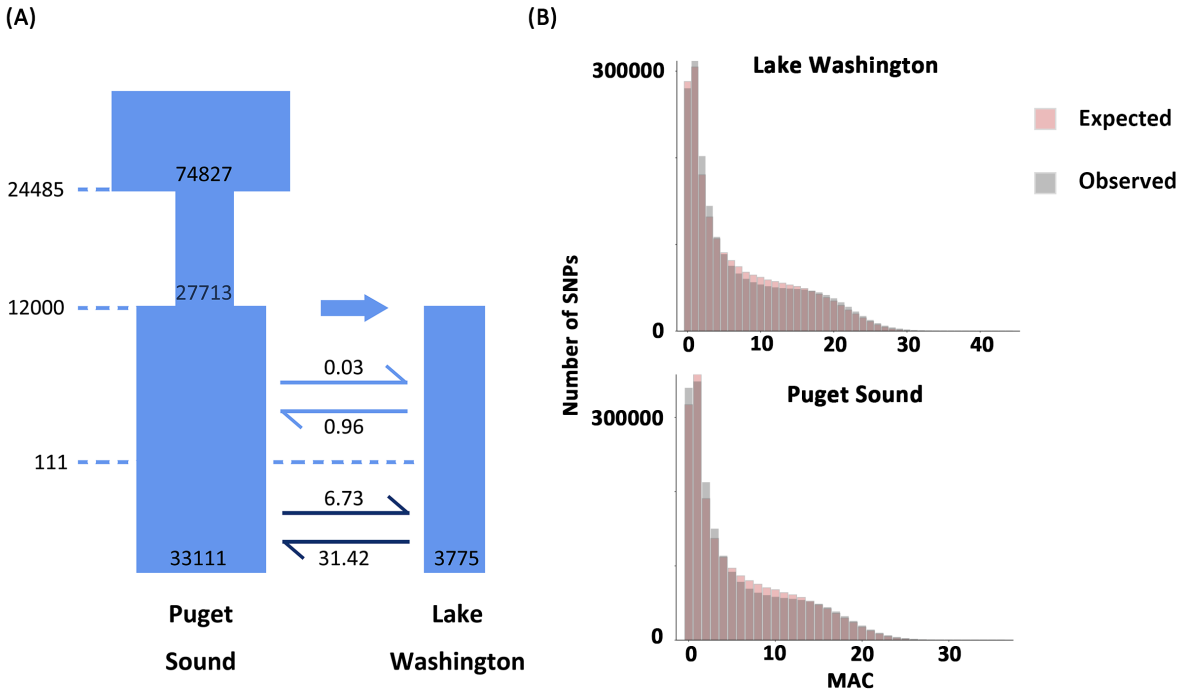


Fig. 4. Signatures of selection in the Lake Washington freshwater and Puget Sound marine populations of *G. aculeatus*. Statistics are shown here for chromosomes 4 (A) and 7 (B), with all chromosomes shown in Supplementary Fig. S5. All statistics were calculated in 20kb sliding windows with a step size 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.



971

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

976