Title:
Predictability and parallelism in the contemporary evolution of hybrid genomes

Authors: Quinn K. Langdon[1,2*], Daniel L. Powell[1,2], Bernard Kim[1], Shreya M. Banerjee[1,2], Cheyenne Payne[1,2], Tristram O. Dodge[1,2], Ben Moran[1,2], Paola Fascinetto-Zago[2,3], Molly Schumer[1,2,4*]

[1]Department of Biology, Stanford University
[2]Centro de Investigaciones Científicas de las Huastecas "Aguazarca", A.C.
[3]Department of Biology, Texas A&M University
[4]Hanna H. Gray Fellow, Howard Hughes Medical Institutes

*Correspondence to: qlangdon@stanford.edu and schumer@stanford.edu

Abstract

Hybridization between species is widespread across the tree of life. As a result, many species, including our own, harbor regions of their genome derived from hybridization. Despite the recognition that this process is widespread, we understand little about how the genome stabilizes following hybridization, and whether the mechanisms driving this stabilization tend to be shared across species. Here, we dissect the drivers of variation in local ancestry across the genome in replicated hybridization events between two species pairs of swordtail fish: *Xiphophorus birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche*. We find unexpectedly high levels of repeatability in local ancestry across the two types of hybrid populations. This repeatability is attributable in part to the fact that the recombination landscape and locations of functionally important elements play a major role in driving variation in local ancestry in both types of hybrid populations. Beyond these broad scale patterns, we identify dozens of regions of the genome where minor parent ancestry is unusually low or high across species pairs. Analysis of these regions points to shared sites under selection across species pairs, and in some cases, shared mechanisms of selection. We show that one such region is a previously unknown hybrid incompatibility that is shared across *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations.

Introduction

In the past decade the number of species with sequenced genomes has risen exponentially [1–3]. This has allowed for a greater exploration of the evolutionary histories of the diversity of life on the planet. One consequence of this work has been a new understanding of the ubiquity of genetic exchange between species [4–6]. Studies have demonstrated the impact of both ancient and recent hybridization on the genomes of extant species from fruit flies to humans [7–14]. While it is now clear that hybridization is commonplace, what is less understood is how the genome changes after admixture and what genetic and evolutionary factors generally shape this process.

An important piece of predicting how the genome will respond after hybridization is understanding the sources of selection that act on hybrids. A large number of studies have shown that hybrids between different species tend to experience fitness consequences; they are often less capable of surviving and reproducing than the species that formed them [15–20]. Several types of selection on hybrids can drive this pattern. Decades of empirical and theoretical work have indicated that Dobzhansky-Muller hybrid incompatibilities (DMIs), or incompatible interactions between mutations that are derived in each of the parental species, are a common cause of inviability and infertility in hybrids [21–24]. Hybrids can also suffer from the effects of intermediate or transgressive traits that place them far from the phenotypic optimum for either parental species [25–28]. In both of these scenarios, selection is generally expected to act against alleles derived from the "minor" parent species, or the species that contributed less to the genome of the hybrids. This is because removing incompatible alleles derived from the minor parent species (or variants that move a hybrid further away from the major parent's phenotypic optimum) is the fastest route to restoring fitness [6]. In addition, recent work has indicated that another mechanism of selection against hybrids can stem from differences in historical effective populations size between the parental species. Such differences can result in the disproportionate accumulation of deleterious mutations along the parental lineage with the smaller historical effective population size and drive the removal of haplotypes from this parental species after hybridization [19,29]. However, this mechanism is most likely to impact hybrids that form between species with dramatically different historical effective population sizes [19,29].

Although it is well-established that hybrids between species frequently experience selection, it is less clear how predictable the outcomes of selection are at a genome-wide scale, both within and between species pairs. While it has been appreciated for decades that certain regions of the genome, such as sex chromosomes, tend to retain less minor parent ancestry [30–34], recent studies have highlighted other patterns that appear to be consistent across several independent hybridization events. These include broad selection against minor parent ancestry across the genome (presumably due to the mechanisms outlined above) and stronger depletion of minor parent ancestry in regions of the genome with a high density of conserved or coding elements [8,12,35,36]. Another key variable highlighted by empirical and theoretical studies is the importance of the interplay between recombination rate and selection, with several groups finding that selection on hybrids generates a positive correlation between local recombination rate and minor parent ancestry [8,35–41]. This correlation is thought to be driven by how the local recombination rate impacts the decoupling of neutral or adaptive alleles from alleles that are deleterious in hybrids. Neutral or adaptive minor parent alleles that occur in regions of the genome with high local recombination rates are less likely to remain linked to deleterious

neighbors, resulting in the retention of more minor parent haplotypes in such regions after selection.

These emerging findings highlight how certain features of genome organization can interact with selection after hybridization in a predictable way across diverse species [6]. This suggests that we can expect some broad scale repeatability in the outcomes of hybridization across species and the types of regions of the genome that are more or less permeable to introgression. In hybridization events between pairs of closely related species we may expect repeatability in the outcomes of genome evolution after hybridization to be even higher. This is in part because close relatives are expected to share genome composition and conserved local recombination maps (in some species; [42,43]). However, in close relatives, repeatability in the outcomes of hybridization could also be driven by shared mechanisms of selection acting on individual sites. For example, close relatives may share DMIs due to shared phylogenetic history [44] or similarities in underlying genetic networks [45] and among some species groups, certain alleles may be globally adaptive, leading to shared ancestry driven by adaptive introgression. Thus, leveraging comparisons of independent natural hybridization events between closely related species pairs is one promising approach for evaluating the factors that drive repeatable evolution after hybridization [12].

Recently formed hybrid populations are an especially powerful tool for addressing these questions because ancestry variation along the genome can reflect the action of strong selection on hybrids. Here, we leverage swordtail fish species (genus *Xiphophorus*) to investigate genome evolution after several recent, independent hybridization events. Natural replicate hybrid populations formed between *X. birchmanni* and *X. malinche* are an emerging model system for studying genome evolution after hybridization. Research on this sister species pair has shown that variation in local ancestry along the genome is best explained by the presence of prevalent hybrid incompatibilities (as opposed to historical effective populations size differences in the parental species or ecological drivers of ancestry variation; [35]), and has documented strong DMIs between them [23]. Here, we expand on our recent description of replicate hybrid populations of the species pair *X. birchmanni* × *X. cortezi* [46] to explore repeatability in local ancestry between these two types of hybrid populations.

X. birchmanni × *X. cortezi* are more deeply diverged than *X. birchmanni* × *X. malinche* (~450k generations diverged and ~250k generations diverged respectively) and unlike *X. birchmanni* and *X. malinche*, they do not differ in their time-averaged historical effective population sizes (Fig. S1) [46]. Hybrid populations between both species pairs formed recently (~100-150 generations ago [20,46]), likely due to human disturbance [47]. Comparisons across these two types of hybridization events provide one of the first windows into the drivers of evolution after hybridization in related species, allowing us to address questions about predictability and parallelism in selection on hybrids at both a local and broad-scale along the genome.

Results

Background

Throughout the manuscript we explore repeatability in genome evolution across two types of hybrid populations, those formed between the *X. birchmanni* × *X. cortezi* species pair and those formed between the *X. birchmanni* × *X. malinche* species pair. These three species are closely related, with divergence between *X. birchmanni* × *X. cortezi* slightly higher than divergence between *X. birchmanni* × *X. malinche* (Fig. 1; sequence divergence 0.6% and 0.4% per basepair respectively). Nucleotide diversity and inferred long-term effective population sizes are similar between *X. birchmanni* and *X. cortezi* ($\pi \sim 0.1\%$ per basepair; [46]) and lower in *X. malinche* ($\pi \sim 0.03\%$ per basepair; [35]) (Fig. S1). *X. birchmanni* is distributed between *X. malinche* and *X. cortezi*, with hybridization occurring in areas of the rivers where the species are sympatric (Fig. 1); *X. malinche* and *X. cortezi* are allopatric throughout their ranges. We have previously shown that replicate hybrid populations in both species pairs formed recently, within the last ~100-150 generations [20,35,46,48], likely due to human-mediated habitat disturbance [47].

For clarity, we refer to the hybridization events between each pair of species as the two "types" of hybridization events. In addition, we analyzed data from two *X. birchmanni* × *X. cortezi* (Huextetitla and Santa Cruz) and three *X. birchmanni* × *X. malinche* (Acuapa, Aguazarca, and Tlatemaco) hybrid populations, each of which occur in different rivers or tributaries (Fig. 1). When discussing comparisons within a hybridization type (e.g. *X. birchmanni* × *X. cortezi* or *X. birchmanni* × *X. malinche*) we refer to these geographically distinct samples as "replicate" hybrid populations. Past work in the *X. birchmanni* × *X. malinche* system has shown that the replicate hybrid populations formed independently [49]. Additionally, they also differ in their genome-wide ancestry: *X. birchmanni* is the minor parent in Tlatemaco, while *X. malinche* is the minor parent in Acuapa and Aguazarca (Fig. 1).

Despite the fact that similar factors (i.e. population disturbance) have likely driven recent hybridization across the two species pairs [46,47], each replicate population occurs in a distinct ecological environment, and varies in the degree of assortative mating observed [35,46,48,49]. For a more detailed description of what is known about the ecological environments and mating dynamics of each population from past work see Supporting Information 1.

The Santa Cruz and Huextetitla hybrid populations occur along geographically separated tributaries of the same river (Fig. 1) and have similar genome-wide admixture proportions and ancestry structure [46]. Hybrids in both populations derive 85-90% of their genomes from the *X. cortezi* parent species (Fig. 1). Using a hidden Markov model (HMM) based approach, we collected low-coverage whole genome sequence data and inferred local ancestry across the genome in 254 hybrids from the two populations (see Methods). While we initially planned to use the two sites to explore repeatability in genome evolution within the replicate populations, we found evidence that they were not completely independent hybridization events (Fig. S2-S3; Supporting Information 2). Thus, we focus our discussion on the Santa Cruz hybrid population but present analyses for the Huextetitla hybrid population (for which results are qualitatively similar) in parts of the main text and supplement.

Inference of demographic histories of hybrid populations

Initial work investigating the demographic history of the *X. birchmanni* × *X. cortezi* hybrid populations at Santa Cruz and Huextetitla [46] suggested that they formed recently and

experience low levels of migration from the parental species [46]. In this paper we expanded our sampling from previous work [46] and explored the demographic history of *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations in more detail. We used simulations and an approximate Bayesian computation approach (ABC, Supporting Information 3) to infer the demographic history of two populations: the Santa Cruz and Acuapa hybrid populations. Drawing from uniform (or log-uniform) prior distributions for parameters of interest (initial admixture proportions, time since initial admixture, population size, and migration rates), we conducted simulations in SLiM and performed rejection sampling based on summary statistics derived from the observed data (Supporting Information 3). These simulations yielded well-resolved posterior distributions for most parameters in both hybrid populations (Fig. S4). We discuss these results in more detail in Supporting Information 3. Briefly, our findings are consistent with our previous work [46]; both hybrid zones are young and experience limited migration from the parental species, but the *X. birchmanni* × *X. cortezi* hybrid population formed less recently (~120 generations ago) than the *X. birchmanni* × *X. malinche* (~80 generations ago) hybrid population (Fig. S4). Together, our ABC results allow us to model plausible demographic histories for each type of hybrid population to explore expected patterns of genome evolution.

Predictable genomic features correlate with minor parent ancestry in X. birchmanni × X. cortezi hybrid populations

Although admixture began recently between *X. birchmanni* and *X. cortezi*, we observe substantial heterogeneity in local ancestry along the genome, with regions fixed for both major and minor parent ancestry (e.g. Fig. S3). While severe bottlenecks could also generate fixation in recently formed hybrid populations ([46]), this variance in local ancestry could also be indicative of selection on hybrids. Indeed, simulations of variation in local ancestry using the demographic parameters we infer with ABC (admixture time, admixture proportion, and hybrid population size) suggest that the patterns of fixation in major and minor parent ancestry we observe are unexpected under the most likely scenario of neutral admixture (Fig. S5; Supporting Information 3).

We next evaluated correlations between minor parent ancestry and genomic features known to interact with selection in other hybrid systems. Studies in a number of systems have found that selection on hybrids can drive a positive correlation between recombination rate and minor parent ancestry [8,35,36,40]. Using averages of minor parent ancestry and recombination rate summarized in non-overlapping windows we found strong evidence that recombination rate is positively correlated with minor parent ancestry at both a local (Fig. 2; Santa Cruz – 100 kb windows: $\rho = 0.44$, $p<10^{-300}$; Huextetitla – 100 kb windows: $\rho = 0.42$, $p<10^{-275}$) and chromosome-wide scale (Supporting Information 4; Fig. S6-7) in replicate hybrid populations of *X. birchmanni* × *X. cortezi*. This pattern is robust across spatial scales (Table S1) and when controlling for a number of possible technical variables; e.g. thinning windows based on linkage disequilibrium and use of recombination rate estimates from observed crossovers in F_2s (see Table S2-S3).

We also investigated how ancestry in *X. birchmanni* × *X. cortezi* hybrid populations covaries with coding and conserved basepairs across the genome. Based on our previous findings in *X. birchmanni* × *X. malinche* hybrid populations [35], we predicted that regions with more coding and conserved basepairs would be depleted in minor parent ancestry. We summarized average minor parent ancestry and number of coding and conserved basepairs in non-overlapping genetic windows, to account for variation in recombination rate, and found depletion of minor

parent ancestry in regions especially dense in coding and conserved basepairs (Fig. 2; Table S4; Santa Cruz – 0.1 cM windows coding basepairs: $\rho = -0.15$, $p<10^{-73}$, conserved basepairs: $\rho = -0.22$, $p<10^{-161}$; Huextetitla – 0.1 cM windows: coding basepairs: $\rho = -0.09$, $p<10^{-27}$, conserved basepairs: $\rho = -0.16$, $p<10^{-79}$). This pattern is robust across a range of window sizes (Table S4), and when controlling for a number of potential technical artifacts (Tables S5-S9). Interestingly, we did not detect an impact of nonsynonymous substitutions between species on minor parent ancestry after controlling for local substitution rates (Supporting Information 5; Fig. S8).

Sources of selection driving variation in minor parent ancestry

Above we confirmed via simulations that the observed variation in local ancestry is not expected by chance or as a consequence of the demographic history of hybrid populations (Supporting Information 3), motivating us to ask about the role different sources of selection might play in driving the broad-scale ancestry patterns we observed in *X. birchmanni* × *X. cortezi* type and *X. birchmanni* × *X. malinche* type hybrid populations. To do so, we used simulations, grounded in the inferred demographic history of the hybrid populations (Supporting Information 3,6).

In our empirical data, minor parent ancestry is consistently depleted in regions of the genome with low recombination rate (and higher coding/conserved density), regardless of *which* species is the minor parent in a given population. Our simulations indicate that these positive correlations between minor parent ancestry and recombination rate across replicate populations and species pairs are best explained by a model where selection acts against minor parent ancestry at many sites across the genome. In simulations we use a model of selection against hybrid incompatibilities, which results in selection against minor parent ancestry at incompatible sites and has been well-documented in swordtails [20,23,35,50]. We note, however, that a subset of other models of selection on hybrids could generate similar results (see discussion in Supporting Information 6; reviewed in [6]). However, the patterns we observe are inconsistent with models of selection against hybridization load or global selection against one of the parental species (Fig. S9; Supporting Information 6). This latter finding indicates that selection on the *X. birchmanni* genome, which is a shared parent species in both hybrid population types, does not underlie the patterns that we document.

Repeatability extends across hybrid population types

Given evidence for selection against minor parent ancestry in hybrid populations of both types, the key question we wish to understand is how repeatable the outcomes of admixture are across different species pairs, and what mechanisms contribute to this repeatability. Past work has suggested that we expect repeatability in local ancestry in replicate hybrid populations formed between the same species pairs [35,51]. This repeatability often exceeds expectations from shared genetic architecture (e.g. recombination rate and functional basepair density), indicating that selection on the same sites in replicate hybrid populations drives concordance in local ancestry [35,51].

We wanted to evaluate whether there was evidence for a similar phenomenon across the two distinct types of hybridization events studied here. The *X. birchmanni* × *X. cortezi* hybridization events are independent from each of the three *X. birchmanni* × *X. malinche* hybrid populations (Fig. 1). In addition, the three replicate *X. birchmanni* × *X. malinche* hybrid populations we analyze (Acuapa, Aguazarca, and Tlatemaco) occur in geographically isolated river systems (Fig. 1) and formed independently of each other [49].

We observe repeatability in local patterns of minor parent ancestry across the genome between most pairs of hybrid populations formed between these distinct species pairs (in 100 kb non-overlapping windows $\rho_{\text{Santa Cruz - Acuapa}} = 0.25$, $p<10^{-93}$; $\rho_{\text{Santa Cruz - Aguazarca}} = 0.17$, $p<10^{-42}$; Fig. 2; Table S10), an intriguing observation given the demographic independence of these hybrid populations. These results were robust to a number of factors, including accounting for the locations of structural differences between the species (see Table S11; Supporting Information 4, 6). One exception to this pattern is in comparisons involving the Tlatemaco population. We speculate that this is driven by the fact that hybrids in this population derive the majority of their genome from *X. malinche*, a species with low historical effective population sizes (Fig. S1). We discuss this pattern in more detail in Supporting Information 6.

Because many features of the architecture of the genome, such as the local recombination rate and the locations of coding and conserved basepairs, are largely conserved across *X. birchmanni*, *X. malinche*, and *X. cortezi* (Supporting Information 4), we wanted to evaluate whether cross-population repeatability exceeded what was expected from these features alone. Partial correlation analysis of local ancestry across pairs of populations, including these features as covariates, indicated that some of the signal of shared local ancestry across hybrid population types can be explained by these features (Table S12). However, after accounting for this shared genomic architecture, we still observed repeatability in patterns of minor parent ancestry between most pairs of hybrid populations, including those formed between distinct species pairs (in 100 kb non-overlapping windows $\rho_{\text{Santa Cruz - Acuapa}} = 0.15$, $p<10^{-33}$; $\rho_{\text{Santa Cruz - Aguazarca}} = 0.09$, $p<10^{-15}$; Table S12). Simulations indicate that the observed correlations in local ancestry across the two types of hybridization events are unlikely to be driven by demographic history but can readily be generated by shared sites under selection (Fig. S10; Supporting Information 3,6).

Shared islands and deserts of minor parent ancestry across hybrid population types

The analyses described above indicate that not all cross-population repeatability in ancestry can be explained by shared genomic architecture between these closely related species pairs (Table S10, S12-S13; Supporting Information 6). We next explored how other factors, such as shared sites where minor parent ancestry is favored or disfavored, contributed to repeatability in local ancestry. Specifically, we asked whether there were certain regions of the genome with shared patterns of minor parent ancestry across hybrid population types, and whether we could identify the mechanisms of selection driving these shared patterns.

We first focused on the Santa Cruz *X. birchmanni* × *X. cortezi* population to identify regions of the genome with unusual minor parent ancestry. We identified minor parent "deserts" by scanning the genome for ancestry informative sites that fell in the lower 2.5% tail of genome-wide ancestry. We then expanded out from these sites to define regions of reduced minor parent ancestry (defined as those falling in the lower <5% tail of genome-wide ancestry; see Methods). These regions were then filtered to exclude very short ancestry tracts, those supported by few ancestry informative sites, and were merged with nearby deserts (see Methods; Fig. S11). This procedure resulted in the identification of 81 minor parent deserts with an average length of 317 kb spread across 18 of the 24 swordtail chromosomes. Of these 81 minor parent ancestry deserts, 21 (26%) overlapped with regions of low minor parent ancestry in one or more replicate *X. birchmanni* × *X. malinche* hybrid population (see Methods).

Using the same approach, we identified 76 regions of unusually high minor parent ancestry (minor parent ancestry "islands"), in the Santa Cruz *X. birchmanni* × *X. cortezi* population, where minor parent ancestry fell in the upper 2.5% tail of the genome wide ancestry

distribution. Approximately 25% of these minor parent islands were also identified in one or more *X. birchmanni* × *X. malinche* hybrid population.

We next explored whether the amount of sharing of both deserts and islands exceeded what would be expected by chance using two approaches. First, we randomly sampled 0.05 cM windows from each focal population and counted the number of windows overlapping with detected deserts and islands that had low or high minor parent ancestry. With this approach we found that sharing of both low and high minor parent ancestry regions in the real data exceeded sharing expected by chance (Fig. 3A; Methods; both deserts and islands $p<0.001$ by permutation). Because this simple approach could introduce artifacts by disrupting autocorrelation in local ancestry along the genome, we also evaluated expected ancestry sharing by chance using an approach where we preserved the structure of local ancestry by shuffling the genome in large chunks (Supporting Information 8, Fig. S12; deserts $p<0.008$, islands $p=0.2$, by permutation). While both approaches indicated that the number of shared minor parent deserts in the real data is unexpected by chance, support for shared islands is weaker.

We also evaluated shared ancestry deserts and islands in light of a number of technical factors that could bias ancestry calls and potentially drive artifacts in the data, such as variation in power to infer ancestry along the genome and coincidence of minor parent deserts and islands with repetitive regions (Supporting Information 8). We do not find evidence for systematic differences in power or expected error rates between shared and unshared regions (Fig. S13-14; Supporting Information 8). We also found that shared minor parent deserts and islands are not unusual in the number of conserved or coding basepairs they contain, suggesting that they are not driven by regions with an especially high or low density of conserved or coding basepairs (Supporting Information 8; Fig. S15). Moreover, minor parent deserts were still ancestry outliers when compared to regions of the genome predicted to have especially high constraint and minor parent islands remained ancestry outliers when compared to regions of the genome predicted to have especially low constraint (Fig. S16; Supporting Information 8). Finally, simulations indicated that the approach we use to identify shared minor parent deserts has good power even at moderate selection coefficients (e.g. s~0.05; Supporting Information 7). Overall, our analyses suggest that our approach to identifying regions of high and low minor parent ancestry between hybrid population types is robust to technical artifacts.

Features coinciding with shared minor parent ancestry deserts and islands

The unexpectedly high number of shared minor parent ancestry deserts and islands is consistent with a model where there are some shared loci under selection across species pairs. These shared ancestry patterns could be driven by a number of shared sources of selection on hybrids across the two species pairs, including selection against hybrid incompatibilities, shared signatures of adaptive introgression, among other possible mechanisms (Supporting Information 6). We evaluated the locations of minor parent deserts and islands across the two types of hybrid populations to begin to explore the mechanisms of selection driving them. Two of the 21 shared minor parent deserts overlap with known segregation distorters in *X. birchmanni* × *X. malinche* that are likely caused by hybrid incompatibilities (see next section), and three shared deserts correspond to structural differences between species on chromosome 7 and chromosome 17. Another shared desert was found in all replicate hybrid populations where *X. birchmanni* is the minor parent. This region spans the first two megabases of chromosome 21, the putative sex chromosome (Fig. S17), and is depleted in *X. birchmanni* ancestry. However, consistent with previous work in swordtails, we do not detect an enrichment of shared minor parent deserts (or

347 islands) on chromosome 21 (Supporting Information 8), possibly because sex chromosomes are
348 young and not strongly differentiated between males and females [52].
349 Of the 19 regions with especially high minor parent ancestry across hybrid population
350 types, two correspond to an inversion that occurs on chromosome 17 and one corresponds to an
351 inversion on chromosome 19 (Fig. S18). Intriguingly, inversions that differed between species
352 had unusually high minor parent ancestry in all hybrid populations on average (Fig. S19), in
353 contrast to what is most frequently observed in other systems ([53–55] but see [40,56,57]).
354 The remaining minor parent deserts and islands do not coincide with known structural
355 differences between species [35,46], mapped hybrid incompatibilities between *X. birchmanni* ×
356 *X. malinche* [20,23,50], or mapped QTL underlying trait differences between the species [23,58].
357 We investigate other features associated with these regions below.
358
359 *Ancestry at sites under selection in X. birchmanni × X. malinche hybrids*
360 We observe some overlap between known hybrid incompatibilities and minor parent
361 ancestry deserts, suggesting that this is one mechanism that can drive cross-population
362 repeatability in ancestry (among several plausible mechanisms; Supporting Information 6). To
363 test the role of hybrid incompatibilities in driving cross-repeatability more formally, we took
364 advantage of sites likely to be involved in hybrid incompatibilities in replicate *X. birchmanni* ×
365 *X. malinche* hybrid populations and asked if they had particularly low minor parent ancestry in
366 *X. birchmanni* × *X. cortezi* hybrid populations.
367 Nearly a decade of work has aimed to identify hybrid incompatibilities between *X.*
368 *birchmanni* × *X. malinche* using natural hybrid populations and artificial crosses [20,23,35,50].
369 We focused on two available datasets: putative segregating DMIs identified in natural hybrid
370 populations between *X. birchmanni* × *X. malinche* [20,50] as well as six segregation distorters
371 identified in F_2 hybrids of lab crosses between *X. birchmanni* × *X. malinche* that show ancestry
372 patterns consistent with the action of strongly selected DMIs (e.g. $s>0.25$; see Methods; Fig.
373 S20). For the segregation distorter dataset, we found that sites under selection in *X. birchmanni* ×
374 *X. malinche* F_2 hybrids had lower minor parent ancestry on average in the replicate *X.*
375 *birchmanni* × *X. cortezi* hybrid populations (Fig. 3B). Moreover, we show that this result is
376 robust to excluding a segregation distorter that is also a shared minor parent ancestry desert
377 across all populations (Fig. S20-21). We do not see an effect of DMIs mapped in *X. birchmanni*
378 × *X. malinche* natural hybrid populations that were inferred to be under weaker selection, which
379 could indicate that they are not enriched for hybrid incompatibilities between *X. birchmanni* × *X.*
380 *cortezi* or that we have low power to detect incompatibilities under weak selection [20].
381
382 *Mapping a newly identified shared hybrid incompatibility*
383 In the previous section we showed that hybrid incompatibilities identified in one species
384 pair are enriched for low minor parent ancestry in natural hybrid populations formed between
385 another species pair, indicating that some hybrid incompatibilities are repeatable across the two
386 types of hybrid populations. One striking signal we observe in our data is the coincidence of an
387 ancestry desert on chromosome 6 shared across all populations where *X. birchmanni* is the minor
388 parent and a strong segregation distorter identified in F_2 hybrids between *X. birchmanni* × *X.*
389 *malinche* (Fig. 3C-D; see Methods). Further examination of this region indicated that it is
390 depleted for minor parent ancestry in all replicate *X. birchmanni* × *X. malinche* populations (Fig.
391 3C-D, though it did not reach the shared desert threshold in all populations).

Since minor parent ancestry in this region appears to be deleterious across all hybrid populations in both species pairs, we predicted that it was likely to be a shared hybrid incompatibility ([35]; Supporting Information 6). By definition, hybrid incompatibilities involve two or more loci, so we set out to identify loci interacting with the chromosome 6 region in hybrids. Because this region appears to be under selection in *X. birchmanni* × *X. malinche* F_2 hybrids, we leveraged our data from *X. birchmanni* × *X. malinche* F_2s to scan for interacting loci. Using deviations from expected two locus genotypes (see Methods), we mapped a locus that interacts with the chromosome 6 region to chromosome 13 (Fig. 4A). Several genotype combinations are found in F_2 hybrids at lower than expected frequencies (Fig. 4).

Examining the chromosome 13 region in natural hybrid populations of both *X. birchmanni* × *X. malinche* and *X. birchmanni* × *X. cortezi*, we found that it also tended to have lower minor parent ancestry across populations than expected by chance, although it was not extreme enough to have been identified as a shared desert in our initial scan. At the center of the associated chromosome 13 region, minor parent ancestry in four out of five of the hybrid populations fell into the lower 10% quantile genome-wide (Fig. S22; empirical p-value based on ancestry in 10 kb non-overlapping windows – $p<0.005$).

The associated regions on chromosome 6 and chromosome 13 identified in F_2 hybrids between *X. birchmanni* × *X. malinche* contain 13 and 29 genes respectively. We used the STRING database [59] to evaluate whether any pairs of genes across the two regions are known to interact, restricting interaction criteria to those supported with experimental evidence, evidence of dimerization, or evidence of co-expression. Only one pair of genes was annotated as interacting based on these criteria across the two regions on chromosomes 6 and 13, the mitochondrial complex I genes *ndufs5* and *ndufa13*. Both these genes are strongly conserved in other *Xiphophorus* species [60], however several nonsynonymous substitutions distinguish these three species and appear to be derived in *X. birchmanni* (Table 1). Intriguingly, all of the hybrid populations analyzed have fixed the mitochondrial haplotype of their major parent species (Fig. S23; see also [60]), and our cross setup means that all F_2 hybrids had *X. malinche* mitochondria since the reverse cross is rarely successful (see Methods). We explore the possibility that these genes form a complex incompatibility with mitochondrial interactors in a companion study [60].

Similar analyses provided some evidence for the presence of additional shared hybrid incompatibilities involving other shared minor parent deserts (see Methods; Supporting Information 9). These included evidence for another pairwise hybrid incompatibility involving a minor parent ancestry desert on chromosome 9 and an interacting locus on chromosome 4 (Fig. 4 C-D) and evidence for a complex hybrid incompatibility involving a minor parent ancestry desert identified on chromosome 5 (Fig. S24; Supporting Information 9), both detected at FPR threshold of 10%.

Evidence from simulations indicated that in contrast to good power to detect minor parent deserts in the population data driven by moderate selection (s~0.05), we have poor power to detect interactions in the F_2 dataset, with the exception of cases in which selection is extremely strong (e.g. $s > 0.25$; Supporting Information 7, 9). Given that we likely lack power to detect many interactions between shared minor parent deserts and other regions of the genome, we also analyzed the data using a broader approach. We asked whether there was evidence that genes in minor parent deserts (or islands) were enriched for known protein-protein interactions or high expression network centrality compared to matched null datasets. We did not find clear evidence that the interaction networks of genes in minor parent deserts (or islands) deviated from the genome-wide background (Supporting Information 10).

438
439 **Table 1.** Nucleotide (Nt) and amino acid (AA) changes between species pairs in interacting
440 genes between the chromosome 6 and chromosome 13 regions. Ratios of rates of
441 nonsynonymous to synonymous substitutions (dN/dS) inferred by codeml [61] are also shown.
442 Comparisons where dN/dS is marked as > 99 are due to all nucleotide changes being
443 nonsynonymous.
444

	ndufs5			***ndufa13***		
Species Comparison	dN/dS	Nt changes	AA changes	dN/dS	Nt changes	AA changes
X. birchmanni - X. cortezi	> 99	5	5	> 99	3	3
X. birchmanni - X. malinche	> 99	5	5	1.2	4	3
X. cortezi – X. malinche	0	0	0	0.001	1	0

445
446
447

Discussion

We now know that species across the tree of life have hybridized with their close relatives and that many species derive substantial proportions of their genomes from past hybridization events. Based on these observations, it is likely that hybridization plays an important role in the evolution of genomes and species, but we still understand relatively little about how this process unfolds. There are major open questions about which mechanisms shape retention and loss of minor parent ancestry after hybridization and how repeatable this process is across species.

While the predictability of evolution has been a classic question in evolutionary biology for decades [62], such questions have been less central to the hybridization literature [6,10,63]. With the expanding number of known hybridization events between closely related species, we are now poised to tackle these questions. One general principle that has emerged is that ancestry from the minor parent species, or the species that contributed less to the genome of hybrids, is more frequently purged by selection after hybridization [6]. This purging can be explained by several mechanisms of selection on hybrids, including selection against hybrid incompatibilities [19,25,28,29,35], and tends to be stronger in regions of the genome with many linked coding or conserved basepairs [8,12,35]. We also uncover these patterns in the *X. birchmanni* × *X. cortezi* hybrid populations studied here, showing a substantial effect of local recombination rate and linked coding and conserved basepairs on minor parent ancestry in hybrids (Fig. 2). This represents at least the sixth system in which these effects have been detected [8,12,35,36,38,40], suggesting that these patterns of selection on minor parent ancestry are predictable outcomes of the process of genome stabilization after hybridization.

Here, we dramatically expand our understanding of how repeatable the outcomes of hybridization are by comparing patterns of local ancestry in hybridization events between *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* species pairs. This allows us to evaluate, for the first time, some of the factors driving repeatability in the outcomes of hybridization across related species pairs with similar histories of admixture. Both hybridization events began recently (in the last ~80-150 generations [20,35]) but occur in different geographic regions between distinct species pairs (Fig. 1). Because we find that broad scale factors like genome structure are important in generating local ancestry variation in both species pairs, this is expected to drive some concordance in local ancestry across the two species pairs. Indeed, we detect repeatability in where in the genome minor parent ancestry is retained or purged across pairs of *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations (Fig. 2; Table S10).

Beyond these broad patterns generated by shared genetic architecture, little is known about whether the same genes or regions of the genome are intolerant of hybridization in related species pairs. Intriguingly, we still observe correlations in local ancestry across the two types of hybrid populations after accounting for shared genomic architecture (Table S12). This indicates that although shared genome structure plays a role in some of the repeatability in local ancestry we observe, other factors are likely to contribute, such as shared sites under selection across the two types of hybridization events. Indeed, when we searched for local signals of shared ancestry across the two hybrid population types, we found evidence that some of the same regions of the genome are under selection in hybrid populations of both species pairs. Compared to expected overlap by chance, *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations are >10X enriched in both shared minor parent deserts and islands.

What factors contribute to this high repeatability of sites under selection across the two types of hybrid populations? From first principles, possible mechanisms driving repeatable local ancestry include shared hybrid incompatibilities or shared regions of adaptive introgression. Given that the genetic architecture of hybrid incompatibilities has been well-studied in *X. birchmanni* × *X. malinche* hybrids, we can directly ask about the role of such sites in contributing to repeatable ancestry across the two types of hybrid populations. In regions identified as likely hybrid incompatibilities in F_2 hybrids between *X. birchmanni* × *X. malinche*, minor parent ancestry was on average lower near these sites in *X. birchmanni* × *X. cortezi* hybrid populations. Only two of these regions overlap with the minor parent deserts discussed above, suggesting that they represent largely complementary signals of shared sites under selection in replicate *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations.

Further comparisons across hybrid population types reveal at least one and possibly multiple shared hybrid incompatibilities between *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche*. We identified a single region on chromosome 6 that was a segregation distorter in early generation hybrids and depleted in minor parent ancestry in all five replicate *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations. We found that this region on chromosome 6 interacts with a locus on chromosome 13 (Fig. 4). Based on gene network analysis, we identified two mitonuclear genes in the chromosome 6 and chromosome 13 regions that physically interact to form part of mitochondrial complex I. Intriguingly, all hybrid populations studied here of both *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* are fixed for mitochondrial haplotypes from the major parent species, and crosses only included individuals with *X. malinche* mitochondria. This raises the possibility that the incompatibility may be driven by complex interactions between the mitochondrial genome and its interacting nuclear counterparts. Indeed, we find evidence that this is the case in *X. birchmanni* × *X. malinche* hybrids in a companion study [60]. Similar analyses of other regions suggest that there are multiple shared hybrid incompatibilities between *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* associated with shared minor parent ancestry deserts (Fig. 4).

While it is clear that hybrid incompatibilities play a role in driving some of the shared minor parent deserts across the two types of hybrid populations, other mechanisms of selection are likely important in generating shared minor parent islands (although we note that minor parent islands are less strongly enriched compared to expectations by chance). In some cases, they may reflect cases where the minor parent haplotype is globally beneficial (i.e. adaptive introgression). For example, *X. birchmanni* is the minor parent in both the Santa Cruz and Tlatemaco populations, hinting that *X. birchmanni* haplotypes may be globally favored within the four shared minor parent islands identified in these two populations. In other pairs of populations, however, when the same regions are enriched in minor parent ancestry this means that different ancestries are favored in the two populations (i.e. in the seven islands shared between Santa Cruz and Aguazarca, *X. birchmanni* ancestry is overrepresented in the former and *X. malinche* in the latter). We speculate that some minor parent islands could correlate with regions under balancing selection such that minor parent ancestry is generally advantageous or other mechanisms that can favor minor parent ancestry such as masking of weakly deleterious alleles [64]. Notably, three shared minor parent islands occur in inversions that differentiate species, conflicting with the prevailing wisdom that inversions should resist introgression (Fig. S18; see [40,56,57] for similar findings).

The shared hybrid incompatibility we identify here in *X. birchmanni* × *X. malinche* and *X. birchmanni* × *X. cortezi* hybrids is one of the first described cases in any species group.

However, this finding should not be surprising since some hybrid incompatibilities are predicted to be shared between related species from first principles. Hybrid incompatibilities are typically described as interactions between derived mutations in each species but incompatibilities can also arise between the ancestral and derived genotypes when multiple derived mutations have accumulated along one lineage (Fig. S25; [44]). This type of ancestral-derived incompatibility may be more likely in cases where there are biased patterns of lineage specific substitutions in particular gene classes, as appears to be the case for the *X. birchmanni* mitochondrially interacting proteins discussed above (Table 1). Our findings add to a handful of known cases where the same or similar genetic interactions are under selection in hybridization events across multiple species pairs [21,23,65,66]. Whether this phenomenon is widespread awaits progress mapping additional hybrid incompatibilities in related species pairs [6]. The comparative approach used here is a powerful strategy for evaluating this question on a genome-wide scale.

Together, our results highlight predictable outcomes of selection after hybridization. We find that the outcomes of hybridization are repeatable across closely related species pairs, and are driven in part by broad scale factors such as shared genetic architecture. Beyond shared genomic architecture, we find that diverse mechanisms contribute to repeatability in local ancestry, including shared hybrid incompatibilities between species and shared sites where minor parent ancestry is favored. This parallelism at both the broad and local scale is an important first step towards untangling the drivers of local variation in ancestry in hybrid genomes.

Methods

Sample Collection

Wild fish were caught using baited minnow traps in the states of Hidalgo and San Luis Potosí, Mexico [46]. Individuals from populations that contain a mixture of pure *X. birchmanni* individuals and *X. birchmanni* × *X. cortezi* hybrids were sampled from two distinct collection sites occurring in separate tributaries of the Río Santa Cruz in northern Hidalgo: Huextetitla (21°9'43.82"N 98°33'27.19"W) and Santa Cruz (21°9'27.63"N 98°31'13.79"W). We excluded pure *X. birchmanni* individuals based on genome-wide ancestry, resulting in 254 hybrid individuals included in this study. Ninety-eight of these individuals were originally sequenced as part of a previous study documenting the hybridization event between *X. cortezi* and *X. birchmanni*, and the remaining individuals were collected from the Santa Cruz hybrid population in July of 2020. Collected fish were anesthetized in a 100 mg/mL buffered solution of MS-222 and water (Stanford APLAC protocol #33071). Fish were then photographed and a small fin clip was taken from each individual's caudal fin. Fin clips were preserved in 95% ethanol for later DNA extraction.

Tn5 libraries for low-coverage whole genome sequencing of X. birchmanni × X. cortezi hybrids

We prepared libraries for low-coverage whole genome sequencing from DNA extracted from fin clips collected from fish caught at the Huextetitla and Santa Cruz populations as described in Powell et al. [46]. Briefly, we extracted DNA from fin tissue using the Agencourt DNAdvance kit (Beckman Coulter, Brea, California) using half the recommended reaction volume but otherwise as specified by the manufacturer. We quantified isolated DNA with a TECAN Infinite M1000 microplate reader (Tecan Trading AG, Switzerland). We then prepared tagmentation-based whole genome libraries for low coverage sequencing by enzymatically shearing DNA diluted to approximately 2.5 ng/ul using the Illumina Tagment DNA TDE1 Enzyme and Buffer Kits (Illumina, San Diego, CA) at 55°C for 5 minutes. We amplified the sheared DNA in dually-indexed PCR reactions for 12 cycles. We purified pooled amplified PCR reactions with 18% SPRI magnetic beads. Library concentrations were measured with a Qubit fluorometer (Thermo Scientific, Wilmington, DE) and library size distributions were measured using Agilent 4200 Tapestation (Agilent, Santa Clara, CA). Libraries were sent for sequencing on an Illumina HiSeq 4000 at Admera Health Services, South Plainfield, NJ.

Local ancestry inference in X. birchmanni × X. cortezi hybrids

We previously developed and validated local ancestry inference methods for *X. birchmanni* × *X. cortezi* hybrids using a combination of demographic inference, simulations, analysis of parental individuals, and analysis of a subset of the hybrid individuals described here [46]. Briefly, we used panels of parental individuals (N=37 for *X. cortezi* and N=55 for *X. birchmanni*) to identify 1.1 million ancestry informative sites genome wide that were nearly fixed between the two parental species (90% frequency difference or greater), yielding approximately 3 ancestry informative sites per 2 kb. Performance was tested on pure parental individuals not included in the reference panel and in a series of simulations [46]. We found that across scenarios tested, accuracy in local ancestry inference was remarkably high (≤0.3% errors per ancestry informative site in all tested scenarios), presumably due to the high density of ancestry informative markers.

We applied a hidden Markov model (HMM) to infer local ancestry based on low-coverage whole genome sequencing data collected from Huextetitla and Santa Cruz hybrids, using the pipeline described above [46,67]. Average coverage across individuals was 1.1X and the minimum coverage cutoff used in this study was 0.15X. We set priors for the number of generations since initial admixture to 150 and the admixture proportion to 85% *X. cortezi* based on the results of previous work [46]. We set the error rate for the HMM to 0.02 based on empirically inferred error and cross-contamination rates for our library preparation protocol [67]. After running the HMM and excluding ancestry informative sites that were not covered in any individual, we recovered posterior probabilities for each ancestry state at 1,050,362 informative sites genome wide in 254 individuals.

For many of our analyses, it was convenient to convert these posterior probabilities for a given ancestry state to hard-calls. We used a posterior probability threshold of 0.9 or greater to assign a site to a given ancestry state (homozygous *X. birchmanni*, heterozygous, or homozygous *X. cortezi*). Sites with lower than 0.9 posterior probability were masked, as were sites that were present in fewer than 25% of individuals. Minor parent ancestry was summarized across individuals by averaging the ancestry hard-calls at each site across a range of non-overlapping window sizes (e.g. 10 kb – 1 Mb; 0.1 – 1 cM).

Historical recombination events in admixed populations are detectable as ancestry transitions in present day individuals [68]. We were also able to use this data to infer the approximate locations of crossover events in a hybrid ancestor that resulted in a transition between ancestry states. We identified the interval over which the posterior probability changes from 0.9 for one ancestry state to 0.9 in support of a different ancestry state. We excluded markers that were not in Hardy-Weinberg equilibrium at a Bonferroni corrected p-value of 0.1, since errors in genotyping and ancestry inference could generate false ancestry switches (excluded sites in this analysis: 84,366 – 8% of ancestry informative markers). This resulted in the identification of 168,858 ancestry transitions genome-wide across the two populations with a median resolution of 22 kb (~29 per individual per chromosome).

The density of ancestry informative sites in swordtails varies with features of interest such as the local recombination rate and the number of linked functional basepairs [35], likely due to processes such as background selection within the parental lineages. As a result, we must be cautious of technical biases that could arise due to the distribution of ancestry informative sites that could mislead us into inferring biological differences when there are in fact none. To address this, we thinned the data with respect to the density of ancestry informative sites and re-inferred local ancestry as described above. On average, an ancestry informative site between *X. birchmanni* × *X. cortezi* occurs once every 395 basepairs. For the ~50% of the genome where ancestry informative sites occurred more frequently, we thinned informative sites to retain only one per 395 basepairs. We inferred ancestry using this thinned set of informative sites (N= 689,983) and performed all of the analyses described on both thinned and unthinned datasets to ask if any results qualitatively change as a result of the input dataset. Results for the thinned datasets are reported in Supporting Information 11.

Local ancestry inference in X. birchmanni × X. malinche hybrids

For the replicate *X. birchmanni* × *X. malinche* hybrid populations analyzed in our study, we took advantage of data collected from previous studies of these replicate populations [35,58]. As previously described, we inferred local ancestry with an HMM and used population specific priors for the admixture proportions and time since initial admixture [23,35]. We limited our

652 analysis to individuals collected over a single sampling year. Briefly, we included 95 individuals
653 collected from Tlatemaco in 2013 (resulting in 629,584 ancestry informative sites), 97
654 individuals collected from Acuapa in 2018 (613,171 sites), and 51 individuals collected in 2016
655 from Aguazarca (628,811 sites).
656 Previous work has explored the accuracy of these local ancestry inference approaches in
657 *X. birchmanni* × *X. malinche* hybrids in great detail [23,67]. Briefly, initial definition of ancestry
658 informative markers relied on 183 parental individuals [23], ancestry informative sites that were
659 not fixed or nearly fixed between the two parental species were excluded (<98% frequency
660 difference). Performance was evaluated in simulations, on pure parental individuals not used in
661 the definition of ancestry informative sites and on early generation F_1 and F_2 hybrids where
662 switches in ancestry over short genetic distances can confidently be inferred to be errors. Based
663 on these analyses we estimated error rates to be ≤0.2% errors per ancestry informative site in all
664 tested scenarios [23,67]. Thus, given high expected accuracy of ancestry inference in both *X.*
665 *birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrids, we do not expect that errors
666 will drive to the global or local patterns that we observe in the empirical data.
667
668 *Comparing minor parent ancestry as a function of local genetic architecture*
669 For the *X. birchmanni* × *X. cortezi* hybrid populations, we wanted to evaluate whether
670 minor parent ancestry was correlated with recombination rate, as had previously been described
671 in other species [35,36,40]. We summarized average ancestry in non-overlapping 50 kb, 100 kb,
672 and 250 kb windows for each population. We previously created a linkage disequilibrium (LD)
673 based recombination map for *X. birchmanni* [35] and calculated recombination rate based on this
674 map over these same spatial scales. Available data suggests that recombination maps are likely to
675 be conserved between *X. birchmanni* and *X. cortezi* (Supporting Information 4; [35]), and
676 between parental species and hybrids [35].
677 Because the local recombination rate covaries with other genomic features in swordtails
678 [35], we also calculated the number of coding (or conserved) basepairs in each window so that
679 we could account for this in our analysis. We performed these calculations using annotations
680 developed for the *X. birchmanni* genome [23]. We then used a Spearman's partial correlation in
681 R to determine if ancestry was correlated with recombination rate, after accounting for the
682 number of coding and conserved basepairs in each window. To aid in interpreting these
683 empirical results, we also analyzed simulations that were performed under the inferred
684 demographic history of *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid
685 populations, with and without selection (Supporting Information 3, 6).
686 After evaluating the relationship between recombination rate and ancestry in *X.*
687 *birchmanni* × *X. cortezi* populations, we were also interested in the relationship between minor
688 parent ancestry and the locations of coding and conserved basepairs along the genome. To
689 control for recombination rate, which is positively correlated with coding and conserved bases in
690 swordtails [35], we summarized average ancestry in 0.25 cM non-overlapping windows. As
691 noted above, we used the gene annotation file for *X. birchmanni* and a set of regions conserved
692 in percomorph fish that we had previously identified using phastCons. Comparisons of
693 alignments across species indicate that the locations of these elements are largely conserved
694 across *X. birchmanni*, *X. malinche*, and *X. cortezi* (Supporting Information 4). We counted the
695 number of coding or conserved basepairs in each window. Using this approach, we evaluated the
696 relationship between ancestry and the number of linked coding or conserved basepairs with a
697 Spearman's correlation (and partial correlation) in R. We chose to present results for 0.25 cM

698 windows in the main text because 50% of minor parent ancestry tracts in Santa Cruz are 0.25
699 cMs in length or longer (75% are 0.1 cMs in length or greater). However, we evaluated these
700 relationships across a range of non-overlapping window sizes (0.1-1 cM) and found the results
701 were qualitatively similar (Table S4).
702 Because power to accurately infer ancestry may be correlated with the number of nearby
703 conserved basepairs, we also repeated these analyses using ancestry probabilities generated from
704 a set of ancestry informative sites thinned to reduce power differences between different regions
705 of the genome (see above, *Local ancestry inference in X. birchmanni × X. cortezi hybrids*). We
706 found that our results were qualitatively unchanged (Table S8). We also repeated analyses of the
707 relationship between the number of conserved and coding basepairs and minor parent ancestry,
708 excluding windows that fell in the lower or upper 25% quantile of the number of ancestry
709 informative sites. We found that excluding these windows, where we expect to have particularly
710 low or high power to infer ancestry, did not impact our results (Table S9).
711 For analyses including recombination rate, we repeated tests using an LD-based map that
712 we inferred using the same methods after thinning SNPs in regions of the genome with high SNP
713 density in the *X. birchmanni* population sample it was originally generated with [35]. Because
714 our ability to reliably estimate recombination rate depends on the density of SNPs, we expect our
715 power to estimate recombination rate to vary along the genome. If power to estimate
716 recombination rate varies with other features of interest, such as density of coding or conserved
717 basepairs, this could mislead us into interpreting variation in power as variation in signal. To
718 evaluate this possibility, we calculated the median distance between SNPs in our dataset, which
719 was 184 basepairs. We thinned the SNPs input into LDhelmet so that when SNPs were denser
720 than one every 184 basepairs, a single SNP was retained per 184 basepairs. We then re-inferred
721 recombination rates using LDhelmet as described previously [35]. We repeated all analyses
722 described above that incorporate recombination rate with this thinned LD map and found the
723 results qualitatively unchanged (Supporting Information 11).
724 As a further validation of the relationship between minor parent ancestry and
725 recombination rate, we repeated analyses with a lower-resolution crossover-based map generated
726 from the 943 F_2 *X. birchmanni* x *X. malinche* hybrids. We observed a total of 34,939 ancestry
727 transitions genome-wide in our dataset, or an average of 1.5 per individual per chromosome. We
728 summarized the number of observed ancestry transitions, reflecting crossovers in the F_1 parents,
729 in windows. On average we observed only ~5 crossovers per 100 kb, so we chose to summarize
730 recombination rate in large windows of 1 Mb and 5 Mb. We repeated analyses of the relationship
731 between recombination rate and minor parent ancestry using this lower resolution crossover map
732 and found that our results were consistent with those recovered using the LD-based maps (Table
733 S3).
734 Finally, we tested the robustness of our results in several additional ways to control for a
735 number of potential technical artifacts (see full description in Supporting Information 11 and
736 Table S2). Since unexpectedly short ancestry tracts may represent errors, we repeated our
737 analyses masking minor parent ancestry tracts that were shorter than 0.004 cM and major parent
738 ancestry tracts that were shorter than 0.035 cM (tract length expectations given population
739 history were derived from [69]). To control for potential autocorrelation in nearby windows we
740 calculated correlations between minor parent ancestry and features of interest when we retained
741 one window per 500 kb (approximately the scale of admixture LD in these hybrid populations;
742 [35,46]). To control for the fact that recombination rates in hybrids are expected to differ within

743 species-specific structural rearrangements, we identified and removed windows spanning known
744 inversions and repeated our analyses.
745
746 *Repeatability of local ancestry between cross types*
747 To evaluate correlations in minor parent ancestry across *X. birchmanni* × *X. cortezi* and
748 *X. birchmanni* × *X. malinche* populations we used summaries of observed ancestry in non-
749 overlapping physical (50-250 kb) and genetic windows (0.1-1 cM), as described above. We first
750 calculated pairwise correlations between populations in minor parent ancestry without
751 accounting for other features (Fig. 2, Table S10). We then repeated these calculations using
752 partial correlation tests in physical windows (50-250 kb) that included the recombination rate,
753 the number of linked coding, and conserved basepairs as covariates (Table S12) and partial
754 correlation tests in genetic windows (0.1-1 cM) that included the number of linked coding and
755 conserved basepairs as covariates (Table S13). This allowed us to infer both whether broad-scale
756 correlations in local ancestry could be detected across populations and whether these correlations
757 could be largely explained by shared genetic architecture (i.e. locations of coding and conserved
758 basepairs and local recombination rate). We also explored this question using simulations. We
759 performed simulations under the inferred demographic histories for *X. birchmanni* × *X. cortezi*
760 and *X. birchmanni* × *X. malinche* hybrid populations with and without selection (Supporting
761 Information 3, 6) and evaluated expected cross-population correlations in ancestry under
762 different scenarios.
763
764 *Identification of minor parent ancestry "deserts" and "islands"*
765 Beyond shared correlations as a result of shared broad-scale architecture of the genome,
766 we were interested if we could identify shared minor parent ancestry "deserts," where minor
767 parent ancestry was unusually low, and shared minor parent ancestry "islands, where minor
768 parent ancestry was unusually high, that were shared across *X. birchmanni* × *X. cortezi* and *X.*
769 *birchmanni* × *X. malinche* hybrid populations.
770 To do so we identified regions of especially high and low minor parent ancestry in Santa
771 Cruz, the *X. birchmanni* × *X. cortezi* hybrid population for which we have the largest dataset (for
772 results of the same analysis using the Huextetitla populations see Supporting Information 12).
773 We first identified ancestry informative sites where the minor or major parent ancestry fell in the
774 lower 2.5% tail of genome-wide ancestry. We expanded out in the 5' and 3' direction from the
775 focal ancestry informative site until we reached a site on each edge that exceeded the 5%
776 ancestry quantile, and treated this interval as the boundary of the minor parent ancestry desert or
777 island. To control for possible technical artifacts, we then evaluated ancestry in these regions in
778 0.05 cM windows. We identified the window that contained the midpoint of the islands or desert
779 and checked that this entire window fell in the 10% minor or major ancestry tail, contained at
780 least 10 ancestry informative sites, and was more than 10 kb in length. We also merged any
781 detected regions that were closer together than 50 kb (deserts N=54, islands N=14) and removed
782 any regions shorter than 10 kb (deserts N=1). We also filtered regions where inferred ancestry
783 varied substantially between inference with unthinned and thinned ancestry informative markers
784 (Fig. S13). Specifically, we filtered islands where inferred ancestry differed by more than 10%
785 between thinned and unthinned data (N=58, based on 0.05 cM windows). This resulted in the
786 identification of 81 minor parent ancestry deserts with an average length of 317 kb, and 76
787 islands with an average length of 242 kb in the Santa Cruz population.

We next wanted to compare these regions across populations. Because each population has a distinct recombination history and distribution of ancestry tract sizes, we took the midpoint of the deserts and islands identified in the Santa Cruz hybrid population and overlapped these regions with average ancestry in other populations, calculated at a fine spatial scale (0.05 cM, or on average 11 kb). We then determined if these regions had unusually high or low minor parent ancestry in other populations (defined as falling in the 10% minor or major ancestry tail in another population). We identified several shared islands and deserts between Santa Cruz and *X. birchmanni* × *X. malinche* populations (Fig. 2, 3). Simulations suggest that this two-step approach has excellent power to detect shared sites under selection, even when selection coefficients are relatively weak (Supporting Information 7).

Next, we asked whether the number of shared minor parent islands and deserts across *X. birchmanni* × *X. malinche* and *X. birchmanni* × *X. cortezi* hybrid populations exceeded what would be expected by chance. To investigate this, we used the same 0.05 cM windows we used above to determine whether minor parent ancestry deserts and islands were shared. We permuted the average ancestry values in *X. birchmanni* × *X. malinche* populations across windows and asked whether any windows that were major or minor parent ancestry outliers in the permuted data overlapped with the ancestry deserts identified in Santa Cruz (defined as falling in the 10% tail of major or minor parent ancestry, as in the real data). We repeated this procedure 1000 times. Based on these permutations, we found that few minor (or major) parent ancestry outliers in *X. birchmanni* × *X. malinche* hybrid populations overlap minor parent deserts (or islands) identified in the Santa Cruz hybrid population by chance (Fig. 3A). For pairs of populations, the expected false positive rate across all simulations was less than 10% (Fig. 3A). When we instead asked about the probability that any minor parent desert or island identified in Santa Cruz overlapped an ancestry outlier in simulations, the false positive rate was higher (~20%).

Our interest in performing permutations was to determine how much overlap in minor parent ancestry deserts and islands is expected by chance, and how much we can attribute to shared sources of selection on hybrids. While our permutations above suggest that little overlap is expected by chance, one challenge of this type of analysis is that permuting the data disrupts all correlations in ancestry across hybrid populations. This means that it will also disrupt correlations driven by interactions between selection and features of the genome such as the local recombination rate and coding/conserved basepair density. Since we are primarily interested in minor parent deserts and islands driven by sites under shared selection rather than windows that happen to be in low recombination rate or high functional density regions of the genome, we attempted to control for this. First, our permutations use genetic windows, which account in part for the relationship between minor parent ancestry and recombination rate variation. To evaluate whether minor parent ancestry deserts and islands were likely driven by functional basepair density, we compared the windows identified in our analysis to the genomic background. We describe these methods in detail in Supporting Information 8 and the results of these analyses are presented in Fig. S14-S16.

Given the importance of this analysis for the conclusions of our manuscript, we also took two additional approaches to determine the expected false positive rates for desert and island regions. In the first, we shuffled ancestry in blocks along the genome to preserve local ancestry structure during permutations but otherwise evaluated false positive rates as described above. We found the results of these permutations to be similar, but with a somewhat higher false positive rate for minor parent deserts and a substantially higher false positive rate for minor parent islands (Supporting Information 8, Fig. S12). In the second, we analyzed expected false positive rates for

each identified desert and island individually. Briefly, for each minor parent desert or island we generated windows of matched cM length, and selected all windows in the genome that had a similar content of coding basepairs. We next used permutations to determine the expected overlap by chance of regions of extreme ancestry within these matched sets (Supporting Information 8).

Identifying loci under selection using controlled crosses and natural populations

Hybridization between *X. birchmanni* × *X. malinche* has been studied for more than a decade. As such, we have generated controlled crosses between *X. birchmanni* × *X. malinche* that are not yet feasible between *X. birchmanni* × *X. cortezi*. We have also mapped a number of hybrid incompatibilities that occur between the two species, including a gene interaction that causes melanoma in hybrids [23] and dozens of pairs of loci that show signatures consistent with selection against DMIs in natural hybrid populations [20,50].

We also recently generated a large mapping panel of F_2 hybrids between *X. birchmanni* × *X. malinche*, which we used for QTL mapping of male sexually selected traits [70]. Here, we reanalyzed this mapping panel, doubling our sample size by including females, to identify strong segregation distorters. Segregation distortion can be generated by selfish genetic elements but in hybrids segregation distortion is often a hallmark of regions of the genome that impact hybrid viability or fertility [31]. Since segregation distorters that we have power to detect in our mapping panel are likely to have a strong impact on viability or fertility of *X. birchmanni* × *X. malinche*, they also represent loci that we were likely unable to map previously since our initial work on hybrid incompatibilities relied on segregating variation in *X. birchmanni* × *X. malinche* hybrid populations [20,23]. Briefly, we generated F_1 hybrids between *X. malinche* females and *X. birchmanni* males, and produced F_2s by crossing this F_1 generation [70]. The reverse cross direction is rarely successful [60]. We inferred local ancestry as described previously for this cross [67,70]. As expected from the cross design, average ancestry across ancestry informative sites in our dataset of 943 males and females indicated that on average 50% of the genome was derived from each parental species. Our previously published dataset for QTL mapping included 568 males [70].

In reanalyzing this data, we next turned to identifying regions of the genome with unusually high *X. birchmanni* or *X. malinche* ancestry, consistent with loci having a major impact on viability or fertility in hybrids (or generating segregation distortion for other reasons). We first thinned the F_2 data to retain one ancestry informative marker per 50 kb for computational efficiency. We then used a binomial test to identify loci that deviated from the expected segregation ratio of 50% *X. birchmanni* and 50% *X. malinche* ancestry. We used a p-value threshold of $p<5^{-4}$, approximately corresponding to the 5% tail of ancestry outliers in our dataset. We note that several of these regions were robust to Bonferroni correction for the number of markers tested (N= 20,011; $p<5^{-6}$ corresponding to FDR of 10%). We consider this analysis conservative due to the extent of ancestry linkage disequilibrium in F_2 hybrids. After identifying regions of the genome with evidence for segregation distortion, we visualized ancestry in each region to ensure that the pattern was consistent with an extended region of segregation distortion rather than a short interval as might be expected from e.g. genotyping error (see Fig. S20).

To explore the strength of selection that would be required to generate segregation distortion at the threshold we impose on the real data, we performed simple simulations. We simulated genotypes formed from 50-50 mixtures of the two parental gametes. Next, we

880 randomly drew a selection coefficient from 0-1 against one of the homozygous genotypes and a
881 dominance coefficient (0-1). We implemented selection on this pool of genotypes, treating 1-*s*
882 (or 1-*hs*) as the probability of survival. We drew 943 individuals from the surviving genotypes
883 and calculated average ancestry at the simulated sites. Simulations where ancestry fell in the 5%
884 tail of the real data were treated as accepted simulations; we repeated this procedure 10,000
885 times. Based on the distribution of accepted simulations (Fig. S26), we predict that very strong
886 selection ($s > 0.25$) is required to generate ancestry patterns as extreme as the loci we focus on in
887 our F_2 dataset. Distributions of the dominance coefficient from accepted simulations mirrored the
888 prior distribution.
889
890 *Role of loci under selection in ancestry variation in natural populations*
891 One of our major questions is whether loci under selection are shared not just across
892 hybrid populations of the same cross type but between hybrid populations of different cross
893 types. To tackle this question we took advantage of what is known about selection in the *X.*
894 *birchmanni* × *X. malinche* system, using the segregation distortion regions identified above, and
895 more weakly selected putative DMIs that have been previously identified in natural hybrid
896 populations between *X. birchmanni* × *X. malinche* [20,50]. Both sets of loci are associated with
897 lower minor parent ancestry in *X. birchmanni* × *X. malinche* hybrid populations [35], as expected
898 by theory.
899 We asked if these loci were also associated with lower minor parent ancestry in *X.*
900 *birchmanni* × *X. cortezi* hybrid populations, and the degree to which they coincided with
901 ancestry deserts that are shared across *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X.*
902 *malinche* hybrid populations. We summarized ancestry in 10 kb non-overlapping windows and
903 used bedtools [71] to determine the distance of each window to the closest selected site in each
904 of the two datasets and plotted average major parent ancestry as a function of that distance (Fig.
905 S27).
906
907 *Identifying potential partners of shared ancestry desert and segregation distortion region*
908 One of the regions that we identified as a significant segregation distorter on
909 chromosome 6 in F_2 hybrids between *X. birchmanni* × *X. malinche* was also a shared minor
910 ancestry desert, and had unusually low minor parent ancestry across all surveyed *X. birchmanni*
911 × *X. malinche* and *X. birchmanni* × *X. cortezi* hybrid populations. The observation that minor
912 parent ancestry is disfavored across populations of diverse ancestry (majority *malinche*, majority
913 *birchmanni*, and majority *cortezi*) indicates that this locus is likely to be involved in a hybrid
914 incompatibility [35], as opposed to being driven by other possible mechanisms of selection on
915 hybrids [6].
916 Based on these observations, we wanted to identify the interacting partner of the
917 chromosome 6 locus. Given the clear expectations for genotype frequencies in F_2 hybrids and
918 apparent shared selection on this locus across species pairs, we used the F_2 data generated
919 between *X. birchmanni* × *X. malinche* to perform this scan. We selected an ancestry informative
920 marker from the center of the shared ancestry desert on chromosome 6, which falls within the
921 peak segregation distortion region on chromosome 6 identified in F_2 hybrids (Fig. 3). Next, we
922 performed a scan against all other ancestry informative markers across the genome and
923 calculated χ^2 statistics for deviations from expected two-locus genotype frequencies (with four
924 degrees of freedom). Importantly, we used the observed genotype frequencies across individuals
925 at each of the two loci to generate this expectation, such that we identified loci with deviations

926 from expected two-locus genotype combinations rather than deviations in ancestry at one locus
927 or the other.
928 To determine the appropriate significance threshold for this scan, we performed
929 simulations. Using the F_2 ancestry data described in the previous paragraph, we simulated
930 genotypes at the focal locus based on each individual's genome-wide ancestry and performed the
931 χ^2 test as described above. We repeated this for every locus in the dataset (i.e. over 20,011
932 markers) and recorded the minimum p-value. We performed this procedure 500 times and took
933 the lower 5% quantile of minimum p-values across simulations. This value served as our false
934 positive rate (FPR) cutoff for analysis of the real data. We describe this approach in more detail
935 in Supporting Information 9. We initially explored an approach using permutations at the focal
936 locus to set the FPR but found that this approach was less conservative; these results are also
937 described in Supporting Information 9. Only one region in the genome (found on chromosome
938 13) passed the FPR threshold determined by permutation.
939 We also repeated this scan for interacting loci for all ancestry deserts shared between *X.*
940 *birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* populations. While we did not identify
941 any interacting loci that passed the 5% genome-wide FPR correction threshold for other regions,
942 we highlight two other cases where interactions are detected with minor parent deserts at a FPR
943 of 10% (Supporting Information 9). Simulations suggest that our power to detect interactions in
944 is relatively low (Supporting Information 9), leaving open the possibility that other minor parent
945 deserts interact with partner loci that we are unable to detect in the current data.
946
947 *Pathway analysis of genes falling in a shared hybrid incompatibility*
948 We wanted to ask whether genes in the identified chromosome 6 and chromosome 13
949 regions might be known to interact from past experimental, fusion, or co-expression data. To
950 evaluate this, we took advantage of the STRING database [59]. We used haltools [72] to liftover
951 coordinates from the *X. birchmanni* reference genome to the *X. maculatus* reference genome for
952 the shared ancestry desert on chromosome 6 and the interacting region on chromosome 13. This
953 allowed us to extract gene IDs for *X. maculatus* in this region, which is included in the STRING
954 database.
955 We input gene names from both chromosomes (chromosome 6 – 13 genes; chromosome
956 13 – 29 genes) into STRING using the protein batch search function. Two pairs of interacting
957 genes with supporting experimental, fusion, or co-expression data were identified. For one of
958 these pairs, both genes fell within the chromosome 6 region but for the other pair, one gene fell
959 within the chromosome 6 region and the other within the chromosome 13 region.
960 Given that our power to map interacting regions of shared minor parent deserts is
961 relatively low (Supporting Information 9), we wanted to perform a broad scan to ask whether
962 there was evidence for enrichment of genes that have many protein-protein interactions or dense
963 regulatory networks within shared minor parent ancestry deserts and islands. We thus performed
964 similar analyses for genes found in other shared minor parent ancestry deserts and islands. These
965 analyses included both STRING database searches similar to those described above, and
966 evaluation of enrichment of hub genes from co-expression analyses in minor parent deserts and
967 islands compared to matched null sets (Supporting Information 10).
968
969 *Investigating other features associated with ancestry variation*
970 Because theory predicts that many of the changes in genome-wide ancestry that occur in
971 hybrid populations happen rapidly after admixture [29,39], we were interested in investigating

correlations between minor parent ancestry and chromosome-level features. We evaluated whether there were correlations between chromosome length, number of genes per chromosome, and minor parent ancestry using a Spearman's rank correlation test.

We also evaluated minor parent ancestry as a function of structural differences detected between the *X. birchmanni*, *X. malinche*, and *X. cortezi* assemblies. To identify chromosomal inversions between species, we compared de-novo assemblies using MUMmer (mummer-4.0.0beta2) [73]. After aligning the 24 largest linkage groups, which correspond to the 24 *Xiphophorus* chromosomes, we excluded alignments under 5 kb before proceeding with the analysis. We focused our analysis on scanning for large inversions (> 100 kb). To identify the putative inversion breakpoints, we used mummer's show-coords command. By aligning the *X. birchmanni* [23] assembly to *X. cortezi* [46] and to *X. malinche* [23], and comparing to an outgroup sequence (*X. maculatus;* [52]), we determined which inversions were derived in each species. For simplicity, we assumed the inversion configuration detected in the reference genomes was fixed at the species level (π in *X. birchmanni*, *X. malinche*, and *X. cortezi* is <0.12%) but we note that some segregating inversions have been detected in *Xiphophorus* [23].

Animal Ethics

All animal use followed protocols that were approved by Stanford IACUC (Stanford APLAC protocol number 33071). We are grateful to the Mexican federal government for permission to collect samples (permit number: PFF/DGOPA-064/20).

Acknowledgements

We thank Stepfanie Aguillon, Yaniv Brandvain, Erin Calfee, Yuki Haba, Priya Moorjani, Rachel Moran, Emilie Richards, Ken Thompson, Leslie Turner, and members of the Brandvain, Coop, Martin, Moorjani, Goldberg and Schumer labs for helpful discussion and/or feedback on earlier versions of this work. We also thank Baruc Zago-Mazzocco for field work support. We are grateful to the Mexican federal government for permission to collect samples. We thank Stanford University and the Stanford Research Computing Center for providing computational support for this project. This work was supported by a Knight-Hennessy Scholars fellowship and NSF GRFP 2019273798 to B. Moran, NRSA F32 GM135998 to B. Kim, a CEHG fellowship and NSF PRFB (2010950) to Q. Langdon, and a Hanna H. Gray fellowship and NIH 1R35GM133774 grant to M. Schumer

Figures

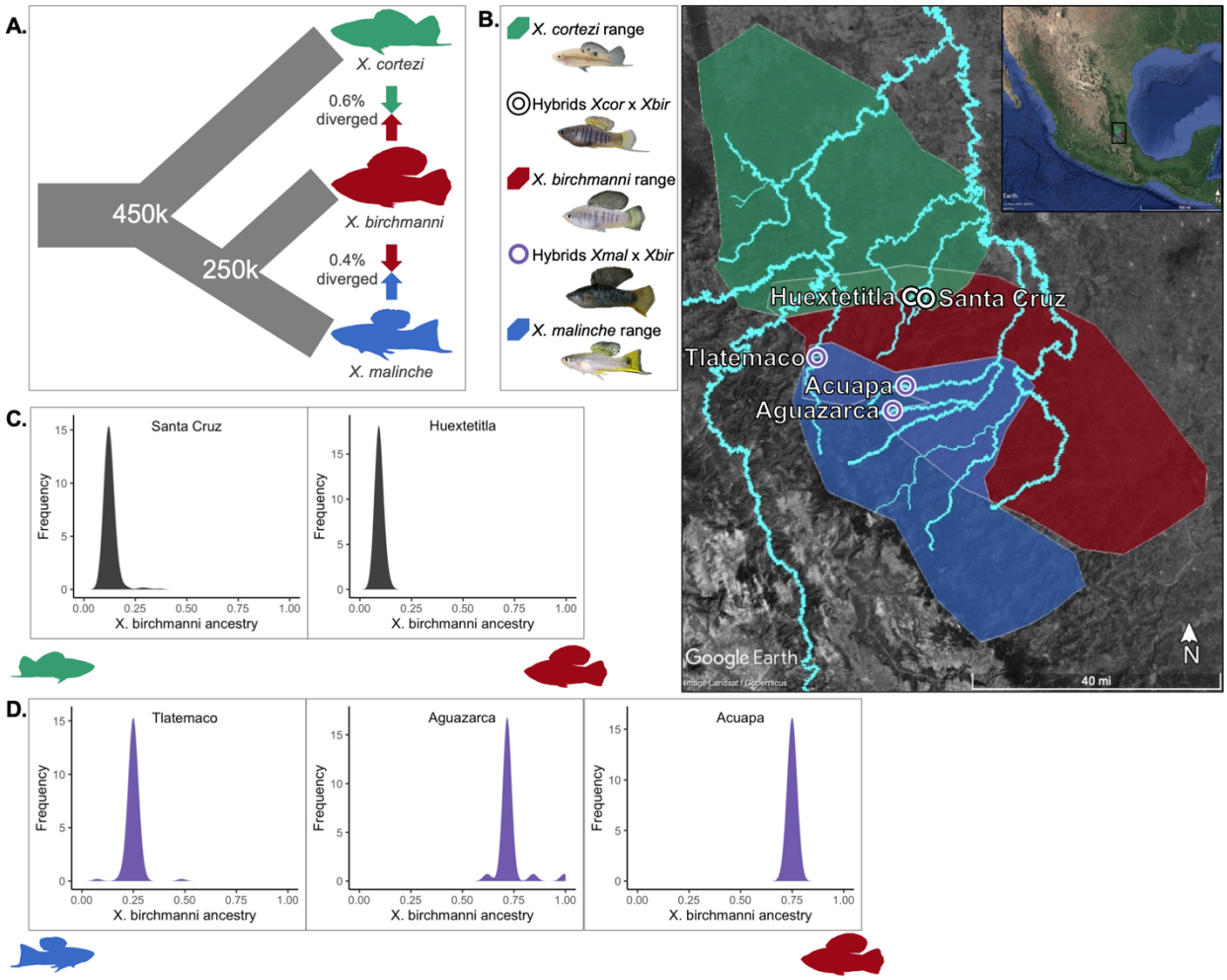

Figure 1. Description of focal species, study sites, and hybrid population ancestry structure. **A.** Phylogenetic relationships and contemporary hybridization events between focal species *X. birchmanni*, *X. malinche*, and *X. cortezi*. **B.** Distributions of *X. birchmanni* (*Xbir*), *X. malinche* (*Xmal*), and *X. cortezi* (*Xcor*), and their hybrids in Hidalgo and San Luis Potosí, Mexico. Blue shading shows the known range of *X. malinche*, red shading shows the known range of *X. birchmanni*, and green shading shows the known range of *X. cortezi*. **C**. Genome-wide admixture proportions of *X. birchmanni* × *X. cortezi* hybrids found in the Santa Cruz and Huextetitla replicate hybrid populations. **D**. Genome-wide admixture proportions of *X. birchmanni* × *X. malinche* hybrids found in the replicate populations at Tlatemaco (N=95), Aguazarca (N=51), and Acuapa (N=97).

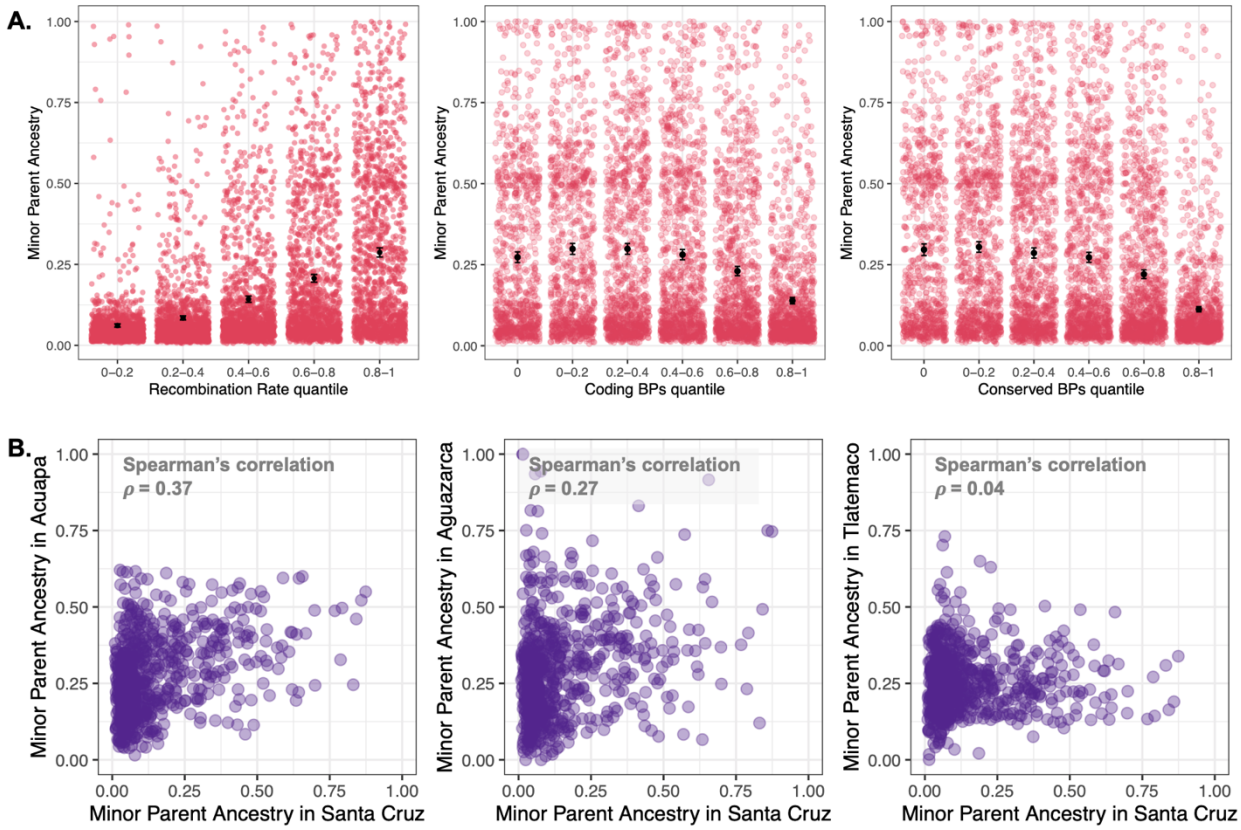

Figure 2. Relationship between minor parent ancestry, recombination rate and functional basepairs in *X. birchmanni* × *X. cortezi* hybrid populations, as well as cross-population type comparisons. **A.** Minor parent ancestry (*X. birchmanni* ancestry) in the Santa Cruz hybrid population is higher in regions of the genome with higher recombination rates. Semi-transparent points show results from individual 100 kb non-overlapping windows, black points and whiskers show the mean ± two standard errors of the mean. After accounting for recombination rate by averaging ancestry in 0.25 cM non-overlapping windows, we also observe a strong negative correlation between minor parent ancestry and the number of linked coding (middle) and conserved basepairs **(**right**)**. We find the same results when analyzing data from the Huextetitla hybrid population and across a range of window sizes (Table S1, S4). Note that results for **A** are shown in quintiles for visualization but statistical tests reported in the text were performed on unbinned data. **B**. Correlations in local ancestry, plotted here in 1 Mb non-overlapping windows. Cross-population ancestry correlations between Santa Cruz (or Huextetila – Table S10) and *X. birchmanni* × *X. malinche* hybrid populations are significant in all comparisons except for the majority *X. malinche* population Tlatemaco (far right; see main text & Supporting Information 6). Some of the ancestry covariance between *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations can be explained by shared features of genome architecture such as the locations of coding and conserved basepairs and the local recombination rate, but substantial covariance remains after accounting for these features (Table S12).

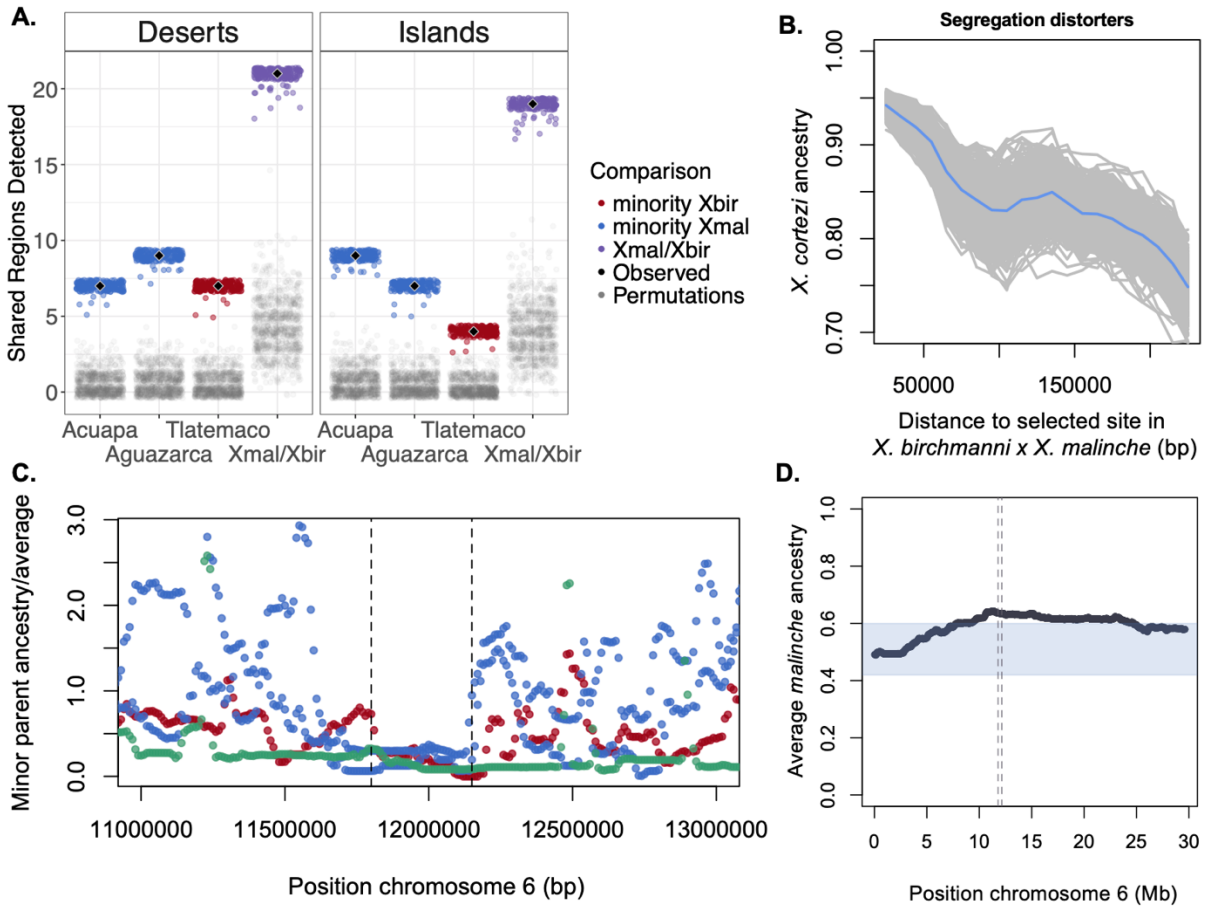

Figure 3. A. Shared minor parent deserts and islands are enriched compared to expectations by chance between *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations. Results shown here indicate the number of shared minor parent deserts (or islands) between the Santa Cruz *X. birchmanni* × *X. cortezi* hybrid population and each *X. birchmanni* × *X. malinche* hybrid population (Acuapa, Aguazarca, and Tlatemaco). Black diamonds show the observed number of shared minor parent deserts or islands and colored points show the results when the data is jack-knife bootstrapped. In contrast, few shared minor parent deserts or islands are expected by chance (results of permutations shown by gray points). The column labeled Xmal/Xbir shows the number of shared deserts or islands between the Santa Cruz *X. birchmanni* × *X. cortezi* hybrid population and any single *X. birchmanni* × *X. malinche* population. See Supporting Information 8 and Fig. S12 for the results of alternative analysis approaches. **B.** Minor parent ancestry in *X. birchmanni* × *X. cortezi* hybrid populations is depleted near 6 segregation distorters identified in artificial crosses between *X. birchmanni* × *X. malinche* (see Results). Results plotted here are for the Santa Cruz hybrid population; gray lines show 500 replicates bootstrap resampling windows in the data, blue line indicates the mean across simulations. **C.** One minor parent ancestry desert on chromosome 6 was detected across all *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations. Points represent minor parent ancestry calculated in 10 kb non-overlapping windows divided by the proportion of the genome derived from the minor parent species. Red – Tlatemaco (*malinche* × *birchmanni*), Blue – Acuapa and Aguazarca (*birchmanni* × *malinche*), Green –Santa Cruz (*birchmanni* ×

1063 *cortezi*). **D**. The shared minor parent ancestry desert on chromosome 6 coincides with a strong
1064 segregation distorter in artificial crosses between *X. birchmanni* × *X. malinche*. Plotted here is
1065 average ancestry at ancestry informative sites across 943 F_2 hybrids, where expected ancestry is
1066 50% *X. birchmanni* and 50% *X. malinche*. The blue envelop indicates the 99% quantile of
1067 variation in ancestry in this cross genome-wide. The dashed lines indicate the location of the
1068 shared ancestry desert in hybrid populations plotted in **C**.

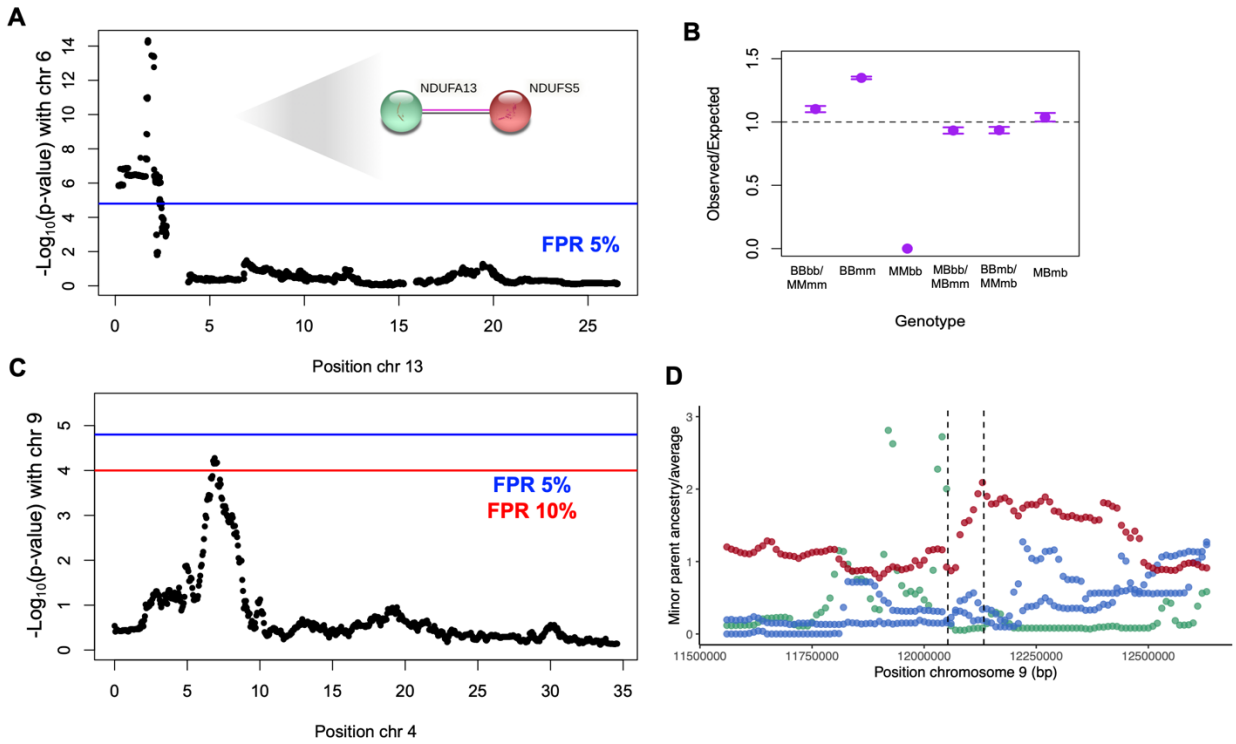

Figure 4. Evidence for repeated hybrid incompatibilities between *X. birchmanni* × *X. malinche* and *X. birchmanni* × *X. cortezi* hybrids. **A**. Interacting locus of the shared chromosome 6 ancestry desert identified in a genome-wide scan. This region was identified scanning for deviations from expected two locus genotypes using a χ^2 test in F_2 hybrids, based on observed genotypes at the shared chromosome 6 desert and genotypes at other loci throughout the genome. Genome-wide significance threshold (FPR 5%; blue line) was determined using permutations, see Methods for details. Inset shows STRING network with experimentally verified interactions between *ndufa13* (contained in the chromosome 6 region) and *ndufs5* (contained in the chromosome 13 region). **B**. Ratio of expected to observed two-locus genotype combinations in F_2 hybrids between *X. birchmanni* × *X. malinche* at the shared chromosome 6 desert and associated chromosome 13 region. Capital letters indicate genotype at chromosome 6 and lowercase letters indicate genotype at the chromosome 13 region. For visualization, certain genotype combinations are collapsed (i.e. homozygous parental – BBbb & MMmm; heterozygous locus 1, homozygous locus 2 – MBbb & MBmm). Note that several genotype combinations were expected to be found at low frequency due to low *X. birchmanni* ancestry on chromosome 13 (MMbb, MBbb, BBbb). **C.** A second interaction between the minor parent desert on chromosome 9 and a region on chromosome 4 is detected at a relaxed FPR of 10%. **D**. This minor parent desert is shared between *X. birchmanni* × *X. cortezi* hybrid populations and the Acuapa *X. birchmanni* × *X. malinche* hybrid population and occurs on chromosome 9 at 12.1 Mb. Points represent minor parent ancestry calculated in 10 kb non-overlapping windows divided by the proportion of the genome derived from the minor parent species. Red – Tlatemaco (*malinche* × *birchmanni*), Blue – Acuapa and Aguazarca (*birchmanni* × *malinche*), Green –Santa Cruz (*birchmanni* × *cortezi*).

1094 **References**
1095

1096 1. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete
1097 and error-free genome assemblies of all vertebrate species. Nature. 2021;592:737-746. doi:
1098 10.1038/s41586-021-03451-0

1099 2. Genereux DP, Serres A, Armstrong J, Johnson J, Marinescu VD, Murén E, et al. A
1100 comparative genomics multitool for scientific discovery and conservation. Nature.
1101 2020;587: 240–245. doi:10.1038/s41586-020-2876-6

1102 3. The Global Invertebrate Genomics Alliance (GIGA): Developing Community Resources to
1103 Study Diverse Invertebrate Genomes. J Hered. 2014;105: 1–18. doi:10.1093/jhered/est084

1104 4. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and
1105 speciation. Journal of Evolutionary Biology. 2013;26: 229–246. doi:10.1111/j.1420-
1106 9101.2012.02599.x

1107 5. Mallet J. Hybridization as an invasion of the genome. Trends Ecol Evol. 2005;20: 229–237.
1108 doi:10.1016/j.tree.2005.02.010

1109 6. Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. The genomic
1110 consequences of hybridization. Wittkopp PJ, editor. eLife. 2021;10: e69016.
1111 doi:10.7554/eLife.69016

1112 7. Turissini DA, Matute DR. Fine scale mapping of genomic introgressions within the
1113 Drosophila yakuba clade. PLOS Genetics. 2017;13: e1006971.
1114 doi:10.1371/journal.pgen.1006971

1115 8. Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. Speciation and Introgression
1116 between Mimulus nasutus and Mimulus guttatus. PLOS Genetics. 2014;10: e1004410.
1117 doi:10.1371/journal.pgen.1004410

1118 9. D'Angiolo M, De Chiara M, Yue J-X, Irizar A, Stenberg S, Persson K, et al. A yeast living
1119 ancestor reveals the origin of genomic introgressions. Nature. 2020;587: 420–425.
1120 doi:10.1038/s41586-020-2889-1

1121 10. Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, et al. Major
1122 ecological transitions in wild sunflowers facilitated by hybridization. Science. 2003;301:
1123 1211–1216. doi:10.1126/science.1086949

1124 11. Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal G. Phylogenomics reveals
1125 extensive reticulate evolution in Xiphophorus fishes. Evolution. 2013;67: 2166–2179.

1126 12. Sankararaman S, Mallick S, Patterson N, Reich D. The Combined Landscape of Denisovan
1127 and Neanderthal Ancestry in Present-Day Humans. Current Biology. 2016;26: 1241–1247.
1128 doi:10.1016/j.cub.2016.03.037

1129 13. Heliconius Genome C. Butterfly genome reveals promiscuous exchange of mimicry
1130 adaptations among species. Nature. 2012;487: 94–8.

1131 14. Trier CN, Hermansen JS, Sætre G-P, Bailey RI. Evidence for Mito-Nuclear and Sex-Linked
1132 Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. PLOS
1133 Genetics. 2014;10: e1004075. doi:10.1371/journal.pgen.1004075

1134 15. Orr HA. The population genetics of speciation: the evolution of hybrid incompatibilities.
1135 Genetics. 1995;139: 1805–1813.

1136 16. Arnegard ME, McGee MD, Matthews B, Marchinko KB, Conte GL, Kabir S, et al.
1137 Genetics of ecological divergence during speciation. Nature. 2014;511: 307–311.
1138 doi:10.1038/nature13301

1139 17. Svedin N, Wiley C, Veen T, Gustafsson L, Qvarnström A. Natural and sexual selection
1140 against hybrid flycatchers. Proceedings of the Royal Society B: Biological Sciences.
1141 2008;275: 735–744. doi:10.1098/rspb.2007.0967

1142 18. Christie K, Strauss SY. Along the speciation continuum: Quantifying intrinsic and extrinsic
1143 isolating barriers across five million years of evolutionary divergence in California
1144 jewelflowers. Evolution. 2018;72: 1063–1079. doi:10.1111/evo.13477

1145 19. Juric I, Aeschbacher S, Coop G. The Strength of Selection against Neanderthal
1146 Introgression. PLOS Genetics. 2016;12: e1006340. doi:10.1371/journal.pgen.1006340

1147 20. Schumer M, Cui R, Powell DL, Dresner R, Rosenthal GG, Andolfatto P. High-resolution
1148 mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. McVean
1149 G, editor. eLife. 2014;3: e02535. doi:10.7554/eLife.02535

1150 21. Chae E, Bomblies K, Kim S-T, Karelina D, Zaidem M, Ossowski S, et al. Species-wide
1151 genetic incompatibility analysis identifies immune genes as hot spots of deleterious
1152 epistasis. Cell. 2014;159: 1341–1351. doi:10.1016/j.cell.2014.10.049

1153 22. Larson EL, Vanderpool D, Sarver BAJ, Callahan C, Keeble S, Provencio LL, et al. The
1154 Evolution of Polymorphic Hybrid Incompatibilities in House Mice. Genetics. 2018;209:
1155 845–859. doi:10.1534/genetics.118.300840

1156 23. Powell DL, García-Olazábal M, Keegan M, Reilly P, Du K, Díaz-Loyo AP, et al. Natural
1157 hybridization reveals incompatible alleles that cause melanoma in swordtail fish. Science.
1158 2020;368: 731–736. doi:10.1126/science.aba5216

1159 24. Zuellig MP, Sweigart AL. A two-locus hybrid incompatibility is widespread, polymorphic,
1160 and active in natural populations of Mimulus*. Evolution. 2018;72: 2394–2405.
1161 doi:10.1111/evo.13596

1162 25. Barton NH. The role of hybridization in evolution. Molecular Ecology. 2001;10: 551–568.
1163 doi:https://doi.org/10.1046/j.1365-294x.2001.01216.x

26. Slatkin M, Lande R. Segregation variance after hybridization of isolated populations. Genet Res. 1994;64: 51–56. doi:10.1017/s0016672300032547

27. Thompson KA. Experimental Hybridization Studies Suggest That Pleiotropic Alleles Commonly Underlie Adaptive Divergence between Natural Populations. The American Naturalist. 2020;196: E16–E22. doi:10.1086/708722

28. Simon A, Bierne N, Welch JJ. Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. Evolution Letters. 2018;2: 472–498. doi:10.1002/evl3.66

29. Harris K, Nielsen R. The Genetic Cost of Neanderthal Introgression. Genetics. 2016;203: 881–891. doi:10.1534/genetics.116.186890

30. Presgraves DC. Sex chromosomes and speciation in Drosophila. Trends in Genetics. 2008;24: 336–343. doi:10.1016/j.tig.2008.04.007

31. Maheshwari S, Barbash DA. The Genetics of Hybrid Incompatibilities. Annual Review of Genetics. 2011;45: 331–355. doi:10.1146/annurev-genet-110410-132514

32. Qvarnström A, Bailey RI. Speciation through evolution of sex-linked genes. Heredity. 2009;102: 4–15. doi:10.1038/hdy.2008.93

33. Trier CN, Hermansen JS, Sætre G-P, Bailey RI. Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. PLoS Genet. 2014;10. doi:10.1371/journal.pgen.1004075

34. Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? Molecular Ecology. 2012;21: 4942–4957. doi:10.1111/mec.12006

35. Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 2018;360: 656. doi:10.1126/science.aar3684

36. Martin SH, Davey JW, Salazar C, Jiggins CD. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLOS Biology. 2019;17: e2006288. doi:10.1371/journal.pbio.2006288

37. Nachman MW, Payseur BA. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos Trans R Soc Lond B Biol Sci. 2012;367: 409–421. doi:10.1098/rstb.2011.0249

38. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507: 354–357. doi:10.1038/nature12961

1197 39. Veller C, Edelman NB, Muralidhar P, Nowak MA. Recombination, variance in genetic
1198 relatedness, and selection against introgressed DNA. bioRxiv. 2019; 846147.
1199 doi:10.1101/846147

1200 40. Calfee E, Gates D, Lorant A, Perkins MT, Coop G, Ross-Ibarra J. Selective sorting of
1201 ancestral introgression in maize and teosinte along an elevational cline. PLoS Genet.
1202 2021;17:e1009810. doi: 10.1371/journal.pgen.1009810

1203 41. Vilgalys TP, Fogel AS, Mututua RS, Warutere JK, Siodi L, Anderson JA, et al. Selection
1204 against admixture and gene regulatory divergence in a long-term primate field study. 2021
1205 Aug p. 2021.08.19.456711. doi:10.1101/2021.08.19.456711

1206 42. Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, et al. Repeated
1207 losses of PRDM9-directed recombination despite the conservation of PRDM9 across
1208 vertebrates. In: eLife [Internet]. 6 Jun 2017 [cited 23 Jul 2019]. doi:10.7554/eLife.24133

1209 43. Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, et al. Stable
1210 recombination hotspots in birds. Science. 2015;350: 928–932. doi:10.1126/science.aad0843

1211 44. Wang RJ, Ané C, Payseur BA. The Evolution of Hybrid Incompatibilities Along a
1212 Phylogeny. Evolution. 2013;67: 2905–2922. doi:https://doi.org/10.1111/evo.12173

1213 45. Yang C-H, Scarpino SV. Reproductive barriers as a byproduct of gene network evolution.
1214 bioRxiv. 2020; 2020.06.12.147322. doi:10.1101/2020.06.12.147322

1215 46. Powell DL, Moran BM, Kim BY, Banerjee SM, Aguillon SM, Fascinetto-Zago P, et al.
1216 Two new hybrid populations expand the swordtail hybridization model system. Evolution.
1217 2021;75: 2524–2539. doi:10.1111/evo.14337

1218 47. Fisher HS, Wong BBM, Rosenthal GG. Alteration of the chemical environment disrupts
1219 communication in a freshwater fish. Proc R Soc London Ser B. 2006;273: 1187–1193.
1220 doi:10.1098/rspb.2005.3406

1221 48. Schumer M, Powell DL, Delclós PJ, Squire M, Cui R, Andolfatto P, et al. Assortative
1222 mating and persistent reproductive isolation in hybrids. Proc Natl Acad Sci USA. 2017;114:
1223 10936. doi:10.1073/pnas.1711238114

1224 49. Culumber ZW, Fisher HS, Tobler M, Mateos M, Barber PH, Sorenson MD, et al.
1225 Replicated hybrid zones of Xiphophorus swordtails along an elevational gradient.
1226 Molecular Ecology. 2011;20: 342–356. doi:10.1111/j.1365-294X.2010.04949.x

1227 50. Schumer M, Brandvain Y. Determining epistatic selection in admixed populations.
1228 Molecular Ecology. 2016;25: 2577–2591. doi:10.1111/mec.13641

1229 51. Matute DR, Comeault AA, Earley E, Serrato-Capuchina A, Peede D, Monroy-Eklund A, et
1230 al. Rapid and Predictable Evolution of Admixed Populations Between Two Drosophila
1231 Species Pairs. Genetics. 2019 [cited 20 Apr 2020]. doi:10.1534/genetics.119.302685

1232 52. Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, et al. The genome of the
1233 platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and
1234 several complex traits. Nature Genetics. 2013;45: 567.

1235 53. Noor MAF, Grams KL, Bertucci LA, Reiland J. Chromosomal inversions and the
1236 reproductive isolation of species. PNAS. 2001;98: 12084–12088.
1237 doi:10.1073/pnas.221274498

1238 54. Kirkpatrick M, Barton N. Chromosome Inversions, Local Adaptation and Speciation.
1239 Genetics. 2006;173: 419–434. doi:10.1534/genetics.105.047985

1240 55. Wellenreuther M, Bernatchez L. Eco-Evolutionary Genomics of Chromosomal Inversions.
1241 Trends in Ecology & Evolution. 2018;33: 427–440. doi:10.1016/j.tree.2018.04.002

1242 56. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, et al. Genomic
1243 architecture and introgression shape a butterfly radiation. Science. 2019;366: 594–599.
1244 doi:10.1126/science.aaw2090

1245 57. Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, et al. Massive
1246 haplotypes underlie ecotypic differentiation in sunflowers. Nature. 2020;584: 602–607.
1247 doi:10.1038/s41586-020-2467-6

1248 58. Powell DL, Payne C, Banerjee SM, Keegan M, Bashkirova E, Cui R, et al. The Genetic
1249 Architecture of Variation in the Sexually Selected Sword Ornament and Its Evolution in
1250 Hybrid Populations. Current Biology. 2021 [cited 28 Jan 2021].
1251 doi:10.1016/j.cub.2020.12.049

1252 59. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING
1253 database in 2017: quality-controlled protein–protein association networks, made broadly
1254 accessible. Nucleic Acids Res. 2017;45: D362–D368. doi:10.1093/nar/gkw937

1255 60. Moran BM, Payne CY, Powell DL, Iverson ENK, Banerjee SM, Langdon QK, et al. A
1256 Lethal Genetic Incompatibility between Naturally Hybridizing Species in Mitochondrial
1257 Complex I. 2021 Jul p. 2021.07.13.452279. doi:10.1101/2021.07.13.452279

1258 61. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol.
1259 2007;24: 1586–1591. doi:10.1093/molbev/msm088

1260 62. Blount ZD, Lenski RE, Losos JB. Contingency and determinism in evolution: Replaying
1261 life's tape. Science. 2018;362. doi:10.1126/science.aam5979

1262 63. Chaturvedi S, Lucas LK, Buerkle CA, Fordyce JA, Forister ML, Nice CC, et al. Recent
1263 hybrids recapitulate ancient hybrid outcomes. Nat Commun. 2020;11: 2179.
1264 doi:10.1038/s41467-020-15641-x

1265 64. Kim BY, Huber CD, Lohmueller K. Deleterious variation shapes the genomic landscape of
1266 introgression. PLoS Genet. 2018;14:e1007741. doi: 10.1371/journal.pgen.1007741

1267 65. Lu Y, Sandoval A, Voss S, Lai Z, Kneitz S, Boswell W, et al. Oncogenic allelic interaction
1268 in Xiphophorus highlights hybrid incompatibility. PNAS. 2020;117: 29786–29794.
1269 doi:10.1073/pnas.2010133117

1270 66. Alcázar R, von Reth M, Bautor J, Chae E, Weigel D, Koornneef M, et al. Analysis of a
1271 plant complex resistance gene locus underlying immune-related hybrid incompatibility and
1272 its occurrence in nature. PLoS Genet. 2014;10: e1004848.
1273 doi:10.1371/journal.pgen.1004848

1274 67. Schumer M, Powell DL, Corbett-Detig R. Versatile simulations of admixture and accurate
1275 local ancestry inference with mixnmatch and ancestryinfer. Mol Ecol Resour. 2020;20:
1276 1141–1151. doi:10.1111/1755-0998.13175

1277 68. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, et al. The landscape of
1278 recombination in African Americans. Nature. 2011;476: 170–175. doi:10.1038/nature10336

1279 69. Gravel S. Population Genetics Models of Local Ancestry. Genetics. 2012;191: 607.
1280 doi:10.1534/genetics.112.139808

1281 70. Powell DL, Payne C, Banerjee SM, Keegan M, Bashkirova E, Cui R, et al. The Genetic
1282 Architecture of Variation in the Sexually Selected Sword Ornament and Its Evolution in
1283 Hybrid Populations. Current Biology. 2021;31: 923-935.e11.
1284 doi:10.1016/j.cub.2020.12.049

1285 71. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
1286 features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

1287 72. Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing
1288 and analyzing multiple genome alignments. Bioinformatics. 2013;29: 1341–1342.
1289 doi:10.1093/bioinformatics/btt128

1290 73. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A
1291 fast and versatile genome alignment system. PLOS Computational Biology. 2018;14:
1292 e1005944. doi:10.1371/journal.pcbi.1005944

1293