EVOLUTIONARY GENOMICS

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

Molly Schumer,[1,2,3,4]* Chenling Xu,[5] Daniel L. Powell,[4,6] Arun Durvasula,[7] Laurits Skov,[8] Chris Holland,[4,6] John C. Blazier,[6,9] Sriram Sankararaman,[7,10] Peter Andolfatto,[11]† Gil G. Rosenthal,[4,6]† Molly Przeworski[3,12]*

To investigate the consequences of hybridization between species, we studied three replicate hybrid populations that formed naturally between two swordtail fish species, estimating their fine-scale genetic map and inferring ancestry along the genomes of 690 individuals. In all three populations, ancestry from the "minor" parental species is more common in regions of high recombination and where there is linkage to fewer putative targets of selection. The same patterns are apparent in a reanalysis of human and archaic admixture. These results support models in which ancestry from the minor parental species is more likely to persist when rapidly uncoupled from alleles that are deleterious in hybrids. Our analyses further indicate that selection on swordtail hybrids stems predominantly from deleterious combinations of epistatically interacting alleles.

Understanding speciation is central to understanding evolution, but so much about the process still puzzles us. Foundational work in evolutionary biology envisioned speciation as an ordered process in which reproductive barriers, once established, prevent gene flow between species (*1*). We now realize, however, that speciation is much more dynamic, with evidence of historical and ongoing hybridization visible in the genomes of myriad species (*2–5*). The ubiquity of hybridization raises the question of how species that interbreed remain distinct.

At least part of the answer lies in widespread selection on hybrid genomes (*1*). Analyses of hominin and swordtail fish hybrids indicate that ancestry from the "minor" parent species (the parent that contributed less to the gene pool of hybrids) is decreased near functionally important elements (*4, 6, 7*), presumably because such regions are enriched for harmful alleles. Aside from these observations, however, little

is known about how hybrid genomes evolve. Decades of experimental work have demonstrated that Bateson-Dobzhansky-Muller incompatibilities (BDMIs) are a central mechanism underlying reproductive isolation once species are formed (*8–10*), but the importance of BDMIs in the evolution of hybrid genomes remains unknown, as does the role of other modes of selection. When there is introgression from a species with a smaller effective population size, hybrids may suffer from increased genetic load ("hybridization load") due to the introduction of weakly deleterious alleles (*6, 11, 12*). Depending on the environment in which hybrids find themselves, alleles that underlie ecological adaptations in the parental species may also be deleterious (*13, 14*). Complicating matters further, the sources of selection on hybrids will likely vary from system to system, depending on the extent of genetic and ecological differentiation between the parental species as well as the differences in their effective population sizes.

Regardless of the source of selection, however, one feature is expected to play a central role in mediating its effects: variation in recombination rates along the genome (*6, 11, 15–17*). In models of BDMIs, neutral ancestry from the minor parent is more likely to persist in regions of higher recombination, where it is more rapidly uncoupled from mutations deleterious in the prevalent (major parent) genetic background (Fig. 1, A and B) (*17*). Similarly, in models of hybridization load, all else being equal, shorter linkage blocks tend to carry fewer weakly deleterious mutations and thus be less rapidly purged by selection (fig. S1) (*6, 11*). Previous studies have reported patterns consistent with these expectations (*18–20*) but without investigating ancestry patterns and their relationship to local recombination rates, distinguishing among selection in

the parental or hybrid lineages, or determining the major sources of selection in hybrid populations.

To address these issues, we took advantage of naturally occurring hybrid populations between sister species of swordtail fish, *Xiphophorus birchmanni* and *X. malinche* (Fig. 1, C to E) (*21*). The species are ~0.5% divergent at the nucleotide level, and because of the small effective population size of *X. malinche*, incomplete lineage sorting between the two is relatively rare (Fig. 1F) (*21*). We focused on three hybrid populations that formed independently between the two species fewer than 100 generations ago (*22*). Previous analyses of hybrids between these species suggested that there are ~100 unlinked BDMI pairs segregating, with estimated selection coefficients of ~0.02 to 0.05, in addition to which there could also be linked BDMIs (*22, 23*).

To infer local ancestry patterns, we generated ~1× coverage whole-genome data for 690 hybrids sampled from the three hybrid populations and then estimated local ancestry patterns by applying a hidden Markov model to ~1 million sites genome-wide (*21, 24*). Two of the hybrid populations derive on average 75 to 80% of their genomes from *X. birchmanni*, whereas individuals in the third population derive on average 72% of their genomes from *X. malinche* (Fig. 1E) (*21*), with median homozygous tract lengths for the minor parent ranging from 84 to 225 kb across the three populations (*21*).

Our previous work (*25*) indicated that local recombination rates should be conserved between *X. birchmanni* and *X. malinche* (*21*). To consider the relationship between local ancestry and recombination rate, we inferred a fine-scale genetic map for *X. birchmanni* from patterns of linkage disequilibrium (table S1) (*21*). We also generated a crossover map from ancestry switch points in hybrids, which was concordant with the one obtained for *X. birchmanni* (fig. S2) (*21*).

In all three hybrid populations, the probability of carrying ancestry from the minor parent increases with the local recombination rate (Fig. 2 and table S2). The relationship remains, irrespective of the choice of scale (fig. S3) and after thinning of the single-nucleotide polymorphism and ancestry data so as to control for possible differences in the reliability of estimated recombination rates or the power to call ancestry across windows (*21*). This pattern is not expected under neutrality (fig. S1) but can readily be generated under several models of selection, including selection against BDMIs or hybridization load or widespread ecological selection against loci from the minor parent (Fig. 1B and fig. S1). Thus, our finding supports models in which minor parent ancestry persists where it was more likely to have been rapidly uncoupled from the deleterious alleles with which it was originally linked (*21*).

In principle, the chance of minor parent ancestry persisting should be a function of the exact number of deleterious alleles to which it was linked since hybridization occurred. Local recombination rates are one proxy for this (unknown) parameter, as is the number of coding or conserved base

[1]Howard Hughes Medical Institute (HHMI), Boston, MA, USA. [2]Harvard Society of Fellows, Harvard University, Cambridge, MA, USA. [3]Department of Biological Sciences, Columbia University, New York, NY, USA. [4]Centro de Investigaciones Científicas de las Huastecas "Aguazarca," Calnali, Hidalgo, Mexico. [5]Center for Computational Biology, University of California at Berkeley, Berkeley, CA, USA. [6]Department of Biology, Texas A&M University, College Station, TX, USA. [7]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [8]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [9]Texas A&M Institute for Genome Sciences and Society, College Station, TX, USA. [10]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA. [11]Department of Ecology and Evolutionary Biology and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [12]Department of Systems Biology, Columbia University, New York, NY, USA.
*Corresponding author. Email: schumer@fas.harvard.edu (M.S.); mp3284@columbia.edu (M.P.) †These authors contributed equally to this work.

pairs nearby. Both features predict average minor parent ancestry (figs. S4 and S5) (*21*), but in our data, recombination is the stronger predictor and remains similarly strong after controlling for the number of coding (or conserved) base pairs (tables S2 and S3).

To investigate the mode of selection on hybrids, we considered correlations in local ancestry between pairs of hybrid populations: Though weaker between populations with different major parent ancestries, the correlation was in all cases significantly positive (with the recombination rate controlled for) (Fig. 3, A and B). These corre-

lation patterns should not arise from ecological selection but are expected from selection against hybridization load as well as, less intuitively, from selection on the same BDMIs (fig. S6) (*21*).

Further evidence about the mode of selection comes from an analysis of genome sequences from *X. malinche* (*3*, *22*) and *X. birchmanni*, which indicates that *X. malinche* has had a smaller long-term effective population size over the last ~20,000 generations (Fig. 1F) (*21*), as reflected in its four times lower heterozygosity (0.03% versus 0.12% per base pair). Accordingly, the *X. malinche* genome carries significantly

more derived putatively deleterious alleles (a 2.5% excess) than that of *X. birchmanni* (*21*, *26*). As a result of this difference, the three hybrid populations of swordtail fish provide an informative contrast: Whereas BDMIs should lead to selection against minor parent ancestry in all three populations, hybridization load should favor the major parent in the first two populations (Totonicapa and Aguazarca) and the minor parent in the third (Tlatemaco) (Figs. 1E and 3C). That minor parent ancestry also increases with recombination in the Tlatemaco population (Fig. 2) then indicates that hybrid
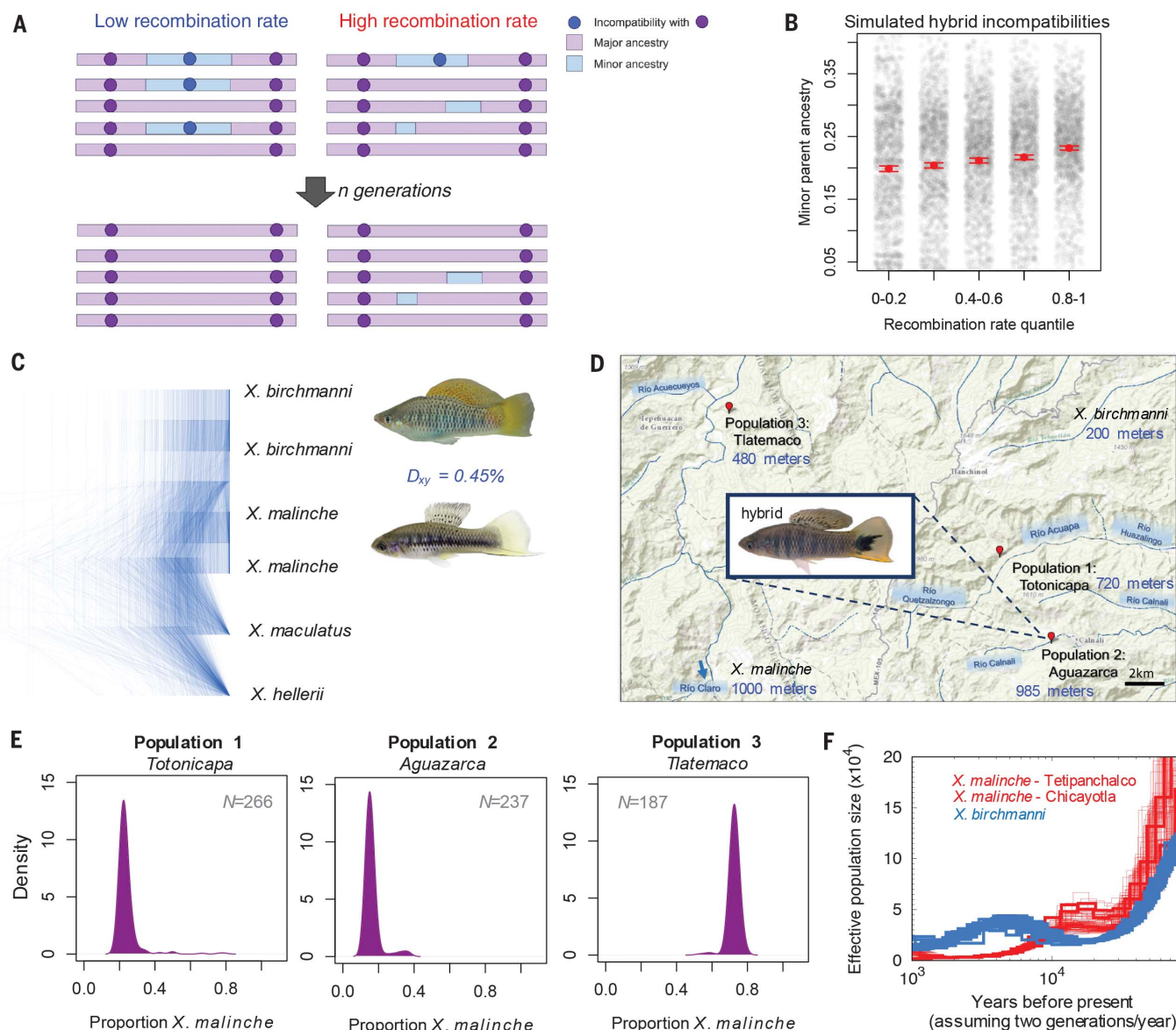
Fig. 1. Predicted relationships between minor parent ancestry and recombination rates and properties of focal swordtail populations.
(**A**) In the presence of hybrid incompatibilities, minor parent ancestry is more likely to persist in regions of high recombination. (**B**) One randomly chosen replicate of simulations under plausible parameters for swordtail species (*21*). Red points indicate the means, and whiskers indicate 2 SEM; gray points are raw data. (**C**) Maximum likelihood trees from RAxML for 1000 alignments of randomly selected 10-kb regions in swordtail species.

D_{xy} refers to the average nucleotide divergence between *X. birchmanni* and *X. malinche*. (**D**) Locations of hybrid populations in river systems in Hidalgo, Mexico. Elevations of the hybrid populations and typical elevations for parental populations are listed in blue. (**E**) Inferred ancestry proportions for individuals (*n*) sampled from each population. (**F**) Effective population sizes inferred from three *X. malinche* genomes (sampled from two populations) and 20 *X. birchmanni* genomes. Fifty bootstraps are shown for one individual from each *X. malinche* population (*21*).

incompatibilities are the dominant source of selection, rather than hybridization load (Fig. 3C and fig. S7) (*21*). In principle, ecological selection favoring the major parent could also produce a positive correlation between recombination and minor parent ancestry (but not positive correla-
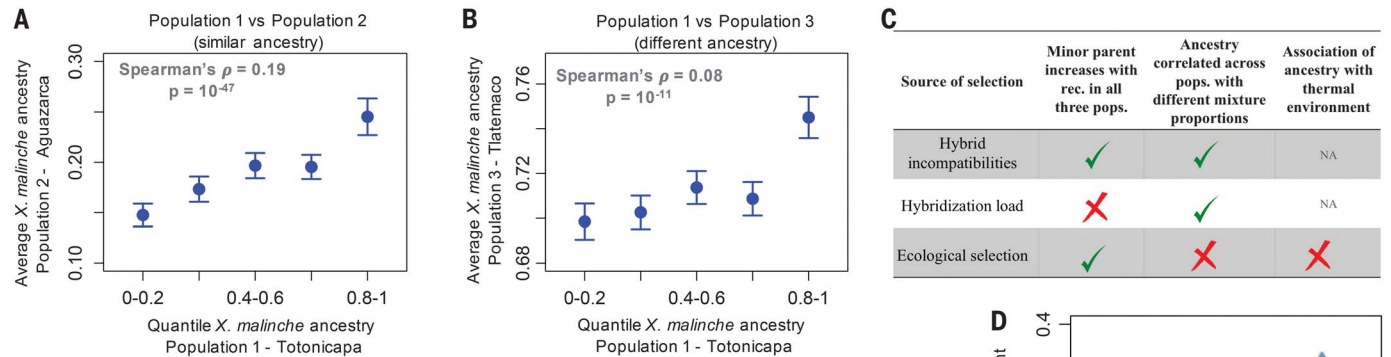
tions in ancestry between populations) (Fig. 3, A and B). However, this explanation would require two of the hybrid populations to occur in more *X. birchmanni*–like environments and one in a more *X. malinche*–like environment, when available evidence suggests otherwise (Fig. 1D) (*21*).

Furthermore, in all populations, minor parent ancestry is unusually low near previously mapped putative BDMIs (*22, 23*). Lower minor parent ancestry does not result from the approach used to identify BDMIs (*21, 22*) but is expected from selection on epistatically interacting alleles (Fig. 3D)
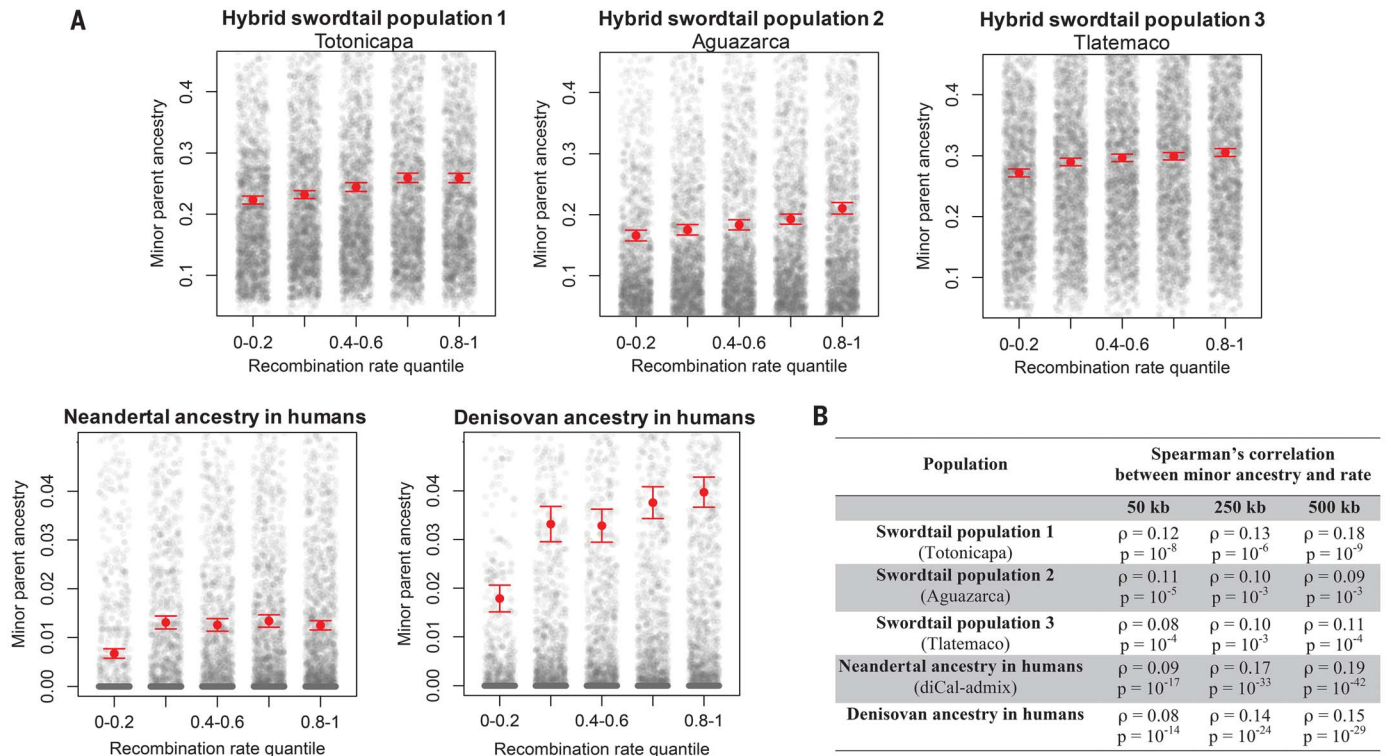

Fig. 2. Relationships between minor parent ancestry and recombination rate in swordtails and hominins. (**A**) Relationship between minor parent ancestry and recombination rate in swordtails and in humans, summarized in 50-kb windows for swordtail analyses and 250-kb windows for humans (fig. S8), so that the numbers of windows are similar. (**B**) Spearman's correlations between average

minor parent ancestry and recombination rate at several scales [complete results are provided in table S2, and details of the Denisovan analysis are provided in (*21*)]. In (A), red points and whiskers indicate the means with 2 SEM determined by bootstrapping; gray points show raw data. Quantile binning is for visualization; statistical tests were performed on the unbinned data.

Fig. 3. Evidence for BDMIs being the major source of selection on hybrids. (**A** and **B**) Correlations in ancestry between independently formed swordtail hybrid populations (in 0.1-centimorgan windows) (fig. S9). Points show the means, and whiskers indicate 2 SEM; correlations were calculated on unbinned data. (**C**) Predictions for different sources of selection on hybrids. rec., recombination; pops., populations; NA, not applicable. (**D**) The average proportion of minor parent ancestry is unusually depleted in 50-kb windows containing putative unlinked BDMIs [red points, from (*23*)] compared with that in 1000 null data sets (blue) (*21*). Lower average minor parent ancestry at putative BDMIs is not expected as a result of the way the BDMIs were originally identified (*21*).
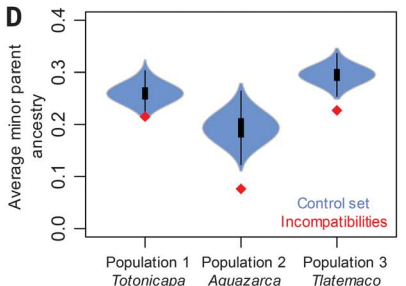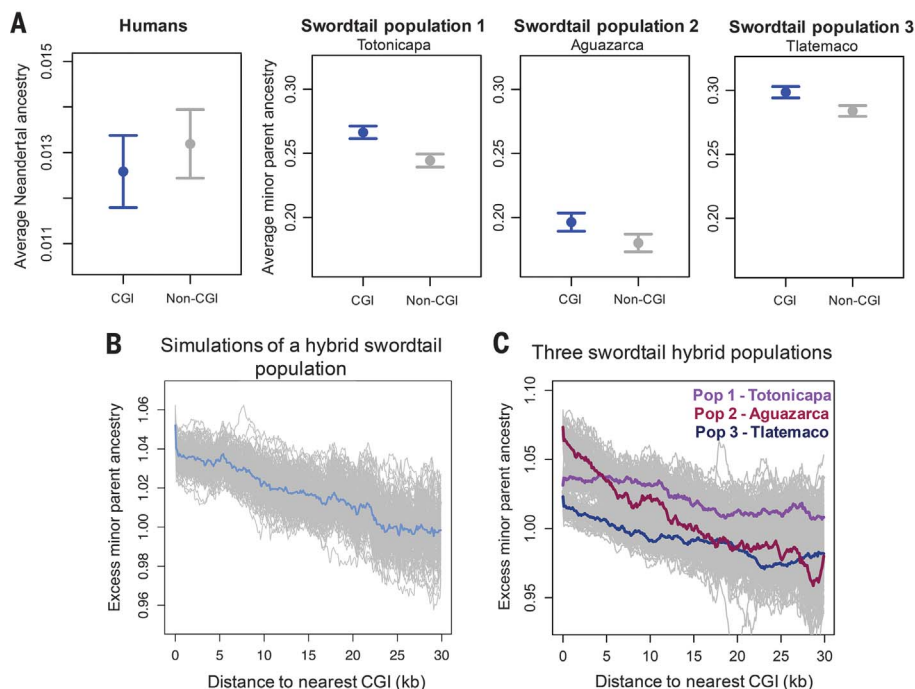
(21). Together, these lines of evidence indicate that BDMIs are the predominant—though not necessarily sole—source of selection filtering minor parent ancestry in these three swordtail hybrid populations (Fig. 3C).

To explore the generality of these relationships, we considered admixture between humans and archaic hominins. Several studies have reported that Neandertal ancestry tends to decrease with the number of linked coding base pairs and with a measure of purifying selection at linked sites (4, 6, 11), patterns for which both BDMIs and hybridization load—in part because of the smaller effective population size of Neandertals (27)—have been proposed as explanations (4, 6, 11). Reanalyzing the data, we found that the proportion of Neandertal ancestry decreases in regions of the human genome with lower recombination rates (Fig. 2 and tables S2 and S4). This relationship is seen by using three different approaches to infer Neandertal ancestry (table S2) and is not explained by variation in power to identify introgression or the number of coding base pairs nearby (table S2) (21). Repeating these analyses for Denisovan ancestry, we obtained the same pattern (Fig. 2 and table S2) (21).

As with the persistence of minor parent ancestry in swordtails, the persistence of archaic hominin ancestry in regions of higher recombination is not expected under neutrality (fig. S1). However, our conclusion about the source of selection reached for swordtails need not hold for hominins—a priori, because modern humans were less diverged from Neandertals and Denisovans when they interbred (28), and because plausible models of hybridization load have been shown to provide a good fit to the distribution of Neandertal ancestry in the human genome (6).

Thus, minor parent ancestry is predicted by the local recombination rate across three replicate admixture events in swordtails, as well as in two cases of admixture in hominins. Together with earlier indications in other species (18–20), our findings show the distribution of minor parent ancestry to be at least in part predictable from genomic features. Knowledge of local recombination should therefore provide a guide to where in the genome minor parent ancestry is expected to be highest. In hominins, meiotic recombination is directed to the genome by binding of the *PRDM9* gene; in swordtails, meiotic recombination is not and tends to occur near promoter-like features (21, 25). Accordingly, minor parent ancestry is higher around promoter-like features in swordtails but not in humans (Fig. 4) (21). Thus, the mechanism by which recombination is directed to the genome affects the distribution of minor parent ancestry.

One implication is that the reliance on PRDM9 to direct recombination may affect reproductive isolation between species not only directly [as in mice (29)] but also indirectly. For example, if epistatic interactions often occur between regulatory and coding regions, hybrids with recombination concentrated in promoter-like regions may experience greater negative selection due to BDMIs but more opportunities for adaptive introgression. As genomic data accumulate for hybridizing species across the tree of life, the consequences of recombination mechanisms for the fate of hybrids can be evaluated systematically.

REFERENCES AND NOTES

1. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Associates, 2004).
2. L. H. Rieseberg, J. Whitton, K. Gardner, *Genetics* **152**, 713–727 (1999).
3. R. Cui et al., *Evolution* **67**, 2166–2179 (2013).
4. S. Sankararaman et al., *Nature* **507**, 354–357 (2014).
5. F. Jacobsen, K. E. Omland, *Mol. Ecol.* **20**, 2236–2239 (2011).
6. I. Juric, S. Aeschbacher, G. Coop, *PLOS Genet.* **12**, e1006340 (2016).
7. M. Schumer, R. Cui, D. L. Powell, G. G. Rosenthal, P. Andolfatto, *Mol. Ecol.* **25**, 2661–2679 (2016).
8. J. P. Masly, D. C. Presgraves, *PLOS Biol.* **5**, e243 (2007).
9. K. Bomblies et al., *PLOS Biol.* **5**, e236 (2007).
10. H.-Y. Lee et al., *Cell* **135**, 1065–1073 (2008).
11. K. Harris, R. Nielsen, *Genetics* **203**, 881–891 (2016).
12. N. Bierne, T. Lenormand, F. Bonhomme, P. David, *Genet. Res.* **80**, 197–204 (2002).
13. L. H. Rieseberg et al., *Science* **301**, 1211–1216 (2003).
14. M. E. Arnegard et al., *Nature* **511**, 307–311 (2014).
15. C. I. Wu, *J. Evol. Biol.* **14**, 851–865 (2001).
16. M. A. F. Noor, K. L. Grams, L. A. Bertucci, J. Reiland, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12084–12088 (2001).
17. M. W. Nachman, B. A. Payseur, *Philos. Trans. R. Soc. London Ser. B* **367**, 409–421 (2012).
18. Y. Brandvain, A. M. Kenney, L. Flagel, G. Coop, A. L. Sweigart, *PLOS Genet.* **10**, e1004410 (2014).
19. M. Carneiro, N. Ferrand, M. W. Nachman, *Genetics* **181**, 593–606 (2009).
20. A. Geraldes, P. Basset, K. L. Smith, M. W. Nachman, *Mol. Ecol.* **20**, 4722–4736 (2011).
21. Materials and methods are available as supplementary materials.
22. M. Schumer et al., *eLife* **3**, e02535 (2014).
23. M. Schumer, Y. Brandvain, *Mol. Ecol.* **25**, 2577–2591 (2016).
24. P. Andolfatto et al., *Genome Res.* **21**, 610–617 (2011).
25. Z. Baker et al., *eLife* **6**, e24133 (2017).
26. R. Do et al., *Nat. Genet.* **47**, 126–131 (2015).
27. K. Prüfer et al., *Nature* **505**, 43–49 (2014).
28. H. A. Orr, M. Turelli, *Evolution* **55**, 1085–1094 (2001).
29. B. Davies et al., *Nature* **530**, 171–176 (2016).

Fig. 4. The recombination mechanism shapes the distribution of minor parent ancestry.
(**A**) Neandertal ancestry is not elevated in 50-kb windows that overlap with CpG islands (CGIs) compared with windows that do not but have similar GC content. The fold difference λ is 0.95 ($P = 0.91$) (21). The same analysis for swordtail hybrids reveals that the proportion of minor parent ancestry is higher in windows that overlap CGIs (population 1, λ = 1.09, $P < 0.005$; population 2, λ = 1.09, $P < 0.005$; population 3, λ = 1.02, $P < 0.005$). Points show the means, and whiskers indicate 2 SEM obtained from 1000 joint bootstraps. (**B**) Simulations of incompatibility selection in swordtails predict an enrichment of minor parent ancestry near CGIs. (**C**) This prediction is met for all hybrid populations. In (B) and (C), gray lines show results of 500 replicate simulations bootstrapping 5-kb windows; colored lines indicate the means for all replicates in sliding 5-kb windows.

Science

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

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

Selection, recombination, and hybrid evolution

Hybridization is an important force in evolution. The effects of hybridization across the whole genome are not understood. Using a fine-scale genetic map, Schumer *et al.* examined local ancestry in replicate natural hybrid populations of swordtail fish. Each parental species contributes different proportions of genetic material to the genomes of their descendants. Genes from the "minor" (less well-represented) parent occur in regions of the genome that are subject to higher recombination rates and where there are fewer potentially deleterious genes. Neanderthal ancestry in human genomes shows similar patterns.

Science, this issue p. 656

ARTICLE TOOLS	http://science.sciencemag.org/content/360/6389/656
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/04/18/science.aar3684.DC1
REFERENCES	This article cites 87 articles, 26 of which you can access for free http://science.sciencemag.org/content/360/6389/656#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service