

# Evidence that genomic incompatibilities and other multilocus processes impact hybrid fitness in a rattlesnake hybrid zone

Zachary L. Nikolakis,<sup>1</sup> Drew R. Schield,<sup>1,2</sup> Aundrea K. Westfall,<sup>1</sup> Blair W. Perry,<sup>1</sup> Kathleen N. Ivey,<sup>1</sup> Richard W. Orton,<sup>1</sup> Nicole R. Hales,<sup>1</sup> Richard H. Adams,<sup>3</sup> Jesse M. Meik,<sup>4</sup> Joshua M. Parker,<sup>5</sup> Cara F. Smith,<sup>6</sup> Zachariah Gompert,<sup>7</sup> Stephen P. Mackessy,<sup>6</sup> and Todd A. Castoe<sup>1,8</sup>

<sup>1</sup>Department of Biology, University of Texas at Arlington, Arlington, Texas 76019

<sup>2</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309

<sup>3</sup>Department of Biological and Environmental Sciences, Georgia College and State University, Milledgeville, Georgia 31061

<sup>4</sup>Department of Biological Sciences, Tarleton State University, Stephenville, Texas 76402

<sup>5</sup>Department of Life Sciences, Fresno City College, Fresno, California 93741

<sup>6</sup>School of Biological Sciences, University of Northern Colorado, Greeley, Colorado 80639

<sup>7</sup>Department of Biology, Utah State University, Logan, Utah 84322

<sup>8</sup>E-mail: todd.castoe@uta.edu

Received September 4, 2021

Accepted August 15, 2022

Hybrid zones provide valuable opportunities to understand the genomic mechanisms that promote speciation by providing insight into factors involved in intermediate stages of speciation. Here, we investigate introgression in a hybrid zone between two rattlesnake species (*Crotalus viridis* and *Crotalus oreganus concolor*) that have undergone historical allopatric divergence and recent range expansion and secondary contact. We use Bayesian genomic cline models to characterize genomic patterns of introgression between these lineages and identify loci potentially subject to selection in hybrids. We find evidence for a large number of genomic regions with biased ancestry that deviate from the genomic background in hybrids (i.e., excess ancestry loci), which tend to be associated with genomic regions with higher recombination rates. We also identify suites of excess ancestry loci that show highly correlated allele frequencies (including conspecific and heterospecific combinations) across physically unlinked genomic regions in hybrids. Our findings provide evidence for multiple multilocus evolutionary processes impacting hybrid fitness in this system.

**KEY WORDS:** Hybridization, linkage disequilibrium, population genomics, selection.

Understanding the mechanisms that generate and maintain species is a central goal of biology. Characterizing the factors that limit introgression between lineages by studying natural hybrid zones can provide valuable insight into the processes that contribute to reduced gene flow (i.e., reproductive isolation) and ultimately drive speciation (Abbott et al. 2013; Harrison and Larson 2014; Taylor and Larson 2019). Multiple evolutionary processes may shape hybrid fitness and genomic patterns of introgression

to varying degrees (Moore 1977; Felsenstein 1981; Barton 1983; Seehausen et al. 2014). A key priority for ongoing work is to understand how such processes interact to collectively shape hybrid zone dynamics and reproductive isolation, and how this varies across the speciation continuum (Coyne and Orr 2004; Seehausen et al. 2014; Moran et al. 2021).

Decades of speciation research have produced a diversity of models to explain hybrid zones and the evolutionary processes

that govern them (Endler 1977; Moore 1977; Barton and Hewitt 1985). The prominent tension zone model, for example, explains hybrid zones as a migration-selection balance between dispersal of parental forms and selection against hybrids (Bazykin 1969; Barton and Hewitt 1985). This model presumes that hybridization results primarily in deleterious incompatibilities, leading to reduced fitness or inviability of hybrids. The best understood form of such incompatibilities, Bateson-Dobzhansky-Muller incompatibilities (BDMIs; Bateson 1909; Dobzhansky 1937; Muller 1942; Coyne and Orr 2004; Cutter 2012), involves deleterious epistatic interactions and can occur as a result of remixing of genotypes that evolved under neutral or adaptive evolutionary processes in isolation and appear common based on empirical evidence (Presgraves et al. 2003; Payseur and Hoekstra 2005). Key features of BDMIs are asymmetric fitness impacts in different parental backgrounds, resulting in a subset of hybrid genotypes that negatively impact hybrid fitness.

In contrast to the tension zone model and its emphasis on intrinsic incompatibilities, other models, including the bounded hybrid superiority model and the ecotonal model, predict that hybrid fitness is environment dependent (i.e., extrinsic), and that hybrids may have increased fitness within the bounds of the hybrid zone (Endler 1973; Moore 1977). Other related models suggest context-dependent selection on hybrids (Seehausen et al. 1997; Craig et al. 2007), including selection for hybrid traits that are uniquely well-suited to environments occupied by hybrids (Stelkens et al. 2009; Selz and Seehausen 2019; Hessenaue et al. 2020). Despite the diversity of models proposed to explain hybrid zone dynamics, the likely reality is that multiple selective processes operate simultaneously within a single hybrid zone, act on individual loci or sets of loci, and may even oppose or amplify one another (Seehausen et al. 2014; Butlin and Smadja 2018; Moran et al. 2021). This suggests that understanding hybridization requires an appreciation of the complex actions and interactions among multiple processes.

A number of gene-centric models have emerged to explain hybrid zone dynamics, including single-locus and multilocus models that differ in their emphasis on a single gene (or trait) versus multilocus targets of selection (Barton 1983; Barton and Bengtsson 1986; Flaxman et al. 2013; Butlin and Smadja 2018) that may promote reproductive isolation, often collectively referred to as “barrier loci” (Seehausen et al. 2014; Butlin and Smadja 2018; Schilling et al. 2018). Key priorities for future work include the integration of existing broad models that describe hybrid zone dynamics with empirical examples of how single- and multilocus patterns of selection collectively contribute to the buildup of reproductive isolation (i.e., the reduction of gene flow between introgressing lineages). The buildup of associations (i.e., linkage disequilibrium; LD) among loci in hybrids may arise through a diversity of processes, es-

pecially during periods of allopatry, such as genetic drift, or via natural selection related to an accumulation of intrinsic and extrinsic barriers, which may result in reproductive isolation (Rundle and Nosil 2005). Accordingly, quantifying the strength (i.e., magnitude of LD) and patterns associated with physically unlinked (e.g., loci on different chromosomes) but nonetheless correlated allelic combinations in hybrids is important for understanding the consequences of hybridization because it has the potential to identify genomic regions with functional or epistatic interactions relevant to hybrid fitness. Furthermore, the precise patterns of LD in hybrids can provide useful evidence for discriminating among evolutionary processes that shape hybrid fitness (Nosil et al. 2021). In hybrid zones, the expectations for the strength and abundance of genetic incompatibilities and other multilocus effects likely depend upon the underlying degree of divergence and variation in demographic history between hybridizing lineages (Abbott et al. 2013; Moran et al. 2021). For example, species that have substantially diverged in allopatry and experience gene flow upon secondary contact may exhibit a large number of intrinsic multilocus interactions in hybrids that reduce hybrid fitness, particularly in genomic regions that are highly differentiated between parental lineages (Schumer et al. 2014). Because hybridization often occurs at the margins of parental lineage ranges, we may expect that other extrinsic factors (e.g., difference in ecological or environmental conditions) related to the unique conditions of the hybrid zone may also impact hybrid fitness (Gompert et al. 2012b; Nosil et al. 2012).

Here, we investigate a hybrid zone between two rattlesnake species, the Prairie Rattlesnake (*Crotalus viridis viridis*) and the Midget Faded Rattlesnake (*C. oreganus concolor*), to characterize genomic patterns of introgression and relate these to potential underlying processes that may have shaped these patterns. These two rattlesnake lineages differ substantially in phenotypic characteristics including venom composition, body size, and coloration. *Crotalus o. concolor* possesses venom that is characterized by the presence of a heterodimeric neurotoxic component (concolor toxin; Modahl and Mackessy 2016), and has a smaller body size with comparatively indistinct color patterning. Conversely, *C. v. viridis* individuals are typically larger bodied with venom that consists mainly of lytic components (Ashton 2001; Mackessy 2010; Saviola et al. 2015). These species also occupy distinct ecological settings, with *C. o. concolor* inhabiting high desert regions of the Colorado Plateau and *C. v. viridis* inhabiting grasslands of the Great Plains region (Campbell et al. 2004; Parker and Anderson 2007). Previous studies have suggested that these species diverged during the Pliocene (~3–5.2 MYA; Fig. S1) and were isolated during Pleistocene glacial cycles, and have since come into secondary contact after range expansion during the Holocene (Pook et al. 2000; Schield et al. 2019b).

Rattlesnakes provide an interesting model for studying hybridization because previous research has shown that hybridization between distantly related species is fairly common in nature, and lab-based crosses of divergent species produce viable offspring, together suggesting incomplete isolation mechanisms (Meik et al. 2008; Smith and Mackessy 2016; Zancolli et al. 2016; Schield et al. 2019b; Myers 2021). As such, investigating the genomic patterns of introgression in rattlesnakes may provide valuable perspectives on the mechanisms and processes that underlie partial reproductive isolation in species that otherwise lack apparent complete reproductive barriers, and possess female heterogametic (ZW) sex determination mechanisms. In this study, we characterize the genomic patterns of introgression in this hybrid zone by incorporating genomic cline models, measures of parental population genetic differentiation, and genotype correlation patterns across physically unlinked genomic regions in hybrids, all evaluated in the context of a chromosome-level reference genome for the Prairie Rattlesnake (Schield et al. 2019a) and recombination maps for each hybridizing species (Schield et al. 2020). Using these approaches and resources, we address the following questions: (1) do loci with high population differentiation between parental lineages show outstanding patterns of introgression in hybrids?, (2) what is the relationship between fine-scale recombination rates in parental lineages and locus-specific patterns of introgression in hybrids?, (3) do we see evidence of elevated levels of nonrandom associations among genomic loci in hybrids consistent with selection for specific combinations of parental alleles, and (4) do patterns of correlated allele frequencies between physically unlinked loci in hybrids support the influence of genomic incompatibilities or other evolutionary processes contributing to partial reproductive isolation between hybridizing lineages?

## Materials and Methods

### SAMPLING, ddRADseq GENERATION, AND VARIANT CALLING

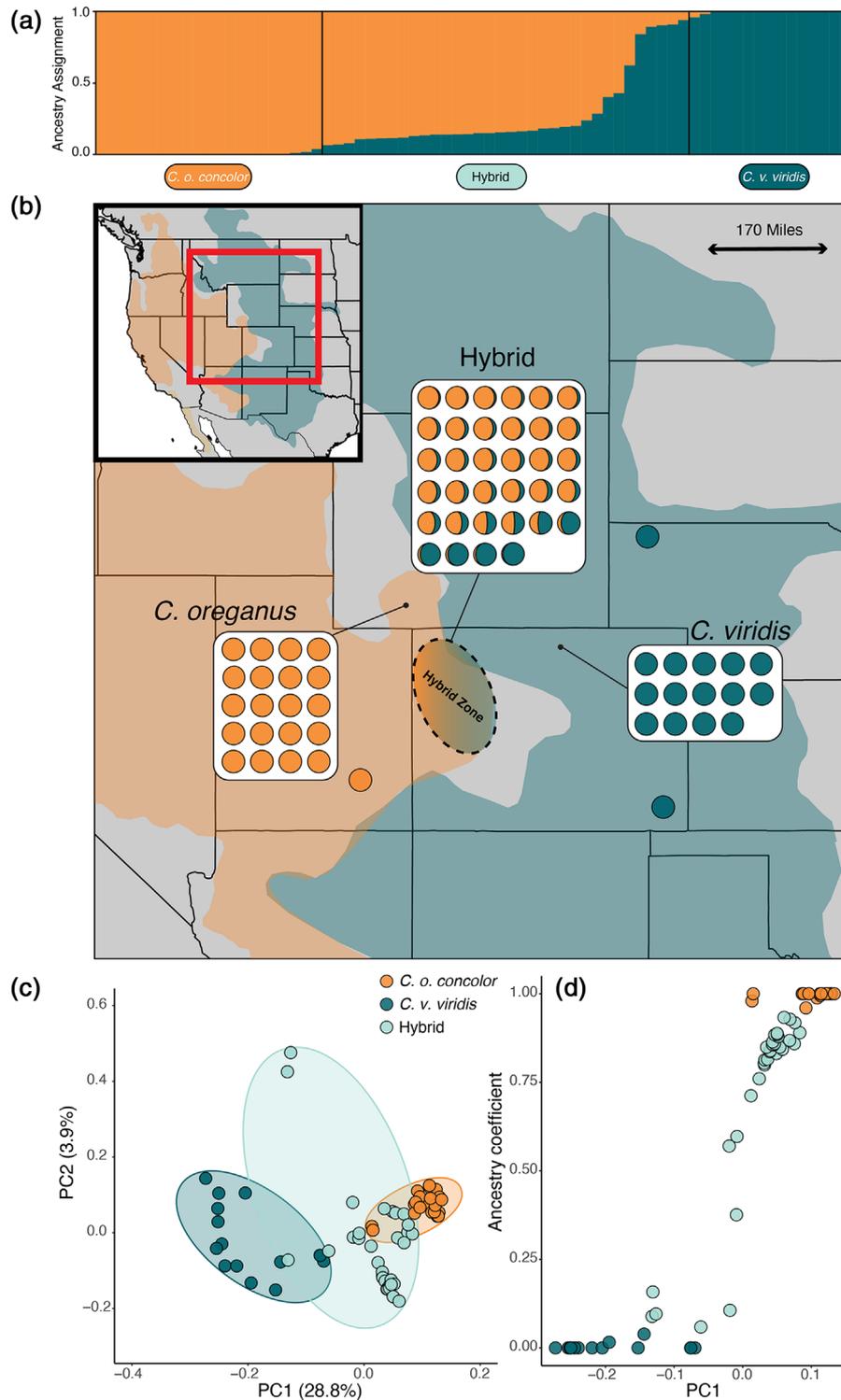
In this study, we integrated new genomic data with data from a previous study ( $n = 50$ ; Schield et al. 2019b) that sampled both parental lineages (*C. v. viridis* and *C. o. concolor*) and an identified hybrid zone between these species. We generated new reduced representation genomic libraries (ddRAD) for an additional 20 samples from the hybrid zone (Table S1; Figs. 1, S1), for a combined total of  $n = 70$  samples used for analyses in this study (see Appendix S1 for detailed method on DNA extraction ddRAD library preparation).

We used the Stacks version 1.42 (Catchen et al. 2013) clone\_filter module to filter PCR clones from raw sequence data, then trimmed 8 bp of adapter sequence from all reads using Fastx-

Toolkit (Hannon 2014). We then demultiplexed trimmed sequences into individual samples using the Stacks process\_radtags module, which demultiplexes samples according to their 6-bp individual barcode and checks for intact restriction cut sites. Processed and demultiplexed data are available from the NCBI SRA (BioProject PRJNA548132). We mapped read data for all samples to the *Crotalus viridis* reference genome (genome size = 1.3 Gb; seven macrochromosomes, 10 microchromosomes, and one sex chromosome [Z]; Schield et al. 2019a) using BWA version 0.7.10 (Li and Durbin 2009) using default settings for the local alignment option “mem.” We then sorted mappings, generated a “pileup” of all sample alignments, and called variants, including multiallelic sites using a combination of SAMtools and BCFtools version 1.9 (Li et al. 2009). We recoded all raw variant calls within individuals with a read depth fewer than five ( $DP < 5$ ) as missing data using the BCFtools filter function. We then filtered variants to retain only biallelic SNPs every 1000 base pairs with a minor allele frequency greater than 0.05 and quality scores (QUAL)  $> 30$  that also had data from at least 20% of samples across the dataset using VCFtools (Danecek et al. 2011) and included only those SNPs that were found on identified chromosomes.

### POPULATION STRUCTURE AND IDENTIFICATION OF HYBRIDS

We used the program STRUCTURE version 2.3.4 (Pritchard et al. 2000; Falush et al. 2003; Hubisz et al. 2009) to infer population structure based on admixture (ancestry) proportions of one or more genetic clusters using the admixture model and correlated allele frequencies with an uninformative prior assumption over the admixture proportions of individuals to  $K$  different populations. The program Strauto (Chhatre and Emerson 2017) was then used to run multiple iterations of  $K$  values ranging from 1 to 4 by using the parallelization function; this included 20 iterations of 25,000 burn-in generations followed by 200,000 logged generations. The program CLUMPP was used to combine all iterations for each  $K$  value using the greedy algorithm to estimate the most likely ancestry proportions. We visualized the results using the program “pophelper” in the statistical package R (Francis 2017). Putative hybrid individuals were identified as those having ancestry coefficients  $< 95\%$  for one of the two parental backgrounds in the  $K = 2$  STRUCTURE model; individuals in this class were also strictly associated with the specific geographic region assumed to be the hybrid zone. We performed a principal component analysis (PCA) on the SNP covariance matrix estimated in the R package “SNPRelate” (Patterson et al. 2006) to visualize the ordination of putative hybrid individuals in relation to parental populations. To validate individual hybrid identification further, we also visualized the relationship between ancestry coefficients from our STRUCTURE



**Figure 1.** Genetic evidence of hybridization between *Crotalus viridis viridis* and *Crotalus oreganus concolor*. (a) STRUCTURE plot showing ancestry coefficients for all sampled individuals. (b) Range map of both parental species with sampling localities used in this study indicated by colored circles corresponding to ancestry coefficients from STRUCTURE. Note that the entire range map is shown for the binomial *C. oreganus*. (c) Principal component analysis (PCA) showing the distribution of genetic variance and separation between parental lineages and hybrid individuals. Each group is delineated by ellipses that were estimated using the Khachiyan algorithm that correspond to sampling colors from the species range map. (d) Correlation between the first principal component (PC1) and ancestry coefficient from the STRUCTURE analysis.

output and the first axis from the principal component analysis (PC1) to test whether population membership between both analyses was congruent. Additionally, to confirm that missing data did not bias population structure or group assignment, we repeated these analyses on a more complete dataset (see methods in Appendix S1).

### GENOMIC CLINE ANALYSIS

Our designations of hybrid and parental *C. v. viridis* and *C. o. concolor* populations were based on ancestry coefficients inferred in STRUCTURE, which also corresponded well with expected ancestry based on the geographic dispersion of samples and relationships between genetic variation and geographic distance (Table S1). Using these designations, we calculated locus-specific allele frequencies for each population and used the genomic cline model implemented in *bgc* (Gompert and Buerkle 2011) to identify loci that deviate from null models of genome-average introgression (see Appendix S1 for detailed methods on genomic cline analyses). Here, we use term “excess ancestry” to refer to loci for which the parental allelic contributions in hybrid populations deviate from the predicted parental allele frequencies based on the remainder of the genome, as quantified by the hybrid index (Buerkle and Lexer 2008; Gompert and Buerkle 2011). That is, these excess ancestry loci exhibit biased allele frequencies favoring one of the two parentals. Departures from 0 in either  $\alpha$ , the genomic cline center parameter that describes cline shifts in relation to hybrid indices, or  $\beta$ , the genomic cline rate parameter that describes the slope of locus-specific clines, denote an increase or decrease in the ancestry probability from one parental species or the rate in transition from one species to another. Genetic drift can cause patterns of introgression that deviate from null, genome-average expectations (i.e.,  $\alpha$  and  $\beta = 0$ ), but such deviations should be more prevalent for loci effected by selection in hybrids, and thus loci with credible departures from  $\alpha$  or  $\beta = 0$  should be enriched for loci experiencing selection (i.e., for barrier loci and other loci in LD with barrier loci). Using our estimates of posterior probabilities of inheritance from parental populations ( $\Phi$ ) inferred from *bgc*, we then compared these to a simplified model of hybridization to test whether we inferred excess ancestry loci in a simulated neutrally introgressing hybrid population (see Appendix S1 for details on simulation methods).

### RELATIONSHIPS BETWEEN GENETIC DIFFERENTIATION AND INTROGRESSION ACROSS GENOMIC REGIONS

To test for relationships between patterns of divergence in allopatry versus excess ancestry loci in *bgc* analyses, we estimated genetic differentiation between parental populations using Weir and Cockerham’s  $F_{ST}$  in VCFtools using a site-based approach (Weir

and Cockerham 1984). We tested the relationship between the absolute values of the genomic cline center parameter,  $\alpha$ , and  $F_{ST}$  values using Pearson’s correlation coefficient in R (R Core Team 2021), with the expectation that a positive correlation indicates that loci that are more highly differentiated between parental lineages also harbor excess ancestry from one parental population or the other in hybrids. This would be expected if such loci were enriched loci involved in BDMIs or otherwise affected hybrid fitness (i.e., by affecting traits under environment-dependent selection), or if they were in high LD with such loci. We also performed an ANOVA to test whether there were significant differences in mean values of  $F_{ST}$  for loci that were inferred to introgress neutrally (i.e.,  $\alpha$  and  $\beta = 0$ ) versus loci with excess *C. v. viridis* or *C. o. concolor* ancestry in hybrids.

Based on prior studies that have shown evidence that chromosome classes or specific chromosomal regions may differ substantially in structure and content (O’Connor et al. 2019; Perry et al. 2021), recombination (Kawakami et al. 2014; Schield et al. 2020), and relative importance in pre- and postmating isolation in speciation (Irwin 2018), we tested if specific chromosomal classes or regions were enriched for excess ancestry loci. To investigate this, we compared differences in the proportion of excess ancestry loci between chromosome classes (e.g., micro-, macro-, and sex-chromosomes) using the “prop.test” function in R to test the null hypothesis of equal proportions. We also tested for enrichment between specific intrachromosomal regions because preliminary analyses suggested that the pseudoautosomal region (PAR) of the Z chromosome is enriched for excess ancestry loci compared to the remaining portion. Additional comparisons of the distribution of excess ancestry loci between distal and proximal autosome regions are provided for context in Appendix S1.

### RELATIONSHIPS BETWEEN RECOMBINATION, PARENTAL LINEAGE DIFFERENTIATION, AND HYBRID EXCESS ANCESTRY LOCI

We leveraged recombination maps estimating  $\rho = 4N_e r$  (Schield et al. 2020) for both parental species (*C. viridis* and *C. oregonus*) to test for relationships between recombination rates, chromosome classes, loci categories, and levels of differentiation between parental lineages across regions of the genome. We used 10-kb windowed recombination maps from *C. viridis* and *C. oregonus* to limit the number of individual loci within a given window, because our sampled ddRAD loci do not encompass the entire genome and included windows that contained neutrally introgressing or excess ancestry loci exclusively. Because previous hybrid zone studies have demonstrated complex relationships between recombination and genome-wide patterns of introgression (Schumer et al. 2018; Calfee et al. 2021), and prior studies in this system have shown significant differences in

recombination across chromosome categories (Schield et al. 2020), we examined the relationship between local recombination rates, loci categories, and parental differentiation across multiple scales (i.e., genome-wide and within and among chromosomes; see Appendix S1 for details of all comparisons).

We investigated evidence for LD between physically unlinked genomic regions (from different chromosomes) in hybrids by comparing allele frequencies for loci in hybrids that were either excess ancestry or neutrally introgressing loci (defined by *bgc* inferences). Because our genomic data were not phased, we quantified allele frequency correlations using a genotype-based LD measure by estimating the squared Pearson correlation coefficient ( $r^2$ ). To do this, we analyzed only hybrid individuals and split the variant dataset into three separate VCF files containing (1) only excess *C. o. concolor* ancestry loci, (2) only excess *C. v. viridis* ancestry loci, and (3) all inferred neutrally introgressing loci from *bgc* and then calculated the squared Pearson's correlation coefficient ( $r^2$ ) for each pairwise interchromosomal SNP combination, using the "interchrom-geno-r2" flag in VCFtools. We also calculated the  $r^2$  value for combinations of *C. o. concolor* and *C. v. viridis* ancestry loci to assess the degree of correlation between heterospecific associations. To remove the effects of physical linkage in these analyses, we only examined pairs of excess ancestry loci that were derived from physically unlinked regions of the genome by only analyzing loci located on different chromosomes (see Appendix S1 for methods detailing the exclusion of structural variation and misassemblies in our dataset). To investigate the relationships between recombination and interchromosomal associations, we performed an ANOVA to compare patterns of mean recombination rates across interchromosomal categories for sets of correlated loci in hybrids (conspecific vs. heterospecific).

To understand the degree and extent of allele frequency associations among physically unlinked loci within hybrid genomes (i.e., different chromosomes), we examined multiple  $r^2$  value thresholds for inter- and intrachromosomal allele correlations. We first explored the results by comparing the overall distribution between parental-specific excess ancestry categories for all pairs of loci with any appreciable level of correlation ( $r^2 > 0.1$ ); statistically significant differences between these distributions were determined using a Welch's two-sample *t*-test. We then visualized the genomic locations of loci that had five or more interchromosomal correlated locus connections with an  $r^2$  value of 0.5 (and where data were observed in at least 15 hybrid individuals). To understand the relative size of suites of correlated loci (regarding loci in LD), we estimated the number of interchromosomal loci pairs of excess ancestry loci with allele frequencies correlated at  $r^2 = 1.0$  in hybrids (see Appendix S1 for methods on determining significance of  $r^2$  values).

## Results

### VARIANT DATASET, POPULATION STRUCTURE, AND PARENTAL DIFFERENTIATION

We analyzed reduced-representation sequencing data (ddRAD-seq) for a total of 70 individuals from parental and hybrid zone populations, averaging 5.8 million mapped reads per sample (Table S1). We retained 8924 SNPs after filtering our dataset to retain biallelic variants separated by at least 1000 bp with genotype quality scores (QUAL)  $>30$ , and minor allele frequencies above 0.05; the majority of retained SNPs were associated with high-quality scores (Fig. S2). This initial set was further filtered to a set of 8590 SNPs that met the following criteria: SNPs mapping to scaffolds that were assigned to chromosomes, loci that were observed in at least one hybrid and at least one individual from both parental lineages, loci that did not differ in neutral introgression or excess ancestry assignment across multiple *bgc* runs, and loci that fell below excess ancestry thresholds (median  $\alpha$  values) based on neutral introgression simulation runs of *bgc* (Table S2). Results from our STRUCTURE analysis indicated that a *K* value of 2 (Fig. 1a) was the best fit model based on the  $\Delta K$  method (Evanno et al. 2005). Individual ancestry coefficients within the hybrid zone varied; however, a majority of hybrids exhibited greater *C. o. concolor* ancestry (average hybrid ancestry = 0.71), with only eight of 34 hybrid samples having greater *C. v. viridis* ancestry, consistent with an overall greater degree of backcrossing with *C. o. concolor* (Figs. 1a,b, S3, S4). Results from our PCA showed separation between the parental species, with hybrid individuals positioned intermediate between the parental species based on PC1 and PC2, which explained 28.8% and 3.9% of the genetic variation, respectively (Fig. 1c). We found a positive relationship between PC1 scores and ancestry coefficients inferred from STRUCTURE, suggesting that the primary axis of SNP variation across samples corresponds strongly with inferred parental ancestry inferred by STRUCTURE (Fig. 1d). Additionally, using our more complete (80% complete) dataset of 1750 SNPs we found no difference in group membership and that most individual ancestry coefficients were nearly identical (Spearman's  $\rho = 0.99$ ;  $P$ -value  $\leq 2.2 \times 10^{-16}$ ) to our full dataset; comparative analysis of this more complete dataset also showed no major shifts in group ordination in PCA analyses (Fig. S5).

Overall genetic differentiation was moderately high between the two parental populations (mean  $F_{ST} = 0.47$ ,  $\sigma = 0.36\pm$ ), with 13% of loci being fixed for different alleles ( $F_{ST} = 1$ ) in the parental populations and 8% with values of zero (Table S2). Micro- and macrochromosomes had very similar levels of differentiation (average  $F_{ST\ Macro} = 0.47$ ,  $\sigma_{Macro} = 0.36\pm$ ,  $F_{ST\ Micro} = 0.46$ ,  $\sigma_{Micro} = 0.35\pm$ ), whereas on average, loci located on the Z chromosome tend to have higher degrees of

differentiation (average  $F_{STZ} = 0.49$ ,  $\sigma_Z = 0.40$ ) with most of these sites occurring on the sex-linked region of the Z (Table S2).

## EMPIRICAL AND SIMULATED PATTERNS OF INTROGRESSION

We found a consistent pattern of relatively equal ratios of excess ancestry loci from each parental populations in our simulations. Here, the minimum and maximum estimates of  $\alpha$  (posterior medians) for simulated excess ancestry loci across our simulation experiment were 0.75 and  $-0.67$ , respectively, which we then used to filter potential false positives for excess ancestry in our empirical dataset (Fig. S6). Visualization of our MCMC output from *bgc* confirmed that all five empirical runs showed convergence and were therefore used in subsequent analyses (Fig. S7). We recovered 530 SNPs ( $\sim 6\%$  of full dataset) with evidence of excess ancestry for one of the two parental lineages (i.e.,  $\alpha$  95% CI  $\neq 0$  and median  $\alpha$  greater or less than maximum and minimum simulated excess ancestry loci); this magnitude of excess ancestry loci is similar to estimates of inferred excess ancestry loci from previous studies that have used *bgc* (Fig. 2; Gompert et al. 2012a; Schield et al. 2017; Baiz et al. 2019). These 530 excess ancestry SNPs included 279 loci with evidence of excess *C. v. viridis* ancestry and 251 with excess *C. o. concolor* ancestry that were distributed across all chromosomes except chromosome 14, which did not contain any excess *C. o. concolor* ancestry loci (Table S2, Figs. 2b, 3a). We did not recover any loci for which the 95%  $\beta$  CI did not contain zero.

Our plots of excess ancestry (excess  $\alpha$ ) loci across chromosomes also suggest relatively small ancestry blocks exist, with adjacent chromosome blocks showing excess ancestry for different parental populations (Fig. 3a), consistent with prolonged admixture in the hybrid zone and with decreased effects of admixture LD under neutral introgression. To further explore the distribution of excess ancestry loci across chromosomes, we regressed the total numbers and proportion of excess ancestry loci per parental lineage against chromosome length (Fig. S8). These results suggest that approximate number of excess ancestry loci per chromosome is correlated with chromosome length (Fig. S8; Pearson's correlation [ $r$ ] for *C. concolor* excess ancestry loci = 0.91,  $P$ -value  $< 0.005$ ; Pearson's correlation [ $r$ ] for *C. viridis* excess ancestry loci = 0.94,  $P$ -value  $\leq 0.005$ ), and highlight differences among chromosomes in the relative proportion of excess ancestry loci from each parental lineage (Fig. S8c,d; see also Fig. 2b).

Microchromosomes were enriched for excess ancestry loci compared to macrochromosomes (based on a proportion test;  $P = 0.03$ ), but there was no evidence of enrichment among chromosomes within each of the two chromosome classes ( $P_{\text{Macro}} = 0.786$ ,  $P_{\text{Micro}} = 0.435$ ). Lineage-specific excess ancestry estimates on the Z chromosome showed a distinct pattern

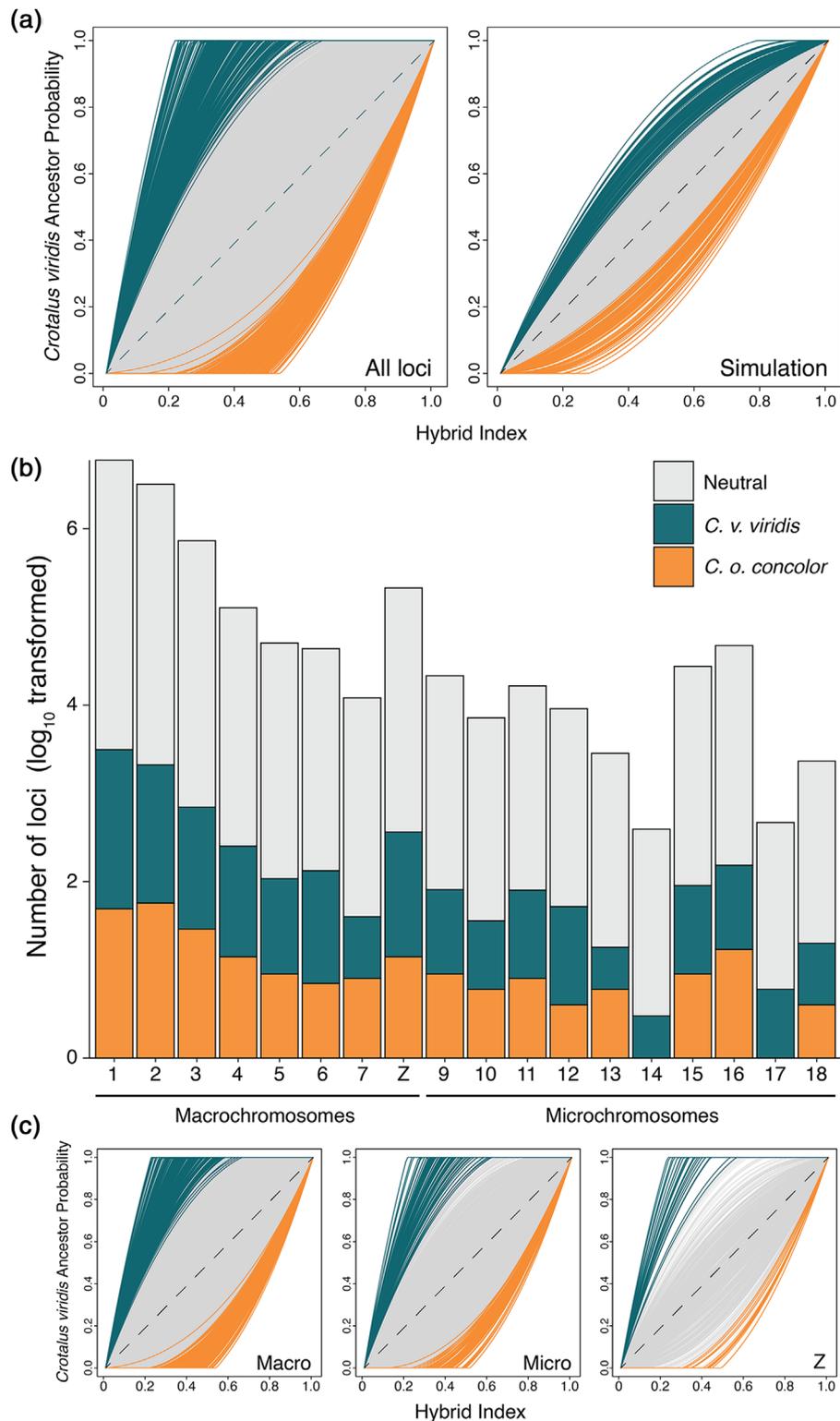
from autosomes, and there were twice as many loci with excess *C. v. viridis* ancestry ( $n = 26$ ) as those with excess *C. o. concolor* ancestry ( $n = 14$ ; Table S2). In contrast, the PAR, the only region of the Z chromosome that recombines in both sexes, contained nearly equal amounts of lineage-specific excess ancestry loci ( $n_{C. v. viridis} = 7$ ,  $n_{C. o. concolor} = 6$ ) and was enriched for excess ancestry loci compared to the sex-linked region of the Z chromosome (proportion test;  $P = 0.0008$ , Fig. 3b).

## RECOMBINATION RATES IN GENOMIC REGIONS WITH EXCESS ANCESTRY LOCI

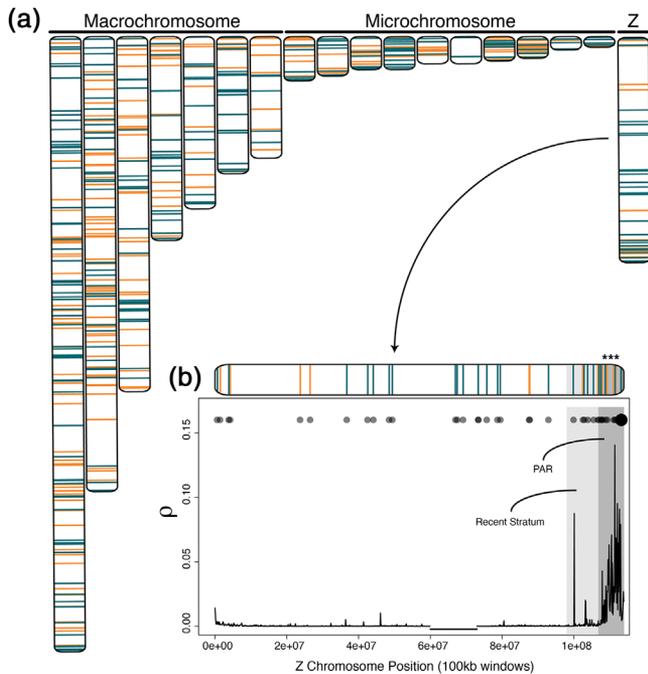
We used the recombination maps for both parental species (Schield et al. 2020) to test for significant differences in recombination rates between windows that contained excess ancestry or neutrally introgressing loci across multiple scales. Despite broad-scale conservation in genome-wide recombination landscapes between the two species (Schield et al. 2020), the locations of most fine-scale recombination hotspots differ due to presence of PRDM9 in the germline. Genome-wide analyses highlight stark differences in the relationships between recombination rates and locus categories depending on the parental recombination map used (Fig. 5). Based on the *C. oregonus* recombination map, we found significant differences between the recombination rates of windows that contained excess ancestry loci versus neutrally introgressing markers ( $t$ -tests,  $P < 0.05$ ; Fig. 5c,d), but found no significant differences based on comparisons with the *C. viridis* recombination map ( $P = 0.79$ ; Fig. 5a,b). In comparisons with the *C. oregonus* map, windows that only contained excess ancestry loci had higher recombination rates on average than neutrally introgressing markers, and also had a higher maximum overall recombination rate (Table S3). When examining these relationships by chromosome class, we find the same pattern of excess ancestry loci tending to occur in genomic windows with higher recombination rates, but find no statistically significant differences between loci categories (Fig. S9). We find similar patterns when comparing recombination rates and excess ancestry categories across chromosome classes, with *C. concolor* excess ancestry loci residing in regions of higher recombination than *C. viridis* excess ancestry loci, yet there is no significant difference (Fig. S10).

## RELATIONSHIP BETWEEN PARENTAL GENETIC DIFFERENTIATION, INTROGRESSION, AND RECOMBINATION

To understand how locus-specific patterns of genetic differentiation in allopatry may impact the process of hybridization, we compared  $F_{ST}$  distributions between allopatric parentals across each locus category (neutrally introgressing, *C. o. concolor* excess ancestry, and *C. v. viridis* excess ancestry). Excess ancestry loci from either parental lineage had significantly higher



**Figure 2.** Genome-wide distribution of excess ancestry loci from Bayesian genomic cline analyses. (a) Genomic clines for all loci and simulated datasets. The dashed line represents the relationship of hybrid indices and ancestry probabilities under the expectation of neutral introgression, whereas the colored lines represent loci that show patterns of excess ancestry within hybrid individuals. (b) Stacked bar plot of log transformed counts of *bgc* inferred neutrally introgressing and excess ancestry loci classified by parental lineage across individual chromosomes indicating the relative contributions from each parental species. (c) Genomic clines showing the probability of *C. v. viridis* ancestry for each chromosome class. The dashed line represents the relationship of hybrid indices and ancestry probabilities under the expectation of neutral introgression.



**Figure 3.** Physical genomic distribution of excess ancestry loci within hybrids, including a focus on the Z-chromosome and pseudoautosomal sex-chromosome region (PAR). (a) Location of excess ancestry loci genome-wide with chromosome classes indicated by bars above plot. Colors represent specific parental lineages as from Figures 1–3. (b) Plot showing the distribution of excess ancestry loci across the Z-chromosome in relation to the recombination landscape with recent stratum and pseudoautosomal regions (PAR) highlighted. Circle size is relative to the number of excess ancestry loci found within the corresponding 100-kb recombination windows. Asterisk indicates significant enrichment ( $P < 0.005$ ) of excess ancestry within the PAR relative to the remaining Z-chromosome. Recombination maps were used from Schield et al. (2020).

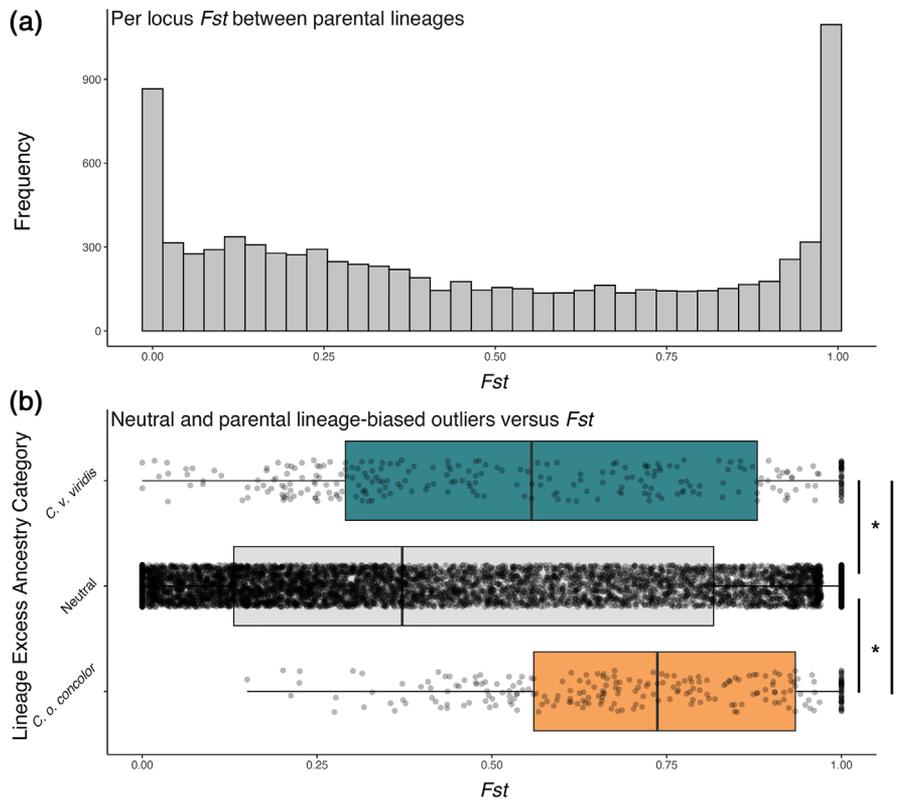
overall  $F_{ST}$  values than neutrally introgressing markers (ANOVA;  $P < 2.0 \times 10^{-16}$ ) and loci that had *C. o. concolor* ancestry ( $\alpha < 0$ ) had higher  $F_{ST}$  on average than loci with excess *C. v. viridis* ancestry ( $\alpha > 0$ , Welch's two-sample  $t$ -test, effect size 0.6;  $P < 2 \times 10^{-16}$ ; Figs. 4, S11). We also found a low but significant positive relationship between locus-specific genetic differentiation of parental populations and absolute values of the genomic cline center parameter (Pearson's correlation  $r = 0.04$ ;  $P < 6.0 \times 10^{-5}$ ; Fig. S11). Additionally, we find that recombination rates and parental differentiation show evidence of a negative relationship for neutrally introgressing loci for recombination maps of both species (Pearson's correlation  $r_{viridis} = -0.069$ ;  $P_{viridis} = 9 \times 10^{-10}$ ;  $r_{oreganus} = -0.05$ ;  $P_{oreganus} = 2.8 \times 10^{-6}$ ; Fig. S12). Conversely, we find a positive, nonsignificant relationship with recombination rates and parental differentiation for excess

ancestry loci (Pearson's correlation  $r_{viridis} = 0.057$ ;  $P_{viridis} = 0.19$ ;  $r_{oreganus} = 0.018$ ;  $P_{oreganus} = 0.67$ ; Fig. S12).

### MEASURES OF CORRELATED ALLELE FREQUENCIES ACROSS LOCI CATEGORIES

Our analyses of correlated allele frequencies among lineage-specific excess ancestry loci showed that squared correlation coefficient values ( $r^2$ ) are significantly higher for *C. o. concolor*-ancestry loci (Welch's two-sample  $t$ -test, effect size 0.32;  $P < 2.2 \times 10^{-16}$ ; Fig. 6a). That is, overall values of the squared Pearson's correlation coefficient ( $r^2$ ) for combinations of *C. o. concolor*-ancestry loci located on different chromosomes were significantly higher than pairwise correlated alleles for loci with excess *C. v. viridis* ancestry. The observation of elevated correlation levels between excess *C. o. concolor* ancestry loci was also consistent across multiple  $r^2$  value filtering thresholds. For example, we observed overall greater numbers of genome-wide coupled connections—multiple physically unlinked loci with correlated allele frequencies in hybrids—for loci with *C. o. concolor* excess ancestry with an  $r^2$  value of  $>0.5$  (Figs. 6b,c, S13). We also found that *C. o. concolor* excess ancestry loci tended to have a greater number of interchromosomal connections (mean = 4.8) with other excess ancestry loci with  $r^2 = 1$  (i.e., alleles that are always co-inherited) than excess *C. v. viridis* ancestry loci (mean = 3.8), although this difference was not statistically significant (Welch's two-sample  $t$ -test, effect size 0.27;  $P = 0.06$ ; Fig. 6d,e). We also find a substantial number of loci with heterospecific allele combinations that exhibit high degrees of interlocus allele correlation ( $r^2 > 0.8$ ; Fig. 6f), but which only involve a limited number of correlated loci connections in contrast to the generally higher multilocus nature of many conspecific associations (Fig. S14). For comparisons between the  $r^2$  distribution for each *bgc* category (i.e., excess ancestry and neutrally introgressing loci), we found that the degree of correlation was higher for excess ancestry loci compared to neutrally introgressing loci (Welch's two-sample  $t$ -test, effect size 0.44;  $P < 2.2 \times 10^{-16}$ ; Fig. S15).

The accumulation of LD, or cline correlations, in hybrids can result from neutral processes (e.g., drift), selection, or a combination of both (Wang et al. 2011), and discerning the roles of selection and drift in driving patterns of LD is challenging. However, neutral processes that generate admixture LD in hybrids are predicted to impact all autosomal chromosomes equally, because they share the same demographic history and inheritance (Harrison and Bogdanowicz 1997), and should result in the buildup of ancestry LD for conspecific correlated alleles. To examine the degree to which patterns of LD in hybrids followed this prediction of neutral introgression, we compared allele correlations across all loci pairs between chromosomes to generate



**Figure 4.** Relationship between parental lineage differentiation and genomic cline outliers in hybrids. (a) Histogram of  $F_{ST}$  values between contributing parental species. (b) Boxplots showing the relative differentiation distribution of loci that are categorized as excess ancestry from each parental lineage or neutral from hybrid individuals. Asterisks indicates significant ( $P < 0.05$ ) differences in pairwise comparisons of  $F_{ST}$  distributions for each category.

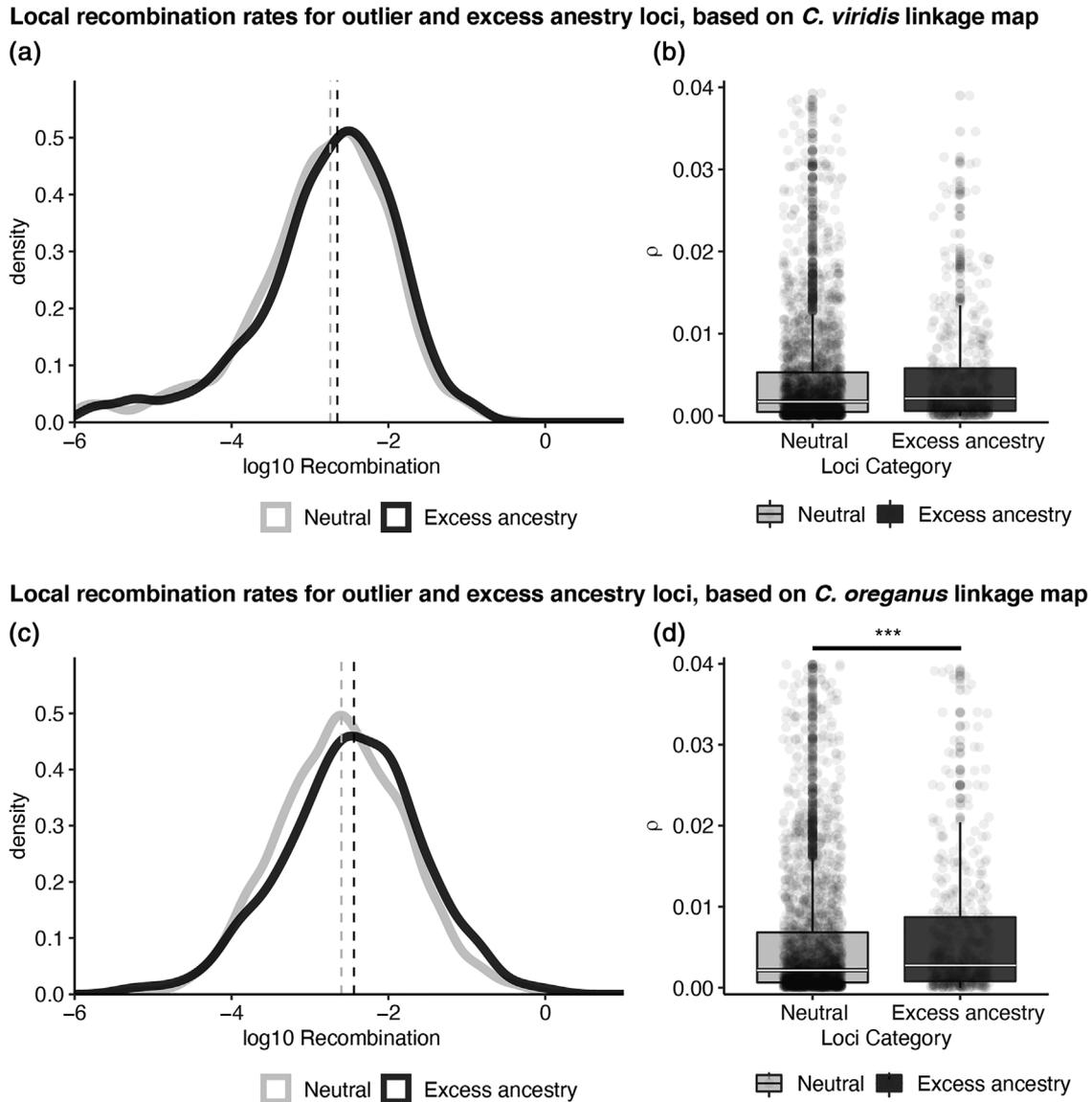
distributions of interchromosomal LD and summarized these patterns for all loci and specifically for excess ancestry loci for each chromosome. As predicted under free introgression, the distributions of the mean and variance of interchromosomal allele correlations ( $r^2$  averaged across all pairwise comparisons for each SNP) for *bgc*-inferred neutral introgressing loci are similar across chromosomes (Fig. S16). However, compared to these neutral introgressing loci distributions, excess ancestry loci from both parental species (especially those from *C. o. concolor*) are consistently outliers, with higher  $r^2$  mean and variance (Fig. S16).

We found that measures of  $r^2$  for intrachromosomal comparisons were significantly higher than interchromosomal pairs in parental populations, supporting that view that inferences of elevated LD between pairs of interchromosomal loci were not misled by genome misassembly or structural variation (Fig. S17). The large number of interchromosomal correlated loci, and dispersion across chromosomes of these loci, also suggests that these inferences are unlikely to be due to genome misassemblies or structural variation between parental lineages, which is consistent with previous evidence for highly correlated recombination landscapes in the two parental species when interpreted based on the *C. viridis* reference genome assembly (Schield et al.

2020). Average  $F_{ST}$  values between parental populations for sets of physically unlinked excess ancestry loci with highly correlated allele frequencies ( $r^2 = 0.8-1.0$ ) were also highest for correlated sets with excess *C. o. concolor* ancestry compared to sets with excess *C. v. viridis* ancestry. Both sets of conspecific highly correlated ( $r^2 > 0.8$ ) and excess ancestry loci had significantly higher average  $F_{ST}$  values than did correlated sets of loci that were identified as neutrally introgressing loci based on *bgc* analyses (Fig. S18).

## Discussion

The evolutionary forces that shape hybrid zone dynamics are diverse and often interact (Barton and Hewitt 1985; Gompert et al. 2017; Schumer et al. 2018; Moran et al. 2021). However, many models and previous approaches consider only single processes in isolation (Schield et al. 2017; Baiz et al. 2019), despite the reality that multiple selective, genomic, and demographic processes likely operate simultaneously on different regions of the genome during introgression and may even oppose or amplify one other (Hvala et al. 2018; Moran et al. 2021). We used Bayesian genomic cline models to characterize genome-wide patterns of

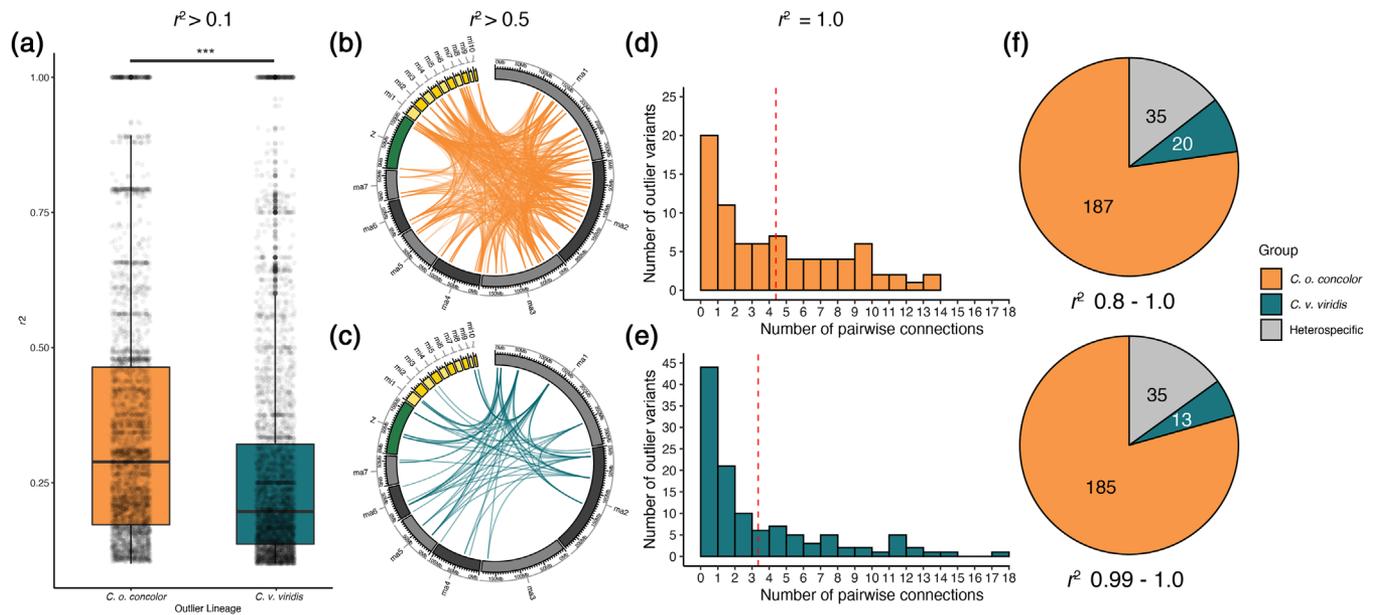


**Figure 5.** Patterns of recombination between excess ancestry and neutrally introgressing loci. (a, c) Density plots of log-transformed recombination rates between excess ancestry and neutrally introgressing loci for *Crotalus viridis* and *Crotalus oregonus* recombination maps with dashed lines representing the median of each category for each recombination map. (b, d) Boxplots of recombination rates between excess ancestry and outlier loci. Asterisks in panel (c) and (d) indicate significant differences between recombination rate distributions between loci categories for *C. oregonus* recombination map ( $P < 0.05$ ).

introgression and find evidence for a large number of excess ancestry loci in hybrids, with a substantial portion of these loci exhibiting correlated allele frequencies in hybrids (i.e., LD) that are on different chromosomes and therefore physically unlinked. Our results indicate that many pairs of physically unlinked loci with highly correlated allele frequencies have conspecific allele combinations from one parental lineage, with a smaller set of correlated loci having heterospecific allele combinations. These patterns of interchromosomal LD in hybrids can be shaped by a combination of neutral and selective processes, and thus a key question is to what extent has selection shaped these patterns. Re-

gardless of their origin, these long-range correlations in ancestry clines will cause ongoing selection on subsets of excess ancestry to spill over (via LD) and affect other correlated loci, creating more of a genome-wide barrier to gene flow.

Both neutral introgression and selection can increase LD between loci on different chromosomes (Wang et al. 2011), which has also been demonstrated by simulation studies (Gompert and Buerkle 2011; Gompert et al. 2012b; Lindtke and Buerkle 2015). Genome misassemblies or structural variants may also lead to spurious inferences of interchromosomal LD. Collectively, the observed abundance and wide distribution of correlated excess



**Figure 6.** Lineage specific combinations of physically unlinked correlated excess ancestry loci among hybrid individuals. (a) Box plot showing the distribution of all  $r^2$  values  $> 0.1$  for each pairwise connection within parental lineage ancestry categories. Asterisks indicate a significant difference between the two categories using a Welch's  $t$ -test ( $P < 2.2 \times 10^{-16}$ ). (b, c) Circos plots showing the genomic localities of connections between pairs of lineage-specific ancestry categories. The lines represent positions in which (i) there were at least 15 sampled individuals for that specific site, (ii) the pairwise connection  $r^2$  is at least 0.5, and (iii) the connection frequency is five or greater. Each circos plot shows the relationship of connections within each parental lineage category with colors corresponding to the lineage colors used through the manuscript. (d, e) Histograms of the connection frequencies for variant connections that have an  $r^2$  of 1.0. For example, panel (d) shows that there are 20 variants that are of *C. o. concolor* ancestry that have at least one connection with an  $r^2$  of 1.0. The dotted line represents the average connections for each lineage category. No additional filtering schemes were applied to this histogram besides a connection  $r^2$  of 1.0 to show the full distribution of perfect genotype correlations between ancestry categories. (f) Pie chart showing the breakdown for the number of pairwise conspecific and heterospecific correlated excess ancestry loci associations that have an  $r^2$  greater than 0.8.

ancestry loci across hybrid chromosomes, the highly correlated recombination landscapes of parental *C. viridis* and *C. oregonus* (Schield et al. 2020), and our comparisons of inter versus intrachromosomal associations in parentals (Fig. S13) all suggest that interchromosomal associations in hybrids are not likely to be due to structural rearrangements or genome misassemblies. Under neutral introgression, patterns of interchromosomal LD in hybrids would be expected to be similar across chromosomes, because they share common historical demographic and inheritance patterns. Our evidence that a subset of excess ancestry loci in hybrids tend to exhibit extreme patterns of interchromosomal allele correlations indicates that neutral processes alone are unlikely to explain the degree of elevated interchromosomal allele correlations, at least for some excess ancestry loci in hybrids. Additionally, neutral admixture is only expected to generate conspecific allele correlations, yet we observe both conspecific and heterospecific allele correlations between excess ancestry loci in hybrids. Considering the predictions of neutral admixture LD, the observed patterns of conspecific and heterospecific allele correla-

tions in hybrids suggest multilocus selective processes impact hybrid fitness and favor specific combinations of hybrid genotypes and are driving allele correlations and long-range LD among physically unlinked loci in hybrids. These conclusions also align with prior simulation studies that have identified similar patterns of LD in hybrids that result from selective processes (including those related to incompatibilities or extrinsic factors; Gompert and Buerkle 2011; Gompert et al. 2012b; Lindtke and Buerkle 2015). Here, our observed patterns of interchromosomal allele correlations in hybrids are consistent with both intrinsic multilocus mechanisms (e.g., incompatibilities) that lead to conspecific combinations of correlated alleles across chromosomes, as well as other likely extrinsic mechanisms (e.g., selection for transgressive phenotypes) underlying heterospecific allelic combinations (Schumer et al. 2014; Knief et al. 2020; Powell et al. 2021). Collectively, our results provide evidence for the roles of both neutral and selective processes shaping rattlesnake hybrids, and for the interaction of multiple selective processes impacting hybrid fitness and contributing to partial reproductive isolation.

## EVIDENCE FOR MULTIPLE INTRINSIC PROCESSES SHAPING GENETIC INCOMPATIBILITIES WITHIN HYBRIDS

Several prominent hybrid zone models (e.g., tension zone, additive hybrid load) predict that hybridization results in reduced hybrid fitness, often through the mixture of intrinsically incompatible genotypes or through exposure of deleterious alleles in hybrids (Barton and Hewitt 1985; Orr 1996; Powell et al. 2020). We identified genomic regions that defy genome-average expectations of free introgression across the hybrid zone and identified pairs of these loci that were physically unlinked yet showed correlated allele frequencies (evidence of LD) in hybrids. Our analyses characterized the nature of these correlations, including the strength of correlation, and whether sets of correlated alleles favored combinations from the same parental species (conspecific associations) or combinations of alleles from different parents (heterospecific combinations). We find that the number of *C. o. concolor* conspecific loci with high interchromosomal correlations ( $r^2 > 0.50$ ) was far greater than those for *C. viridis* conspecific correlated loci, despite nearly equal numbers of excess ancestry loci from each parental lineage (Table S2; Fig. 6b,c). Although some fraction of these conspecific correlated loci are likely the result of neutral introgression and admixture LD, our analyses suggest that neutral introgression alone is unlikely to explain the full extent of these conspecific associations observed in hybrids. Instead, the large number of physically unlinked loci with highly correlated conspecific alleles suggests a role for selection acting on numerous suites of loci in hybrids to maintain conspecific sets of alleles.

Intrinsic incompatibilities would be expected to favor the inheritance of suites of conspecific alleles in hybrids (Johnson 2000; Schumer et al. 2015). Previous studies suggest that effective population size is approximately similar for these two parental lineages (Schield et al. 2019b), indicating that a major influence of hybrid load biases toward the parental lineage with higher  $N_e$  is unlikely to have driven these conspecific associations (Lohr and Haag 2015). Intrinsic incompatibilities may also be associated with asymmetric fitness impacts in different parental backgrounds, selection against minor parent ancestry, and only a subset of hybrid genotypes negatively impacting hybrid fitness (Orr 1996; Coyne and Orr 2004).

Our results identify correlations between collectively hundreds of conspecific loci across different chromosomes, mostly composed of sets of relatively few loci with very strong between-locus correlations (Fig. 6). This pattern is most consistent with expectations of selection against incompatibilities that maintains conspecific associations in hybrids. We also observe conspecific correlations among larger suites of physically unlinked loci, some of which with high or moderate correlations, consistent with polygenic selection in hybrids. Focusing on conspecific loci cor-

relations among suites of five or more loci, we find a much higher frequency of these locus sets with *C. o. concolor*-biased ancestry, corresponding with the genomic background of most hybrid individuals being biased toward *C. o. concolor*, potentially consistent with selection against minor parental ancestry (an expectation of BDMs; Schumer et al. 2014) driving biased inheritance, or perhaps also suggesting that a subset of these associations are the result of neutral admixture. Considering that these correlated locus sets involve variable numbers of individual loci, it remains an open question which of these correlated locus sets may represent selection on incompatibilities, or selection related to other factors (e.g., environment-dependent selection on polygenic traits), and precisely what fraction of these may be driven by neutral admixture LD versus selection. We also note that there were several suites of *C. viridis* conspecific correlated loci involving a high number of loci (up to 18; Fig. 6e). Overall, the differences between parental biases in the numbers of correlated loci and the strength of correlations are striking, given the similar number of excess ancestry loci estimated for each parental lineage was nearly equal. These findings highlight the major differences in patterns, and parent-specific biases in patterns between single-locus processes (e.g., dominating excess ancestry loci inferences) compared to multilocus effects (e.g., correlated allele patterns).

## EVIDENCE OF SELECTION FOR HETEROSPECIFIC ALLELE COMBINATIONS IN HYBRIDS

In contrast to tension zone and related models, hypotheses for other evolutionary scenarios that govern hybrid zone dynamics, including the bounded hybrid superiority and ecotonal models, predict that hybrid fitness is environmentally dependent and reflects local conditions (Bert and Arnold 1995; Rolán-Alvarez et al. 1997). These models suggest that environmental factors may favor genotypes in hybrids that involve novel combinations of alleles from both parental lineages (Birchler et al. 2006; Fitzpatrick and Shaffer 2007). Our results are generally consistent with these predictions for a subset of excess ancestry loci, based on evidence for strong correlations of heterospecific allele combinations that span different chromosomes that appear favored in hybrids (e.g., Fig. S14). These patterns may be partially driven by selection favoring novel allele combinations from both parental species that manifest in high degrees of association in hybrid individuals. We inferred moderate numbers of such heterospecific associations (Fig. 6f); many had nearly perfect allelic correlations ( $r^2 > 0.99$ ) between physically unlinked loci suggesting that multiple suites of novel allelic combinations are favored in the hybrid population (Fig. 6f). Notably, heterospecific correlated loci tend to occur in sets that involve few loci with correlated frequencies, indicating such favored heterospecific combinations involve relatively simple genetic architectures (Fig. S9), in contrast to some conspecific patterns highlighted above that

appear to be highly polygenic (involving many loci on different chromosomes). Because neutral admixture LD is not predicted to result in heterospecific combinations, the observations of these heterospecific correlations in hybrids are strong evidence of selection on hybrids that is more easily disentangled from neutral processes, compared to conspecific allele correlations. These heterospecific correlated allele patterns likely represent selection favoring transgressive phenotypes in hybrids, which may also represent adaptive introgression in at least a subset of these loci. We also note that our sampling included only single geographically adjacent parental populations, which limits the ability to detect adaptively introgressed loci that may have penetrated more deeply into parental populations (Simon et al. 2021). Overall, we expect that a substantial fraction of heterospecific allele combinations observed in hybrids are likely driven by selection due to extrinsic factors; however, a subset may also be due to other processes, including the masking of linked deleterious alleles (e.g., associative overdominance; Ohta 1971; Zhao and Charlesworth 2016) in hybrids.

#### RELATIONSHIPS BETWEEN ALLOPATRIC DIVERGENCE, RECOMBINATION, AND INTROGRESSION

Determining the roles of selection and drift before hybridization may provide insight into processes in allopatric populations that ultimately may shape hybrid zone dynamics (Gompert et al. 2012a; Schield et al. 2017; Baiz et al. 2019). Here, we used measures of differentiation (i.e.,  $F_{ST}$ ) between parental populations to test if loci resistant to introgression are also associated with (1) elevated levels of lineage differentiation in parental populations and (2) increased recombination rates. We found a significant positive correlation, although with weak effect, between the genomic cline center parameter  $\alpha$  and single-locus estimates of parental  $F_{ST}$  (Fig. S7), similar to some previous studies (Gompert et al. 2012a; Schield et al. 2017; Baiz et al. 2019). We further tested this relationship by measuring the average parental  $F_{ST}$  between sets of excess ancestry interchromosomal correlated loci with conspecific alleles. We found that conspecific correlated sets of loci had significantly higher  $F_{ST}$  values, compared with the set of interchromosomal correlated loci identified by *bgc* as neutrally introgressing (Figs. S10, S13). These findings suggest that conspecific correlated loci tend to be more highly differentiated between parental populations, similar to trends observed in other hybrid systems (Schumer et al. 2014).

Additionally, we observe contrasting relationships between recombination rates and measures of parental differentiation across loci categories that are consistent with broad predictions for the interactions among recombination, introgression, and selection in hybrids.  $F_{ST}$  between parental lineages for loci that were inferred to be neutrally introgressing showed a negative re-

lationship with recombination rates (Fig. S12), whereas  $F_{ST}$  for *bgc*-inferred excess ancestry loci showed a positive relationship with a comparatively weaker effect (Fig. S12). Although these results are generally consistent with the prediction that highly differentiated genomic regions in parental lineages also tend to be resistant to introgression in hybrids, high parental population differentiation seems to be a relatively poor indicator of which loci may be under selection in hybrids, further supporting the view that loci under selection and drift in allopatry are likely both important for hybrid fitness (Cutter 2012; Nosil et al. 2012; Cutter and Payseur 2013).

In addition to multilocus processes shaping hybrid genotypes, variation in recombination rate across the genome also shapes patterns of introgression in hybrid populations (Burri 2017; Schumer et al. 2018; Powell et al. 2021), and the genome-wide relationships between recombination and introgression can be quite complex (Calfee et al. 2021). To explore these relationships, we analyzed recombination maps from both *C. viridis* and *C. oreganus* to test for differences in local recombination rates among genomic windows that contained either neutrally introgressing or excess ancestry loci in hybrids genome-wide and broken down by chromosome class. Our results indicate that excess ancestry loci tend to occur in regions of higher recombination genome-wide and across chromosome classes, including recombining regions of the Z chromosome (i.e., PAR), yet the recombination map for only one parental lineage (*C. oreganus*) showed significant correlations between recombination rates of excess ancestry versus neutrally introgressing loci (Figs. 5, S9, S10, S19). Prior evidence that recombination hotspots in these species tend to occur in GC-rich genomic regions (Schield et al. 2020) further highlights relationships among genomic nucleotide content, recombination, and excess ancestry in hybrids. Although our findings are limited by the relatively sparse genomic sampling of our ddRADseq data, they are consistent with the hypothesis that elevated local recombination rates may lead to increased efficiency of natural selection on excess ancestry loci in hybrids. Although similar relationships between loci with biased ancestry in hybrids and recombination have been observed in other systems, our evidence here for single parent-recombination bias contrasts prior studies in which parental lineages had highly similar patterns of recombination (Schumer et al. 2018; Powell et al. 2020). Our results therefore highlight the value of considering both parental lineage contributions in the context of introgression if parental lineages appear to have divergent landscapes.

## Conclusions

Recent studies have accumulated substantial evidence that hybridization in rattlesnakes tends to be fairly ubiquitous, even between relatively distantly related lineages (Glenn and Straight

1990; Meik et al. 2008; Myers 2021), yet most hybrid zones tend to be limited geographically to narrow zones where species' ranges come into contact (Zancolli et al. 2016). This suggests that reproductive isolation between rattlesnake species tends to be incomplete and raises the question of what postzygotic mechanisms may exist that ultimately limit hybridization. Here, we find evidence for genomic patterns of introgression that vary across regions of the genome. These varying patterns support multiple distinct models of hybrid zone dynamics and implicate multiple evolutionary processes that appear to synergistically shape hybrid fitness. We find patterns consistent with tension zone-like models, where hybrid genotypes experience reduced fitness due to predominantly intrinsic multilocus processes resulting in conspecific associations of parental allele combinations across physically unlinked genomic regions. We also find evidence consistent with other models of hybrid zone dynamics (e.g., bounded hybrid superiority) that result in multilocus processes driven by extrinsic factors (potentially including ecological or context-dependent selection, or adaptive introgression for transgressive phenotypes), which result in selection for heterospecific allele combinations in hybrids. We also find that elevated divergence in allopatry (i.e., parental  $F_{ST}$ ) and local recombination rates tend to correlate with excess ancestry loci in hybrids, suggesting the relevance of processes in allopatry and genome structure-function in shaping the genomic architecture of hybrids (Noor and Bennett 2009; Schumer et al. 2018; Powell et al. 2020).

Our findings reinforce the view of hybrid zones as windows into the complex, synergistic, and even antagonistic evolutionary processes that ultimately lead to speciation (Gompert et al. 2017; Moran et al. 2021; Nosil et al. 2021). Although our study is somewhat limited given our lower sampling of hybrid individuals and parental population localities, our findings ultimately underscore the importance of studying hybrid zones with the expectation that multiple processes are likely operating and interacting simultaneously. Evidence for multiple multilocus processes in this system lends support to the broad concept of genomic coupling and genome congealing (Feder et al. 2014; Burri et al. 2015)—or the accumulation of multilocus effects that may collectively represent an increasingly strong reproductive barrier—as a key transition along the path to speciation. Our results here point to this system being at an intermediate stage of this process, in which multiple multilocus effects appear to indeed impact reproductive isolation while not completely preventing hybridization. Interestingly, we find evidence that hybrids experience what appear to be antagonistic genomic interactions that both select for and against suites of admixed genotypes within the hybrid zone. Accordingly, our results provide intriguing evidence that helps explain why this hybrid zone exists yet appears to be confined geographically, due to the interactions of intrinsic and extrinsic processes that select for and against hybrid genotypes within this restricted zone.

## AUTHOR CONTRIBUTIONS

ZLN, DRS, and TAC designed the study. ZLN, DRS, AKW, RHA, and BWP conducted analyses. ZLN, DRS, AKW, BWP, RWO, NRH, JMM, KNI, JP, CFS, SPM, and TAC contributed to data collection. ZLN and TAC wrote the initial draft, and all authors edited subsequent versions.

## ACKNOWLEDGMENTS

We thank S. Spear for tissue samples. Support for this work was provided by National Science Foundation Grant DEB-1655571 to JMM, SPM, and TAC.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA ARCHIVING

All raw data are publicly available on NCBI under the BioProject ID PRJNA548132. Processed variant files and *bgc* formatting scripts are available on Dryad (<https://doi.org/10.5061/dryad.d2547d859>).

## REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., et al. (2013) Hybridization and speciation. *J. Evol. Biol.*, 26, 229–246.
- Ashton, K.G. (2001) Body size variation among mainland populations of the western rattlesnake (*Crotalus viridis*). *Evolution*, 55, 2523–2533.
- Baiz, M.D., Tucker, P.K. & Cortés-Ortiz, L. (2019) Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Mol. Ecol.*, 28, 1056–1069.
- Barton, N.H. (1983) Multilocus clines. *Evolution; international journal of organic evolution*, 37, 454–471.
- Barton, N.H. & Hewitt, G.M. (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, 16, 113–148.
- Barton, N. & Bengtsson, B.O. (1986) The barrier to genetic exchange between hybridising populations. *Heredity*, 57, 357–376.
- Bateson, W. (1909) *Heredity and variation in modern lights*. Darwin and modern science. Cambridge Univ. Press, Cambridge, U.K.
- Bazykin, A.D. (1969) Hypothetical mechanism of speciation. *Evolution; international journal of organic evolution*, 23, 685–687.
- Bert, T.M. & Arnold, W.S. (1995) An empirical test of predictions of two competing models for the maintenance and fate of hybrid zones: both models are supported in a hard-clam hybrid zone. *Evolution; international journal of organic evolution*, 49, 276–289.
- Birchler, J.A., Yao, H. & Chudalayandi, S. (2006) Unraveling the genetic basis of hybrid vigor. *Proceedings of the National Academy of Sciences*, 103, 12957–12958.
- Buerkle, C.A. & Lexer, C. (2008) Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution*, 23, 686–694.
- Burri, R. (2017) Interpreting differentiation landscapes in the light of long-term linked selection: differentiation and long-term linked selection. *Evolution Letters*, 1, 118–131.
- Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., Suh, A., Dutoit, L., Bureš, S. & Garamszegi, L.Z. (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, 25, 1656–1665.
- Butlin, R.K. & Smadja, C.M. (2018) Coupling, reinforcement, and speciation. *American Naturalist*, 191, 155–172.

- Calfee, E., Gates, D., Lorant, A., Perkins, M.T., Coop, G. & Ross-Ibarra, J. (2021) Selective sorting of ancestral introgression in maize and teosinte along an elevational cline. *PLoS Genetics*, 17, e1009810.
- Campbell, J.A., Lamar, W.W. & Brodie, E.D. (2004) The venomous reptiles of the Western Hemisphere. Comstock Pub. Associates, Ithaca, NY.
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140.
- Chhatre, V.E. & Emerson, K.J. (2017) StrAuto: automation and parallelization of STRUCTURE analysis. *BMC bioinformatics*, 18, 192.
- Coyne, J.A. & Orr, H.A. (2004) Speciation. Sinauer Associates, Sunderland, MA.
- Craig, T.P., Itami, J.K. & Craig, J.V. (2007) Host plant genotype influences survival of hybrids between *Eurosta solidaginis* host races. *Evolution; international journal of organic evolution*, 61, 2607–2613.
- Cutter, A.D. (2012) The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends in Ecology & Evolution*, 27, 209–218.
- Cutter, A.D. & Payseur, B.A. (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, 14, 262–274.
- 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., et al. (2011) The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, 27, 2156–2158.
- Dobzhansky, T. (1937) Genetic nature of species differences. *The American Naturalist*, 71, 404–420.
- Endler, J.A. (1973) Gene flow and population differentiation. *Science (New York, N.Y.)*, 179, 243–250.
- . (1977) Geographic variation, speciation and clines. Princeton Univ. Press, Princeton, NJ.
- Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611–2620.
- Falush, D., Stephens, M. & Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Feder, J.L., Nosil, P., Wacholder, A.C., Egan, S.P., Berlocher, S.H. & Flaxman, S.M. (2014) Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *Journal of Heredity*, 105, 810–820.
- Felsenstein, J. (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution; international journal of organic evolution*, 35, 124–138.
- Fitzpatrick, B.M. & Shaffer, H.B. (2007) Hybrid vigor between native and introduced salamanders raises new challenges for conservation. *Proceedings of the National Academy of Sciences*, 104, 15793–15798.
- Flaxman, S.M., Feder, J.L. & Nosil, P. (2013) Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution; international journal of organic evolution*, 67, 2577–2591.
- Francis, R.M. (2017) pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17, 27–32.
- Glenn, J.L. & Straight, R.C. (1990) Venom characteristics as an indicator of hybridization between *Crotalus viridis viridis* and *Crotalus scutulatus scutulatus* in New Mexico *Toxicon: official journal of the International Society on Toxinology*, 28, 857–862.
- Gompert, Z. & Buerkle, C.A. (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, 20, 2111–2127.
- Gompert, Z., Lucas, L.K., Nice, C.C., Fordyce, J.A., Forister, M.L. & Buerkle, C.A. (2012a) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution; international journal of organic evolution*, 66, 2167–2181.
- Gompert, Z., Parchman, T.L. & Buerkle, C.A. (2012b) Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 439–450.
- Gompert, Z., Mandeville, E.G. & Buerkle, C.A. (2017) Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, 48, 207–229.
- Hannon, G.J. (2014) FASTX-Toolkit. Available via [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/).
- Harrison, R.G. & Bogdanowicz, S.M. (1997) Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution; international journal of organic evolution*, 51, 493–505.
- Harrison, R.G. & Larson, E.L. (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105, 795–809.
- Hessenaue, P., Fijarczyk, A., Martin, H., Prunier, J., Charron, G., Chapuis, J., Bernier, L., Tanguay, P., Hamelin, R.C. & Landry, C.R. (2020) Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. *Nature Ecology Evolution*, 4, 626–638.
- Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9, 1322–1332.
- Hvala, J.A., Frayer, M.E. & Payseur, B.A. (2018) Signatures of hybridization and speciation in genomic patterns of ancestry. *Evolution; international journal of organic evolution*, 72, 1540–1552.
- Irwin, D.E. (2018) Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*, 27, 3831–3851.
- Johnson, N.A. (2000) Speciation: Dobzhansky–Muller incompatibilities, dominance and gene interactions. *Trends in Ecology & Evolution*, 15, 480–482.
- Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C.F., Olason, P. & Ellegren, H. (2014) A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology*, 23, 4035–4058.
- Knief, U., Bossu, C.M. & Wolf, J.B. (2020) Extra-pair paternity as a strategy to reduce the costs of heterospecific reproduction? Insights from the crow hybrid zone. *Journal of Evolutionary Biology*, 33, 727–733.
- Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics (Oxford, England)*, 25, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009) the sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)*, 25, 2078–2079.
- Lindtke, D. & Buerkle, C.A. (2015) The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution; international journal of organic evolution*, 69, 1987–2004.
- Lohr, J.N. & Haag, C.R. (2015) Genetic load, inbreeding depression, and hybrid vigor covary with population size: an empirical evaluation of theoretical predictions. *Evolution; international journal of organic evolution*, 69, 3109–3122.
- Mackessy, S.P. (2010) Evolutionary trends in venom composition in the western rattlesnakes (*Crotalus viridis* sensu lato): toxicity vs. tenderizers. *Toxicon: official journal of the International Society on Toxinology*, 55, 1463–1474.
- Meik, J.M., Fontenot, B.E., Franklin, C.J. & King, C. (2008) Apparent natural hybridization between the rattlesnakes *Crotalus atrox* and *C. horridus*. *The Southwestern Naturalist*, 53, 196–200.

- Modahl, C.M. & Mackessy, S.P. (2016) Full-length venom protein cDNA sequences from venom-derived mRNA: exploring compositional variation and adaptive multigene evolution. *PLoS Neglected Tropical Diseases*, 10, e0004587.
- Moore, W.S. (1977) An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology*, 52, 263–277.
- Moran, B.M., Payne, C., Langdon, Q., Powell, D.L., Brandvain, Y. & Schumer, M. (2021) The genomic consequences of hybridization. *eLife*, 10, e69016.
- Muller, H.J. (1942) Isolating mechanisms, evolution and temperature. *Biology Symposium*, 6, 71–125.
- Myers, E.A. (2021) Genome-wide data reveal extensive gene flow during the diversification of the western rattlesnakes (Viperidae: Crotalinae: *Crotalus*). *Molecular Phylogenetics and Evolution*, 165, 107313.
- Noor, M.A. & Bennett, S.M. (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103, 439–444.
- Nosil, P., Parchman, T.L., Feder, J.L. & Gompert, Z. (2012) Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in *Timema* stick insects. *BMC Evolutionary Biology*, 12, 1–12.
- Nosil, P., Feder, J.L. & Gompert, Z. (2021) How many genetic changes create new species? *Science (New York, N.Y.)*, 371, 777–779.
- O'Connor, R.E., Kiazim, L., Skinner, B., Fonseka, G., Joseph, S., Jennings, R., Larkin, D.M. & Griffin, D.K. (2019) Patterns of microchromosome organization remain highly conserved throughout avian evolution. *Chromosoma*, 128, 21–29.
- Ohta, T. (1971) Associative overdominance caused by linked detrimental mutations. *Genetics Research*, 18, 277–286.
- Orr, H.A. (1996) Dobzhansky, bateson, and the genetics of speciation. *Genetics*, 144, 1331–1335.
- Parker, J.M. & Anderson, S.H. (2007) Ecology and behavior of the midget faded rattlesnake (*Crotalus oreganus concolor*) in Wyoming. *Journal of Herpetology*, 41, 41–51.
- Patterson, N., Price, A.L. & Reich, D. (2006) Population structure and eigenanalysis. *PLoS Genetics*, 2, e190.
- Payseur, B.A. & Hoekstra, H.E. (2005) Signatures of reproductive isolation in patterns of single nucleotide diversity across inbred strains of mice. *Genetics*, 171, 1905–1916.
- Perry, B.W., Schield, D.R., Adams, R.H. & Castoe, T.A. (2021) Microchromosomes exhibit distinct features of vertebrate chromosome structure and function with underappreciated ramifications for genome evolution. *Molecular Biology and Evolution*, 38, 904–910.
- Pook, C.E., Wüster, W. & Thorpe, R.S. (2000) Historical biogeography of the western rattlesnake (Serpentes: Viperidae: *Crotalus viridis*), inferred from mitochondrial DNA sequence information. *Molecular Phylogenetics and Evolution*, 15, 269–282.
- Powell, D.L., García-Olazábal, M., Keegan, M., Reilly, P., Du, K., Díaz-Loyo, A.P., Banerjee, S., Blakkan, D., Reich, D. & Andolfatto, P. (2020) Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science (New York, N.Y.)*, 368, 731–736.
- Powell, D.L., Moran, B., Kim, B., Banerjee, S.M., Aguilón, S.M., Fascinetto-Zago, P., Langdon, Q. & Schumer, M. (2021) Two new hybrid populations expand the swordtail hybridization model system. *Evolution; International Journal of Organic Evolution*, 75, 2524–2539.
- Presgraves, D.C., Balagopalan, L., Abmayr, S.M. & Orr, H.A. (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature*, 423, 715–719.
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- R Core Team (2021) R: a language and environment for statistical computing (R version 4.0.3). R Foundation for Statistical Computing, Vienna.
- Rolán-Alvarez, E., Johannesson, K. & Erlandsson, J. (1997) The maintenance of a cline in the marine snail *Littorina saxatilis*: the role of home site advantage and hybrid fitness. *Evolution; International Journal of Organic Evolution*, 51, 1838–1847.
- Rundle, H.D. & Nosil, P. (2005) Ecological speciation. *Ecology Letters*, 8, 336–352.
- Saviola, A.J., Pla, D., Sanz, L., Castoe, T.A., Calvete, J.J. & Mackessy, S.P. (2015) Comparative venomomics of the Prairie Rattlesnake (*Crotalus viridis viridis*) from Colorado: identification of a novel pattern of ontogenetic changes in venom composition and assessment of the immunoreactivity of the commercial antivenom CroFab®. *Journal of Proteomics*, 121, 28–43.
- Schild, D.R., Adams, R.H., Card, D.C., Perry, B.W., Pasquesi, G.M., Jezkova, T., Portik, D.M., Andrew, A.L., Spencer, C.L. & Sanchez, E.E. (2017) Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. *Ecology and Evolution*, 7, 3951–3966.
- Schild, D.R., Card, D.C., Hales, N.R., Perry, B.W., Pasquesi, G.M., Blackmon, H., Adams, R.H., Corbin, A.B., Smith, C.F. & Ramesh, B. (2019a) The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Research*, 29, 590–601.
- Schild, D.R., Perry, B.W., Adams, R.H., Card, D.C., Jezkova, T., Pasquesi, G.I.M., Nikolakis, Z.L., Row, K., Meik, J.M. & Smith, C.F. (2019b) Allopatric divergence and secondary contact with gene flow: a recurring theme in rattlesnake speciation. *Biological Journal of the Linnean Society*, 128, 149–169.
- Schild, D.R., Pasquesi, G.I.M., Perry, B.W., Adams, R.H., Nikolakis, Z.L., Westfall, A.K., Orton, R.W., Meik, J.M., Mackessy, S.P. & Castoe, T.A. (2020) Snake recombination landscapes are concentrated in functional regions despite PRDM9. *Molecular Biology and Evolution*, 37, 1272–1294.
- Schilling, M.P., Mullen, S.P., Kronforst, M., Safran, R.J., Nosil, P., Feder, J.L., Gompert, Z. & Flaxman, S.M. (2018) Transitions from single- to multi-locus processes during speciation with gene flow. *Genes*, 9, 274.
- Schumer, M., Cui, R., Powell, D.L., Dresner, R., Rosenthal, G.G. & Andolfatto, P. (2014) High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *eLife*, 3, e02535.
- Schumer, M., Cui, R., Rosenthal, G.G. & Andolfatto, P. (2015) Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS genetics*, 11, e1005041.
- Schumer, M., Xu, C., Powell, D.L., Durvasula, A., Skov, L., Holland, C., Blazier, J.C., Sankararaman, S., Andolfatto, P. & Rosenthal, G.G. (2018) Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science (New York, N.Y.)*, 360, 656–660.
- Seehausen, O., van Alphen, J.J.M. & Witte, F. (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science (New York, N.Y.)*, 277, 1808–1811.
- Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.-P., Bank, C., Brännström, Å., et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics*, 15, 176–192.
- Selz, O.M. & Seehausen, O. (2019) Interspecific hybridization can generate functional novelty in cichlid fish. *Proceedings of the Royal Society B*, 286, 20191621.
- Simon, A., Fraïsse, C., El Ayari, T., Liautard-Haag, C., Strelkov, P., Welch, J.J. & Bierne, N. (2021) How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology*, 34, 208–223.

- Smith, C.F. & Mackessy, S.P. (2016) The effects of hybridization on divergent venom phenotypes: characterization of venom from *Crotalus scutulatus scutulatus* × *Crotalus oreganus helleri* hybrids. *Toxicon : Official Journal of the International Society on Toxinology*, 120, 110–123.
- Stelkens, R.B., Schmid, C., Selz, O. & Seehausen, O. (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evolutionary Biology*, 9, 283.
- Taylor, S.A. & Larson, E.L. (2019) Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, 3, 170–177.
- Wang, L., Luzynski, K., Pool, J.E., Janoušek, V., Dufková, P., Vyskočilová, M.M., Teeter, K.C., Nachman, M.W., Munclinger, P. & Macholán, M. (2011) Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Molecular Ecology*, 20, 2985–3000.
- Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution; International Journal of Organic Evolution*, 38, 1358–1370.
- Zancolli, G., Baker, T.G., Barlow, A., Bradley, R.K., Calvete, J.J., Carter, K.C., De Jager, K., Owens, J.B., Price, J.F. & Sanz, L. (2016) Is hybridization a source of adaptive venom variation in rattlesnakes? A test, using a *Crotalus scutulatus* × *viridis* hybrid zone in southwestern New Mexico. *Toxins*, 8, 188.
- Zhao, L. & Charlesworth, B. (2016) Resolving the conflict between associative overdominance and background selection. *Genetics*, 203, 1315–1334.

Associate Editor: W. Haerty  
Handling Editor: T. Chapman

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1.** Evolutionary relationships of parental taxa.

**Figure S2.** SNP quality score distribution.

**Figure S3.** Ancestry coefficients for all sampled individuals under a  $K=3$  model estimated in STRUCTURE.

**Figure S4.** Hybrid index distribution.

**Figure S5.** Population assignment and clustering for an 80% complete data matrix.

**Figure S6.** Simulation of Bayesian genomic cline analysis.

**Figure S7.** Bayesian genomic cline run convergence.

**Figure S8.** Relationship between excess ancestry loci and chromosome length.

**Figure S9.** Patterns of recombination between excess ancestry and neutrally introgressing loci across chromosome classes and parental recombination maps.

**Figure S10.** Relationship between excess ancestry categories, parental recombination maps, and chromosome classes.

**Figure S11.** Relationship between genomic cline center parameter and parental differentiation.

**Figure S12.** Relationship between recombination rates and  $F_{st}$  for each parental recombination map and loci categories.

**Figure S13.** Histogram of  $r^2$  values for all pairwise interchromosomal connections.

**Figure S14.** Genomic distribution of heterospecific interchromosomal connections.

**Figure S15.** Comparison of excess ancestry and neutral  $r^2$  distributions for pairs of interchromosomal loci in hybrids.

**Figure S16.** Distribution of  $r^2$  values for interchromosomal allele correlations in hybrids.

**Figure S17.** Distribution of  $r^2$  values across lineage and chromosomal comparison categories.

**Figure S18.** Relationship between near perfect  $r^2$  intervals and parental differentiation.

**Figure S19.** Recombination patterns across categories of interchromosomal comparisons.

**Table S1.** Sample data for group membership, project ID, locality, NCBI accession, reference study, and number of mapped reads. Group membership was classified based on ancestry coefficients from STRUCTURE runs, ordination of genetic variance, and geographic locality.

**Table S2.** Summary of fixed SNPs and excess ancestry loci across individual chromosomes and Z specific regions

**Table S3.** Summary measures of 10kb windowed recombination maps for neutral and outlier windows from each parental lineage recombination landscape.

**Appendix S1** – supplementary methods