

# Local adaptation and the evolution of genome architecture in threespine stickleback

Qiushi Li<sup>1</sup>, Dorothea Lindtke<sup>1</sup>, Carlos Rodríguez-Ramírez<sup>2</sup>, Ryo Kakioka<sup>3</sup>, Hiroshi Takahashi<sup>4</sup>,  
Atsushi Toyoda<sup>5</sup>, Jun Kitano<sup>6</sup>, Rachel L. Ehrlich<sup>7</sup>, Joshua Chang Mell<sup>7</sup>, Sam Yeaman<sup>\*1</sup>

\*corresponding author: [samuel.yeaman@ucalgary.ca](mailto:samuel.yeaman@ucalgary.ca)

1. Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Canada, T2N 1N4
2. Division of Evolutionary Ecology, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland
3. Tropical Biosphere Research Center, University of the Ryukyus, Nishihara, Nakagami-gun, Okinawa, 903-0213, Japan
4. National Fisheries University, 2-7-1 Nagata-honmachi, Shimonoseki, Yamaguchi, 759-6595, Japan
5. Comparative Genomics Laboratory, National Institute of Genetics, Mishima, Shizuoka, 411-8540, Japan
6. Ecological Genetics Laboratory, National Institute of Genetics, Mishima, Shizuoka, 411-8540, Japan
7. Department of Microbiology & Immunology, Drexel University College of Medicine, Philadelphia, USA. 19102

Keywords: genome evolution, chromosomal rearrangement, local adaptation, transposable element, gene flow

## Significance statement

The architecture of the genome can evolve through chromosomal rearrangements, duplications, and deletions, but this is thought to be a largely random process, with selection purging deleterious changes. Here, we explore whether such changes tend to evolve most rapidly in regions of the genome involved in local adaptation to freshwater vs. saltwater in the threespine stickleback. We find enrichment of several types of rearrangement in these regions, which often involve movement or duplication of genes that are differentially expressed in freshwater- vs. saltwater-adapted genotypes. As clustering of causal loci is theoretically favoured under local adaptation, clustering of these rearrangements suggests that evolution may be actively reshaping the genome to favour a higher-fitness architecture.

1 **Abstract**

2 Theory predicts that local adaptation should favour the evolution of a concentrated genetic  
3 architecture, where the alleles driving adaptive divergence are tightly clustered on chromosomes.  
4 Adaptation to marine vs. freshwater environments in threespine stickleback has resulted in an  
5 architecture that seems consistent with this prediction: divergence among populations is mainly  
6 driven by a few genomic regions harbouring multiple quantitative trait loci (QTL) for  
7 environmentally adapted traits, as well as candidate genes with well-established phenotypic  
8 effects. One theory for the evolution of these “genomic islands” is that rearrangements remodel  
9 the genome to bring causal loci into tight proximity, but this has not been studied explicitly. We  
10 tested this theory using synteny analysis to identify micro- and macro-rearrangements in the  
11 stickleback genome and assess their potential involvement in the evolution of genomic islands.  
12 To identify rearrangements, we conducted a *de novo* assembly of the closely-related tubesnout  
13 (*Aulorhynchus flavidus*) genome and compared this to the genomes of threespine stickleback and  
14 two other closely related species. We found that small rearrangements, within-chromosome  
15 duplications, and Lineage-Specific Genes (LSGs) were enriched around genomic islands, and that  
16 all three chromosomes harbouring large genomic islands have experienced macro-  
17 rearrangements. We also found that duplicates and micro-rearrangements are 9.9x and 2.9x more  
18 likely to involve genes differentially expressed between marine and freshwater genotypes. While  
19 not conclusive, these results are consistent with the explanation that strong divergent selection on  
20 candidate genes drove the recruitment of rearrangements to yield clusters of locally adaptive loci.  
21

ACCEPTED MANUSCRIPT

## 1 **Introduction**

2 Many species inhabit heterogeneous environments where spatial differences in the direction of  
3 natural selection drive adaptation to the local environment (Hedrick et al. 1976; Hereford 2009).  
4 When migration rate among populations is sufficiently high, an evolutionary tension develops  
5 with divergent selection that can profoundly affect the genetic architecture of local adaptation.  
6 Because weakly-selected alleles are susceptible to “swamping” by migration under these  
7 conditions (Haldane 1930; Lenormand 2002), there is a general advantage for alleles with larger  
8 effects and/or tightly linked clusters of alleles with smaller effects (Yeaman and Otto 2011;  
9 Yeaman and Whitlock 2011). This advantage of such “concentrated” genetic architectures is  
10 expected to favour the evolution of clustering of causal alleles (Feder et al. 2012; Via 2012)  
11 which can occur via three broad types of mechanism: 1) differential probability of establishment,  
12 persistence time, or competition favouring alleles that are more tightly linked (Yeaman and  
13 Whitlock 2011; Aeschbacher and Buerger 2014; Yeaman et al. 2016); 2) modifiers reducing the  
14 rate of recombination between existing loosely-linked alleles (*e.g.* by establishment of an  
15 inversion capturing the alleles; (Noor et al. 2001; Rieseberg 2001; Kirkpatrick and Barton 2006);  
16 3) fixation of a chromosomal rearrangement moving a causal locus into close proximity with  
17 other causal loci (Yeaman 2013; Guerrero and Kirkpatrick 2014).

18 While evidence in some species seems consistent with the concentrated architectures  
19 hypothesis, much remains unclear about which mechanisms drive their evolution. Empirical work  
20 has revealed a wide range of patterns in the genomic landscape of differentiation underlying local  
21 adaptation, with some studies finding large clusters of loci that are highly differentiated between  
22 populations (“genomic islands”), but others finding little evidence for such patterns (Nosil et al.  
23 2009; Cruickshank and Hahn 2014; Yeaman 2021). Unfortunately, when genomic islands are  
24 found it is typically unclear which loci within them are selected vs. neutral, so it is difficult to  
25 infer if this is evidence for clustering of causal loci. Furthermore, in many cases genomic islands  
26 could also be explained as artefacts arising from linkage and background or positive selection  
27 (Noor and Bennett 2009; Cruickshank and Hahn 2014; Booker et al. 2021). If genomic islands do  
28 in fact represent concentrated architectures, it is particularly interesting to know whether  
29 rearrangements contributed to their evolution, because this constitutes a durable change in the  
30 architecture of the genome. The other mechanisms of architecture evolution (1 & 2) depend on  
31 the segregation of alleles or inversions, which could be lost following an extreme population  
32 bottleneck.

33 Clear evidence of clustering has been found for the genes involved in secondary  
34 metabolic pathways in many plants (Nützmann and Osbourn 2014; Slot and Gluck-Thaler 2019),  
35 but it is unclear whether such clustering has evolved to reduce recombination or for some other

1 more proximate benefit, such as coordination of gene expression or translation. Some fascinating  
2 examples of “supergene” architectures with tightly clustered alleles have been found in species  
3 experiencing local adaptation or negative frequency dependent selection within populations  
4 (Schwander et al. 2014; Thompson and Jiggins 2014; Charlesworth 2016), such as in the social  
5 chromosomes in ants (Wang et al. 2013; Purcell et al. 2014), wing-color pattern in *Heliconius*  
6 butterflies (Joron et al. 2011), floral architecture in petunia (Hermann et al. 2013), and coloration  
7 in stick insects (Villoutreix et al. 2020). However, in most cases it is unclear whether such  
8 supergenes evolved through allelic replacement (mechanism 1, above) or rearrangement of  
9 underlying loci (mechanism 3; (Charlesworth and Charlesworth 1975)), and in many cases these  
10 supergenes are also associated with inversions.

11 Here, we approach this question from the other direction, beginning with regions of the  
12 genome known to be involved in local adaptation, and asking whether such regions have  
13 experienced more rapid evolution in genome organization and architecture. While most  
14 rearrangements likely evolve under the balance between mutation, drift, and purifying selection,  
15 an increased occurrence in the genomic regions involved in local adaptation would be unlikely to  
16 occur under this null model. By contrast, if local adaptation has favoured the fixation of  
17 rearrangements to create clusters of causal loci with increased linkage (Yeaman 2013), we would  
18 expect to see an enrichment of such events in genomic regions driving local adaptation. We also  
19 study changes in macro-scale chromosomal architecture, as fusions can bring together larger  
20 regions of the genome harbouring multiple genomic islands, due to a similar advantage for local  
21 adaptation (Guerrero and Kirkpatrick 2014).

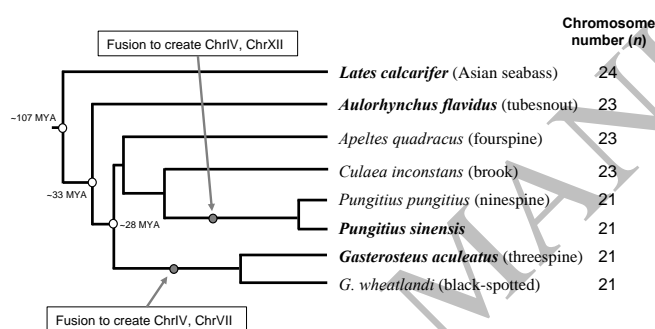
22 We explore this question by studying the evolution of genome architecture in threespine  
23 stickleback (*Gasterosteus aculeatus*), a model species for the study of ecological adaptation.  
24 Extensive study has revealed regions of the genome that are disproportionately involved in local  
25 adaptation for freshwater vs. saltwater environments, harbouring large numbers of linked QTL for  
26 a range of ecologically important traits (Miller et al. 2014; Peichel and Marques 2017; Erickson et  
27 al. 2018) that tend to co-occur with genomic islands of differentiation (Hohenlohe et al. 2010;  
28 Jones et al. 2012; Samuk et al. 2017; Kingman et al. 2021). Importantly, threespine stickleback  
29 has been undergoing repeated bouts of local adaptation to freshwater through many cycles of  
30 extirpation and recolonization over millions of years (Bell and Foster 1994; Schluter and Conte  
31 2009; Nelson and Cresko 2018), while outgroup species such as tubenout and seabass are  
32 obligately marine. While it is unclear exactly when this lineage began colonizing freshwater,  
33 other species in the stickleback clade also inhabit both marine and brackish or freshwater  
34 environments (Kawahara et al. 2009, indicating that this is an old adaptive strategy. Given that  
35 theory shows that genome evolution in response to these evolutionary pressures is likely to be

1 slow (Yeaman 2013), it is important to test this theory in a clade that has experienced a prolonged  
2 evolutionary history of inhabiting a strongly heterogenous selection environment, making the  
3 threespine stickleback a strong candidate.

4 Previous studies have revealed changes in karyotype within the stickleback clade (Ross et  
5 al. 2009; Urton et al. 2011; Rastas et al. 2016; Varadharajan et al. 2019), indicating some kinds of  
6 macro-rearrangements. Despite being less closely related (Figure 1), the *Gasterosteus* and  
7 *Pungitius* sticklebacks are more similar in their karyotype than the other close relatives of  
8 *Pungitius* (fourspine and brook stickleback). Both *Gasterosteus* and *Pungitius* have  $n = 21$   
9 chromosomes (compared to  $n = 23$  in fourspine and brook) and have syntenic arrangements for  
10 ChrIV, which is homologous to two smaller chromosomes in fourspine stickleback (Ross et al.  
11 2009; Urton et al. 2011; Rastas et al. 2016; Varadharajan et al. 2019). *Gasterosteus* and *Pungitius*  
12 differ in ChrVII: in *Gasterosteus* it is homologous to two smaller chromosomes in fourspine  
13 stickleback, but in *Pungitius* one of these smaller chromosomes has fused with the chromosome  
14 ancestral to ChrXII in threespine (Urton et al. 2011; Rastas et al. 2016; Varadharajan et al. 2019).  
15 Thus, it is unclear how karyotype has evolved within this group of species and which architecture  
16 more closely resembles the ancestral form, although it is evident that at least two of the three  
17 chromosomes most commonly involved in local adaptation have experienced some large-scale  
18 rearrangements.

19 To study the interplay between genome evolution and local adaptation in threespine  
20 stickleback, we reconstruct the history of macro- and micro-rearrangements by comparing the  
21 genomic position of orthologs among closely related species. As this requires comparison with an  
22 outgroup species, we construct the first chromosome-scale *de novo* genome assembly of  
23 tubesnout (*Aulorhynchus flavidus*), a closely-related and obligately marine outgroup of the  
24 stickleback clade, and compare this to the recently-published assembly of another stickleback  
25 (*Pungitius sinensis*; (Yamasaki et al. 2020)). We use Asian seabass (*Lates calcarifer*; (Vij et al.  
26 2016)) and additional outgroup species to improve orthology reconstruction and identify whether  
27 putative rearrangements happened in the tubesnout or stickleback lineage. For small-scale  
28 changes in genome architecture, we characterize three types of events, which we collectively  
29 refer to as Micro Genome Evolution Events (MGEEs): within-chromosome gene duplications,  
30 inter-chromosomal rearrangement of one or more adjacent genes, and Lineage-Specific Genes  
31 (LSGs) suggestive of *de novo* gene birth. We then test whether these MGEEs tend to be enriched  
32 within and around genomic islands for marine vs. freshwater divergence identified by Kingman et  
33 al. (2021). While *de novo* gene birth is not a rearrangement, if it results in a novel adaptive  
34 function and occurs in a beneficial linkage relationship to other locally adapted loci in a genomic  
35 island, this would favour recruitment of LSGs within genomic islands above the background rate.

1 Our analysis of the distribution of both macro and micro-rearrangements placed *a priori* focus on  
 2 chromosomes IV, VII, and XXI, as numerous lines of evidence from QTL studies and genome  
 3 scans show they tend to be over-represented in their contributions to marine vs. freshwater local  
 4 adaptation (Hohenlohe et al. 2010; Jones et al. 2012; Miller et al. 2014; Peichel and Marques  
 5 2017; Erickson et al. 2018; Samuk et al. 2017), and harbour a number of candidate genes  
 6 identified by fine-scale mapping, including *Eda* (Colosimo et al. 2005), *Msx2a* (Howes et al.  
 7 2017), *Wnt7b* (Jones et al. 2012), *Pitx1* (Shapiro et al. 2004), *Tfap2a* (Erickson et al. 2018), and  
 8 *Bmp6* (Cleves et al. 2014; see Table S1 and Supplementary materials for further details about  
 9 methods development).  
 10  
 11



12  
 13 Figure 1. Phylogeny of the stickleback and closely related species. Chromosome numbers are  
 14 derived from (Urton et al. 2011; Vij et al. 2016), and the current study. Divergence times were  
 15 estimated by taking the median across a number of studies using Timetree (Kumar et al. 2017),  
 16 which places the *Gasterosteus* and *Pungitius-Apeltes* split at ~27.8 MYA (confidence interval, CI  
 17 = 19.0 – 32.1 MYA), the stickleback and tubesnout split at ~33 MYA (CI = 25 – 43 MYA), and  
 18 the split with Asian seabass at ~107 MYA (CI = 94 – 115 MYA). Redrawn based on Kawahara et  
 19 al. (2009); a more recent phylogeny based on genome-wide data also groups both *A. quadracus*  
 20 and *C. inconstans* with the *Pungitius* clade with 100% bootstrap support (Figure 2A in Guo et al.  
 21 2019); branch lengths are not drawn to scale.

## 22 23 Results

### 24 *First draft de novo assembly of the tubesnout genome*

25 The estimated genome size of the male tubesnout used in this study was 468.9 Mb based on the  
 26 16 k-mer frequency counting result using Illumina sequence (Figure S1). The kmer-individual  
 27 heterozygous ratio is about 0.032, indicating the high heterozygosity of the sample, and 12.18%  
 28 of the genome was categorized as the repetitive content. We used Pacbio (RS II) long reads (50.3

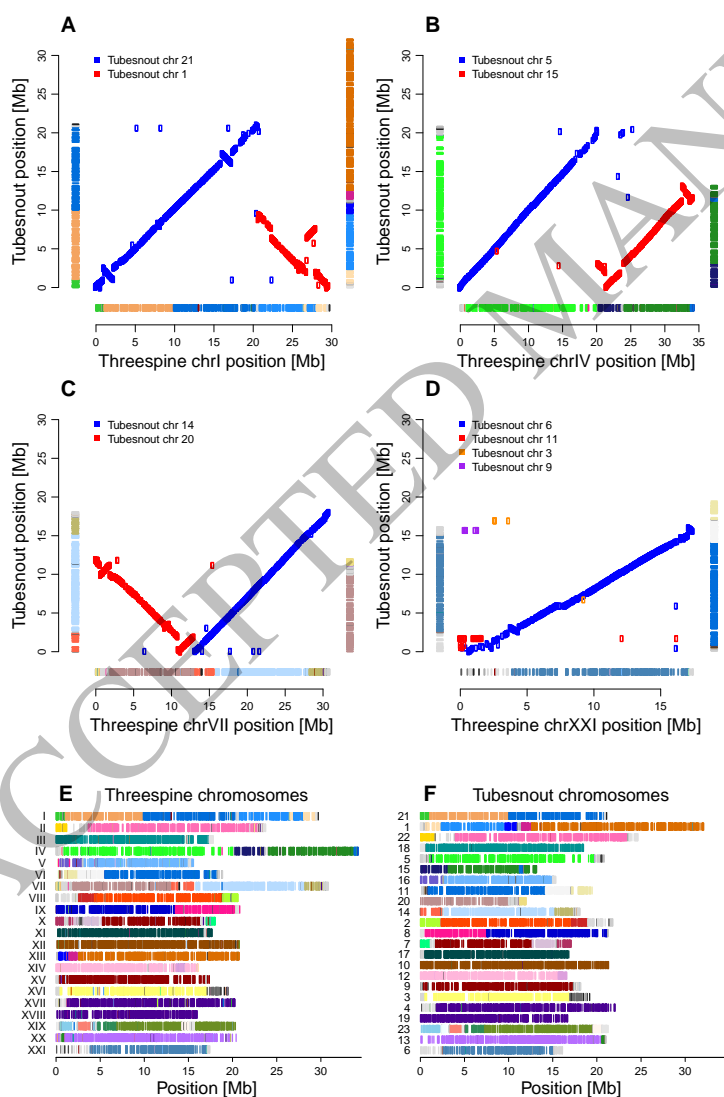
1 Gb, > 100x coverage) generated from a 20kb insert-size SMRTbell library for the contig-level  
2 assembly to ensure accuracy. We obtained 1,118 phased haplotigs with a total length of 488.5 Mb  
3 and N50 length of 2.2 Mb, which was subsequently polished with 226x Illumina short reads and  
4 used as the input for Hi-C scaffolding. Mis-joins and duplicates of the haplotigs were solved  
5 based on the chromatin conformation information captured from same individual. Finally, contigs  
6 totalling 445.6 Mb, accounting for 97.1% of the total 458.8 Mb assembled genome sequences,  
7 were clustered into 23 chromosome-scale scaffolds in the Hi-C scaffolding step (Figure S2).  
8 BUSCO assessment with the Actinopterygii database composed of 4584 BUSCOs revealed a  
9 completeness summary of complete orthologs: 94.3% (single-copy: 92.1%, duplicated: 2.2%),  
10 fragmented orthologs: 2.6%, missing orthologs: 3.1%. The 23 chromosome scaffolds show great  
11 consistency in contiguity when compared to the 10 longest reads (>100kb) assembled following  
12 the orthogonal linked-reads strategy (Figure S3). All these data demonstrate a high-quality  
13 genome assembly of the tubesnout.

#### 14 *Macro-rearrangements in stickleback*

15 Both methods (1 & 2; see Methods) for identifying rearrangements revealed that both  
16 chromosomes IV and VII had undergone fusions somewhere on the threespine stickleback  
17 lineage, as the homologous regions in both seabass and tubesnout are present as two separate  
18 chromosomes in each case (Figure 1, 2, S4-S6). Consistent with the prediction of these macro-  
19 rearrangements being driven by an advantage for local adaptation, the fusions creating both  
20 chromosomes IV and VII involved regions of the genome harbouring genomic islands strongly  
21 implicated in local adaptation in threespine stickleback (Figure 3). These regions would have  
22 been on separate chromosomes and therefore freely recombining prior to the ancestral fusion.  
23 These results show that fusions of the same two parent chromosomes independently created  
24 ChrIV in both *Pungitius* and *Gasterosteus*, which would be very unlikely to happen by chance.  
25 Previous data showed that ChrIV is syntenic but not collinear in these species (Rastas et al. 2016;  
26 Varadharajan et al. 2019) and our comparison with tubesnout and seabass shows that the non-  
27 fused architecture of this chromosome was most likely ancestral, and it is also presumably shared  
28 with fourspine stickleback, which has the same number of chromosomes as tubesnout (Figure 1;  
29 Urton et al. 2011). Another analysis of the genomes of several stickleback species, including a *de*  
30 *novo* fourspine stickleback genome assembly, has come to similar conclusions about their macro-  
31 evolutionary history (Liu et al. 2021). Also on the threespine stickleback lineage, our  
32 reconstructions show that chromosome I experienced a large translocation (involving homologs  
33 to tubesnout chromosomes 1 & 21; Figure 2) and chromosome XXI experienced a series of  
34 complex rearrangements at one end involving three tubesnout homologs (Supplementary Results;

1 Figure S6), while the other 17 chromosomes are broadly conserved in their synteny between  
 2 stickleback and tubesnout (Table S2).

3 Taken together, these results show that all three of the chromosomes most commonly  
 4 involved in local adaptation in threespine stickleback have undergone macro-scale  
 5 rearrangements in the threespine stickleback lineage, but that only one other chromosome has  
 6 done so. If the chance of each of the 21 stickleback chromosomes undergoing such macro-  
 7 rearrangement is equal, the probability that all 3 chromosomes with pronounced genomic islands  
 8 experienced rearrangement is  $p = 0.003$ , given 4 random draws from 21 without replacement.  
 9 Alternatively, if the chance of rearrangement is proportional to chromosome length in threespine  
 10 stickleback, then this probability is  $p = 0.0064$  (by 100,000 random draws).  
 11



12



1 Figure 2. Patterns of synteny between threespine stickleback and tubesnout. Panels A-D show dot  
2 plots for based on the positions of orthologs reconstructed by method 2, while colored bars on the  
3 sides A-D and in E-F indicate patterns of synteny identified by method 1, with the largest 47  
4 CARs plotted in color and the remaining 202 minor CARs plotted in grayscale. Note that  
5 stickleback ChrXIII is homologous to part of tubesnout chromosome 1.

### 6 *Characteristics of MGEEs*

7 We observed a total of 154 micro-rearrangements, but in some cases we could not conclusively  
8 determine whether they had occurred in the tubesnout or threespine stickleback lineage. If all of  
9 these occurred on the stickleback lineage, the long-term rate of occurrence of such events would  
10 be approximately 4.7/million years, given the divergence time of 33 MY (Kumar et al. 2017),  
11 although this might be overestimated by up to ~2x if some events in tubesnout were mis-  
12 attributed to stickleback. We observed a total of 288 LSGs common to both threespine  
13 stickleback and *P. sinensis* (70 of which were high confidence). Given estimated divergence  
14 times of ~27.8MYA between threespine stickleback and *P. sinensis* and 33MYA between the  
15 sticklebacks and tubesnout, this suggests a burst of LSGs in the early stages of stickleback  
16 evolution (288 over ~4.2 million years). We observed 248 duplications in stickleback not found  
17 in tubesnout, which would correspond to a rate of 7.5/million years. The size of the genes  
18 involved in micro-rearrangements (mean = 831.1 bp), LSGs (472.4 bp), and duplications (1008.3  
19 bp) tended to be significantly smaller than for genes that have not undergone such events (mean =  
20 1582.9 bp; Wilcoxon rank sum test,  $p < 10^{-15}$  in all cases). We were able to annotate 238 out of the  
21 248 duplicated genes and 128 out of 182 of the re-arranged genes. We conducted a test of GO  
22 enrichment in these genes and interestingly found a significant enrichment of genes related to  
23 olfactory receptors and the hemoglobin complex on the duplicated genes, and an enrichment of  
24 genes related to the dynein complex on the re-arranged genes (Table S4).

### 26 *Genomic distribution of MGEEs*

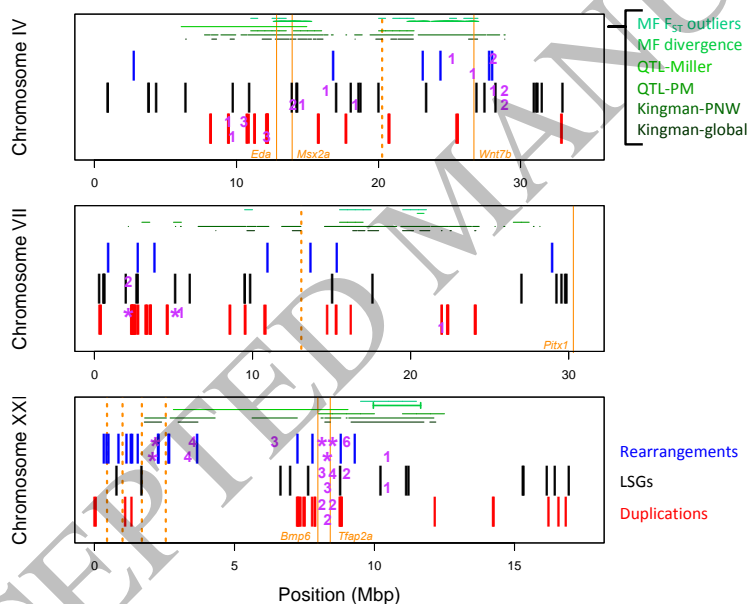
27 In order to test for enrichment of MGEEs around genomic islands, it was first necessary to  
28 characterize the broad-scale patterns in their distribution to develop null models for enrichment  
29 testing. In chromosomes without a history of macro-rearrangement, the density of rearrangements  
30 and LSGs tended to be higher towards the ends of chromosomes, whereas duplications exhibited  
31 a less consistent pattern with reduced density near the ends of chromosomes (Figure S7). We  
32 found no consistent differences in patterns of MGEE occurrence between acro-, meta-, telo-  
33 centric chromosomes (Figure S8), and these patterns did not consistently covary with gene  
34 density (Figure S7).

1 To test for enrichment around genomic islands, we compared observed counts of MGEEs  
2 to expectations under a null model based on their respective density patterns (Figure S7), applied  
3 separately on either side of the ancestral breakpoint in chromosomes that had undergone macro-  
4 rearrangement (the ‘double-adj’ model, Figure S9). We tested enrichment of each type of MGEE  
5 in windows up to 3Mbp around the genomic islands identified by Kingman et al. (2021) for both  
6 the Pacific North West (PNW) and global regions. At the whole genome scale, we found that both  
7 duplications and rearrangements were enriched ( $p < 0.05$ ) in significantly more genomic islands  
8 than expected (*i.e.*  $>> 5\%$  of cases, by a binomial test) for the PNW set, but not for the global set  
9 (Figure 4A). Similar patterns of enrichment were found for rearrangements around genomic  
10 islands in the focal chromosomes (IV, VII, and XXI), but duplications showed somewhat reduced  
11 enrichment that was non-significant, perhaps due to the lower power associated with a smaller  
12 number of genomic islands (Figure 4B). LSGs showed less consistent patterns that were only  
13 significantly enriched at a few window sizes (Figure 4). These patterns were largely robust to the  
14 null model with similar results found using the gene density, flat, and single-adj models, with the  
15 exception of a loss of significance for rearrangements in the flat model (Figure S10).

16 Examining patterns within the focal chromosomes, we found particularly strong  
17 signatures of enrichment for all three types MGEE in genomic islands near *Bmp6* and *Tfap2a* on  
18 ChrXXI (Figure 3), which are genes known to be involved with tooth gain and craniofacial  
19 architecture in stickleback (Cleves et al. 2014; Erickson et al. 2018). Micro-rearrangements were  
20 also significantly enriched in genomic islands near the complex macro-rearrangements on Chr  
21 XXI. Less pronounced signatures of enrichment were found in genomic islands on chromosomes  
22 IV and VII, with the strongest of these being for duplications in genomic islands near the *Eda* and  
23 *Msx2a* genes, which are involved with local adaptation to freshwater (Colosimo et al. 2005;  
24 Howes et al. 2017; Schluter et al. 2021). The degree of significance of these patterns of  
25 enrichment varies with the choice of null model, but the broad patterns remain significant  
26 regardless (Figures S11-13), so the choice of density model does not seem to be driving our  
27 results. Similar patterns of enrichment were also found applying this method to test enrichment  
28 around the main candidate genes on the focal chromosomes (Table S4). Examining patterns of  
29 MGEE distribution irrespective of genomic islands, there was significant enrichment of  
30 duplications on chromosomes XIX, X, and XI, of rearrangements on XXI and X, and of LSGs on  
31 XII (Figure S14). Thus, the above patterns of enrichment within genomic islands on the focal  
32 chromosomes do not arise from an overall higher rate of MGEEs on these chromosomes.

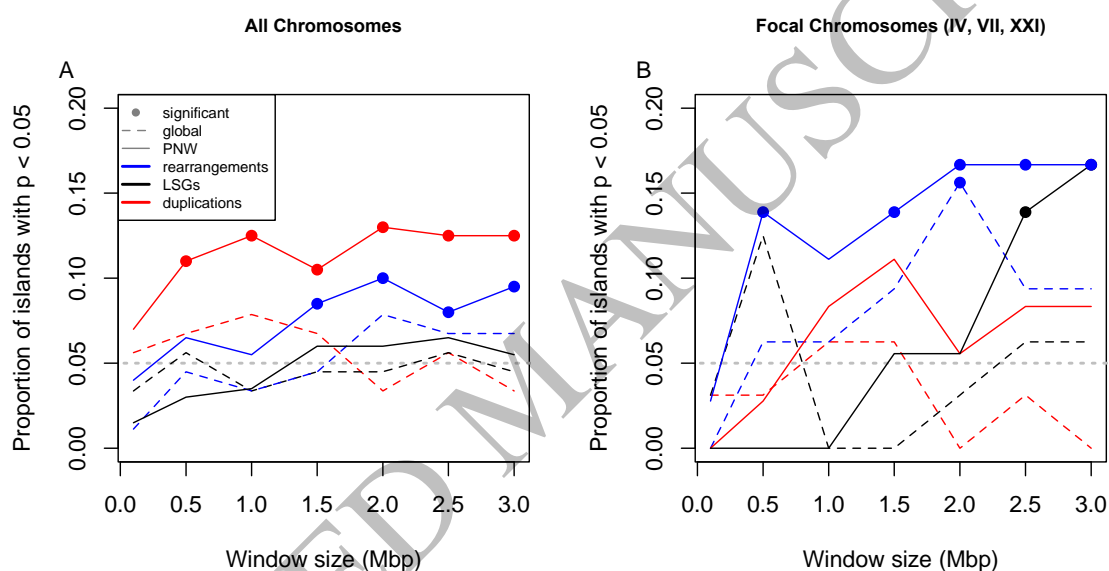
33 It is possible that the above patterns of enrichment of MGEEs around genomic islands  
34 were driven by an increased rate of occurrence near macro-rearrangement breakpoints, if genomic  
35 islands happen to also be close to these breakpoints. To test this alternative hypothesis, we

1 applied the same approach to testing enrichment of MGEEs around the ancestral chromosomal  
 2 breakpoints on each of the four chromosomes with macro-rearrangements, finding that both Chr I  
 3 & XXI showed significant increases in micro-rearrangement near their ancestral breakpoints, but  
 4 Chr IV and VII did not (Table S3). LSGs and duplications were not enriched near any ancestral  
 5 breakpoints (Table S3). Most of the genomic islands that are significantly enriched for micro-  
 6 rearrangements are not near the ancestral breakpoints on the focal chromosomes (Figure 3, with  
 7 the exception of two islands on Chr XXI), and none of the islands on Chr I contributed to the  
 8 significance of the genome-wide patterns (Figure 4). Enrichment driven by a higher rate of  
 9 MGEEs near ancestral breakpoints therefore does not seem to be a general explanation for the  
 10 patterns we found, but might explain the enrichment found in the two islands that overlap the  
 11 breakpoints on Chr XXI.  
 12



13  
 14 Figure 3. Chromosomal distribution of three types of Micro-Genomic Evolution Event (MGEE):  
 15 micro-rearrangements (blue), Lineage Specific Genes (LSGs; black), and duplications (red)  
 16 across the three focal chromosomes commonly involved in local adaptation. Shaded rectangles  
 17 indicate regions that are significantly enriched for each type of MGEE ( $p < 0.05$ ) around the  
 18 Kingman PNW (above) and global (below) sets of genomic islands. Purple numbers indicate how  
 19 many of the 7 window sizes were found to be significant ( $p < 0.05$ ), with “\*” indicating an island  
 20 that was significant following FDR correction across all islands tested ( $q < 0.05$ ) for at least one  
 21 window size. The locations of candidate genes for local adaptation are shown with solid orange  
 22 lines; orange dashed lines indicate the approximate location of breakpoints for ancestral macro-

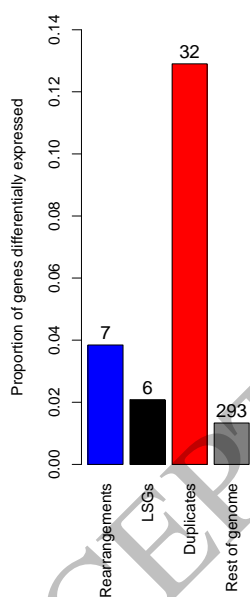
1 rearrangements. Lines along the top of each panel indicate positions where: mean  $F_{ST}$  between  
 2 marine-freshwater populations from Samuk et al. (2017) falls in the top 5% of the distribution;  
 3 extreme marine-freshwater divergence identified by Jones *et al.*, (2012; with the inversion on  
 4 ChrXXI identified as a bounded line); QTL identified by Miller et al. (2014); number of QTL  
 5 from the meta-analysis of Peichel and Marques (Peichel and Marques 2017) falls into the top 5%  
 6 of the distribution (dark green), and the PNW and global sets of genomic islands from Kingman  
 7 et al (2021).  
 8



9  
 10 Figure 4. The proportion of genomic islands in the PNW and global sets of “ecopeaks” from  
 11 Kingman et al. (2021) with significant enrichment of different types of MGEE. Enrichment  
 12 analysis was conducted using the double-adj model within windows of various sizes around each  
 13 genomic island. Results are shown for the 200 (PNW) and 89 (global) genomic islands across all  
 14 chromosomes (A) and the 36 (PNW) and 32 (global) islands within the focal chromosomes most  
 15 commonly involved in local adaptation to marine vs. freshwater (B). Significance indicated by  
 16 the filled dots occurs when the number of windows with  $p < 0.05$  exceeds the 95<sup>th</sup> percentile of a  
 17 binomial distribution, with the null expectation of 5% indicated by a horizontal dashed grey line.  
 18 *MGEEs are enriched for marine vs. freshwater differential expression*  
 19 To examine whether the Micro Genome Evolution Events (MGEEs) tend to involve genes that  
 20 are functionally important, we used a recently published dataset on differential gene expression  
 21 among freshwater and marine stickleback ecotypes raised in a common environment, assayed in  
 22 gill tissue (Verta and Jones 2019). Out of the 21,855 genes in our high confidence set that had not

1 experienced an MGEE, 293 were identified as being differentially expressed between these  
 2 ecotypes (1.3%; Figure 5). We found significantly higher rates of differential expression in genes  
 3 involved in micro-rearrangements (3.8%; binomial test  $p = 0.02$ ) and duplications (12.9%;  $p < 10^{-19}$ )  
 4 but not LSGs (2.1%;  $p = 0.26$ ). As there were very few of these genes overall, it was not  
 5 possible to test enrichment within chromosomes, however some intriguing patterns are apparent.  
 6 Of the seven differentially expressed genes on ChrIV that were involved in an MGEE, six of  
 7 them cluster within the significant regions near *Eda/Msx2a* identified in Figure 3 (3 genes  
 8 involved in groups of duplications and 3 LSGs). Similarly, all five of the MGEEs on ChrVII that  
 9 were also differentially expressed are found within the first 3.5Mbp of the chromosome, where  
 10 there is significant enrichment of duplications within a PNW genomic island. By contrast,  
 11 differential expression was not found in any of the genes involved in the MGEEs within the  
 12 genomic islands on ChrXXI near *Tfap2a* and *Bmp6*.

13



14

15

16 Figure 5. Proportion of genes involved in the three types of MGEE that are differentially  
 17 expressed among freshwater vs. saltwater threespine stickleback ecotypes. Numbers above each  
 18 bar indicate the number of genes; for duplications, all genes in group of related duplicates are  
 19 counted as a single gene.

20

21

22

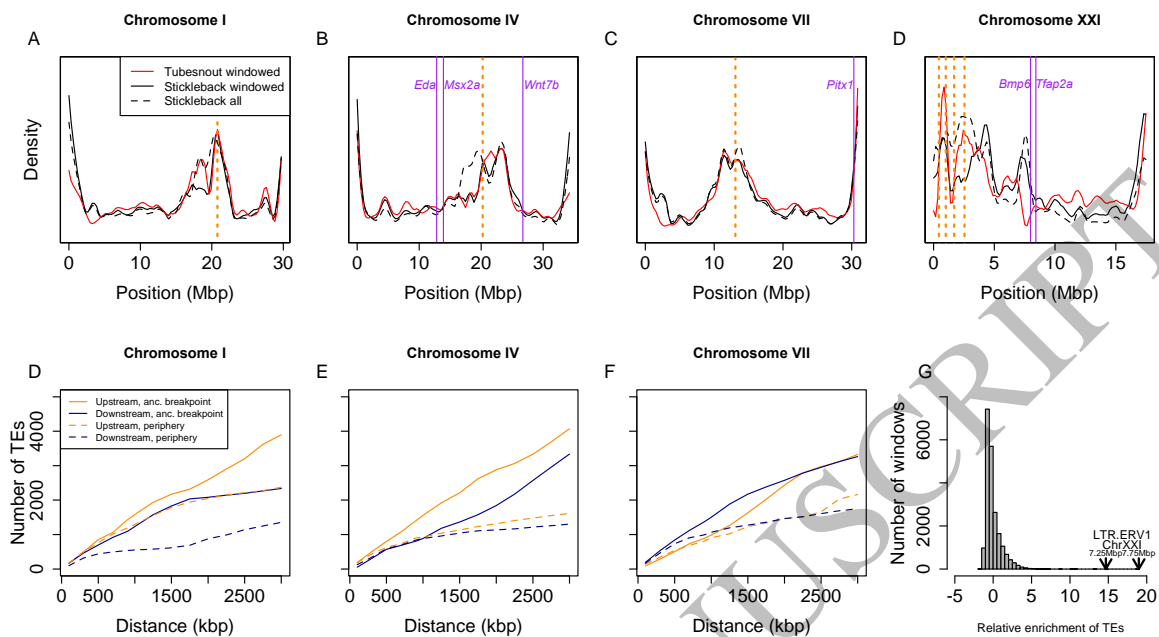
23

## 1 *Chromosomal distribution of transposable elements*

2 Transposable elements can drive genome evolution, either by directly moving genes during  
3 transposition or by facilitating rearrangements through unequal recombination (Petes and Hill  
4 1988; Mani and Chinnaiyan 2010; Lisch 2013). Across the whole genome, the numbers of micro-  
5 rearrangements and duplications correlated strongly with the density of transposable elements  
6 (TEs) when calculated on 500kbp moving windows (Kendall's  $\tau = 0.21$ ,  $p < 10^{-14}$  and  $\tau = 0.11$ ,  
7  $p < 10^{-4}$ , respectively), with patterns of elevated TE density in the peripheral chromosomal  
8 regions (Figure S7D). TEs were also significantly enriched near the ancestral breakpoints of the  
9 macro-rearrangements on chromosomes I, IV, and VII (Figure 6A-D, E-G), even after correcting  
10 for the increased density expected if the macro-rearrangements had not occurred and these had  
11 remained as peripheral regions (Table S5). By contrast, LSGs did not correlate with TE density ( $\tau$   
12  $= 0.03$ ,  $p = 0.25$ ).

13 To compare patterns of TE occurrence in threespine stickleback vs. tubesnout, we used a  
14 500kb moving-window analysis in threespine stickleback, identifying all genes within each  
15 window with syntenic and collinear mappings in tubesnout. For each of these genes, we included  
16 all TEs mapped within 50kb, and then counted all unique TEs within each 500kb window (see  
17 Methods for details). We found similar patterns of overall TE density in both species, with small  
18 localized increases or decreases in density found in many homologous chromosome regions in  
19 both species (Figure 6A-D). The most striking departure from this similarity is observed just  
20 upstream from the region of *Bmp6* and *Tfap2a* on ChrXXI that harbours significant enrichment of  
21 all three types of MGEE, where there is dramatic enrichment of LTR.ERV1 elements found only  
22 in threespine stickleback (Figure 6D). Two 500kbp windows centered around 7.25 and 7.75 Mbp  
23 on ChrXXI show the greatest enrichment observed for any common type of TE anywhere in the  
24 genome, with 162 and 208 LTR.ERV1 elements respectively, which is  $>14$  standard deviations  
25 above the mean of 4.9 per window (Figure 6H).

26



1

2 Figure 6. Distributions of Transposable Elements (TEs) in threespine stickleback and tubesnout in  
 3 the 4 chromosomes with macro-rearrangements. Panels A-D show the density of TEs from the  
 4 windowed-ortholog analysis in both species, as well as the raw density of all TEs in threespine  
 5 stickleback, which includes those that do not fall within the windows around identified orthologs.  
 6 Panels E-G show the number of TEs in threespine stickleback within a given distance upstream  
 7 and downstream from the ancestral breakpoint of the macro-rearrangement, and compare this to  
 8 the number of TEs found in the same distance from either end of the contemporary periphery of  
 9 each chromosome. Panel G shows the relative abundance of different classes of TE within 500kb  
 10 windows, with the highest enrichment shown for LTR.ERV1 elements in two windows on  
 11 ChrXXI, highlighted in the colored box, with the corresponding region highlighted in panel D.

## 12 Discussion

13 Our aim was to test whether macro- and micro-rearrangements have affected the genome  
 14 architecture of loci that contribute to local adaptation. The evidence on this question is mixed: we  
 15 found no MGEE "smoking guns" directly involving known candidate genes for local adaptation,  
 16 which were all found as single copies in syntenic locations in all species. However, we did find  
 17 many patterns that are consistent with local adaptation causing evolution in genome architecture,  
 18 with significant enrichment of micro-rearrangements and duplications around genomic islands  
 19 (Figure 4), pronounced enrichment of all three types of MGEE around genomic islands on the  
 20 focal chromosomes (Figure 3), increased involvement of differentially expressed genes in  
 21 duplications and micro-rearrangements (Figure 5), and macro-rearrangements in all 3 focal  
 22 chromosomes (Figure 2).

1 On the macro-scale, we identified a previously unknown pattern of complex  
2 rearrangements in the first 2.5Mbp on ChrXXI, which shows that all three focal chromosomes  
3 commonly involved in local adaptation have undergone either fusions or complex  
4 rearrangements, which is unlikely to have occurred at random ( $p \leq 0.003$ ). This is particularly  
5 noteworthy given that in both threespine and ninespine stickleback, chromosome IV has likely  
6 been created by fusions of the same ancestral chromosomes, which would be very unlikely to  
7 happen by chance. These patterns are consistent with another analysis of macro-rearrangements  
8 using a *de novo* assembly of the fourspine stickleback (Liu et al. 2021), and strongly suggest an  
9 adaptive mechanism driving macro-rearrangements in stickleback. Finding macro-rearrangements  
10 associated with local adaptation is consistent with population genetic predictions: if two locally  
11 adapted loci experience selection of  $s_a$  and  $s_b$  and are separated by recombination at rate  $r$ , then  
12 they will experience an advantage due to linkage whenever  $r < s_a s_b / m$ , where  $m$  is the migration  
13 rate (Yeaman et al. 2016). Given that *Eda* experiences particularly strong selection, with  
14 estimates of  $s \sim 0.5$  (Schluter et al. 2021), an advantage for linkage with *Eda* would extend to  
15 other locally-adapted alleles (with  $s \sim m$ ) across the length of ChrIV (*i.e.* at distances up to  $r =$   
16 0.5). As such, selection acting on *Eda* and any locally adapted alleles on the other pre-fusion  
17 chromosome could have yielded a benefit of linkage strong enough to drive the fixation of these  
18 ancestral fusions (as per Guerrero and Kirkpatrick 2014).

19 On the micro-scale, we found significant patterns of enrichment of rearrangements and  
20 duplications within genomic islands of local adaptation (Figure 4), with particular enrichment of  
21 all three types of MGEE near *Tfap2a* and *Bmp6* and enrichment of duplications and LSGs in  
22 genomic islands in the region of *Eda* and *Msx2a* (Figure 3), although the significance of this latter  
23 pattern was more pronounced under the “single-adj” and gene density models (Figure S11, S13).  
24 It is unlikely that such enrichment would happen under a null model of MGEEs being driven only  
25 by drift and purifying selection. Similarly, the genes involved in small rearrangements or  
26 duplications were, respectively, 2.9 and 9.9 times more likely to be differentially expressed  
27 between marine and freshwater ecotypes than non-MGEEs (Figure 5), which is very unlikely to  
28 happen if such events occur at random. This could potentially be explained if purifying selection  
29 to eliminate new MGEEs is weaker when they involve genes with evolutionarily labile expression  
30 – the observed enrichment could then be driven by a lack of MGEEs involving genes with  
31 conserved patterns of gene expression. Finally, we found that genes involved in duplications  
32 tended to be enriched for GO terms related to olfactory receptor activity and hemoglobin gas  
33 transport (Table S4), both of which may be important for local adaptation to marine vs.  
34 freshwater. Duplication of genes involved in olfaction is common among vertebrates (Niimura et  
35 al. 2014; Vandewege et al. 2016) and recent evidence has found increases in copy number of



1 certain subfamilies of olfactory receptors in freshwater fish species, relative to marine ones (Liu  
2 et al. 2021). Similarly, duplication of globin genes allows for synthesis of different hemoglobin  
3 forms (Storz 2016), and in fish the evolution of pH-specific globin isoforms is thought to help  
4 them colonize a wide variety of aquatic environments (Randall et al. 2014; Storz 2016). For  
5 example, in red drum (*Sciaenops ocellatus*), there is evidence for changes in the expression of  
6 hemoglobin isoforms during acclimation to hypoxia, supporting the idea that an increased  
7 repertoire of hemoglobin genes can help species deal with environmental challenges (Pan et al.  
8 2017). Taken together, many MGEEs have evidence consistent with a role in local adaptation to  
9 freshwater vs. marine environments, but functional characterization of the genes involved is  
10 needed to more concretely establish this.

11 The large number of putative Lineage-Specific Genes (LSGs) identified here suggests  
12 that *de novo* gene birth might be important in stickleback (particularly near *Tfap2a* and *Bmp6*),  
13 but our confidence in these results is limited by difficulties involved with correct identification of  
14 LSGs. A recent analysis showed how even under a model of uniform evolutionary rate,  
15 homology-based search approaches could fail to detect true orthologs and therefore often  
16 misidentify a shared gene as an LSG (Weisman et al. 2020). On the other hand, there are some  
17 well-substantiated examples of *de novo* gene birth (Schlötterer 2015; Van Oss and Carvunis  
18 2019) and when this happens, presumably the LSG would still share some nucleotide homology  
19 with other closely related species (in the region where the gene was “born”), but without the  
20 expression and function that are hallmarks of a “real” gene. To allow for this latter possibility, we  
21 conducted our enrichment analyses with a permissive filtering criterion to allow for partial  
22 homology. However, we caution that many of these putative LSGs require further validation by  
23 studying function and expression more deeply, and consider the “stringent” list of genes included  
24 in the archived data as the higher confidence set of putative LSGs.

25 It seems likely that transposable elements are at least partly responsible for these patterns  
26 in MGEEs, whether through promoting higher rates of rearrangement through unequal  
27 recombination (Petes and Hill 1988; Mani and Chinnaiyan 2010), or more directly through  
28 transposon-mediated movement (Lisch 2013). The region just upstream of *Tfap2a* and *Bmp6* on  
29 ChrXXI harbours the greatest enrichment of any TE anywhere in the genome, with 33-42x the  
30 average number of LTR.ERV1 elements (Figure 6H). While it is possible that these patterns are  
31 the neutral result of rearrangement rate, this would not explain why such a concentration happens  
32 to occur adjacent to these two candidate genes which also harbour significant enrichment of  
33 MGEEs in genomic islands.

34 Taken together, these results and those of another comparative study using fourspine  
35 stickleback (Liu et al. 2021) are consistent with the patterns expected if local adaptation drives

1 genome evolution, but are not conclusive, given the retrospective nature of the analysis.  
2 Functional analysis of the genes involved in MGEEs would help strengthen the evidence for local  
3 adaptation as a driving force shaping these rearrangements. Further studies on other species that  
4 experience strong and persistent divergent selection and local adaptation over millions of years  
5 would help establish whether this pattern is the result of common process or is particular to the  
6 stickleback clade. Given that rates of rearrangement are particularly high in plants (Zhao and  
7 Schranz 2019), it seems possible that local adaptation in plants would more readily result in this  
8 kind of evolution in genome architecture.

9

## 10 **Methods**

### 11 *De novo assembly of the tubesnout genome*

12 A single male tubesnout specimen supplied by Living Elements (Vancouver BC) was used for all  
13 genome sequencing and assembly-related experiments in this study. The genome assembly was  
14 performed by GCEv1.0. with 18.5 Gb (~40x) error corrected Pacbio reads. High molecular  
15 weight genomic DNA was isolated and purified with the QIAGEN Genomic-Tip from muscle  
16 tissue stored at -80 °C. One 20kb insert size SMRTbell library was constructed with the Pacbio  
17 P6 v2 binding Kit, and was sequenced on 53 SMRTcells. SMRTanalysis V4.0 was used for  
18 processing and filtering the raw reads to get reads-of-insert (ROI). The ROIs longer than 3.5kb  
19 were chosen as seed reads to generate error corrected consensus sequences with higher accuracy  
20 for genome assembly.

21 We employed the diploid aware “FALCON + FALCON-unzip” approach to assemble the  
22 phased haploid genome sequences of tubesnout (59; see Dryad archive for config files).  
23 FALCON v0.5 was first used to produce the sets of primary/associated contigs representing the  
24 divergent allelic variants. All the contigs were then conveyed to the FALCON-unzip module,  
25 during which the phased haplotigs were separated based on the information of heterozygous  
26 SNPs identified by mapping the ROIs to the FALCON primary/associated contigs.

27 The Proximo Hi-C library with the insert size of the sheared ligations of ~600bp was  
28 constructed from 95% ethanol preserved muscle tissue by Phase Genomics, and was sequenced  
29 on the Illumina Hiseq 4000 platform, generating 113,119,916 paired-end reads with 100bp read-  
30 length. The Hi-C scaffolding was performed with the 3D *de novo* assembly (3D DNA) pipeline.  
31 Firstly, the Hi-C reads were mapped to the draft-assembled contigs with Juicer to generate the Hi-  
32 C contact matrix. Then we ran the 3d-DNA analysis to create an interactive heatmap, which was  
33 manually revised for the few remaining errors like haplotigs residual and incorrect placement of  
34 the contigs. The final 23 chromosome-scale super scaffolds were exported with the run-asm-

1 pipeline-post-review.sh script. The contigs that couldn't be assigned to any super scaffolds were  
2 concatenated into chromosome UN with 500 Ns separating each contig.

### 3 *Gene identification and de novo TE annotation*

4 The threespine stickleback genome (Peichel et al. 2017) and tubesnout genome were first soft-  
5 masked for repeats using Repeatmasker with Repbase and custom libraries created by  
6 RepeatModeler (v1.0.11; <http://www.repeatmasker.org/RepeatModeler>). For gene structure  
7 annotation of the threespine stickleback genome, we followed the Braker2 pipeline (Brüna et al.  
8 2020) using online RNA-seq data from different tissues (SRR5237998, SRR5420700,  
9 SRR4116640, SRR1390640, SRR1390630, SRR5420689) and the protein sequences from the  
10 existing Ensembl annotation ([ftp://ftp.ensembl.org/pub/release-90/fasta/gasterosteus\\_aculeatus/](ftp://ftp.ensembl.org/pub/release-90/fasta/gasterosteus_aculeatus/))  
11 to train the gene prediction tools GeneMark-ET (Lomsadze et al. 2014) and AUGUSTUS (Stanke  
12 et al. 2006). For the tubesnout genome, muscle RNA-seq data and the threespine stickleback  
13 protein sequences identified from the prior re-annotation were used in Braker for this novel  
14 genome. In threespine stickleback, the 25,439 identified genes were validated by either presence  
15 in the Broad annotation, or > 0.1 TPM RNA-seq reads from tissues of brain, liver, gill, kidney,  
16 head kidney, spleen, muscle, skin, eye, heart and testis tissues, or the result of target restricted  
17 assembler, aTRAM (Allen et al. 2018) with the same RNA-seq dataset. We used an automated  
18 software package (EDTA; (Ou et al. 2019)) for *de novo* genome-scale TE annotation in  
19 threespine stickleback. We then assessed if any of the above 25,439 genes were likely mis-  
20 annotated TEs using two methods. First, we used protein BLAST+ (Camacho et al. 2009) to map  
21 gene sequences against the TE library generated from EDTA, and removed any gene with >50%  
22 of its sequence having hits to TEs with >75% nucleotide identity. Second, we assessed the  
23 overlap between TE annotations and gene annotations and removed any gene with >10% of its  
24 exon sequence overlapping with TE annotations. We used these cutoffs as the default approach to  
25 curate a final annotation, yielding 23,185 high confidence genes, which are used for all  
26 downstream analyses unless specifically noted.

### 27 *Identifying macro-rearrangements by ancestral genome reconstruction (Method 1)*

28 To reconstruct macro-rearrangements, we conducted rigorous identification of orthologs with 10  
29 fish species (including tubesnout), using OMA standalone v2.2.0 (Altenhoff et al. 2015), coupled  
30 with genome reconstruction of the threespine stickleback – tubesnout ancestor using ANGES  
31 v1.01 (Jones et al. 2012) (see Supplementary Methods), and identified 19,563 orthologs with high  
32 confidence. Genome maps for these species plus threespine stickleback and tubesnout were  
33 prepared based on gene position information extracted from the gff3 or gtf files. Gene start and  
34 end positions were calculated as the average of CDS midpoints  $\pm$  1 base pair to avoid the  
35 occurrence of overlapping gene positions, which are not supported by the genome reconstruction

1 software ANGES v1.01 (Jones et al. 2012). A few remaining overlaps between gene positions  
2 were resolved manually to obtain an unambiguous order of genes for each genome. ANGES input  
3 files were generated from these genome maps and from the best-scoring phylogenetic tree  
4 computed with a set of 2,504 common one-to-one orthologs. The genome of the *G. aculeatus* - *A.*  
5 *flavidus* ancestor was reconstructed using the ANGES master pipeline (anges\_CAR.py) and  
6 options markers\_doubled 1 (infer ancestral marker orientation), markers\_unique 2 (no duplicated  
7 markers), markers\_universal 1 (no missing markers in ingroup), c1p\_telomeres 0 (no telomeres),  
8 and c1p\_heuristic 1 (using a greedy heuristic), including as outgroup all nine additional species.  
9 A total of 46,363 *Ancestral Contiguous Sets* (ACS; Jones et al., 2012) were identified by  
10 ANGES, of which 42,993 ACS were organized into 249 CARs (3,370, or 7.3%, of ACS were  
11 discarded by the program). These CARs comprised a total of 14,461 ancestral markers, and  
12 12,474 of them (86.3%) were grouped into the 23 largest CARs (i.e., major ancestral  
13 chromosomes). Putative fusions between ancestral chromosomes were identified by visually  
14 inspecting assignments of CARs to threespine stickleback and tubesnout chromosomes (Figure 2)  
15 using the R package rearrvisr (Lindtke and Yeaman 2020).

#### 16 *Identifying macro-rearrangements and MGEEs by homolog mapping (Method 2)*

17 To reconstruct the history of macro-rearrangements and MGEEs in the threespine stickleback  
18 lineage, we first used gmap (Wu and Watanabe 2005) to map all 23,185 putative genes from  
19 threespine stickleback to identify their closest homologs in *P. sinensis* (21,885 mappings),  
20 tubesnout (20,995 mappings), and seabass (18,989 mappings) genomes (with >100bp of sequence  
21 matching at >75% ID). Because the tubesnout and seabass genomes are assembled to near  
22 chromosome scale, we used these for our main analysis, and used the *P. sinensis* genome, which  
23 is somewhat more fragmented, to aid in resolving uncertain synteny relationships. Macro-  
24 rearrangements were identified by visual inspection of chromosomal synteny plots (Figures S4-  
25 6). For micro-rearrangements, three types of non-correspondence in the spatial organization of  
26 these homologs were identified at the gene-level: (A) Lineage-specific genes (LSGs) unique to  
27 sticklebacks (i.e. present in *P. sinensis* and threespine stickleback but absent from tubesnout and  
28 seabass); (B) genes where the homolog is present on a non-syntenic chromosome in at least one  
29 species, which we call putative rearrangements; (C) cases where multiple genes on a single  
30 chromosome in threespine stickleback map to a single homolog in tubesnout, which we refer to as  
31 duplications.

32 For the Lineage Specific Genes (case A), we failed to identify any match for 1340 of the  
33 23,185 high confidence stickleback genes in either tubesnout or seabass, 608 of which could also  
34 be successfully mapped to the *P. sinensis* genome. There are four plausible explanations for their  
35 occurrence: (i) they are bioinformatic errors and not real genes, (ii) they have a true homolog in

1 another species but evolved rapidly in their sequence, thereby obscuring orthology relationships,  
2 (iii) they are genes that evolved through *de novo* “gene birth” in stickleback, or (iv) homologous  
3 genes were lost independently in both the tubesnout and seabass lineages. We assume that (iv) is  
4 unlikely to have occurred commonly, so we do not further consider it here. To attempt to identify  
5 stickleback LSGs and rule out the first two explanations, we conducted additional filtering,  
6 removing any genes with a successful BLAST+ (Camacho et al. 2009) hit to the NR boneyfish  
7 database using a permissive threshold ( $e < 0.001$ ), leaving 299 genes that appear to be unique to  
8 clade including *Gasterosteus* and *Pungitius* stickleback. To further check whether the above steps  
9 failed detect a true homolog in tubesnout, we identified the homologous genomic region between  
10 tubesnout orthologs of the genes flanking the putative LSG in threespine stickleback. We then  
11 conducted a BLAST+ search of the putative LSG against this restricted region ( $e < 0.001$ ) and  
12 excluded any cases with >90% coverage (“permissive” filter; keeping 288 genes) or any cases  
13 with a hit at any level of coverage (“stringent” filter; keeping 70 genes). For all enrichment  
14 testing below, we report results based on the permissively filtered set of 288 genes.

15 For among-chromosome mismatches (case B), these could arise from true  
16 rearrangements, genome mis-assembly in one or more species, or as bioinformatic errors in  
17 ortholog identification, but we refer to them as rearrangements for simplicity. For each threespine  
18 stickleback chromosome, the homologous chromosome(s) were identified in tubesnout and  
19 individual genes in threespine stickleback were considered as putative micro-rearrangements if  
20 none of the best-hit mappings from gmap were in a syntenic location (we considered up to five  
21 mappings for each gene). This included both genes that had a one-to-one mapping relationship  
22 and genes where multiple stickleback genes on different chromosomes mapped to the same  
23 location in tubesnout (many-to-one). Several steps were then taken to exclude cases that more  
24 likely arose from bioinformatic errors and to attempt to infer where in the phylogeny the  
25 rearrangement may have occurred, excluding cases that could confidently be ascribed to have  
26 occurred in the tubesnout lineage (See Supplementary Methods).

27 For both LSGs (A) and putative rearrangements (B), we conducted a follow-up filter  
28 using BLAST+ (tblastx) to attempt to map each putative LSG or rearranged gene to the  
29 homologous area of the tubesnout genome, spanning the region between the closest syntenic  
30 neighbouring orthologs upstream and downstream of the focal gene (as per Weisman et al.  
31 (2020)). For any case with a BLAST+ hit of  $e < 0.001$  within this restricted region, the putative  
32 LSG/rearranged gene was excluded from further analysis. When flanking orthologs could not be  
33 identified readily, as would occur in areas with complex macro-rearrangements (*i.e.* ChrXXI),  
34 this final test was not conducted.

1 For case (C) we identified duplications as cases where at least two genes on the same  
2 chromosome in stickleback have their highest mappings to a single-copy gene in tubesnout and  
3 are also single-copy in seabass. Putative duplications were removed if they did not also present as  
4 a duplication when the same analysis was repeated comparing stickleback to seabass, as this  
5 would be more parsimoniously explained by a deletion in tubesnout.

### 6 7 *Testing enrichment of MGEEs and TEs*

8 To assess whether MGEEs or TEs were enriched near particular regions of the genome, for each  
9 type of event, we counted all occurrences within  $< x$  bp upstream and downstream of the region  
10 of interest, and allowed  $x$  to vary from 100kb to 3Mbp (7 increments), in order to examine  
11 clustering at different scales. For each  $x$ , we constructed a null distribution by randomly re-  
12 drawing the chromosome based on the number of genes, randomly re-drawing the start positions  
13 of all events of the same type (according to one of four density distributions) and recording how  
14 many events fall within the same increment, using 10,000 replicates. The empirical  $p$ -value was  
15 calculated as the proportion of null distribution replicates that equaled or exceeded the  
16 observation. Where a rearrangement event included more than one gene, this was counted as a  
17 single event; for a duplication event, any adjacent copies separated by  $< 1$ Mbp were counted as a  
18 single event to discount a signal of clustering caused by multiple tandem duplicates of the same  
19 gene.

20 As we observed that the distribution of both MGEEs and TEs tended to be nonuniform  
21 across the chromosome (Figure S7), we repeated the above approach under four models  
22 specifying the probability of event occurrence based on the relative position along the  
23 chromosome. First, we visually assessed whether there were differences between the spatial  
24 distribution of MGEEs among chromosome morphologies (i.e. acrocentric, metacentric,  
25 telocentric, as per Urton et al. (2011)). As there were no striking differences between these types  
26 (Figure S8), we opted to treat all types of chromosomes equally, given uncertainties in  
27 centromere position and how to rescale the relative position in such cases. We constructed  
28 probability density models for each type of event based on observations from all chromosomes  
29 that had not undergone macro-rearrangements and folded each chromosome in half, such that the  
30 relative positions scale from 0 to 0.5 (“reflected rescaling”). For each type of event we fit a  
31 bounded density model to the reflected rescaled data using the bde library in R (v1.01, with  
32 “boundarykernel” and  $b = 0.15$ ), which we termed the “single-adj” density model. Given that  
33 chromosomes I, IV, and VII experienced simple fusions or translocations (rather than the  
34 complex rearrangements found in ChrXXI), for these chromosomes we also fit the above bde  
35 model to each side of the breakpoint of the macro-rearrangement individually, scaled to the

1 length of each segment (which we term the “double-adj” model; see Figure S9). The parameters  
2 for density models were determined based on subjective visual assessment of the goodness of fit,  
3 prior to running the enrichment tests and no further alteration of these parameters was made, to  
4 avoid *p*-hacking. In the main body of the manuscript we use the “double-adj” model for  
5 chromosomes I, IV, and VII (as these chromosomes are fusion products) and the “single-adj”  
6 model for chromosome XXI (as it has only a small region of complex rearrangement at one end).  
7 We report results for the single-adj model, a uniform distribution (“flat” model), and a model  
8 scaling MGEE occurrence by gene density in the supplementary materials.

### 9 *Differential expression*

10 We were interested in assessing overlap between our rearrangements, LSGs, and duplications  
11 with the genes identified as “parallel diverged” in their expression between marine vs. freshwater  
12 ecotypes by Verta and Jones (2019). As their analysis used the BROAD S1 genome, it was  
13 necessary to map the nucleotide sequences for these genes to the Peichel et al. (2017) annotations  
14 used here, which was done using BLAST against the cDNA. These mappings were sorted by z-  
15 score and e-value and the best match was determined based on highest sequence overlap.

16 Analysis of enrichment for duplicated genes treated all copies of a duplicate as a single gene,  
17 which was counted as differentially expressed if at least one of the duplicates had a best-hit  
18 mapping from the Verta and Jones candidates.

### 19 *Transposable Element density*

20 To compare the chromosomal landscapes of TE density between threespine stickleback and  
21 tubesnout, we used the simple gene mappings from method 2 that were not involved in any  
22 MGEE and were thus collinear and syntenic between the two species. We conducted our analysis  
23 in 500kb windows; within each window we identified all TEs that fell within 50kb upstream or  
24 downstream of each collinear and syntenic gene in each species, and then counted the number of  
25 unique TEs within each 500kb window. To study the chromosomal distribution of each type of  
26 TE in threespine stickleback, we excluded any TEs with a mean density of <1 copy per 500kb  
27 window, and then converted their relative density within each window to a Z-score based on the  
28 mean and standard deviation of occurrences per window across the whole genome.

### 29 **Data availability**

30 The genomic resources, data, and scripts needed to conduct the main analyses in this paper are  
31 included in the Dryad repository (doi: <https://doi.org/10.5061/dryad.1c59zw3w3>).

### 32 **Acknowledgements**

33 We would like to thank Mackenzie Urquhart-Cronish, Philip Bruecker (Living Elements), and the  
34 Vancouver Aquarium for help obtaining tubesnout samples, JP Verta for sharing data on gene  
35 expression, and Kieran Samuk for sharing  $F_{ST}$  data. Thanks also to Katie Peichel, Dolph Schluter,  
36 and many other researchers in the stickleback community for providing advice during the design  
37 and execution of the analysis, and to Jason Holliday, Kay Hodgins, and anonymous reviewers for  
38 helpful comments. Sequencing was conducted at the Drexel University College of Medicine  
39 Genomics Core Facility and the Genome Quebec sequencing center, and Compute Canada  
40 provided computational support. This work was funded by an Alberta Innovates grant and  
41 NSERC Discovery grant to SY, and JSPS grants (19H01003 to JK and 16H06279 to AT).

## 1 **References**

- 2 Aeschbacher S, Buerger R. 2014. The effect of linkage on establishment and survival of locally  
3 beneficial mutations. *Genetics* 197:317–336.
- 4 Allen JM, LaFrance R, Folk RA, Johnson KP, Guralnick RP. 2018. aTRAM 2.0: An Improved,  
5 Flexible Locus Assembler for NGS Data. *Evol Bioinform Online* 14:117693431877454.
- 6 Altenhoff AM, Škunca N, Glover N, Train C-M, Sueki A, Piližota I, Gori K, Tomiczek B, Müller  
7 S, Redestig H, et al. 2015. The OMA orthology database in 2015: function predictions,  
8 better plant support, synteny view and other improvements. *Nucleic Acids Research*  
9 43:D240–D249.
- 10 Bell MA, Foster SA. 1994. The evolutionary biology of the Threespine Stickleback. Oxford, UK:  
11 Oxford University Press
- 12 Booker TR, Yeaman S, Whitlock MC. 2021. Global adaptation complicates the interpretation of  
13 genome scans for local adaptation. *Evolution Letters* 5:4–15.
- 14 Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2020. BRAKER2: Automatic  
15 Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a  
16 Protein Database. Bioinformatics Available from:  
17 <http://biorxiv.org/lookup/doi/10.1101/2020.08.10.245134>
- 18 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.  
19 BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- 20 Charlesworth D. 2016. The status of supergenes in the 21st century: recombination suppression in  
21 Batesian mimicry and sex chromosomes and other complex adaptations. *Evol Appl*  
22 9:74–90.
- 23 Charlesworth D, Charlesworth B. 1975. Theoretical genetics of batesian mimicry II. Evolution of  
24 supergenes. *Journal of Theoretical Biology* 55:305–324.
- 25 Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R,  
26 Figueroa-Balderas R, Morales-Cruz A, et al. 2016. Phased diploid genome assembly  
27 with single-molecule real-time sequencing. *Nat Methods* 13:1050–1054.



- 1 Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, Miller CT. 2014.  
2 Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6.  
3 *Proceedings of the National Academy of Sciences* 111:13912–13917.
- 4 Colosimo PF. 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of  
5 Ectodysplasin Alleles. *Science* 307:1928–1933.
- 6 Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due  
7 to reduced diversity, not reduced gene flow. *Mol Ecol* 23:3133–3157.
- 8 Erickson PA, Baek J, Hart JC, Cleves PA, Miller CT. 2018. Genetic Dissection of a Supergene  
9 Implicates *Tfap2a* in Craniofacial Evolution of Threespine Sticklebacks. *Genetics*  
10 209:591–605.
- 11 Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends in*  
12 *Genetics* 28:342–350.
- 13 Guerrero RF, Kirkpatrick M. 2014. Local Adaptation and the Evolution of Chromosome Fusions.  
14 *Evolution* 68:2747–2756.
- 15 Guo B, Fang B, Shikano T, Momigliano P, Wang C, Kravchenko A, Merilä. 2019. A  
16 phylogenomic perspective on diversity, hybridization and evolutionary affinities  
17 in the stickleback genus *Pungitius*. 28:4046-4064.
- 18 Haldane JBS. 1930. A mathematical theory of natural and artificial selection. (Part VI, Isolation.).  
19 *Math. Proc. Camb. Phil. Soc.* 26:220–230.
- 20 Han MV, Zmasek CM. 2009. phyloXML: XML for evolutionary biology and comparative  
21 genomics. *BMC Bioinformatics* 10:356.
- 22 Hedrick PW, Ginevan ME, Ewing EP. 1976. Genetic Polymorphism in Heterogeneous  
23 Environments. *Annu. Rev. Ecol. Syst.* 7:1–32.
- 24 Hereford J. 2009. A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. *The*  
25 *American Naturalist* 173:579–588.
- 26 Hermann K, Klahre U, Moser M, Sheehan H, Mandel T, Kuhlemeier C. 2013. Tight Genetic  
27 Linkage of Prezygotic Barrier Loci Creates a Multifunctional Speciation Island in  
28 *Petunia*. *Current Biology* 23:873–877.

- 1 Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA. 2010. Population Genomics of  
2 Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS*  
3 *Genetics* 6:23.
- 4 Howes TR, Summers BR, Kingsley DM. 2017. Dorsal spine evolution in threespine sticklebacks  
5 via a splicing change in *MSX2A*. *BMC Biol* 15:115.
- 6 Jones BR, Rajaraman A, Tannier E, Chauve C. 2012. ANGES: reconstructing ANcestral  
7 GENomeS maps. *Bioinformatics* 28:2388–2390.
- 8 Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody  
9 MC, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks.  
10 *Nature* 484:55–61.
- 11 Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter  
12 SW, Ferguson L, et al. 2011. Chromosomal rearrangements maintain a polymorphic  
13 supergene controlling butterfly mimicry. *Nature* 477:203–206.
- 14 Kingman, GAR, Vyas DN, Jones FC, Brady SD, Chen HI, et al. 2021. Predicting future from  
15 past: The genomic basis of recurrent and rapid stickleback evolution. *Science*  
16 *Advances*. 7:eabg5285.
- 17 Kirkpatrick M, Barton N. 2006. Chromosome Inversions, Local Adaptation and Speciation.  
18 *Genetics* 173:419–434.
- 19 Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines,  
20 Timetrees, and Divergence Times. *Molecular Biology and Evolution* 34:1812–1819.
- 21 Lenormand T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*  
22 17:183–189.
- 23 Lindtke D, Yeaman S. 2020. *rearrvisr*: an R package to detect, classify, and visualize genome  
24 rearrangements. *Bioinformatics* Available from:  
25 <http://biorxiv.org/lookup/doi/10.1101/2020.06.25.170522>
- 26 Lisch D. 2013. How important are transposons for plant evolution? *Nat Rev Genet* 14:49–61.
- 27 Liu H, Chen C, Lv M, Liu N, Hu Y, Zhang H, Enbody ED, Gao Z, Andersson L, Wang W. 2021.  
28 A Chromosome-Level Assembly of Blunt Snout Bream ( *Megalobrama amblycephala* )

- 1 Genome Reveals an Expansion of Olfactory Receptor Genes in Freshwater Fish.  
2 *Molecular Biology and Evolution*:msab152.
- 3 Liu Z, Roesti M, Marques D, Hiltbrunner M, Saladin V, Peichel CL (2021) Chromosomal fusions  
4 facilitate adaptation to divergent environments in threespine stickleback. *In review*.
- 5 Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into  
6 automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research*  
7 42:e119–e119.
- 8 Mani R-S, Chinnaiyan AM. 2010. Triggers for genomic rearrangements: insights into genomic,  
9 cellular and environmental influences. *Nat Rev Genet* 11:819–829.
- 10 Miller CT, Glazer AM, Summers BR, Blackman BK, Norman AR, Shapiro MD, Cole BL,  
11 Peichel CL, Schluter D, Kingsley DM. 2014. Modular Skeletal Evolution in  
12 Sticklebacks Is Controlled by Additive and Clustered Quantitative Trait Loci. *Genetics*  
13 197:405–420.
- 14 Nelson TC, Cresko WA. 2018. Ancient genomic variation underlies repeated ecological  
15 adaptation in young stickleback populations. *Evolution Letters* 2:9–21.
- 16 Niimura Y, Matsui A, Touhara K. 2014. Extreme expansion of the olfactory receptor gene  
17 repertoire in African elephants and evolutionary dynamics of orthologous gene groups  
18 in 13 placental mammals. *Genome Res.* 24:1485–1496.
- 19 Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role  
20 of restricted recombination in maintaining species. *Heredity* 103:439–444.
- 21 Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the  
22 reproductive isolation of species. *Proceedings of the National Academy of Sciences*  
23 98:12084–12088.
- 24 Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic  
25 divergence. *Molecular Ecology* 18:375–402.
- 26 Nützmann H-W, Osbourn A. 2014. Gene clustering in plant specialized metabolism. *Current*  
27 *Opinion in Biotechnology* 26:91–99.

- 1 Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D,  
2 Peterson T, et al. 2019. Benchmarking transposable element annotation methods for  
3 creation of a streamlined, comprehensive pipeline. *Genome Biol* 20:275.
- 4 Pan YK, Ern R, Morrison PR, Brauner CJ, Esbaugh AJ. 2017. Acclimation to prolonged hypoxia  
5 alters hemoglobin isoform expression and increases hemoglobin oxygen affinity and  
6 aerobic performance in a marine fish. *Sci Rep* 7:7834.
- 7 Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity  
8 in sticklebacks. *Phil. Trans. R. Soc. B* 372:20150486.
- 9 Peichel CL, Sullivan ST, Liachko I, White MA. 2017. Improvement of the Threespine  
10 Stickleback Genome Using a Hi-C-Based Proximity-Guided Assembly. *Journal of*  
11 *Heredity* 108:693–700.
- 12 Petes TD, Hill CW. 1988. Recombination Between Repeated Genes in Microorganisms. *Annu.*  
13 *Rev. Genet.* 22:147–168.
- 14 Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M. 2014. Convergent Genetic Architecture  
15 Underlies Social Organization in Ants. *Current Biology* 24:2728–2732.
- 16 Randall DJ, Rummer JL, Wilson JM, Wang S, Brauner CJ. 2014. A unique mode of tissue  
17 oxygenation and the adaptive radiation of teleost fishes. *Journal of Experimental*  
18 *Biology* 217:1205–1214.
- 19 Rastas P, Calboli FCF, Guo B, Shikano T, Merilä J. 2016. Construction of Ultradense Linkage  
20 Maps with Lep-MAP2: Stickleback F<sub>2</sub> Recombinant Crosses as an Example. *Genome*  
21 *Biol Evol* 8:78–93.
- 22 Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology and*  
23 *evolution* 16:351–358.
- 24 Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of Sex Chromosomes in  
25 the Stickleback Fishes (Gasterosteidae). *PLoS Genet* 5:e1000391.
- 26 Samuk K, Owens GL, Delmore KE, Miller SE, Rennison DJ, Schluter D. 2017. Gene flow and  
27 selection interact to promote adaptive divergence in regions of low recombination. *Mol*  
28 *Ecol* 26:4378–4390.

- 1 Schlötterer C. 2015. Genes from scratch – the evolutionary fate of de novo genes. *Trends in*  
2 *Genetics* 31:215–219.
- 3 Schluter D, Conte GL. 2009. Genetics and ecological speciation. *Proceedings of the National*  
4 *Academy of Sciences* 106:9955–9962.
- 5 Schluter D, Marchinko KB, Arnegard ME, Zhang H, Brady SD, Jones FC, Bell MA, Kingsley  
6 DM. 2021. Fitness maps to a large-effect locus in introduced stickleback populations.  
7 *Proceedings of the National Academy of Sciences* 118:e1914889118.
- 8 Schwander T, Libbrecht R, Keller L. 2014. Supergenes and Complex Phenotypes. *Current*  
9 *Biology* 24:R288–R294.
- 10 Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D,  
11 Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction  
12 in threespine sticklebacks. *Nature* 428:717–723.
- 13 Slot JC, Gluck-Thaler E. 2019. Metabolic gene clusters, fungal diversity, and the generation of  
14 accessory functions. *Current Opinion in Genetics & Development* 58–59:17–24.
- 15 Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio  
16 prediction of alternative transcripts. *Nucleic Acids Research* 34:W435–W439.
- 17 Storz JF. 2016. Gene Duplication and Evolutionary Innovations in Hemoglobin-Oxygen  
18 Transport. *Physiology* 31:223–232.
- 19 Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity* 113:1–8.
- 20 Urton JR, McCann SR, Peichel CL. 2011. Karyotype Differentiation between Two Stickleback  
21 Species (Gasterosteidae). *Cytogenet Genome Res* 135:150–159.
- 22 Van Oss SB, Carvunis A-R. 2019. De novo gene birth. *PLoS Genet* 15:e1008160.
- 23 Vandewege MW, Mangum SF, Gabaldón T, Castoe TA, Ray DA, Hoffmann FG. 2016.  
24 Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile  
25 and bird genomes. *Genome Biol Evol*:evw013.
- 26 Varadharajan S, Rastas P, Löytynoja A, Matschiner M, Calboli Federico C F, Guo B, Nederbragt  
27 AJ, Jakobsen KS, Merilä J. 2019. A high-quality assembly of the nine-spined  
28 stickleback (*Pungitius pungitius*) genome. *Genome Biology and Evolution* 11:3291-  
29 3308.
- 30 Verta J-P, Jones FC. 2019. Predominance of cis-regulatory changes in parallel expression  
31 divergence of sticklebacks. *eLife* 8:e43785.

- 1 Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological  
2 speciation-with-gene-flow. *Phil. Trans. R. Soc. B* 367:451–460.
- 3 Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Heusden PV, Singh S,  
4 Thevasagayam NM, Prakki SRS, Purushothaman K, et al. 2016. Chromosomal-Level  
5 Assembly of the Asian Seabass Genome Using Long Sequence Reads and Multi-  
6 layered Scaffolding. *PLOS Genetics*:35.
- 7 Villoutreix R, de Carvalho CF, Soria-Carrasco V, Lindtke D, De-la-Mora M, Muschick M, Feder  
8 JL, Parchman TL, Gompert Z, Nosil P. 2020. Large-scale mutation in the evolution of a  
9 gene complex for cryptic coloration. *Science* 369:460–466.
- 10 Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker D, Keller L.  
11 2013. A Y-like social chromosome causes alternative colony organization in fire ants.  
12 *Nature* 493:664–668.
- 13 Weisman CM, Murray AW, Eddy SR. 2020. Many, but not all, lineage-specific genes can be  
14 explained by homology detection failure. *PLoS Biol* 18:e3000862.
- 15 Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA  
16 and EST sequences. *Bioinformatics* 21:1859–1875.
- 17 Yamasaki YY, Kakioka R, Takahashi H, Toyoda A, Nagano AJ, Machida Y, Møller PR, Kitano  
18 J. 2020. Genome-wide patterns of divergence and introgression after secondary contact  
19 between *Pungitius* sticklebacks. *Phil. Trans. R. Soc. B* 375:20190548.
- 20 Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci.  
21 *Proceedings of the National Academy of Sciences* 110:E1743–E1751.
- 22 Yeaman S, Aeschbacher S, Bürger R. 2016. The evolution of genomic islands by increased  
23 establishment probability of linked alleles. *Mol Ecol* 25:2542–2558.
- 24 Yeaman S, Otto SP. 2011. Establishment and maintenance of adaptive genetic divergence under  
25 migration, selection, and drift. *Evolution* 65:2123–2129.
- 26 Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration-selection  
27 balance: the genetic architecture of local adaptation. *Evolution* 65:1897–1911.
- 28 Yeaman S. 2021. Evolution of polygenic traits under global vs. local adaptation. In press:  
29 *Genetics*.
- 30 Zhao T, Schranz ME. 2019. Network-based microsynteny analysis identifies major differences  
31 and genomic outliers in mammalian and angiosperm genomes. *Proc Natl Acad Sci USA*  
32 116:2165–2174.