**Title:**

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

**One sentence summary:** The persistence of hybrid ancestry is predictable from local recombination rates, in three replicate hybrid populations as well as in humans.

**Authors:**

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

[1] Hanna H. Gray Fellow, Howard Hughes Medical Institute
[2] Harvard Society of Fellows, Harvard University
[3] Department of Biological Sciences, Columbia University
[4] Centro de Investigaciones Científicas de las Huastecas "Aguazarca"
[5] Center for Computational Biology, University of California at Berkeley
[6] Department of Biology, Texas A&M University
[7] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles
[8] Bioinformatics Research Centre, Aarhus University, Aarhus C., Denmark
[9] Department of Computer Science, University of California, Los Angeles
[10] Department of Ecology and Evolutionary Biology and Lewis-Sigler Institute for Integrative Genomics, Princeton University
[11] Department of Systems Biology, Columbia University

* Correspondence to: schumer@fas.harvard.edu or mp3284@columbia.edu
§ Indicates equal contribution

30 **Abstract**

31

32 While hybridization between species is increasingly appreciated to be a common occurrence,

33 little is known about the forces that govern the subsequent evolution of hybrid genomes. We

34 considered this question in three independent, naturally-occurring hybrid populations formed

35 between swordtail fish species *Xiphophorus birchmanni* and *X. malinche.* To this end, we built a

36 fine-scale genetic map and inferred patterns of local ancestry along the genomes of 690

37 individuals sampled from the three populations. In all three cases, we found hybrid ancestry to be

38 more common in regions of high recombination and where there is linkage to fewer putative

39 targets of selection. These same patterns are also apparent in a reanalysis of human-Neanderthal

40 admixture. Our results lend support to models in which ancestry from the "minor" parental

41 species persists only where it is rapidly uncoupled from alleles that are deleterious in hybrids,

42 and show the retention of hybrid ancestry to be at least in part predictable from genomic features.

43 Our analyses further indicate that in swordtail fish, the dominant source of selection on hybrids

44 stems from deleterious combinations of epistatically-interacting alleles.

45    **Main text**

46

47    Understanding speciation is central to understanding evolution, but so much about the

48    process still puzzles us. The foundational work in evolutionary biology envisioned speciation as

49    an ordered process during which reproductive barriers, once established, prevent gene flow

50    between species (*1*). We now realize, however, that speciation is much more dynamic, with

51    hybridization occurring both during and after the evolution of reproductive barriers and evidence

52    of past hybridization with close relatives still visible in the genomes of myriad animal and plant

53    species (*2-9*). The ubiquity of hybridization raises the question of how species that hybridize

54    remain genetically and ecologically differentiated.

55    At least part of the answer is likely that selection filters out deleterious hybrid ancestry

56    from the genome (*1*). For instance, in hominins and swordtail fish, individuals are less likely to

57    carry hybrid ancestry near functionally important elements (*6, 10, 11*), presumably because it is

58    especially deleterious in such regions. Aside from these observations, however, little is known

59    about how admixed genomes evolve—or, in some cases, stabilize—following hybridization. Our

60    understanding of the evolution of hybrid genomes is complicated by the existence of many

61    possible modes of selection and by the fact that, in most systems, the location of sites under

62    selection is unknown. Decades of experimental work have demonstrated that Dobzhansky-

63    Muller incompatibilities (DMIs) are a central mechanism underlying reproductive isolation once

64    species are formed (*12-16*), but the importance of DMIs in shaping the evolution of hybrid

65    genomes remains unknown, as does the role of other modes of selection. Notably, it was recently

66    pointed out that when there is introgression from a species with a lower effective population size,

67    hybrids may suffer from increased genetic load ("hybridization load") due to the introduction of

68    weakly deleterious alleles (*10, 17*). Depending on the environment in which hybrids find

69    themselves, alleles that underlie ecological adaptations in the parental species may also be

70    deleterious in hybrids (*18, 19*). Complicating matters yet further, modes of selection on hybrid

71    ancestry will likely vary from system to system, depending on the extent of genetic divergence

72    and ecological differentiation between the parental species, as well as long-term differences in

73    their effective population sizes.

74    One feature, however, is expected to play a central role in all these models: variation in

75    recombination rates along the genome (*10, 17, 20-22*). Theory predicts that selection is more

3

76    likely to weed out hybrid ancestry in regions of low recombination (*23-25*). Specifically, in

77    models of DMIs, minor parent ancestry will persist preferentially in regions of higher

78    recombination because it is more rapidly uncoupled from mutations that are incompatible with

79    the prevalent (i.e., major parent) genetic background (Fig. 1). Similarly, in models of

80    hybridization load, all else being equal, shorter linkage blocks will carry fewer weakly

81    deleterious mutations and therefore be will be less rapidly purged by selection (*10, 17*; Fig. S1).

82    Previous studies have reported patterns potentially consistent with these expectations (*26, 27*),

83    but without directly investigating ancestry patterns and their relationship with local

84    recombination rates (*28*).

85

86    *Recombination shapes ancestry in swordtail fish.* To test predictions about the role of

87    recombination in filtering hybrid ancestry, we took advantage of a set of naturally occurring

88    hybrid populations between two swordtail fish species, *Xiphophorus birchmanni* and a closely

89    related species, *X. malinche* (Fig. 2; Supporting Information 1-3). The two species are ~0.5%

90    divergent at the nucleotide level and incomplete lineage sorting between the two is relatively rare

91    (*29*; Fig. 2A). We focused on three hybrid populations that formed independently between the

92    two fewer than 100 generations ago (*29*), likely as a result of human-mediated habitat

93    disturbance (*30*). Previous analyses of hybrid zones between these two species, including two of

94    the three populations analyzed here, suggested that there are on the order of 100 pairs of

95    unlinked DMIs segregating in hybrids (*29, 31*), with estimated selection coefficients ~0.03-0.05

96    (*29*), and potentially many more linked DMIs, indicating that swordtail hybrids may be

97    experiencing widespread selection on DMIs.

98    To infer local ancestry patterns, we generated ~1X low coverage whole genome data for

99    690 hybrids sampled from the three hybrid populations (Supporting Information 1). We

100    estimated local ancestry patterns for the 690 hybrids by applying a hidden Markov model (*32*);

101    this approach is predicted to have high accuracy for these hybrid populations, given the marker

102    density and time since mixture (*29, 32, 33*). Using ancestry calls at 1-1.2 million sites genome

103    wide, we inferred that two of the hybrid populations derive on average 75-80% of their genomes

104    from *X. birchmanni*, whereas individuals in the third population derive on average 72% of their

105    genomes from *X. malinche* (Fig. 2; Supporting Information 1; *34*). The median homozygous tract

106    length for the minor parent ranges from 84 kb to 225 kb across the three populations, roughly

4

107    matching expectations for hybrid populations of these ages and mixture proportions (Supporting

108    Information 4).

109         To consider the relationship between local ancestry and recombination rates, we inferred

110    a fine-scale genetic map for *X. birchmanni* from patterns of linkage disequilibrium (LD) in

111    unrelated individuals (Table S1; Supporting Information 2, 4-5). Based on our previous work on

112    recombination in this taxon (*35*), we had a strong prior expectation that local recombination rates

113    should be conserved between *X. birchmanni* and *X. malinche* (Supporting Information 6). We

114    also generated crossover maps from hybrids based on inferred switch points between the two

115    ancestries. Overall the hybrid and parental maps are consistent (Fig. S2), with the correlations

116    between maps roughly comparable to what would be expected if the maps were in fact identical

117    (Supporting Information 7).

118         In all three hybrid populations of swordtail fish, the probability of carrying ancestry from

119    the minor parent increases with the local recombination rate (Fig. 3, Table 1). This association

120    remains irrespective of the choice of scale (Fig. S3) and after thinning the SNP and ancestry

121    variation data to control for possible differences in the ability to reliably infer recombination

122    rates or the power to call hybrid ancestry across windows (Supporting Information 4). The

123    preferential persistence of minor parent ancestry in regions of higher recombination is not

124    expected under neutrality (Fig. S1) and instead indicates that minor parent ancestry was retained

125    where it was more likely to have been rapidly uncoupled from the deleterious alleles with which

126    it was originally linked (Supporting Information 5). This qualitative pattern can be generated

127    under several models of selection, including selection against DMIs, selection against weakly

128    deleterious alleles introduced by hybridization, or widespread ecological selection against loci

129    that derive from the minor parent (Fig. 1, Fig. S1).

130         In principle, the retention of hybrid ancestry should be most accurately predicted from the

131    exact number of deleterious alleles to which a minor parent segment was linked since

132    hybridization occurred. Local recombination rates are one proxy for this (unknown) parameter,

133    as are the number of coding base pairs nearby. In these data, both factors predict minor parent

134    ancestry (Fig. S4; Fig. S5; Supporting Information 4), but local recombination is a stronger

135    predictor and remains a predictor after controlling for the number of coding base pairs (Table 1;

136    Table S2). These findings are consistent with those obtained in simulations mimicking the data

137    structure (Supporting Information 4-5), presumably because the number of coding base pairs

138    nearby is an extremely noisy proxy.

139

140    *The source of selection*. Controlling for the recombination rate, local ancestry is positively

141    correlated between all pairs of hybrid populations, with weaker but significant correlations seen

142    even between populations with different major parent ancestries (Fig. 4). These findings are

143    expected from selection on the same underlying loci in independently formed populations

144    (Supporting Information 8). Whereas both DMIs and hybridization load are predicted to drive

145    positive correlations in local ancestry across populations regardless of the admixture proportions

146    (Fig. S6), ecological selection against minor parent ancestry should lead to negative correlations

147    in local ancestry and is thus inconsistent with the observed patterns (Fig. 4; Supporting

148    Information 8).

149         Comparison among the three hybrid populations also provides a means to distinguish

150    between the remaining two hypotheses. Analyzing genome sequences from *X. malinche* (*5, 29*)

151    and *X. birchmanni,* we found that *X. malinche* has had a lower long-term effective population

152    size than *X. birchmanni* (Fig. 2; Supporting Information 3), as seen both in the approximately

153    four-fold lower average heterozygosity in *X. malinche* (0.03% vs 0.12% per base pair,

154    respectively) and in estimates of effective population sizes over time from high coverage genome

155    sequences (Fig. 2, Table S1). Consistent with a lower long-term effective population size, *X.*

156    *malinche* carries significantly more putative deleterious alleles relative to the inferred ancestral

157    sequence than does *X. birchmanni*, as measured by the number of derived, non-synonymous

158    substitutions per haploid genome (a 2.5% excess, p=0.016 based on 1,000 bootstrap resamples;

159    see Supporting Information 3; *36, 37*). Because *X. birchmanni* and *X. malinche* source

160    populations differ in the number of putatively deleterious variants, the three hybrid populations

161    of swordtail fish provide an informative contrast: whereas DMIs should lead to selection against

162    minor parent ancestry in all three populations, hybridization load should favor the major parent

163    in populations 1 and 2 and the minor parent in population 3 (Fig. 2; Fig. 4).

164         In this regard, the fact that minor parent ancestry also increases with recombination in the

165    third hybrid population, which derives most of its genome from the parental species that has

166    lower effective population size (Fig. 2, 3), indicates that hybrid incompatibilities are the

167    dominant mode of selection shaping ancestry in the genome in these hybrid populations, rather

6

168    than selection against hybridization load (Fig. 4; Fig. S7; Supporting Information 4-5). In

169    principle, ecological selection favoring the major parent could also produce a positive correlation

170    between recombination rate and ancestry (though not the positive correlations in ancestry across

171    populations; Fig. 4). However, this scenario would require two of the hybrid populations to occur

172    in more *birchmanni*-like environments and one in a more *malinche*-like environment, when

173    available evidence suggests otherwise—notably, all of the hybrid populations are found in

174    thermal environments that are mismatched to the environment where their major parent is found

175    (Fig. 2; Supporting Information 5).

176         Moreover, in all three hybrid populations, minor parent ancestry is unusually low near

177    previously mapped DMIs between the two parental species (*29, 31*), a pattern that should not

178    arise from the approach used to identify DMIs (Supporting Information 5), but is expected from

179    selection on epistatically-interacting alleles (Fig. 4; Supporting Information 4-5). Taken together,

180    these lines of evidence indicate that DMIs are the main (though not necessarily sole) source of

181    selection shaping the retention of hybrid ancestry in these three swordtail fish hybrid populations

182    (Fig. 4).

183

184    *Ancestry also interacts with the local recombination rate in hominins*. To evaluate the generality

185    of the relationship between recombination rate and ancestry seen in swordtails, we considered

186    the only other case with similar genomic data available: admixture between humans and archaic

187    hominins. Several studies have reported that the average proportion of Neanderthal ancestry

188    decreases with the number of closely linked coding base pairs and with a measure of the strength

189    of purifying selection at linked sites (*6, 10, 17, 38*), patterns for which both DMIs and

190    hybridization load (due to the smaller effective population size of Neanderthals, *39*) have been

191    proposed as explanations (*6, 10, 17*). Reanalyzing the data, we found that the proportion of

192    Neanderthal ancestry (the minor parental species) decreases in regions of the human genome

193    with lower recombination rates (Fig. 3D; Table 1; Table S3). This relationship is seen for

194    different window size choices and with any of three approaches to infer Neanderthal ancestry in

195    the human genome (Table 1), and is not expected as a result of variation in the power to identify

196    introgression along the genome (Supporting Information 9). The effect of local recombination

197    rate on Neanderthal ancestry also persists after accounting for the number of coding base pairs

198    nearby (Table 1; Supporting Information 9). Interestingly, the relationship between Neanderthal

199    ancestry and local recombination rate is especially strong when excluding regions of unusually

200    high frequency Neanderthal ancestry (e.g. top 1%; Fig. S8), possibly because these regions are

201    enriched for cases of adaptive introgression (*6, 38, 40, 41*). Repeating these analyses for

202    Denisovan ancestry, for which there is lower power to identify ancestry tracts, there is a much

203    weaker but consistent trend (Table 1; Supporting Information 9).

204        As in swordtails, the persistence of Neanderthal ancestry in regions of higher

205    recombination is not expected under neutrality (Fig. S1) but could be generated by selection

206    against DMIs, weakly deleterious alleles introduced by Neanderthals (*6, 10, 17, 38*), or

207    widespread ecological selection against Neanderthal ancestry (Fig. 1, Fig. S1). Unlike in the case

208    of swordtails, however, these causes cannot be distinguished based on these data alone (*6, 10, 17,*

209    *38, 42*). Moreover, the conclusion about the source of selection reached for swordtail fish need

210    not hold for hominins, in particular because modern humans and Neanderthals (Denisovans)

211    were less diverged when they are thought to have interbred, and thus may have accumulated

212    many fewer DMIs (*43, 44*).

213

214    *The predictability of hybrid ancestry.* Hybrid ancestry is predicted by the local recombination

215    rate across three replicate admixture events between the same species pair in swordtail fish, as

216    well as in two cases of admixture in hominins. In swordtail fish hybrids, several lines of evidence

217    indicate that selection against hybrid incompatibilities is the dominant force shaping minor

218    parent ancestry in the genome. In hominins, the source of selection remains unclear. Regardless

219    of the precise mechanisms of selection on hybrids, the generality of these patterns reveals the

220    retention of hybrid ancestry to be at least in part predictable from genomic features.

221        The relationship of minor parent ancestry to local recombination thus provides a useful

222    tool for predicting where in the genome we might expect hybrid ancestry to persist preferentially.

223    In particular, in hominins, meiotic recombination events are directed to the genome by binding of

224    the *PRDM9* gene, whereas in swordtail fish, they are not and instead are concentrated around

225    CpG islands and other promoter-like features (*35*; Supporting Information 5-6). Accordingly, we

226    found that in swordtail fish, minor parent ancestry is higher around CpG islands and transcription

227    start sites whereas in humans, it is not (Fig. 5; Supporting Information 5).  In other words, the

228    mechanism by which recombination is directed to the genome shapes the retention of hybrid

229    ancestry.

8

230     One implication is that the reliance on PRDM9 to direct recombination may not only

231     impact reproductive isolation between species directly (as in mice, *45*), but also indirectly. For

232     example, if DMIs tend to occur between neighboring genes (*46*), hybrids between species with

233     PRDM9-independent recombination may experience greater negative selection than species that

234     use PRDM9, because recombination events are more likely to uncouple negatively-interacting

235     alleles. On the other hand, in species with PRDM9-independent recombination, genic regions

236     have higher recombination rates and thus may be more likely to be uncoupled from a deleterious

237     background, potentially providing more opportunities for adaptive introgression.  As genomic

238     data accumulates for hybridizing species across the tree of life (*47-53*), the importance of

239     recombination mechanisms for the fate of hybrids can soon be systematically evaluated.

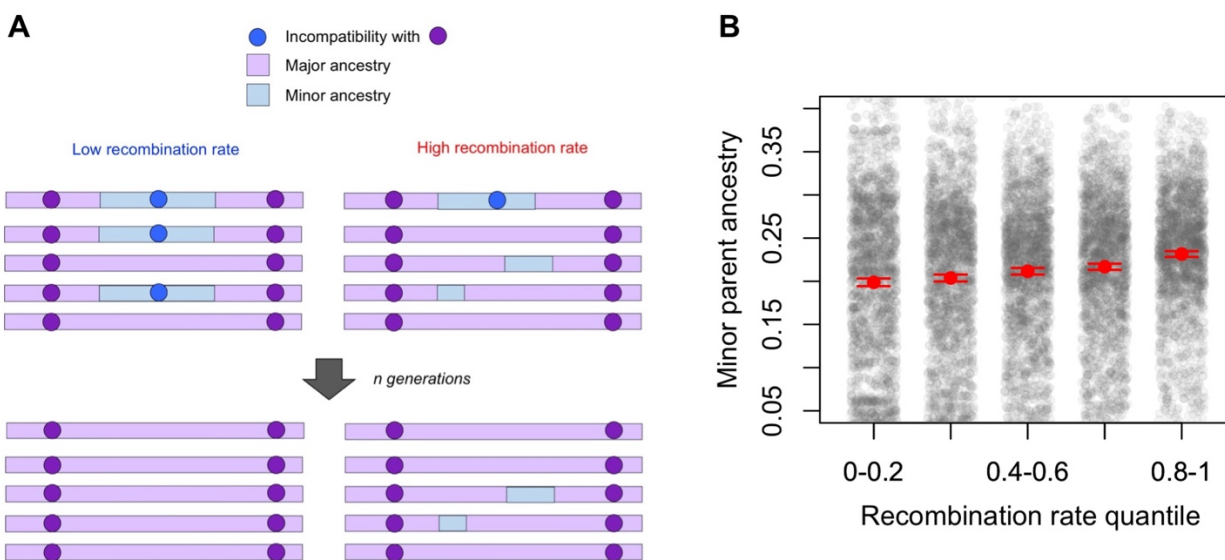240

241

250

251 **Figures**

252



253
254
255 **Figure 1. Predicted relationship between minor parent ancestry and recombination rates in**
256 **the presence of hybrid incompatibilities.** (A) In the presence of hybrid incompatibilities, the
257 local recombination rate influences the rate at which neutral regions will be uncoupled from
258 nearby hybrid incompatibility loci; as a result, more minor parent ancestry should be retained in
259 regions of high recombination. (B) This prediction is borne out in simulations. Shown here are
260 simulation results with two pairs of randomly placed hybrid incompatibilities per chromosome
261 and $s$=0.1, plausible parameters for *Xiphophorus* species (*29*, Supporting Information 5). In these
262 simulations, 250 individuals were sampled at generation 70 and ancestry along 25 Mb
263 chromosomes was summarized in 50 kb windows. Shown here is a replicate selected at random,
264 with the number of simulated chromosomes chosen to mimic the amount of data used in our
265 analyses; 74% of simulations had a significantly positive relationship between minor parent
266 ancestry and recombination rate at the 5% level. Red points and whiskers indicate the mean
267 minor parent ancestry with two standard errors of the mean determined by bootstrapping
268 windows; gray points show raw data. Note that the y-axis range is truncated.
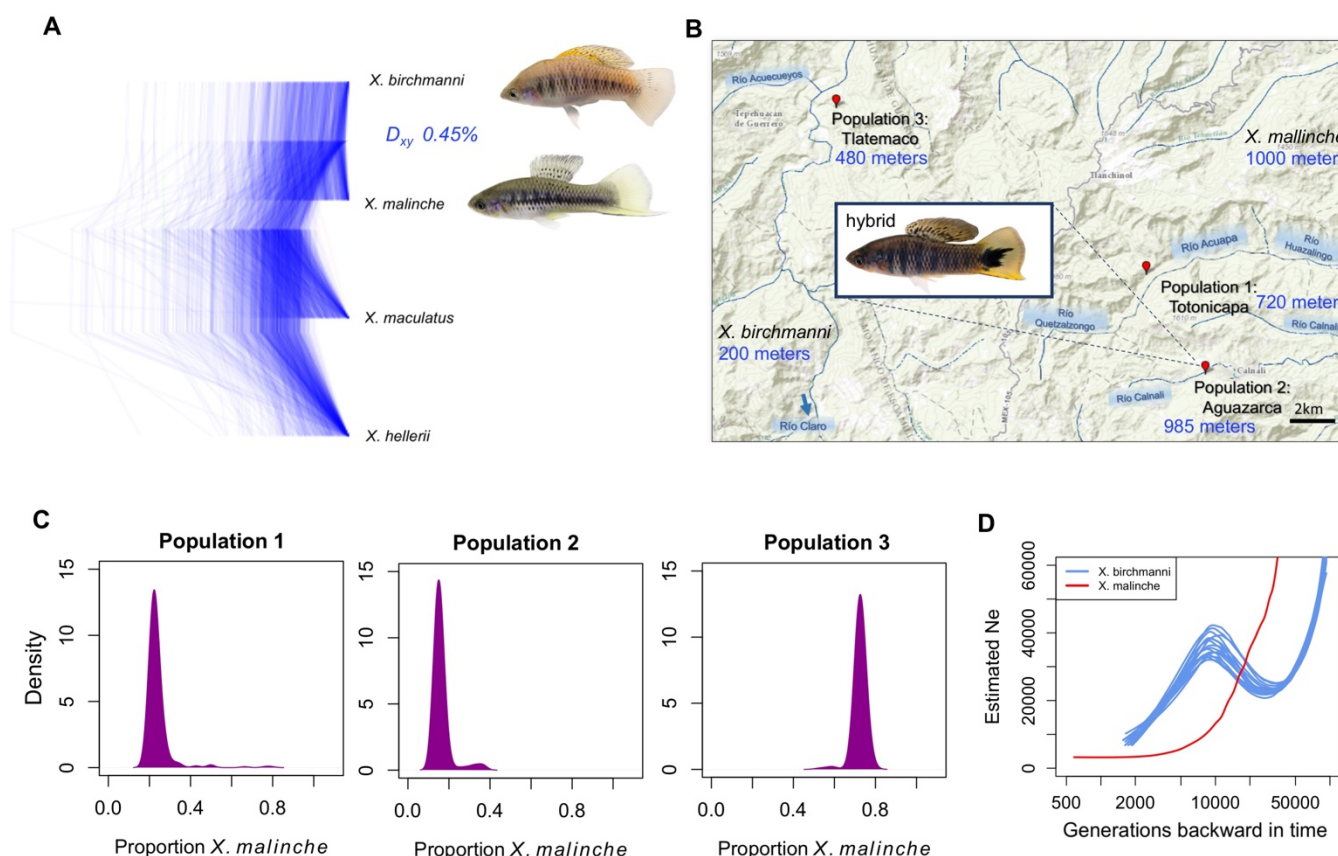
269



270
271 **Figure 2. Hybridization between sister species *X. birchmanni* and *X. malinche*.** (A)
272 Maximum likelihood trees produced by RAxML (*54*) for 1,000 alignments of randomly selected
273 10 kb regions. $D_{xy}$ refers to the average nucleotide divergence between *X. birchmanni* and *X.*
274 *malinche*. Outgroup genome sequences were obtained from previous work (*55, 56*). (B)
275 Locations of the three hybrid populations on which we focused, all sampled from different river
276 systems in Hidalgo, Mexico; listed in blue are elevations of the sampled hybrid populations and
277 typical elevations for the parental populations. (C) Inferred ancestry proportions of individuals
278 from each of the three hybrid populations (see Methods for details). (D) Loess fit to effective
279 population size estimates inferred by MSMC from one *X. malinche* genome and each of the 20 *X.*
280 *birchmanni* genomes collected for this study, assuming a mutation rate of 3.5 x $10^{-9}$ (*57*; see
281 Supporting Information 3). The time interval overlapping with zero is not plotted.

11

282
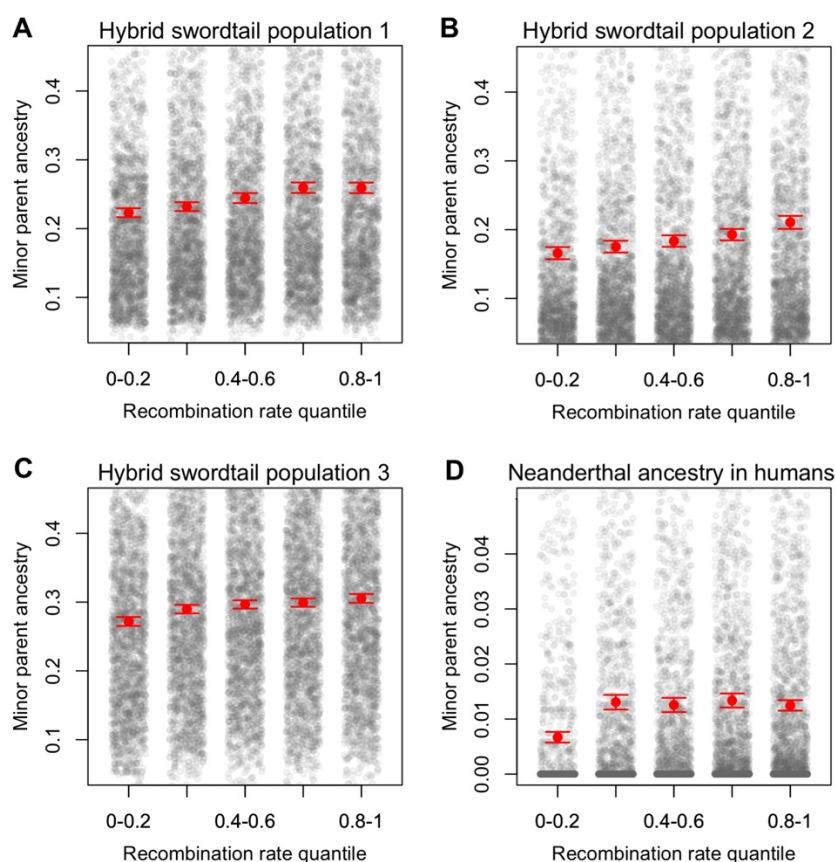


283
284 **Figure 3. Minor parent ancestry is significantly decreased in regions of low recombination,**
285 **in swordtail fish and humans.** (A-C) Minor parent ancestry is rarest in regions of the swordtail
286 genome with the lowest recombination rates. (D) Neanderthal ancestry is likewise rarest in
287 regions of the human genome with the lowest recombination rates. Removing windows of
288 particularly high Neanderthal ancestry, which may have experienced adaptive introgression,
289 further strengthens this relationship (Fig. S8). Data are summarized in 50 kb windows in
290 swordtail analyses and 250 kb windows in analysis of human data; results are similar for a range
291 of window sizes (summarized in Table 1). Quantile binning is for visual representation only; all
292 statistical tests reported in Table 1 were performed on the unbinned data. Red points and
293 whiskers indicate the mean minor parent ancestry and two standard errors of the mean, obtained
294 by bootstrapping windows; gray points show raw data. Note that the y-axis is truncated in all
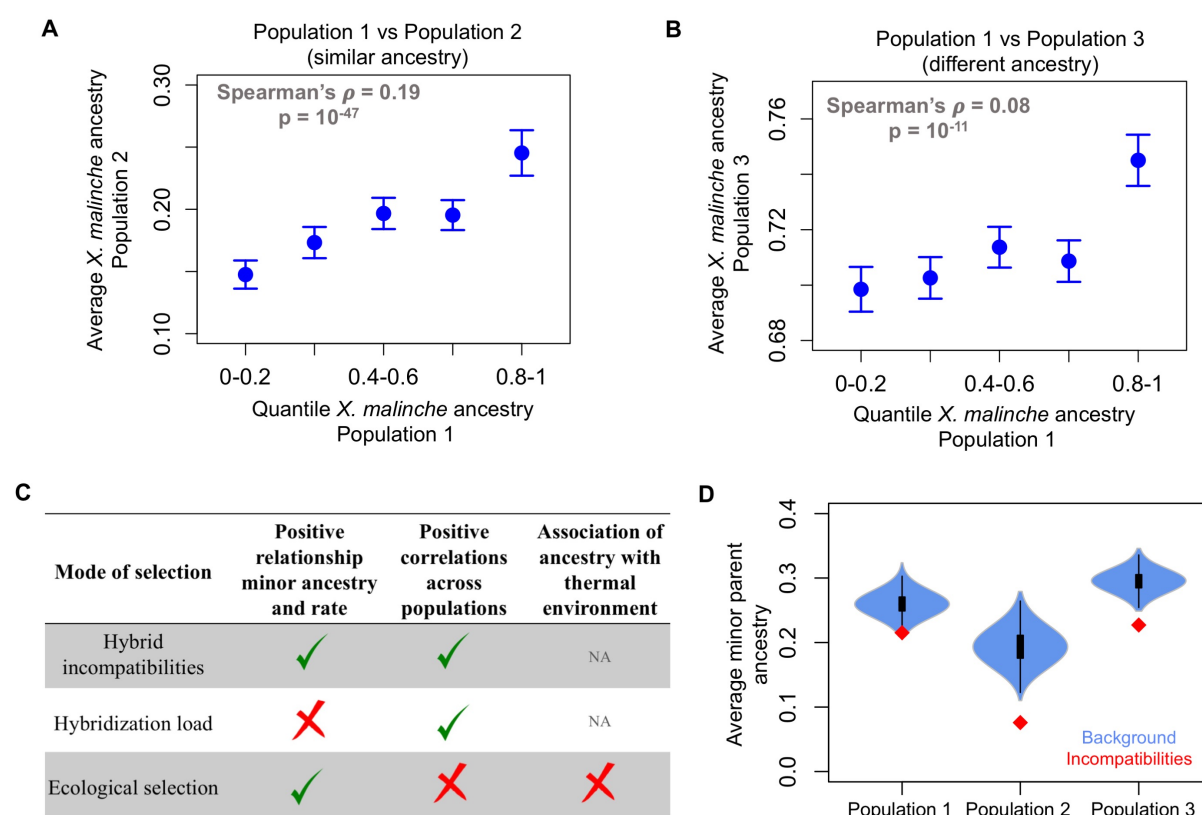295 panels.

296



297
298  **Figure 4. Evidence for DMIs as the major source of selection on hybrids.** (A) Local ancestry
299  is strongly and significantly correlated between independently formed swordtail hybrid
300  populations with similar mixture proportions (using 0.1 cM windows; Supporting Information 8).
301  (B) Local ancestry is also correlated between swordtail hybrid populations with distinct major
302  parent ancestry, but the relationship is weaker. Similar results for cross-population correlations in
303  ancestry are observed when controlling for the number of coding base pairs in a window using a
304  partial correlation analysis (when comparing population 1 and 2, $\rho = 0.19$, $p=10^{-46}$; in turn, $\rho$
305  0.04-0.08, $p<0.005$ for populations 1 and 2 versus population 3). Points show the mean ancestry
306  and whiskers indicate two standard errors of the mean. (C) Predictions for different modes of
307  selection on hybrids. Predictions met in the data are shown with a check and not met with a red
308  cross. See text for details. (D) Average minor parent ancestry is unusually depleted in 50 kb
309  windows containing previously mapped, unlinked DMIs (red points, from *31*) compared to 1,000
310  null datasets generated by randomly sampling the same number of windows from the
311  background (blue distribution). Simulations suggest that observed ancestry departures at DMIs
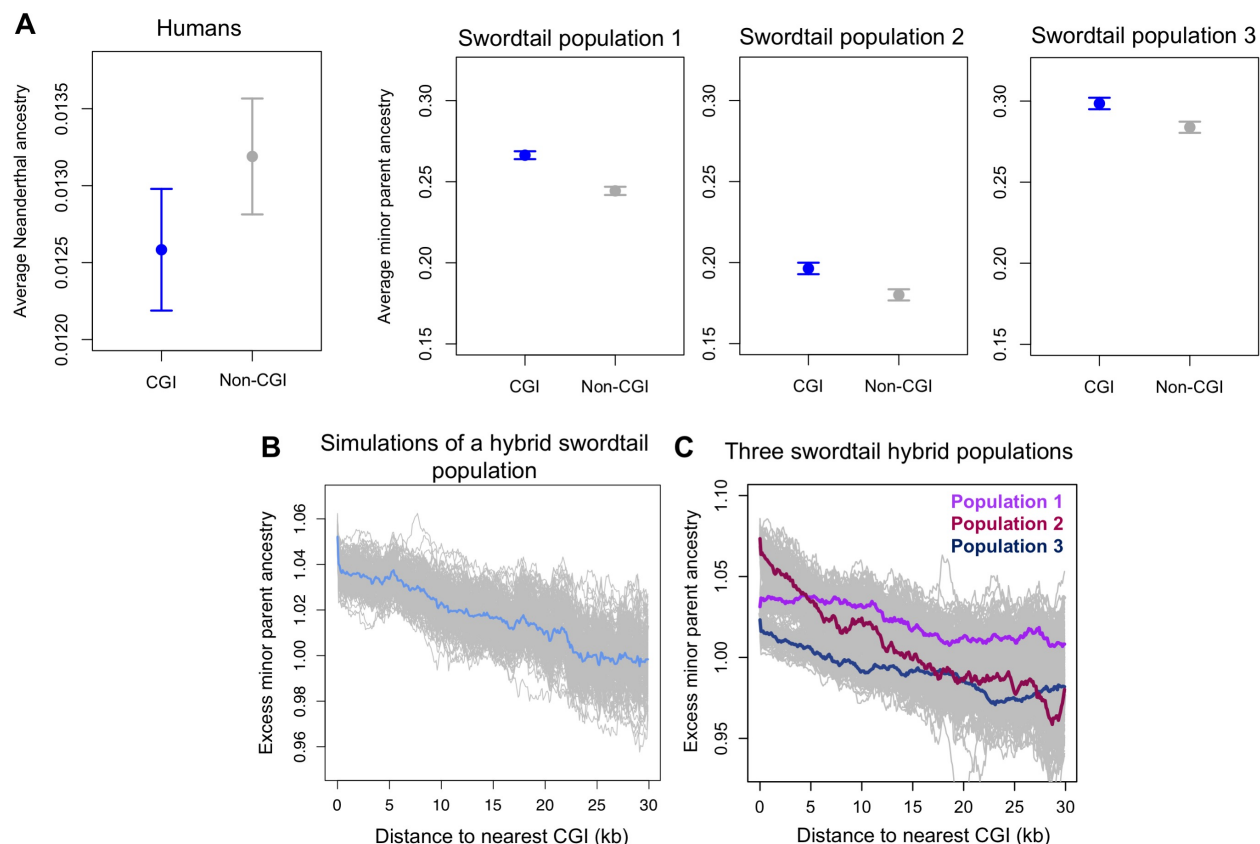312  are not expected as a result of ascertainment (Supporting Information 5).
313

314



315

316 **Figure 5. The mechanism by which recombination is directed to the genome shapes the**
317 **genomic location of minor parent ancestry.** (A) Neanderthal ancestry in the human genome is
318 not significantly elevated in 50 kb windows that overlap with CpG islands (CGIs), when
319 compared to windows that do not overlap CGIs but have similar GC base pair composition (the
320 fold difference is 0.95, p=0.91). Points show the mean of each group and whiskers indicate one
321 standard error of the mean obtained by 1,000 joint bootstraps resampling the data. (A) In
322 contrast, in all three swordtail fish populations, minor parent ancestry is significantly elevated in
323 windows that overlap CGIs compared to windows that do not but have similar GC base pair
324 composition (in population 1, the fold-difference is 1.09, p<0.005; in population 2, the fold-
325 difference is 1.09, p<0.005; and in population 3 the fold-difference is 1.02, p<0.005). See
326 Supporting Information 5 for details. (B) Simulations of incompatibility selection in swordtails
327 predict high minor parent ancestry near CGIs (Supporting Information 5). (C) This prediction is
328 met for all three hybrid populations. In B and C, gray lines show results of 500 replicate
329 simulations obtained by bootstrapping 5 kb windows and re-calculating the relationship between
330 distance to the nearest CGI and minor parent ancestry; colored lines indicate the mean of all
331 replicates in sliding 5 kb windows. The number of simulated chromosomes shown in B was
332 chosen to mimic the amount of data used in our analyses, and is one replicate chosen at random.

333
334

14

**Table 1. Relationship between minor parent ancestry and recombination rate.** Results for Spearman's correlations between minor parent ancestry and recombination rate at several scales, in three swordtail fish hybrid populations and for archaic hominin ancestry in the human genome. Also shown are partial correlation results, controlling for the number of coding base pairs. Because different filtering approaches were applied to the archaic hominin ancestry datasets in the original studies, results reported here are based on windows included in all datasets. P-values are obtained after thinning windows, to minimize correlations among nearby windows (see Supporting Information 4). Additional information and analyses can be found in Supporting Information 4 and Supporting Information 9.

| Population | Spearman's correlation between minor ancestry and rate | | | Spearman's partial correlation between minor ancestry and rate including coding base pairs | | |
| --- | --- | --- | --- | --- | --- | --- |
| | 50 kb | 250 kb | 500 kb | 50 kb | 250 kb | 500 kb |
| **Swordtail population 1** | $\rho = 0.14$ $p = 10^{-10}$ | $\rho = 0.15$ $p = 10^{-8}$ | $\rho = 0.18$ $p = 10^{-10}$ | $\rho = 0.15$ $p = 10^{-11}$ | $\rho = 0.16$ $p = 10^{-8}$ | $\rho = 0.18$ $p = 10^{-10}$ |
| **Swordtail population 2** | $\rho = 0.1$ $p = 10^{-4}$ | $\rho = 0.10$ $p = 10^{-4}$ | $\rho = 0.08$ $p = 0.007$ | $\rho = 0.1$ $p = 10^{-4}$ | $\rho = 0.11$ $p = 10^{-4}$ | $\rho = 0.08$ $p = 0.008$ |
| **Swordtail population 3** | $\rho = 0.11$ $p = 10^{-6}$ | $\rho = 0.11$ $p = 10^{-4}$ | $\rho = 0.12$ $p = 10^{-5}$ | $\rho = 0.11$ $p = 10^{-6}$ | $\rho = 0.11$ $p = 10^{-4}$ | $\rho = 0.11$ $p = 10^{-4}$ |
| **Neanderthal ancestry in humans** (diCal-admix) | $\rho = 0.08$ $p = 10^{-14}$ | $\rho = 0.17$ $p = 10^{-35}$ | $\rho = 0.19$ $p = 10^{-42}$ | $\rho = 0.08$ $p = 10^{-13}$ | $\rho = 0.17$ $p = 10^{-34}$ | $\rho = 0.19$ $p = 10^{-43}$ |
| **Neanderthal ancestry in humans** (Sankararaman et al. 2014) | $\rho = 0.06$ $p = 10^{-8}$ | $\rho = 0.15$ $p = 10^{-26}$ | $\rho = 0.16$ $p = 10^{-34}$ | $\rho = 0.06$ $p = 10^{-8}$ | $\rho = 0.14$ $p = 10^{-25}$ | $\rho = 0.17$ $p = 10^{-34}$ |
| **Neanderthal ancestry in humans** (Skov et al. 2017) | $\rho = 0.07$ $p = 10^{-10}$ | $\rho = 0.14$ $p = 10^{-24}$ | $\rho = 0.15$ $p = 10^{-30}$ | $\rho = 0.07$ $p = 10^{-9}$ | $\rho = 0.14$ $p = 10^{-24}$ | $\rho = 0.16$ $p = 10^{-30}$ |
| **Denisovan ancestry in humans** (Sankararaman et al. 2016) | $\rho = 0.02$ $p = 0.06$ | $\rho = 0.05$ $p = 10^{-4}$ | $\rho = 0.07$ $p = 10^{-6}$ | $\rho = 0.02$ $p = 0.07$ | $\rho = 0.05$ $p = 10^{-4}$ | $\rho = 0.07$ $p = 10^{-6}$ |

15

**References**

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

2. E. H. Stukenbrock, F. B. Christiansen, T. T. Hansen, J. Y. Dutheil, M. H. Schierup, Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 10954-10959 (2012).

3. D. Bachtrog, K. Thornton, A. Clark, P. Andolfatto, Extensive introgression of mitochondrial DNA relative to nuclear genes in the Drosophila yakuba species group. *Evolution* **60**, 292-302 (2006).

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

5. R. Cui *et al.*, Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution* **67**, 2166–2179 (2013).

6. S. Sankararaman *et al.*, The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354-357 (2014).

7. F. Jacobsen, K. E. Omland, Increasing evidence of the role of gene flow in animal evolution: hybrid speciation in the yellow-rumped warbler complex. *Molecular Ecology* **20**, 2236-2239 (2011).

8. J. W. Poelstra *et al.*, The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**, 1410-1414 (2014).

9. B. M. Fitzpatrick, H. B. Shaffer, Hybrid vigor between native and introduced salamanders raises new challenges for conservation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 15793-15798 (2007).

10. I. Juric, S. Aeschbacher, G. Coop, The Strength of Selection against Neanderthal Introgression. *Plos Genetics* **12**, (2016).

11. M. Schumer, R. F. Cui, D. L. Powell, G. G. Rosenthal, P. Andolfatto, Ancient hybridization and genomic stabilization in a swordtail fish. *Molecular Ecology* **25**, 2661-2679 (2016).

12. J. P. Masly, D. C. Presgraves, High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS Biology* **5**, 1890-1898 (2007).

377   13.   D. C. Presgraves, A fine-scale genetic analysis of hybrid Incompatibilities in drosophila.
378         *Genetics* **163**, 955-972 (2003).

379   14.   S. Tang, D. C. Presgraves, Evolution of the Drosophila Nuclear Pore Complex Results in
380         Multiple Hybrid Incompatibilities. *Science* **323**, 779-782 (2009).

381   15.   K. Bomblies *et al.*, Autoimmune response as a mechanism for a Dobzhansky-Muller-type
382         incompatibility syndrome in plants. *PLoS Biology* **5**, 1962-1972 (2007).

383   16.   H.-Y. Lee *et al.*, Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid
384         Sterility between Two Yeast Species. *Cell* **135**, 1065-1073 (2008).

385   17.   K. Harris, R. Nielsen, The Genetic Cost of Neanderthal Introgression. *Genetics* **203**, 881-
386         891 (2016).

387   18.   L. H. Rieseberg *et al.*, Major ecological transitions in wild sunflowers facilitated by
388         hybridization. *Science* **301**, 1211-1216 (2003).

389   19.   M. E. Arnegard *et al.*, Genetics of ecological divergence during speciation. *Nature* **511**,
390         307-311 (2014).

391   20.   C. I. Wu, The genic view of the process of speciation. *Journal of Evolutionary Biology*
392         **14**, 851-865 (2001).

393   21.   A. A. Hoffmann, L. H. Rieseberg, Revisiting the Impact of Inversions in Evolution: From
394         Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual*
395         *Review of Ecology Evolution and Systematics* **39**, 21-42 (2008).

396   22.   M. A. F. Noor, K. L. Grams, L. A. Bertucci, J. Reiland, Chromosomal inversions and the
397         reproductive isolation of species. *Proceedings of the National Academy of Sciences of the*
398         *United States of America* **98**, 12084-12088 (2001).

399   23.   M. W. Nachman, B. A. Payseur, Recombination rate variation and speciation: theoretical
400         predictions and empirical results from rabbits and mice. *Philosophical Transactions of*
401         *the Royal Society B-Biological Sciences* **367**, 409-421 (2012).

402   24.   N. Barton, B. O. Bengtsson, The barrier to genetic exchange between hybridizing
403         populations. *Heredity* **57**, 357-376 (1986).

404   25.   D. Ortiz-Barrientos, J. Engelstadter, L. H. Rieseberg, Recombination Rate Evolution and
405         the Origin of Species. *Trends in Ecology & Evolution* **31**, 226-236 (2016).

406   26.   Y. Brandvain, A. M. Kenney, L. Flagel, G. Coop, A. L. Sweigart, Speciation and
407         Introgression between Mimulus nasutus and Mimulus guttatus. *Plos Genetics* **10**,  (2014).

17

408  27.  A. M. Kenney, A. L. Sweigart, Reproductive isolation and introgression between
409       sympatric Mimulus species. *Molecular Ecology* **25**, 2499-2517 (2016).

410  28.  S. Renaut *et al.*, Genomic islands of divergence are not affected by geography of
411       speciation in sunflowers. *Nature Communications* **4**,  (2013).

412  29.  M. Schumer *et al.*, High-resolution Mapping Reveals Hundreds of Genetic
413       Incompatibilities in Hybridizing Fish Species. *eLife* **3**,  (2014).

414  30.  H. S. Fisher, B. B. M. Wong, G. G. Rosenthal, Alteration of the chemical environment
415       disrupts communication in a freshwater fish. *Proceedings of the Royal Society London*
416       *Series B* **273**, 1187-1193 (2006).

417  31.  M. Schumer, Y. Brandvain, Determining epistatic selection in admixed populations.
418       *Molecular Ecology* **25**, 2577-2591 (2016).

419  32.  P. Andolfatto *et al.*, Multiplexed shotgun genotyping for rapid and efficient genetic
420       mapping. *Genome Research* **21**, 610-617 (2011).

421  33.  M. Schumer, R. Cui, G. Rosenthal, P. Andolfatto, simMSG: an experimental design tool
422       for high-throughput genotyping of hybrids. *Molecular Ecology Resources* **16**,  (2015).

423  34.  M. Schumer *et al.*, Assortative mating and persistent reproductive isolation in hybrids.
424       *Proceedings of the National Academy of Sciences of the United States of America*
425       **114** 10936–10941 (2017).

426  35.  Z. Baker *et al.*, Repeated losses of PRDM9-directed recombination despite the
427       conservation of PRDM9 across vertebrates. *eLife* **6**,  (2017).

428  36.  R. Do *et al.*, No evidence that selection has been less effective at removing deleterious
429       mutations in Europeans than in Africans. *Nature Genetics* **47**, 126-131 (2015).

430  37.  Y. B. Simons, G. Sella, The impact of recent population history on the deleterious
431       mutation load in humans and close evolutionary relatives. *Current Opinion in Genetics &*
432       *Development* **41**, 150-158 (2016).

433  38.  B. Vernot, J. M. Akey, Resurrecting Surviving Neandertal Lineages from Modern Human
434       Genomes. *Science* **343**, 1017-1021 (2014).

435  39.  K. Prufer *et al.*, The complete genome sequence of a Neanderthal from the Altai
436       Mountains. *Nature* **505**, 43-49 (2014).

437  40.  D. Enard, D. Petrov,  RNA viruses drove adaptive introgressions between Neanderthals
438       and modern humans. *bioRxiv*,  (2017).

18

439    41.    F. Racimo, S. Sankararaman, R. Nielsen, E. Huerta-Sanchez, Evidence for archaic
440           adaptive introgression in humans. *Nature Reviews Genetics* **16**, 359-371 (2015).

441    42.    S. Sankararaman, S. Mallick, N. Patterson, D. Reich, The Combined Landscape of
442           Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology* **26**, 1241-
443           1247 (2016).

444    43.    H. A. Orr, The population genetics of speciation - the evolution of hybrid
445           incompatibilities. *Genetics* **139**, 1805-1813 (1995).

446    44.    H. A. Orr, M. Turelli, The evolution of postzygotic isolation: Accumulating Dobzhansky-
447           Muller incompatibilities. *Evolution* **55**, 1085-1094 (2001).

448    45.    B. Davies *et al.*, Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in
449           mice. *Nature* **530**, 171-176 (2016).

450    46.    Y. Yang, W. Cao, S. Wu, W. Qian, Genetic interaction network as an important
451           determinant of gene order in genome evolution. *Molecular Biology and Evolution*
452           https://doi.org/10.1093/molbev/msx264,  (2017).

453    47.    J. I. Meier *et al.*, Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature
454           Communications* **8**,  (2017).

455    48.    A. Brelsford, B. Mila, D. E. Irwin, Hybrid origin of Audubon's warbler. *Molecular
456           Ecology* **20**, 2380-2389 (2011).

457    49.    B. M. vonHoldt *et al.*, A genome-wide perspective on the evolutionary history of
458           enigmatic wolf-like canids. *Genome Research* **21**, 1294-1305 (2011).

459    50.    J. B. Leducq *et al.*, Speciation driven by hybridization and chromosomal plasticity in a
460           wild yeast. *Nature Microbiology* **1**,  (2016).

461    51.    E. B. Taylor *et al.*, Speciation in reverse: morphological and genetic evidence of the
462           collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Molecular
463           Ecology* **15**, 343-355 (2006).

464    52.    L. M. Turner, M. A. White, D. Tautz, B. A. Payseur, Genomic Networks of Hybrid
465           Sterility. *PLoS Genet* **10**,  (2014).

466    53.    G. Li, B. W. Davis, E. Eizirik, W. J. Murphy, Phylogenomic evidence for ancient
467           hybridization in the genomes of living cats (Felidae). *Genome Research* **26**, 1-11 (2016).

468    54.    A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with
469           thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).

19

470    55.    M. Schumer *et al.*, An evaluation of the hybrid speciation hypothesis for *Xiphophorus*

471            *clemenciae* based on whole genome sequences. *Evolution* **67**, 1155-1168 (2013).

472    56.    M. Schartl *et al.*, The genome of the platyfish, Xiphophorus maculatus, provides insights

473            into evolutionary adaptation and several complex traits. *Nat Genet* **45**, 567-U150 (2013).

474    57.    M. Malinsky *et al.*, Whole genome sequences of Malawi cichlids reveal  multiple

475            radiations interconnected  by gene flow. *bioRxiv*,  (2017).

476

477    **Supplementary Materials**

478    Materials and Methods

479    Table S1-S5

480    Figures S1-S25

481