www.sciencemag.org/cgi/content/full/science.aar3684/DC1

Supplementary Materials for

Natural selection interacts with recombination to shape the evolution of hybrid genomes

Molly Schumer,* Chenling Xu, Daniel L. Powell, Arun Durvasula, Laurits Skov, Chris Holland, John C. Blazier, Sriram Sankararaman, Peter Andolfatto,† Gil G. Rosenthal,† Molly Przeworski*

*Corresponding author. Email: schumer@fas.harvard.edu (M.S.); mp3284@columbia.edu (M.P.)
†These authors contributed equally to this work.

Published 19 April 2018 on *Science* First Release
DOI: 10.1126/science.aar3684

This PDF file includes:

Natural selection interacts with recombination to shape the evolution of hybrid genomes

Supplementary Online Material

Molly Schumer, Chenling Xu, Daniel L. Powell, Arun Durvasula, Laurits Skov, Chris Holland, John C. Blazier, Sriram Sankararaman, Peter Andolfatto, Gil G. Rosenthal, Molly Przeworski

Contents

1. Materials and Methods

1.1 Ancestry assignment in *X. birchmanni x X. malinche* hybrids

To understand the relationship between minor parent ancestry and recombination in swordtail hybrids, we generated fine-scale estimates of ancestry throughout the genomes of natural hybrids from three different hybrid populations (Fig. 1E).

1.1.1 Sample collection and sequencing of *X. birchmanni x X. malinche* hybrids

DNA was extracted from 276 individuals collected from the Totonicapa hybrid population in 2013, 2014, and 2015; 193 individuals collected from the Tlatemaco hybrid populations in 2012, 2013 and 2015; and 243 individuals collected from the Aguazarca hybrid population in 2010, 2013, and 2015. Libraries for these individuals were prepared following (*30*). Briefly, three to ten nanograms of DNA was mixed with Tn5 transposase enzyme pre-charged with custom adapters and incubated at 55 °C for 15 minutes. The reaction was stopped by adding 0.2% SDS and incubating at 55 °C for an additional seven minutes. One of 96 custom indices were added to each sample in a plate with an individual PCR reaction including 1 μl of the tagmented DNA; between 13-16 PCR cycles were used. After amplification, 5 μl of each reaction was pooled and purified using Agencourt AMPpure XP beads. Library size distribution and quality was visualized on the Bioanalyzer 1000 (Agilent, Santa Clara, California) and size selected by Princeton's Lewis-Sigler Institute Genomics Core Facility to be between 350-750 base pairs (bp). Libraries were sequenced on the Illumina HiSeq 4000 at Weill Cornell Medical Center across three lanes to collect paired-end 100 nucleotide reads.

1.1.2 Ancestry assignment in hybrids

Ancestry assignment in hybrids was performed using the Multiplexed Shotgun Genotyping ("MSG") pipeline (*24*). We previously performed extensive simulations and sequencing to demonstrate that MSG is predicted to perform well in inferring local ancestry for the parameters of the hybrid populations on which we focused here (*22, 31*). However, in this study, we collected substantially more whole genome data at deeper coverage (see Materials and Methods 1.2) and thus detected many new polymorphisms in *X. birchmanni.* Because these sites can cause errors in MSG if they are shared between species and are not masked before ancestry inference (see next section), we masked polymorphic sites identified in the 25 *X. birchmanni* genomes sequenced here from the reference genomes input into MSG, as well as sites detected in two newly sequenced *X. malinche* genomes (SRP130891;SRP018918). Polymorphic sites were detected using the GATK pipeline (see Materials and Methods 1.2) and masked in the reference genomes using the mutfa function of the program seqtk.

For ancestry inference in hybrids, raw data was parsed by barcode and trimmed to remove low-quality bps (Phred quality score <20). Reads with fewer than 30 nucleotides after trimming were discarded. Because of prohibitively long computational times, reads from individuals with more than 16 million reads were subsampled to 16 million before running the MSG pipeline. The minimum number of reads for an individual to be included was set to 300,000, since ancestry inference with fewer reads is predicted to have lower accuracy based on simulations (*31*). This procedure resulted in the inclusion

of 690 individuals in our final analysis (187-266 individuals per population), with an average coverage of 7.6 million reads or ~1X coverage genome-wide.

The parameters used in the MSG run were based on previous work on these hybrid populations (*22, 32*). Data from each population were run separately with population-specific priors. Namely, the expected number of recombination events per chromosome (recRate) was set to 8-17 based on a prior expectation of approximately 35-60 generations of admixture and assuming initial admixture proportions of 75% of the genome derived from one parent and 25% derived from other parent (with the major parent varying by population). Similarly, priors for each ancestry state were set using previously estimated mixture proportions for each population, assuming Hardy-Weinberg equilibrium (*32*). The recombination rate scaling factor (rfac) was set to the default value of 1.

MSG outputs ancestry in the form of posterior probabilities. We converted these to hard ancestry calls, requiring a posterior probability of 0.95 or greater to assign a site to a given ancestry state. This threshold was used because past work suggested that posterior probabilities may be poorly calibrated, and we wanted our call set to be conservative. Sites with lower than 0.95 posterior probability were masked, as were sites that were covered in fewer than 25% of individuals in a given population sample. For each population, we then summarized minor parent ancestry by averaging the ancestry calls at each site, for a range of window sizes (e.g. 5 kb – 1 Mb).

In summarizing ancestry across the genome, we identified two regions with unusually low variance in ancestry, on chromosomes 17 and 24. Investigating these signals further, we found that they had high admixture LD and few detectable crossovers, leading us to conclude that these are most likely fixed inversions between species (Fig. S10). We excluded the two chromosomes containing these putative inversions from all subsequent ancestry analyses.

1.1.3 Allele sharing between *X. birchmanni* and *X. malinche* and its impact on ancestry assignment

Shared polymorphisms between species can cause errors in ancestry assignment if they are incorrectly included as ancestry informative sites (see previous section). *A priori*, we expected few shared polymorphisms due to incomplete lineage sorting because the coalescence times in *X. malinche* should be much shorter than the divergence to *X. birchmanni* (Fig. 1F; Materials and Methods 1.3). However, given the larger coalescence times of *X. birchmanni*, *X. malinche* diversity could still fall within the diversity of *X. birchmanni*, i.e., the genealogies of the *X. birchmanni* and *X. malinche* lineages will not necessarily be reciprocally monophyletic.

Examining the genome sequences we had collected (see previous section), we found that the number of shared polymorphisms between the two species was higher than expected from a model with no gene flow: considering 20 unrelated *X. birchmanni* genomes and two *X. malinche* genomes, 15% of 5 kb windows contained SNPs shared between the two species. We hypothesized that the shared SNPs were due to gene flow, either between the two species since they split or from other species in the *Xiphophorus* group (*3*). To explore this possibility, we ran *TreeMix* (*33*), including other available *Xiphophorus* genome sequences (*7, 34*). Results were sensitive to the choice of parameters (notably to k, the number of SNPs per window) and the species included; we

therefore considered a number of implementations (Fig. S11). Importantly, in no implementation was gene flow inferred directly between *X. birchmanni* and *X. malinche* since their split; however, both species appear to have been the recipients of gene flow from closely related outgroup species, events that are presumably the source of the shared SNPs.

As noted in the previous section, when inferring local ancestry in hybrids, we had excluded all shared SNPs between *X. malinche* and *X. birchmanni* (see previous section). Simulations suggest that ancestry inference with MSG is relatively robust to undetected shared polymorphism (*31*); accordingly, the method correctly assigns parental individuals as homozygous for ancestry throughout their genomes (*22, 32*). In light of the discovery of gene flow from other species, in addition to excluding the shared SNPs themselves, we reanalyzed the data excluding any 10 kb windows with shared SNPs (almost all of which also contained fixed differences between samples). Results were unchanged (see Materials and Methods 1.4).

1.2 Generating a linkage disequilibrium-based map for *X. birchmanni*

To infer fine-scale estimates of recombination rates throughout the genome, we generated a linkage disequilibrium (LD)-based recombination map for *X. birchmanni*. To this end, we generated high coverage sequence data for a *X. birchmanni* population sample. Eighteen unrelated individuals were collected from the Coacuilco *X. birchmanni* population in Hidalgo, Mexico in 2010 using baited minnow traps. Individuals were stored in 95% ethanol. DNA was extracted from the liver using the Qiagen DNeasy kit (Qiagen, Valencia, CA), following the manufacturer's instructions. DNA was quantified on a Qubit fluorometer (Thermo Scientific, Wilmington, DE) and assessed for quality on a Nanodrop 1000 (Thermo Scientific, Wilmington, DE). DNA was sheared to approximately 400 base pair (bp) fragments using a Covaris LE220 sonicator (Covaris, Woburn, MA). Fragmented DNA was prepared for sequencing following the protocol of Quail et al. (*35*). Sheared DNA was end-repaired and A-tailed; custom adapters were ligated and libraries were size selected (400-600 bp) on a 1% agarose gel. Purified fragments were amplified using the Phusion high fidelity PCR kit (NEB, Ipswich, MA) for 10-11 cycles and purified with Agencourt AMPpure XP beads (Beckman Coulter, Brea, California). Libraries were evaluated for size distribution and quality on a Bioanalyzer 2100 (Agilent, Santa Clara, California). Sequencing was performed with Illumina HiSeq 4000 chemistry at Weill Cornell Medical Center across six lanes to collect paired-end 100 reads. Average coverage per individual after alignment is reported in Table S1.

DNA was also extracted from a Coacuilco population male, a female, and five of their lab-generated offspring, using the Qiagen DNeasy kit (Qiagen, Valencia, CA). DNA was sent to the New York Genome Center for PCR-free library preparation. Libraries were sequenced with Illumina HiSeq 2500 chemistry to collect paired-end 150 bp reads. All sequence data are available through NCBI's SRA (SRP130891;SRP018918). Average coverage per individual for this pedigree after alignment is also reported in Table S1.

1.2.1 SNP calling

Raw reads were parsed by index, using a custom python script (https://github.com/JaneliaSciComp/msg/blob/master/barcode_splitter.py). Reads were trimmed to remove adapter sequences using the program cutadapt v1.9 (*36*) and mapped to the outgroup *X. maculatus* reference genome (1.5% diverged; *34, 37*) using bwa, with the *bwa-mem* algorithm (*38*). Mapping statistics are provided in Table S1. GATK (v3.4; *39*) was used to convert sam files to bam files and picard tools (v1.118) was used to mark and remove duplicates for those libraries that underwent PCR amplification (this step was omitted for PCR-free libraries). Insertion-deletion differences (indels) were realigned using GATK. Variant calling was performed using GATK with the HaplotypeCaller algorithm in the GVCF mode. We lacked access to a high-quality variant set for variant recalibration, so did not perform variant recalibration with GATK.

Instead, we applied hard call thresholds recommended by GATK (https://software.broadinstitute.org/gatk/documentation/article.php?id=3225) and additionally masked a 5 bp window around indels and any sites with greater than or less than 2X average genome-wide coverage. Previous simulations evaluating the approach with parameters matching our data suggested that it should be reliable (*7*). Moreover, since we collected high quality sequence data for two parents and five offspring from a family, we were able to quantify Mendelian error rates in this family using plink (v1.07; *40*) and we therefore used them to evaluate the reliability of a hard call approach to identify variants.

When applying GATK recommended hard calls, we found that our estimated Mendelian error rate was fairly low: 0.54% of SNPs had an error in one or more of five offspring. Using this information, we further attempted to improve our hard call filtering. To this end, we examined the distribution of scores for various quality metrics in the family data for calls that were Mendelian errors versus for scores at the same position in another offspring in the family that did not have a Mendelian error. We found that two quality metrics had significantly different distributions (Fig. S12); based on these results, we re-adjusted the required QD score to ≥10 and the required FS score to ≤10. Our error rate with this more stringent filtering is therefore likely lower than 0.54%.

Guided by these analyses, for the population sample of *X. birchmanni*, we removed Mendelian errors identified in the *X. birchmanni* pedigree and masked sites with greater than two-fold the genome-wide average coverage or less than half the genome-wide average coverage (*41*), all indels, all repetitive regions, and any sites within a 5 bp window of an indel. In addition, we excluded sites with coverage less than 10 (DP); low quality score (variant quality - GQ or invariant quality - RGQ <20) or quality by depth score (QD<10); low mapping quality (MQ < 40); high Fisher strand score (FS>10); high strand odds ratio (SOR > 4); low read position rank sum score (ReadPosRankSum< -8); and low mapping quality rank sum score (MQRankSum < -12.5). Note that for invariant sites, only RGQ and DP filters can be used. This filtered dataset was used in subsequent steps.

1.2.2 Inference of ancestral sequences and mutation matrix

LDhelmet, the program that we used to build a linkage disequilibrium (LD) based recombination map, requires a quadra-allelic mutation model and thus relies on a 4x4 mutation transition matrix (*42*). To infer this matrix, we needed to generate ancestral

sequences for *X. birchmanni*. We tried two methods to infer ancestral sequences, relying on the root state probabilities output by RAxML (v7.2.8; *43*) and those produced by Phylofit (v1.3; *44*). To test the performance of these two approaches, we used macs (*45*) to simulate sequences with phylogenetic relationships corresponding to those for the available swordtail genomes (Fig. S13). Split times were estimated from the relationship $T_{div(4N)}=0.5(D_{xy}/\theta - 1)$, where θ is the population mutation rate and D_{xy} is the average pairwise divergence between species. We assumed $\theta = 0.001$ per site for all species, which is approximately the estimated θ in most swordtail species studied to date (*7, 46*). We simulated 1,000 one Mb sequences and input the macs output into seq-gen (*47*) to generate nucleotide sequences using the observed base composition in swordtails. We then inferred the swordtail root sequence with both approaches (*3*). We compared the accuracy of the inferred ancestral sequence to the true simulated ancestral sequence, only considering bases at which an allele was assigned a probability ≥0.99 of being ancestral. Both programs significantly outperformed a strict parsimony approach (Fig. S14) and had comparable performance on the parameters tested.

For the real data, we inferred the ancestral sequence using Phylofit and the available whole genome sequences for swordtails (Fig. S13). Because Phylofit does not accommodate polymorphic sites, we used a base by base coin flip to pick an allele at known polymorphic bases. To generate the mutation matrix for LDhelmet, we counted the number of mutations of each possible class in the population data that we had collected for *X. birchmanni*, given the inferred ancestral sequence (following *42*). We then calculated mutation frequencies and converted the mutation frequencies to a mutation matrix.

1.2.3 Evaluating the reliability of the approach taken to estimate recombination rates in *X. birchmanni*

To evaluate how reliable we should expect an LD map for *X. birchmanni* built with LDhelmet to be, and to inform our choice of parameters, we performed simulations. Because of the computational resources required by these simulations, we ran only 15 replicates.

Specifically, we simulated ten Mb chromosomes from one population for a sample of 40 individuals using macs with $\theta = 0.001$ and a background population recombination rate of $\rho = 0.0006$ per bp (i.e., the median ρ per bp estimated from the data; see below). We placed four hotspots per Mb, with heat drawn from a random uniform distribution of heat from 10x background to 50x the background. In order to incorporate the ancestral allele inference step into our simulations, we inferred ancestral sequences as described above, and converted macs output into sequences using the seq-gen program. Next, we selected four haplotypes from the simulated 40 *X. birchmanni* sequences, two to serve as the maternal and two as the paternal chromosomes. We then used custom scripts to generate five "offspring" between these parental haplotypes to mimic the pedigree that we had generated. For each offspring, we randomly chose a genetic distance; converted it into a physical location on the genome of each parent; introduced a crossover between the haplotypes at this location and generated diploid offspring with one recombinant paternal and maternal chromosome (ignoring in particular the negligible number of *de novo* mutations that are expected). We collapsed pairs of haplotypes into diploid genotypes and used these data in subsequent steps.

LDhelmet requires phased haplotypes. Phased haplotypes can increase the resolution and reliability of LD-based recombination maps, but errors introduced by incorrect phasing can potentially result in spurious inferences of recombination. As a result, we also evaluated this tradeoff in our simulations. Specifically, using the simulated data, we compared how well phasing performed using one of two programs, shapeit2 (*48*) or impute2 (*49*). In order to apply shapeit2, we converted the sequence files to ped format, indicating known family relationships, and then ran shapeit2 with the duohmm flag (*48*). In turn, to run impute2, we converted the sequence files to ped format and then used hapi to generate haplotypes for the family (*48-50*). We used these haplotypes as input for impute2 (*49*), treating them as known without error, then inferring haplotypes for the population data. With these two sets of phased haplotypes from the population data and the ancestral sequence and mutation matrix obtained from the same simulation (see previous section), we were now ready to run LDhelmet.

One key parameter in LDhelmet is the block penalty, which is the penalty imposed for switching recombination rates. A higher block penalty will result in fewer rate switches and a smoother recombination landscape, whereas a lower block penalty will result in more rate switches and a less smooth inferred rate landscape. The LDhelmet documentation recommends using a block penalty of 5 for humans and a block penalty of 50 for *Drosophila melanogaster*, but in general the optimal choice depends on the true and unknown heterogeneity in recombination rates. Since swordtails appear to have hotspots based on previous work (*25*), a sensible guess is that 5 would be the appropriate block penalty. We tested this notion in simulations by running LDhelmet on the simulated sequences described in the previous paragraph for a range of block penalties (5, 20 and 50).

To evaluate performance in these simulations, we considered LDhelmet inferred rates in 50 kb windows and compared them to the true rates in 50 kb windows. Overall, the inferred LD map was strongly correlated with the true map in simulations (Fig. S15). Both in terms of estimated rates in 50 kb windows and estimated hotspot heats, LDhelmet performed better in simulations with a block penalty of 5 or 20 than a block penalty of 50 (Fig. S15). On that basis, we used a block penalty of 5 in our analysis. For phasing pipelines, we found that with both phasing approaches, the average correlation between the true and simulated map for all simulations was 0.66. In the real data, we chose to use the simpler, one step pipeline of shapeit2.

Since LDhelmet requires phased haplotypes, the above simulations allowed us to evaluate the performance of LDhelmet with different phasing pipelines, but not whether phasing improves the reliability of rate inference compared to using unphased data. For a direct comparison, we used the program LDhat, which (in contrast to LDhelmet) can generate LD maps using both genotypes and haplotypes (*51*). In data analyses however, we used LDhelmet, because it was found to outperform LDhat (*42*). In simulations with shapeit2 phasing, the correlations between the true and inferred maps were higher when using inferred haplotypes rather than genotypes (Fig. S16), suggesting a net gain in map reliability despite phasing errors (assuming LDhelmet has similar properties to LDhat in this regard, as seems sensible). We therefore used inferred haplotypes when analyzing the real data.

1.2.4 An LD-based recombination map for *X. birchmanni*

Given the approaches that worked better in simulations, we used phased haplotypes from all unrelated individuals (in total, 40 haplotypes) to construct an LD-based recombination map with the program LDhelmet (*42*). We generated ancestral sequences using several previously sequenced swordtail genomes (Fig. S13), the program Phylofit (*44*) and inferred a mutation matrix following Chan *et al. (42)*, as described above. We computed a likelihood lookup table for a grid of ρ values for each set of haplotypes for each chromosome. To infer recombination rates, we used the rjMCMC procedure with a block penalty of 5 and a burn-in of 100,000 and ran the Markov chain for 1,000,000 iterations. After excluding SNPs for which recombination rate was estimated to be implausibly high (ρ/bp ≥ 0.4), we summarized recombination rates in 5, 50, 1000 and 5000 kb windows. For 5 and 50 kb windows, we removed windows that overlapped with a contig boundary, as rate estimates may be inaccurate for adjacent SNPs across a contig boundary. We proceeded with this LD-based map for subsequent analyses.

1.3 Demographic history of *X. birchmanni* and *X. malinche* and its consequences for load

Understanding differences in the demographic history of both *X. birchmanni* and *X. malinche* is important in evaluating the nature of selection on hybrids (Materials and Methods 1.5; *6, 11*). We therefore used multiple approaches to better understand their demographic history.

1.3.1 Estimation of nucleotide diversity and the mutation rate

We estimated average pairwise nucleotide diversity (π) to be 0.0012 per bp for *X. birchmanni* and 0.0003 per bp for *X. malinche* using biallelic sites (almost all polymorphic positions in our samples) in data from 20 unrelated *X. birchmanni* individuals from a single population and two *X. malinche* individuals from a single population. To understand if low levels of genetic diversity in *X. malinche* are a general feature of this species, we analyzed an individual from an *X. malinche* population in an independent river system, and found comparably low levels of genome-wide π (an average of 0.00036 per bp). Recent work in cichlid fish, which are ~100 million years diverged from swordtails reported mutation rates for three species of around 3.5×10^{-9} per bp per generation, similar to estimates from another fish species (*52*). We used this estimate in our analyses.

Since *X. birchmanni* and *X. malinche* differ markedly in levels of genetic diversity, we asked whether there was evidence that the mutation rate differs between species using the relative rates test (*53*), assuming that the two species have the same generation time. We computed the proportion of sites that differed between the *X. birchmanni* sequence and *X. maculatus* sequence and *X. malinche* sequence and *X. maculatus* sequence in 1 Mb windows. For polymorphic sites, we randomly selected one of the alleles. We then calculated the relative branch lengths of *X. malinche* and *X. birchmanni* in terms of pairwise divergence:

$l_A = d_{Xmal\_Xbir} + d_{Xmal\_Xmac} - d_{Xbir\_Xmac}$

$l_B = d_{Xmal\_Xbir} + d_{Xbir\_Xmac} - d_{Xnal\_Xmac}$

Given the above assumptions, we expect that $l_A - l_B$ will be zero if mutation rates are the same. Indeed, the branch length for *X. birchmanni* was only 1.007x the branch length for *X. malinche*. This finding suggests that the mutation rates in these two lineages do not differ appreciably and that differences in π reflect long-term differences in effective population sizes.

Assuming a standard neutral model of a constant population size, and relying on a mutation rate of 3.5×10^{-9} per bp per generation, we would therefore estimate that the effective population size of *X. malinche*, $N_{malinche}$, is approximately 21,000, and the effective population size of *X. birchmanni*, $N_{birchmanni}$, is approximately 86,000.

1.3.2 Demographic inference in *X. birchmanni* and *X. malinche* and possible effects on estimates of the population recombination rate

To investigate the demographic history of *X. birchmanni* and *X. malinche*, we ran PSMC (*54*) on two (unphased) diploid individuals of *X. malinche* from separate river populations (*22*; this study - SRP018918) and on all 20 unrelated *X. birchmanni* genomes collected for this study from a single population. We set the –r parameter, the ratio of θ to ρ, to 2; the –g parameter, the number of years per generation, to 0.5 (*55*); and the mutation rate to 3.5×10^{-9} per bp per generation (see previous section). We present results truncated at approximately $2N_{birchmanni}$ generations or 80,000 years, but note that bootstrapping results suggest that for *X. malinche*, population size estimates for times beyond 10^4 years ago may not be reliable. Based on the bootstrapping results, we also do not consider timepoints more recent than 1 Kya (Fig. 1F). Results are roughly consistent with those obtained from diversity levels under a constant population size model, in that the harmonic means of the PSMC estimated population sizes are similar to $N_{malinche}$ and $N_{birchmanni}$.

The PSMC results further confirmed our expectation, based on levels of genetic diversity, that *X. malinche* has had much lower historical effective population sizes than *X. birchmanni*, approximately four-fold lower for the past twenty thousand generations (Fig. 1F). This finding suggests that historical population sizes in *X. malinche* could have allowed for the accumulation and fixation of weakly deleterious mutations that would have been effectively purged from the larger *X. birchmanni* population. We therefore tested this possibility, as described below in *Evidence for higher load in X. malinche*.

We also used the PSMC results to ask whether the demographic history of *X. birchmanni* is likely to have a major impact on the accuracy of recombination rate estimates generated by LDhelmet, which were inferred under the assumption of a constant size, panmictic model. In species where this demographic model is clearly violated, such as humans, this type of misspecification has been reported to affect the reliability of estimates (*56*). For simplicity, in simulations we used step-wise changes in population size with the –en option in macs (Fig. S17). We chose a value of θ for simulations that yields roughly the observed present-day diversity levels (Fig. S18). To specify the recombination map, we used the observed LDhelmet recombination rates for the first 10 Mb of chromosome 1. In other respects, simulations followed the pipeline

described above (see *Evaluating the reliability of the approach taken to estimate recombination rates in X. birchmanni*). When we simulated data under the inferred demographic history for *X. birchmanni* but estimated recombination rates under a model of constant population size, there was little impact on the reliability of the map at the scale of 50 kb. Specifically, in the 10 simulations we ran, the Spearman's correlations between the simulated and true maps were no worse than those observed in 10 simulations of a constant population size (0.74-0.81 vs 0.73-0.8, respectively). We therefore considered the estimates obtained under a constant population size model in our analyses.

1.3.3 Evidence for higher load in *X. malinche*

Smaller historical population sizes can lead to the fixation of deleterious mutations that would be efficiently purged in larger populations. Notably, results of the PSMC analysis highlight that population sizes have been low in *X. malinche* for approximately the last $N_{malinche}$ generations, the relevant timescale for such mutations. We wanted to examine if there was indeed evidence for higher load on the *X. malinche* lineage, as expected given four-fold differences in total diversity levels. We therefore used the approach of Do et al. (*26, 57*) to ask whether a single diploid *X. malinche* genome shows an excess of derived non-synonymous sites relative to a single diploid *X. birchmanni* genome ($R_{birchmanni/malinche}$), compared to the inferred ancestral northern swordtail sequence (see Materials and Methods 1.2). Based on this measure, we estimated $R_{birchmanni/malinche}$ to be 0.975±0.004, with the standard error estimated from jack-knife bootstrapping of the variant data divided into 100 contiguous blocks. The ratio is significantly different from 1 ($p = 0.016$, estimated by bootstrapping), indicating an excess of non-synonymous substitutions in *X. malinche* relative to *X. birchmanni*. Given that selection is less effective in regions of low recombination, we might expect an even larger difference in regions of the genome with low recombination.

1.4 Evaluating the relationship between minor parent ancestry and recombination rate in three swordtail hybrid populations

1.4.1 Minor parent ancestry and recombination

Having inferred local recombination rates and ancestry throughout the genome, we set out to understand their relationship in each of the three swordtail hybrid populations. Selection on hybrids is predicted to generate correlations between minor parent ancestry and recombination rates (Fig. 1; Materials and Methods 1.5). The appropriate size scale for this analysis depends on a number of parameters, notably the number of generations over which selection has operated since hybridization. To guide our analysis, we performed admix'em (*58*) simulations of hybrid incompatibilities, as described in detail in Materials and Methods 1.5 (simulations with 4 BDMIs, s=0.1, h=0.5), and examined how ancestry changed over time. Though a decrease in average minor parent ancestry near hybrid incompatibilities was observed early on in simulations (e.g., after 10 generations, when ancestry blocks are Mbs in length), it became more pronounced and localized over time (Fig. S19). Motivated by these simulations, and the observation that that the median length of ancestry tracts that are homozygous for the minor parent in the

swordtail hybrid populations is 84-225 kb (Materials and Methods 1.5, 1.7), we considered relationships between local recombination rates and ancestry over the scale of 100s of kb in the empirical data (Table S2).

At some of the scales considered, nearby windows are likely not independent, due to LD. As an illustration, the expected correlation in allele frequencies, $E(r^2)$, falls to approximately 0.1 within 50 kb in *X. birchmanni*. To address this issue, we thinned analyses in 50 and 250 kb windows such that we sampled approximately one window every 250 kb, a distance over which there is little or no significant pairwise LD (i.e., we picked one window of every six in the analysis of 50 kb windows and one window of every two in the analysis of 250 kb windows). When we apply this thinning in neutral simulations (see Materials and Methods 1.5), ~5% of simulations show a significant relationship (at $\alpha = 0.05$) between minor parent ancestry and recombination rate, suggesting that this thinning results in properly calibrated p-values.

We identified significant positive correlations between minor parent ancestry and the recombination rate in all three hybrid swordtail populations, regardless of our choice of window size (main text; Table S2). However, recombination rates are also correlated with other features in the genome that may influence selection on hybrids, including the number of coding and conserved bps. To investigate these effects, we examined minor parent ancestry in a focal 50 kb window as a function of the number of coding or conserved bps within 0.1 cM on either side of the window (which seems an appropriate scale based on the data; Fig. S20), using the reported *Xiphophorus* chromosome lengths in cM from a previous pedigree study (*37*). Using this approach, we found a significant negative relationship between minor parent ancestry and the number of linked coding or conserved bps in all populations (Table S5). However, these relationships became weaker and in some cases non-significant when the data were thinned as described in the previous paragraph.

To consider minor parent ancestry, recombination, and putatively functional bps jointly, we calculated the partial correlation between local ancestry and the recombination rate, controlling for the number of coding (or conserved) bp in a window of a given physical size, using the "ppcor" package in R (considering 50 kb, 250 kb and 500 kb windows). Regardless of the choice of window size, we consistently identified an association between minor parent ancestry and recombination, after controlling for other features, and recombination rate was always the stronger predictor of minor parent ancestry (Table S2; Table S3).

1.4.2 Taking into account possible differences in power among windows

Our ability to reliably estimate recombination rates in *X. birchmanni* depends on SNP density and our power to detect ancestry switches in hybrids depends on divergent site density. Since both SNPs and divergent sites vary in density along the genome, so may the reliability of our results. Moreover, if they vary in ways that are correlated with features of interest, it could mislead us into thinking there are interesting biological differences when in fact there are just differences in error rates. To evaluate these possibilities, we thinned the data with respect to these features and re-inferred both recombination rates and ancestry.

For the LDhelmet map, we evaluated the distribution of polymorphic sites in 50 kb windows. We calculated the median number of polymorphic sites in all 50 kb

windows, which was 58. Next, for all windows with more than 58 polymorphic sites, we randomly subsampled 58 polymorphic positions for that window. We re-ran LDhelmet as described in Materials and Methods 1.2 to infer recombination rates based on these thinned data. We then re-analyzed the correlations between recombination rates and local ancestry.

For local ancestry inference, we evaluated the distribution of ancestry informative sites in 10 kb windows. As above, we calculated the median number of fixed ancestry informative sites in all 10 kb windows, which was 20. For windows with greater than 20 ancestry informative sites, we randomly sampled 20 of these sites and re-ran the HMM inference step of MSG. As with the polymorphism data, we asked whether relationships between ancestry and rate based on these thinned data were qualitatively changed.

Our results remained qualitatively unchanged after these thinning procedures. Specifically, the relationship between minor parent ancestry and recombination rate remained when the recombination rate was quantified using the thinned LD map (Fig. S21) and when ancestry was quantified based on inferences from the thinned ancestry calls (Fig. S21). Furthermore, this relationship was still observed when we reanalyzed the data in 10 kb windows, considering only the 50% of windows in the thinned dataset in which ancestry was inferred based on the exact same number of markers (20 markers; Fig. S22), and when considering the top 50% of windows where we expect to have the best power for rate inference with LDhelmet (Fig. S22). Finally, to guard against errors in ancestry assignment, in addition to excluding any shared polymorphic sites before ancestry inference, we re-analyzed the data excluding entire windows (of 10 kb) that contain any shared polymorphisms. The relationship is qualitatively unchanged (Fig. S22).

1.4.3 Possible effects of linked selection on the relationship of minor parent ancestry to recombination

LD-based methods to infer the recombination rate provide an estimate of ρ, the recombination rate scaled by the effective population size N_e. As a result, estimates of the local recombination rate will potentially confound variation in N_e with variation in recombination rate. In particular, selection on genic regions can reduce N_e nearby, potentially leading to under-estimates of ρ. In practice, there is only a weak correlation between distance to the nearest exon and levels of genetic diversity, which should also reflect variation in N_e (Fig. S23), and the relationship between π and ρ is driven only by the lowest recombination rate windows (Fig. S23). Nonetheless, we checked the robustness of our results by excluding any 50 kb windows with values for ρ in the lowest 25% quantile (i.e., < 0.00026 per bp; Fig. S23). We compared this analysis to correlations obtained when excluding windows with values for ρ in the highest 25% rate quantile, to account for the effect of excluding a quarter of windows. A similar correlation between minor parent ancestry and recombination rate was obtained when analyzing both sets (Table S6).

1.5 Modeling interactions between selection, the local recombination rate, and minor parent ancestry

To better understand the expected relationships between minor parent ancestry and recombination rate under neutrality and under different models of selection, we performed a series of simulations, using the hybrid population simulator admix'em (*58*).

1.5.1 Determining simulation parameters for the neutral simulations

Previous analyses based on the decay of admixture LD suggested that swordtail hybrid populations formed approximately 35-60 generations ago (*22*). We therefore performed simulations for times since admixture around those estimates and compared results to the real data. We simulated two 25 Mb chromosomes and placed crossover events along each chromosome according to the probabilities of the LDhelmet inferred map for chromosome 1 and chromosome 2 in swordtails, so that local rates in simulations mimicked the rate variation of the real data for swordtails. We simulated mixture proportions of 30:70 to approximately match observed mixture proportions in the data (Fig. 1E) and simulated a hybrid population size of 10,000. We found that simulations of 70 generations of admixture generated results similar to the data for one of the populations, and were not too far off for the other two. For instance, for these simulated parameters, the median homozygous tract length for the minor parent was 117 kb, with 95% central intervals ranging from 85 to 159 kb; by comparison, in population 1 (Totonicapa), the median length was 84 kb; in population 2 (Aguazarca), 225 kb; and in population 3 (Tlatemaco), 103 kb. In addition, the number of observed ancestry transitions in these simulations was similar to the observed data. We observed 103,819 - 388,963 events genome-wide in the three swordtail hybrid populations, compared to a 95% central interval of 217,712 - 331,134 events predicted genome-wide based on simulations. Thus, we proceeded with 70 generations in simulations as a rough approximation, but note that the presence of selection and demographic factors not considered in the simulations will influence tract lengths in the real data; we note further that these simulations are a better match to populations 1 and 3 than population 2, plausibly because of assortative mating in the latter (*32*).

Based on these results, we performed simulations of neutral admixture as described above; at generation 70, we randomly sampled 250 individuals from the population and summarized ancestry in 50 kb windows. Ancestry analyses and window thinning were performed as described for the real data (Materials and Methods 1.4). We note that in these simulations, ancestry is true ancestry, not inferred ancestry. We performed 200 neutral simulations for comparison with simulations of selection (see next section). Since our simulations mimic only ~10% of the genome and are computationally expensive, to match the approximate amount of data used in our actual analyses, we sampled 11 simulations without replacement 200 times. This procedure generated a distribution of Spearman correlation coefficients for the relationship between minor parent ancestry and recombination rate under neutrality (Fig. S1).

1.5.2 Simulations of selection on hybrids

We performed simulations under three models of selection on hybrids: 1) *BDM hybrid incompatibilities*, a model in which selection occurs because of epistatic

interactions between pairs of loci with different parental ancestries; 2) *hybridization load*, in which purifying selection on hybrids occurs due to long-term population size differences between the parental species (*6, 11, 12*); and 3) *ecological selection*, a model in which selection acts against minor parent ancestry because hybrids are in an environment more like that of the major parent and the two parental species are ecologically diverged. These three alternative models are obviously an oversimplification of the dynamics in natural populations and need not be mutually exclusive. In particular, in cases in which the traits underlying ecological specialization interact (e.g. *14*), loci under ecological selection may behave more like the BDM incompatibilities modeled here.

All three models of selection on hybrids predict that minor parent ancestry should depend on linkage to deleterious mutations, and that the recombination rate and the location of deleterious mutations should be key parameters shaping the distribution of minor parent ancestry in the genome. For each model, we picked selection coefficients so that the total selection on F_1 and F_2 hybrids was similar across scenarios ($w_{F1} = 0.9$ and average $w_{F2} \approx 0.9$). Under all models, effects of selected loci on individual fitness were multiplicative. Simulations and analyses were performed as with the neutral simulations described in the previous section, except that selection occurred every generation. In addition, we only performed 100 simulations under the hybridization load model, as these simulations were computationally extremely expensive.

In simulations of *BDM incompatibilities*, we modeled BDMIs as arising due to neutral fixation (i.e., we considered the derived alleles in the parental species to have the same fitness as the ancestral alleles) but, as we show below, the same patterns result from other types of incompatibility models. We simulated four pairs of hybrid incompatibilities with s=0.1 and h=0.5, placing their positions randomly in physical distance on one of the two simulated chromosomes. Extrapolating this number to the whole genome would suggest approximately 56 incompatible pairs across the genome, which is on the low end of what has been estimated for swordtail hybrid populations previously (based on two of the three populations included here; *22, 23*).

In simulations of *hybridization load*, we simulated 1,000 sites under weak selection, with $s=2\text{x}10^{-4}$ and h=0.5; their positions were randomly chosen along the physical maps of the two chromosomes. While this choice of dominance coefficient is arbitrary, we expect it to capture the effects of a range of h values (*14*); the exception is complete recessivity (i.e., h=0), in which case introgression could lead to heterosis in hybrids and patterns quite distinct to those observed (*14*). We chose this selection coefficient because demographic inference suggests *X. birchmanni* had a long-term population size four times that of *X. malinche* for the last $N_{malinche}$ generations (Fig. 1F). Simulating a population size of 10,000 for *X. birchmanni*, mutations with this selective disadvantage ($N_e s$ = -2) would be effectively purged in *X. birchmanni* (as single mutations) but behave nearly neutrally in *X. malinche* (in which $N_e s$ = -0.5). These mutations are the types that are hypothesized to lead to selection against *X. malinche* ancestry in hybrids under a load scenario (*6, 11*). Note that, for computational efficiency, the population size used here is lower than the likely historical population size for *X. birchmanni* (Fig. 1F); this choice will not impact simulation results, as the salient parameters are the compound values of $N_e s$ and the relative difference in population size.

In simulations of *ecological selection* against minor parent ancestry, we simulated four loci that reduced fitness if the allele is derived from the minor parent, each with s=0.05 and h=0.5. As above, we randomly placed their positions in physical distance on one of the two chromosomes.

The results of these simulations show that all three models of selection can induce a correlation between minor parent ancestry and recombination (Fig. 1; Fig. S1). In cases of incompatibility selection and ecological selection, minor parent ancestry is more likely to persist in regions of the genome with higher recombination rates. In the model of hybridization load, the direction of the correlation between minor parent ancestry and the recombination rate depends on which parent species has higher genetic load (Fig. S7).

Because the genetic architecture of hybrid incompatibility loci is not well understood, we also performed simulations under alternate hybrid incompatibility models, namely BDM incompatibilities that arise due to coevolution between loci or the fixation of beneficial alleles in the parental species. These incompatibility models differ from models in which the BDMIs fixed due to genetic drift alone (*59*). We performed simulations as described above, except that we modified selection coefficients (s=0.05 and h=0.5) so that average fitness of F_1 and F_2 hybrids matched the simulations described above. The results of these simulations indicate the relationship between minor parent ancestry and recombination rate is not dependent on the specific hybrid incompatibility model used (Fig. S24).

Similarly, because BDMIs between species may have a range of dominance coefficients (*9, 60, 61*), we performed simulations as described above, except with h=0 (setting s = 0.2 such that average fitness in the F_2 generation matched the previous simulations). In this case, a significantly positive relationship between minor parent ancestry and recombination rate was seen somewhat less often (in 52% of simulations; Fig. S24). This finding suggests that while the qualitative prediction of the model is similar, power to detect this relationship may be lower when BDMIs are fully recessive, presumably because fewer genotype combinations are under selection in this scenario. In practice, we probably expect a combination of dominance coefficients and even when h< 0.5, it may be rare for it to be precisely 0, as modeled.

Together, these simulation results demonstrate that local recombination rate is an important parameter in predicting patterns of minor parent ancestry in the genome, regardless of the major source of selection on hybrids.

1.5.3 Additional demographic models

In the previous section, we simulated a simple demographic history consisting of a single admixture event. To confirm that these expectations also held under more complex demographic histories, we considered two additional scenarios that are plausible: 1) multiple pulses of migration from the minor parent species (as seen in some swordtail hybrid populations; *32*) and 2) a strong, continual bottleneck in the hybrid population (Fig. S25). For simulations with multiple pulses of migration, we simulated a pulse of migration from the minor parent every ten generations with a migration rate of 1%. To generate final mixture proportions similar to other simulations, we set the initial mixture proportions to 20% parent 1 and 80% parent 2. For simulations of a bottlenecked population, we set the hybrid population size to 200 individuals, and otherwise performed 200 simulations as described in the previous section. We found that neither of these

scenarios resulted in a significant relationship between minor parent ancestry and recombination more often than expected by chance (i.e., by a two-tailed test, 8% of 200 simulations with multiple admixture pulses and 6% of simulations with a strong bottleneck were significant at the 5% level).

1.5.4 Impact of recombination rate and functional elements

The probability of minor parent ancestry is expected to depend on the number of deleterious alleles to which a region has been linked since hybridization occurred. Although this number is unknown, we expect that it should depend both on the local recombination rate and the number of selected sites nearby. To verify this intuition, we ran simulations relying on the locations of exons on chromosomes 1 and 2 of swordtails and placing selected sites within exons uniformly. We performed replicate simulations for BDMI, hybridization load, and ecological selection scenarios (as described in *Simulations of selection on hybrids*; 100 replicates for load simulations, 200 for other simulations). When considered jointly, both the local recombination rate in a window and the number of coding bps in a window predicted average minor parent ancestry in these simulations, over a range of scales (Table S7). In practice, selected sites may occur outside of exons; in the case of BDMIs, for instance, it is known that non-coding DNA and selfish genetic elements can be involved in incompatibilities between species (reviewed in *62*), and such regions may also play a role in hybridization load and ecological selection on hybrids. As expected from these simulations, when analyzed jointly, both the recombination rate and number of coding (or conserved) bps in a physical window are correlated with average minor parent ancestry in swordtail hybrids, though depending on the scale, the number of coding (or conserved) bps is not always a significant predictor (Table S2; Table S3; Table S8).

We further examined if the number of coding (and conserved) bps within a given genetic window—a measure that combines both features—is a better predictor of minor parent ancestry in the swordtail hybrid data than recombination rates or the number of coding base pairs within a physical distance alone. At the 0.1 cM scale at least, we found that minor parent ancestry was more strongly correlated with the local recombination rate alone than the number of coding or conserved bps at a given genetic distance (Table S2; Table S5; 0.1 cM windows were on average 56 kb). We therefore focus on results for partial correlations with recombination rates and coding or conserved base pairs as separate predictors (Table S2; Table S3).

1.5.5 Implications for the source of selection on swordtail hybrids

Recent work on the human-Neanderthal admixture event suggested that because Neanderthals had a persistently lower effective population size, they may have accumulated weakly deleterious mutations that were then selected against in the modern human gene pool following admixture (*6, 11*). Because Neanderthals are the minor parent species and also the population with lower effective population size, we expect that selection against BDMIs, selection due to hybridization load, and ecological selection against minor parent ancestry could all lead to increased minor parent ancestry in regions of high recombination (Fig. S1, Fig. S7; see Materials and Methods 1.9). Thus, based only on the relationship between average minor parent ancestry and recombination rate,

we cannot distinguish between the different modes of selection that could have led to this correlation, and indeed there may be multiple operating.

However, when the major parent is the species that carries more deleterious mutations, it is possible to distinguish between a hybridization load model and other models. Thus, in swordtail fish, our access to three populations with different mixture proportions (Fig. 1, 2) allowed us to explicitly evaluate the hybridization load hypothesis. As discussed above, the *X. malinche* source population has a much lower effective population size (Fig. 1), and has accumulated more non-synonymous mutations along its lineage, raising the possibility that selection on hybrids may be driven by weakly deleterious mutations carried by *X. malinche*. However, our simulation results suggest that if selection on hybridization load were the major source of selection on swordtail hybrids, population 3 should show a negative correlation between minor parent ancestry and recombination rate rather than a positive correlation (compare Fig. 2 to Fig. S7). The fact that a positive correlation is observed instead indicates that incompatibility selection (or ecological selection) is the dominant (though not necessarily sole) force shaping the relationship between recombination rate and minor parent ancestry in the three swordtail hybrid populations (Fig. 3).

Because the simulations described above in *Simulations of selection on hybrids* were designed to capture properties of all three of the swordtail hybrid populations, we reran simulations specifically for population 3, which is most informative in distinguishing between BDMI and hybridization load as sources of selection. To this end, we repeated BDMI and hybridization load simulations as described previously, except that we set initial mixture proportions such that the average minor parent ancestry after selection was 28%, matching observed mixture proportions in population 3 (Fig. 1E). Patterns were largely unchanged from what we had previously obtained. We observed a significant positive correlation between minor parent ancestry and recombination rate in BDMI simulations in 62% of simulations. In hybridization load simulations, we observed a significant positive correlation between minor parent ancestry and recombination rate when the minor parent species had higher load in 81% of simulations, and a significant negative correlation between minor parent ancestry and recombination rate when the major parent species had higher load in 86% of simulations.

In principle, strong ecological selection on hybrids could also generate the observed relationship between minor parent ancestry and recombination rate, but in practice there is no clear concordance between the ecologies of the swordtail hybrid populations and the parent species from which they derive most of their genome. In particular, population 3 (Tlatemaco), which derives the majority of its genome from the high-altitude and cold tolerant *X. malinche* parent, is found at lower altitudes (480 meters) than either of the hybrid populations with *X. birchmanni* as the major parent (population 1 – Totonicapa: 720 meters; population 2 – Aguazarca: 985-1000 meters). This mismatch between the ecological environment of the hybrid populations and their major parent populations can also be seen in available temperature data. In population 3 (for which the major parent is *X. malinche*), the average temperature in June 2015 was 24.1±1.4°C and in December 2015 was 20.1±1.1°C, whereas in population 2 (where the major parent is *X. birchmanni*), the average temperature in June was 20.8±1.2°C and in December was 16.2±1.3°C (temperatures are summarized from temperature loggers recording four times daily). Thus, the hybrid population with greater *X. malinche* ancestry is found in a

warmer habitat, whereas at least one of the populations where *X. birchmanni* is the major parent species is found in a colder habitat, when the opposite is true of the parental species (*63, 64*). These patterns run counter to our expectations if ecological selection due to temperature were responsible for the observed relationship between minor parent ancestry and recombination rate (Fig. 3).

Together, these results strongly suggest that BDMIs are a source of selection on swordtail hybrids, but it could be that both BDMIs and hybridization load are operating. To explore this possibility, we simulated both BDMIs and load, considering two scenarios for load: one in which the major parent species was the species that harbored fewer weakly deleterious mutations and one in which the major parent species was not. Consistent with previous simulations, we set selection coefficients such that the fitness of F_1 hybrids due to all sources of selection was ~0.9 (simulating two BDMI pairs with s=0.1, h=0.5 and 500 weakly selected loci with s=2x10^{-4}, h=0.5), performing 100 replicate simulations as described above. The results of these simulations suggest that when the impact of BDMIs and hybridization load on hybrid fitness are similar, the same qualitative relationship between minor parent ancestry and recombination rate is observed when the major parent is the low load parent, but no relationship between minor parent ancestry and recombination rate is observed when the major parent is the high load parent (Fig. S26). Given that minor parent ancestry increases with recombination rate regardless of whether it is the parent with greater or smaller load (Fig. 2), we conclude that selection on BDMIs is the predominant source of selection shaping ancestry patterns. We note, however, that there are necessarily many models with multiple sources of selection on hybrids that we do not explore.

Providing further support for the importance of BDMIs, deviations in minor parent ancestry are more pronounced near putative hybrid incompatibilities. Previous work in this system used signals of ancestry LD in two swordtail hybrid populations (populations 2 and 3, Aguazarca and Tlatemaco) to identify on the order of 100 pairs of unlinked putative BDMIs (*22, 23*). While the focus was on unlinked BDMIs because these are easier to identify without ambiguity, there could also be many (unmapped) linked BDMIs, especially if genes with related functions are often clustered in the genome (*65, 66*). To examine ancestry patterns at the unlinked, putative BDMIs, we selected pairs of loci previously identified in the 1% tail of ancestry LD in populations 2 and 3 (*23*) and used bedtools2 (*67*) to identify 50 kb windows overlapping with these sites, removing duplicate windows. We then compared average minor parent ancestry in these windows to that from 1,000 null datasets, generated by randomly selecting the same number of windows from the background and calculating the average minor parent ancestry. In all three hybrid populations, putative BDMIs had unusually low minor parent ancestry compared to that expected by chance (Fig. 3). We confirmed that this pattern is expected in simulations of hybrid incompatibilities (Fig. S27), based on 500 simulations sampled after 50 generations of admixture, with s = 0.1. Importantly, low minor parent ancestry at simulated incompatibilities is obtained whether or not we condition on a significant p-value for ancestry LD (at $p < 0.05$), so this observation is not due to the way in which the putative BDMIs were originally identified. Moreover, while low minor parent ancestry is seen in these simulations at truly epistatically interacting loci, it is not observed in false positives (e.g., loci in ancestry LD at $p<0.05$ that are not in fact under selection). Thus, the observation of lower minor parent ancestry than expected at putative

BDMIs is further support for selection against minor parent ancestry driven by linkage to a BDMI.

Together, these results suggest that hybrid incompatibilities are playing a predominant role in shaping the relationship between minor parent ancestry and recombination rate in these swordtail fish hybrids.

1.5.6 The impact of recombination localization mechanisms

If higher recombination rates predict the persistence of minor parent ancestry, then in species that use PRDM9, such as humans, minor parent ancestry should persist in distinct regions of the genome compared to in species that do not, such as swordtails (see Materials and Methods 1.6). In particular, because recombination is concentrated near CpG islands and TSSs in swordtail fish (Fig. S2; Materials and Methods 1.6), as in other species lacking PRDM9-directed recombination, then all else being equal, we might expect an enrichment of minor parent ancestry near these functional annotations. To test this prediction, we performed 100 simulations with admix'em as described above, placing hybrid incompatibilities uniformly within exons. Using the functional annotations of the simulated chromosomes, and roughly matching the amount of real data, we found that selection against hybrid incompatibilities leads to significantly higher minor parent ancestry near CpG islands in 99% of simulations at the 5% level (although the median Spearman's ρ between minor parent ancestry and distance from a CpG island in these simulations was -0.06; Fig. 4).

To compare patterns of minor parent ancestry near and away from functional elements in the data, we summarized average minor parent ancestry in 50 kb windows and used the program bedops (*68*) to determine the distance of each window to the nearest CpG island (see Materials and Methods 1.9 for details of the ancestry analysis in humans). For each window that overlapped a CpG island, we identified another window on the same chromosome that did not overlap a CpG island but was within 1% of the GC content of the focal window based on the hg19 and *X. birchmanni* reference genome sequences. We repeated this analysis for the TSS (Fig. S28).

To test whether the enrichment in minor parent ancestry observed at CpG islands (and at the TSS) is statistically significant, we randomly chose a set of windows in the genome, matching the number of windows that overlapped a CpG island (or TSS) in the actual data, then paired each randomly chosen window to a window with similar GC content. We repeated this procedure 200 times to obtain 200 sets of paired windows. We then asked in how many of the 200 sets the fold-enrichment in minor parent ancestry between the focal and matched set was greater than or equal to what is observed in the real data. By this procedure and as expected from simulations, we found that hybrid swordtail populations have significantly higher minor parent ancestry in windows overlapping CpG islands and TSSs, whereas humans do not (Fig. 4; Fig. S28).

1.6 Broad-scale patterns of recombination in *X. birchmanni* and expected rate conservation across species

In our analyses of the relationship between minor parent ancestry and recombination rate, we did not attempt to evaluate fine-scale rates in *X. malinche*, which

has very low diversity levels. Instead, we used recombination rates as estimated from LD data for *X. birchmanni* (Materials and Methods 1.4, 1.5), and assumed that recombination rates are conserved between *X. birchmanni* and *X. malinche*, as well as in their hybrids. This assumption is justified by what is known about mechanisms of recombination in these species. Notably in species that do not use PRDM9, including birds and yeasts, recombination hotspots have been shown to be remarkably stable, even over millions of years of divergence (*69, 70*), likely due to slow evolution of genomic features with which rates are associated. Our previous work with a lower-resolution crossover map (based on ancestry switchpoints in hybrids) suggested that swordtail fish behave like PRDM9-knockouts with regard to the genomic localization of recombination events, despite carrying a partial ortholog (*25*).

With our higher resolution population recombination map based on LD, we revisited patterns of recombination along the genome. We used the program bedops (*68*) and annotations available from the *X. maculatus* reference genome to assess how estimated recombination rates vary with distance to the nearest annotated transcriptional start site, CpG island, and H3K4me3 peak inferred from native Chip-seq of *X. birchmanni* testis (*25*). We also identified potential PRDM9 binding sites in the *X. birchmanni* genome based on the computationally predicted motif with the polynomial SVM model (*71*), using the program FIMO with a binding score cutoff of 5 (*72*). Based on these analyses, we confirmed our previous results of a strong relationship between the recombination rate and distance to CpG islands and testis H3K4me3 peaks (Fig. S2), and again found that recombination is not locally enriched near the computationally-predicted PRDM9 binding motif (Fig. S29). These findings support the hypothesis that the partial PRDM9 ortholog carried by swordtail fish is not involved in recombination (*25*).

Moreover, should PRDM9 play some role that we did not detect, the predicted PRDM9 binding motif is identical between *X. birchmanni* and *X. malinche* (*25*). Finally, the hybrid map also looks quite similar to the one obtained for *X. birchmanni* (see below). Together, these findings made us confident that fine-scale recombination rates are highly similar between the parental species.

1.7 A hybrid recombination map and differences to the parental map

We expected that hybrid and parental recombination maps will be highly similar (Materials and Methods 1.6) in the absence of strong selection on hybrid recombinants. To test this prediction, we generated a recombination map for hybrids using ancestry switchpoints and a novel Markov Chain Monte Carlo (MCMC) based approach and compared the hybrid map to the LD-based map generated for *X. birchmanni*.

1.7.1 Building a high resolution hybrid map for *Xiphophorus* hybrids

In inferring hybrid recombination maps using our ancestry data, we focused on two of the hybrid populations (the Totonicapa and Tlatemaco populations, populations 1 and 3 respectively), in which the data we had collected for ancestry inference was high enough coverage to infer precise locations of switches between parental species in the hybrids and built an admixture-based crossover map. The output of the ancestry inference program, MSG, are the posterior probabilities for each ancestry state. To identify

switches that correspond to crossover events in a hybrid ancestor, we considered the interval over which the posterior probability changed from ≥0.95 in support of one ancestry state to ≥0.95 in support of a different ancestry state. Before identifying these intervals, we excluded all markers that were not in Hardy-Weinberg equilibrium after a genome-wide Bonferroni correction, concerned that genotyping errors could generate spurious switchpoints (1-1.2 million markers remained in the two populations). We also excluded any breakpoints within 10 kb of a contig edge, as we suspected that ancestry inference may be less accurate there. Because a subset of these ancestry switch calls may still be spurious, we assessed how the choice of filters impacts overall recombination patterns and correlations between hybrid and parental maps (see below).

This procedure resulted in 388,963 candidate intervals within which we infer a recombination event to have occurred in population 1 and 379,119 in population 3. The inferred recombination intervals varied substantially in their lengths, i.e., in the resolution of the crossover event. The median interval resolution was 15 kb in population 1, with 75% of breakpoints resolved within 35 kb or less, and a median resolution of 22 kb in population 3, with 68% of breakpoints resolved with 35 kb or less. The frequency of these switchpoints in a given window can be used to estimate a recombination rate. However, because of the large average size of these intervals, we expect the resulting map to be fairly low resolution.

To improve the resolution of the hybrid recombination maps, we developed and applied a novel MCMC based approach for the inference of recombination rate using observed intervals of ancestry switching. The input to our inference was a set of recombination events detected by an inferred switch of ancestry, and a genomic interval in which the switching occurs. There are two types of parameters that our procedure tried to infer. The first type of parameter is a recombination rate for each genomic window. Because recombination events are rare, we can approximate the process of recombination as a Poisson process. Namely, if, at each bp, there is a small probability of a recombination event that is independent of other events, the number of events happening in each genomic window is Poisson-distributed. The parameter of this Poisson distribution is its recombination rate. The second type of parameter is the exact position of the recombination event for each recombination interval. This parameter is necessary because the uncertainty in the location of the ancestry switchpoints is large. If we simply assumed that the event is equally likely to happen anywhere within the region of ancestry switching, we would lose information and obtain a noisier estimate. These two types of parameters are dependent on each other. Intuitively, if the exact positions of all recombination events are known, then due to the properties of the Poisson process, the recombination rate is equal to the expected number of recombination events. On the other hand, if the recombination rate is known, then the distribution of the exact position of the recombination event is simply a uniform distribution weighted by the recombination rate. Thus, we used a Gibbs Sampler procedure to sample from the posterior distribution. The idea of the Gibbs Sampler is to sample each variable conditional on the current values of all other parameters. Although the current values of each parameter might not represent the true values, this algorithm guarantees that the stationary distribution of the Markov Chain is equal to the joint probability distribution of all parameters. We implemented our algorithm in R. The output of the R program is a series of recombination rate estimates at each iteration, and the likelihood of all observed recombination intervals given the

recombination rate. The distribution of the recombination rate at each window can be approximated by the empirical distribution of the recombination rate estimates after discarding the burn-in samples before the likelihood reaches its stable level.

The computational details of our model are described as follows. Let there be N genomic windows and M recombination intervals. We defined genomic windows to be the same as those used to summarize parental rates at a given size scale. Let the i^{th} window be written as W_i and the recombination rate of W_i be r_i. The prior of r_i is a gamma distribution of rate parameter of 1 and shape parameter of average genome-wide recombination rate. Let pos_j be the position of the j^{th} recombination event, and L_j be the observed interval within which the breakpoint lies.

INITIATION

$prior \sim Gamma\left(alpha = \frac{M}{genome\ size * K}, beta = 1\right)$

$O_j^i = width(\ overlap(\ W_i\ , L_j\ )\ )$

$rate_0 = sample\ (size = N, prior)$

ITERATION:

$For\ k\ in\ \#iterations$:

$P(pos_j \in W_i) =\ r_i * O_j^i$

$place\ event\ j\ in\ a\ genomic\ window\ according\ to\ its\ probability$

$C_i =\ number\ of\ events\ in\ W_i$

$r_i \sim Gamma(alpha = \frac{C_i}{width(W_i)} + \frac{M}{genome\ size * K}, beta = 1 + K)$

$rate_k =\ sample\ r_i\ from\ its\ posterior\ distribution$

$likelihood_k =\ \prod_{j=1}^{M}(\sum_{i=1}^{N} r_i * O_j^i)$

TERMINATION:

$E(r_i) =\ mean(r_i^k)\ for\ all\ k > burnin\ iterations$

$Var(r_i) =\ var(r_i^k)\ for\ all\ k > burnin\ interations$

Since recombination events are inferred based on switches of ancestry, a recombination event can only be observed if, at the time of the recombination event, the ancestral individual was heterozygous for ancestry. Thus, we can think of our estimated r_i not as the true recombination Poisson Process rate, but as the rate of a thinned Poisson Process, where the event is only observed with probability p, where p is the probability of being heterozygous integrated over all generations since time of admixture. This probability p is unknown; we approximated it by using the current admixture proportion genome-wide and letting $p = 2f(1-f)$. By properties of thinned Poisson Process, the original rate before thinning is then $\frac{r_i}{p}$.

We applied this MCMC approach to ancestry switch intervals from populations 1 and 3, inferring rates separately for each population. We estimated hybrid rates in 5, 10, and 50 kb intervals and compared these maps to the *X. birchmanni* LD-based map (see *Correlations between hybrid and parental maps*).

1.7.2 Evaluating the accuracy of the MCMC approach using simulations

Since this MCMC approach is a new method for inferring recombination rates in hybrids, we wanted to evaluate its performance. To this end, we simulated 10 Mb

sequences with the program macs (*45*), with parameters matching estimates in *X. birchmanni*. We used a similar approach as described in Materials and Methods 1.2, except that we used inferred recombination rates from the first 10 Mb of chromosome 1 as the recombination map input into macs to better match the observed rates and rate variation in *X. birchmanni*. We also simulated sequences for *X. malinche*, setting divergence time between the species in units of $4N_{birchmanni}$ generations (where $N_{birchmanni}$ is the effective population size of *X. birchmanni* at present) to 2. Because *X. malinche* has approximately one fourth the diversity levels of *X. birchmanni* (see Materials and Methods 1.3), we set the relative effective population size of *X. malinche* to 1/4th that of *X. birchmanni* for $N_{birchmanni}$ generations. Although PSMC results (Fig. 1) support a more gradual decrease in effective population size for *X. malinche*, this choice leads to a rough match to the observed number of ancestry informative markers between *X. malinche* and *X. birchmanni* and thus seems sensible. In the simulations, we generated a hybrid population between simulated *X. birchmanni* and *X. malinche* populations with mixture proportions of 75:25 approximately 70 generations ago (Materials and Methods 1.4, 1.5).

Next, we sampled 500 haplotypes to generate 250 hybrid individuals and generated fastq files for these individuals with the program wgsim (v0.3.1-r13), with a sequencing error rate of 0.01. We simulated 100 bp single end reads and generated 100,000 reads per individual, equivalent to 1X coverage of this region (our average coverage in the real hybrid data). Using these simulated reads and the parental genomes, we ran MSG to infer local ancestry and ancestry transition intervals. Because in practice, inferred intervals in these simulations were better resolved than those observed in the real data, for each breakpoint identified in the simulated data, we sampled an interval size from the distribution of observed breakpoint intervals in population 1. We then applied the MCMC approach to infer hybrid rates. We performed 10 replicate simulations of this entire pipeline.

We used simulations to compare the performance of the MCMC approach to an approach where, when a breakpoint spanned multiple windows, we picked a window at random and placed the breakpoint in it uniformly. Since most intervals were shorter than 50 kb, the MCMC approach was (as expected) only modestly more correlated to the true map at the 50 kb scale: Spearman's ρ was 0.5 – 0.6 whereas after applying the MCMC, it was 0.51 – 0.65. However, at the 5 kb scale, the MCMC approach provided a substantial improvement over the uniform placement of breakpoints within windows: Spearman's ρ was 0.14 – 0.20, while after applying the MCMC, it was 0.28 – 0.43.

1.7.3 Correlations between hybrid and parental maps

In species without PRDM9-directed recombination, parental recombination maps are expected to be highly similar (Materials and Methods 1.6), and thus hybrids are expected to have the same underlying recombination rates as observed in the parental species, unless there is strong and ubiquitous selection on recombinants or BDMIs that involve the recombination machinery. Because the hybrid and parental maps were generated using different approaches, however, which vary in power and specificity, it is unclear how correlated maps obtained from the two approaches should be, even in the absence of true differences among them, and therefore it is unclear how to interpret any apparent differences between them. To explore this question, we performed simulations in which the two approaches were applied to the same underlying map. We note that our

simulations do not incorporate several, potentially important sources of error (e.g., mapping errors), so we view the expected correlations from these simulations as an upper bound on the expected correlations between maps.

We performed macs simulations as described above (*Evaluating the accuracy of the MCMC approach using simulations*). From the simulated *X. birchmanni* population, we sampled 40 haplotypes as before and generated sequences with seq-gen, then inferred an LD map as described in Materials and Methods 1.2. We generated the hybrid recombination map for each simulation as described above (in *Evaluating the accuracy of the MCMC approach using simulations*). We compared the resulting hybrid rates to estimates from the simulated LDhelmet map in 50 kb windows. Because we used Spearman's correlation to evaluate map correlations in the real data, we also used it here.

Since these simulations were highly computationally intensive, we only performed 10 replicate simulations. However, results from these simulations suggest that even in this best case scenario, we only expect hybrid and LD maps to be moderately correlated: ρ varied from 0.5 to 0.68 across the 10 replicates. In general, simulated LD maps were better correlated with the true map than were hybrid maps: in comparisons of the LD map vs. the true map, ρ varied from 0.73-0.8 whereas in comparisons of the hybrid map vs. the true map, ρ ranged from 0.51 to 0.65.

This finding justifies our choice to investigate the relationships between recombination rates, functional annotations and local ancestry in hybrids using the fine-scale map obtained in *X. birchmanni*. Not only is this map predicted to be more reliable based on these simulations, it presents the advantage of being independently derived from the ancestry correlations that we are interested in investigating in hybrids.

1.7.4 Filtering approaches applied to the hybrid maps

To compare inferred hybrid and parental recombination rates, we quantified the recombination rate in each window as the rate in that window divided by the sum of the rate in all windows in the genome with rate estimates for both parentals and hybrids. We observed that, excluding the putative large species-specific inversions on chromosome 17 and chromosome 24 (Fig. S10), correlations between both hybrid maps and the parental map at the 50 kb scale were moderate (Spearman's ρ in population 1 = 0.4 and in population 3 = 0.36). Although the observed correlations between hybrid and parental maps were somewhat lower than might be expected based on simulations (where we obtained a range from $\rho = 0.5$ to 0.68 across 10 replicates), our simulations lacked sources of error likely present in the real data. Further, much of the difference between hybrid and parental maps is predicted by local ancestry variation in the hybrid populations (in population 1, $\rho = 0.26$, $p < 10^{-100}$; in population 3, $\rho = 0.15$, $p < 10^{-100}$). Thus, part of the explanation may also stem from local ancestry variation in hybrids impacting our power to detect ancestry transitions (*73*).

To evaluate whether additional filters might yield better correlations between the hybrid and parental maps, we tried different approaches. To this end, we focused our analysis on the hybrid map generated from population 1, since it is the highest resolution map and is slightly more strongly correlated to the parental LD map than the hybrid map generated from population 3. We asked if the correlation to the parental map was improved when requiring a breakpoint to be supported by 5, 10, or 100 markers on its 5' and 3' edges and then re-inferring rates based on these filtered breakpoints. None of these

approaches substantially improved the correlation between the hybrid and parental map, yielding a range of ρ from 0.35 to 0.41 for 100 to 5 markers, respectively.

A switch in ancestry that is shortly followed by a reversion to the previous ancestry state could likewise indicate genotyping errors or gene conversion events that are mistaken for recombination events. We generated a version of the hybrid map for population 1 where we removed all such switches when one of the flanking tracts was ≤5 kb. As above, this filtering did not improve correlations between the hybrid and parental maps (the resulting ρ=0.35).

Six individuals included in our analysis of population 1 have unusual ancestry (with hybrid indexes of 50%-86% of the genome derived from *X. malinche*, compared to ~25% for the rest of the population); these individuals are likely descendants of recent migrants from upstream *X. malinche* populations. We inferred the hybrid map excluding breakpoints identified in these individuals but found that it did not affect the correlation between maps.

1.7.5 Understanding local deviations between hybrid and parental maps

Although overall parental and hybrid maps are only slightly less correlated than expected under a best case scenario, hybrid maps in populations 1 and 3 had significantly lower estimated recombination rates than did *X. birchmanni* around the TSS, H3K4me3 peaks identified in the testis, and CpG islands. The greatest deviation between maps was seen at CpG islands, where hybrids have 11-13% lower heat in 5 kb windows overlapping CpG islands. This signal could be indicative of selection against recombination events, or could be a technical artifact of differing error profiles or sensitivity of the methods used to generated LD or admixture-based maps.

To ask if our estimation methods alone could generate these types of deviations, we simulated the first 10 Mb of the observed swordtail map for chromosomes 1-5 in macs *(45)*, following the simulation procedure described above, and inferred the parental and hybrid maps in 5 kb windows using LDhelmet and MSG and our MCMC-based approach, as in the real data. Because these simulations are computationally intensive, we only performed five replicate simulations of each chromosome. We then asked whether this procedure resulted in any deviation in inferred hybrid rates near CpG islands. We found that the differences in map inference approach did not result in higher inferred parental rates near CpG islands (Fig. S30). Although there are many sources of error that we did not model and could contribute to the difference between hybrid and parental maps at CpG islands, these results suggest that the map differences are not due to biases in our estimators.

We therefore explored other scenarios that could result in a rate depression at CpG islands such as the one that we observe in the real data with admix'em simulations. As before, we simulated a hybrid population generated from 30:70 mixture between the parental species and allowed admixture to occur for 70 generations. Finding no evidence that the estimated rate was decreased in CpG islands in the simulation scenarios described in Materials and Methods 1.5 with few BDMIs, we performed simulations with 50 pairs of weakly selected BDMIs per chromosome (s=0.01; average F_2 fitness ~0.9), with both randomly placed BDMIs and linked pairs in separate simulations (each interacting locus separated by 1 Mb). We also considered a scenario with a strong

bottleneck, where the hybrid population was formed with 200 individuals and maintained at this small population size.

Simulating many selected loci resulted in a modest but significant decrease in the estimated recombination rate in 5 kb windows overlapping with CpG islands (when BDMIs were unlinked, by 0.035 ± 0.006; when they were linked, by 0.02 ± 0.005). Simulations of strong genetic drift also resulted in significant rate decreases near CpG islands (0.06 ± 0.01). Thus, some combination of selection and genetic drift could explain the observation that estimated recombination rates in hybrids are lower near CpG islands. Importantly, however, simulations of genetic drift alone did not result in elevated minor parent ancestry at windows overlapping with CpG islands (Fig. S31), or generate a relationship between recombination rate and minor parent ancestry (Fig. S25).

1.8 Correlations in ancestry across independently formed hybrid populations

We observed local correlations in ancestry across all three independently formed swordtail hybrid populations (see main text, Fig. 3). To understand whether observed correlations in ancestry across populations are expected under different models of selection, we performed simulations using admix'em as described in Materials and Methods 1.5, simulating three hybrid populations formed by admixture of the two parental species (two at 25:75 mixture and one 75:25), with the same underlying selected loci in each set of simulations.

In the case of selection on hybrid incompatibilities, we expected that selection should induce local ancestry correlations between independently formed hybrid populations based on results from previous work modeling hybrid incompatibilities (*59*). Specifically, in populations with the same mixture proportions, the same loci are expected to fix for the major parent over time, inducing strong correlations in local ancestry (Fig. S6). Correlations in local ancestry are also expected for populations with different mixture proportions, because selection on hybrids to resolve the BDMI initially shifts ancestry at the loci involved in the incompatibility in the same direction, although to differing extents (Fig. S6; *59*). This counterintuitive behavior is illustrated in Fig. S6B and stems from that fact that, at least initially, resolution of the BDMI involves a fixation event at only one of the two loci involved in the incompatibility and an increase in the ancestral allele frequency at the other.

Consistent with these predictions from models of BDMIs, simulation results show that selection against the same incompatibilities induces positive correlations in local ancestry between independently formed hybrid populations, even if they differ in mixture proportion. Between simulated populations with the same mixture proportions (25:75 and 25:75), positive correlations in local ancestry in 0.1 cM windows were significant at the 5% level in 92% of 100 simulations, with Spearman's correlation coefficients ranging from ρ=0.14 to 0.78. In populations with opposite mixture proportions (25:75 and 75:25), positive correlations were significant in 68% of the 100 replicate simulations, with Spearman's correlation coefficients ranging from ρ=0.11 to 0.7 (whereas significant negative correlations were observed in 3% of simulations). In practice, correlations in ancestry as strong as those we observed in simulations may not be expected given the other forces influencing ancestry variation in independently formed hybrid populations, such as differences in demographic history among populations.

Repeating these simulations for a hybridization load scenario (as described in Materials and Methods 1.5), we also found positive correlations in local ancestry across hybrid populations with different mixture proportions. Because of the computational demands of these simulations, we only performed 50 simulations for each mixture scenario. In simulations of populations with the same mixture proportions, positive correlations in local ancestry were significant at the 5% level in 52% of simulations, with correlation coefficients ranging from 0.12 to 0.31. In simulations of populations with different mixture proportions, positive correlations in local ancestry were significant in 62% of simulations, with correlation coefficients ranging 0.11-0.32. Thus, both incompatibility selection and selection against hybridization load could drive the cross-population correlations in ancestry that we observe, but several other lines of evidence argue for selection against hybrid incompatibilities being the dominant process (Fig. 3C; Materials and Methods 1.5).

In contrast, simulations of repeated selection against minor parent ancestry in hybrids (e.g., due to ecological differences between the parental species) should lead to a distinct outcome. While local ancestry was positively correlated at the 5% level in 99% of simulations when comparing populations with the same mixture proportions (with correlation coefficients ranging from 0.15 to 0.74), negative correlations in local ancestry were significant in 100% of simulations comparing populations with different mixture proportions (with correlation coefficients ranging from -0.15 to -0.73). This finding suggests that ecological selection against minor parent ancestry does not explain the observed positive correlations in local ancestry between populations with different major parents.

1.9 Re-analysis of previously collected data on archaic ancestry in the human genome

One of the few other cases in which both local ancestry in the genome and fine-scale recombination rates are well characterized is the admixture of modern humans and Neanderthals. This case differs in many respects from that of *X. birchmanni* and *X. malinche*. For example, human-Neanderthal admixture occurred approximately 2,000 generations ago (*74*) vs. fewer than 100 generations (*22*), so that selection has had longer to act on segments of minor parent ancestry. In addition, the starting mixture proportions were likely much more skewed, with some studies suggesting that at most ~10% of the genome was initially derived from Neanderthals (*6, 11*). However, these admixture events also have interesting similarities. As is the case with swordtails (Materials and Methods 1.3), Neanderthals had a much smaller long term effective population size than modern humans, and previous work has implicated this difference as a plausible cause of the distribution of Neanderthal ancestry in the human genome (*6, 11*). Specifically, a model of selection against weakly deleterious alleles introduced by hybridization alone provides a good fit to the distribution of Neanderthal ancestry along the human genome (*6*). Other work proposed instead that hybrid incompatibilities played a role in shaping the distribution of Neanderthal ancestry in the human genome (*4, 75*).

Previous work investigating the distribution of Neanderthal ancestry in the human genome showed that it tends to be lower in regions linked to more coding sites (*6, 11*)

and that the strength of background selection, as measured by the B-statistic (*76*), is correlated with the frequency of Neanderthal ancestry (*4, 75*). However, B is an estimate of the strength of within-species selection against strongly deleterious mutations at linked sites, a phenomenon which should be operating in both parental species. As a result, it is not obvious a priori why it should be associated with low minor parent ancestry, apart from being a function of underlying features that are predicted to interact with selection after admixture, i.e., the local recombination rate and the presence of deleterious mutations. Thus, we focused our analyses directly on the local recombination rate and the number of linked coding bps. As with the swordtail data, we thinned the data such that one window was sampled every 500 kb (Materials and Methods 1.4).

We explored several options for quantifying Neanderthal ancestry in the human genome. We initially relied on estimates that have been widely used (e.g. *6, 77, 78*), namely posterior probabilities from Sankararaman et al. (*4*), calling a site as Neanderthal when the posterior probability exceeded 0.9. However, we found that these calls differed substantially depending on the choice of prior on the local recombination rate (specifically, a rate prior based on the human LD-based genetic map, (*4*) versus a uniform recombination rate prior), raising the concern that the relationship of Neanderthal ancestry to recombination could reflect in part the prior. We then considered the proportion of Neanderthal haplotypes, obtained by averaging the frequency of Neanderthal haplotypes at each site in a window (from *4*). These calls remained relatively insensitive to the choice of prior: the correlation between haplotype-based ancestry estimates generated from calls under a uniform rate prior versus prior from the combined human LD-based genetic map is $\rho = 0.95$ in 50 kb windows ($\rho = 0.98$ in 500 kb windows). We thus proceeded with the previously published haplotype-based measure of Neanderthal ancestry (from *4*). We note that using haplotype-based estimates of Neanderthal ancestry likely reduces the power to detect Neanderthal ancestry tracts in the human genome (as suggested by an estimated genome-wide mixture proportion for Europeans (CEU) of 2.5% based on posterior probabilities versus 1.1% based on haplotypes). To estimate the recombination rate, we used the human LD-based genetic map relied on by (*4*), which is combined across current human populations. Results are reported in Table S2.

One caveat is the greater power to detect Neanderthal ancestry tracts in some regions of the genome. However, if there were no true relationship of Neanderthal ancestry to recombination, then if anything, we would expect to see a negative correlation between minor parent ancestry and recombination rate, since there should be greater power to detect introgression in low recombination rate regions of the human genome (*4*). As discussed in the main text, our analysis instead reveals a positive relationship between recombination rate and Neanderthal ancestry (Fig. 2; Fig. S8). Thus, to the extent that systematic differences in power contribute to the relationship, we predict that the underlying signal is even stronger than what we report.

As a second approach to quantifying Neanderthal ancestry, we analyzed a call set developed for the diCal-admix project (*79*; http://dical-admix.sourceforge.net). Importantly, these Neanderthal ancestry calls were inferred with a uniform recombination prior across each chromosome. To obtain an estimate of the proportion of Neanderthal ancestry in a window, we used the "strict" mappability filter call set; converted posterior probabilities to ancestry calls (0 or 1 for Neanderthal ancestry) at a posterior probability

threshold of Neanderthal ancestry of 0.42; took the average ancestry over sites in the window for each individual and then took the average over individuals (*80*). We found that these calls were highly correlated with the haplotype-based calls of Sankararaman et al. (at the 50 kb scale, Spearman's $\rho = 0.88$; at the 500 kb scale, $\rho = 0.94$), and we found the same qualitative relationship between Neanderthal ancestry and recombination rate using these ancestry calls (Table S2).

Finally, we analyzed data from a reference-free approach for identifying archaic ancestry. Using archaic reference genomes will increase sensitivity to detecting introgressed haplotypes from populations closely related to the reference individual, but could introduce bias in the case of introgression from more distantly related populations. The approach of Skov et al. (https://github.com/LauritsSkov/Introgression-detection) uses a HMM applied to variants that are identified only in non-Africans relative to Africans. We used a posterior probability cutoff of 0.5 for data from the 1000 genomes project analyzed by Skov et al. (https://github.com/LauritsSkov/Introgression-detection), treating 1 kb regions with greater than >0.5 posterior probability as introgressed from archaic hominins and regions <0.5 posterior probability as not, then took the average ancestry over sites in the window for each individual and then the average over individuals. We excluded regions that were not analyzed for archaic ancestry due to poor callability (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible_genome_masks/StrictMask/) or because of repeat masking (hgdownload.cse.ucsc.edu/goldenpath/hg19/bigZips/chromFaMasked.tar.gz). We then repeated analyses as described above. We again found the same qualitative relationship between archaic ancestry and recombination rate alone and when accounting for the number of coding bps in the window (Table S2). Because different filtering approaches were applied to the hominin datasets in the studies described above, results reported in Table S2 are based on windows included in all datasets.

Due to rapid turnover of the fine-scale recombination landscape in species with PRDM9 (*81-83*), local recombination rates on the scale of 10s of kbs likely differ among modern human populations (even though at broader scales they remain similar, *81*) and could possibly have changed in the 47–65 Kya since humans and Neanderthals are thought to have interbred. We therefore repeated our analysis using multiple human recombination maps: the deCode map for Icelandic individuals based on pedigree data, the Yoruba LD-based map, and an African-American admixture map (downloaded from http://www.well.ox.ac.uk/~anjali/AAmap/). Our results are qualitatively similar across maps for a range of scales: regardless of the human recombination map used, Neanderthal ancestry is most depleted from the regions of the genome with the lowest recombination rates (Table S4).

Several groups have reported evidence for adaptive introgression of Neanderthal ancestry into modern human populations (*4, 78, 84, 85*). Because positive selection for Neanderthal ancestry could act as a countervailing force against genome-wide patterns of selection against Neanderthal ancestry, we also analyzed the relationship between minor parent ancestry and recombination when excluding windows in the top 1% Neanderthal ancestry. Interestingly, the relationship is strengthened (Fig. 2; Fig. S8).

Local ancestry data for Denisovan introgression into human populations are also available (75, 86). Because levels of introgression are reported to be highest in Oceanic populations, we focused our analysis on data from these populations. We analyzed data

from Sankararaman et al. (*75*) and Skov et al. (https://github.com/LauritsSkov/Introgression-detection; *87*). The approach of Skov et al. detects archaic ancestry tracts, then in a second step assigns an origin to the introgressed segments; thus, it does not initially distinguish between Neanderthal and Denisovan ancestry. To distinguish between these sources, we removed archaic ancestry tracts that are found outside of Oceanic populations, as all non-African human populations share an admixture event with Neanderthals (*27, 88*), but only Oceanic populations derive a substantial proportion of their genomes from admixture with Denisovans. Importantly, calls from Skov et al. are expected to be less sensitive to divergence between the Denisovan reference sequence and the hybridizing Denisovan population because of the reference-free approach (see above). We performed analyses as described for Neanderthal introgression except that for the calls from Sankararaman et al. (*75*), we used a posterior probability cutoff of 0.5, the threshold used in the original study.

We found a significant positive relationship between Denisovan ancestry and recombination rate using the calls of Skov et al. (*87*), over all scales considered (Fig. 2; Table S2). Using the calls of Sankararaman et al. (*75*), which rely on the Denisovan reference, we found a much weaker but similar relationship (Table S2).

2 Appendix of representative commands

Trimming adapter sequences from reads with cutadapt

```
./cutadapt  -b AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -b
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -b
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT -b
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT -o
Family_father_read_1.trimmed.fq.gz -p
Family_father_read_2.trimmed.fq.gz Father-
family_TCCGGAGA_HVF33CCXX_L007_001.R1.fastq.gz Father-
family_TCCGGAGA_HVF33CCXX_L007_001.R2.fastq.gz
```

Mapping with bwa

```
./bwa mem -t 3 -M -R
'@RG\tID:Fatherfamily\tSM:Fatherfamily\tPL:illumina\tLB:Fatherfamilylib
\tPU:NYGCIllumina' xma_washu_4.4.2-jhp_0.1_combined-unplaced-mito.fa
Family_father_read_1.trimmed.fq.gz Family_father_read_2.trimmed.fq.gz >
Family-father_RG.sam
```

Processing with picardtools

```
java -jar ./picard-tools-1.118/SortSam.jar INPUT=Family-father_RG.sam
OUTPUT=Family-father_RG.sorted.bam SORT_ORDER=coordinate

java -jar ./picard-tools-1.118/BuildBamIndex.jar INPUT=Family-
father_RG.sorted.bam
```

Realign indels with GATK

```
java -jar ./GATK3.4/GenomeAnalysisTK.jar -T RealignerTargetCreator -R
xma_washu_4.4.2-jhp_0.1_combined-unplaced-mito.fa -I Family-
father_RG.sorted.bam -o target_intervals_fatherfamily.list

java -jar ./GATK3.4/GenomeAnalysisTK.jar -T IndelRealigner -R
xma_washu_4.4.2-jhp_0.1_combined-unplaced-mito.fa -I Family-
father_RG.sorted.bam  -targetIntervals
target_intervals_fatherfamily.list -o Family-
father_RG.sorted.realigned.bam
```

Call variants with GATK

```
java -jar ./GATK3.4/GenomeAnalysisTK.jar -T HaplotypeCaller -R
xma_washu_4.4.2-jhp_0.1_combined-unplaced-mito.fa -I Family-
father_RG.sorted.realigned.bam --genotyping_mode DISCOVERY -L group1 -
stand_emit_conf 10 -stand_call_conf 30 -ERC GVCF -o Family-
father_raw_variants_group1.g.vcf

java -jar ./GATK3.4/GenomeAnalysisTK.jar -T GenotypeGVCFs -R
xma_washu_4.4.2-jhp_0.1_combined-unplaced-mito.fa --variant Family-
father_raw_variants_group1.g.vcf --sample_ploidy 2 --
max_alternate_alleles 4 --includeNonVariantSites --
standard_min_confidence_threshold_for_calling 30 -o Family-
father_GVCF_group1.g.vcf
```

Create hard-call files with custom script

```
perl create_insnp_oneindiv_GATK3_4_v10.pl Family-
father_GVCF_group1.g.vcf 20 20 10 40 2 60 4 -12.5 -8.0 10 100 25
```

Generate updated genomes

```
./seqtk mutfa xma_washu_4.4.2-jhp_0.1_combined-rm-mask.fa Family-
father_allgroups_maxcov.g.vcf.insnp > Xbirchmanni_COAC-family-
father_backbonejhp_0.1_rm-mask.fa
```

Identify Mendelian errors

```
./plink-1.07-x86_64/plink --noweb --file family-genomes_group1.phy --
mendel --out group1_family
```

Generate ancestral sequences

```
./phastcons/phast-1.3/bin/phyloFit --tree
"(((((Xmal_group1,Xbir_group1),Xcor_group1),Xmont_group1),(Xvar_group1,
Xmac_group1)),Xhel_group1)" allspecies_group1_nopoly.fa
```

Phasing

```
./shapeit_precompiled/bin/shapeit --input-ped combine_final-filter-
freq_family_population_data_group1.ped combine_final-filter-
freq_family_population_data_group1.map --input-map prior_map_group1.map
--output-max combine_final-filter-
freq_family_group1.shapeit2.phased.haps combine_final-filter-
freq_family.shapeit2.phased.sample --duohmm -W 5 --output-graph
duohmm.graph
```

LDhelmet

```
./LDhelmet_v1.7/ldhelmet find_confs --num_threads 10 -w 50 -o
combine_final-filter-freq_family_group1.conf combine_final-filter-
freq_family_group1.shapeit2.phased.haps.fa

./LDhelmet_v1.7/ldhelmet table_gen --num_threads 10 -c combine_final-
filter-freq_family_group1.conf -t 0.001 -r 0.0 0.01 1.0 1.0 10.0 -o
combine_final-filter-freq_family_group1.lk

./LDhelmet_v1.7/ldhelmet pade --num_threads 10 -c combine_final-filter-
freq_family_group1.conf -t 0.001 -x 11 -o combine_final-filter-
freq_family_group1.pade

cut -f 4 combine_final-filter-freq_family_population_data_group1.map >
snp_positions_birchmanni_group1.txt

./LDhelmet_v1.7/ldhelmet rjmcmc --num_threads 5 -w 5 -l combine_final-
filter-freq_family_group1.lk -p combine_final-filter-
freq_family_group1.pade -b 50.0 --snps_file combine_final-filter-
freq_family_group1.shapeit2.phased.haps.fa --pos_file
snp_positions_birchmanni_group1.txt -a ancestral.Xmal_group1-
Xhel_group1.probs_snps_probs -m mutation_matrix --burn_in 100000 -n
1000000 -o combine_final-filter-freq_family_group1.post
```

Run ancestry inference

```
perl msg/msgCluster.pl
```

PSMC command line

```
./psmc –N25 -t150 -r2 -p "4+25*2+4+6" -o xbir_COAC1.psmc
xbir_COAC1.psmc.fa
```

Admix'em selection file generation and simulation command line

```
Rscript generate_neutral_DMI_rec_exons.R 4 0.1

./admixemp admixsimul.cfg
```

Running MCMC rate inference

```
Rscript run_group1_MCMC_rate_inference.R
```

3 Supplementary Figures

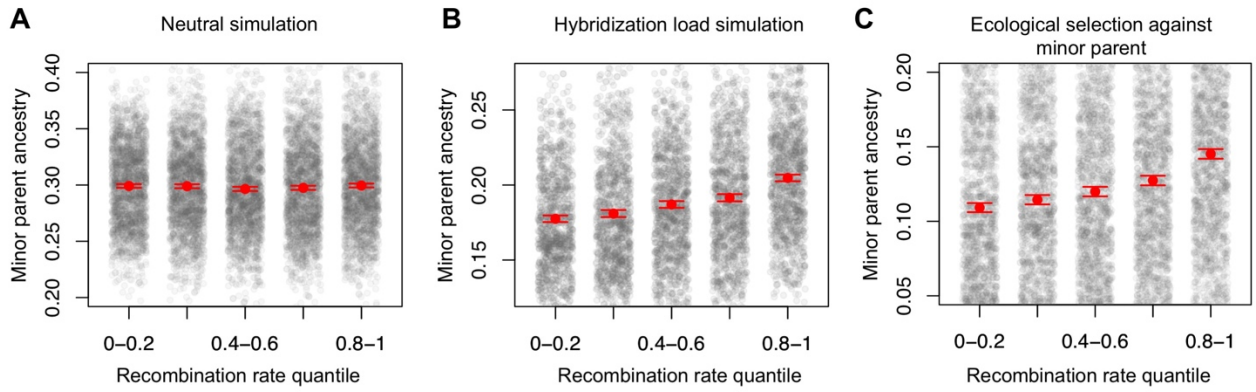

Fig. S1. Expected relationship between minor parent ancestry and recombination rate for different scenarios of selection on hybrids. Each simulation mimics the amount of real data; results plotted here are from the last simulation. Red points and whiskers indicate the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows and gray points show raw data; note that the y-axis is truncated. A relationship between minor parent ancestry and recombination rate is not expected in a neutral admixture scenario (Panel A; 6% of 200 simulations were significant at the 5% level). In contrast, a positive correlation is expected in simulations of selection on hybrid incompatibilities (Fig. 1; 74% of simulations were significantly positive at the 5% level); in simulations of hybridization load with a lower long-term effective population size of the minor parent (B; 85% of simulations were significantly positive at the 5% level); or when there is ecological selection against minor parent ancestry (C; 90% of simulations were significantly positive at the 5% level). See Materials and Methods 1.5 for details of the simulations.

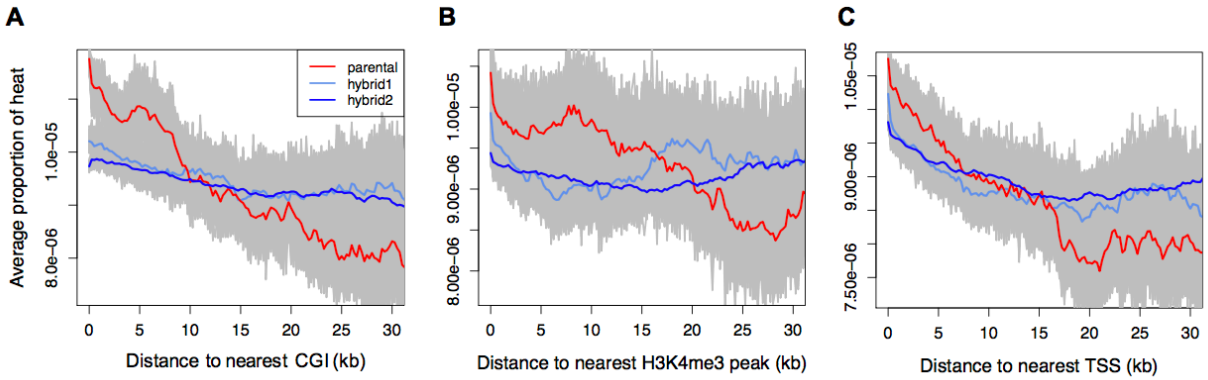

Fig. S2. Recombination events are elevated around promoter-like features. Recombination rates in the *X. birchmanni* LD-based map (red) and in two maps based on ancestry switches in hybrids (both blue) are higher around CpG islands (CGI) and transcriptional start sites (TSS), similar to what is observed in other species that do not use PRDM9-directed recombination and as expected from previous work in these swordtail fish (*25*). Gray lines show results of 500 replicate simulations jointly bootstrapping windows from the hybrid and parental maps. See Materials and Methods 1.7 for a discussion of the slight rate differences between maps.

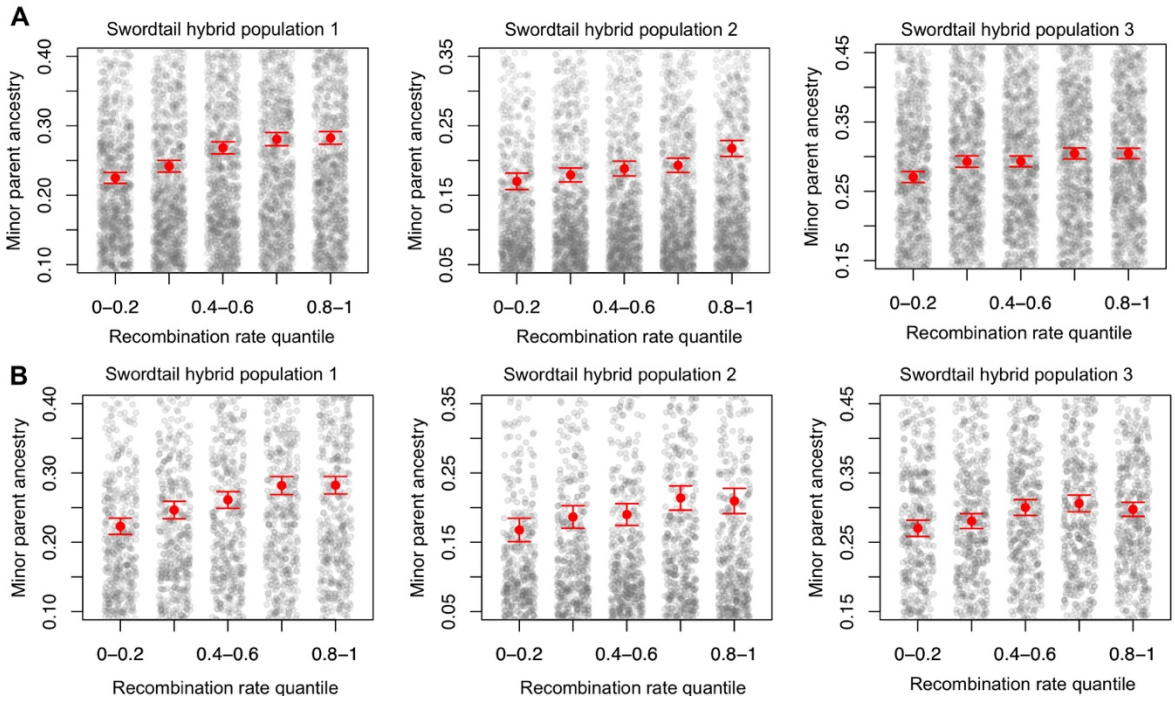

Fig. S3. Relationship between minor parent ancestry and recombination rate in swordtail hybrid populations, for different size scales. Results in the main text (Fig. 2) summarize the data in 50 kb windows. In A, results are shown for 100 kb windows and in B, for 250 kb windows. Red points and whiskers indicate the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated. Quantile binning is for visual representation only; all statistical tests reported in Table S2 were performed on the unbinned data. See Materials and Methods 1.4 for details.

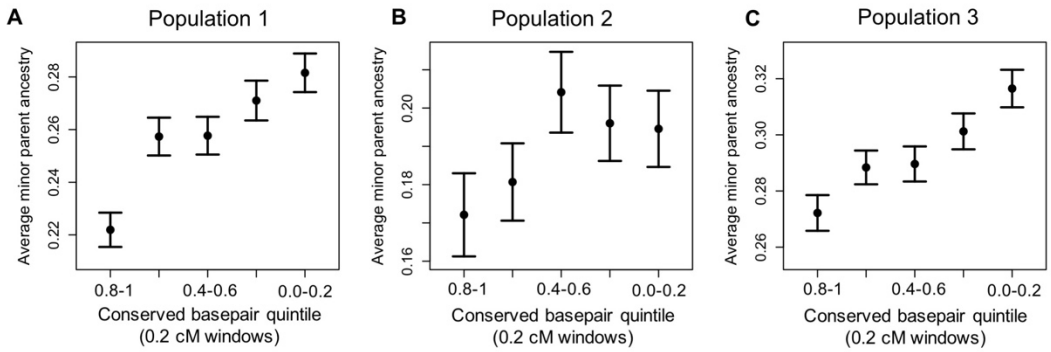

Fig. S4. Minor parent ancestry tends to increase in regions linked to fewer linked conserved base pairs. Shown here is the relationship between minor parent ancestry in a 50 kb window and the number of linked conserved elements (from *7*) within 0.1 cM of either side of the window. Spearman's correlation coefficient ρ ranges from -0.09 to -0.12, depending on the population (see Materials and Methods 1.4 and Table S5 for statistics for each population). The window size was chosen based on average tract length (Fig. S20). Points show the mean minor parent ancestry and whiskers indicate two standard errors of the mean, estimated from 1,000 replicates bootstrap resampling the data. A qualitatively similar relationship is seen for coding base pairs (Fig. S5). Quantile binning is for visual representation only; all statistical tests reported in Materials and Methods 1.4 were performed on the unbinned data.

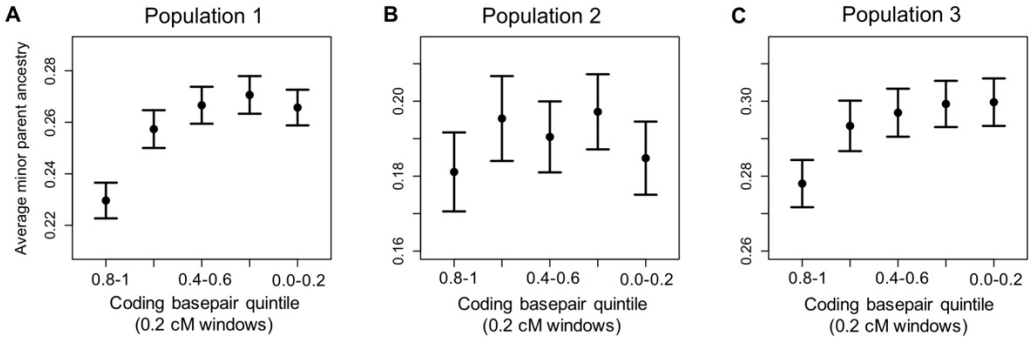

Fig. S5. Minor parent ancestry tends to increase in regions linked to fewer linked coding bps. Shown here is the relationship between minor parent ancestry in a 50 kb window and the number of linked coding bps within 0.1 cM of either side of the window. Spearman's correlation coefficient, calculated on raw, unbinned data, ranges from $\rho = -0.05$ to -0.1, depending on the population (see Materials and Methods 1.4 and Table S5 for details). The window size was chosen based on average tract length (Fig. S20). Points show the mean minor parent ancestry and whiskers indicate two standard errors of the mean, estimated from 1,000 replicates bootstrap resampling the data. A qualitatively similar relationship is seen for the number of linked conserved base pairs (Fig. S4). Quantile binning is for visual representation only; all statistical tests reported in Materials and Methods 1.4 were performed on the unbinned data.

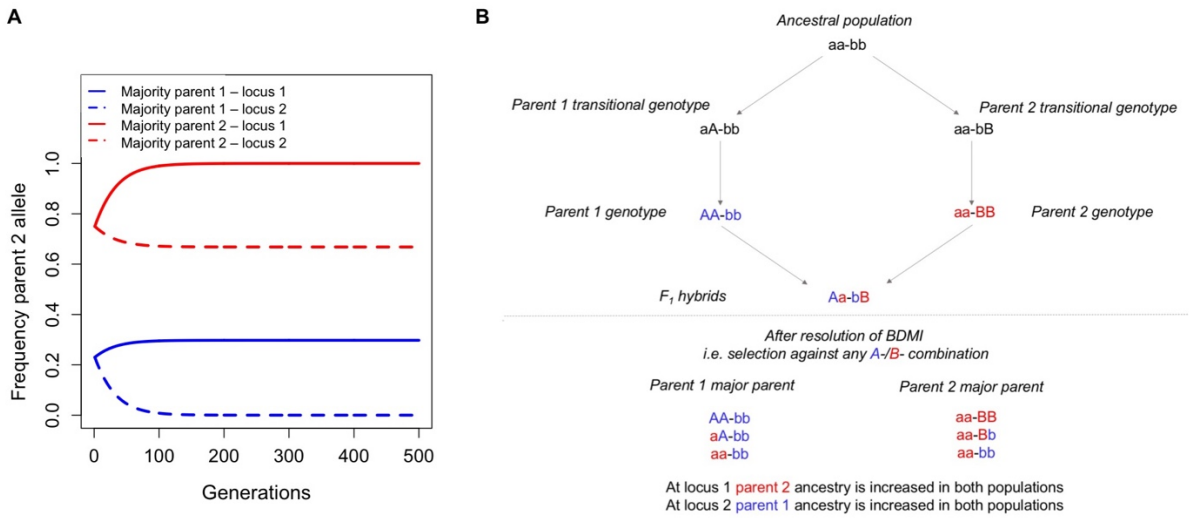

Fig. S6. Positive correlations in ancestry among hybrid populations with different admixture proportions are expected as a result of "neutral" BDMIs, where derived genotypes in the parental species do not have a fitness advantage compared to the ancestral genotype. (A) Frequency changes of each locus involved in a neutral BDMI hybrid incompatibility pair, in a deterministic model showing two hybrid populations with different starting admixture proportions. The dashed line tracking ancestry at locus 1 over time decreases in parent 2 ancestry in both populations, whereas the solid line tracking ancestry at locus 2 over time increases in parent 2 ancestry in both populations. As a result, regardless of mixture proportions, loci involved in BDMIs shift in ancestry in the same direction (though to differing extents; *59*), resulting in correlations in local ancestry even between hybrid populations that differ in their initial admixture proportions (Materials and Methods 1.8). (B) The mechanism driving this dynamic can be seen by examining the evolutionary history of a hypothetical BDMI. Ancestry for parent 1 is highlighted in blue and for parent 2 is highlighted in red. Because ancestral and transitional genotypes are equally fit under the neutral BDMI model, after selection on hybrids purges the BDMI from the population, they are expected to be fixed for the allele of the major parent at one of the loci involved in the BDMI and retain minor parent alleles at the other locus. This resolution will lead to positively correlated changes in ancestry in the two populations, despite their differing mixture proportions. In the case of "non-neutral" BDMIs, where derived genotypes in the parental species have a fitness advantage compared to the ancestral genotype (*89*), the parental genotypes are ultimately expected to fix in hybrid populations (e.g., the populations will be fixed for AA-bb and aa-BB), but the dynamics in recently formed hybrid populations should initially be similar to those presented here.

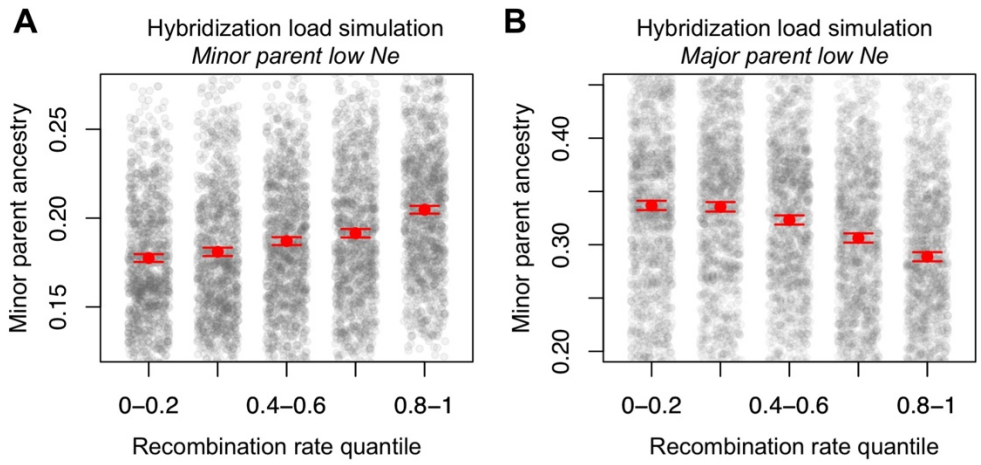

Fig. S7. Correlation between minor parent ancestry and the local recombination rate in simulations of hybridization load. Red points and whiskers indicate the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated. Each simulation mimics the amount of real data; results plotted here show data from a randomly chosen simulation (the last one). Selection against hybrid incompatibilities, ecological selection against minor parent ancestry and selection against hybridization load can all induce positive correlations between minor parent ancestry and recombination rates (Fig. 1, Fig. S1). In cases where hybridization load is the source of selection on hybrids, the direction of the relationship between minor parent ancestry and recombination rate depends on which parent species has had the lower long-term effective population size. In Panel A, 85% of 100 simulations had a significantly positive relationship between minor parent ancestry and rate at the 5% level. In B, 89% of 100 simulations had a significantly negative relationship between minor parent ancestry and rate at the 5% level. See Materials and Methods 1.5 for details.

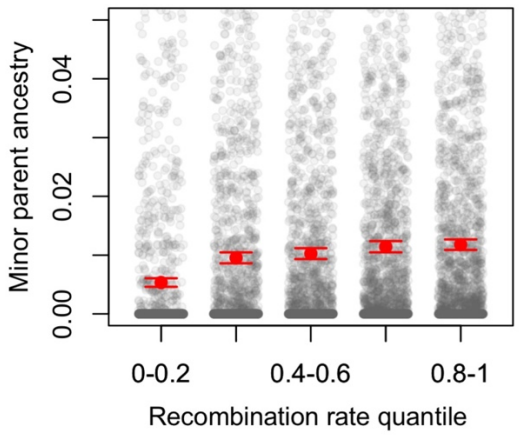

Fig. S8. Correlation between Neanderthal ancestry and local recombination rates in the human genome, when filtering out windows of unusually high Neanderthal ancestry (top 1%), which may be enriched for regions that have undergone adaptive introgression (*47, 51,56-58*). In data shown here, rate and ancestry are summarized in 250 kb windows but results are similar across size scales (Table S2). Red points and whiskers show the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated. Data are quantile-binned for visualization purposes; all statistical analyses were performed on the unbinned data. The resulting correlation (Spearman's $\rho = 0.23$, $p = 10^{-14}$) is stronger than when not excluding the top 1% (compare to Fig. 2).

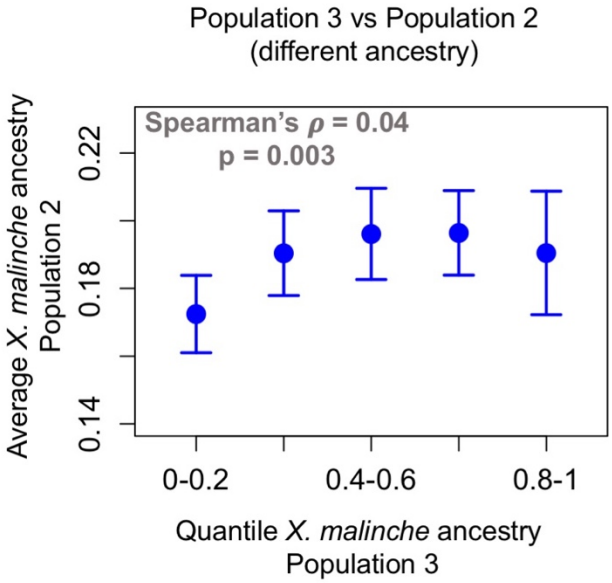

Fig. S9. Correlations in local ancestry between swordtail populations 2 and 3 in 0.1 cM windows (Aguazarca and Tlatemaco, respectively). Local ancestry is more strongly correlated between hybrid swordtail populations with similar genome-wide ancestry proportions (Fig. 3A,B). Points show the mean ancestry and whiskers indicate two standard errors of the mean. Quantile binning is for visualization only; all analyses were performed on the unbinned data.

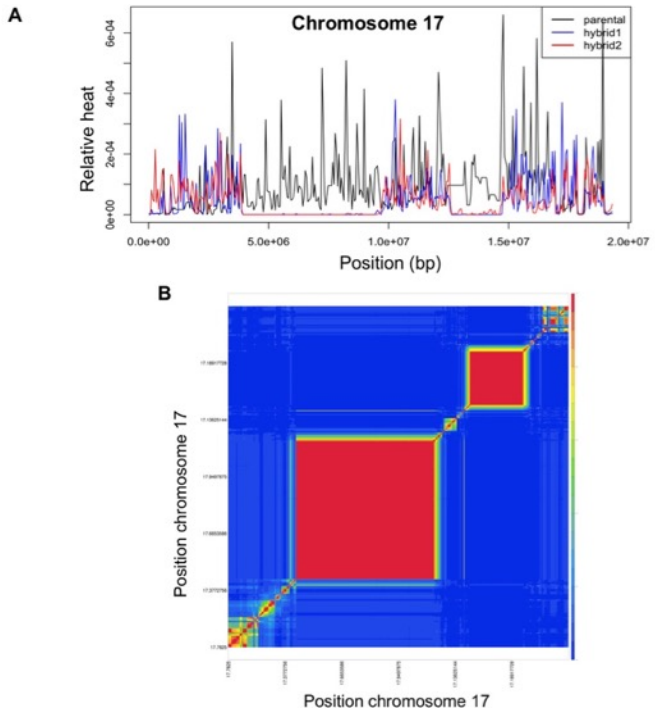

Fig. S10. Analysis of putative species-specific inversions on chromosome 17. Hybrid (red, blue) and parental maps (black) differ most strikingly in regions where few hybrid recombination events are observed—likely due to fixed inversions between species (A). These putative inversions can also be visualized in patterns of ancestry linkage disequilibrium, shown here for chromosome 17 (B). In B, red coloration indicates correlations in ancestry close to 1 and blue indicates correlations in ancestry close to 0.

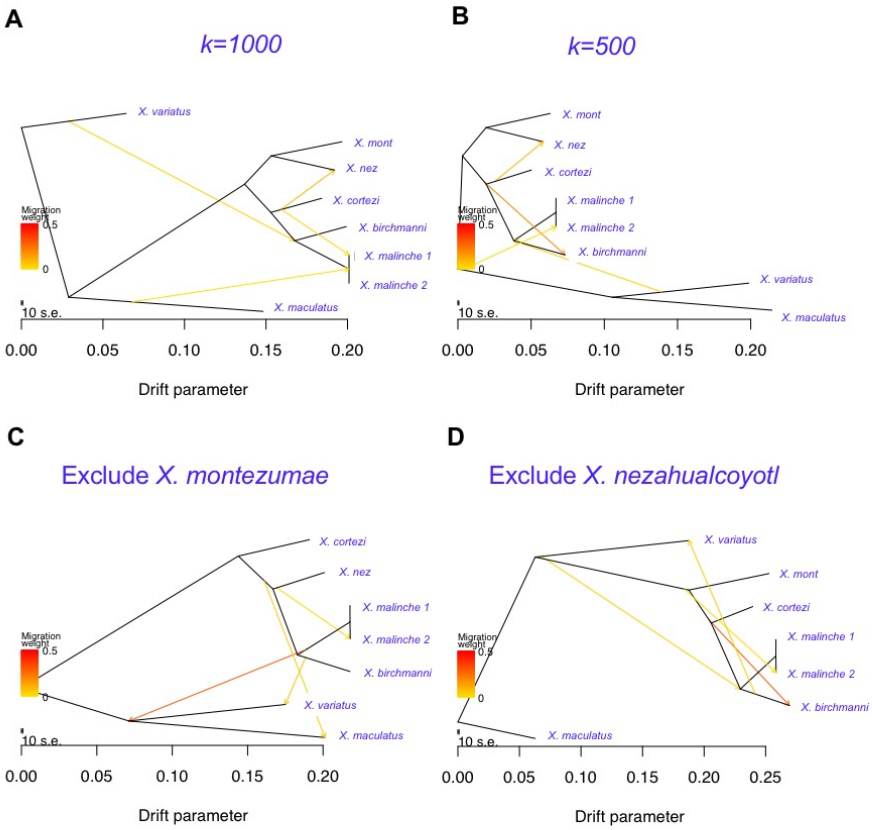

Fig. S11. Results from several implementations of *Treemix*. Each panel shows the maximum likelihood tree for swordtail species with four migration events inferred by *TreeMix*. The k parameter refers to the number of SNPs included in a window; it was set to 1,000 in all analyses, except for implementation shown in panel B. Arrows indicate inferred migration events, with the color proportional to the inferred weight of the migration event. Migration weights are related to the proportion of alleles in the recipient population derived from the migration event, and range from 0.008 to 0.09 in panel A. In no case is gene flow inferred between *X. malinche* and *X. birchmanni* since they split, but in all implementations, they both receive some gene flow from related species.

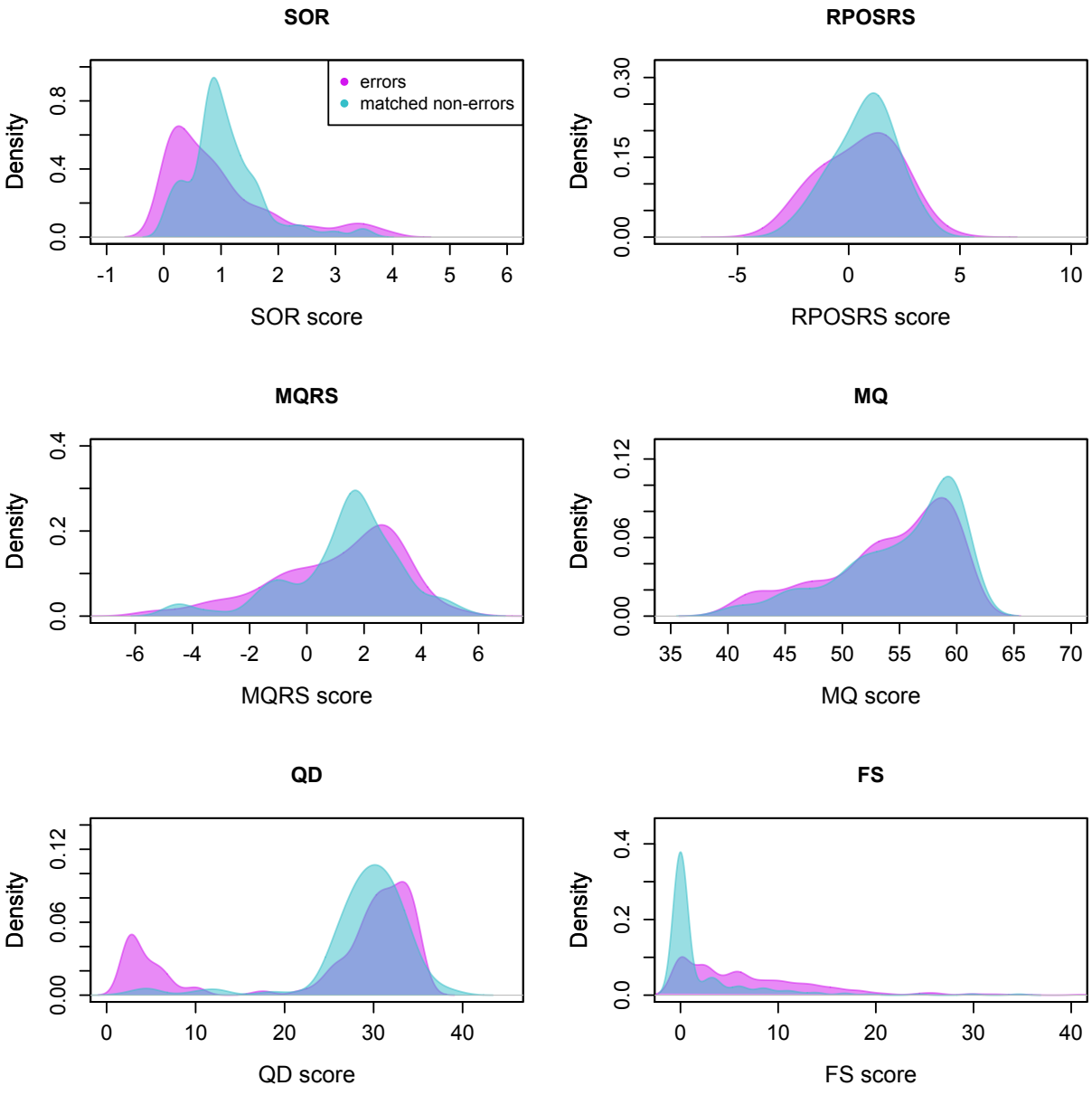

Fig. S12. Distributions of various quality scores for Mendelian errors identified through analysis of the family data (pink) versus matched non-errors from another individual in the family dataset (blue), at the same set of sites. These distributions were used to guide cutoffs in hard-call variant filtering; specifically, we modified our hard call cutoffs for the QD and FS metrics (see Materials and Methods 1.2).

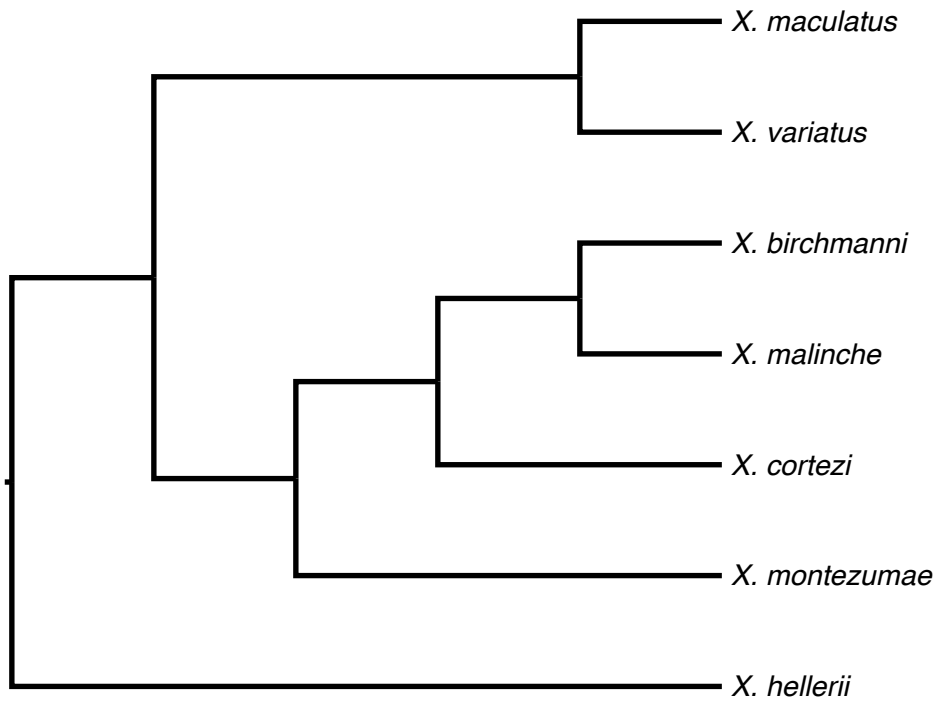

Fig. S13. Phylogenetic relationships between swordtail species (*5*) for which genomes were available and of sufficient quality for inclusion when inferring the ancestral sequence of swordtail species. Except for *X. maculatus*, genome sequences are pseudo-genomes generated by alignment to the *X. maculatus* reference.

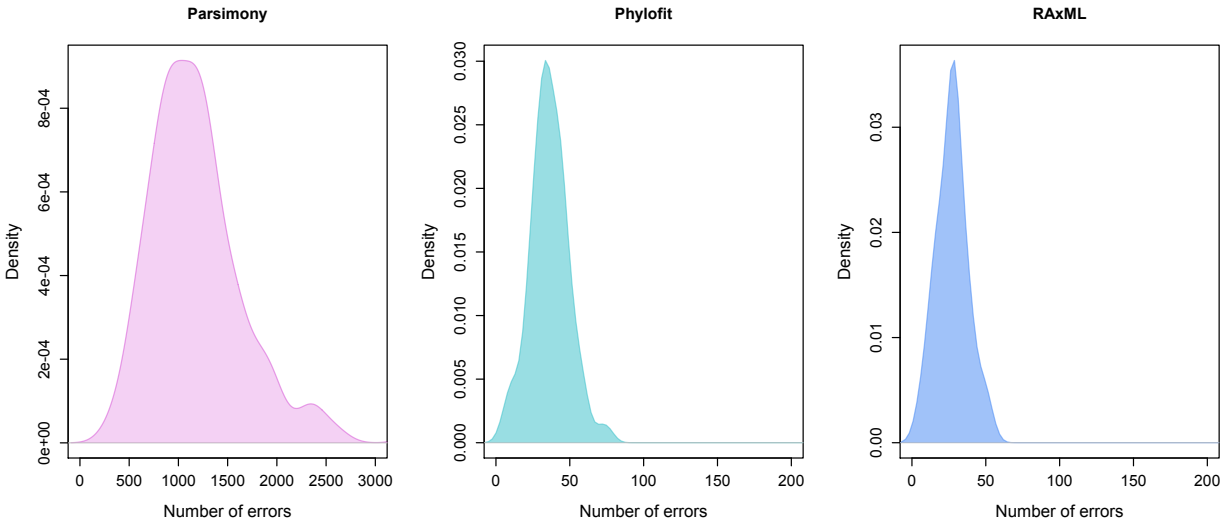

Fig. S14. Distribution of the number of errors in ancestral sequence inference per simulated 1 Mb sequence in 1,000 simulations. Both Phylofit and RAxML ancestral sequence inference outperformed a parsimony-based approach in simulations (see Materials and Methods 1.2).

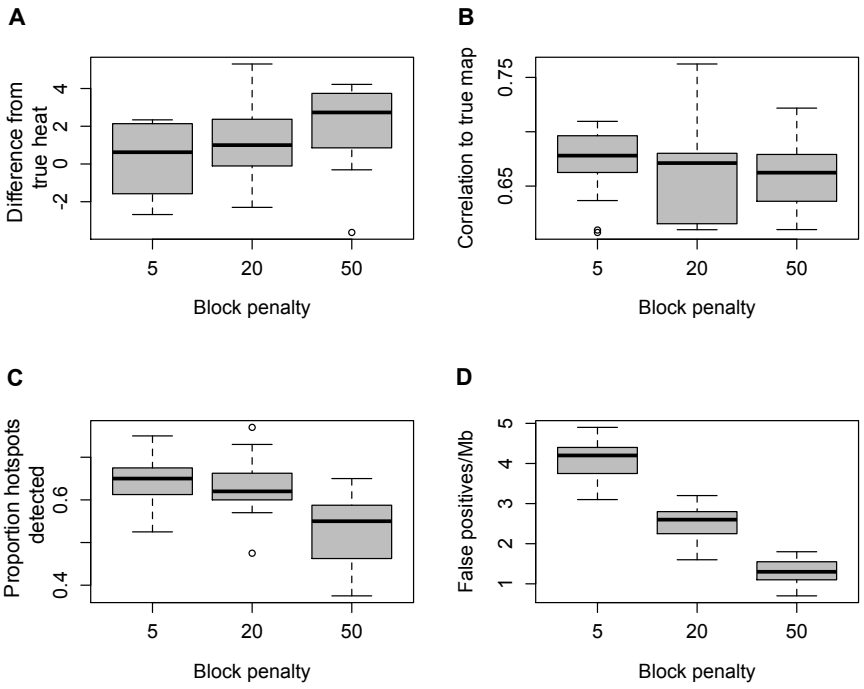

Fig. S15. Performance of LDhelmet in 15 replicate simulations with a block penalty of either 5, 20, or 50. The block penalty is the penalty imposed for switching recombination rates. Based on these simulations and other considerations, we used a block penalty of 5.

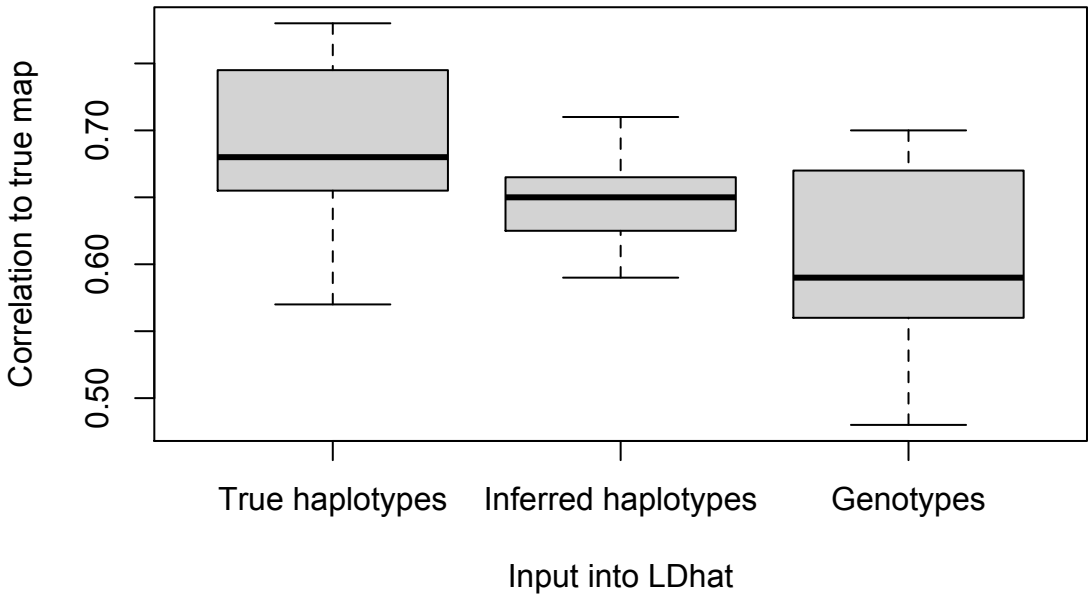

Fig. S16. Reliability of map inference in LDhat with different input data types in 15 replicate simulations. We ran these simulations to evaluate the impact of statistical phasing of haplotypes on the reliability of the resulting fine-scale genetic map. We report Spearman's correlation between the true and inferred maps in 50 kb windows. These simulations used LDhat, rather than LDhelmet as in the main analyses, because LDhelmet does not accommodate unphased data.

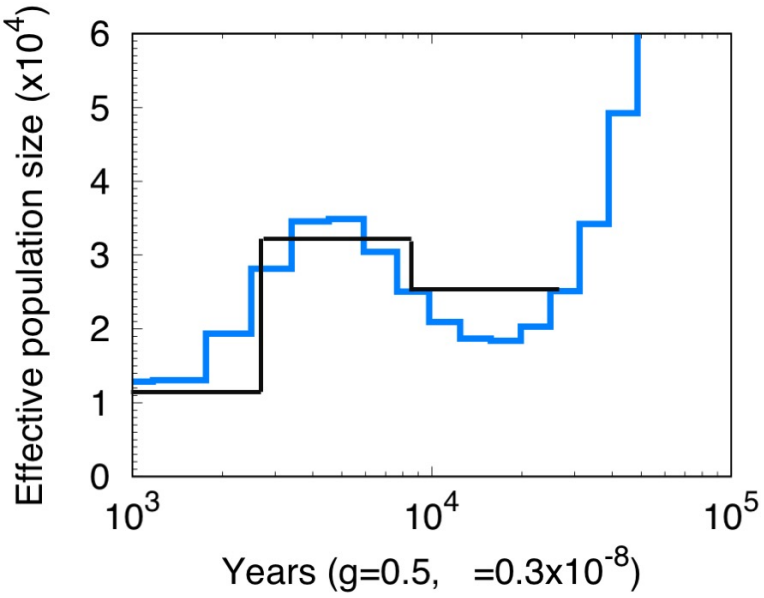

Fig. S17. Schematic of the approach used when evaluating the performance of LDhelmet under a more complex demographic history. Solid black lines show the simplified demographic history used in simulations (black) alongside the actual PSMC results for an *X. birchmanni* individual (blue); see Materials and Methods 1.3 for more details.

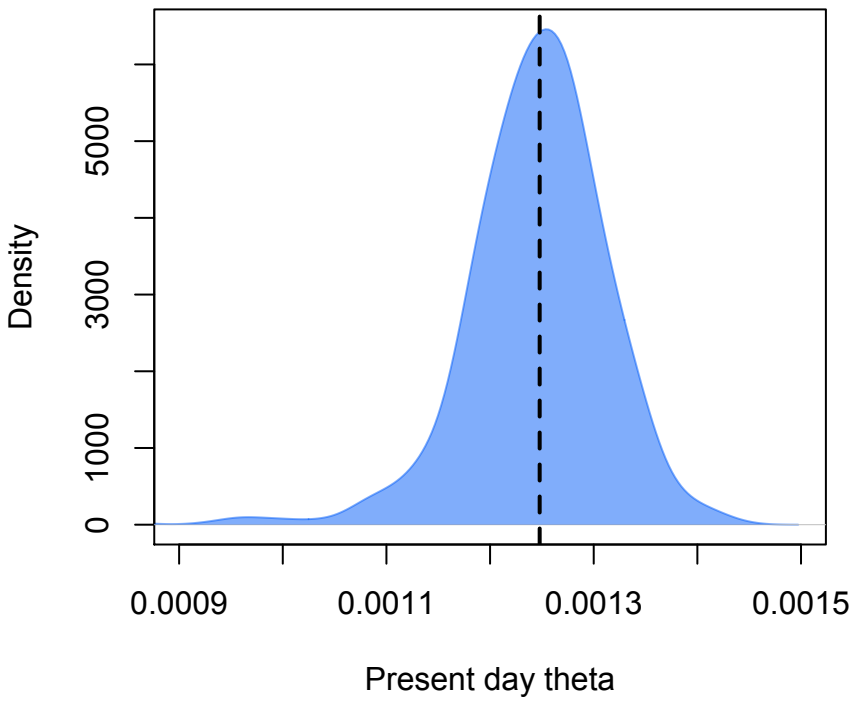

Fig. S18. Choosing the population mutation rate for simulations of more realistic demographic histories (see Fig. S17; Materials and Methods 1.3). The mode of the distribution of observed θ in 1,000 simulations of 10 Mb segment under this demographic history closely matches the observed θ in present day *X. birchmanni* when simulated θ is 0.00065.

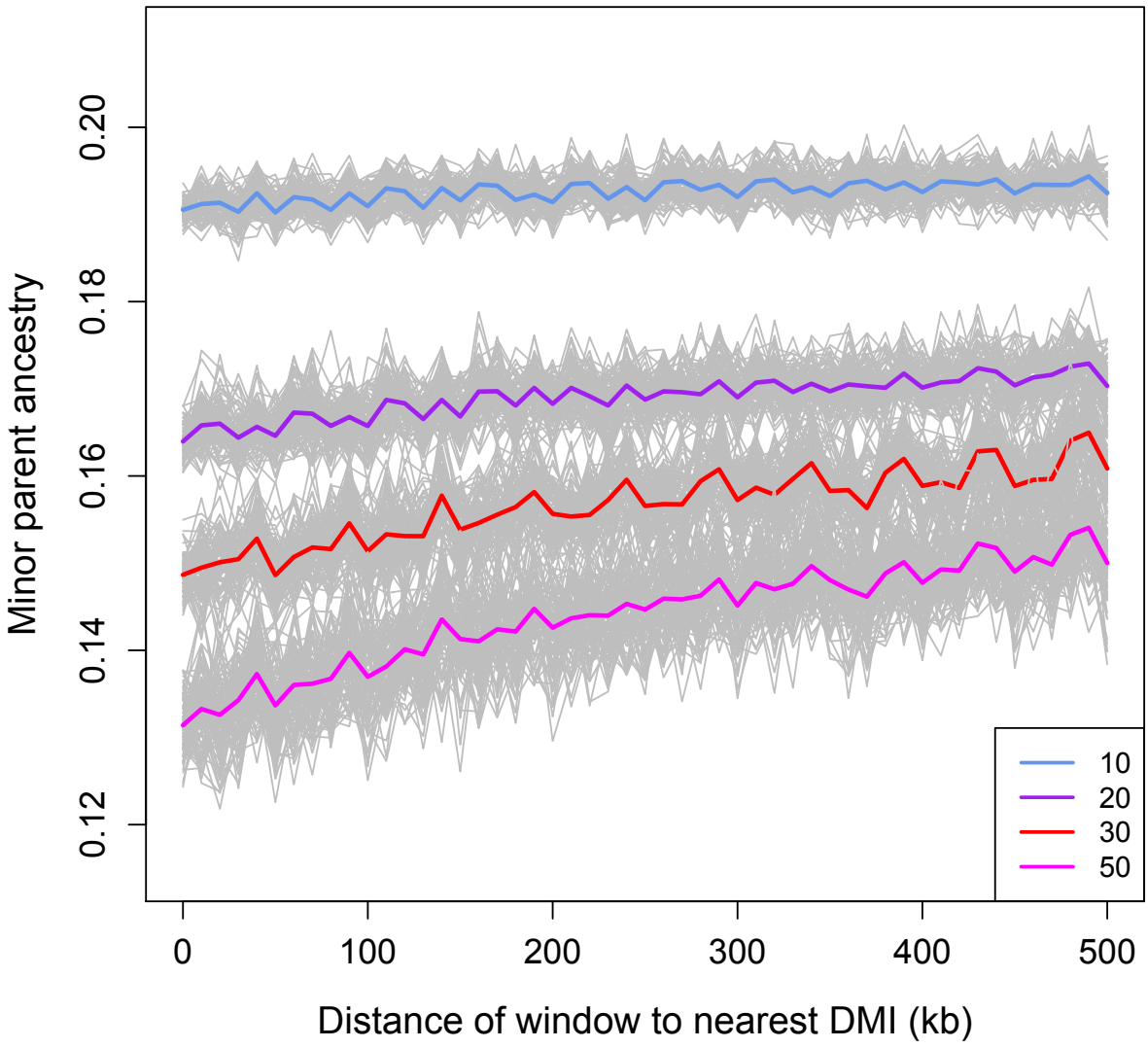

Fig. S19. The decrease in minor parent ancestry near hybrid incompatibility loci becomes more extreme and more localized over time. Shown here is average minor parent ancestry as a function of distance from a locus involved in a BDMI pair, sampled 10, 20, 30 and 50 generations after hybridization in 100 admix'em simulations (*58*). Gray lines show mean minor parent ancestry from 100 replicates bootstrapping windows.

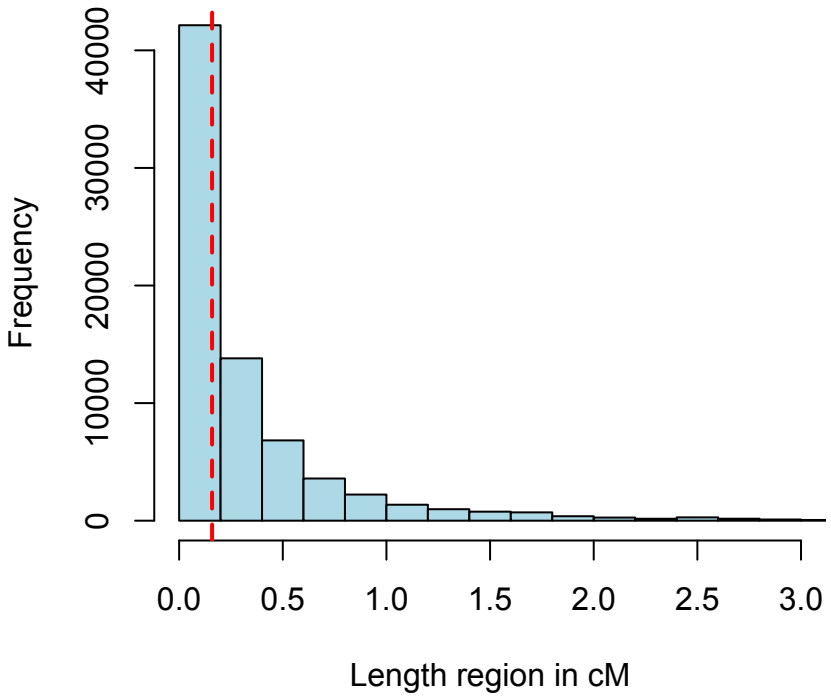

Fig. S20. The distribution of tracts homozygous for the minor parent in hybrid population 1 (Totonicapa). The red dashed line indicates the median tract length (estimated to be ~0.15 cM). This length is roughly the size scale that we focus on in analyses of linked coding and conserved base pairs (Materials and Methods 1.4).

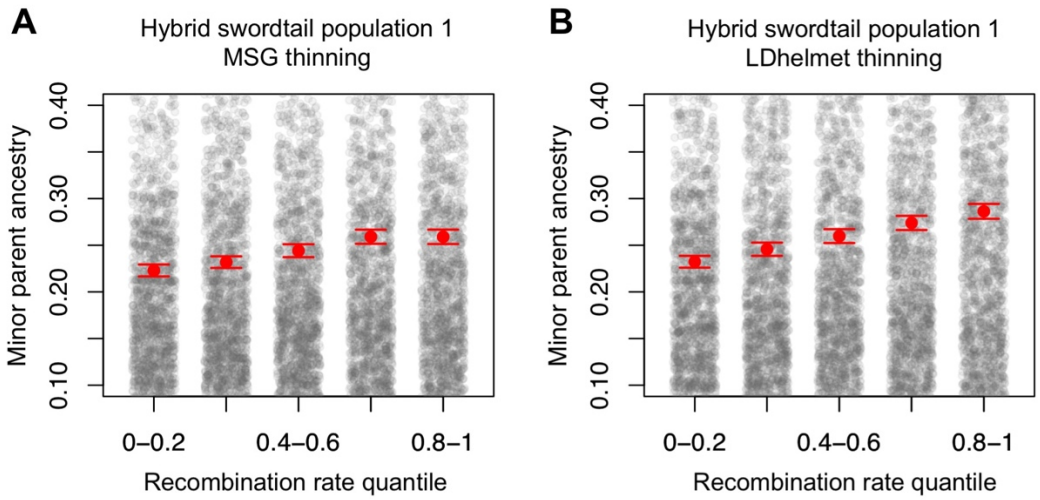

Fig. S21. Relationship between minor parent ancestry and local recombination rate after thinning. Specifically, we thinned ancestry informative markers before ancestry inference and the SNP data before LD-based rate inference. Thinning data before MSG inference of ancestry (shown in Panel A) or LDhelmet inference of recombination rate (B) does not change the qualitative relationships between minor parent ancestry and recombination rate: in unthinned analyses, Spearman's $\rho = 0.12$, $p = 10^{-8}$ whereas in the MSG thinned analyses, $\rho = 0.10$, $p = 10^{-4}$ and in the LDhelmet thinned analyses, $\rho = 0.10$, $p = 10^{-4}$. Ancestry proportions and rates are summarized in 50 kb windows. Red points and whiskers indicate the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated. Quantile binning is for visual representation only; all statistical tests were performed on the unbinned data.

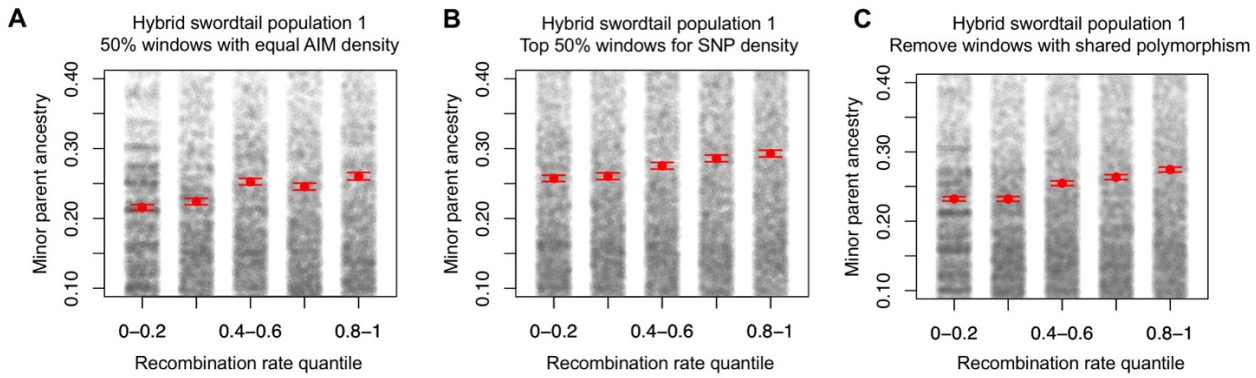

Fig. S22. Results considering only windows with lower predicted error rates for each analysis. (A) Results are shown for the 50% of 10 kb windows where ancestry was inferred based on exactly the same number of markers. (B) Results are shown for the 50% of 10 kb windows with highest SNP density, where LDhelmet estimates should be most reliable. (C) Results are shown excluding 10 kb windows that overlap shared polymorphisms between *X. birchmanni* and *X. mallinche*; although shared polymorphisms were masked prior to ancestry inference, this analysis guards against higher error rates in those windows. The correlations are similar to those for the entire dataset: without filtering, Spearman's $\rho = 0.11$, $p = 10^{-157}$; with equal AIM density in 10 kb windows, $\rho = 0.12$, $p = 10^{-46}$; using high SNP density 10 kb windows, $\rho = 0.1$, $p = 10^{-50}$; and removing 10 kb windows with shared polymorphism, $\rho = 0.11$, $p = 10^{-114}$. Note that in contrast to the p-values reported in Table S2, these p-values are reported based on all windows. Red points and whiskers indicate the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated. Quantile binning is for visual representation only; all statistical tests were performed on the unbinned data.

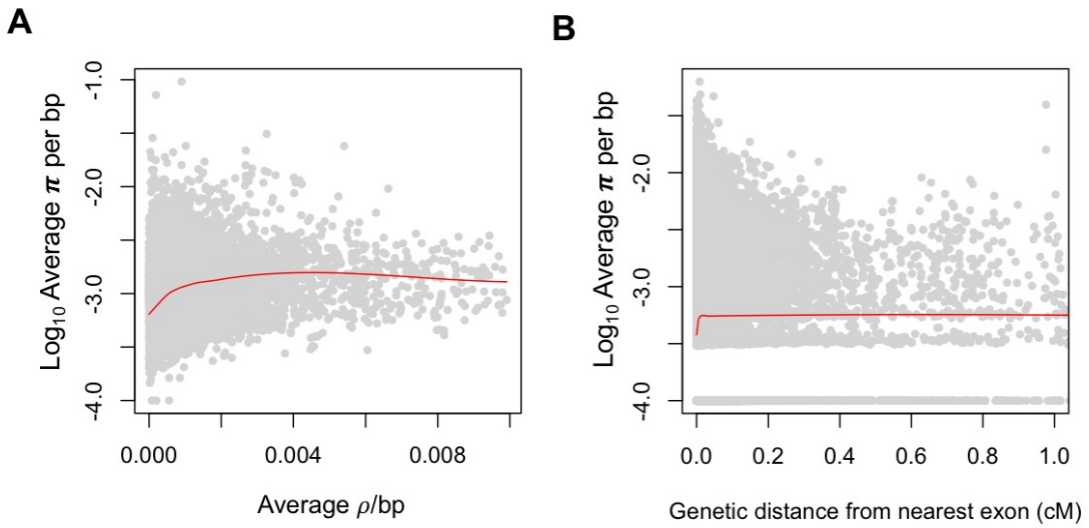

Fig. S23. Relationship between the heterozygosity (π) per bp, the population recombination rate, and the genetic distance to the nearest exon. Gray points show the raw data; red lines show a loess fit to these points. (A) As in other species, there is a positive relationship between π and the population recombination rate ρ/bp, summarized in 50 kb windows. We note that this relationship is expected from the fact that both π and ρ depend on the effective population size, but is also likely to reflect linked selection reducing diversity levels in regions of very low recombination. (B) Diversity is decreased within 0.05 cM of an exon, but not discernibly at farther distances.

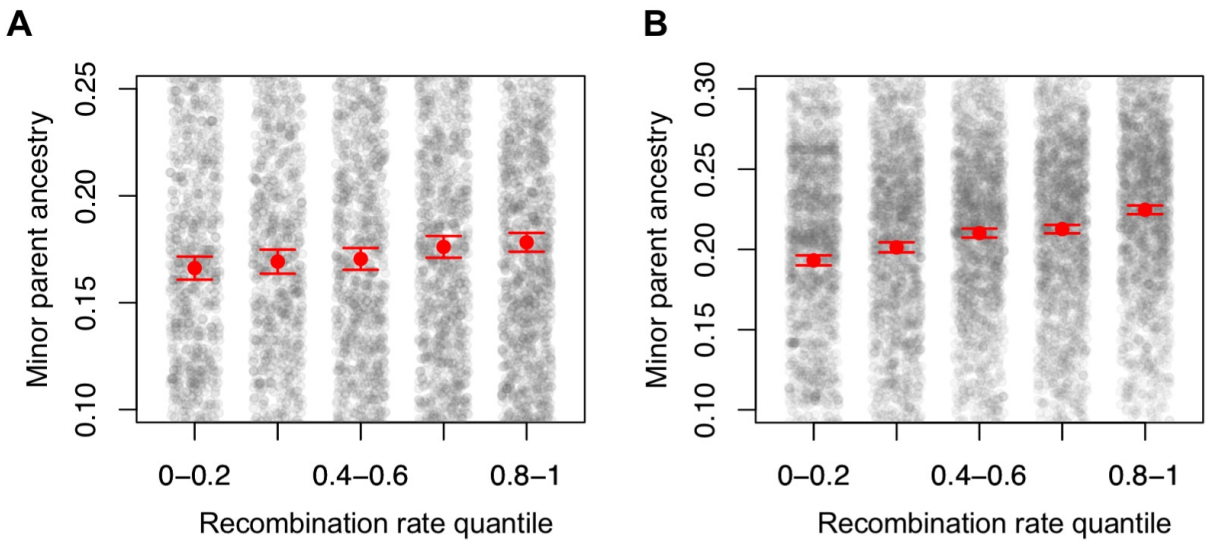

Fig. S24. Predicted relationship between minor parent ancestry and the local recombination rate under different models of hybrid incompatibilities. Red points and whiskers show the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated. The simulations shown here mimic the amount of real data; results plotted show data from the last of 200 simulations. In A, results are shown for simulations in which BDMIs are recessive (h=0). In this case, a significant positive relationship between minor parent ancestry and recombination rate still arises, but less frequently, with 52% of 200 simulations significant at the 5% level. In B, results are shown for simulations under a different hybrid incompatibility model. Specifically, in simulations of BDMIs shown in the main text (Fig. 1), we considered the case where the ancestral genotype has equal fitness to the derived genotypes (see Fig. S6). In contrast, under a coevolution model of hybrid incompatibilities, the ancestral genotype is also selected against (*59*). As illustrated here (B), this case is also predicted to generate a positive correlation between minor parent ancestry and local recombination rates; there was a significant positive correlation between rate and ancestry at the 5% level in 92% of 200 simulations. See Materials and Methods 1.5 for more details on these simulations.

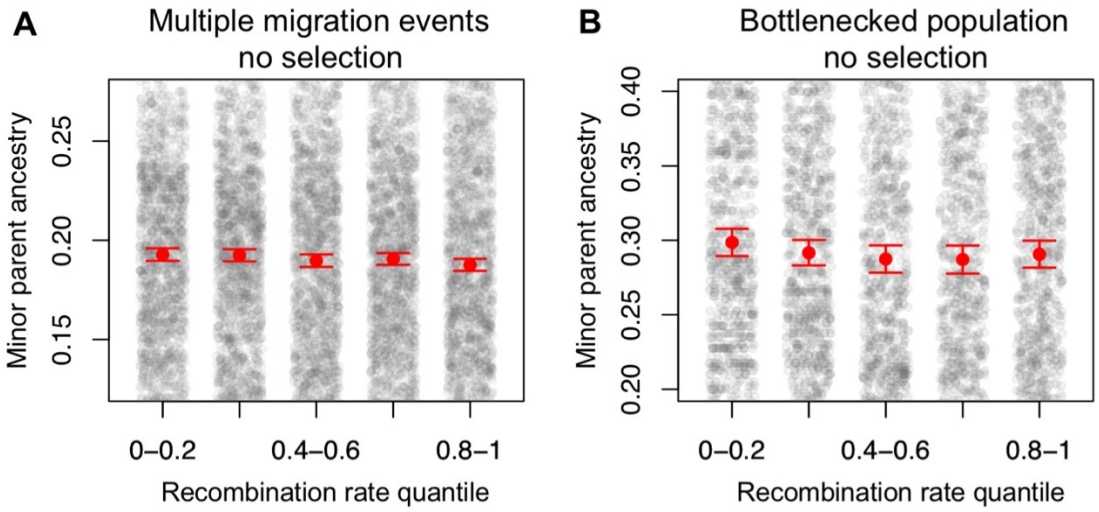

Fig. S25. Neutral simulations under additional demographic scenarios. In simulations with multiple pulses of admixture (A) and of a highly bottlenecked hybrid population (B), no relationship between minor parent ancestry and recombination is observed in the absence of selection. In Panel A, 8% of 200 simulations had a significant relationship between minor parent ancestry and recombination at the 5% level. In B, 6% of 200 simulations had a significant relationship between minor parent ancestry and recombination at the 5% level. Red points and whiskers indicate the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated. Each simulation mimics the amount of data used in our analyses of swordtail populations; results plotted here show data from the last simulation. See Materials and Methods 1.5 for details.

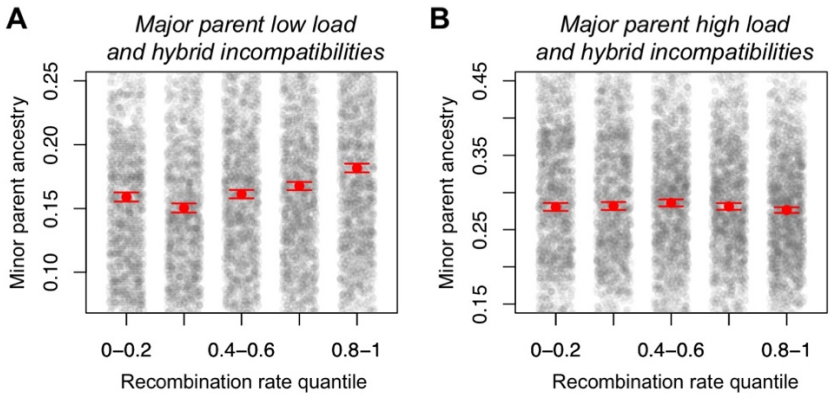

Fig. S26. Relationships between minor parent ancestry and recombination in the presence of both BDMIs and hybridization load. (A) In simulations with both BDMIs and hybridization load in which the major parent had lower load, positive correlations between minor parent ancestry and recombination rate were observed in 70% of 100 simulations at the 5% level. (B) In simulations with both BDMIs and hybridization load in which the major parent had higher load, positive correlations between minor parent ancestry and recombination rate were observed in only 6% of 100 simulations at the 5% level. Red points and whiskers indicate the mean minor parent ancestry with two standard errors of the mean determined by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated. Each simulation mimics the amount of real data; results plotted here show data from the last simulation. See Materials and Methods 1.5 for simulation details.

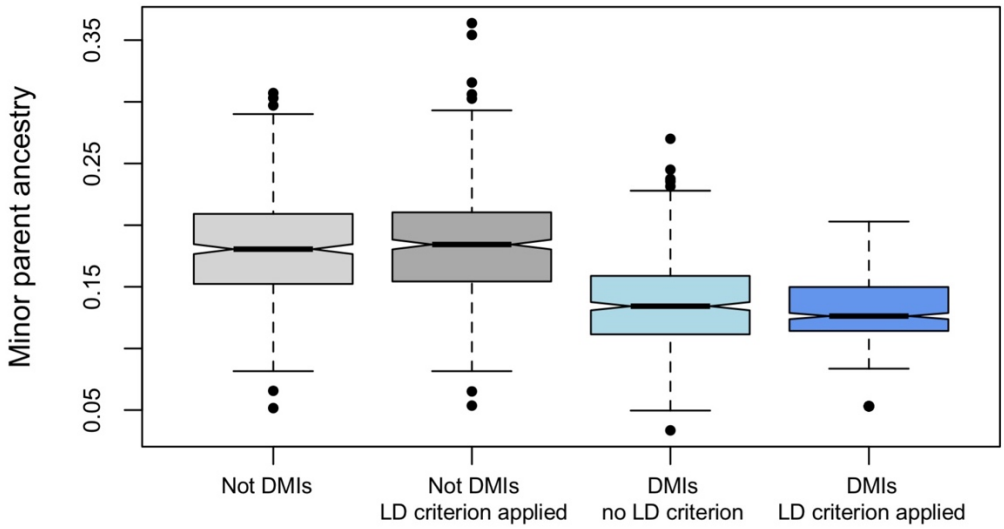

Fig. S27. In simulations of BDMIs, minor parent ancestry is unusually low at loci involved in BDMIs compared to loci that are not under selection. Shown here are results of 500 simulations, each including one "neutral" BDMI pair (s=0.1; h=0.5). To generate non-BDMI loci, 500 pairs were randomly selected from the background; to generate non-BDMI loci that further satisfy an admixture LD criterion, pairs were randomly selected from the background until a pair with admixture LD at $p<0.05$ was found. Pairs of loci randomly selected from the background have higher minor parent ancestry than BDMIs, whether or not they show a signal of admixture LD at $p<0.05$ (light and dark gray, respectively). In turn, pairs of loci involved in BDMIs have lower than average minor parent ancestry, whether or not they show a signal of admixture LD at $p<0.05$ (light and dark blue, respectively). Results for the real data are shown in Fig. 4.

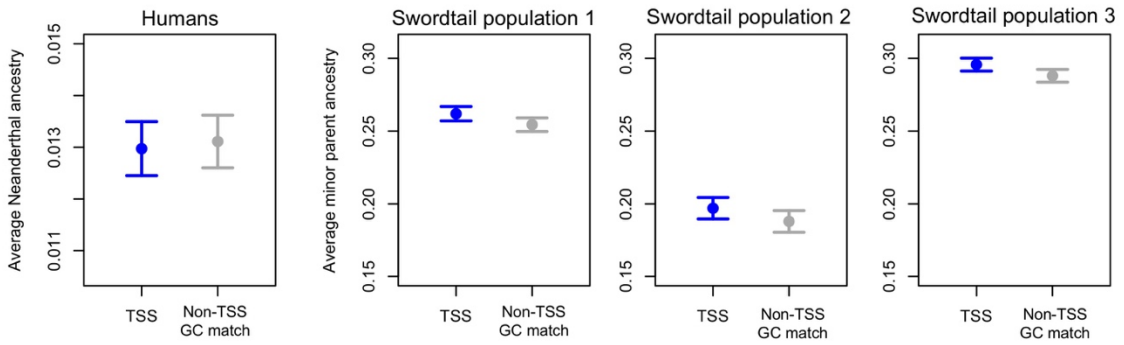

Fig. S28. Comparison of average minor parent ancestry in 50 kb windows that overlap a transcription start site (TSS) and windows with similar GC content (within 1%) that do not. Neanderthal ancestry in the human genome is not significantly elevated in 50 kb windows that overlap with the TSS compared to windows that do not (the fold difference is 0.99, one-tailed p=0.65). In contrast, swordtail populations show a significant elevation of minor parent ancestry in windows overlapping the TSS: for population 1, the fold-difference is 1.03, $p<0.005$ for population 2, the fold-difference is 1.05, $p<0.005$; and for population 3, the fold-difference is 1.02, $p<0.005$. See Materials and Methods 1.5 for a description of these analyses. Points show the mean of each group and whiskers show two standard errors of the mean determined by 1,000 joint bootstrap samples of the data. The difference is smaller than that observed for windows overlapping CGIs (Fig. 4), as expected from recombination rates being more elevated around CGIs than TSSs in these swordtail fish species (*25*).

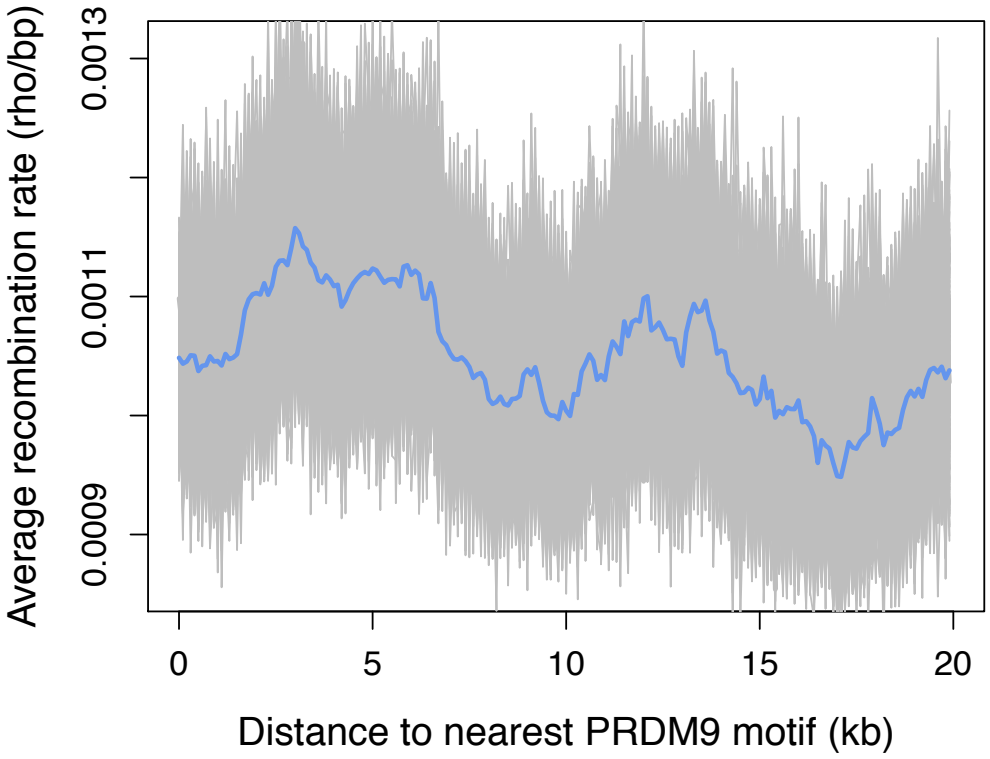

Fig. S29. Recombination rates are not locally elevated around computationally predicted PRDM9 binding sites. Shown here is recombination rate in 5 kb windows as a function of distance from predicted PRDM9 binding motifs in the *X. birchmanni* genome. The average rate is show in blue and results of 500 replicate bootstraps resampling the data in gray. The Spearman's correlation coefficient between distance from the predicted PRDM9 binding motif and rate is $\rho = -0.01$ ($p=0.19$).

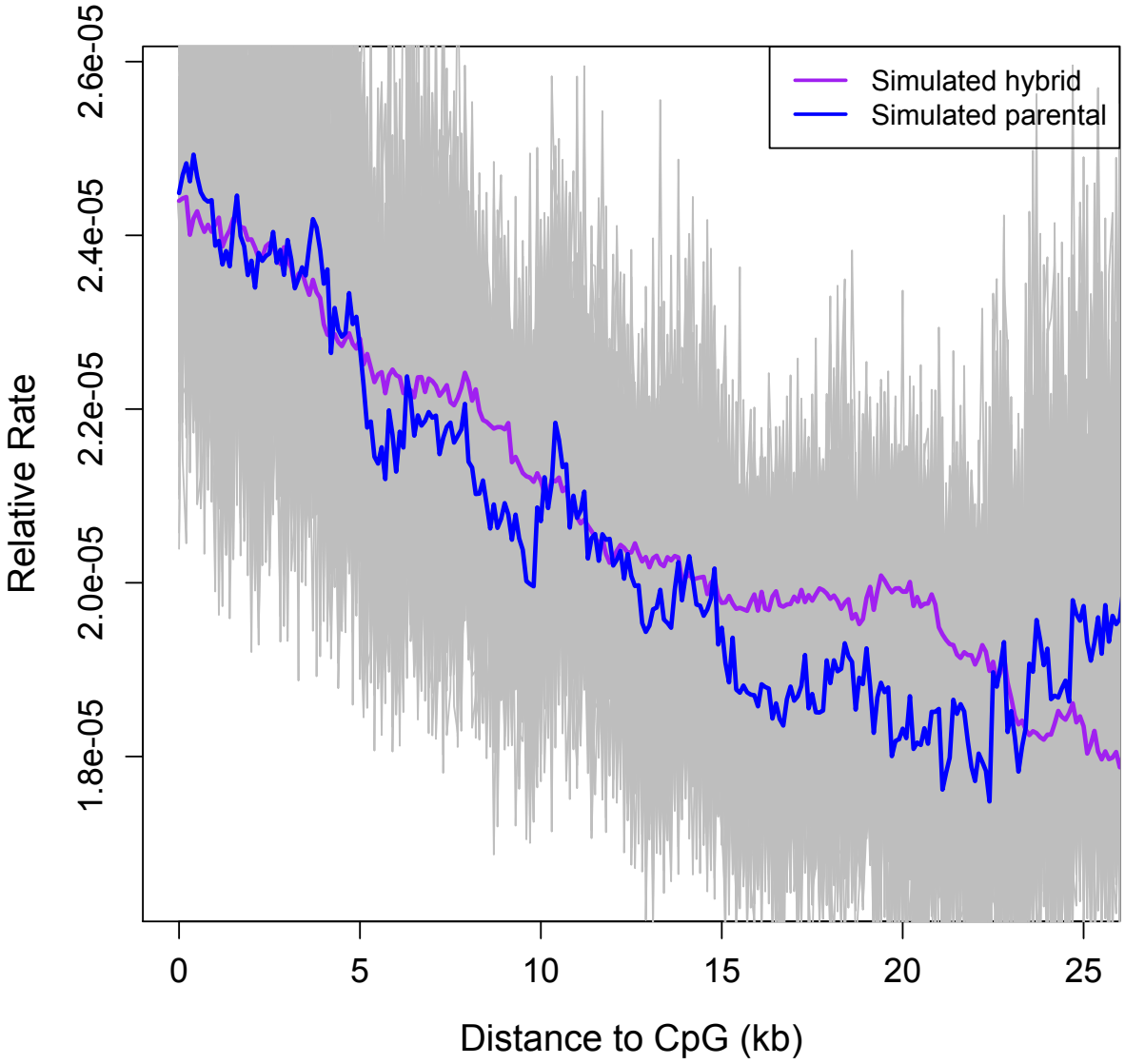

Fig. S30. No evidence that observed differences between parental and hybrid recombination maps are due to differences in power and errors in the two approaches (LD-based and ancestry-based) used to infer recombination rates. Simulations of LD and hybrid maps based on five replicate simulations of the first 10 Mb of chromosomes 1-5 show that the map inferences should not produce the observed depression in hybrid rates around CGIs (Fig. S2), at least as modeled; see Materials and Methods 1.7 for simulation details and discussion.

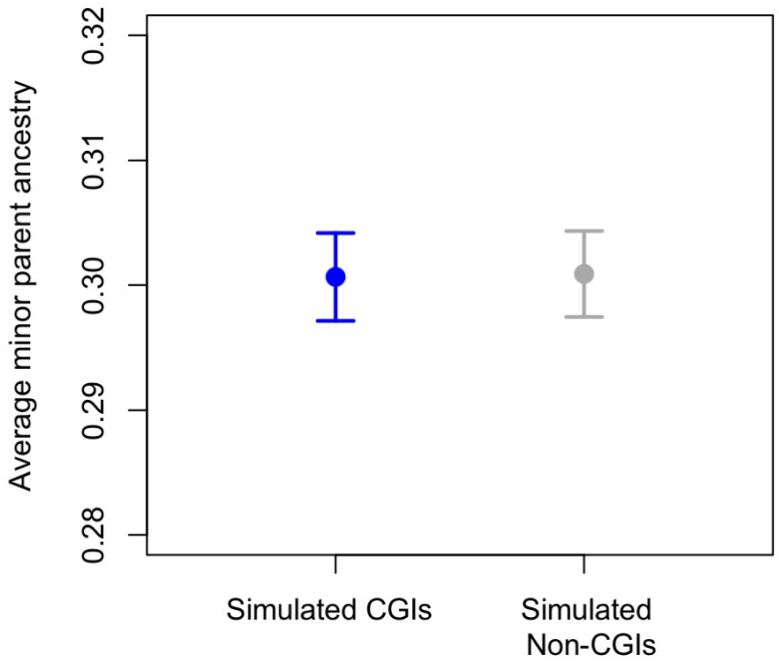

Fig. S31. Elevated minor parent ancestry is not expected around CpG islands due to strong genetic drift alone. To check whether genetic drift alone generates higher minor parent ancestry at windows that overlap with CpG islands, we relied on admix'em simulations of bottlenecked populations (described in Materials and Methods 1.5), which used the observed recombination rates from swordtail chromosomes 1 and 2. The non-CpG island comparison set used here was the same set of windows used for chromosomes 1 and 2 in the real data. Our results suggest that genetic drift alone does not lead to a signal of increased minor parent ancestry around CpG islands observed in swordtail hybrid populations (Fig. 4). Points show average minor parent ancestry and whiskers indicate two standard errors from 1,000 bootstraps resampling the data. The number of simulations sampled was chosen to mimic the amount of real data (Fig. 4).

4 Supplementary Tables

Table S1. Number of reads per individual, mapping statistics, and per bp coverage after alignment and filtering. Paired-end 150 bp reads were collected for individuals included for family sequencing; for all other individuals, we obtained paired-end 100 bp reads.

Individual id	Number of reads	Number reads mapped	Average coverage per basepair
COAC10F	191005850	184051735	18.6
COAC12F	191390021	185686732	21.3
COAC14F	184842328	179344024	19.7
COAC15F	172964665	165568247	16.7
COAC16F	212780097	204529674	21.5
COAC17F	210315871	202209760	18.8
COAC19F	182415266	175069512	16.7
COAC1F	210257267	203299687	21.9
COAC2010F06	259945683	252095137	26.7
COAC2010F09	162419570	153329498	15.3
COAC2010F18	273717455	266364176	27.9
COAC2010M01	185480458	178111891	16.1
COAC2010M02	230737120	222223567	20.9
COAC2010M03	263799977	255861972	26.5
COAC7F	186498562	180913544	19.6
COAC8F	206167003	200104405	21.3
COAC9F	295955137	278688874	26.2
COAC-ref	323438540	312972455	31.9
COAC-family-father	281936948	268019429	49.4
COAC-family-mother	240956081	229658118	43.5
COAC-family-offspring11	199819622	190226809	36.8
COAC-family-offspring12	212314024	203558618	38.6
COAC-family-offspring13	166798052	159974927	31.3
COAC-family-offspring14	243458992	233719087	43.4
COAC-family-offspring15	166730135	159946152	31.2

Table S2. Relationship between minor parent ancestry and recombination rate. Spearman's correlations between average minor parent ancestry and recombination rate at several scales, in swordtail hybrid populations and for archaic ancestry in the human genome. Also shown are partial correlation results, controlling for the number of coding base pairs (for conserved base pairs, see Table S3). P-values were obtained after thinning windows, to minimize correlations among nearby windows (see Materials and Methods 1.4).

Population	Spearman's correlation between minor ancestry and rate			Spearman's partial correlation between minor ancestry and rate, including coding base pairs		
	50 kb	**250 kb**	**500 kb**	**50 kb**	**250 kb**	**500 kb**
Swordtail population 1 (Totonicapa)	$\rho = 0.12$ $p = 10^{-8}$	$\rho = 0.13$ $p = 10^{-6}$	$\rho = 0.18$ $p = 10^{-9}$	$\rho = 0.12$ $p = 10^{-8}$	$\rho = 0.14$ $p = 10^{-6}$	$\rho = 0.18$ $p = 10^{-10}$
Swordtail population 2 (Aguazarca)	$\rho = 0.11$ $p = 10^{-5}$	$\rho = 0.10$ $p = 10^{-3}$	$\rho = 0.09$ $p = 0.001$	$\rho = 0.10$ $p = 10^{-5}$	$\rho = 0.10$ $p = 10^{-3}$	$\rho = 0.09$ $p = 0.001$
Swordtail population 3 (Tlatemaco)	$\rho = 0.08$ $p = 10^{-4}$	$\rho = 0.10$ $p = 10^{-3}$	$\rho = 0.11$ $p = 10^{-4}$	$\rho = 0.08$ $p = 10^{-4}$	$\rho = 0.10$ $p = 10^{-3}$	$\rho = 0.11$ $p = 10^{-4}$
Neanderthal ancestry in humans (*80*)	$\rho = 0.09$ $p = 10^{-17}$	$\rho = 0.17$ $p = 10^{-33}$	$\rho = 0.19$ $p = 10^{-42}$	$\rho = 0.09$ $p = 10^{-16}$	$\rho = 0.17$ $p = 10^{-33}$	$\rho = 0.19$ $p = 10^{-43}$
Neanderthal ancestry in humans (*4*)	$\rho = 0.07$ $p = 10^{-9}$	$\rho = 0.14$ $p = 10^{-25}$	$\rho = 0.17$ $p = 10^{-33}$	$\rho = 0.06$ $p = 10^{-9}$	$\rho = 0.14$ $p = 10^{-25}$	$\rho = 0.17$ $p = 10^{-34}$
Neanderthal ancestry in humans (*87*)	$\rho = 0.07$ $p = 10^{-11}$	$\rho = 0.13$ $p = 10^{-21}$	$\rho = 0.15$ $p = 10^{-30}$	$\rho = 0.07$ $p = 10^{-11}$	$\rho = 0.13$ $p = 10^{-21}$	$\rho = 0.16$ $p = 10^{-30}$
Denisovan ancestry in humans (*75*)	$\rho = 0.02$ $p = 0.46$	$\rho = 0.04$ $p = 10^{-3}$	$\rho = 0.07$ $p = 10^{-6}$	$\rho = 0.02$ $p = 0.62$	$\rho = 0.04$ $p = 10^{-3}$	$\rho = 0.07$ $p = 10^{-6}$
Denisovan ancestry in humans (*87*)	$\rho = 0.08$ $p = 10^{-14}$	$\rho = 0.14$ $p = 10^{-24}$	$\rho = 0.15$ $p = 10^{-29}$	$\rho = 0.08$ $p = 10^{-14}$	$\rho = 0.14$ $p = 10^{-24}$	$\rho = 0.15$ $p = 10^{-29}$

Table S3. Relationship between average minor parent ancestry and the recombination rate in partial correlation analyses including the number of conserved bps per window. Results for Spearman's partial correlations between average minor parent ancestry and recombination rate at several size scales in three swordtail fish hybrid populations, including as a covariate the number of conserved bps in that window. For analysis details see Materials and Methods 1.4; results for the same analysis with coding bps are reported in Table S2. P-values are obtained after thinning windows, to minimize correlations among nearby windows (see Materials and Methods 1.4).

Population	**Spearman's correlation between minor ancestry and rate including conserved base pairs**		
	50 kb	**250 kb**	**500 kb**
Swordtail population 1 (Totonicapa)	$\rho = 0.12$ $p = 10^{-8}$	$\rho = 0.14$ $p = 10^{-6}$	$\rho = 0.18$ $p = 10^{-10}$
Swordtail population 2 (Aguazarca)	$\rho = 0.11$ $p = 10^{-5}$	$\rho = 0.10$ $p = 0.001$	$\rho = 0.09$ $p = 0.001$
Swordtail population 3 (Tlatemaco)	$\rho = 0.09$ $p = 10^{-4}$	$\rho = 0.09$ $p = 0.001$	$\rho = 0.11$ $p = 10^{-4}$

Table S4. Spearman's correlations between average Neanderthal ancestry and recombination rate, using different human recombination maps. Maps were downloaded from http://www.well.ox.ac.uk/~anjali/AAmap/. Correlations listed here are based on Neanderthal ancestry estimates from diCal-admix (*79-80*). P-values are obtained after thinning windows, to minimize correlations among nearby windows (see Materials and Methods 1.4).

Size scale	deCode map	YRI LD map	African-American map
50 kb	$\rho = 0.06$ $p = 10^{-9}$	$\rho = 0.08$ $p = 10^{-12}$	$\rho = 0.07$ $p = 10^{-12}$
250 kb	$\rho = 0.13$ $p = 10^{-21}$	$\rho = 0.17$ $p = 10^{-36}$	$\rho = 0.16$ $p = 10^{-32}$
500 kb	$\rho = 0.15$ $p = 10^{-28}$	$\rho = 0.19$ $p = 10^{-44}$	$\rho = 0.18$ $p = 10^{-50}$

Table S5. Spearman's correlation between average minor parent ancestry in 50 kb windows and linked coding (or conserved bps) within 0.1 cM upstream and downstream of that window in swordtail hybrid populations. Note that analyses shown here are for all windows (i.e., not thinned); relationships become weaker and in some cases non-significant with thinning. See Materials and Methods 1.4 for details.

Population	Minor parent ancestry and linked coding base pairs	Minor parent ancestry and linked conserved base pairs
Swordtail population 1 (Totonicapa)	ρ = -0.1 p = 10^{-24}	ρ = -0.12 p = 10^{-36}
Swordtail population 2 (Aguazarca)	ρ = -0.05 p = 10^{-6}	ρ = -0.09 p = 10^{-16}
Swordtail population 3 (Tlatemaco)	ρ = -0.05 p = 10^{-6}	ρ = -0.1 p = 10^{-24}

Table S6. Spearman's correlation between average minor parent ancestry and recombination rate excluding 50 kb windows in the lowest 25% quantile of recombination rate and windows in the highest 25% quantile of recombination rate. Note that analyses shown here are for unthinned windows (see Materials and Methods 1.4 for details).

Population	Exclude lowest 25% rate quantile	Exclude highest 25% rate quantile
Swordtail population 1 (Totonicapa)	$\rho = 0.08$ $p = 10^{-13}$	$\rho = 0.09$ $p = 10^{-18}$
Swordtail population 2 (Aguazarca)	$\rho = 0.07$ $p = 10^{-8}$	$\rho = 0.06$ $p = 10^{-6}$
Swordtail population 3 (Tlatemaco)	$\rho = 0.06$ $p = 10^{-8}$	$\rho = 0.07$ $p = 10^{-11}$

Table S7. Relationship between average minor parent ancestry and recombination rate in simulations, as assessed by a partial correlation analysis including the number of coding bps in a window. Shown are the median Spearman's ρ from the 200 simulations (100 in the case of hybridization load) and the fraction of simulations that were significant at the 5% level. In these simulations, which are described in Materials and Methods 1.5, selected sites were placed only in exons. P-values are obtained after thinning windows, to minimize correlations among nearby windows (see Materials and Methods 1.4).

Simulation scenario	**Spearman's partial correlation between minor ancestry and rate including coding base pairs**		
	50 kb	**100 kb**	**250 kb**
BDMI simulation	median ρ_{rate} = 0.08 84% significant	median ρ_{rate} = 0.08 89% significant	median ρ_{rate} = 0.08 82% significant
	median ρ_{coding} = -0.05 43% significant	median ρ_{coding} = -0.06 74% significant	median ρ_{coding} = -0.09 83% significant
Hybridization load simulation	median ρ_{rate} = 0.10 87% significant	median ρ_{rate} = 0.12 100% significant	median ρ_{rate} = 0.17 99% significant
	median ρ_{coding} = -0.07 94% significant	median ρ_{coding} = -0.08 97% significant	median ρ_{coding} = -0.15 100% significant
Ecological selection simulation	median ρ_{rate} = 0.11 95% significant	median ρ_{rate} = 0.12 99% significant	median ρ_{rate} = 0.13 99% significant
	median ρ_{coding} = -0.07 90% significant	median ρ_{coding} = -0.09 97% significant	median ρ_{coding} = -0.16 99% significant

Table S8. Relationships between average minor parent ancestry and the number of coding (or conserved) bps in physical windows and the local recombination rate. Shown for three swordtail fish hybrid populations is the Spearman's partial correlation of the average minor parent ancestry and the number of coding or conserved bps at several size scales, including the recombination rate as a covariate (for details, see Materials and Methods 1.4). Analogous analyses of the effect of recombination rates are reported in Table S2 and Table S3; correlations between average minor parent ancestry and the number of coding bps in a genetic window are shown in Table S5. P-values shown here are based on all windows (i.e., not thinned; see Materials and Methods 1.4).

Population	**Spearman's correlation between minor ancestry and coding (conserved) bps including recombination rate**		
	50 kb	**250 kb**	**500 kb**
Population 1			
Coding bps	ρ = -0.01 p = 0.3	ρ = -0.03 p = 0.2	ρ = -0.03 p = 0.3
Conserved bps	ρ = -0.02 p= 0.03	ρ = -0.03 p= 0.12	ρ = -0.03 p= 0.25
Population 2			
Coding bps	ρ = -0.017 p= 0.08	ρ = -0.03 p = 0.2	ρ = -0.001 p = 0.9
Conserved bps	ρ = -0.02 p = 0.04	ρ = -0.05 p = 0.02	ρ = -0.07 p= 0.02
Population 3			
Coding bps	ρ = 0.02 p = 0.03	ρ = 0.01 p = 0.63	ρ = 0.001 p = 0.9
Conserved bps	ρ = -0.04 p = 0.0003	ρ = -0.04 p = 0.08	ρ = -0.06 p = 0.08

References and Notes

1. J. A. Coyne, H. A. Orr, *Speciation* (Sinaeur Associates, 2004).

2. L. H. Rieseberg, J. Whitton, K. Gardner, Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**, 713–727 (1999). Medline

3. R. Cui, M. Schumer, K. Kruesi, R. Walter, P. Andolfatto, G. G. Rosenthal, Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution* **67**, 2166–2179 (2013). doi:10.1111/evo.12099 Medline

4. S. Sankararaman, S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, D. Reich, The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014). doi:10.1038/nature12961 Medline

5. F. Jacobsen, K. E. Omland, Increasing evidence of the role of gene flow in animal evolution: Hybrid speciation in the yellow-rumped warbler complex. *Mol. Ecol.* **20**, 2236–2239 (2011). doi:10.1111/j.1365-294X.2011.05120.x Medline

6. I. Juric, S. Aeschbacher, G. Coop, The strength of selection against Neanderthal introgression. *PLOS Genet.* **12**, e1006340 (2016). doi:10.1371/journal.pgen.1006340 Medline

7. M. Schumer, R. Cui, D. L. Powell, G. G. Rosenthal, P. Andolfatto, Ancient hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* **25**, 2661–2679 (2016). doi:10.1111/mec.13602 Medline

8. J. P. Masly, D. C. Presgraves, High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLOS Biol.* **5**, e243 (2007). doi:10.1371/journal.pbio.0050243 Medline

9. K. Bomblies, J. Lempe, P. Epple, N. Warthmann, C. Lanz, J. L. Dangl, D. Weigel, Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLOS Biol.* **5**, e236 (2007). doi:10.1371/journal.pbio.0050236 Medline

10. H.-Y. Lee, J.-Y. Chou, L. Cheong, N.-H. Chang, S.-Y. Yang, J.-Y. Leu, Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**, 1065–1073 (2008). doi:10.1016/j.cell.2008.10.047 Medline

11. K. Harris, R. Nielsen, The genetic cost of Neanderthal introgression. *Genetics* **203**, 881–891 (2016). doi:10.1534/genetics.116.186890 Medline

12. N. Bierne, T. Lenormand, F. Bonhomme, P. David, Deleterious mutations in a hybrid zone: Can mutational load decrease the barrier to gene flow? *Genet. Res.* **80**, 197–204 (2002). doi:10.1017/S001667230200592X Medline

13. L. H. Rieseberg, O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, C. Lexer, Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211–1216 (2003). doi:10.1126/science.1086949 Medline

14. M. E. Arnegard, M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte, S. Kabir, N. Bedford, S. Bergek, Y. F. Chan, F. C. Jones, D. M. Kingsley, C. L. Peichel, D. Schluter,

Genetics of ecological divergence during speciation. *Nature* **511**, 307–311 (2014). doi:10.1038/nature13301 Medline

15. C. I. Wu, The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001). doi:10.1046/j.1420-9101.2001.00335.x

16. M. A. F. Noor, K. L. Grams, L. A. Bertucci, J. Reiland, Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12084–12088 (2001). doi:10.1073/pnas.221274498 Medline

17. M. W. Nachman, B. A. Payseur, Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. London Ser. B* **367**, 409–421 (2012). doi:10.1098/rstb.2011.0249 Medline

18. Y. Brandvain, A. M. Kenney, L. Flagel, G. Coop, A. L. Sweigart, Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLOS Genet.* **10**, e1004410 (2014). doi:10.1371/journal.pgen.1004410 Medline

19. M. Carneiro, N. Ferrand, M. W. Nachman, Recombination and speciation: Loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* **181**, 593–606 (2009). doi:10.1534/genetics.108.096826 Medline

20. A. Geraldes, P. Basset, K. L. Smith, M. W. Nachman, Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Mol. Ecol.* **20**, 4722–4736 (2011). doi:10.1111/j.1365-294X.2011.05285.x Medline

21. Materials and methods are available as supplementary materials.

22. M. Schumer, R. Cui, D. L. Powell, R. Dresner, G. G. Rosenthal, P. Andolfatto, High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *eLife* **3**, e02535 (2014). doi:10.7554/eLife.02535 Medline

23. M. Schumer, Y. Brandvain, Determining epistatic selection in admixed populations. *Mol. Ecol.* **25**, 2577–2591 (2016). doi:10.1111/mec.13641 Medline

24. P. Andolfatto, D. Davison, D. Erezyilmaz, T. T. Hu, J. Mast, T. Sunayama-Morita, D. L. Stern, Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* **21**, 610–617 (2011). doi:10.1101/gr.115402.110 Medline

25. Z. Baker, M. Schumer, Y. Haba, L. Bashkirova, C. Holland, G. G. Rosenthal, M. Przeworski, Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife* **6**, e24133 (2017). doi:10.7554/eLife.24133 Medline

26. R. Do, D. Balick, H. Li, I. Adzhubei, S. Sunyaev, D. Reich, No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat. Genet.* **47**, 126–131 (2015). doi:10.1038/ng.3186 Medline

27. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E.

Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014). doi:10.1038/nature12886 Medline

28. H. A. Orr, M. Turelli, The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. *Evolution* **55**, 1085–1094 (2001). doi:10.1111/j.0014-3820.2001.tb00628.x Medline

29. B. Davies, E. Hatton, N. Altemose, J. G. Hussin, F. Pratto, G. Zhang, A. G. Hinch, D. Moralli, D. Biggs, R. Diaz, C. Preece, R. Li, E. Bitoun, K. Brick, C. M. Green, R. D. Camerini-Otero, S. R. Myers, P. Donnelly, Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* **530**, 171–176 (2016). doi:10.1038/nature16931 Medline

30. S. Picelli, Å. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn*5* transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014). doi:10.1101/gr.177881.114 Medline

31. M. Schumer, R. Cui, G. Rosenthal, P. Andolfatto, simMSG: an experimental design tool for high-throughput genotyping of hybrids. *Mol. Ecol. Resour.* **16**, 183–192 (2015). doi:10.1111/1755-0998.12434

32. M. Schumer, D. L. Powell, P. J. Delclós, M. Squire, R. Cui, P. Andolfatto, G. G. Rosenthal, Assortative mating and persistent reproductive isolation in hybrids. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10936–10941 (2017). Medline

33. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* **8**, e1002967–e1002967 (2012). doi:10.1371/journal.pgen.1002967 Medline

34. M. Schartl, R. B. Walter, Y. Shen, T. Garcia, J. Catchen, A. Amores, I. Braasch, D. Chalopin, J.-N. Volff, K.-P. Lesch, A. Bisazza, P. Minx, L. Hillier, R. K. Wilson, S. Fuerstenberg, J. Boore, S. Searle, J. H. Postlethwait, W. C. Warren, The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat. Genet.* **45**, 567–572 (2013). doi:10.1038/ng.2604 Medline

35. M. A. Quail, H. Swerdlow, D. J. Turner, Improved protocols for the Illumina Genome Analyzer sequencing system. *Curr. Protoc. Hum. Genet.* **62**, 18.2.1–18.2.27 (2009). doi:10.1002/0471142905.hg1802s62

36. J. W. Li, K. Robison, M. Martin, A. Sjödin, B. Usadel, M. Young, E. C. Olivares, D. M. Bolser, The SEQanswers wiki: A wiki database of tools for high-throughput sequencing analysis. *Nucleic Acids Res.* **40** (D1), D1313–D1317 (2012). doi:10.1093/nar/gkr1058 Medline

37. A. Amores, J. Catchen, I. Nanda, W. Warren, R. Walter, M. Schartl, J. H. Postlethwait, A RAD-tag genetic map for the platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish. *Genetics* **197**, 625–641 (2014). doi:10.1534/genetics.114.164293 Medline

38. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp324 Medline

39. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi:10.1101/gr.107524.110 Medline

40. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi:10.1086/519795 Medline

41. D. M. Altshuler, R. A. Gibbs, L. Peltonen, D. M. Altshuler, R. A. Gibbs, L. Peltonen, E. Dermitzakis, S. F. Schaffner, F. Yu, L. Peltonen, E. Dermitzakis, P. E. Bonnen, D. M. Altshuler, R. A. Gibbs, P. I. W. de Bakker, P. Deloukas, S. B. Gabriel, R. Gwilliam, S. Hunt, M. Inouye, X. Jia, A. Palotie, M. Parkin, P. Whittaker, F. Yu, K. Chang, A. Hawes, L. R. Lewis, Y. Ren, D. Wheeler, R. A. Gibbs, D. M. Muzny, C. Barnes, K. Darvishi, M. Hurles, J. M. Korn, K. Kristiansson, C. Lee, S. A. McCarrol, J. Nemesh, E. Dermitzakis, A. Keinan, S. B. Montgomery, S. Pollack, A. L. Price, N. Soranzo, P. E. Bonnen, R. A. Gibbs, C. Gonzaga-Jauregui, A. Keinan, A. L. Price, F. Yu, V. Anttila, W. Brodeur, M. J. Daly, S. Leslie, G. McVean, L. Moutsianas, H. Nguyen, S. F. Schaffner, Q. Zhang, M. J. R. Ghori, R. McGinnis, W. McLaren, S. Pollack, A. L. Price, S. F. Schaffner, F. Takeuchi, S. R. Grossman, I. Shlyakhter, E. B. Hostetter, P. C. Sabeti, C. A. Adebamowo, M. W. Foster, D. R. Gordon, J. Licinio, M. C. Manca, P. A. Marshall, I. Matsuda, D. Ngare, V. O. Wang, D. Reddy, C. N. Rotimi, C. D. Royal, R. R. Sharp, C. Zeng, L. D. Brooks, J. E. McEwen, International HapMap 3 Consortium, Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010). doi:10.1038/nature09298 Medline

42. A. H. Chan, P. A. Jenkins, Y. S. Song, Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1003090 (2012). doi:10.1371/journal.pgen.1003090 Medline

43. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006). doi:10.1093/bioinformatics/btl446 Medline

44. A. Siepel, D. Haussler, Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* **21**, 468–488 (2004). doi:10.1093/molbev/msh039 Medline

45. G. K. Chen, P. Marjoram, J. D. Wall, Fast and flexible simulation of DNA sequence data. *Genome Res.* **19**, 136–142 (2009). doi:10.1101/gr.083634.108 Medline

46. M. Schumer, R. Cui, B. Boussau, R. Walter, G. Rosenthal, P. Andolfatto, An evaluation of the hybrid speciation hypothesis for *Xiphophorus clemenciae* based on whole genome sequences. *Evolution* **67**, 1155–1168 (2013). doi:10.1111/evo.12009 Medline

47. A. Rambaut, N. C. Grassly, Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**, 235–238 (1997).

48. J. O'Connell, D. Gurdasani, O. Delaneau, N. Pirastu, S. Ulivi, M. Cocca, M. Traglia, J. Huang, J. E. Huffman, I. Rudan, R. McQuillan, R. M. Fraser, H. Campbell, O. Polasek, G. Asiki, K. Ekoru, C. Hayward, A. F. Wright, V. Vitart, P. Navarro, J.-F. Zagury, J. F.

Wilson, D. Toniolo, P. Gasparini, N. Soranzo, M. S. Sandhu, J. Marchini, A general approach for haplotype phasing across the full spectrum of relatedness. *PLOS Genet.* **10**, e1004234 (2014). doi:10.1371/journal.pgen.1004234 Medline

49. B. N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genet.* **5**, e1000529 (2009). doi:10.1371/journal.pgen.1000529 Medline

50. A. L. Williams, D. E. Housman, M. C. Rinard, D. K. Gifford, Rapid haplotype inference for nuclear families. *Genome Biol.* **11**, R108 (2010). doi:10.1186/gb-2010-11-10-r108 Medline

51. A. Auton, G. McVean, Recombination rate estimation in the presence of hotspots. *Genome Res.* **17**, 1219–1227 (2007). doi:10.1101/gr.6386707 Medline

52. C. Feng, M. Pettersson, S. Lamichhaney, C.-J. Rubin, N. Rafati, M. Casini, A. Folkvord, L. Andersson, Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. *eLife* **6**, e23907 (2017). doi:10.7554/eLife.23907 Medline

53. C. I. Wu, W. H. Li, Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1741–1745 (1985). doi:10.1073/pnas.82.6.1741 Medline

54. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011). doi:10.1038/nature10231 Medline

55. C. A. Walling, N. J. Royle, N. B. Metcalfe, J. Lindström, Green swordtails alter their age at maturation in response to the population level of male ornamentation. *Biol. Lett.* **3**, 144–146 (2007). doi:10.1098/rsbl.2006.0608 Medline

56. J. A. Kamm, J. P. Spence, J. Chan, Y. S. Song, Two-locus likelihoods under variable population size and fine-scale recombination rate estimation. *Genetics* **203**, 1381–1399 (2016). doi:10.1534/genetics.115.184820 Medline

57. Y. B. Simons, G. Sella, The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Curr. Opin. Genet. Dev.* **41**, 150–158 (2016). doi:10.1016/j.gde.2016.09.006 Medline

58. R. Cui, M. Schumer, G. G. Rosenthal, Admix'em: A flexible framework for forward-time simulations of hybrid populations with selection and mate choice. *Bioinformatics* **32**, 1103–1105 (2016). doi:10.1093/bioinformatics/btv700 Medline

59. M. Schumer, R. Cui, G. G. Rosenthal, P. Andolfatto, Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLOS Genet.* **11**, e1005041 (2015). doi:10.1371/journal.pgen.1005041 Medline

60. D. C. Presgraves, A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics* **163**, 955–972 (2003). Medline

61. A. L. Sweigart, Simple Y-autosomal incompatibilities cause hybrid male sterility in reciprocal crosses between *Drosophila virilis* and *D. americana*. *Genetics* **184**, 779–787 (2010). doi:10.1534/genetics.109.112896 Medline

62. S. Maheshwari, D. A. Barbash, The genetics of hybrid incompatibilities. *Annu. Rev. Genet.* **45**, 331–355 (2011). doi:10.1146/annurev-genet-110410-132514 Medline

63. Z. W. Culumber, H. S. Fisher, M. Tobler, M. Mateos, P. H. Barber, M. D. Sorenson, G. G. Rosenthal, Replicated hybrid zones of *Xiphophorus* swordtails along an elevational gradient. *Mol. Ecol.* **20**, 342–356 (2011). doi:10.1111/j.1365-294X.2010.04949.x Medline

64. Z. W. Culumber, D. B. Shepard, S. W. Coleman, G. G. Rosenthal, M. Tobler, Physiological adaptation along environmental gradients and replicated hybrid zone structure in swordtails (Teleostei: *Xiphophorus*). *J. Evol. Biol.* **25**, 1800–1814 (2012). doi:10.1111/j.1420-9101.2012.02562.x Medline

65. Y. F. Yang, W. Cao, S. Wu, W. Qian, Genetic interaction network as an important determinant of gene order in genome evolution. *Mol. Biol. Evol.* **34**, 3254–3266 (2017). doi:10.1093/molbev/msx264 Medline

66. L. D. Hurst, C. Pál, M. J. Lercher, The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**, 299–310 (2004). doi:10.1038/nrg1319 Medline

67. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi:10.1093/bioinformatics/btq033 Medline

68. S. Neph, M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, M. T. Maurano, J. Vierstra, S. Thomas, R. Sandstrom, R. Humbert, J. A. Stamatoyannopoulos, BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012). doi:10.1093/bioinformatics/bts277 Medline

69. S. Singhal, E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean, M. Przeworski, Stable recombination hotspots in birds. *Science* **350**, 928–932 (2015). doi:10.1126/science.aad0843 Medline

70. I. Lam, S. Keeney, Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. *Science* **350**, 932–937 (2015). doi:10.1126/science.aad0814 Medline

71. A. V. Persikov, M. Singh, De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* **42**, 97–108 (2014). doi:10.1093/nar/gkt890 Medline

72. T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S. Noble, MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37** (Web Server), W202–W208 (2009). doi:10.1093/nar/gkp335 Medline

73. S. Gravel, Population genetics models of local ancestry. *Genetics* **191**, 607–619 (2012). doi:10.1534/genetics.112.139808 Medline

74. S. Sankararaman, N. Patterson, H. Li, S. Pääbo, D. Reich, The date of interbreeding between Neandertals and modern humans. *PLOS Genet.* **8**, e1002947 (2012). doi:10.1371/journal.pgen.1002947 Medline

75. S. Sankararaman, S. Mallick, N. Patterson, D. Reich, The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr. Biol.* **26**, 1241–1247 (2016). doi:10.1016/j.cub.2016.03.037 Medline

76. G. McVicker, D. Gordon, C. Davis, P. Green, Widespread genomic signatures of natural selection in hominid evolution. *PLOS Genet.* **5**, e1000471 (2009). doi:10.1371/journal.pgen.1000471 Medline

77. M. Dannemann, A. M. Andrés, J. Kelso, Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human Toll-like receptors. *Am. J. Hum. Genet.* **98**, 22–33 (2016). doi:10.1016/j.ajhg.2015.11.015 Medline

78. D. Enard, D. Petrov, RNA viruses drove adaptive introgressions between Neanderthals and modern humans. *bioRxiv* 120477 [Preprint]. 24 March 2017. doi:10.1101/120477

79. M. Steinrücken, J. Kamm, Y. Song, Inference of complex population histories using whole-genome sequences from multiple populations. *bioRxiv* 026591 [Preprint]. 16 September 2015. doi:10.1101/026591

80. M. Steinrücken, J. P. Spence, J. A. Kamm, E. Wieczorek, Y. S. Song, Model-based detection and analysis of introgressed Neanderthal ancestry in modern humans. *Mol. Ecol.* 10.1111/mec.14565 (2018). doi:10.1111/mec.14565

81. A. G. Hinch, A. Tandon, N. Patterson, Y. Song, N. Rohland, C. D. Palmer, G. K. Chen, K. Wang, S. G. Buxbaum, E. L. Akylbekova, M. C. Aldrich, C. B. Ambrosone, C. Amos, E. V. Bandera, S. I. Berndt, L. Bernstein, W. J. Blot, C. H. Bock, E. Boerwinkle, Q. Cai, N. Caporaso, G. Casey, L. A. Cupples, S. L. Deming, W. R. Diver, J. Divers, M. Fornage, E. M. Gillanders, J. Glessner, C. C. Harris, J. J. Hu, S. A. Ingles, W. Isaacs, E. M. John, W. H. Kao, B. Keating, R. A. Kittles, L. N. Kolonel, E. Larkin, L. Le Marchand, L. H. McNeill, R. C. Millikan, A. Murphy, S. Musani, C. Neslund-Dudas, S. Nyante, G. J. Papanicolaou, M. F. Press, B. M. Psaty, A. P. Reiner, S. S. Rich, J. L. Rodriguez-Gil, J. I. Rotter, B. A. Rybicki, A. G. Schwartz, L. B. Signorello, M. Spitz, S. S. Strom, M. J. Thun, M. A. Tucker, Z. Wang, J. K. Wiencke, J. S. Witte, M. Wrensch, X. Wu, Y. Yamamura, K. A. Zanetti, W. Zheng, R. G. Ziegler, X. Zhu, S. Redline, J. N. Hirschhorn, B. E. Henderson, H. A. Taylor Jr., A. L. Price, H. Hakonarson, S. J. Chanock, C. A. Haiman, J. G. Wilson, D. Reich, S. R. Myers, The landscape of recombination in African Americans. *Nature* **476**, 170–175 (2011). doi:10.1038/nature10336 Medline

82. L. S. Stevison, A. E. Woerner, J. M. Kidd, J. L. Kelley, K. R. Veeramah, K. F. McManus, C. D. Bustamante, M. F. Hammer, J. D. Wall; Great ape genome project, the time scale of recombination rate evolution in great apes. *Mol. Biol. Evol.* **33**, 928–945 (2016). doi:10.1093/molbev/msv331 Medline

83. A. Auton, A. Fledel-Alon, S. Pfeifer, O. Venn, L. Ségurel, T. Street, E. M. Leffler, R. Bowden, I. Aneas, J. Broxholme, P. Humburg, Z. Iqbal, G. Lunter, J. Maller, R. D. Hernandez, C. Melton, A. Venkat, M. A. Nobrega, R. Bontrop, S. Myers, P. Donnelly, M. Przeworski, G. McVean, A fine-scale chimpanzee genetic map from population sequencing. *Science* **336**, 193–198 (2012). doi:10.1126/science.1216872 Medline

84. F. Racimo, S. Sankararaman, R. Nielsen, E. Huerta-Sánchez, Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359–371 (2015). doi:10.1038/nrg3936 Medline

85. B. Vernot, J. M. Akey, Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017–1021 (2014). doi:10.1126/science.1245938 Medline

86. B. Vernot, S. Tucci, J. Kelso, J. G. Schraiber, A. B. Wolf, R. M. Gittelman, M. Dannemann, S. Grote, R. C. McCoy, H. Norton, L. B. Scheinfeldt, D. A. Merriwether, G. Koki, J. S. Friedlaender, J. Wakefield, S. Pääbo, J. M. Akey, Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016). doi:10.1126/science.aad9416 Medline

87. L. Skov, R. Hui, A. Hobolth, A. Scally, M. H. Schierup, R. Durbin, Detecting archaic introgression without archaic reference genomes. *bioRxiv* 283606 [Preprint]. 23 March 2018. doi:10.1101/283606

88. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Y. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010). doi:10.1126/science.1188021 Medline

89. S. Gavrilets, Hybrid zones with Dobzhansky-type epistatic selection. *Evolution* **51**, 1027–1035 (1997). doi:10.1111/j.1558-5646.1997.tb03949.x Medline