



Widespread intersex differentiation across the stickleback genome – The signature of sexually antagonistic selection?

Mirjam Bissegger¹ | Telma G. Laurentino¹ | Marius Roesti² | Daniel Berner¹ 

¹Department of Environmental Sciences, Zoology, University of Basel, Basel, Switzerland

²Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

Correspondence

Daniel Berner, Department of Environmental Sciences, Zoology, University of Basel, 4051 Basel, Switzerland.
Email: daniel.berner@unibas.ch

Funding information

Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 31003A_165826 and P300PA_174344

Abstract

Females and males within a species commonly have distinct reproductive roles, and the associated traits may be under perpetual divergent natural selection between the sexes if their sex-specific control has not yet evolved. Here, we explore whether such sexually antagonistic selection can be detected based on the magnitude of differentiation between the sexes across genome-wide genetic polymorphisms by whole-genome sequencing of large pools of female and male threespine stickleback fish. We find numerous autosomal genome regions exhibiting intersex allele frequency differences beyond the range plausible under pure sampling stochasticity. Alternative sequence alignment strategies rule out that these high-differentiation regions represent sex chromosome segments misassembled into the autosomes. Instead, comparing allele frequencies and sequence read depth between the sexes reveals that regions of high intersex differentiation arise because autosomal chromosome segments got copied into the male-specific sex chromosome (Y), where they acquired new mutations. Because the Y chromosome is missing in the stickleback reference genome, sequence reads derived from DNA copies on the Y chromosome still align to the original homologous regions on the autosomes. We argue that this phenomenon hampers the identification of sexually antagonistic selection within a genome, and can lead to spurious conclusions from population genomic analyses when the underlying samples differ in sex ratios. Because the hemizygous sex chromosome sequence (Y or W) is not represented in most reference genomes, these problems may apply broadly.

KEYWORDS

allele frequency, duplication, *Gasterosteus aculeatus*, genome assembly, population genomics, repetitive DNA, sex chromosome

1 | INTRODUCTION

In organisms with distinct sexes, different female and male reproductive strategies may imply that selective trait optima differ between the sexes (Arnqvist & Rowe, 2005; Darwin, 1871; Shine, 1989; Slatkin, 1984). Because the sexes share most of their genome and alleles typically have similar effects in both sexes (Poissant, Wilson, & Coltman, 2010), this can result in a conflict

in that alleles improving a trait in one sex may push the same trait away from its optimum in the other sex (Arnqvist & Rowe, 2005; Rice & Chippindale, 2001). Such sexually antagonistic selection (SAS) may weaken with the emergence of stable sex-specific gene expression and the associated sexual dimorphism. The resolution of sexual antagonism will typically involve the establishment of a link between a preexisting molecular signal derived from the sex-determination pathway, and a newly gained binding site for that

sex-specific signal controlling the level of expression of the selected gene (Stewart, Pischedda, & Rice, 2010; Williams & Carroll, 2009). This resolution process, requiring at least one highly specific mutation, is suggested to be slow (Stewart et al., 2010) and often appears incomplete in natural populations (Cox & Calsbeek, 2009). Moreover, the presence and strength of SAS may plausibly vary over time and between ecologically different environments. For these reasons, genetic polymorphisms under SAS may well be widespread across the genomes of natural populations and may make a substantial contribution to maintaining genetic variation within these populations (Connallon & Clark, 2014; Cox & Calsbeek, 2009; Rice & Chippindale, 2001).

Recent genomic investigations, performed mainly in genetic model organisms, indeed seem to support the notion that loci under SAS are common within the genome (Cheng & Kirkpatrick, 2016; Dutoit et al., 2018; Griffin, Dean, Grace, Ryden, & Friberg, 2013; Innocenti & Morrow, 2010; Lucotte, Laurent, Heyer, Ségurel, & Toupance, 2016). These investigations typically infer genes putatively under SAS based on the skew in the magnitude of gene expression between the sexes, as estimated by transcriptome analysis. Challenges with this approach include ambiguity in the extent to which sex-biased gene expression indicates current intersexual conflict, and methodological difficulties in estimating sex-bias in gene expression reliably (Mank, 2017; Stewart et al., 2010). In principle, a conceptually simple approach to exploring SAS across a genome without using gene expression data exists: if sexual antagonism occurs throughout ontogeny and thus causes divergent viability selection between the sexes (Cox & Calsbeek, 2009; Rice & Chippindale, 2001; Shine, 1989; Slatkin, 1984), the underlying loci should display allele frequency differentiation between the sexes in the adult stage, with female-beneficial alleles enriched in females and male-beneficial alleles enriched in males. In the beginning of every new generation, however, this intersex differentiation should be erased due to the random assortment of female- and male-beneficial alleles during reproduction. Whether allele frequency differentiation due to divergent viability selection between females and males can be detected in genome-wide screens should depend on the number of antagonistically selected loci, and on the strength of selection on these, thus representing an empirical issue. An analysis in humans suggests that a genome-wide signature of SAS can be detected based on female-male differentiation data alone (Lucotte et al., 2016), but evidence from further organisms is needed.

Here, we investigate potential signatures of SAS based on genome-wide differentiation in allele frequencies between the sexes in threespine stickleback fish (*Gasterosteus aculeatus*). The motivation for this study is two-fold. First, in threespine stickleback, males and females play distinct reproductive roles (Östlund-Nilsson, Mayer, & Huntingford, 2007): during the reproductive period, females allocate resources primarily into egg production, whereas males hold territories and perform brood care. The sexes also appear to exploit distinct ecological niches, as indicated by sexual dimorphism in parasite communities (Reimchen & Nosil, 2001), in predator defence traits

(Reimchen & Nosil, 2004), and in trophic morphology (Aguirre & Akinpelu, 2010; Berner, Roesti, Hendry, & Salzburger, 2010; Bolnick & Lau, 2008; Kitano, Mori, & Peichel, 2007; Kristjansson, Skulason, & Noakes, 2002; Spoljaric & Reimchen, 2008). Sexual dimorphism in trophic morphology is particularly pronounced in habitats in which disruptive selection due to intraspecific resource competition is inferred to be strongest (Bolnick & Lau, 2008). Divergence between the sexes in trophic traits cannot plausibly be ascribed to sexual selection and must therefore reflect differential trait optimization by natural selection within each sex (Darwin, 1871; Rice & Chippindale, 2001; Selander, 1966; Shine, 1989; Slatkin, 1984). The opportunity for sexual antagonism mediated by divergent viability selection during ontogeny thus seems given in this species.

The second impetus to our study was the observation of a few autosomal single-nucleotide polymorphisms (SNPs) showing substantial differentiation between female and male stickleback in a preliminary genomic screen (M. Roesti & D. Berner, unpublished data; example shown in Figure S1). This analysis, however, used sequence data with reduced genomic representation (RAD sequencing) (Roesti, Kueng, Moser, & Berner, 2015) and was based on a low number of individuals from each sex (12 females, 13 males), thus making pattern interpretation difficult. Here, we overcome these methodological limitations by a formal analysis of intersex genetic differentiation across the full stickleback genome based on large sample sizes. As we will show, regions exhibiting strong intersex genetic differentiation are abundant across the stickleback's autosomal genome. Scrutinizing the cause for intersex differentiation in these regions, however, highlights a general methodological challenge to evolutionary genomic analysis, rather than providing evidence of SAS.

2 | MATERIALS AND METHODS

2.1 | Study design, sampling and DNA extraction

Our approach to investigating genomic regions potentially showing signatures of SAS in stickleback was to generate a female and a male pool of DNA, each representing a large number of individuals, to perform whole-genome sequencing of these pools, and to subject the resulting polymorphism data to a genome-wide screen for the magnitude of intersex differentiation.

We used adult, reproductive stickleback individuals sampled from Lake Constance (Switzerland) at the ROM study site (Berner et al., 2010; Moser, Roesti, & Berner, 2012) from April to June 2016 for a behavioural experiment (Berner et al., 2017). Sample size for each sex-specific DNA pool was 120 individuals (i.e., 240 haploid genomes per sex). To standardize the contribution of individual DNA to the final pool, we pierced a disk of 2 mm diameter from the spread caudal fin of each individual by using a biopsy puncher (KAI Medical). Within each sex, these individual tissue samples were combined into 12 subpools of 10 individuals per sex, and the sub-pools subjected to DNA extraction with the Qiagen DNeasy Blood & Tissue kit (Qiagen), including an RNase treatment.

2.2 | DNA pool preparation, sequencing and SNP discovery

After DNA quantitation of the 24 total subpools with a Qubit fluorometer (Thermo Fisher Scientific), they were combined without PCR enrichment at equimolar amounts to a single pool per sex. These pools were barcoded and whole-genome paired-end sequenced to 151 bases in two lanes of an Illumina HiSeq2500 instrument, each lane containing female and male DNA in similar parts. The raw sequence reads were demultiplexed by sex, pooled across the two sequencing lanes, and aligned to the third-generation assembly of the 447 Mb threespine stickleback reference genome (Glazer, Killingbeck, Mitros, Rokhsar, & Miller, 2015; hereafter “reference genome”) by using Novoalign v3.00 (<http://www.novocraft.com/products/novoalign>; settings: -t540, -g40, -x12). The `RSAMTOOLS` R package (Morgan, Pages, Obenchain, & Hayden, 2018) was then used to convert the alignments to BAM files, and to perform nucleotide counts at each base position using the `pileup` function (raw genome-wide nucleotide counts for each sex are provided on the Dryad repository). Median read depth across all genome-wide autosomal positions was 125 for females and 137 for males. Combined with the large number of individuals used for sequencing pool preparation within each sex, this high read depth was expected to allow estimating allele frequencies highly accurately (Ferretti, Ramos-Onsins, & Pérez-Enciso, 2013; Gautier et al., 2013). Next, the nucleotide counts of both sexes were pooled to determine if a given position was variable. SNPs were accepted if they displayed a read depth greater than 100 and lower than 800 across the female-male pool (median: 262), and if the minor allele frequency (MAF) in the pool was at least 0.15. The latter filter effectively removed sequencing errors and excluded SNPs having low sensitivity to capture selective shifts (Roesti, Salzburger, & Berner, 2012). A total of 1.63 million autosomal SNPs passed our read depth and MAF filtering, yielding an expected average marker density of one SNP per 255 bp.

2.3 | Quantifying intersex differentiation through genome scans and simulations

We started our analysis of genomic differentiation between females and males by quantifying and visualizing the magnitude of intersex differentiation, expressed by the absolute allele frequency difference (AFD; Berner, 2019), across all chromosomes (the sex chromosome was included for completeness, although our focus lies on the autosomal genome; data available on Dryad). This genome scan revealed numerous genomic regions showing strong intersex differentiation (see Results and Discussion). Therefore, we next used simulations to compare the magnitude of intersex differentiation observed in the genome-wide scan to levels of differentiation expected under pure sampling stochasticity. Here we thus aimed to develop a sense for the differentiation plausible in the absence of any deterministic factor driving sex bias in allele frequencies, such as SAS. We sampled alleles at random with replacement from a female and from a male pool at a SNP with two alleles occurring at

the same frequency of 0.5 in both sexes. This assumption of the highest possible MAF led to conservative results because it maximized the sampling variance in allele frequencies, thus allowing for maximal intersex differentiation (see Figure 4 in Berner, 2019). The two samples were then used to calculate intersex AFD. Two sample sizes were considered: 50 per sex, approximating the minimum read depth required during SNP calling, and 120 per sex, approximating the median read depth observed (see above). In concordance with our empirical differentiation scan, the simulation included 1.63 million AFD estimates for each sample size.

2.4 | Assessing the role of genome misassembly as cause for high intersex differentiation

Before considering that the genomic regions of high intersex differentiation observed in the above genome scan represented genuine signatures of SAS, it was essential to rule out methodological explanations. In a first step, we performed two analyses based on realignment of our sequence reads. Specifically, threespine stickleback display divergent sex chromosomes (Peichel et al., 2004; Roesti, Moser, & Berner, 2013), with the females representing the homogametic (XX) and males the heterogametic (XY) sex. Strong intersex differentiation may thus simply emerge because homologous X and Y chromosome segments harbouring single-nucleotide differences erroneously align to autosomal regions. This may occur due to either genome sequence divergence between our focal population (derived from an Atlantic marine ancestor) and the reference genome (representing an individual derived from a Pacific ancestor; Jones et al., 2012), or the incorrect placement of sex chromosome segments on autosomes in the reference genome assembly. To explore these possibilities, we assessed whether regions of strong differentiation still persisted when performing more stringent alignment (i.e., tolerating much lower sequence mismatch: -t200), which should reduce the likelihood of sex chromosome segments to erroneously align to autosomes. The sequence alignments resulting from this alternative alignment approach were used for a genome-wide scan for the magnitude of intersex differentiation as described above.

In addition, we aligned our raw sequence reads to a new threespine stickleback genome sequenced and assembled *de novo* (Berner et al., 2019), using the initial alignment settings. This new genome was derived from an individual sampled from the same watershed as our study population, thus ensuring minimal sequence divergence. The resulting sequence alignments were again used for a genome scan for intersex differentiation, which also indicated numerous regions of high differentiation. To assess whether these regions in the *de novo* genome corresponded to high-differentiation regions in our original scan, we chose a 151 bp sequence overlapping a high-differentiation SNP from a dozen of strongly differentiated regions located on different scaffolds of the *de novo* genome. We then evaluated visually the magnitude of differentiation in the 50 kb (kilobases) neighbourhood around the alignment position of these sequences within the reference genome.

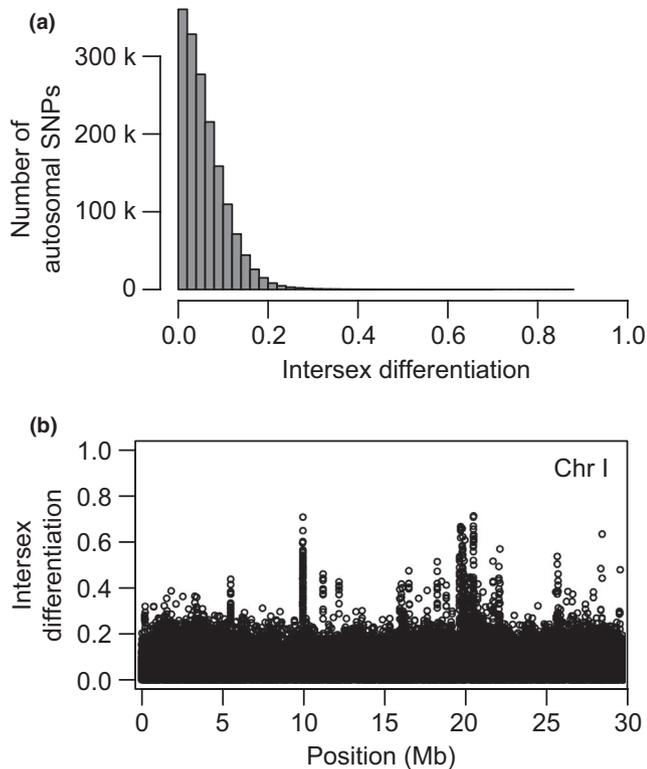


FIGURE 1 (a) Distribution of the magnitude of genetic differentiation between female and male stickleback, as quantified by the absolute allele frequency difference AFD, across 1.63 million autosomal single-nucleotide polymorphisms. (b) Intersex differentiation mapped along a representative autosome

2.5 | Testing if high intersex differentiation is driven by the lack of the Y chromosome sequence in the reference genome

After examining the possibility that autosomal regions of high differentiation emerged because of erroneous alignment of X and Y chromosome segments to autosomes, we evaluated a second methodological explanation. We here considered that both the reference genome and the new *de novo* genome are derived from a female (XX) individual. The Y chromosome is therefore necessarily missing in these genome assemblies. DNA segments closely related between autosomes and the Y chromosome may thus cause the alignment of Y-specific alleles to autosomes, thus potentially producing SNPs showing high intersex differentiation. This scenario leads to two testable predictions (see also Dou et al., 2012; McKinney, Waples, Seeb, & Seeb, 2017; Tsai, Evans, Noorai, Starr-Moss, & Clark, 2019). First, the SNPs defining regions of high differentiation on autosomes should display a systematically higher MAF in the male than female pool because only males harbour the Y-specific allele that makes the given genome position polymorphic. Second, these SNPs should represent exclusively autosomal DNA in the females but autosomal plus Y chromosome segments in the males, and hence exhibit higher read depth in the male than female pool.

To test these two predictions, we first delimited a focal set of autosomal regions exhibiting high intersex differentiation (HIDRs).

Based on the distribution of intersex differentiation values observed empirically on the one hand, and the simulated distribution of differentiation under pure sampling stochasticity on the other (see below), HIDRs were required to harbour at least five SNPs showing AFD of 0.5 or greater within a window of 5 kb. HIDRs further needed to be spaced by at least 100 kb from any other such region to ensure independence. Given these criteria, we identified a total of 38 autosomal HIDRs. For each HIDR, we next selected at random a single representative high-differentiation SNP (AFD ≥ 0.5) exhibiting a sex-specific read depth of at least 50-fold, hereafter called HIDR SNP. To obtain negative controls for statistical analysis, we also selected a control SNP for each HIDR, defined as the SNP closest to the genomic position located 30 kb upstream of the corresponding HIDR SNP and passing the same read depth thresholds. For both SNP classes (i.e., HIDR and control), we then explored if there was sex-related skew in the MAF, and in read depth (quantified as read depth ratio, i.e., the nucleotide count of the male pool divided by the count of the female pool). The MAF data were analyzed visually based on histograms, while for the read depth ratio we calculated median values for each SNP class along with their 95% bootstrap confidence intervals generated by 10,000 resamples (Manly, 2006).

2.6 | Simulations exploring intersex differentiation in relation to selection strength

The above empirical analyses indicated that our detected HIDRs represented methodological artefacts (see Results and Discussion). To complement this evidence by theory, we additionally performed stochastic individual-based simulations exploring the magnitude of intersex differentiation resulting from SAS of different strengths on a single locus. The objective of this simulation analysis was not a comprehensive theoretical treatment, but to gain qualitative insight into the (im)plausibility of our HIDRs to reflect signatures of SAS.

We implemented a model starting with a population of 100,000 diploid individuals showing a balanced sex ratio. The locus under selection was biallelic with one allele favorable in females and the other allele favorable in males (we thus assumed perfectly symmetric divergent selection, recognizing that in reality, the strength of selection on a polymorphism may differ between the sexes). The starting frequency of both alleles was 0.5. We modelled viability selection, as required if SAS should drive intersex differentiation within a generation, by making access to mating dependent on the genotype at the locus under selection (Berner & Roesti, 2017; Berner & Thibert-Plante, 2015). Specifically, an individual's probability of surviving to the reproductive stage was a stochastic function of the individual's deviation from the sex-specific optimum genotype. This deviation was determined by the number of unfavorable alleles times the selection coefficient, resulting in additive fitness. The genotypes of the females and males surviving to the reproductive stage were used to quantify the magnitude of intersex AFD observed after SAS within the focal generation. These individuals then mated at random, each pair producing a constant

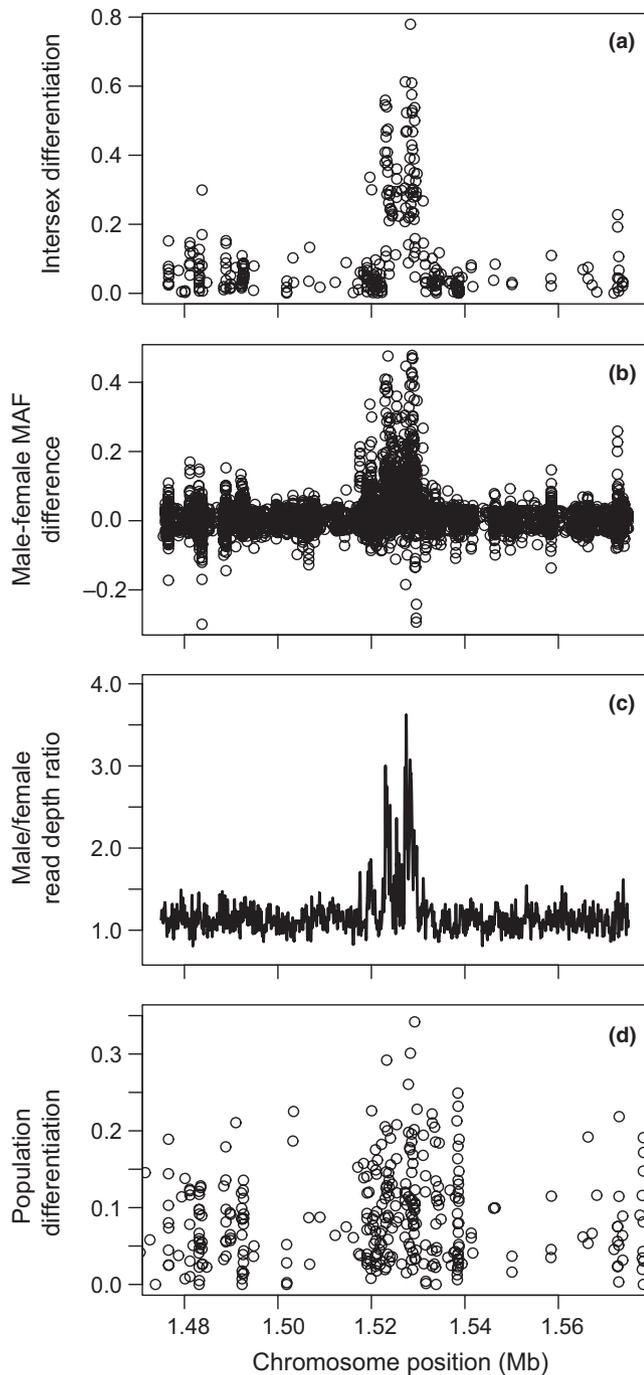


FIGURE 2 Characterization of a 100 kb segment on chromosome XI containing a representative region of high differentiation between female and male stickleback. (a) Genetic differentiation (AFD) between the sexes at single-nucleotide polymorphisms (SNPs) showing a pooled minor allele frequency (MAF) of at least 0.15. (b) Difference between the sexes in the MAF, considering all SNPs passing a pooled MAF threshold of 0.01. Positive values indicate that the two alleles at a SNP occur in more balanced proportion in males than in females. (c) Read depth in males standardized by the depth in females. High values indicate that male reads are relatively overrepresented in the sequencing output overlapping the corresponding genome positions. Note that because this statistic is calculated for every base position (not just the SNPs), a smoother (LOESS; moving average with a span of 0.002) was chosen for visualization to reduce complexity. (d) Genetic differentiation (AFD) between two population samples with symmetrical sex bias in opposite directions generated by resampling empirically observed female and male nucleotide data at each SNP

3 | RESULTS

3.1 | Regions of strong intersex differentiation are widespread across stickleback autosomes

Allele frequency differentiation (AFD) between stickleback females and males showed a median magnitude of 0.053 across all genome-wide autosomal SNPs, but the distribution tapered off to a long tail reaching values up to 0.87 (Figure 1a). The latter strong intersex differentiation cannot be explained by pure sampling stochasticity, as revealed by comparing the empirical distribution of differentiation values to simulated distributions: even when modeling minimal sample size ($N = 50$) for each sex, and hence low precision in allele frequency estimation, differentiation values above 0.5 did not emerge across the 1.63 million replications (Figure S2). Assuming sample sizes more typical of the read depth of our data set ($N = 120$), the top differentiation value observed in the simulations dropped to 0.32. Given that the simulations assumed the highest MAF possible (0.5; i.e., both alleles occurring in perfectly balanced proportions), even the latter upper simulation limit for differentiation due to sampling variation alone must be considered cautiously high. Nevertheless, we used an AFD threshold of 0.5 for the identification of high-differentiation regions (HIDRs) in the analyses below.

Exploring the physical distribution of intersex differentiation values along chromosomes revealed narrow regions (typically a few kb wide) of high differentiation standing out clearly against background differentiation on all autosomes (Figure 1b; Figure 2a shows a representative example in high physical resolution, reanalyzed using F_{ST} as differentiation metric in Figure S3a; the complete differentiation plots for all chromosomes are presented as Figure S4).

3.2 | Reference genome misassembly is not the cause for high intersex differentiation on autosomes

A chromosome exhibiting particularly extensive intersex differentiation along almost its entire length was the sex chromosome

number of offspring ($N = 10$; using four or 20 offspring produced similar results; details not presented) that overall exactly re-established initial population size. Offspring sex was assigned at random. We considered selection coefficients of 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5, the latter representing the complete unviability (zero fitness) of individuals homozygous for the unfavourable allele. For each selection coefficient, we carried out 10 replicate simulations, each running for 20 generations. We thus obtained a total of 200 estimates of within-generation intersex differentiation for a given selection strength. The simulation code is available on Dryad. Unless specified otherwise, all analyses were performed with the R language (R Development Core Team, 2017).

(chromosome XIX, Figure S4). Along this chromosome, differentiation primarily reflects the evolutionary divergence between the nonrecombining regions of the X and Y sequences, with an additional contribution from reduced precision in allele frequency estimation in the hemizygous males (i.e., in males, the X chromosome occurs in a single copy only, thus causing systematically lighter read depth in the male pool). This observation motivated investigating whether regions of high intersex differentiation may be explained by the incorrect placement of DNA segments homologous but polymorphic between the X and Y chromosome into autosomes during reference genome assembly. Inconsistent with this idea, a genome scan for intersex differentiation based on sequence reads aligned to the reference genome with more stringent alignment settings did not produce results differing qualitatively from our initial genome scan: although read alignment success dropped from 81% to 69% with more stringent alignment, genomic regions showing high intersex differentiation in the initial genome scan were generally still present (details not presented). Similarly, aligning our sequence reads to a *de novo* stickleback genome assembly derived from an individual originating from the same watershed as our study population still revealed numerous genomic regions of high intersex differentiation. These regions consistently coincided with autosomal regions of high differentiation in our initial genome scan based on the reference genome (three examples are shown in Figure S5). Together, these two analyses using alternative alignment strategies make clear that the incorrect placing of sex chromosome segments within autosomes in the stickleback reference genome assembly fails as a general explanation for autosomal regions of high intersex differentiation.

3.3 | High intersex autosomal differentiation arises from DNA segments shared between autosomes and the Y chromosome

Having ruled out reference genome misassembly as an explanation for strong autosomal differentiation between the sexes, we addressed a second hypothesis focused on reference genome incompleteness: that DNA segments similar to autosomal chromosome regions occur on the Y chromosome that is not part of any current genome assembly, and that these segments harbour private genetic variants that cause intersex differentiation when aligning to their autosomal counterparts (see also Tsai et al., 2019). Consistent with this idea, we observed that SNPs located within HIDRs showed a systematically reduced MAF in the female relative to the male pool (Figure 2b). More specifically, the majority of HIDR SNPs showed a female MAF of zero (i.e., monomorphism for one allele), while the male frequency was near 0.5 (i.e., the two SNP alleles occurred at relatively balanced frequency) (Figure 3 top). By contrast, the control SNPs showed a relatively uniform distribution of MAFs in both sexes (Figure 3 bottom). These observations make clear that the polymorphisms driving HIDRs arise from derived alleles restricted to the males.

The most plausible explanation for such male-specific alleles is that the DNA segments harbouring these alleles are located on the Y

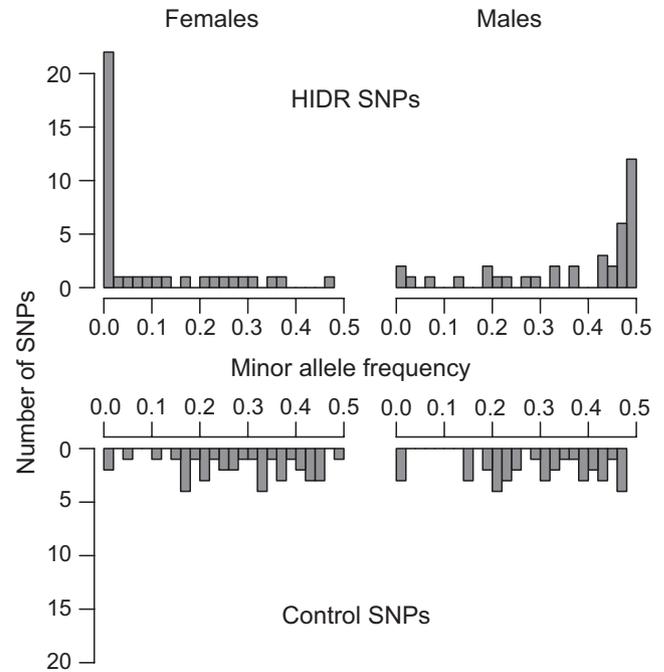


FIGURE 3 Frequency of the minor allele in females and males at 38 SNPs representing independent regions of high intersex differentiation (HIDR SNPs), and at their associated control SNPs

chromosome. A unique prediction derived from this scenario is that the chromosome segments around HIDR SNPs should display elevated read depth in the male relative to the female sex. The reason is that only in males, these segments should recruit truly autosomal plus Y-chromosomal sequence reads aligning to the same location in the genome assembly. This prediction was confirmed unambiguously: the SNPs driving HIDRs consistently exhibited elevated read depth in males compared to females (Figure 2c and 4). Such bias was absent in the control SNPs. (Note that the slight imbalance between the sexes at the control SNPs in Figure 4 is expected because the male DNA pool was sequenced to approximately 10% higher read depth, see Materials and Methods). Interestingly, for the HIDR SNPs, the male-female read depth ratio showed a median of 2.18 (control SNPs: 1.06), with several SNPs displaying values beyond three. If an autosomal segment was present as a single copy on the Y chromosome, however, one would expect a read depth ratio of 1.5. This leads us to propose a model in which an autosomal DNA segment is first copied to the Y chromosome (see also Koerich, Wang, Clark, & Carvalho, 2008; Tsai et al., 2019), experiences mutation at the new location, and then, to variable extent, experiences further copy number expansion on the Y chromosome (Figure 5). Consistent with this model, the male-female read depth ratios of the HIDR SNPs tended to form distinct clusters overlapping with 1.5, 2, and 2.5 (Figure 4), as expected for autosomal segments falling into discrete copy number classes on the Y chromosome. Although the Y chromosome sequence of threespine stickleback is not yet available, our conceptual model is supported by the indication of an exceptionally high proportion of repeated DNA on a preliminary Y chromosome assembly as compared to all autosomes (M. White & C.

Peichel, personal communication; see also Chalopin, Volff, Galiana, Anderson, & Schartl, 2015; Hobza et al., 2017). As a definitive validation of our model, it would be worthwhile to determine the number of alignment sites of DNA segments representative of our HIDRs in a future Y chromosome assembly.

3.4 | Simulations confirm the implausibility of sexually antagonistic selection as a cause for high autosomal intersex differentiation

Our empirical analyses clearly identified a methodological, nonselective explanation for regions of strong differentiation between the sexes across the stickleback genome. To nevertheless develop a sense for the magnitude of intersex differentiation in allele frequencies that viability selection could drive within a single generation, we used simulations of SAS on a single locus. We found that under the strongest selection considered - a heterozygous selection coefficient of 0.5 - the sexes reach an allele frequency differentiation

of 0.4 within each generation (Figure S6). Under such strong selection, a quarter of all individuals within each sex are expected to be excluded from reproduction (that is, to die during juvenile life) because of their maladaptive genotype at a single locus. Given that we observed dozens of genome regions showing even stronger intersex differentiation (Figures 1a, 2a, Figure S4), it becomes clear from a purely theoretical perspective that SAS fails as a viable explanation for widespread intersex differentiation in our stickleback system; the total selection imposed by dozens of loci under such strong selection would be so intense that the population would go extinct rapidly.

4 | DISCUSSION

4.1 | Analytical implications

Our investigation has identified an alternative to sexually antagonistic selection as a cause for strong and widespread intersex allelic differentiation across autosomes: the copying of autosomal chromosome segments into a sex chromosome not represented in the reference genome assembly ('autosomal' here includes the pseudoautosomal region of the sex chromosome, as this regions also harboured SNPs exhibiting high intersex differentiation; Figure S4). Our work in no way challenges the notion that SAS could be widespread across the genome. However, the above and previous (Kasimatis, Nelson, & Phillipps, 2017) theoretical considerations indicate that intersex differentiation maintained by continuous sexually antagonistic viability selection within a population should be subtle in magnitude. The much stronger intersex differentiation arising artificially from incomplete genome assembly is thus likely to preclude the reliable investigation of the genomic consequences of SAS based exclusively on intersex differentiation data in this and analogous study systems. Although one could consider filtering genome regions based on the difference in MAF and/or imbalance in read depth between the sexes, we doubt that this would completely eliminate spurious autosomal signals of SAS. The reason is that sex-related genetic differentiation and differences in MAF and read depth due to the mechanism described in Figure 5 may well remain subtle if an autosomal segment

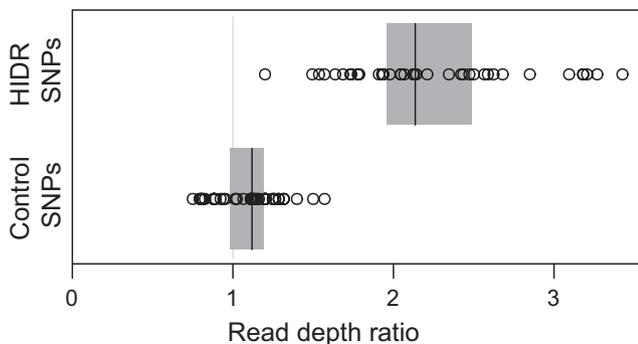


FIGURE 4 Ratio of the male by female read depth at the HIDR and control SNPs. The raw data points are shown along with their median (black vertical line) and the 95% bootstrap confidence interval for the median (grey box) within each SNP class. The grey vertical line indicates balanced read depth between the sexes (note that all observed read depth ratios are slightly biased upward due to deeper sequencing of the male pool). To increase visual resolution, a single HIDR SNP showing an extreme read depth ratio (4.89) was omitted

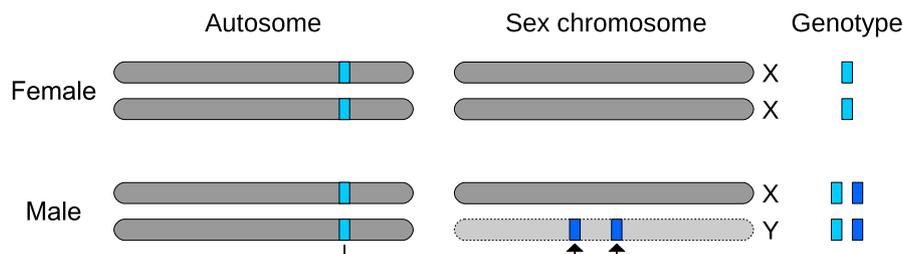


FIGURE 5 Schematic of the model explaining the emergence of HIDRs when a sex chromosome is missing in the genome assembly. First, an autosomal DNA segment (light blue) is copied into the Y chromosome. Mutation then generates polymorphisms distinguishing the original autosomal segment from its copy on the Y (indicated by the distinct blue shades). The Y-copy may then become multiplied further on that chromosome. As a consequence, DNA sequences from both the autosomal segment and its copies on the Y align to the same autosomal location when the reference genome lacks the Y chromosome. The analytical outcome is that males tend to display a more variable genotype (hence a higher MAF) than the females, and hence that the sexes show substantial allele frequency differentiation, at the distinctive polymorphisms. Moreover, male read depth is elevated across the entire focal DNA segment relative to females [Colour figure can be viewed at wileyonlinelibrary.com]

harbouring a distinct genetic variant was copied relatively recently to the Y chromosome and still segregates at low frequency in the new chromosomal location. The availability of a complete genome assembly including both sex chromosomes, and the rigorous elimination of sequences aligning to any of them, may potentially allow detecting genome-wide signatures of SAS based on intersex differentiation data alone (Lucotte et al., 2016), although the reliability of such approaches awaits validation. We also note that if the transfer of autosomal sequences to the Y chromosome includes genes that retain expression in the new location (Mahajan & Bachtrog, 2017; Tsai et al., 2019), autosomal genes may appear to show concurrent intersex differences in both allele frequency and gene expression levels when ignoring copies on a missing sex chromosome.

In the vast majority of organisms used for genomic investigations, the Y (or W) chromosome sequence is not available or incomplete, thus providing the opportunity for spurious intersex differentiation due to sex chromosome evolution. This has immediate implications to population genomics: in marker-based comparisons of populations, localized genome regions exhibiting high differentiation – often interpreted as hotspots harbouring polymorphisms targeted by divergent selection between the populations – may emerge simply because the population samples differ in their proportion of females and males, and hence in the proportion of the two sex chromosomes (Benestan et al., 2017). To illustrate this point in our system, we drew 42 total nucleotides without replacement from the female and male nucleotide pool at all SNPs located within the chromosome window shown in Figure 2. Next, we combined 14 nucleotides from the female pool with 28 nucleotides from the male pool to obtain a first population sample, while the exactly opposite sexual representation was chosen for the second population sample. The outcome thus mimicked two random samples of 21 total diploid individuals from the same biological population, differing, beyond stochasticity in allele sampling, only in the sex ratio. We then calculated the magnitude of population differentiation across this chromosome window and observed, as expected, that the SNPs showing the highest population differentiation co-localized with the peaks in intersex differentiation (compare Figure 2d to 2a; Figure S3 shows this comparison based on F_{ST}). Ignoring imbalance in sex ratio may thus mislead the interpretation of patterns in population differentiation. This echoes an analogous caveat raised recently in a study of two species (American lobster and Arctic Char) in which sex-specific differentiation outliers were observed in genome scans comparing the sexes (Benestan et al., 2017). However, in that study, reference genomes for the focal species were not available. HIDRs were therefore interpreted to reflect divergence between chromosome regions evolving sex-specifically, but the HIDRs could not be physically localized reliably. Our stickleback work extends these insights: even in an organism with a well characterized sex determination system and an identified sex chromosome, HIDRs can occur on autosomes when one sex chromosome is missing (or incomplete) in the genome assembly and population samples differ in sex ratios. We also highlight the possibility that under these conditions, HIDRs may be influential enough to bias marker-based genomic analyses beyond simple differentiation, such as phylogenies or demographic reconstruction.

ACKNOWLEDGEMENTS

We thank Dario Moser for supporting field sampling, Elodie Burcklen and Christian Beisel for carrying out the sequencing at the Department of Biosystems Science and Engineering (D-BSEE, ETH Zürich), and the developers of Novocraft for sharing their aligner. Analysis of whole-genome data was performed at the Center for Scientific Computing at the University of Basel (sciCORE), and was aided by Francisco Pina-Martins. Katie Peichel and Astrid Böhne provided valuable suggestions for analysis. This study was supported by the Swiss National Science Foundation SNF (grants 31003A_165826 to DB and P300PA_174344 to MR), and by the University of Basel.

AUTHOR CONTRIBUTIONS

M.B. performed wet laboratory work, carried out analyses, and wrote a first manuscript draft. T.G.L. performed wet laboratory work, wrote code for and performed cluster computation. M.R. devised the study and performed preliminary analyses. D.B. designed and supervised the experiment, wrote analytical code, analyzed and interpreted data, and wrote the final manuscript, with feedback from M.R. and T.G.L.

DATA AVAILABILITY STATEMENT

Raw sequence reads for the female and male pool are available from NCBI's sequence read archive under the BioSample accession numbers SAMN12777444 and SAMN12777445. Nucleotide counts across all genome-wide positions for the female and male pool, female-male differentiation across all genome-wide SNPs, descriptive information on the HIDR SNPs, and the R code used for simulations of intersex differentiation are available on Dryad (<https://doi.org/10.5061/dryad.pzgmsbcfn>).

ORCID

Daniel Berner  <https://orcid.org/0000-0003-3480-9046>

REFERENCES

- Aguirre, W. E., & Akinpelu, O. (2010). Sexual dimorphism of head morphology in threespine stickleback (*Gasterosteus aculeatus*). *Journal of Fish Biology*, 77, 802–821.
- Arnqvist, G., & Rowe, L. (2005). *Sexual Conflict. Monographs in Behavior and Ecology*. Princeton, NJ: Princeton University Press.
- Benestan, L., Moore, J.-S., Sutherland, B. J. G., Le Luyer, J., Maaroufi, H., Rougeux, C., ... Bernatchez, L. (2017). Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex determination systems. *Molecular Ecology*, 26, 6767–6783. <https://doi.org/10.1111/mec.14217>
- Berner, D. (2019). Allele frequency difference AFD – An intuitive alternative to FST for quantifying genetic population differentiation. *Genes*, 10, 308. <https://doi.org/10.3390/genes10040308>
- Berner, D., Ammann, M., Spencer, E., Rüegg, A., Lüscher, D., & Moser, D. (2017). Sexual isolation promotes divergence between parapatric lake and stream stickleback. *Journal of Evolutionary Biology*, 30, 401–411. <https://doi.org/10.1111/jeb.13016>

- Berner, D., & Roesti, M. (2017). Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. *Molecular Ecology*, *26*, 6351–6369. <https://doi.org/10.1111/mec.14373>
- Berner, D., Roesti, M., Bilobram, S., Chan, S. K., Kirk, H., Pandoh, P., ... DeFaverj, J. (2019). *De novo* sequencing, assembly, and annotation of four threespine stickleback genomes based on microfluidic partitioned DNA libraries. *Genes*, *10*, 426. <https://doi.org/10.3390/genes10060426>
- Berner, D., Roesti, M., Hendry, A. P., & Salzburger, W. (2010). Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. *Molecular Ecology*, *19*, 4963–4978. <https://doi.org/10.1111/j.1365-294X.2010.04858.x>
- Berner, D., & Thibert-Plante, X. (2015). How mechanisms of habitat preference evolve and promote divergence with gene flow. *Journal of Evolutionary Biology*, *28*, 1641–1655. <https://doi.org/10.1111/jeb.12683>
- Bolnick, D. I., & Lau, O. L. (2008). Predictable patterns of disruptive selection in stickleback in postglacial lakes. *American Naturalist*, *172*, 1–11. <https://doi.org/10.1086/587805>
- Chalopin, D., Volf, J.-N., Galiana, D., Anderson, J. L., & Schartl, M. (2015). Transposable elements and early evolution of sex chromosomes in fish. *Chromosome Research*, *23*, 545–560. <https://doi.org/10.1007/s10577-015-9490-8>
- Cheng, C. D., & Kirkpatrick, M. (2016). Sex-specific selection and sex-biased gene expression in humans and flies. *PLOS Genetics*, *12*, e1006170. <https://doi.org/10.1371/journal.pgen.1006170>
- Connallon, T., & Clark, A. G. (2014). Balancing selection in species with separate sexes: Insights from Fisher's geometric model. *Genetics*, *197*, 991–1006. <https://doi.org/10.1534/genetics.114.165605>
- Cox, R., & Calsbeek, R. (2009). Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *The American Naturalist*, *173*, 176–187. <https://doi.org/10.1086/595841>
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London, UK: John Murray.
- Dou, J., Zhao, X., Fu, X., Jia, W., Wang, N., Zhang, L., ... Bao, Z. (2012). Reference-free SNP calling: Improved accuracy by preventing incorrect calls from repetitive genomic regions. *Biology Direct*, *7*, 17. <https://doi.org/10.1186/1745-6150-7-17>
- Dutoit, L., Mugal, C. F., Bolivar, P., Wang, M., Nadachowska-Brzyska, K., Smeds, L., ... Ellegren, H. (2018). Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (*Ficedula albicollis*) genome. *Molecular Ecology*, *27*, 3572–3581.
- Ferretti, L., Ramos-Onsins, S. E., & Pérez-Enciso, M. (2013). Population genomics from pool sequencing. *Molecular Ecology*, *22*, 5561–5576. <https://doi.org/10.1111/mec.12522>
- Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., ... Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: Pool-versus individual-based genotyping. *Molecular Ecology*, *22*, 3766–3779. <https://doi.org/10.1111/mec.12360>
- Glazer, A. M., Killingbeck, E. E., Mitros, T., Rokhsar, D. S., & Miller, C. T. (2015). Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3-Genes Genomes Genetics*, *5*, 1463–1472. <https://doi.org/10.1534/g3.115.017905>
- Griffin, R. M., Dean, R., Grace, J. L., Ryden, P., & Friberg, U. (2013). The shared genome is a pervasive constraint on the evolution of sex-biased gene expression. *Molecular Biology and Evolution*, *30*, 2168–2176. <https://doi.org/10.1093/molbev/mst121>
- Hobza, R., Cegan, R., Jesionek, W., Kejnovsky, E., Vyskot, B., & Kubat, Z. (2017). Impact of repetitive elements on the Y chromosome formation in plants. *Genes*, *8*, 302. <https://doi.org/10.3390/genes8110302>
- Innocenti, P., & Morrow, E. H. (2010). The sexually antagonistic genes of *Drosophila melanogaster*. *PLOS Biology*, *8*, e1000335. <https://doi.org/10.1371/journal.pbio.1000335>
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*, 55–61. <https://doi.org/10.1038/nature10944>
- Kasimatis, K. R., Nelson, T. C., & Philipps, P. C. (2017). Genomic signatures of sexual conflict. *Journal of Heredity*, *108*, 780–790. <https://doi.org/10.1093/jhered/esx080>
- Kitano, J., Mori, S., & Peichel, C. L. (2007). Sexual dimorphism in the external morphology of the threespine stickleback (*Gasterosteus aculeatus*). *Copeia*, *2007*, 336–349. [https://doi.org/10.1643/0045-8511\(2007\)7\[336:SDITEM\]2.0.CO;2](https://doi.org/10.1643/0045-8511(2007)7[336:SDITEM]2.0.CO;2)
- Koerich, L. B., Wang, X. Y., Clark, A. G., & Carvalho, A. B. (2008). Low conservation of gene content in the *Drosophila* Y chromosome. *Nature*, *456*, 949–951. <https://doi.org/10.1038/nature07463>
- Kristjánsson, B. K., Skulason, S., & Noakes, D. L. G. (2002). Morphological segregation of Icelandic threespine stickleback (*Gasterosteus aculeatus* L.). *Biological Journal of the Linnean Society*, *76*, 247–257. <https://doi.org/10.1046/j.1095-8312.2002.00063.x>
- Lucotte, E. A., Laurent, R., Heyer, E., Ségurel, L., & Toupance, B. (2016). Detection of allelic frequency differences between the sexes in humans: A signature of sexually antagonistic selection. *Genome Biology and Evolution*, *8*, 1489–1500. <https://doi.org/10.1093/gbe/evw090>
- Mahajan, S., & Bachtrog, D. (2017). Convergent evolution of Y chromosome gene content in flies. *Nature Communications*, *8*, 785. <https://doi.org/10.1038/s41467-017-00653-x>
- Mank, J. E. (2017). The transcriptional architecture of phenotypic dimorphism. *Nature Ecology and Evolution*, *1*, 6. <https://doi.org/10.1038/s41559-016-0006>
- Manly, B. F. J. (2006). *Randomization, bootstrap and monte carlo methods in biology* (3rd ed.). Boca Raton, FL: Chapman & Hall.
- McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Research*, *17*, 656–669. <https://doi.org/10.1111/1755-0998.12613>
- Morgan, M., Pages, H., Obenchain, V., & Hayden, N. (2018). *Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import*. R package version 1.3.0. Retrieved from <http://bioconductor.org/packages/release/bioc/html/Rsamtools.html>
- Moser, D., Roesti, M., & Berner, D. (2012). Repeated lake-stream divergence in stickleback life history within a Central European lake basin. *PLoS ONE*, *7*, e50620. <https://doi.org/10.1371/journal.pone.0050620>
- Östlund-Nilsson, S., Mayer, I., & Huntingford, F. A. (2007). *Biology of the three-spined stickleback*. Boca Raton, FL: CRC.
- Peichel, C. L., Ross, J. A., Matson, C. K., Dickson, M., Grimwood, J., Schmutz, J., ... Kingsley, D. M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Current Biology*, *14*, 1416–1424. <https://doi.org/10.1016/j.cub.2004.08.030>
- Poissant, J., Wilson, A. J., & Coltman, D. W. (2010). Sex-specific genetic variance and the evolution of sexual dimorphism: A systematic review of cross-sex genetic correlations. *Evolution*, *64*, 97–107. <https://doi.org/10.1111/j.1558-5646.2009.00793.x>
- R Core Team, (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Reimchen, T. E., & Nosil, P. (2001). Ecological causes of sex-biased parasitism in threespine stickleback. *Biological Journal of the Linnean Society*, *73*, 51–63. <https://doi.org/10.1111/j.1095-8312.2001.tb01346.x>
- Reimchen, T. E., & Nosil, P. (2004). Variable predation regimes predict the evolution of sexual dimorphism in a population of threespine stickleback. *Evolution*, *58*, 1274–1281. <https://doi.org/10.1111/j.0014-3820.2004.tb01706.x>

- Rice, W. R., & Chippindale, A. K. (2001). Intersexual ontogenetic conflict. *Journal of Evolutionary Biology*, 14, 685–693. <https://doi.org/10.1046/j.1420-9101.2001.00319.x>
- Roesti, M., Kueng, B., Moser, D., & Berner, D. (2015). The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications*, 6, 8767.
- Roesti, M., Moser, D., & Berner, D. (2013). Recombination in the threespine stickleback genome – Patterns and consequences. *Molecular Ecology*, 22, 3014–3027. <https://doi.org/10.1111/mec.12322>
- Roesti, M., Salzburger, W., & Berner, D. (2012). Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, 12, 94. <https://doi.org/10.1186/1471-2148-12-94>
- Selander, R. K. (1966). Sexual dimorphism and differential niche utilization in birds. *Condor*, 68, 113–151. <https://doi.org/10.2307/1365712>
- Shine, R. (1989). Ecological causes for the evolution of sexual dimorphism – A review of the evidence. *The Quarterly Review of Biology*, 64, 419–461. <https://doi.org/10.1086/416458>
- Slatkin, M. (1984). Ecological causes of sexual dimorphism. *Evolution*, 38, 622–630. <https://doi.org/10.1111/j.1558-5646.1984.tb00327.x>
- Spoljaric, M. A., & Reimchen, T. E. (2008). Habitat-dependent reduction of sexual dimorphism in geometric body shape of Haida Gwaii threespine stickleback. *Biological Journal of the Linnean Society*, 95, 505–516. <https://doi.org/10.1111/j.1095-8312.2008.01068.x>
- Stewart, A. D., Pischedda, A., & Rice, W. R. (2010). Resolving intralocus sexual conflict: Genetic mechanisms and time frame. *Journal of Heredity*, 101, S94–S99. <https://doi.org/10.1093/jhered/esq011>
- Tsai, K. L., Evans, J. M., Noorai, R. E., Starr-Moss, A. N., & Clark, L. A. (2019). Novel Y chromosome retrocopies in canids revealed through a genome-wide association study for sex. *Genes*, 10, 320. <https://doi.org/10.3390/genes10040320>
- Williams, T. M., & Carroll, S. B. (2009). Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nature Reviews Genetics*, 10, 797–804. <https://doi.org/10.1038/nrg2687>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Bissegger M, Laurentino TG, Roesti M, Berner D. Widespread intersex differentiation across the stickleback genome – The signature of sexually antagonistic selection? *Mol Ecol*. 2020;29:262–271. <https://doi.org/10.1111/mec.15255>