

UC Davis

UC Davis Previously Published Works

Title

Effects of the neo-X chromosome on genomic signatures of hybridization in *Rumex hastatulus*

Permalink

<https://escholarship.org/uc/item/1rr3t7t1>

Journal

Molecular Ecology, 31(13)

ISSN

0962-1083

Authors

Beaudry, Felix EG

Rifkin, Joanna L

Peake, Amanda L

et al.

Publication Date

2022-07-01

DOI

10.1111/mec.16496

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed

Effects of the neo-X chromosome on genomic signatures of hybridization in *Rumex hastatulus*

Felix E. G. Beaudry  | Joanna L. Rifkin  | Amanda L. Peake | Deanna Kim |
Madeline Jarvis-Cross | Spencer C. H. Barrett  | Stephen I. Wright 

Department of Ecology and Evolutionary Biology, The University of Toronto, Toronto, Ontario, Canada

Correspondence

Felix E. G. Beaudry, The University of Toronto, Department of Ecology and Evolutionary Biology, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada.
Email: felix.beaudry@mail.utoronto.ca

Funding information

Discovery Grants to SIW and SCHB from the Natural Sciences and Engineering Research Council of Canada (NSERC); Canada Research Chair to SIW from the Natural Sciences and Engineering Research Council of Canada (NSERC)

Handling Editor: Abbott Richard

Abstract

Natural hybrid zones provide opportunities for studies of the evolution of reproductive isolation in wild populations. Although recent investigations have found that the formation of neo-sex chromosomes is associated with reproductive isolation, the mechanisms remain unclear in most cases. Here, we assess the contemporary structure of gene flow in the contact zone between largely allopatric cytotypes of the dioecious plant *Rumex hastatulus*, a species with evidence of sex chromosome turn-over. Males to the west of the Mississippi river, USA, have an X and a single Y chromosome, whereas populations to the east of the river have undergone a chromosomal rearrangement giving rise to a larger X and two Y chromosomes. Using reduced-representation sequencing, we provide evidence that hybrids form readily and survive multiple backcross generations in the field, demonstrating the potential for ongoing gene flow between the cytotypes. Cline analysis of each chromosome separately captured no signals of difference in cline shape between chromosomes. However, principal component regression revealed a significant increase in the contribution of individual SNPs to inter-cytotype differentiation on the neo-X chromosome, but no correlation with recombination rate. Cline analysis revealed that the only SNPs with significantly steeper clines than the genome average were located on the neo-X. Our data are consistent with a role for neo-sex chromosomes in reproductive isolation between *R. hastatulus* cytotypes. Our investigation highlights the importance of studying plant hybrid zones for understanding the evolution of sex chromosomes.

KEYWORDS

genomics, hybridization, neo-sex chromosomes, plants

1 | INTRODUCTION

Where species meet and breed successfully in the wild, hybrid zones can form. In these zones of admixture, segregation and recombination shuffle the genomes of closely related species (Barton and Hewitt, 1985; Gompert et al., 2017; Harrison, 1993). Allele frequencies often display a geographic pattern of greater admixture near the center of the hybrid zone, resulting in allele frequency clines (Barton, 1979; Hewitt, 1989; Rieseberg and Wendel, 1993). Under continuous gene flow, hybrid genomes are expected to

contain a mixture of alleles from both species at all sites (Cruzan and Hendrickson, 2020; Harrison, 1993; Hewitt, 1989). However, allele frequency clines are often heterogeneous across the genome of hybrid individuals (Aeschbacher et al., 2017; Brandvain et al., 2014; Gibson and Moyle, 2019; Rieseberg et al., 1995; Rifkin et al., 2019). Genomic regions resistant to the shuffling of alleles can maintain reproductive incompatibilities between species, and loci at which the allele frequency cline is steep can be an indication of linkage with such incompatibilities (Barton 1983; Barton and Bengtsson, 1986; Cruzan et al., 2021; Gompert et al., 2017). Hybrid zones can

therefore provide opportunities for insights into the genetic mechanisms governing reproductive isolation in natural populations.

Chromosomal rearrangements between closely related species or diverging populations, including translocations and inversions, often appear as islands of increased genetic divergence with steeper allele frequency change across hybrid zones compared with genome averages (Coughlan and Willis, 2019; Lee et al., 2017; Noor et al., 2001; Todesco et al., 2019). Rearrangements are thought to be involved in hybrid breakdown because of their impact on failures in pairing and recombination of homologous chromosomes at meiosis. In early hybrid generations, failures in pairing can lead to hybrid offspring with missing or extra chromosomes, with strongly deleterious effects on fitness (Federley, 1913; Torgasheva and Borodina, 2016). In addition, failures in pairing of homologous chromosomes may play a role in restricting recombination locally in the genome without causing hybrid failure (Rieseberg, 2001). Variation in recombination rate is generally thought to be an important driver of speciation and local adaptation because it can both keep alleles involved in adaptation together (Todesco et al., 2019), or contribute to more efficient selection against deleterious incompatibilities (Schumer et al., 2018).

Sex chromosomes are disproportionately associated with genomic divergence and reproductive isolation (Coyne and Orr, 1989; Haldane, 1922; Presgraves, 2018), and the X chromosome is frequently found to show steeper clines across hybrid zones than other genomic regions (Gerald et al., 2006; Hooper et al., 2018; Moran et al., 2018; Tucker et al., 1992). These effects also appear in ZW sex

chromosome systems (such as for birds and butterflies), but here we focus on XY systems. Sex chromosomes may contribute to species incompatibilities when recessive X alleles are unmasked by Y chromosome degeneration, or if genes with sex-specific fitness effects tend to evolve incompatibilities faster than other genes and disproportionately accumulate on the X and Y (Lasne et al., 2017; Wu, 2001). Growing empirical evidence supports an association between sex chromosome rearrangements, such as the formation of neo-sex chromosomes, and the speciation process in animals (Benatti et al., 2010; Bracewell et al., 2017; Franchini et al., 2018; Kitano et al., 2009; Smith et al., 2016; Wang et al., 2022) and plants (Beaudry et al., 2019). However, the mechanisms underlying their association with reproductive isolation remain unclear, especially in plants. Studies of hybrids between species that differ in sex chromosome rearrangement are therefore of particular interest for disentangling the intrinsic factors governing the genetics of reproductive isolation.

In the dioecious plant *Rumex hastatulus* (Polygonaceae), males to the west of the Mississippi river USA have an X and a single Y chromosome (hereafter XX/XY cytotype), whereas males from populations to the east of the river have a larger X and two Y chromosomes (hereafter XX/XY₁Y₂ cytotype). The second Y in the XX/XY₁Y₂ cytotype arose from a chromosomal translocation between the ancestral X and an autosome (Autosome 3), leaving the autosomal homologue of the translocated X to segregate as a second Y chromosome resulting in XYY males (Kasjaniuk et al., 2018; Rifkin et al., 2021; Smith, 1963; Figure 1a). Hereafter, we refer to the newly sex-linked

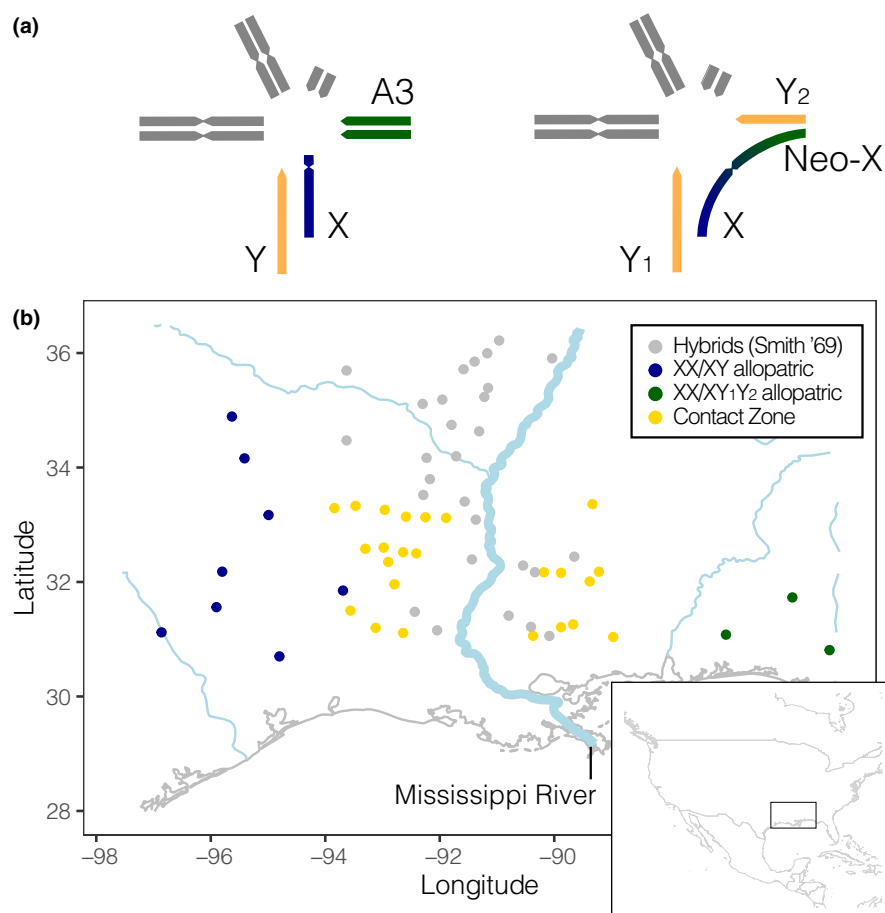


FIGURE 1 *Rumex hastatulus* across the hybrid zone (a) male idiogram for the two *R. hastatulus* cytotypes: The XX/XY karyotype on the left and XX/XY₁Y₂ karyotype on the right. (b) Distribution of sampling sites across the Mississippi river: New samples ("contact zone") in yellow, subset of allopatric XX/XY (blue) and XX/XY₁Y₂ (green) samples previously sequenced in Beaudry et al. (2019). Samples identified as hybrids in Smith (1969) in grey, which were not sequenced. Coastline as grey line, rivers as blue lines; Mississippi river is bolded and labelled. Inset: Map of North America with black box showing region sampled

region as the "neo-X" and the chromosome that incorporates both the ancestral- and neo-X as the "fused-X". The role of neo-sex chromosomes in reproductive isolation and the consequences for the hybridization process in the contact zone of *R. hastatulus* cytotypes have remained unresolved.

The first evidence that the different geographic cytotypes of *R. hastatulus* are in the early stages of developing reproductive isolation came from experimental crosses between individuals of different cytotypes. This research revealed that F_1 hybrid pollen had reduced viability, likely associated with uneven segregation of the sex chromosomes in hybrids (Kasjaniuk et al., 2018). Additional evidence that the neo-sex chromosome plays a key role in reproductive isolation came from demographic inference from allopatric populations in which a temporal association between the sex chromosome rearrangement and a loss of gene flow between the two cytotypes was observed (Beaudry et al., 2019). However, earlier cytotype analysis reported a substantial zone of hybridization around the Mississippi river region (Jackson, 1967; Smith, 1969). Whether gene flow in the zone of contact is ongoing past the F_1 generation remained unclear thus motivating our study.

Genomic studies in contact zones can provide insight into the prevalence of hybrids, the viability of hybrid offspring, and the degree of heterogeneity in gene flow across the genome (Abbott, 2017; Barton and Hewitt, 1985; Gompert et al., 2017). Although investigations of this type can provide critical insights into the process of speciation, hybrid zones of plant species with sex chromosomes are generally under-investigated at the genomic level (Pickup et al., 2019). Here, we assessed the contemporary structure of gene flow between the cytotypes of *R. hastatulus* by analysing individuals in the contact zone originally identified by Smith (1969). Although previous work in allopatric populations of the cytotypes focused on demographic inference from population genetic summary statistics (Beaudry et al., 2019), here we directly investigated the propensity of individuals to hybridize and the outcomes of such hybridization in contact zone populations. The primary goal of our study was therefore to establish the genomic structure of gene flow across the *R. hastatulus* cytotype contact zone and in so doing address the following questions: (1) Is gene flow occurring and do hybrids form? (2) If so, what is the location and breadth of the hybrid zone? (3) What are the relative roles of recombination, chromosomal rearrangements, and sex-linkage on the extent of gene flow in the hybrid zone? During our analyses, we identified evidence of additional chromosome structural rearrangements contributing to population differentiation within the XX/XY-cytotype, which may indicate a broader role for rearrangements in population differentiation of *R. hastatulus*.

2 | MATERIALS AND METHODS

2.1 | Population-level sampling

As cytotype hybrids are not reliably distinguished by observable phenotypic traits (Jackson, 1967; Smith, 1969), we used genetic data to quantify the extent of hybridization in the zone of contact between

the cytotypes by collecting samples from the hybrid zone. In June 2018, we collected seed from a minimum of 20 open-pollinated female plants in each of 30 populations approximately equally distributed among three transects positioned across both sides of the Mississippi river in the US states of Mississippi, Louisiana, and Arkansas (yellow circles in Figures 1b and S1).

For subsequent genomic analysis, we sampled 20 seeds from each of 20 maternal families among the 30 populations (12,000 seeds from 600 maternal families in total; see Figure S2 for experimental design). Seeds were bleached, rinsed, chilled at 4°C, and germinated on moist filter paper in growth chambers for 2 weeks. We transferred germinated seeds to 10 cm pots following earlier protocols and these were grown until they began flowering in the glasshouse (see Pickup and Barrett, 2013 for details). One individual from each maternal family was chosen for DNA extraction, to a maximum of 10 individuals from different families from each population. As not all individuals survived to first flower, this resulted in 143 leaf tissue samples among 24 populations (see Table S1 for population-level survival information). In six populations, all individuals died before flowering (Figure S1). For individuals that survived to flowering, we sexed samples using floral phenotype (presence of anthers vs stigma) and collected tissue from young leaves for DNA extraction.

2.2 | DNA extractions and sequencing

We extracted DNA using a DNeasy Plant Mini Kit (QIAGEN GmbH) following the manufacturer's protocol. The Elshire group in Palmerston, NZ performed library preparation of the DNA samples using the genotype-by-sequencing (GBS) method (Elshire et al., 2011); briefly, reads were fragmented using the restriction enzyme PstI and ligated to barcodes. Libraries were sequenced over 1 lane of 2 × 150 bp Illumina HiSeq X at the Australian Genome Research Facility (AGRF) in Westmead, Australia. Sequenced libraries were demultiplexed with axe-demux (Murray and Borevitz, 2018). Barcodes and adapters were removed using trimalore v.0.4.3 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) according to default parameters.

2.3 | SNP calling

We combined our sequencing data with 94 GBS libraries from 21 allopatric populations (Figures 1b and S1) previously sequenced in Beaudry et al. (2019). Hereafter, the samples sequenced in this study are referred to as "contact zone" whereas those from Beaudry et al. (2019) are referred to "allopatric". We aligned sequencing reads from all 237 samples (94 allopatric and 143 contact zone individuals) from 45 populations (21 allopatric and 24 contact zone populations) in our dataset to the repeat-masked *R. hastatulus* XX/XY-cytotype genome assembly; see Rifkin et al. (2020) for genome assembly and repeat-masking methods. We performed alignments using default parameters for paired

reads in NextGenMap v.0.5.5 (Sedlazeck et al., 2013). Bam files were sorted and read groups added with Picard v.2.23.8 (<http://broadinstitute.github.io/picard>). After filtering individuals based on a 5x minimum mean genome-wide read coverage cutoff, we retained 140 contact zone samples and all 94 allopatric samples (234 samples total). Variant sites were called using the commands *mpileup* followed by *call* in bcftools v. 1.9-67-g626e46b (Li, 2011) using default bcftools *call* parameters, printing the GQ field (*-f gq*) and outputting a zipped tabix index (*-m0 z*), but applying no filtering at this stage. We filtered the resulting .vcf using vcftools v. 0.1.15 (Danecek et al., 2011) for biallelic SNPs with sequencing information for more than 60% of individuals, a minimum PHRED quality score of 10, an all-sample mean site depth of greater than 5 reads, a minimum genotype PHRED quality score of 5 and a minor allele frequency (MAF) of greater than 1% (i.e., alleles present in only two copies or less were removed). We also removed all SNPs where all individuals were heterozygous, or where all individuals from any one population were missing data. This filtering yielded 13,578 SNPs.

2.4 | Removing X-Y fixed differences

Because of the divergence of the Y chromosome from the X chromosome and of the neo-Y from the neo-X (see Hough et al., 2014), some loci identified as SNPs may be fixed differences between X and Y (or neo-X and neo-Y). This would be misleading for inferences of hybridization because these loci inherently show high heterozygosity in males, as well as strong frequency differences between the sexes. To remove X-Y fixed differences, we used differences in heterozygosity between males and females by calculating expected heterozygosity ($2pq$) in all males or all females, respectively, in each of four categories (from west to east): allopatric XX/XY, western contact zone, eastern contact zone and allopatric XX/XY₁Y₂. We then subtracted male heterozygosity from female heterozygosity to estimate excess female heterozygosity on autosomes, the X and neo-X chromosomes (Figure S3). Taking the set of SNPs with excess male heterozygosity >0.25 (–0.25 in Figure S3) as potentially capturing Y-specific polymorphism and fixations, we removed 263 SNPs yielding a total of 13,315 SNPs for our primary analysis. The one part of our study where these SNPs were not removed was the analysis of clines SNP-by-SNP (using *bgc*, see below) in order to avoid excluding highly differentiated X-linked SNPs.

2.5 | Individual-level genotype-based hybrid assessment

We first performed a Principal Component Analysis (PCA) on individual genotypes (coded as [0,1,2], using hard genotype calls, ignoring uncertainty in genotype) using FactoMineR v2.4 (Husson et al., 2010); missing genotype values were imputed by the mean

genotype by FactoMineR and genotypes were not standardized (not given equal variances). We next used the probabilistic model analysis software STRUCTURE v. 2.3.4 (Pritchard et al., 2000) to cluster populations using multilocus genotype data to infer population structure, with clusters (K) ranging from 2 to 5, and ran six MCMC chains for a 1000 generation burn-in followed by 10,000 generations of analysis (11,000 iterations total) for each value of K . We aligned STRUCTURE runs by clusters using permutations in CLUMPP v.1.1.2 (Jakobsson and Rosenberg, 2007).

We examined the relations between heterozygosity and ancestry, using graphical approaches derived from the de Finetti diagram (Cannings and Edwards, 1968), to infer the generation at which hybrids may be inviable or infertile and, thus, to assess the genetic characteristics of hybrid incompatibilities (see Turelli and Orr, 2000). This approach required predefined hybrid and parental populations: parental populations were chosen as the east and west allopatric populations from Beaudry et al. (2019) respectively, as best fit demographic model suggested no gene flow between them in that study. We jointly inferred Hybrid Index (H.I.) and heterozygosity using Hiest v. 2.0 (Fitzpatrick, 2012). For a set of SNPs for which alleles were diagnostic to each population, first-generation hybrids are expected to show peak heterozygosity and a 50/50 ancestry mix, whereas individuals from populations without gene flow are expected to show both low heterozygosity and exclusive ancestry. We began our analysis using all SNPs, but these results did not conclusively distinguish between the cytotypes, likely due to the predominance of shared SNPs between cytotypes and low-frequency unique SNPs (Beaudry et al., 2019). To increase our resolution to detect signatures of hybridization, we repeated the analysis using only the 100 SNPs with the largest contributions to PC1.

2.6 | Population-level allele-frequency-based hybrid assessment

Using the same method as outlined in the preceding section, but on population-level allele frequencies (where population is defined by sampling site), we performed PCA and assessed population-level clustering using CONSTRUCT v1.0.4 (Bradburd et al., 2018) in R. We ran both spatial and non-spatial runs of CONSTRUCT for K -values between 1–6 and visualized CONSTRUCT model comparisons using the *calculate.layer.contribution* function.

To investigate rates of effective migration across the landscape in two-dimensions, we also evaluated the Estimated Effective Migration Surface v.0.0.0.9 (EEMS; Petkova et al., 2016). Here, as above, we defined population by sampling site. We converted the .bed file to .diffs using *bed2diffs_v2* in EEMS; v.2 includes a likelihood estimate for genotypes to account for missing genotypes. We ran six MCMC chains for 2,000,000 iterations with identical parameters among chains: a burn-in length of 1,000,000 iterations, a thinning interval of 9999 iterations, and a model with 800 demes.

For both individual and population-level PCA, we calculated the correlation between PC1 and longitude in R's *corr* function. We

similarly calculated the correlation between PC1 and Hybrid Index for individual and population-level using R's *corr* function.

2.7 | Cline fitting

As canonical cline-based analyses depend on diagnostic fixed differences between groups, the absence of fixed differences in our genetic markers made the classic approach to cline analysis intractable. Instead, by assuming Hybrid Index is continuous across the hybrid zone, we fitted allele frequency clines to Hybrid Index values (see Gompert et al., 2017; Hooper et al., 2018). Because CONSTRUCT incorporates uncertainty of parental genotypes, it is a flexible tool to estimate Hybrid Index (Bradburd et al., 2018; Pritchard et al., 2000). For our Hybrid Index estimate, we used $K = 2$ and a non-spatial model. To increase our sampling density of the putative hybrid zone, we combined data from all three transects. We confirmed our results with cline fitting to normalized population-level (allele-frequency) PC1 (i.e., using the same data as in CONSTRUCT) to ensure that our results were not influenced by over-fitting (see Wang et al., 2019).

Using HZAR (Derryberry et al., 2014), we fitted six cline models (plus a null model) for p , our standardized values of PC1 or H.I. estimated by (non-spatial) CONSTRUCT (see Table S2 for model summaries). Our null model (model 0) was that longitude explained no change in allele frequency, i.e., p is best explained as a horizontal line across the hybrid zone. We fitted two sets of models that allowed p to increase with longitude, the first of which had p starting at 0 and increasing to 1 (model 1–3) and a second that fitted a minimum ($p_{\min}$) and maximum ($p_{\max}$) value to p , according to the data (model 4–6). Models 1 and 4 followed a logit function and were defined by their centre and width. In models 2–3 and 5–6, change in p across longitude can depart from a logit function according to the parameters δ , the distance from the cline center where exponential decay in p starts, and τ , the slope of linear decay in p . For our models including δ and τ , we fitted a model where both δ and τ take the same value at each end of the hybrid zone (mirrored; models 2 and 5), whereas a second model set allowed for separate values of δ and τ on each side of the cline (models 3 and 6). Because we fitted models to both the PC1 and H.I. values, this resulted in a total of 14 models. We used the convention of a two-point AIC difference for significance (Gay et al., 2008; Raufaste et al., 2005) and then repeated the above analyses on each chromosome, splitting each chromosome into a region of low (<0.1 cM/mB) or high (≥ 0.1 cM/mB) recombination, separately (see below for methods for recombination rate estimation).

2.8 | Genomic heterogeneity in hybridization

First, we calculated the loading of each SNP to PC1 from PCAMiner. We then evaluated the effect of recombination density [recombination rate (cM) per megabase (mB)] on differentiation between the

cytotypes by estimating the recombination intensity (cM/mB) in 1mB windows along the genome using the anchored linkage map from Rifkin et al. (2020). To estimate recombination rate, we fitted polynomials onto the anchored linkage map, using the method and adapted code from Corbett-Detig et al. (2015).

We estimated heterogeneity in clines SNP-by-SNP using Bayesian genomic cline *bgc* (Gompert and Buerkle, 2011). We converted SNP files from vcf to *bgc* format using the *vcf2bgc.pl* script in *bgc* v.1.0. This was then run for 50,000 MCMC steps ($-x$), discarding the first 10,000 steps as burnin ($-n$) and thinned for every 5th step ($-t$) using the genotype uncertainty model ($-N$) and including precision parameters ($-p$) and quantiles ($-q$). We returned estimates from all MCMC runs using the *estpost* v 1.0 with credible interval set to default (i.e., 0.95) and then ran the linear models with *lm* in R for neo-X and PAR.

2.9 | XX/XY-cytotype population substructure

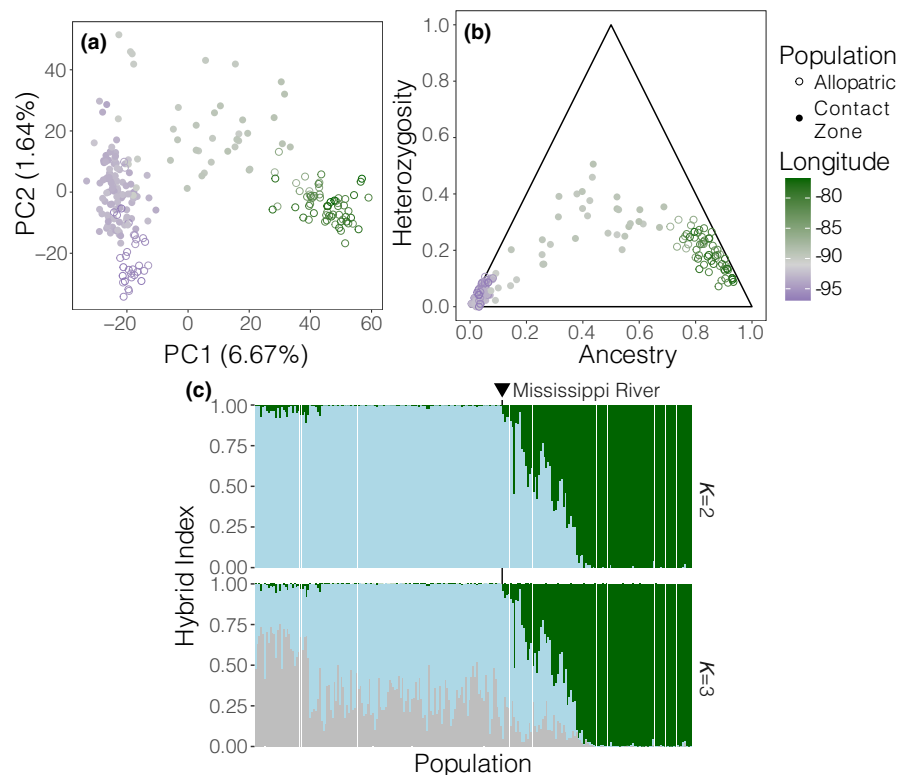
Both results from STRUCTURE and PCA suggested population substructure specific to the XX/XY cytotype. Therefore, to investigate this finding further, we replicated PCA only on individuals from these populations, using the methods from the *Individual-level genotype-based hybrid assessment* section only on samples from the west side of the Mississippi river. We estimated linkage disequilibrium using *plink* v1.9 (Purcell et al., 2007), with the flags $--r2$ *interchr* and $--ld-window-r2$ 0. Finally, we ran PCA for XX/XY-cytotype individuals for SNPs within each of two known inversions (Rifkin et al., 2020): for the A2 inversion, we selected SNPs between mB 5 and 50 (745 SNPs), and for the A3 inversion we selected SNPs between mB 145 and 170 (704 SNPs). To confirm clustering of genotypes, we ran a Bayesian clustering analysis assuming an underlying normal distribution along PC1 using MClust (Scrucca et al., 2016) in R. We chose the best fit number of clusters using Bayesian Information Criterion (BIC).

3 | RESULTS

3.1 | Genomic evidence for hybrid individuals

To identify the occurrence of hybrid individuals in the zone of contact between the *R. hastatus* cytotypes, we first performed genetic clustering analyses using 13,315 SNPs for 234 individuals from 45 populations (21 allopatric and 24 contact zone; Figures 1b and S1). We began by performing PCA, and found that PC1 grouped all individuals from allopatric populations into two separate clusters (empty circles in Figure 2a). PC1 was strongly correlated with longitude (Pearson Product-Moment Correlation: 0.924; $t = 36.83$, D.F. = 232, $p < 2.2e-16$). In contrast, individuals from a region of about 200 km on the eastern side of the Mississippi river were continuously distributed between the two cytotypes (full circles in Figure 2b). The occurrence of continuous variation in genotype-space in the

FIGURE 2 Inferring hybridization between cytotypes of *Rumex hastatulus*. (a) Principal component analysis. (b) Hybrid generation assessment for top 100 SNPs of PC1 by assessing estimated ancestry compared to heterozygosity. In both plots, individuals are coloured by the longitude at which they were sampled [purple longitude = 100° W, grey longitude = 90° W (Mississippi River), green longitude 75° W.] and shape represents whether individuals were collected from the contact zone or allopatric populations. (c) Individual-based STRUCTURE analysis, with populations of *R. hastatulus* sorted west (left) to east (right) for $K = 2$ and $K = 3$. White vertical lines are breaks between populations



geographic region between the allopatric populations is consistent with the hypothesis that hybrids occur in this range. Hereafter, we therefore refer to this geographic region as the hybrid zone. Most of the variance in PC2 was restricted to individuals from the hybrid zone (Figure 2a), consistent with simulation work showing hybrid populations tend to harbour unique genotypic variation (Gompert and Buerkle, 2016).

The results of our STRUCTURE analysis (assuming two clusters i.e., $K = 2$) separated individuals into geographic groups corresponding to the inferred division between the cytotypes (Figure 2c). Individuals from the hybrid zone exhibited Hybrid Index (H.I.) values between 0.04 and 0.75, whereas all individuals from allopatric populations had scores higher than 0.95 or lower than 0.05. With $K = 3$, the data suggest geographic substructure specific to the XX/XY cytotpe, discussed further below. Hybrid Index values were highly correlated with PC1 (Pearson Product-Moment Correlation: 0.989; $t = 101.26$, D.F. = 234, $p < 2.2 \times 10^{-16}$; Figure S4A).

We next jointly inferred Hybrid Index (ancestry) and heterozygosity to assess the number of generations for which hybridization has successfully proceeded. Using all SNPs, the two cytotypes were not conclusively distinguished (Figure S5). To increase our resolution to detect signatures of hybridization, we repeated the analysis using only the 100 SNPs with the largest contributions to PC1 (Figure 2b). In this second analysis, allopatric populations tended towards the corners of the triangle, whereas most individuals from populations in the hybrid zone were in the center of the triangle. Our results do not conclusively separate individuals into generations; however, the continuous patterns of ancestry and heterozygosity (Figure 2) are consistent with late-generation hybrids.

3.2 | Gene flow across the hybrid zone

In contrast to the individual-level genotype-based analyses of the previous section, we next conducted analyses on population-level allele frequencies. PCA of allele frequencies replicated the individual level analyses: PC1 was strongly correlated with longitude (Pearson Product-Moment Correlation: 0.925; $t = 15.98$, D.F. = 43, $p < 2.2 \times 10^{-16}$; Figure S6). Population-level clustering analysis (using CONSTRUCT; Bradburd et al., 2018) separated the cytotypes into separate clusters (Figure S7). Non-spatial CONSTRUCT model comparisons suggested three clusters ($K = 3$) partitioned by the largest portion of genotype variance (Figure S7A), whereas spatial runs suggested increased values of K do not explain more genotype variance. CONSTRUCT at $K = 3$ displayed a pattern concordant with the STRUCTURE results, splitting by cytotpe and suggesting population substructure specific to the XX/XY-cytotype in both spatial and non-spatial runs (Figure S7B,D). Regardless of K or whether the model was spatial or not, CONSTRUCT results confirmed that the majority of populations in the 200 km region east of the Mississippi river are composed of hybrid individuals. Calculated at the population level, Hybrid Index was correlated with PC1 using population allele frequency data (Pearson Product-Moment Correlation: 0.996; $t = 77.38$, D.F. = 43, $p < 2.2 \times 10^{-16}$; Figure S4B).

We next fitted allele frequency clines for seven models for normalized allele-frequency PC1 and the Hybrid Index estimated by CONSTRUCT assuming a non-spatial model and $K = 2$ (Figure 3; Table S2 for model summaries). Briefly, parameters were sequentially added to the cline equation, starting from a null model, to include two slope parameters (center and width; model 1), two decay

parameters (δ , the distance from the cline center where exponential decay in p starts, and τ , the slope of linear decay in allele frequency; model 2), independent decay parameters for each allopatric population (model 3) and a parameter for allele frequency in each allopatric population (rather than assume alleles were alternately fixed; model 4–6). For CONSTRUCT, model 2 (slope parameter and decay parameter; AICc = 24.03) was marginally more likely than 4 (slope parameter [center and width], and variable allele frequencies in allopatric populations; AICc = 24.20), but both models were significantly more supported than any other model (Table 1; purple in Figure 3). For normalized PC1, model 4 (AICc = 17.59) was marginally better supported than model 2 (AIC = 17.65) and both models were significantly more supported than other models (Table 1; red in Figure 3). Whereas model 2 suggests fixed differences in allopatry and a decay term, model 4 suggests no decay term but differences between allopatric populations are not fixed. Based on the 95% confidence intervals, clines fit to both CONSTRUCT and PCA were concordant for both cline center and width (Table 1; Figure S9); indeed, the shape of both clines were very similar (Figure 3).

Assuming each sampling site to be a population, the Estimated Effective Migration Surface (EEMS; Petkova et al., 2016) revealed significant signatures of reduced gene flow across the Mississippi river [95% posterior distribution $\log(m) < 0$; Figure S8]. We also found significantly higher than average rates of gene flow between populations on the west side of the river compared to the rest of the range [95% posterior distribution $\log(m) > 0$]. Although gene flow was prevalent between populations on the west side of the river (which were characterized as having low Hybrid Index values in the previous section), the EEMS analysis indicated that gene flow was

restricted across the river and between populations on the east side of the river, where we identified the hybrid zone. This result therefore suggests that gene flow is lower between populations in the hybrid zone compared to other geographic areas that we sampled.

3.3 | Genomic heterogeneity in hybridization

We next aimed to disentangle the relative roles of sex-linkage and recombination in hybridization between the XX/XY and XX/XY₁Y₂ cytotypes. First, we ran CONSTRUCT followed by cline fitting of Hybrid Index for each chromosome separately and, within each chromosome, we split the chromosome into low ($cM/mb < 0.1$) and high ($cM/mb \geq 0.1$) recombination regions (Table S3 and Figure S10). The estimated center of the cline was consistent across chromosomes, between -88.51° and -89.06° (east of the Mississippi river), and also consistent with the genome-wide average. Because all confidence intervals overlapped, we concluded that there was no significant difference between our parameter estimates for either cline width or cline center position.

We assessed the relative contribution of SNPs to genotype (individual-level) PC1. To this end, we calculated the average effect size (loading) of SNPs on PC1 in windows across the genome. To jointly evaluate the impact of recombination intensity, we fitted a linear model to calculate the effect of chromosome and recombination rate on PC1 effect size. We found the neo-X/A3 had a significantly increased contribution to PC1 effect size (Estimate = +0.0065, SD = 0.0015, t -value = 4.46, $p = 9.98e-06$), as did Autosome 4 (Estimate = +0.0043, SD = 0.0011, t -value = 4.04, $p = 6.07e-05$)

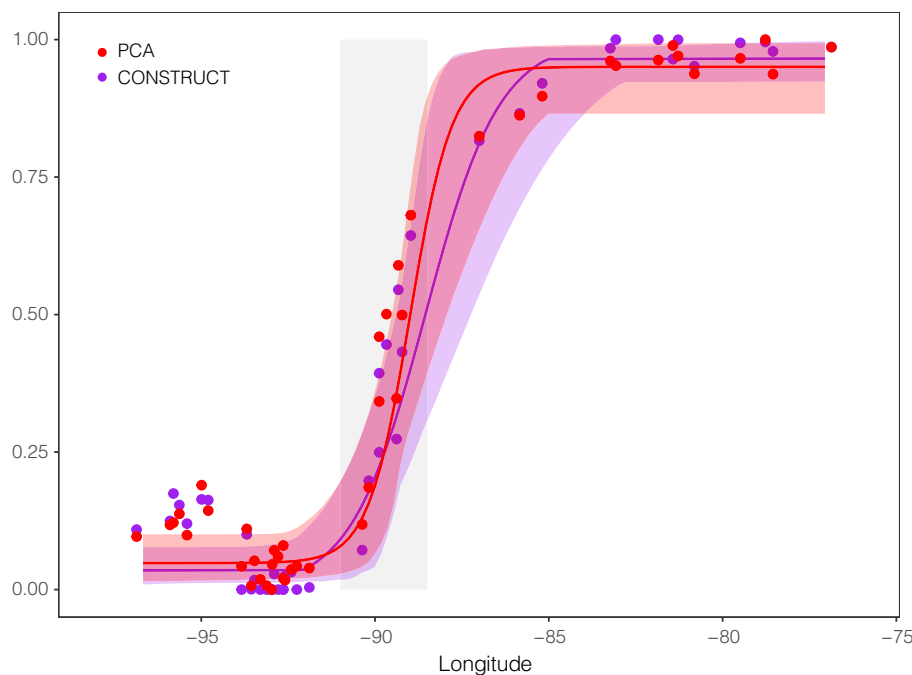


FIGURE 3 The rate of change in hybridization estimate across longitude for population samples of *Rumex hastatulus*. Points are sampling locations, solid line is the best fit estimated by HZAR and the shaded ribbon is the confidence interval for PC1 (blue) and CONSTRUCT (red). Grey shaded area is the approximate position of Mississippi river

TABLE 1 Results of cline fitting models with HZAR using both normalized PC1 and hybrid index (H.I.) estimated at the population-level by CONSTRUCT in *Rumex hastatulus*

Cline	Model	AICc	Parameter	Estimate	Confidence interval
CONSTRUCT H.I.	2	24.03	Center (Lon.)	-88.55°	[-89.30; -87.53]
			Width (km)	385.79	[122.45; 615.94]
			δ (km)	318.40	[84.37; 505.38]
			τ	0.0041	[7.41e-06; 0.34]
Norm. PC1	4	17.59	Center (Lon.)	-89.01°	[-89.47; -88.03]
			Width (km)	212.06	[93.80; 492.84]
			$p_{\min}$	0.05	[0.019; 0.095]
			$p_{\max}$	0.95	[0.88; 0.99]

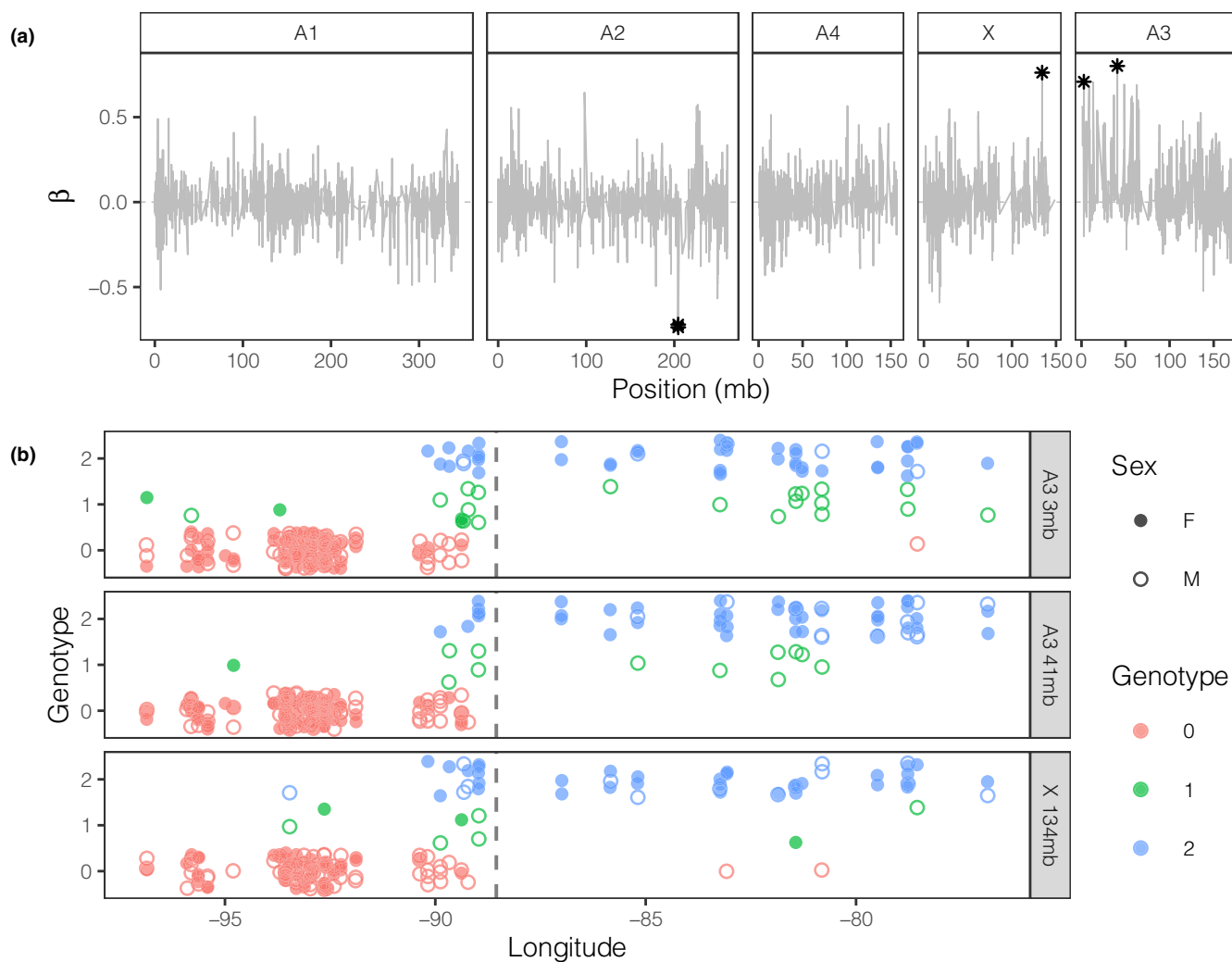


FIGURE 4 (a) β values across *Rumex hastatulus* chromosomes. Positive values indicate narrow clines (less introgression than expected) and negative values indicate shallow clines (more introgression than expected); starred loci mark outliers. (b) genotypes for three outlier positive β SNPs across longitude. Circles represent sex (filled female, empty male) and colour and y-axis are the same data for genotype. The dashed line is the best fit of the cline center

compared to other chromosomes. However, we found no significant impact of recombination rate on PC1 loading ($p = 0.45$).

We next estimated heterogeneity in clines SNP-by-SNP. First, we evaluated broader patterns of male–female heterozygosity across chromosomes: excess heterozygosity in males compared to females (<-0.3) likely represents polymorphism on the X or Y, whereas -0.5

should represent fixed X-Y differences (see *Removing X-Y fixed differences* in methods for details). Excess heterozygosity in females likely represents genes/sites missing from the Y (i.e., males are hemizygous at those loci). In three locations (XX/XY allopatric, western contact zone, and eastern contact zone), we observed more heterozygote excess on the X than on autosomes for both males and females, whereas

we found a paucity of female heterozygote excess in allopatric XX/XY₁Y₂. In both allopatric and western contact zone populations, neo-X and its autosomal homologue in the XX/XY-cytotype (Autosome 3) had identical patterns of heterozygosity as the autosomes, as would be expected because these sites are segregating autosomally in these locations. In the hybrid zone, females exhibited an excess of heterozygosity compared to males on the neo-X/A3 and above the levels recorded for the autosomal loci. In contrast, male-biased heterozygosity was not substantially different from the autosomal patterns on the neo-X/A3 in the hybrid zone. In allopatric XX/XY₁Y₂, we observed the highest increase in male-biased heterozygosity on the neo-X/A3. Given this latter pattern, high male excess heterozygosity may reflect greater levels of neo-X differentiation between the cytotypes.

We assessed clines SNP-by-SNP with Bayesian genomic cline (*bgc*) analysis, including our set of SNPs with excess male heterozygosity, as these may be due to high cytotype divergence on the neo-X. Outliers of α were evenly distributed across the genome. We detected three SNPs with positive outlier cline steepness values (β ; steeper clines than genome median). These SNPs were all sex-linked, either on the X or neo-X (Figure 4a). Investigation of genotype frequencies in males and females supported that these SNPs are linked to the X or neo-X rather than the Y or neo-Y, given the pattern of ohomozygous genotype differences in females across the hybrid zone (Figure 4b). In a linear model evaluating the impacts of sex-linkage and chromosome on β , sex-linked X was significantly higher than the genome average (Estimate = 0.0152, SD = 0.00521, t -value = 2.923, p = .00347), the PAR of the neo-X was significantly higher (Estimate = 0.0148, SD = 0.00247, t -value = 6.017, p = 1.84e-09) and the sex-linked portion of the neo-X had the largest effect (Estimate = 0.0463, SD = 0.00498, t -value = 9.299, p < 2e-16; Figure 4a). We identified a pair of SNPs 10bp apart with negative outlier β values (cline shallowness) on autosome 2; at this distance, these alleles are likely to be linked and to represent the same haplotype.

3.4 | XX/XY cytotype substructure

Our STRUCTURE analysis suggested substructure in the XX/XY cytotype and therefore we performed PCA on genotypes specific to samples from this cytotype. Because the hybrid zone is located exclusively on the east side of the Mississippi, we included only individuals from the west side in this analysis. XX/XY-exclusive PC1 correlates with PC1 from the all-samples PCA (Pearson's product-moment correlation = 0.57, t = 7.96, D.F. = 131, p = 7.24e-13), indicating we are identifying the same pattern in both PCAs. Also, it is evident that the XX/XY-exclusive PC1 correlates with longitude (0.70, t = 11.34, D.F. = 131, p < 2.2e-16). To locate the genomic source of XX/XY substructure, we investigated the effect size of SNPs on PC1 of the XX/XY-exclusive PCA. We found a sizeable region with many SNPs of large effect on the left end of autosome 2 (Figure 5a). The region on the left is noteworthy for containing two nested inversions, based on comparative linkage mapping from mapping populations of the two cytotypes (Figure 5b, data from Rifkin

et al., 2020). Furthermore, based on our SNP set, we found high levels of linkage disequilibrium around the region of the inversion on A2 (Figure 5c). A similar pattern appears for lower principal components: SNPs that contribute most to PCs 2–4 also align with regions where LD is disproportionately high.

Rifkin et al. (2020) identified inversions between the cytotypes in regions overlapping the largest peaks; the noteworthy exception was the increase in PC loadings in the pericentromeric region of Autosome 1 (A1).

Finally, we ran PCA only using SNPs within the rearrangements. For both PCAs, Gaussian mixture modelling suggested the underlying distribution was best modelled as three clusters. For the A2 rearrangement, the three clusters are entirely discriminant and for the A3 rearrangement, one cluster is discriminant from the other two (Figure S11); these three clusters along PC1, with the central cluster showing no variance (Dim. 1 = 0), may represent three possible genotypes at the rearrangement (two homozygous and one heterozygous genotype). Although we have no evidence that the inversions that appear in our comparative maps are segregating in the XX/XY-cytotype, the signal of population substructure specific to the XX/XY cytotype aligns well with these inversions and the regions of high linkage disequilibrium.

4 | DISCUSSION

Genomic analysis of the *Rumex hastatulus* hybrid zone offers rare insights into the evolution of reproductive isolation involving neo-sex chromosomes. Because inter-cytotype hybrids show reduced fertility from controlled experimental crosses (Kasjaniuk et al., 2018) and previous work inferred an absence of gene flow between allopatric populations of each cytotype (Beaudry et al., 2019), we might have expected to find little or no hybridization in the zone of contact. Instead, consistent with cytogenetic studies by Smith (1969), we found evidence of extensive hybridization between the two *R. hastatulus* cytotypes in the zone of contact. We detected geographic variation in the amount of hybridization and in the extent of gene flow across the genome. These findings allowed us to analyse both the structure of the hybrid zone and the role of neo-sex chromosomes in reproductive isolation, as we discuss below.

4.1 | Structure of the hybrid zone

The geographic region in which we identified hybrids is roughly consistent with cytological evidence from the 1960s (Smith 1969). In contrast with Smith's findings, we observed almost total inviability of individuals from populations closest to the west side of the Mississippi river. In the glasshouse, none of the plants that originated from MLRA-O (Mississippi Delta Cotton and Feed Grains Region) survived. Also, *R. hastatulus* individuals were absent from herbarium collections from the MLRA-O region (SERNEC Data Portal 2022; see Supplemental Methods). The major characteristic of the

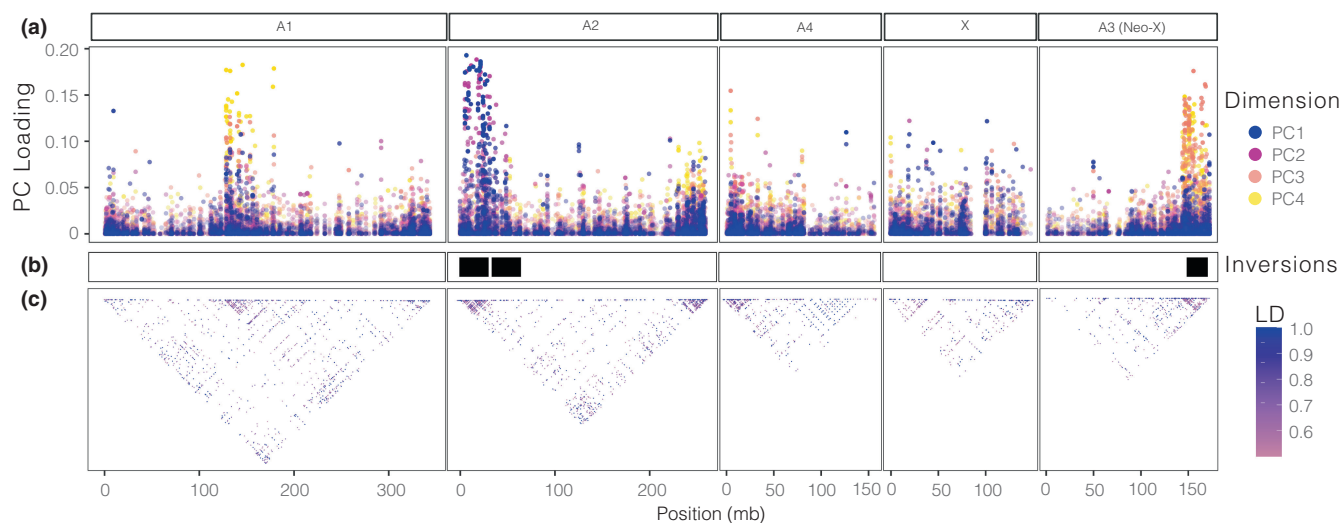


FIGURE 5 Genomic localization of SNPs participating in XX/XY-cytotype population substructure in *Rumex hastatulus*. (a) Distribution of SNP effect sizes for XX/XY-cytotype PC dimensions 1–4 (colour) along the chromosomes. (b) Inversions between cytotypes shown as black blocks (data from Rifkin et al. 2020). (c) Linkage disequilibrium greater than 0.5 (colour) within XX/XY

MLRA-O area is a high-water table and extensive flooding (NRCS, 2006), ecological conditions that are generally not suitable for the growth of *R. hastatulus* populations, which prefer drier soils (Conn and Blum, 1981). Smith (1969) found signals of hybridization in this region whereas we only identified hybrids on the east side of the river. This leaves open the question as to whether individuals from populations in the MLRA-O region do not survive in the glasshouse because of intrinsic hybrid inviability and/or because of the growing conditions in the glasshouse. Although the question of whether hybrid individuals persist on the west side of the Mississippi river remains unclear, our study indicates that there is a strong signal of hybridization on the east side of the Mississippi river.

Patterns of hybridization are often used to infer the genetics of reproductive incompatibilities (Turelli and Moyle, 2007; Turelli and Orr, 2000). Indeed, the generation (F_1 , F_2 , etc) at which hybrids are inviable or infertile can be harnessed to understand the role of dominance in hybrid breakdown (Turelli and Orr, 2000). Given the broad range of ancestry proportions and heterozygosity that we detected in our analyses, our results suggest most hybrid individuals appear to be late generation hybrids. Our findings showing strong evidence of a hybrid zone suggest that introgression is unlikely to be stalled at the F_1 generation (e.g., due to significant lethality or sterility), thus rejecting earlier inferences on a complete loss of gene flow between the cytotypes based on allopatric populations (see Beaudry et al., 2019). The pattern could be, in part, ecological, including for example influences of wind on pollen and seed dispersal, as seen in a recent meta-analysis of genetic diversity in trees (Kling and Ackerly, 2021). However, previous experimental work suggests a role for genetics, as inter-cytotype crosses also suggest correspondingly asymmetric barriers to gene flow, which have parent-of-origin effects on hybrid male genotypes (Kasjaniuk et al., 2018). Overall, the characteristics of hybrids in the hybrid zone is suggestive of a heterogeneous pool of hybrids, but is not directly informative about the genetic architecture of reproductive isolation.

Principal component regression suggested that loci across the genome show differentiation between the cytotypes and a cline in ancestry. This finding supports our underlying assumption that we are observing hybridization rather than a more complex demographic scenario, such as an eastern expansion of *R. hastatulus* associated with the spread of the neo-X. If the on-going fixation of the fusion event explained the patterns, we would not expect to see signs of mixed ancestry elsewhere in the genome in the hybrid zone.

When fitting clines to whole-genome estimates of Hybrid Index, we found that different estimates of hybridization returned different cline fits (Table 1). As our clinal analysis was performed on a composite summary of many loci, our interpretation cannot rely on theory developed for selection at single loci (e.g., Barton, 1979; Barton, 1983) but rather on phenotypic clines. The different best-fitting models likely reflect differences between the assumptions of CONSTRUCT versus those of PCA. If we assumed the discrete classification of individuals into K ancestral groups, cline fitting returned a model supporting less gene flow across the hybrid zone. In contrast, fitting PC1 suggested a continuous distribution of genotype space across the entire range, including some shared polymorphism between allopatric populations as reflected in non-zero values of p . Models 2 and 4 were not significantly different in both hybridization estimates, reflecting the challenges of distinguishing low levels of gene flow from ancestral polymorphism (Kern and Hey, 2017). However, the Estimated Effective Migration Surface (EEMS) makes few assumptions about discrete populations as the number of demes increases, and indicated reduced gene flow in the hybrid zone, suggesting that gene flow is indeed reduced. Together, these results paint a picture broadly consistent with ongoing but limited hybridization with barriers to gene flow.

Empirical studies of dioecious plant species indicate that gene flow is often low in hybrid zones (Pickup et al., 2019). In hybrid zones in Switzerland of *Silene latifolia* and *S. dioica*, differentiation genome-wide was low with late generation hybrids common but early generation hybrids mostly absent, suggesting historical introgression without

extensive recent hybridization (Minder et al., 2007). The difference between the *Silene* and *Rumex* results may arise because *S. latifolia* and *S. dioica* are much more divergent in reproductive biology, phenology and habitat preferences (Baker, 1948) than the *R. hastatulus* cytotypes we investigated here (Jackson, 1967; Puixeu et al., 2019). Minder et al. (2007) did not investigate the impact of the sex chromosomes on reproductive isolation of *S. latifolia* and *S. dioica*, as the study was based on unanchored AFLPs. Given the rarity of analyses of hybrid zones in dioecious plants with sex chromosomes, it is clearly too early to make general conclusions on the nature of hybrid zones and their genomic consequences. Nonetheless, genomic heterogeneity in hybridization involving the sex chromosomes is common in animals, most often showing steeper clines at sex chromosomes than elsewhere in the genome (Hewitt, 1975; Hooper et al., 2018; Macholan et al., 2011). Plants generally show variation in gene flow across the genome (Abbott, 2017; Pickup et al., 2019) so it is possible that overall patterns in plants will be similar to those in animals.

4.2 | Neo-sex chromosomes and reproductive isolation

Given the genomic evidence for the presence of ongoing introgression between cytotypes in the hybrid zone of *R. hastatulus*, we were particularly interested in assessing whether the sex chromosome and/or neo-sex chromosome showed evidence for reduced introgression. In scanning for the location of SNPs with large effects on genotypic differentiation between the cytotypes, we found that SNPs on the neo-X chromosome had disproportionately large loadings on the first principal component (PC1). This pattern is suggestive of an enrichment for loci with high differentiation between the cytotypes on the neo-X, which could indicate a role for the neo-X in reproductive isolation. Furthermore, when fitting clines SNP-by-SNP with *bgc*, we found that the few SNPs with significantly steeper clines than the genome-wide average were on the X or neo-X chromosomes. In contrast, there was no signal of a disproportionate role of the neo-X in the genomic-regional clines. Fitting clines separately for each chromosome, split between regions of low and high recombination, the confidence intervals for cline widths mostly overlapped.

The discordance between results for SNP-by-SNP clines and the genomic-regional clines may be due to a lack of power to differentiate between genomic regions when fitting many cline parameters (six in our best fit models) to entire genomic regions; *bgc* avoids this need by significantly simplifying the model to two parameters (Gompert and Buerkle, 2011). Alternatively, a few outliers may be on the X and neo-X but their signal is overwhelmed by a chromosome-wide analysis. Indeed, the X chromosome of *R. hastatulus* is very large (~20% of the 1.3GB genome) and therefore likely exhibits significant variation in gene flow among sites.

Our results are consistent with a role for the neo-sex chromosome in the evolution of reproductive isolation but leave the specific mechanisms unresolved. The appearance of late generation hybrids in our data suggests that reduced fertility from aneuploidy is unlikely to be the sole driver of reproductive isolation. Instead, the rearrangements

could play a role in reproductive isolation by restricting recombination. Regions with high rates of recombination may be able to actively purge genetic incompatibilities, for example, as seen in swordfish hybrids (Schumer et al., 2018). However, while our data support an important role for neo-X linked SNPs in cytotype differentiation, we found no evidence that variation in recombination rate was involved.

Alternatively, theory predicts that X-linked genes are more likely to contribute to local adaptation because of their relatively lower movement between populations of X chromosomes compared to autosomes (Lasne et al., 2017). In a large-scale geographical study of population differentiation in *R. hastatulus* involving the two cytotypes grown under uniform conditions in the glasshouse, Puixeu et al. (2019) demonstrated significant differentiation in sexually dimorphic traits associated with abiotic climatic variables, such as mean annual temperature and precipitation, suggesting local adaptation could be partly sex-specific. Linkage of sex chromosomes with alleles having locally sex-specific benefits would be consistent with the absence of hybrid breakdown in the contact zone and the lack of evidence for gene flow between allopatric populations (Beaudry et al., 2019). Although alleles can move between the cytotypes, albeit at a reduced rate, local adaptation may limit the success of migrating alleles across a broader geographic range. Mapping of quantitative trait loci for sexually differentiated traits in crosses between the cytotypes would be a promising approach to further understand the impact of sex-linkage and sex-specific local adaptation. Evidence for reduced female heterozygosity specifically for neo-sex linked SNPs in the XX/XY₁Y₂ allopatric range and for fixed differences between females across the range on the neo-X outlier SNP (Figure S3) are consistent with a potential role for a selective sweep in the neo-X. Further disentangling the evolutionary forces that first led to the spread of the neo-X in populations in the eastern range of *R. hastatulus* may also provide valuable insights on the role of the neo-X in reproductive isolation.

Our cline-based estimates should be interpreted with caution, however, as our work revealed what appears to be a gap in theory: it remains unclear if differences in effective population size across the genome can impact cline estimates. Following an increase in studies looking for genetic differences between species by comparing differentiation statistics, a note of caution has been raised highlighting how differentiation between species can be artificially increased by low diversity within species. For example, heterogeneity in effective population size across the genome alone can cause apparent heterogeneity in differentiation between species (Charlesworth et al., 1997; Cruickshank and Hahn, 2014). Even though the use of effective population size is critical to cline theory (Barton, 1983), the impact of variation across the genome in this parameter seems to have not yet been transferred to cline analysis. This is a particularly crucial area of future study for the analysis of neo-sex chromosomes because the effective population size is asymmetric across the cline: under equal sex ratios, the X chromosome inherently has $\frac{3}{4}$ the effective population size of the autosomes (Charlesworth, 2001). It is therefore unclear whether it is a justifiable test to fit clines for loci which are autosomally-linked at one cline end but sex-linked at the other. Extensions of analytical theory or simulation-based evaluation will be required to further resolve this.

4.3 | XX/XY substructure and chromosomal rearrangements

Unexpectedly, we observed three regions with disproportionately large contributions to population substructure within the XX/XY cytotype. Comparison to previous comparative mapping suggested two of these regions aligned with structural chromosomal rearrangements between the cytotypes. Indeed, large rearrangements appear to contribute to differentiation among geographic regions in many plant species (Rieseberg, 2001). Often, such rearrangements are associated with differences in ecology and habitat differentiation, leading to the hypothesis that they play an important role in local adaptation (Connallon et al., 2018; Kirkpatrick and Barton, 2006; Todesco et al., 2019). Whether the rearrangements between the cytotypes of *R. hastatulus* are segregating within the XX/XY cytotype is unknown, but linkage disequilibrium data suggest lower recombination in these genomic regions and local PCA (within each rearrangement) suggested strong population substructuring.

We did not observe any signals that the regions containing these inversions play a role in affecting rates of gene flow or in limiting hybridization between the cytotypes. However, the differentiated XX/XY populations on the far west end of our sampling range seem to have some genotype frequencies which are more similar to XX/XY₁Y₂ cytotype populations than to the nearer XX/XY populations (Figures S6 and S7). Whether this represents biological similarities between populations at both ends of the range remains unclear.

The finding that several large rearrangements contribute to population differentiation within the XX/XY-cytotype helps frame future questions on the role of the neo-sex chromosomes in reproductive isolation: how similar are the population-level patterns on the neo-sex chromosomes to the autosomal rearrangements? Long-read whole genome resequencing should be used to investigate finer scale patterns of differentiation perhaps capturing segregating inversions. Investigating whether genetic variation within chromosomal regions containing structural variation is associated with ecological gradients across the geographical range of the XX/XY-cytotype range should provide insight into the potential role of chromosomal rearrangements in population differentiation of *R. hastatulus*.

In conclusion, our findings indicate that despite large scale chromosomal rearrangements between the cytotypes of *R. hastatulus*, hybrids can be viable and sufficiently fertile to participate in gene exchange over several generations in populations along a 200–300km cline. Our results are consistent with a low but significant level of gene flow across the hybrid zone. We found a disproportionate role of the neo-X chromosome in restricting gene flow. Together, our results highlight the importance of genomic studies of hybrid zones to identify the role of neo-sex chromosomes in reproductive isolation in wild plants.

AUTHOR CONTRIBUTIONS

SCHB, SIW, FEGB and JLR conceived and designed the study; FEGB and JLR performed the field sampling; FEGB, DK and MJ-C

prepared the samples for sequencing, FEGB, JLR, AP and SIW conducted the analyses; and all authors contributed to the writing of the manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank Yunchen Gong for server help, Bill Cole for glasshouse assistance, Baharul Choudhury for wet lab assistance and Elsie Shogren for comments on the manuscript. We thank Richard Abbott and three anonymous reviewers for their thoughtful feedback. This research was funded by Discovery Grants from the Natural Sciences and Engineering Research Council to SIW and SCHB. JLR was supported by an EEB post-doctoral fellowship and by NSERC grants to SIW and SCHB and FB was supported by NSERC grants to SIW.

CONFLICT OF INTEREST

The authors have no conflict of interests to declare.

DATA AVAILABILITY STATEMENT

All code developed and used in this script are available on github (github.com/felixbeaudry/rumexHybridZone). All new data (raw reads for 143 contact zone individuals) are available for download on NCBI's SRA (BioProject PRJNA806769).

ORCID

Felix E. G. Beaudry  <https://orcid.org/0000-0002-1384-382X>

Joanna L. Rifkin  <https://orcid.org/0000-0003-1980-5557>

Spencer C. H. Barrett  <https://orcid.org/0000-0002-7762-3455>

Stephen I. Wright  <https://orcid.org/0000-0001-9973-9697>

REFERENCES

- Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution*, 55, 238–258.
- Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences, USA*, 114, 7061–7066.
- Baker, H. G. (1948). Stages in invasion and replacement demonstrated by species of *Melandrium*. *Journal of Ecology*, 36, 96–119.
- Barton, N. H. (1979). Gene flow past a cline. *Heredity*, 43, 333–339.
- Barton, N. H. (1983). Multilocus clines. *Evolution*, 37, 454–471.
- Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, 57, 357–376.
- Barton, N., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Reviews in Ecology and Systematics*, 16, 113–148.
- Beaudry, F. E. G., Barrett, S. C. H., & Wright, S. I. (2019). Ancestral and neo-sex chromosomes contribute to population divergence in a dioecious plant. *Evolution*, 74, 256–269.
- Benatti, T. R., Valicente, F. H., Aggarwal, R., Zhao, C., Walling, J. G., Chen, M.-S., Cambron, S. E., Schemerhorn, B. J., & Stuart, J. J. (2010). A neo-sex chromosome that drives postzygotic sex determination in the hessian Fly (*Mayetiola destructor*). *Genetics*, 184, 769–777.
- Bracewell, R. R., Bentz, B. J., Sullivan, B. T., & Good, J. M. (2017). Rapid neo-sex chromosome evolution and incipient speciation in a major forest pest. *Nature Communications*, 8, e1593.

- Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, 210, 33–52.
- Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genetics*, 10, e1004410.
- Cannings, C., & Edwards, A. W. F. (1968). Natural selection and the de Finetti diagram. *Annals of Human Genetics*, 31, 421–428.
- Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetic Research*, 70, 155–174.
- Charlesworth, B. (2001). The effect of life-history and mode of inheritance on neutral genetic variability. *Genetic Research*, 77, 153–166.
- Conn, J. S., & Blum, U. (1981). Sex ratio of *Rumex hastatulus*: The effect of environmental factors and certation. *Evolution*, 35, 1108–1116.
- Connallon, T., Olito, C., Dutoit, L., Papoli, H., Ruzicka, F., & Yong, L. (2018). Local adaptation and the evolution of inversions on sex chromosomes and autosomes. *Philosophical Transactions of Royal Society B*, 373, 20170423.
- Corbett-Detig, R. B., Hartl, D. L., & Sackton, T. B. (2015). Natural selection constrains neutral diversity across a wide range of species. *PLoS Biology*, 13, e1002112.
- Coughlan, J. M., & Willis, J. H. (2019). Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. *Molecular Ecology*, 28, 1343–1357.
- Coyne, J. A., & Orr, H. A. (1989). Two rules of speciation. In D. Otte & A. Endler (Eds.), *Speciation and its consequences* (pp. 180–207). Sinauer Associates.
- Cruikshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23, 3133–3157.
- Cruzan, M. B., & Hendrickson, E. C. (2020). Landscape genetics of plants: challenges and opportunities. *Plants Communications*, 1, 100100.
- Cruzan, M. B., Thompson, P. G., Diaz, N. A., Hendrickson, E. C., Gerloff, K. R., Kline, K. A., Machiorlete, H. M., & Persinger, J. M. (2021). Weak coupling among barrier loci and waves of neutral and adaptive introgression across an expanding hybrid zone. *Evolution*, 75, 3098–3114.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.
- Southeast Regional Network of Expertise and Collections (SERNEC). (2022). Data Portal. <http://sernecportal.org/index.php>. Accessed on January 20 2022.
- Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, 14, 652–663.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6, e19379.
- Federley, H. (1913). Das verhalten der chromosome bei der spermatogenese der schmetterlinge *Pygaera anachoreta*, *curtula* und *pigra* sowie einiger ihrer bastarde. *Zeitschrift für inductive abstammungs und vererbungslehre*, 9, 1–110.
- Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, 12, e131.
- Franchini, P., Jones, J. C., Xiong, P., Kneitz, S., Gompert, Z., Warren, W. C., Walter, R. B., Meyer, A., & Scharl, M. (2018). Long-term experimental hybridisation results in the evolution of a new sex chromosome in swordtail fish. *Nature Communications*, 9, e5136.
- Gay, L., Crochet, P.-A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution*, 62, 2789–2806.
- Geraldes, A., Ferrand, N., & Nachman, M. W. (2006). Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, 173, 919–933.
- Gibson, M. J. S., & Moyle, L. C. (2019). Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species. *Molecular Ecology*, 29, 2204–2217.
- Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, 20, 2111–2127.
- Gompert, Z., & Buerkle, C. A. (2016). What, if anything, are hybrids: Enduring truths and challenges associated with population structure and gene flow. *Evolution Applications*, 9, 909–923.
- Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annual Reviews in Ecology and Evolution*, 48, 207–229.
- Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, 12, 101–109.
- Harrison, R. G. (1993). Hybrids and hybrid zones: Historical perspective. In R. G. Harrison (Ed.), *Hybrid zones and the evolutionary process* (pp. 3–12). Oxford University Press.
- Hewitt, G. M. (1975). A sex-chromosome hybrid zone in the grasshopper *Podisma pedestris* (orthoptera: Acrididae). *Heredity*, 35, 375–387.
- Hewitt, G. M. (1989). The subdivision of species by hybrid zones. In D. Otte & J. A. Endler (Eds.), *Speciation and its consequences* (pp. 85–110). Sinauer Associates.
- Hooper, D. M., Griffith, S. C., & Price, T. D. (2018). Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. *Molecular Ecology*, 28, 1246–1262.
- Hough, J., Hollister, J. D., Wang, W., Barrett, S. C. H., & Wright, S. I. (2014). Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proceedings of the National Academy of Sciences, USA*, 111, 7713–7718.
- Husson, F., Le, S., & Pagès, J. (2010). *Exploratory multivariate analysis by example using R*. CRC Press.
- Jackson, S.W. (1967). A biosystematic study of *Rumex hastatulus* [Unpublished doctoral dissertation]. North Carolina State University.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806.
- Kasjaniuk, M., Grabowska-Joachim, A., & Joachim, A. J. (2018). Testing the translocation hypothesis and Haldane's rule in *Rumex hastatulus*. *Protoplasma*, 256, 237–247.
- Kern, A. D., & Hey, J. (2017). Exact calculation of the joint allele frequency spectrum for isolation with migration models. *Genetics*, 207, 241–253.
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversion, local adaptation and speciation. *Genetics*, 173, 419–434.
- Kitano, J., Ross, J. A., Mori, S., Kume, M., Jones, F. C., Chan, Y. F., Absher, D. M., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., & Peichel, C. L. (2009). A role for a neo-sex chromosome in stickleback speciation. *Nature*, 461, 1079–1083.
- Kling, M. M., & Ackerly, D. D. (2021). Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. *Proceedings of the National Academy of Science, USA*, 118, e2017317118.
- Lasne, C., Sgrò, C. M., & Connallon, T. (2017). The relative contributions of the X chromosome and autosomes to local adaptation. *Genetics*, 205, 1285–1304.
- Lee, C.-R., Wang, B., Mojica, J. P., Mandáková, T., Prasad, K. V. S. K., Goicoechea, J. L., Perera, N., Hellsten, U., Hundley, H. N., Johnson, J., Grimwood, J., Barry, K., Fairclough, S., Jenkins, J. W., Yu, Y., Kudrna, D., Zhang, J., Talag, J., Golser, W., ... Mitchell-Olds, T. (2017). Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nature Ecology and Evolution*, 1, e119.

- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27, 2987–2993.
- Macholan, M., Baird, S. J. E., Dufkova, P., Munclinger, P., Bimova, B. V., & Pialek, J. (2011). Assessing multilocus introgression patterns: A case study on the mouse X chromosome in Central Europe. *Evolution*, 65, 1428–1446.
- Minder, A. M., Rothenbuehler, C., & Widmer, A. (2007). Genetic structure of hybrid zones between *Silene latifolia* and *Silene dioica* (Caryophyllaceae): Evidence for introgressive hybridization. *Molecular Ecology*, 16, 2504–2516.
- Moran, P. A., Pascoal, S., Cezard, T., Risse, J. E., Ritchie, M. G., & Bailey, N. W. (2018). Opposing patterns of intraspecific and interspecific differentiation in sex chromosomes and autosomes. *Molecular Ecology*, 27, 3905–3924.
- Murray, K. D., & Borevitz, J. O. (2018). Axe: Rapid, competitive sequence read demultiplexing using a trie. *Bioinformatics*, 34, 3924–3925.
- NRCS. (2006). Land resource regions and major land resource areas of the United States, the Caribbean, and the Pacific basin. In *United States Department of Agriculture Handbook* p. 296.
- Noor, M. A., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences, USA*, 98, 12084–12088.
- Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, 48, 94–100.
- Pickup, M., & Barrett, S. C. H. (2013). The influence of demography and local mating environment on sex ratios in a wind-pollinated dioecious plant. *Ecology and Evolution*, 3, 629–639.
- Pickup, M., Brandvain, Y., Fraise, C., Yakimowski, S. B., Barton, N. H., Dixit, T., Lexer, C., Cereghetti, E., & Field, D. L. (2019). Mating system variation in hybrid zones: Facilitation, barriers and asymmetries to gene flow. *New Phytologist*, 224, 1035–1047.
- Presgraves, D. C. (2018). Evaluating genomics signatures of 'the large X-effect' during complex speciation. *Molecular Ecology*, 27, 3822–3830.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Puixeu, G., Pickup, M., Field, D. L., & Barrett, S. C. H. (2019). Variation in sexual dimorphism in a wind-pollinated plant: The influence of geographical context and life-cycle dynamics. *New Phytologist*, 224, 1108–1120.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81, 559–575.
- Raufaste, N., Orth, A., Belkhir, K., Senet, D., Smadja, C., Baird, S. J. E., Bonhomme, F., Dod, B., & Boursot, P. (2005). Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal Linnean Society of London*, 84, 593–616.
- Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, 16, 351–358.
- Rieseberg, L. H., Linder, C. R., & Seiler, G. J. (1995). Chromosomal and generic barriers to introgression in *helianthus*. *Genetics*, 141, 1163–1171.
- Rieseberg, L. H., & Wendel, J. F. (1993). Introgression and the consequences in plants. In R. G. Harrison (Ed.), *Hybrid zones and the evolutionary process* (pp. 70–111). Oxford University Press.
- Rifkin, J. L., Beaudry, F. E. G., Humphries, Z., Choudhury, B. I., Barrett, S. C. H., & Wright, S. I. (2020). Widespread recombination suppression facilitates plant sex chromosome evolution. *Molecular Biology and Evolution*, 38, 1018–1030.
- Rifkin, J. L., Hnatovska, S., Meng, Y., Sacchi, B., Choudhury, B., Gong, Y., Rastas, P., Barrett, S. C. H., & Wright, S. (2021). Recombination landscape dimorphism contributes to sex chromosome evolution in the dioecious plant *Rumex hastatulus*. *Philosophical Transactions of the Royal Society B*, 377(1850), 20210226.
- Rifkin, J. L., Castillo, A. S., Liao, I. T., & Rauscher, M. D. (2019). Gene flow, divergent selection and resistance to introgression in two species of morning glories (*Ipomoea*). *Molecular Ecology*, 28, 1709–1729.
- Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, 360, 656–660.
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *The R Journal*, 8, 289–317.
- Sedlazeck, F. J., Rescheneder, P., & von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*, 29, 2790–2791.
- Smith, B. W. (1963). The mechanism of sex determination in *Rumex hastatulus*. *Genetics*, 48, 1265–1288.
- Smith, B. W. (1969). Evolution of sex-determining mechanisms in *Rumex*. *Chromosomes Today*, 2, 172–182.
- Smith, D. A. S., Gordon, I. J., Traut, W., Herren, J., Collins, S., Martins, D. J., Saitoti, K., Ireri, P., & Ffrench-Constant, R. (2016). A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. *Proceedings of the Royal Society B: Biological Sciences*, 283, e20160821.
- Todesco, M., Owens, G. L., Bercovich, N., & Légaré, J. S. (2019). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, 584, 602–607.
- Torgasheva, A. A., & Borodin, P. M. (2016). Cytological basis of sterility in male and female hybrids between sibling species of grey voles *Microtus arvalis* and *M. levis*. *Scientific Reports*, 6, e36564.
- Tucker, P. K., Sage, R. D., Warner, J., Wilson, A. C., & Eicher, E. M. (1992). Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution*, 46, 1146–1163.
- Turelli, M., & Moyle, L. C. (2007). Asymmetric post-mating isolation: Darwin's corollary to Haldane's rule. *Genetics*, 176, 1059–1088.
- Turelli, M., & Orr, H. A. (2000). Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*, 154, 1663–1679.
- Wang, S., Nalley, M. J., Chatla, K., & Aldaimalani, R. (2022). Neo-sex chromosome evolution shapes sex-dependent asymmetrical introgression barrier. *Proceedings of the National Academy of Sciences, USA*, 119, e2119382119. <https://doi.org/10.1101/2021.07.13.452191>
- Wang, S., Rohwer, S., Delmore, K., & Irwin, D. E. (2019). Cross-decades stability of an avian hybrid zone. *Journal of Evolutionary Biology*, 32, 1242–1251.
- Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14, 851–865.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Beaudry, F. E. G., Rifkin, J. L., Peake, A. L., Kim, D., Jarvis-Cross, M., Barrett, S. C. H., & Wright, S. I. (2022). Effects of the neo-X chromosome on genomic signatures of hybridization in *Rumex hastatulus*. *Molecular Ecology*, 31, 3708–3721. <https://doi.org/10.1111/mec.16496>