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 surprisingly 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 [8,35–40]. 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; [41,42]). 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 [43] or similarities in underlying genetic networks [44] and among close relatives, 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 the presence of strong DMIs between them [23]. Here, we expand on our recent description of replicate hybrid populations of the species pair *X. birchmanni* × *X. cortezi* [45] 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) [45]. Hybrid populations between both species pairs formed recently (*X. birchmanni* × *X. cortezi* hybrid populations: ~150 generations ago; *X. birchmanni* × *X. malinche* hybrid populations: ~100 generations ago [20,45]), likely due to human disturbance [46]. 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; [45]) and lower in *X. malinche* ($\pi \sim 0.03\%$ per basepair; [35]). *X. birchmanni*'s 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 [20,35,45,47], likely due to human-mediated habitat disturbance [46].

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 [48]. 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 (population disturbance) have likely driven recent hybridization across the two species pairs [45,46], each replicate population occurs in a distinct ecological environment, and varies in the degree of assortative mating observed [35,45,47,48]. We describe what is known about the ecological and mating environments of each population from past work in Supporting Information 1.

Description of X. birchmanni × X. cortezi hybrid populations at the Huextetitla and Santa Cruz sites

Initial work exploring the demographic history of the *X. birchmanni* × *X. cortezi* hybrid populations at Santa Cruz and Huextetitla suggested that both hybrid populations formed recently (~140-170 generations ago, or ~45-85 years ago), have not been severely bottlenecked since their formation, and experience low levels of migration (m<0.2%) from the *X. birchmanni* parental species (see [45] and Supporting Information 2 for more information). We used a larger sample of individuals from both populations to explore this history in greater detail. The Santa Cruz and Huextetitla hybrid populations occur along geographically separated tributaries of the same river (Supporting Information 2) and have similar genome-wide admixture proportions and ancestry structure [45]. 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 242 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 similar analyses for the Huextetitla hybrid population (for which results are qualitatively similar) in parts of the main text and supplement.

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 (Fig. S3). Since demographic analyses suggested that the *X. birchmanni* × *X. cortezi* hybrid populations have not been severely bottlenecked (see above, Supporting Information 3, [45]), this variance in local ancestry could be indicative of selection on hybrids. Indeed, simulations using the inferred demographic parameters for the Santa Cruz *X. birchmanni* × *X. cortezi* hybrid population [45] suggest that the patterns of fixation we observe for both major and minor parent ancestry are unexpected under the most likely scenario of neutral admixture (Fig. S4; Supporting Information 3).

We next evaluated correlations between minor parent ancestry and 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]. We found strong evidence in replicate hybrid populations of *X. birchmanni* × *X. cortezi* 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. S5-6). This pattern is robust across spatial scales, to thinning windows based on linkage disequilibrium, and when controlling for a number of possible technical variables (Table S1, S2).

We also investigated how ancestry across the genome in replicate *X. birchmanni* × *X. cortezi* hybrid populations covaries with coding and conserved basepairs. Similar to our previous findings in *X. birchmanni* × *X. malinche* hybrid populations, both coding and conserved basepairs are more weakly predictive of local ancestry than the local recombination rate (Table S1-S3). After accounting for variation in recombination rate, we find that minor parent ancestry is more depleted in regions of the genome with a higher number of linked coding and conserved basepairs (Fig. 2; 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 S3). 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. S7).

Sources of selection driving variation in minor parent ancestry

After confirming in 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), we asked about the role different sources of selection might play in driving broad-scale ancestry patterns 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 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 [20,23,35,49]. 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. S8; 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 [35,50]. 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 [35,50].

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 [48].

We observe surprisingly high repeatability in patterns of minor parent ancestry between most pairs of hybrid populations formed between these distinct species pairs (in 100 kb windows $\rho_{\text{Santa Cruz - Acuapa}} = 0.25$, $p<10^{-93}$; $\rho_{\text{Santa Cruz - Aguazarca}} = 0.17$, $p<10^{-42}$; Fig. 2; Table S4). 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.

Partial correlation analysis indicated that some of this signal of shared ancestry across hybrid population types can be explained by shared local recombination rates and the locations of coding and conserved basepairs (Table S5), since these features are largely conserved across *X. birchmanni*, *X. malinche*, and *X. cortezi* (Supporting Information 4). However, after accounting for 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 windows $\rho_{\text{Santa Cruz - Acuapa}} = 0.15$, $p<10^{-33}$; $\rho_{\text{Santa Cruz - Aguazarca}} = 0.09$, $p<10^{-15}$; Table S5). 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. S9; 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 S4, S5, S6; 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 types 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 81 minor parent "deserts" as genomic regions where the minor parent ancestry fell in the lower 2.5% tail of genome-wide ancestry (see Methods; Fig. S10). 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). 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), and the number of shared regions is greater than would be expected to be identified by chance (see Methods, Supporting Information 7, Fig. 3).

We similarly 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. Sharing of both low and high minor parent ancestry regions exceeded sharing expected by chance (Fig. 3A; Methods, Supporting Information 7; $p<0.001$ by simulation).

We 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. S11-12; 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. S13). 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. S14; Supporting Information 8).

Features coinciding with shared minor parent ancestry deserts and islands

The unexpectedly high number of shared minor parent ancestry deserts and islands indicate that 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. S15), 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 islands) on chromosome 21 (Supporting Information 8), possibly because sex chromosomes are young and not strongly differentiated between males and females [51].

Of the 19 regions with especially high minor parent ancestry across hybrid population types, two correspond to an inversion that occurs on chromosome 17 and one corresponds to an inversion on chromosome 19 (Fig. S16). Intriguingly, inversions that differed between species had unusually high minor parent ancestry in all hybrid populations on average (Fig. S17), in contrast to what is most frequently observed in other systems ([52–54] but see [40,55,56]).

The remaining minor parent deserts and islands do not coincide with known structural differences between species [35,45], mapped hybrid incompatibilities between *X. birchmanni* × *X. malinche* [20,23,49], or mapped QTL underlying trait differences between the species [23,57]. We investigate other features associated with these regions below.

Ancestry at sites under selection in X. birchmanni × X. malinche hybrids

We observe some overlap between known hybrid incompatibilities and minor parent ancestry deserts, suggesting that this is one mechanism that can drive cross-population repeatability in ancestry (among several plausible mechanisms; Supporting Information 6). To test the role of hybrid incompatibilities in driving cross-repeatability more formally, we took advantage of sites likely to be involved in hybrid incompatibilities in replicate *X. birchmanni* × *X. malinche* hybrid populations and asked if they had particularly low minor parent ancestry in *X. birchmanni* × *X. cortezi* hybrid populations.

Nearly a decade of work has aimed to identify hybrid incompatibilities between *X. birchmanni* × *X. malinche* using natural hybrid populations and artificial crosses [20,23,35,49]. We focused on two available datasets: putative segregating DMIs identified in natural hybrid populations between *X. birchmanni* × *X. malinche* [20,49] as well as six segregation distorters in F_2 hybrids of lab crosses between *X. birchmanni* × *X. malinche* that show ancestry patterns consistent with the action of strongly selected DMIs (e.g. s>0.25; see Methods; Fig. S17). For the segregation distorter dataset, we found that sites under selection in *X. birchmanni* × *X. malinche* F_2 hybrids had lower minor parent ancestry in the replicate *X. birchmanni* × *X. cortezi* hybrid populations (Fig. 3B). Moreover, we show that this result is robust to excluding a segregation distorter that is also a shared minor parent ancestry desert across all populations (Fig. S18-19). We do not see an effect of DMIs inferred to be under weaker selection mapped in *X. birchmanni* × *X. malinche* natural hybrid populations, which could indicate that they are not enriched for hybrid incompatibilities between *X. birchmanni* × *X. cortezi* or that we have low power to detect incompatibilities under weak selection [20].

Mapping a newly identified shared hybrid incompatibility

In the previous section we showed that hybrid incompatibilities identified in one species pair are enriched for low minor parent ancestry in natural hybrid populations formed between another species pair, indicating that some hybrid incompatibilities are repeatable cross the two types of hybrid populations. One striking signal we observe in this data is the coincidence of an ancestry desert on chromosome 6 shared across all populations where *X. birchmanni* is the minor parent and a strong segregation distorter identified in F_2 hybrids between *X. birchmanni* × *X. malinche* (Fig. 3C-D; see Methods). Further examination of this region indicated that it is depleted for minor parent ancestry in all replicate *X. birchmanni* × *X. malinche* populations (Fig. 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. S20; empirical p-value based on ancestry in 10 kb 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 [58] 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 genes have several nonsynonymous substitutions that distinguish the 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. S21; see also Moran et al. 2021), 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 (Moran et al. 2021).

Analyses of other regions provided some evidence for the presence of additional shared hybrid incompatibilities among 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. S22; 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. $s > 0.5$; 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).

Table 1. Nucleotide (Nt) and amino acid changes between species pairs (AA) in interacting genes between the chromosome 6 and chromosome 13 regions. Ratios of rates of nonsynonymous to synonymous substitutions (dN/dS) inferred by codeml [59] are also shown.

	ndufs5			***ndufa13***		
Species Comparison	dN/dS	Nt changes	AA changes	dN/dS	Nt changes	AA changes
X. birchmanni - X. cortezi	102	5	5	98	3	3
X. birchmanni - X. malinche	102	5	5	1.2	4	3
X. cortezi – X. malinche	0	0	0	0.001	1	0

411 **Discussion**
412
413 We now know that species across the tree of life have hybridized with their close
414 relatives and that many species derive substantial proportions of their genomes from past
415 hybridization events. Based on these observations, it is likely that hybridization plays an
416 important role in the evolution of genomes and species, but we still understand relatively little
417 about how this process unfolds. There are major open questions about which mechanisms shape
418 retention and loss of minor parent ancestry after hybridization and how repeatable this process is
419 across species.
420 While the predictability of evolution has been a classic question in evolutionary biology
421 for decades [60], such questions have been less central to the hybridization literature [6]. With
422 the expanding number of known hybridization events between closely related species, we are
423 now poised to tackle these questions. One general principle that has emerged is that ancestry
424 from the minor parent species, or the species that contributed less to the genome of hybrids, is
425 more frequently purged by selection after hybridization [6]. This purging can be explained by
426 several mechanisms of selection on hybrids, including selection against hybrid incompatibilities
427 [19,25,28,29,35], and tends to be stronger in regions of the genome with many linked coding or
428 conserved basepairs [8,12,35]. We also uncover these patterns in the *X. birchmanni* × *X. cortezi*
429 hybrid populations studied here, showing a substantial effect of local recombination rate and
430 linked coding and conserved basepairs on minor parent ancestry in hybrids (Fig. 2). This
431 represents at least the sixth system in which these effects have been detected [8,12,35,36,38,40],
432 suggesting that these patterns of selection on minor parent ancestry are predictable outcomes of
433 the process of genome stabilization after hybridization.
434 Here, we dramatically expand our understanding of how repeatable the outcomes of
435 hybridization are by comparing patterns of local ancestry in hybridization events between *X.*
436 *birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* species pairs. This allows us to
437 evaluate, for the first time, some of the factors driving repeatability in the outcomes of
438 hybridization across related species pairs with similar histories of hybridization. Both
439 hybridization events began recently (in the last ~100-150 generations [20,35]) but occur in
440 different geographic regions between distinct species pairs (Fig. 1). Because we find that broad
441 scale factors like genome structure are important in generating local ancestry variation in both
442 species pairs, this is expected to drive some concordance in local ancestry across the two species
443 pairs. Indeed, we detect repeatability in where in the genome minor parent ancestry is retained or
444 purged across pairs of *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid
445 populations (Fig. 2; Table S4).
446 Beyond these broad patterns generated by shared genetic architecture, little is known
447 about whether the same genes or regions of the genome are intolerant of hybridization in related
448 species pairs. Intriguingly, we still observe correlations in local ancestry across the two types of
449 hybrid populations after accounting for shared genomic architecture (Table S5). This indicates
450 that although shared genome structure plays a role in some of the repeatability in local ancestry
451 we observe, other factors are likely to contribute, such as shared sites under selection across the
452 two types of hybridization events. Indeed, when we searched for local signals of shared ancestry
453 across the two hybrid population types, we found evidence that some of the same regions of the
454 genome are under selection in hybrid populations of both species pairs. Compared to expected
455 overlap by chance, *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid

populations are ~25-27X enriched in shared minor parent deserts and ~18-25X enriched in shared minor parent 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 (Moran et al. 2021). 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. 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. cortezi* 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 such as masking of weakly deleterious alleles that can favor minor parent ancestry [61]. Notably, three shared minor parent islands occur in inversions that differentiate species, conflicting with the prevailing wisdom that inversions should resist introgression (Fig. S16; see [40,55,56] 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. However, incompatibilities can also arise between the ancestral and derived genotypes when multiple derived mutations have accumulated along one lineage (Fig. S23; [43]). 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 [21,23,62,63]. 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 surprisingly 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 [45]. 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. [45]. 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 [45]. 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 [45]. 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 [45,64]. 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 [45]. 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 [64]. 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 window sizes (e.g. 50 kb – 1 Mb; 0.1 – 1 cM).

Historical recombination events in admixed populations are detectable as ancestry transitions in present day individuals [65]. 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 [35,57]. 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

617 analysis to individuals collected over a single sampling year. Briefly, we included 95 individuals
618 collected from Tlatemaco in 2013 (resulting in 629,584 ancestry informative sites), 97
619 individuals collected from Acuapa in 2018 (613,171 sites), and 51 individuals collected in 2016
620 from Aguazarca (628,811 sites).
621 Previous work has explored the accuracy of these local ancestry inference approaches in
622 *X. birchmanni* × *X. malinche* hybrids in great detail [23,64]. Briefly, initial definition of ancestry
623 informative markers relied on 183 parental individuals [23], ancestry informative sites that were
624 not fixed or nearly fixed between the two parental species were excluded (<98% frequency
625 difference). Performance was evaluated in simulations, on pure parental individuals not used in
626 the definition of ancestry informative sites and on early generation F_1 and F_2 hybrids where small
627 switches in ancestry can confidently be inferred to be errors. Based on these analyses we
628 estimated error rates to be ≤0.2% errors per ancestry informative site in all tested scenarios
629 [23,64]. Thus, given high expected accuracy of ancestry inference in both *X. birchmanni* × *X.*
630 *cortezi* and *X. birchmanni* × *X. malinche* hybrids, we do not expect that errors will drive to the
631 global or local patterns that we observe in the empirical data.
632
633 *Comparing minor parent ancestry as a function of local genetic architecture*
634 For the *X. birchmanni* × *X. cortezi* hybrid populations, we wanted to evaluate whether
635 minor parent ancestry was correlated with recombination rate, as had previously been described
636 in other species [35,36,40]. We summarized average ancestry in 50 kb, 100 kb, and 250 kb
637 windows for each population. We previously created a linkage disequilibrium (LD) based
638 recombination map for *X. birchmanni* [35] and calculated recombination rate based on this map
639 over these same spatial scales. Available data suggests that recombination maps are likely to be
640 conserved between *X. birchmanni* and *X. cortezi* (Supporting Information 4; [35]), and between
641 parental species and hybrids [35]. Because the local recombination rate covaries with other
642 genomic features in swordtails [35], we also calculated the number of coding (or conserved)
643 basepairs in each window so that we could account for this in our analysis. We performed these
644 calculations using annotations developed for the *X. birchmanni* genome [23]. We then used a
645 Spearman's partial correlation in R to determine if ancestry was correlated with recombination
646 rate, after accounting for the number of coding and conserved basepairs in each window. To aid
647 in interpreting these empirical results, we also analyzed simulations that were performed under
648 the inferred demographic history of *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche*
649 hybrid populations, with and without selection (Supporting Information 3,6).
650 After evaluating the relationship between recombination rate and ancestry in *X.*
651 *birchmanni* × *X. cortezi* populations, we were also interested in the relationship between minor
652 parent ancestry and the locations of coding and conserved basepairs along the genome. To
653 control for recombination rate, which is positively correlated with coding and conserved bases in
654 swordtails [35], we summarized average ancestry in 0.25 cM windows. As noted above, we used
655 the gene annotation file for *X. birchmanni* and a set of regions conserved in percomorph fish that
656 we had previously identified using phastCons. Comparisons of alignments across species
657 indicate that the locations of these elements are largely conserved across *X. birchmanni*, *X.*
658 *malinche*, and *X. cortezi* (Supporting Information 4). We counted the number of coding or
659 conserved basepairs in each window. Using this approach, we evaluated the relationship between
660 ancestry and the number of linked coding or conserved basepairs with a Spearman's correlation
661 (and partial correlation) in R. We chose to present results for 0.25 cM windows in the main text
662 because 50% of minor parent ancestry tracts in Santa Cruz are 0.25 cMs in length or longer (75%

663 are 0.1 cMs in length or greater). However, we evaluated these relationships across a range of
664 window sizes (0.1-1 cM) and found the results were qualitatively similar (Table S3).
665 For analyses including recombination rate, we repeated tests using a similar LD-based
666 map that we inferred using the same methods [35] after thinning SNPs in regions of the genome
667 with high SNP density in our *X. birchmanni* population sample. Because our ability to reliably
668 estimate recombination rate depends on the density of SNPs, we expect our power to estimate
669 recombination rate to vary along the genome. If power to estimate recombination rate varies with
670 other features of interest, such as density of coding or conserved basepairs, this could mislead us
671 into interpreting variation in power as variation in signal. To evaluate this possibility, we
672 calculated the median distance between SNPs in our dataset, which was 184 basepairs. We
673 thinned the SNPs input into LDhelmet so that when SNPs were more dense than one every 184
674 basepairs, a single SNP was retained per 184 basepairs. We then re-inferred recombination rates
675 using LDhelmet as described previously [35]. We repeated all analyses described above that
676 incorporate recombination rate with this thinned LD map and found the results qualitatively
677 unchanged (Supporting Information 11).
678
679 *Repeatability of local ancestry between cross types*
680 To evaluate correlations in minor parent ancestry across *X. birchmanni* × *X. cortezi* and
681 *X. birchmanni* × *X. malinche* populations we used summaries of ancestry in physical (50-250 kb)
682 and genetic windows (0.1-1 cM), as described above. We first calculated pairwise correlations
683 between populations in minor parent ancestry without accounting for other features (Fig. 2) and
684 then repeated these calculations using partial correlation tests that included the number of linked
685 coding and conserved basepairs as covariates (and recombination rate where applicable; Table
686 S4-S5). This allowed us to infer both whether broad-scale correlations in local ancestry could be
687 detected across populations and whether these correlations could be largely explained by shared
688 genetic architecture (i.e. locations of coding and conserved basepairs and local recombination
689 rate). We also explored this question using simulations. We performed simulations under the
690 inferred demographic histories for *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche*
691 hybrid populations with and without selection (Supporting Information 3,6) and evaluated
692 expected cross-population correlations in ancestry under different scenarios.
693 Beyond shared correlations as a result of shared broad-scale architecture of the genome,
694 we were interested if we could identify shared minor parent ancestry "deserts," where minor
695 parent ancestry was unusually low, and shared minor parent ancestry "islands, where minor
696 parent ancestry was unusually high, that were shared across *X. birchmanni* × *X. cortezi* and *X.*
697 *birchmanni* × *X. malinche* hybrid populations.
698 To do so we identified regions of especially high and low minor parent ancestry in Santa
699 Cruz, the *X. birchmanni* × *X. cortezi* hybrid population for which we have the largest dataset (for
700 results of the same analysis using the Huextetitla populations see Supporting Information 12).
701 We first identified ancestry informative sites where the minor or major parent ancestry fell in the
702 lower 2.5% tail of genome-wide ancestry. We expanded out in the 5' and 3' direction from the
703 focal ancestry informative site until we reached a site on each edge that exceeded the 5%
704 ancestry quantile in Santa Cruz, and treated this interval as the boundary of the minor parent
705 ancestry desert or island. To control for possible technical artifacts, we then evaluated ancestry in
706 these regions in 0.05 cM windows. We identified the window that contained the midpoint of the
707 islands or desert and checked that this entire window fell in the 10% minor or major ancestry tail,
708 contained at least 10 ancestry informative sites, and was larger than 10 kb in length. We also

merged any detected regions that were closer together than 50 kb (deserts N=54, islands N=14) and removed any regions shorter than 1 kb (deserts N=1). We also filtered regions where inferred ancestry varied dramatically between inference with unthinned and thinned ancestry informative markers (Fig. S11). Specifically, we filtered islands where inferred ancestry differed by more than 10% between thinned and unthinned data (N=58, based on 0.05 cM windows). This resulted in the identification of 81 minor parent ancestry deserts with an average length of 317 kb, and 76 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 5% (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 (~40%).

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 change, 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. S12-S14.

Given the importance of this analysis for the conclusions of our manuscript, we also took a second approach to determining the expected false positive rate for desert and island regions that is described in detail in Supporting Information 7. 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 and found that the results were qualitatively similar (Supporting Information 7).

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 [20,49].

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 [66]. 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 [66]. The reverse cross direction is rarely successful (Moran et al. 2021). We inferred local ancestry as described previously for this cross [64,66]. 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 [66].

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. S18).

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 randomly drew a selection coefficient from 0-1 against one of the homozygous genotypes and a dominance coefficient (0-1). We implemented selection on this pool of genotypes, treating 1-*s* (or 1-*hs*) as the probability of survival. We drew 943 individuals from the surviving genotypes and calculated average ancestry at the simulated sites. Simulations where ancestry fell in the 5% tail of the real data were treated as accepted simulations; we repeated this procedure 10,000 times. Based on the distribution of accepted simulations (Fig. S24), we predict that very strong selection ($s > 0.25$) is required to generate ancestry patterns as extreme as the loci we focus on in our F_2 dataset. Distributions of the dominance coefficient from accepted simulations mirrored the prior distribution.

Role of loci under selection in ancestry variation in natural populations

One of our major questions is whether loci under selection are shared not just across hybrid populations of the same cross type but between hybrid populations of different cross types. To tackle this question we took advantage of what is known about selection in the *X. birchmanni* × *X. malinche* system, using the segregation distortion regions identified above, and more weakly selected putative DMIs that have been previously identified in natural hybrid populations between *X. birchmanni* × *X. malinche* [20,49]. Both sets of loci are associated with lower minor parent ancestry in *X. birchmanni* × *X. malinche* hybrid populations [35], as expected by theory.

We asked if these loci were also associated with lower minor parent ancestry in *X. birchmanni* × *X. cortezi* hybrid populations, and the degree to which they coincided with ancestry deserts that are shared across *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations. We summarized ancestry in 10 kb windows and used bedtools [67] to determine the distance of each window to the closest selected site in each of the two datasets and plotted average major parent ancestry as a function of that distance (Fig. S25).

Identifying potential partners of shared ancestry desert and segregation distortion region

One of the regions that we identified as a significant segregation distorter on chromosome 6 in F_2 hybrids between *X. birchmanni* × *X. malinche* was also a shared minor ancestry desert, and had unusually low minor parent ancestry across all surveyed *X. birchmanni* × *X. malinche* and *X. birchmanni* × *X. cortezi* hybrid populations. The observation that minor parent ancestry is disfavored across populations of diverse ancestry (majority *malinche*, majority *birchmanni*, and majority *cortezi*) indicates that this locus is likely to be involved in a hybrid incompatibility [35], as opposed to being driven by other possible mechanisms of selection on hybrids [6].

Based on these observations, we wanted to identify the interacting partner of the chromosome 6 locus. Given the clear expectations for genotype frequencies in F_2 hybrids and apparent shared selection on this locus across species pairs, we used the F_2 data generated between *X. birchmanni* × *X. malinche* to perform this scan. We selected an ancestry informative marker from the center of the shared ancestry desert on chromosome 6, which falls within the peak segregation distortion region on chromosome 6 identified in F_2 hybrids (Fig. 3). Next, we performed a scan against all other ancestry informative markers across the genome and calculated χ^2 statistics for deviations from expected two-locus genotype frequencies (with three

846 degrees of freedom). Importantly, we used the observed genotype frequencies across individuals
847 at each of the two loci to generate this expectation, such that we identified loci with deviations
848 from expected two-locus genotype combinations rather than deviations in ancestry at one locus
849 or the other.
850 To determine the appropriate significance threshold for this scan, we performed
851 simulations. Using the F_2 ancestry data described in the previous paragraph, we simulated
852 genotypes at the focal locus based on each individual's genome-wide ancestry and performed the
853 χ^2 test as described above. We repeated this for every locus in the dataset (i.e. over 20,011
854 markers) and recorded the minimum p-value. We performed this procedure 500 times and took
855 the lower 5% quantile of minimum p-values across simulations. This value served as our false
856 positive rate (FPR) cutoff for analysis of the real data. We describe this approach in more detail
857 in Supporting Information 9. We initially explored an approach using permutations at the focal
858 locus to set the FPR but found that this approach was less conservative; these results are also
859 described in Supporting Information 9. Only one region in the genome (found on chromosome
860 13) passed the FPR threshold determined by permutation.
861 We also repeated this scan for interacting loci for all ancestry deserts shared between *X.*
862 *birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* populations. While we did not identify
863 any interacting loci that passed the 5% genome-wide FPR correction threshold for other regions,
864 we highlight two other cases where interactions are detected with minor parent deserts at a FPR
865 of 10% (Supporting Information 9). Simulations suggest that our power to detect interactions in
866 is relatively low (Supporting Information 9), leaving open the possibility that other minor parent
867 deserts interact with partner loci that we are unable to detect in the current data.
868
869 *Pathway analysis of genes falling in a shared hybrid incompatibility*
870 We wanted to ask whether genes in the identified chromosome 6 and chromosome 13
871 regions might be known to interact from past experimental, fusion, or co-expression data. To
872 evaluate this, we took advantage of the STRING database [58]. We used haltools [68] to liftover
873 coordinates from the *X. birchmanni* reference genome to the *X. maculatus* reference genome for
874 the shared ancestry desert on chromosome 6 and the interacting region on chromosome 13. This
875 allowed us to extract gene IDs for *X. maculatus* in this region, which is included in the STRING
876 database.
877 We input gene names from both chromosomes (chromosome 6 – 13 genes; chromosome
878 13 – 29 genes) into STRING using the protein batch search function. Two pairs of interacting
879 genes with supporting experimental, fusion, or co-expression data were identified. For one of
880 these pairs, both genes fell within the chromosome 6 region but for the other pair, one gene fell
881 within the chromosome 6 region and the other within the chromosome 13 region.
882 Given that our power to map interacting regions of shared minor parent deserts is
883 relatively low (Supporting Information 9), we wanted to perform a broad scan to ask whether
884 there was evidence for enrichment of genes that have many protein-protein interactions or dense
885 regulatory networks within shared minor parent ancestry deserts and islands. We thus performed
886 similar analyses for genes found in other shared minor parent ancestry deserts and islands. These
887 analyses included both STRING database searches similar to those described above, and
888 evaluation of enrichment of hub genes from co-expression analyses in minor parent deserts and
889 islands compared to matched null sets (Supporting Information 10).
890
891 *Investigating other features associated with ancestry variation*

892 Because theory predicts that many of the changes in genome-wide ancestry that occur in
893 hybrid populations happen rapidly after admixture [29,39], we were interested in investigating
894 correlations between minor parent ancestry and chromosome-level features. We evaluated
895 whether there were correlations between chromosome length, number of genes per chromosome,
896 and minor parent ancestry using a Spearman's rank correlation test.
897 We also evaluated minor parent ancestry as a function of structural differences detected
898 between the *X. birchmanni*, *X. malinche*, and *X. cortezi* assemblies. To identify chromosomal
899 inversions between species, we compared de-novo assemblies using MUMmer (mummer-
900 4.0.0beta2) [69]. After aligning the 24 largest linkage groups, which correspond to the 24
901 *Xiphophorus* chromosomes, we excluded alignments under 5 kb before proceeding with the
902 analysis. We focused our analysis on scanning for large inversions (> 100 kb). To identify the
903 putative inversion breakpoints, we used mummer's show-coords command. By aligning the *X.*
904 *birchmanni* [23] assembly to *X. cortezi* [45] and to *X. malinche* [23], and comparing to an
905 outgroup sequence (*X. maculatus;* [51]), we determined which inversions were derived in each
906 species. For simplicity, we assumed the inversion configuration detected in the reference
907 genomes was fixed at the species level (π in *X. birchmanni*, *X. malinche*, and *X. cortezi* is
908 <0.12%) but we note that some segregating inversions have been detected in *Xiphophorus* [23].
909

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

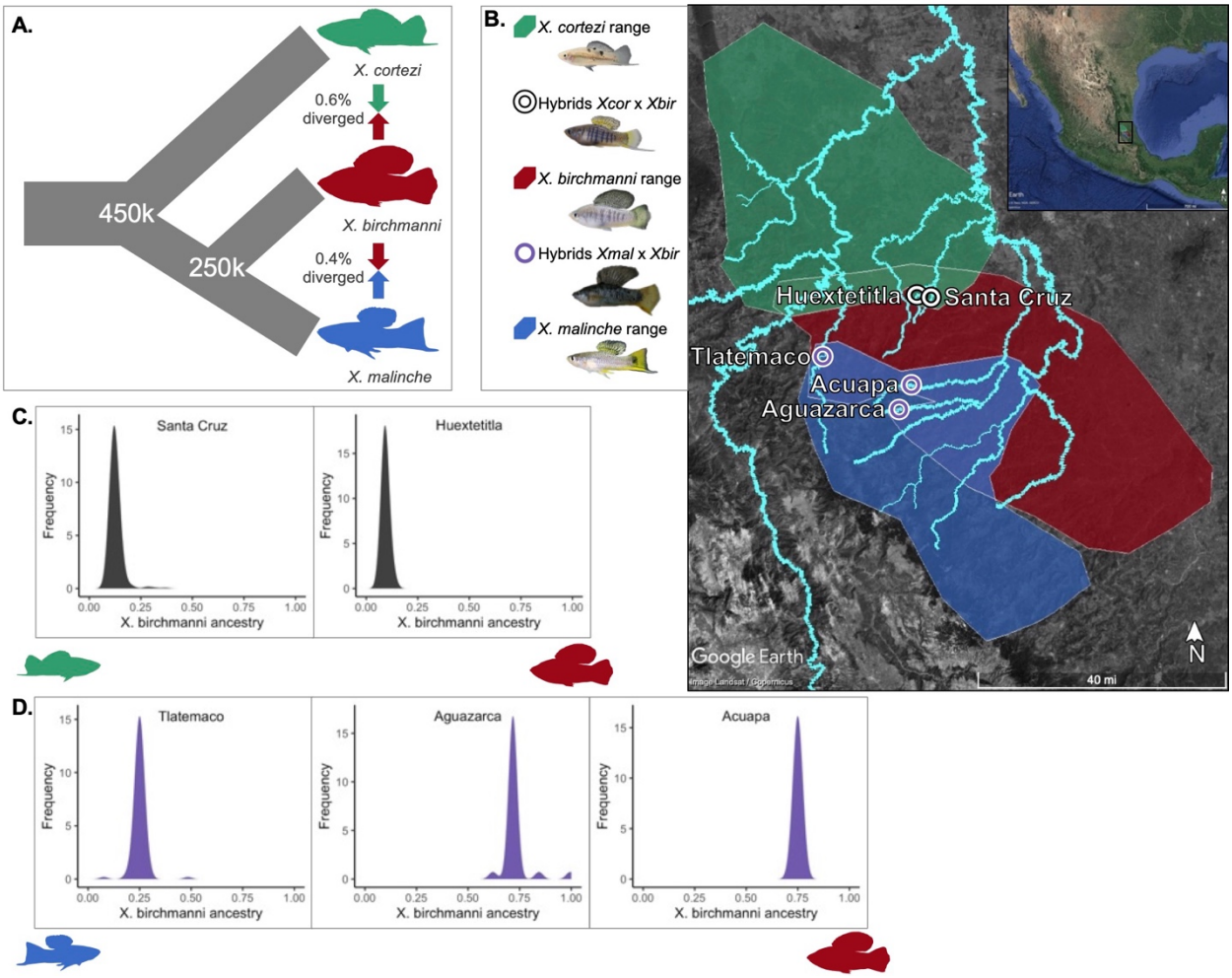

Figure 1. Description of focal species, study sites, and hybrid population ancestry structure. **A.** Phylogenetic relationships and contemporary hybridization events between focal species *X. birchmanni*, *X. malinche*, and *X. cortezi*. **B.** Distributions of *X. birchmanni* (*Xbir*), *X. malinche* (*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).

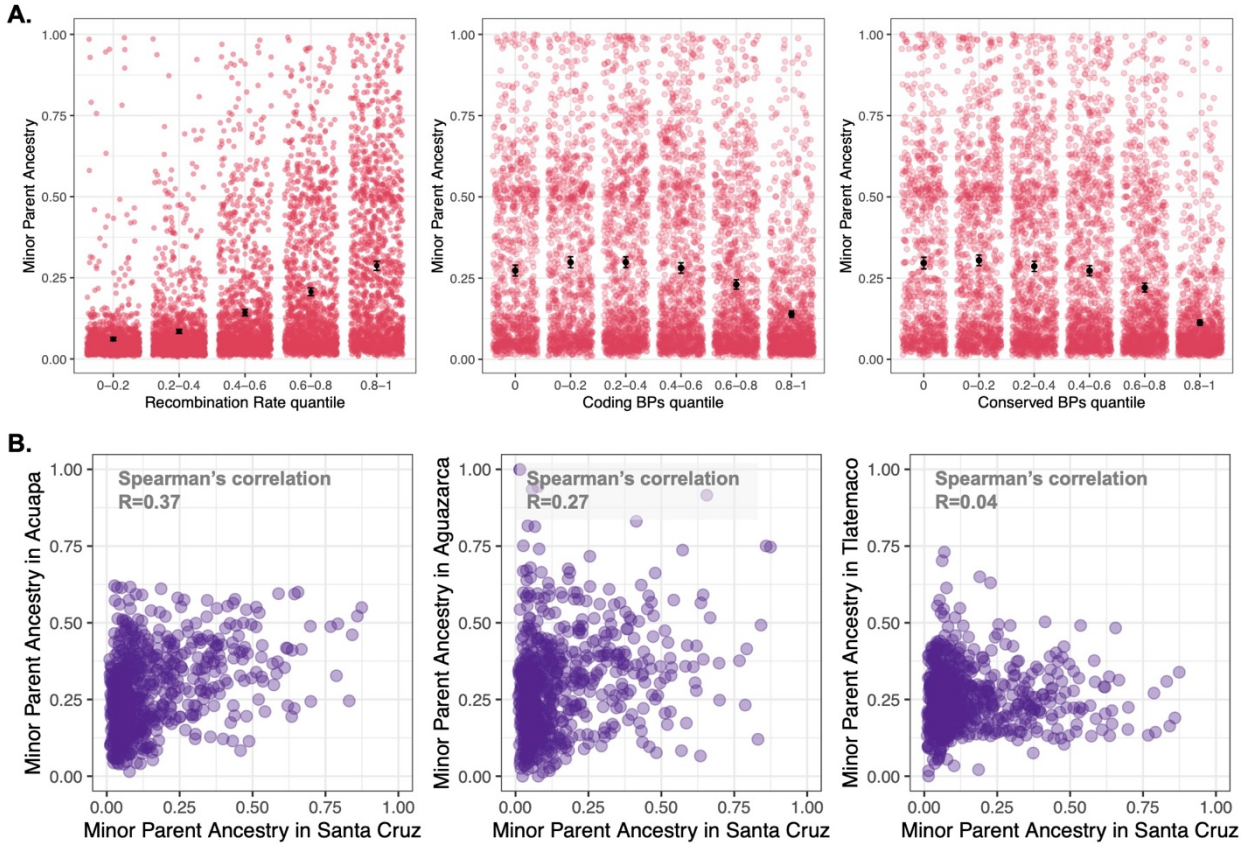

Figure 2. Relationship between minor parent ancestry, recombination rate and functional basepairs in *X. birchmanni* × *X. cortezi* hybrid populations, as well as cross-population type comparisons. **A.** Minor parent ancestry (*X. birchmanni* ancestry) in the Santa Cruz hybrid population is higher in regions of the genome with higher recombination rates. Semi-transparent points show results from individual 50 kb 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 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, S3). 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 windows. Cross-population ancestry correlations between Santa Cruz (or Huextetila – Table S4) 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 S5).

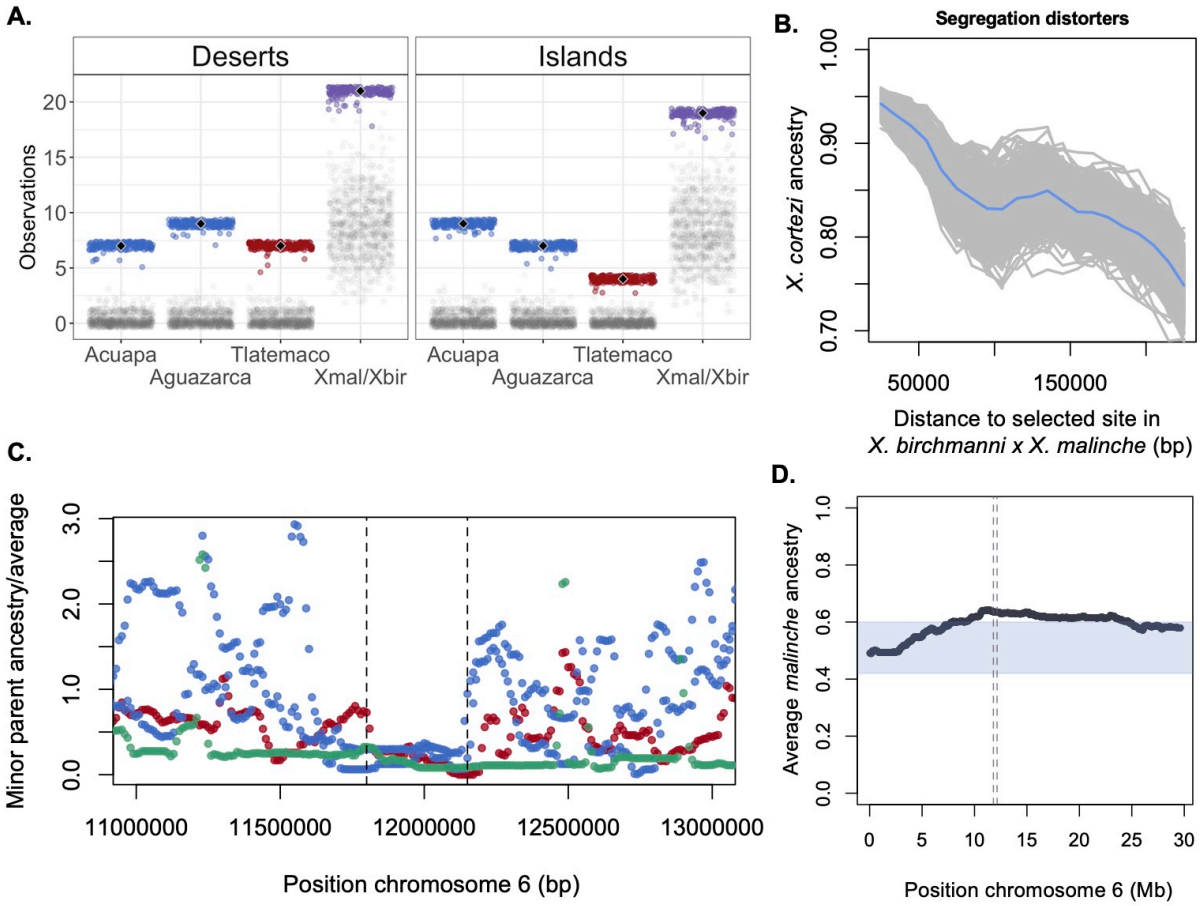

Figure 3. A. Shared minor parent deserts and islands are enriched compared to expectations by chance between *X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche* hybrid populations. Results shown here indicate the number of shared minor parent deserts (or islands) between the Santa Cruz *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. **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 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*). **D**. The shared minor parent ancestry desert on chromosome 6 coincides with a strong segregation distorter in artificial crosses between *X. birchmanni* × *X. malinche*. Plotted here is average ancestry at ancestry

980 informative sites across 943 F_2 hybrids, where expected ancestry is 50% *X. birchmanni* and 50%
981 *X. malinche*. The blue envelop indicates the 99% quantile of variation in ancestry in this cross
982 genome-wide. The dashed lines indicate the location of the shared ancestry desert in hybrid
983 populations plotted in **C**.

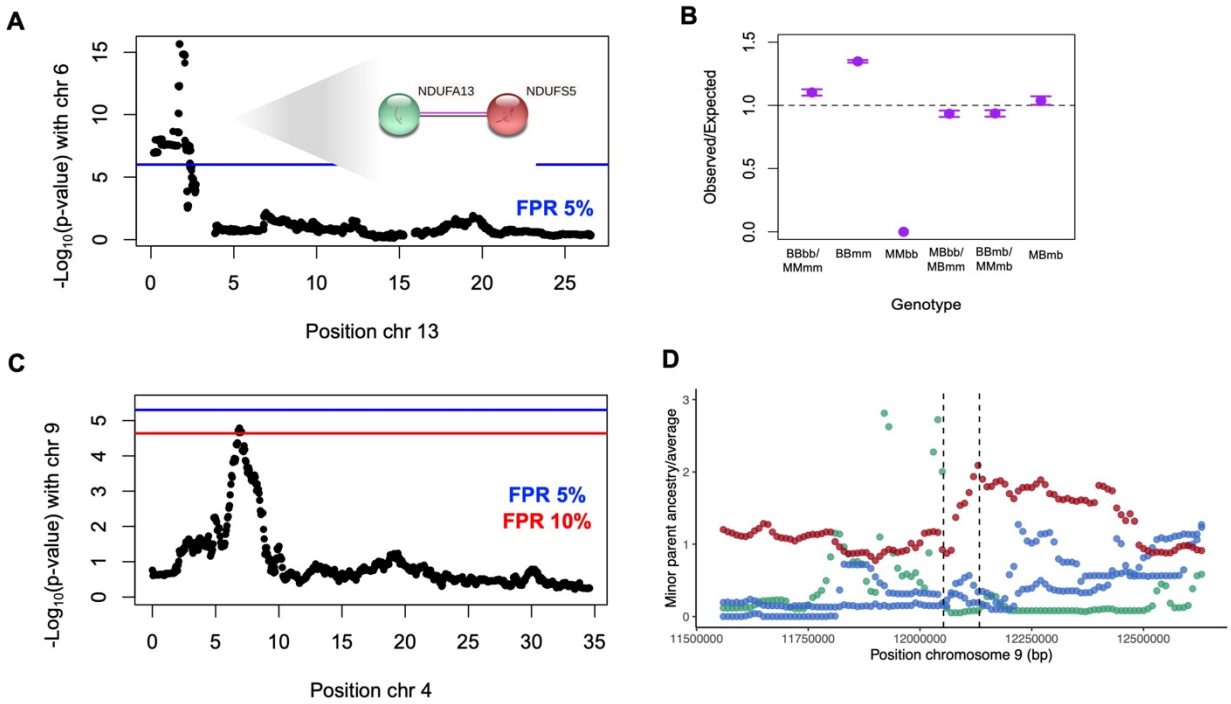

Figure 4. Evidence for repeated hybrid incompatibilities between *X. birchmanni* × *X. malinche* and *X. birchmanni* × *X. cortezi* hybrids. **A**. Interacting locus of the shared chromosome 6 ancestry desert identified in a genome-wide scan. This region was identified scanning for deviations from expected 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 a 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 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*).

1010 **References**
1011

1012 1. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete
1013 and error-free genome assemblies of all vertebrate species. bioRxiv. 2020;
1014 2020.05.22.110833. doi:10.1101/2020.05.22.110833

1015 2. Genereux DP, Serres A, Armstrong J, Johnson J, Marinescu VD, Murén E, et al. A
1016 comparative genomics multitool for scientific discovery and conservation. Nature.
1017 2020;587: 240–245. doi:10.1038/s41586-020-2876-6

1018 3. The Global Invertebrate Genomics Alliance (GIGA): Developing Community Resources to
1019 Study Diverse Invertebrate Genomes. J Hered. 2014;105: 1–18. doi:10.1093/jhered/est084

1020 4. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and
1021 speciation. Journal of Evolutionary Biology. 2013;26: 229–246. doi:10.1111/j.1420-
1022 9101.2012.02599.x

1023 5. Mallet J. Hybridization as an invasion of the genorne. Trends Ecol Evol. 2005;20: 229–237.
1024 doi:10.1016/j.tree.2005.02.010

1025 6. Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. The genetic
1026 consequences of hybridization. arXiv:201204077 [q-bio]. 2020 [cited 27 Feb 2021].
1027 Available: http://arxiv.org/abs/2012.04077

1028 7. Turissini DA, Matute DR. Fine scale mapping of genomic introgressions within the
1029 Drosophila yakuba clade. PLOS Genetics. 2017;13: e1006971.
1030 doi:10.1371/journal.pgen.1006971

1031 8. Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. Speciation and Introgression
1032 between Mimulus nasutus and Mimulus guttatus. PLOS Genetics. 2014;10: e1004410.
1033 doi:10.1371/journal.pgen.1004410

1034 9. D'Angiolo M, De Chiara M, Yue J-X, Irizar A, Stenberg S, Persson K, et al. A yeast living
1035 ancestor reveals the origin of genomic introgressions. Nature. 2020;587: 420–425.
1036 doi:10.1038/s41586-020-2889-1

1037 10. Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, et al. Major
1038 ecological transitions in wild sunflowers facilitated by hybridization. Science. 2003;301:
1039 1211–1216. doi:10.1126/science.1086949

1040 11. Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal G. Phylogenomics reveals
1041 extensive reticulate evolution in Xiphophorus fishes. Evolution. 2013;67: 2166–2179.

1042 12. Sankararaman S, Mallick S, Patterson N, Reich D. The Combined Landscape of Denisovan
1043 and Neanderthal Ancestry in Present-Day Humans. Current Biology. 2016;26: 1241–1247.
1044 doi:10.1016/j.cub.2016.03.037

1045 13. Heliconius Genome C. Butterfly genome reveals promiscuous exchange of mimicry
1046 adaptations among species. Nature. 2012;487: 94–8.

1047 14. Trier CN, Hermansen JS, Sætre G-P, Bailey RI. Evidence for Mito-Nuclear and Sex-Linked
1048 Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. PLOS
1049 Genetics. 2014;10: e1004075. doi:10.1371/journal.pgen.1004075

1050 15. Orr HA. The population genetics of speciation: the evolution of hybrid incompatibilities.
1051 Genetics. 1995;139: 1805–1813.

1052 16. Arnegard ME, McGee MD, Matthews B, Marchinko KB, Conte GL, Kabir S, et al. Genetics of
1053 ecological divergence during speciation. Nature. 2014;511: 307–311.
1054 doi:10.1038/nature13301

1055 17. Svedin N, Wiley C, Veen T, Gustafsson L, Qvarnström A. Natural and sexual selection
1056 against hybrid flycatchers. Proceedings of the Royal Society B: Biological Sciences.
1057 2008;275: 735–744. doi:10.1098/rspb.2007.0967

1058 18. Christie K, Strauss SY. Along the speciation continuum: Quantifying intrinsic and extrinsic
1059 isolating barriers across five million years of evolutionary divergence in California
1060 jewelflowers. Evolution. 2018;72: 1063–1079. doi:10.1111/evo.13477

1061 19. Juric I, Aeschbacher S, Coop G. The Strength of Selection against Neanderthal
1062 Introgression. PLOS Genetics. 2016;12: e1006340. doi:10.1371/journal.pgen.1006340

1063 20. Schumer M, Cui R, Powell DL, Dresner R, Rosenthal GG, Andolfatto P. High-resolution
1064 mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. McVean
1065 G, editor. eLife. 2014;3: e02535. doi:10.7554/eLife.02535

1066 21. Chae E, Bomblies K, Kim S-T, Karelina D, Zaidem M, Ossowski S, et al. Species-wide genetic
1067 incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. Cell.
1068 2014;159: 1341–1351. doi:10.1016/j.cell.2014.10.049

1069 22. Larson EL, Vanderpool D, Sarver BAJ, Callahan C, Keeble S, Provencio LL, et al. The
1070 Evolution of Polymorphic Hybrid Incompatibilities in House Mice. Genetics. 2018;209:
1071 845–859. doi:10.1534/genetics.118.300840

1072 23. Powell DL, García-Olazábal M, Keegan M, Reilly P, Du K, Díaz-Loyo AP, et al. Natural
1073 hybridization reveals incompatible alleles that cause melanoma in swordtail fish. Science.
1074 2020;368: 731–736. doi:10.1126/science.aba5216

1075 24. Zuellig MP, Sweigart AL. A two-locus hybrid incompatibility is widespread, polymorphic,
1076 and active in natural populations of Mimulus*. Evolution. 2018;72: 2394–2405.
1077 doi:10.1111/evo.13596

1078 25. Barton NH. The role of hybridization in evolution. Molecular Ecology. 2001;10: 551–568.
1079 doi:https://doi.org/10.1046/j.1365-294x.2001.01216.x

1080 26. Slatkin M, Lande R. Segregation variance after hybridization of isolated populations. Genet
1081 Res. 1994;64: 51–56. doi:10.1017/s0016672300032547

1082 27. Thompson KA. Experimental Hybridization Studies Suggest That Pleiotropic Alleles
1083 Commonly Underlie Adaptive Divergence between Natural Populations. The American
1084 Naturalist. 2020;196: E16–E22. doi:10.1086/708722

1085 28. Simon A, Bierne N, Welch JJ. Coadapted genomes and selection on hybrids: Fisher's
1086 geometric model explains a variety of empirical patterns. Evolution Letters. 2018;2: 472–
1087 498. doi:10.1002/evl3.66

1088 29. Harris K, Nielsen R. The Genetic Cost of Neanderthal Introgression. Genetics. 2016;203:
1089 881–891. doi:10.1534/genetics.116.186890

1090 30. Presgraves DC. Sex chromosomes and speciation in Drosophila. Trends in Genetics.
1091 2008;24: 336–343. doi:10.1016/j.tig.2008.04.007

1092 31. Maheshwari S, Barbash DA. The Genetics of Hybrid Incompatibilities. Annual Review of
1093 Genetics. 2011;45: 331–355. doi:10.1146/annurev-genet-110410-132514

1094 32. Qvarnström A, Bailey RI. Speciation through evolution of sex-linked genes. Heredity.
1095 2009;102: 4–15. doi:10.1038/hdy.2008.93

1096 33. Trier CN, Hermansen JS, Sætre G-P, Bailey RI. Evidence for Mito-Nuclear and Sex-Linked
1097 Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. PLoS
1098 Genet. 2014;10. doi:10.1371/journal.pgen.1004075

1099 34. Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky–Muller
1100 incompatibilities? Molecular Ecology. 2012;21: 4942–4957. doi:10.1111/mec.12006

1101 35. Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, et al. Natural selection
1102 interacts with recombination to shape the evolution of hybrid genomes. Science.
1103 2018;360: 656. doi:10.1126/science.aar3684

1104 36. Martin SH, Davey JW, Salazar C, Jiggins CD. Recombination rate variation shapes barriers
1105 to introgression across butterfly genomes. PLOS Biology. 2019;17: e2006288.
1106 doi:10.1371/journal.pbio.2006288

1107 37. Nachman MW, Payseur BA. Recombination rate variation and speciation: theoretical
1108 predictions and empirical results from rabbits and mice. Philos Trans R Soc Lond B Biol Sci.
1109 2012;367: 409–421. doi:10.1098/rstb.2011.0249

1110 38. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic
1111 landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507: 354–357.
1112 doi:10.1038/nature12961

1113 39. Veller C, Edelman NB, Muralidhar P, Nowak MA. Recombination, variance in genetic
1114 relatedness, and selection against introgressed DNA. bioRxiv. 2019; 846147.
1115 doi:10.1101/846147

1116 40. Calfee E, Gates D, Lorant A, Perkins MT, Coop G, Ross-Ibarra J. Selective sorting of
1117 ancestral introgression in maize and teosinte along an elevational cline. bioRxiv. 2021;
1118 2021.03.05.434040. doi:10.1101/2021.03.05.434040

1119 41. Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, et al. Repeated losses
1120 of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates.
1121 In: eLife [Internet]. 6 Jun 2017 [cited 23 Jul 2019]. doi:10.7554/eLife.24133

1122 42. Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, et al. Stable
1123 recombination hotspots in birds. Science. 2015;350: 928–932.
1124 doi:10.1126/science.aad0843

1125 43. Wang RJ, Ané C, Payseur BA. The Evolution of Hybrid Incompatibilities Along a Phylogeny.
1126 Evolution. 2013;67: 2905–2922. doi:https://doi.org/10.1111/evo.12173

1127 44. Yang C-H, Scarpino SV. Reproductive barriers as a byproduct of gene network evolution.
1128 bioRxiv. 2020; 2020.06.12.147322. doi:10.1101/2020.06.12.147322

1129 45. Powell DL, Moran B, Kim B, Banerjee SM, Aguillon SM, Fascinetto-Zago P, et al. Two new
1130 hybrid zones expand the swordtail hybridization model system. Evolution *(in press);*

1131 46. Fisher HS, Wong BBM, Rosenthal GG. Alteration of the chemical environment disrupts
1132 communication in a freshwater fish. Proc R Soc London Ser B. 2006;273: 1187–1193.
1133 doi:10.1098/rspb.2005.3406

1134 47. Schumer M, Powell DL, Delclós PJ, Squire M, Cui R, Andolfatto P, et al. Assortative mating
1135 and persistent reproductive isolation in hybrids. Proc Natl Acad Sci USA. 2017;114: 10936.
1136 doi:10.1073/pnas.1711238114

1137 48. Culumber ZW, Fisher HS, Tobler M, Mateos M, Barber PH, Sorenson MD, et al. Replicated
1138 hybrid zones of Xiphophorus swordtails along an elevational gradient. Molecular Ecology.
1139 2011;20: 342–356. doi:10.1111/j.1365-294X.2010.04949.x

1140 49. Schumer M, Brandvain Y. Determining epistatic selection in admixed populations.
1141 Molecular Ecology. 2016;25: 2577–2591. doi:10.1111/mec.13641

1142 50. Matute DR, Comeault AA, Earley E, Serrato-Capuchina A, Peede D, Monroy-Eklund A, et al.
1143 Rapid and Predictable Evolution of Admixed Populations Between Two Drosophila Species
1144 Pairs. Genetics. 2019 [cited 20 Apr 2020]. doi:10.1534/genetics.119.302685

1145 51. Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, et al. The genome of the
1146 platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and
1147 several complex traits. Nature Genetics. 2013;45: 567.

1148 52. Noor MAF, Grams KL, Bertucci LA, Reiland J. Chromosomal inversions and the reproductive
1149 isolation of species. PNAS. 2001;98: 12084–12088. doi:10.1073/pnas.221274498

1150 53. Kirkpatrick M, Barton N. Chromosome Inversions, Local Adaptation and Speciation.
1151 Genetics. 2006;173: 419–434. doi:10.1534/genetics.105.047985

1152 54. Wellenreuther M, Bernatchez L. Eco-Evolutionary Genomics of Chromosomal Inversions.
1153 Trends in Ecology & Evolution. 2018;33: 427–440. doi:10.1016/j.tree.2018.04.002

1154 55. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, et al. Genomic
1155 architecture and introgression shape a butterfly radiation. Science. 2019;366: 594–599.
1156 doi:10.1126/science.aaw2090

1157 56. Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, et al. Massive
1158 haplotypes underlie ecotypic differentiation in sunflowers. Nature. 2020;584: 602–607.
1159 doi:10.1038/s41586-020-2467-6

1160 57. Powell DL, Payne C, Banerjee SM, Keegan M, Bashkirova E, Cui R, et al. The Genetic
1161 Architecture of Variation in the Sexually Selected Sword Ornament and Its Evolution in
1162 Hybrid Populations. Current Biology. 2021 [cited 28 Jan 2021].
1163 doi:10.1016/j.cub.2020.12.049

1164 58. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING
1165 database in 2017: quality-controlled protein–protein association networks, made broadly
1166 accessible. Nucleic Acids Res. 2017;45: D362–D368. doi:10.1093/nar/gkw937

1167 59. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 2007;24:
1168 1586–1591. doi:10.1093/molbev/msm088

1169 60. Blount ZD, Lenski RE, Losos JB. Contingency and determinism in evolution: Replaying life's
1170 tape. Science. 2018;362. doi:10.1126/science.aam5979

1171 61. Kim BY, Huber CD, Lohmueller K. Deleterious variation mimics signatures of genomic
1172 incompatibility and adaptive introgression. 2017. doi:10.1101/221705 %J bioRxiv

1173 62. Lu Y, Sandoval A, Voss S, Lai Z, Kneitz S, Boswell W, et al. Oncogenic allelic interaction in
1174 Xiphophorus highlights hybrid incompatibility. PNAS. 2020;117: 29786–29794.
1175 doi:10.1073/pnas.2010133117

1176 63. Alcázar R, von Reth M, Bautor J, Chae E, Weigel D, Koornneef M, et al. Analysis of a plant
1177 complex resistance gene locus underlying immune-related hybrid incompatibility and its
1178 occurrence in nature. PLoS Genet. 2014;10: e1004848. doi:10.1371/journal.pgen.1004848

1179 64. Schumer M, Powell DL, Corbett-Detig R. Versatile simulations of admixture and accurate
1180 local ancestry inference with mixnmatch and ancestryinfer. Mol Ecol Resour. 2020;20:
1181 1141–1151. doi:10.1111/1755-0998.13175

1182 65. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, et al. The landscape of
1183 recombination in African Americans. Nature. 2011;476: 170–175.
1184 doi:10.1038/nature10336

1185 66. Powell DL, Payne C, Keegan M, Banerjee SM, Cui R, Andolfatto P, et al. The genetic
1186 architecture of the sexually selected sword ornament and its evolution in hybrid
1187 populations. bioRxiv. 2020; 2020.07.23.218164. doi:10.1101/2020.07.23.218164

1188 67. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
1189 Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

1190 68. Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing and
1191 analyzing multiple genome alignments. Bioinformatics. 2013;29: 1341–1342.
1192 doi:10.1093/bioinformatics/btt128

1193 69. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and
1194 versatile genome alignment system. PLOS Computational Biology. 2018;14: e1005944.
1195 doi:10.1371/journal.pcbi.1005944

1196