Figures and figure supplements

High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species

Molly Schumer, et al.
Figure 1. Hybrids between X. malinche and X. birchmanni. (A) Parental (X. malinche top, X. birchmanni bottom) and (B) hybrid phenotypes with sample MSG genotype plots for linkage groups 1–3 (see Figure 1—figure supplement 1 for more examples) for each population shown in the right panel. In MSG plots, solid blue indicates homozygous malinche, solid red indicates homozygous birchmanni and regions of no shading indicate heterozygosity. Hybrid individuals from Tlatemaco (B top) have malinche-biased ancestry (solid blue regions) while hybrids from Calinali (B bottom) have birchmanni-biased ancestry (solid red regions).

DOI: 10.7554/eLife.02535.003
**Figure 1—figure supplement 1.** MSG ancestry plots for parental and hybrid individuals.
DOI: 10.7554/eLife.02535.004
**Figure 2.** $R^2$ distribution and p-value distributions of the sites analyzed in this study. (A) Genome-wide distribution of randomly sampled $R^2$ values for markers on separate chromosomes (see [Figure 2—figure supplement 1](#) for $R^2$ decay by distance; [Figure 2—figure supplement 2](#) for a genome-wide plot). Blue indicates the distribution in Tlatemaco while yellow indicates the distribution in Calnali. Regions of overlapping density are indicated in green. The average genome-wide $R^2$ in Tlatemaco is 0.003 and in Calnali is 0.006. (B) qq-plots of $-\log_{10}(p$-value) for a randomly selected subset of unlinked sites analyzed in this study in each population; expected p-values are drawn from p-values of the permuted data.

DOI: 10.7554/eLife.02535.005

**Figure 2—figure supplement 1.** Decay in linkage disequilibrium.

DOI: 10.7554/eLife.02535.006
Figure 2—figure supplement 2. Genome-wide linkage disequilibrium plot for Tlatemaco.
DOI: 10.7554/eLife.02535.007
Figure 3. Number of unlinked pairs in significant linkage disequilibrium and expected false discovery rates. Plot showing number of pairs of sites in significant LD in both populations in the stringent and relaxed data sets (light blue). The expected number of false positives in each data set is shown in dark blue, and was determined by simulation (see main text; Figure 3—figure supplement 1). DOI: 10.7554/eLife.02535.008
Figure 3—figure supplement 1. False discovery rate (FDR) at different p-value thresholds.
DOI: 10.7554/eLife.02535.009
Figure 4. Distribution of sites in significant linkage disequilibrium throughout the Xiphophorus genome. Schematic of regions in significant LD in both populations at FDR 5%. Regions in blue indicate regions that are positively associated in both populations (conspecific in association), regions in black indicate associations with different signs of R in the two populations, while regions in red indicate those that are negatively associated in both populations (heterospecific in association). Chromosome lengths and position of LD regions are relative to the length of the assembled sequence for that linkage group; most identified LD regions are <50 kb (Figure 4—figure supplement 1, Figure 4—figure supplement 2, Figure 4—figure supplement 3). Analysis of local LD excludes mis-assemblies as the cause of these patterns (Figure 4—figure supplement 4).
DOI: 10.7554/eLife.02535.010

Figure 4—figure supplement 1. Log₁₀ distribution of LD region length in base pairs.
DOI: 10.7554/eLife.02535.011
Figure 4—figure supplement 2. Plot of the number of recombination breakpoints detected along linkage group 2.

DOI: 10.7554/eLife.02535.012
Figure 4—figure supplement 3  Example of the use of data from two populations to narrow candidate regions in cross-chromosomal LD.
DOI: 10.7554/eLife.02535.013

Figure 4—figure supplement 4  Regions in cross-chromosomal LD are also in LD with their neighbors.
DOI: 10.7554/eLife.02535.014
Figure 5. Loci in significant conspecific linkage disequilibrium show patterns consistent with selection against hybrid incompatibilities. (A) Posterior distributions of the selection coefficient and hybrid population size from ABC simulations for Tlatemaco and (B) Calnali. The range of the x-axis indicates the range of the prior distribution, maximum a posteriori estimates (MAP) and 95% CI are indicated in the inset. (C) Departures from expectations under random mating in the actual data (top—blue points indicate LD pairs, black points indicate random pairs from the genomic background) and samples generated by posterior predictive simulations (bottom, see ‘Materials and methods’). The mean is indicated by a dark blue point; in the real data (top) smears denote the distribution of means for 1000 simulations while in the simulated data (bottom) smears indicate results of each simulation. Genotypes with the same predicted deviations on average under the BDM model have been collapsed (Figure 5—figure supplement 1, but see Figure 5—figure supplement 3) and are abbreviated in the format locus1_locus2. These simulations show that the observed deviations are expected under the BDM model. The posterior distributions for $s$ and hybrid population size are correlated at low population sizes (Figure 5—figure supplement 2). Deviations in Calnali also follow expectations under the BDM model (Figure 5—figure supplement 3).

DOI: 10.7554/eLife.02535.017
Figure 5—figure supplement 1. Different fitness matrices associated with selection against hybrid incompatibilities.

DOI: 10.7554/eLife.02535.018
Figure 5—figure supplement 2. Joint posterior distribution of hybrid population size and selection coefficient.
DOI: 10.7554/eLife.02535.019
Figure 5—figure supplement 3. Deviations in genotype combinations compared to expected values under a two-locus selection model in both populations.
DOI: 10.7554/eLife.02535.020
Figure 6. Divergence of LD pairs compared to the genomic background in two species comparisons. (A) Regions identified in *X. birchmanni* and *X. malinche* and (B) orthologous regions in *X. hellerii* and *X. clemenciae*. The blue point shows the average divergence for genomic regions within significant LD pairs, and whiskers denote a 95% confidence interval estimated by resampling genomic regions with replacement. The histogram shows the distribution of the average divergence for 1000 null data sets generated by resampling the genomic background without replacement.
DOI: 10.7554/eLife.02535.021