Title: A Lethal Genetic Incompatibility between Naturally Hybridizing Species in Mitochondrial Complex I

One sentence summary: Accelerated evolution of three interacting nuclear and mitochondrial genes underlies a lethal incompatibility in swordtail hybrids.

Authors: Benjamin M. Moran[1,2]*, Cheyenne Y. Payne[1,2], Daniel L. Powell[1,2], Erik N. K. Iverson[3], Shreya M. Banerjee[1], Quinn K. Langdon[1], Theresa R. Gunn[1], Fang Liu[4], Rowan Matney[4], Kratika Singhal[4], Ryan D. Leib[4], Osvaldo Hernandez-Perez[2], Russell Corbett-Detig[5,6], Manfred Schartl[7,8], Justin C. Havird[3], Molly Schumer[1,2,9]*

Affiliations:

[1]Department of Biology, Stanford University, Stanford, CA, USA

[2]Centro de Investigaciones Científicas de las Huastecas "Aguazarca", A.C., Calnali, Hidalgo, Mexico

[3]Department of Integrative Biology, University of Texas, Austin, TX, USA

[4]Stanford University Mass Spectrometry Core, Stanford University, Stanford, CA, USA

[5]Genomics Institute, University of California, Santa Cruz, CA, USA

[6]Department of Biomolecular Engineering, University of California, Santa Cruz, CA, USA

[7]The Xiphophorus Genetic Stock Center, Texas State University, San Marcos, TX 78666

[8]Developmental Biochemistry, Biozentrum, University of Würzburg, 97070 Würzburg, Germany

[9]Hanna H. Gray Fellow, Howard Hughes Medical Institute, Stanford, CA, USA

*Correspondence: benmoran@stanford.edu, schumer@stanford.edu

Abstract

The evolution of reproductive barriers is fundamental to the formation of new species and can help us understand the diversification of life on Earth. These reproductive barriers often take the form of "hybrid incompatibilities," where genes derived from two different species no longer interact properly. Theory predicts that incompatibilities involving multiple genes should be common and that rapidly evolving genes will be more likely to cause incompatibilities, but empirical evidence has lagged behind these predictions. Here, we describe a mitonuclear incompatibility involving three genes within respiratory Complex I in naturally hybridizing swordtail fish. Individuals with specific mismatched protein combinations fail to complete embryonic development while those heterozygous for the incompatibility have reduced function of Complex I and unbalanced representation of parental alleles in the mitochondrial proteome. We localize the protein-protein interactions that underlie the incompatibility and document accelerated evolution and introgression in the genes involved. This work thus provides a precise characterization of the genetic architecture, physiological impacts, and evolutionary origin of a multi-gene incompatibility impacting naturally hybridizing species.

Main text

Biologists have long been fascinated by the question of how new species are formed and what mechanisms maintain isolation between them. One key factor in the formation and maintenance of new species is the emergence of genetic incompatibilities that reduce viability or fertility in hybrids. When species diverge from each other, they accumulate unique sets of mutations (*1*). As originally described by the Dobzhansky-Müller model of hybrid incompatibility (DMI model; *2*, *3*), when these mutations are brought together in hybrids, they may interact poorly, given that they have never been tested against one another by selection. Due to the technical challenges of identifying these interactions (*4*), only a handful of genes involved in hybrid incompatibilities have been precisely mapped (*5*) and exploration of the functional and evolutionary causes of hybrid incompatibilities has been limited to a small number of cases in model organisms (*4*).

As a consequence of this knowledge gap, key predictions about the genetic architecture of hybrid incompatibilities and the evolutionary processes that drive their emergence remain untested. For one, theory suggests that incompatibilities should be more common within dense gene networks, both because the number of potentially incompatible genotypes explodes as the complexity of the genetic interaction increases and because genes involved in such interactions are expected to be tightly co-evolving (*6*, *7*). Consistent with this prediction, mutagenesis experiments have highlighted the sensitivity of multi-protein interactions to changes in any of their components (*6*). However, genetic interactions are notoriously difficult to detect empirically except in systems with especially powerful genetic tools (*8*), and this problem is exacerbated with complex genetic interactions (*9*, *10*). Such technical challenges may explain

their rarity in the empirical literature (*6*, but see *11–14*). Another largely untested prediction is that rapid molecular evolution will increase the rate at which incompatibilities accumulate between species (*4*, *5*, *15*). While several incompatibilities identified to date show signatures of positive selection, it is unclear how unusual rates of protein evolution are in genes involved in hybrid incompatibilities relative to the genomic background (*5*, *15*).

Another open question is the degree to which the genes involved in hybrid incompatibilities may be predictable from their molecular or evolutionary properties. The mitochondrial genome, in particular, has been proposed as a hotspot for the accumulation of genetic incompatibilities (*16*, *17*). Mitochondria are essential for energy production in nearly all eukaryotic organisms (*18*). In addition to this critical role, the particularities of mitochondrial inheritance and function might drive the rapid evolution of hybrid incompatibilities between species. Uniparental inheritance of mitochondria is predominant in animals, plants, and many fungi (*19*), creating the potential for sexually antagonistic selection (*20*, *21*). In many animals, mitochondrial genomes also experience elevated mutation rates relative to the nuclear genome which, combined with reduced effective population size and a lack of recombination, results in up to ~25X higher mitochondrial substitution rates in some species (*22–24*). At the same time, nuclear and mitochondrial gene products must directly interact with each other in key steps of ATP synthesis, increasing the likelihood of coevolution between these genomes (*25*, *26*). These molecular and evolutionary factors suggest that interactions between mitochondrial- and nuclear-encoded proteins could play an outsized role in the emergence of hybrid incompatibilities (*16*).

Although few studies have successfully identified the individual genes underlying hybrid incompatibilities (*4*, *5*), crosses in numerous species have provided indirect evidence for the prevalence of mitonuclear incompatibilities, since hybrid viability often depends on the identity

of the maternal species (*27–30*). However, the field has struggled to move beyond these coarse-scale patterns, especially in non-model systems where large mapping experiments can be infeasible. Despite predictions that mitonuclear incompatibilities play a disproportionate role in the evolution of reproductive isolation, few studies have mapped mitonuclear incompatibilities to the single gene level (*31–34*) and none of those identified to date have been studied in species that naturally hybridize.

As we make progress identifying the individual genes underlying hybrid incompatibilities, the next frontier is evaluating the processes that drive their evolution. Here, we use an integrative approach to precisely map the genetic basis of a lethal mitonuclear hybrid incompatibility in swordtail fish and to uncover its evolutionary history. Sister species *Xiphophorus birchmanni* and *X. malinche* began hybridizing in the last ~100 generations in multiple river systems (*35*) after premating barriers were disrupted by habitat disturbance (*36*), and are a powerful system to study the emergence of hybrid incompatibilities in young species. Given their recent divergence (~250,000 generations; 0.5% divergence per basepair; *37*), some hybrids between *X. birchmanni* and *X. malinche* are viable and fertile, while others experience strong selection against incompatibilities (*37*, *38*). Leveraging data from controlled laboratory crosses and natural hybrid zones, we pinpoint three genes that generate a lethal mitonuclear hybrid incompatibility, characterize its developmental and physiological effects, and trace its evolutionary history.

Admixture Mapping Reveals a Lethal Mitonuclear Incompatibility

To identify loci under selection in *X. birchmanni* × *X. malinche* hybrids, we generated ~1X low-coverage whole-genome sequence data for 943 individuals from controlled laboratory crosses and 281 wild-caught hybrid adults, and applied a hidden Markov model to data at 629,524 ancestry-informative sites in order to infer local ancestry (~1 per kb; *39*, *40*; Supplementary Materials 1.1.1-1.1.4). Using these results, we found evidence for a previously unknown incompatibility between the nuclear and mitochondrial genomes of *X. birchmanni* and *X. malinche* (Supplementary Materials 1.1.4-1.1.7). Our first direct evidence came from controlled laboratory crosses (Supplementary Materials 1.1.1, 1.1.2). Because the cross is largely unsuccessful in the opposite direction, all lab-bred hybrids were the offspring of F_1 hybrids generated between *X. malinche* females and *X. birchmanni* males and harbored a mitochondrial haplotype derived from the *X. malinche* parent species. Offspring of F_1 intercrosses are expected to derive on average 50% of their genome from each parent species. This expectation is satisfied genome-wide and locally along most chromosomes in F_2 hybrids (on average 50.3% *X. malinche* ancestry; Fig. S1). However, we detected severe segregation distortion along a 6.5 Mb block of chromosome 13, where average ancestry among surviving individuals was 67% *X. malinche*. Closer examination of genotypes in a 3.75 Mb subregion showed that almost none of the surviving individuals harbored homozygous *X. birchmanni* ancestry in this region of chromosome 13 (Fig. 1A; Fig. S2; 0.5% observed vs 25% expected). This pattern is unexpected even in the case of a lethal nuclear-nuclear incompatibility (where we should recover homozygous *X. birchmanni* ancestry in 6% of individuals; Supplementary Materials 1.1.2), but is consistent with a lethal mitonuclear incompatibility. Using approximate Bayesian computation

(ABC) approaches we asked what strength of selection against *X. birchmanni* ancestry in this region was consistent with the genotypes and ancestry deviations observed. We estimated posterior distributions of selection and dominance coefficients and found that inferred selection on this genotype in F_2s is largely recessive and essentially lethal (maximum a posteriori estimate: $h = 0.12$ and $s = 0.996$, 95% credible interval $h = 0.010$-0.194 and $s = 0.986$-0.999; Fig. 1D, Supplementary Materials 1.2.1-1.2.2).

To directly evaluate evidence for a mitonuclear incompatibility and pinpoint the region involved on chromosome 13, we leveraged data from natural hybrid populations. Most naturally occurring *X. birchmanni* × *malinche* hybrid populations are fixed for mitochondrial haplotypes from one parental species (Supplementary Materials 1.1.6). However, a few segregate for the mitochondrial genomes of both parental types, and we focused on one such population (the "Calnali Low" population, hereafter the admixture mapping population). Admixture mapping of mitochondrial ancestry (controlling for genome-wide ancestry) revealed a sharp peak of association that spanned approximately 380 kb within the region of chromosome 13 identified using F_2 crosses (Fig. 1B; Fig. S3; Supplementary Materials 1.1.5). We replicated this signal in another hybrid population (Fig. S4).

Of the 32 genes in this region (Table S1), two had known mitochondrial functions, but only one, the NADH dehydrogenase ubiquinone iron-sulfur protein 5 (*ndufs5*), directly interacts with mitochondrially encoded proteins. Using three natural hybrid populations that had fixed the mitochondrial haplotype of one of the parental species (Fig. S5), we fine-mapped the incompatible mitochondrial interactor within the 380 kb admixture mapping peak. All populations showed a depletion of non-mitochondrial parent ancestry at *ndufs5* (Supplementary

Materials 1.1.6), implicating it as the nuclear component of the mitonuclear incompatibility (Fig. 1C).

Given that hybrids with *X. malinche* mitochondria and homozygous *X. birchmanni* ancestry at *ndufs5* suffer hybrid lethality, we investigated evidence for selection on the opposite genotype combination (*X. birchmanni* mitochondria and homozygous *X. malinche* ancestry at *ndufs5*) using hybrids from the admixture mapping population (Supplementary Materials 1.2.1; Fig. S6). We found that this genotype combination was also observed at unexpectedly low frequencies (Fig. 1D, 1G; $p<10^{-4}$ by simulation, Supplementary Materials 1.2.1) and inferred to be under strong selection, albeit much weaker than the opposite ancestry combination (maximum a posteriori estimate: 0.17, 95% credible intervals 0.08-0.55; Fig. 1D, Supplementary Materials 1.2.2). We focus primarily on the incompatibility involving the *X. malinche* mitochondria in our results below (Fig. 1G) but discuss this reciprocal incompatibility in Supplementary Materials 1.2.3.

Mitonuclear incompatibility in *X. birchmanni* × *X. malinche* hybrids involves three genes

Mitochondrial protein respiratory Complex I is the first component of the electron transport chain that ultimately allows the cell to generate ATP. In vertebrates, this complex has a total of 47 subunits and assembly factors, including *ndufs5* (Table S2). While the genome-wide scan that we used to identify *ndufs5* is an unbiased method to search for mitochondrially interacting loci, it has low power given our sample size. After determining that genes in Complex I were involved in the incompatibility, we performed targeted analyses to determine if other genes in this complex appear to be under selection in hybrids (Fig. S7-S8; Supplementary Materials 1.1.7).

178 This analysis identified a significant interaction between mitochondrial ancestry and another
179 member of Complex I, the *ndufa13* gene on chromosome 6 (Supplementary Materials 1.1.7;
180 Bonferroni corrected p-value = 0.02). This indicates that the *X. malinche* mitochondria is actually
181 incompatible with at least two other genes (Fig. 1G), although we discuss uncertainty about the
182 exact architecture of the interaction in Supplementary Materials 1.1.8.
183 Depletion of non-mitochondrial parent ancestry at *ndufa13* was unidirectional (Fig. 1E),
184 consistent with selection acting only against the combination of the *X. malinche* mitochondria
185 with homozygous *X. birchmanni* ancestry at *ndufa13* (see Supplementary Materials 1.2.3 for
186 more detail). Notably, we also detect selection on *ndufa13* in our dataset of lab-bred hybrids that
187 harbor *X. malinche* mitochondria. The proportion of individuals with homozygous *X. birchmanni*
188 ancestry at *ndufa13* is 3% (compared to 0.5% at *ndufs5* and 25% expected genome-wide).
189 However, ABC approaches indicate that selection against the mismatch between *X. malinche*
190 mitochondria and homozygous *X. birchmanni* ancestry at *ndufa13* is also severe (maximum a
191 posteriori estimate: 0.91, 95% credible interval 0.87-0.94; Fig. 1E, Supplementary Materials
192 1.2.2).
193
194 **Lethal Effect of Incompatibility in Early Development**
195
196 The incompatibility involving the *X. malinche* mitochondria appears to be lethal by the
197 time individuals reach adulthood. To pinpoint the developmental timing of the incompatibility,
198 we genotyped pregnant females from the admixture mapping population and recorded the
199 developmental stages of their embryos (*41*; swordtails are livebearing fish). We focused on the
200 interaction between *X. malinche* mitochondria and homozygous *X. birchmanni* ancestry at

ndufs5, given that we did not detect an effect of ancestry at *ndufa13* on developmental stage (Fig. S9-10; Supplementary Materials 1.3.1). While developmental asynchrony is typically on the scale of 0-2 days in pure species (*42*; Supplementary Materials 1.3.1), we observed much greater variation in broods collected from the admixture mapping population where the mitochondrial incompatibility is segregating (e.g. stages normally separated by 12 days of development found in the same brood; Supplementary Materials 1.3.1; Fig. 2A-B). Genotyping results revealed that lethal mitonuclear genotype combinations were present in embryos at early developmental stages, but that these embryos failed to reach a phenotype beyond that typical of the first seven days of gestation (the full length of gestation is ~21 days in *Xiphophorus*; Fig. 2A). Comparing siblings with incompatible and compatible genotypes revealed a nearly universal lag in development between individuals with incompatible genotypes and the most fully developed individual in their brood (Fig. 2B, 2C). This developmental lag could itself cause mortality, since *Xiphophorus* embryos that do not complete embryonic development within the mother fail to survive more than a few days after birth (Supplementary Materials 1.3.1).

Mitochondrial Biology in Hybrids

The phenotypes of developing embryos suggest that the incompatibility may inhibit somatic growth and development, likely through mitochondrial dysfunction. Given that individuals with the incompatibility involving the *X. malinche* mitochondria generally do not complete embryonic development, we turned to F_1 hybrids between *X. birchmanni* and *X. malinche* to further explore the hybrid incompatibility *in vivo*. Since F_1 hybrids, which derive their mitochondria from *X. malinche* and are heterozygous for ancestry at *ndufs5* and *ndufa13*,

224 are largely viable, we asked whether there was evidence for compensatory nuclear or
225 mitochondrial regulation that might be protective in F_1 hybrids. We found no evidence for allele-
226 specific expression of *ndufs5* or *ndufa13* in F_1 hybrids (Supplementary Materials 1.3.2; Fig. S11-
227 12) or significant differences in mitochondrial copy number in F_1 hybrids (Supplementary
228 Materials 1.3.3; Fig. S13).

229 With no clear indication of a compensatory regulatory response, we reasoned that we
230 might be able to detect reduced mitochondrial function, particularly in respiratory Complex I, in
231 hybrids heterozygous for the incompatibility. To examine mitochondrial function in *X.*
232 *birchmanni*, *X. malinche*, and hybrids harboring the *X. malinche* mitochondria and heterozygous
233 ancestry at *ndufs5* and *ndufa13,* we quantified respiratory phenotypes in isolated mitochondria
234 using a multiple substrate, uncoupler, and inhibitor titration protocol with the Oroboros O2K
235 respirometer (Fig. S14; Supplementary Materials 1.3.4). We found that Complex I efficiency was
236 significantly lower in hybrids compared to the two parental species (Fig. 2D; Fig. S15, P =
237 0.023, n = 7 per genotype), although overall levels of mitochondrial respiration were unchanged
238 (Fig. 2E, P = 0.97; Supplementary Materials 1.3.4). This points to reduced function of Complex I
239 in heterozygous individuals, as well as possible physiological compensation by other
240 components in the oxidative phosphorylation pathway.

241 Given physiological evidence for reduced Complex I function in hybrids heterozygous at
242 *ndufs5* and *ndufa13*, we predicted that there might be an altered frequency of protein complexes
243 incorporating both *X. malinche* mitochondrial proteins and *X. birchmanni* proteins at *ndufs5* and
244 *ndufa13* in F_1 hybrids. To test this prediction, we took a mass spectrometry based quantitative
245 proteomics approach. We used stable isotope-labeled peptides to distinguish between the *X.*
246 *birchmanni* and *X. malinche ndufs5* and *ndufa13* peptides in mitochondrial proteomes extracted

from F_1 hybrids (see Supplementary Materials 1.4.1-1.4.4). While native *ndufa13* peptides were too rare to quantify accurately, we found consistent deviations from the expected 50-50 ratio of *X. birchmanni* to *X. malinche* peptides for *ndufs5* in F_1 hybrids, with a significant overrepresentation of *ndufs5* derived from *X. malinche* in the mitochondrial proteome (Fig. 2G-H; Supplementary Materials 1.4.5). Since we did not observe allele-specific expression of *ndufs5* (Fig 2F; Supplementary Materials 1.3.2; Fig. S11-12), this result is consistent with disproportionate degradation of *X. birchmanni*-derived *ndufs5* peptides in the mitochondrial proteome or differences in translation of *ndufs5* peptides from the two species.

Substitutions in two mitochondrial proteins contact *ndufs5* and *ndufa13*

While we can leverage the independent recombination histories of natural hybrid populations to pinpoint the nuclear components of the hybrid incompatibility, we cannot take this approach to distinguish among the 37 genes in the swordtail mitochondrial genome, which do not undergo meiotic recombination. To investigate the possible mitochondrial partners of *ndufs5* and *ndufa13*, we therefore turned to protein modeling, relying on high quality cryo-EM based structures (*43–45*).

Barring a hybrid incompatibility generated by regulatory divergence, which we address in Supplementary Materials 1.3.2, we expect hybrid incompatibilities to be driven by amino acid changes in the interacting proteins (*46*). We used the program RaptorX (*47*) to generate predicted structures of *X. birchmanni* and *X. malinche ndufs5, ndufa13*, and nearby Complex I mitochondrial and nuclear genes, which we aligned to a mouse cryo-EM Complex I structure (Fig. 3A; Fig. S16; *43*). Using this structure, we detected amino acid substitutions between *X.*

birchmanni and *X. malinche* at the interfaces of *ndufs5, ndufa13* and four mitochondrially encoded genes: *nd2, nd3, nd4l,* and *nd6* (Fig. 3A). Notably, one *ndufs5* substitution directly contacts a substitution in *nd2* (Fig. 3B, Fig. S16). An unstructured loop in *nd6* that passes between *ndufs5* and *ndufa13* harbors substitutions between *X. birchmanni* and *X. malinche* in 4 out of 22 amino acids that make up the loop, including one that is predicted to be in close proximity to a substitution in *ndufs5* (Fig 3B, Fig. S16; Supplementary Materials 1.4.6). Results of these structural predictions were robust to a number of technical variables (Supplementary Materials 1.4.6). The direct contact between species-specific mitochondrial and nuclear substitutions point to *nd2* and *nd6* as the genes most likely to be involved in the mitochondrial component of the hybrid incompatibility (Fig. 3A, 3B; Fig. S17).

Rapid evolution of Complex I proteins

Theory predicts that hybrid incompatibilities are more likely to arise in rapidly evolving genes (*4*, *5*, *15*). Consistent with this hypothesis, *ndufs5* is among the most rapidly evolving genes genome-wide between *X. birchmanni* and *X. malinche* (Fig. 3C, 3D). Aligning the *ndufs5* coding sequences of *X. birchmanni, X. malinche*, and twelve other swordtail species revealed that all four amino acid substitutions that differentiate *X. birchmanni* and *X. malinche* at *ndufs5* were derived on the *X. birchmanni* branch (Fig. 3C). Phylogenetic tests indicate that there has been accelerated evolution of *ndufs5* on this branch ($dN/dS > 99$, $N = 4$, $S = 0$, CodeML branch test $P = 0.005$, Fig. 3C). Similar patterns of rapid evolution are observed at *ndufa13*, which also showed evidence for accelerated evolution in *X. birchmanni* (Fig. 3E, $dN/dS = 1.2$, $N = 3$, $S = 1$, CodeML branch test $P = 0.002$).

Rapid evolution of *ndufs5* and *ndufa13* may be the result of coevolution with mitochondrial substitutions, a mechanism that has been proposed to explain the outsized role of the mitochondria in hybrid incompatibilities (*16*, *48*). Indeed, there is an excess of derived substitutions in the *X. birchmanni* mitochondrial protein *nd6*, one of the proteins that interacts with *ndufs5* and *ndufa13* (Table S3, Fig. S18; CodeML branch test $P = 0.005$). Moreover, a number of the substitutions in both mitochondrial and nuclear genes are predicted to have functional consequences (Supplementary Materials 1.5.1; Table S4, Fig. S19-20) and occur in regions of direct protein-protein interaction between *ndufs5, ndufa13, nd2*, and *nd6* (Fig. 3A, 3B). Such colocalization of substitutions predicted to affect protein structure is precisely what would be expected under classic models of hybrid incompatibility.

Introgression of genes underlying mitonuclear incompatibility

The presence of a mitonuclear incompatibility in *Xiphophorus* is especially intriguing, given previous reports that mitochondrial genomes may have introgressed between species (*49*). While *X. malinche* and *X. birchmanni* are sister species based on the nuclear genome, they are mitochondrially divergent, with *X. malinche* and *X. cortezi* grouped as sister species based on the mitochondrial phylogeny (Fig. 4A; *49*). As we show, all *X. cortezi* mitochondria sequenced to date are nested within *X. malinche* mitochondrial diversity (Fig. 4B; Supplementary Materials 1.5.2; Fig. S19-S21), including the interacting mitochondrial genes *nd2* and *nd6* (Fig. 3A). Simulations indicate that gene flow, rather than incomplete lineage sorting, likely drove replacement of the *X. cortezi* mitochondria with the *X. malinche* sequence ($P < 0.002$ by simulation; Fig. 4C; Supplementary Materials 1.5.3-1.5.4).

316 The introgression of the mitochondrial genome from *X. malinche* into *X. cortezi* suggests the
317 possibility that other Complex I genes may have co-introgressed (*50*). Indeed, the nucleotide
318 sequence for *ndufs5* is identical between *X. malinche* and *X. cortezi*, and the sequence of *ndufa13*
319 differs by a single synonymous mutation (although conservation of both genes is high throughout
320 *Xiphophorus*; Fig. S22-23). Identical amino acid sequences at the genes underlying the hybrid
321 incompatibility between *X. malinche* and *X. birchmanni* suggest that *X. cortezi* and *X.*
322 *birchmanni* are likely to harbor the same mitonuclear incompatibility, as a result of ancient
323 introgression between *X. malinche* and *X. cortezi* (Supplementary Materials 1.5.3-1.5.4). This
324 inference is supported by analysis of ancestry in two contemporary *X. birchmanni* × *X. cortezi*
325 hybrid zones, which reveals a significant depletion of non-mitochondrial parent ancestry at
326 *ndufs5* and *ndufa13* (Fig. 4D, Supplementary Materials 1.5.5; *51*). These results are consistent
327 with the mitonuclear incompatibility observed in *X. birchmanni* × *X. malinche* being active in *X.*
328 *birchmanni* × *X. cortezi* populations (see also *51*). This exciting finding hints that genes
329 underlying hybrid incompatibilities can introgress together, transferring incompatibilities
330 between related species.
331
332 **Implications**
333
334 An open question in the field is what genetic and evolutionary forces drive the emergence of
335 hybrid incompatibilities, especially between closely related species. We identified a mitonuclear
336 incompatibility that involves at least three genes and causes hybrid lethality in lab and wild
337 populations. Theory predicts that hybrid incompatibilities involving multiple genes should be
338 common (*6*, *7*), but with few exceptions (*8*, *12–14*), they remain virtually uncharacterized at the

339 genic level (*6*). Moreover, we show that there has been exceptionally rapid evolution in both
340 mitochondrial and interacting nuclear genes in *X. birchmanni*, which may have introduced
341 mutations that are incompatible in hybrids (Fig. 4). Whether driven by rapid adaptation or some
342 other mechanism, this is consistent with the hypothesis that the coevolution of mitochondrial and
343 nuclear genes can drive the overrepresentation of mitonuclear interactions in hybrid
344 incompatibilities (*25*, *26*, *48*).

345 Characterizing the incompatibility across multiple scales of organization allowed us to
346 identify the mechanisms through which it acts (*52–54*). Our results suggest that the lethal form of
347 the hybrid incompatibility is driven by dysfunctional protein-protein interactions in Complex I of
348 the mitochondrial electron transport chain that cause breakdown in bioenergetic pathways. These
349 physiological impacts are observed in hybrids heterozygous for the incompatibility, even though
350 these individuals escape its lethal effects.

351 Finally, this mitonuclear incompatibility provides a new case in which the same genes are
352 involved in incompatibilities across multiple species (*55–57*). Surprisingly, we found that
353 introgression has resulted in the transfer of genes underlying the incompatibility from *X.*
354 *malinche* to *X. cortezi*, and evidence from *X. birchmanni* × *X. cortezi* hybrid populations
355 indicates that the incompatibility is likely under selection in these populations as well. The
356 possibility that hybridization could transfer incompatibilities between species has not been
357 previously recognized, perhaps due to an underappreciation of the frequency of hybridization.
358 The importance of past hybridization in the structure of present reproductive barriers is a
359 promising area for future inquiry.
360
361

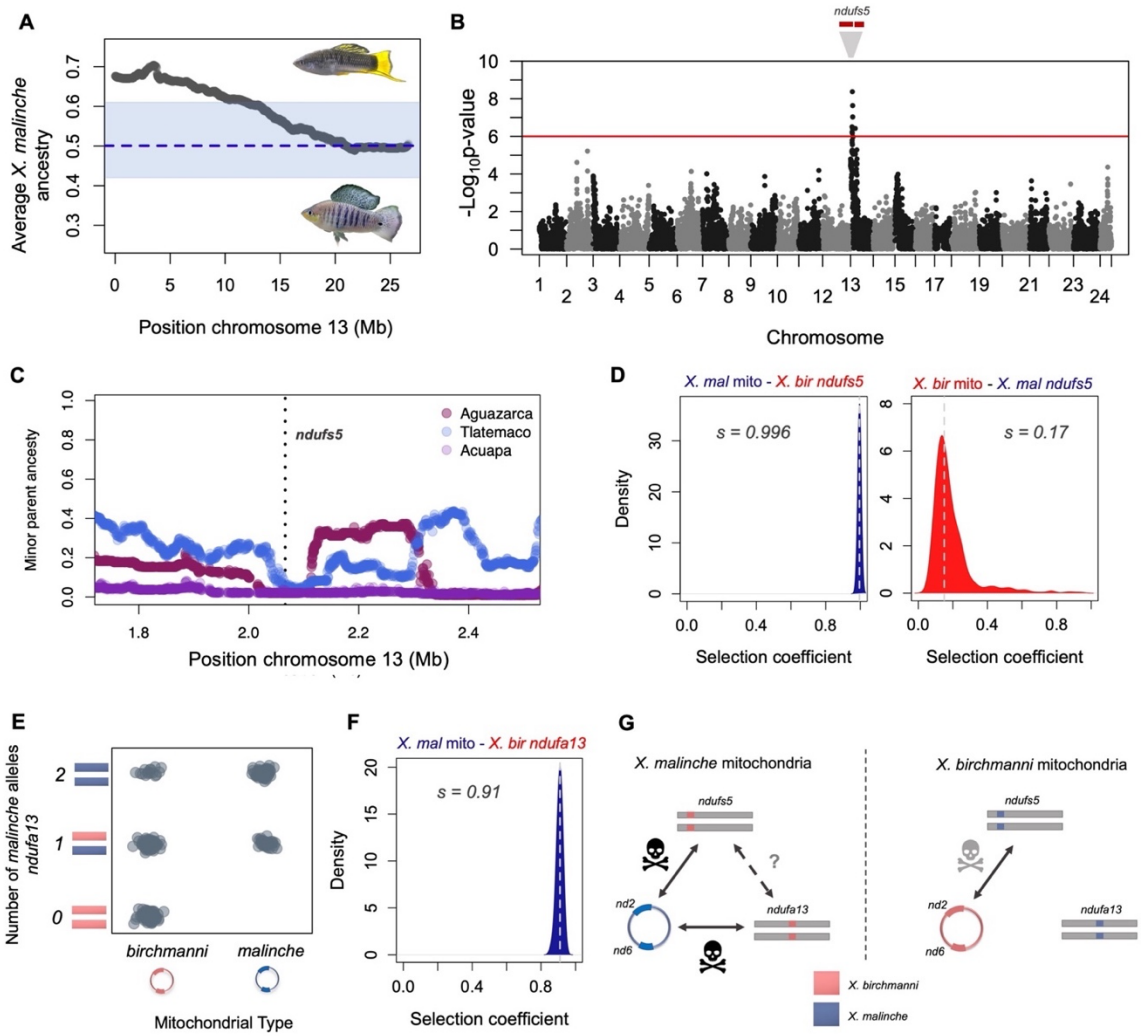

Figure 1. Admixture mapping pinpoints a mitonuclear incompatibility in *Xiphophorus*. (**A**) F_2 hybrid offspring of a laboratory cross between *X. malinche* (top) and *X. birchmanni* (bottom) have segregation distortion towards *X. malinche* ancestry across a large region of chromosome 13. Blue envelop shows the 99% quantiles of *X. malinche* ancestry at all ancestry informative sites genome wide. Dashed line shows the expected *X. malinche* ancestry for this cross. (**B**) Admixture mapping results for association between nuclear ancestry and mitochondrial haplotype in natural hybrids, controlling for genome-wide ancestry. Red line indicates genome-wide significance threshold determined by simulations. (**C**) Non-mitochondrial parent ancestry in natural hybrid populations fixed for *X. birchmanni* (Aguazarca, Acuapa) or *X. malinche* (Tlatemaco) mitochondrial haplotypes. Vertical dashed line represents the position of *ndufs5*, one of three genes in a shared ancestry desert at the center of the admixture mapping region. (**D**)

Results of approximate Bayesian computation (ABC) simulations estimating selection coefficients against heterospecific ancestry combinations at *ndufs5* and the *X. birchmanni* or *X. malinche* mitochondrial haplotypes. Shown here are posterior distributions from accepted simulations; vertical lines indicate the maximum a posteriori estimates for selection coefficient in each direction. (**E**) Observed genotype frequencies at *ndufa13* in the admixture mapping population. (**F**) Results of ABC simulations estimating the strength of selection against the combination of *X. malinche* mitochondria with *X. birchmanni* ancestry at *ndufa13*. Shown here is the posterior distribution of accepted simulations; the vertical line indicates the maximum a posteriori estimate for the selection coefficient. (**G**) Schematic of identified interactions with both the *X. malinche* and *X. birchmanni* mitochondrial genomes from our mapping data and strength of selection underlying each interaction in hybrids (gray – moderate, black – near lethal). We discuss evidence for possible interactions between *ndufs5* and *ndufa13* (indicated by the dashed line) in Supplementary Materials 1.1.8.

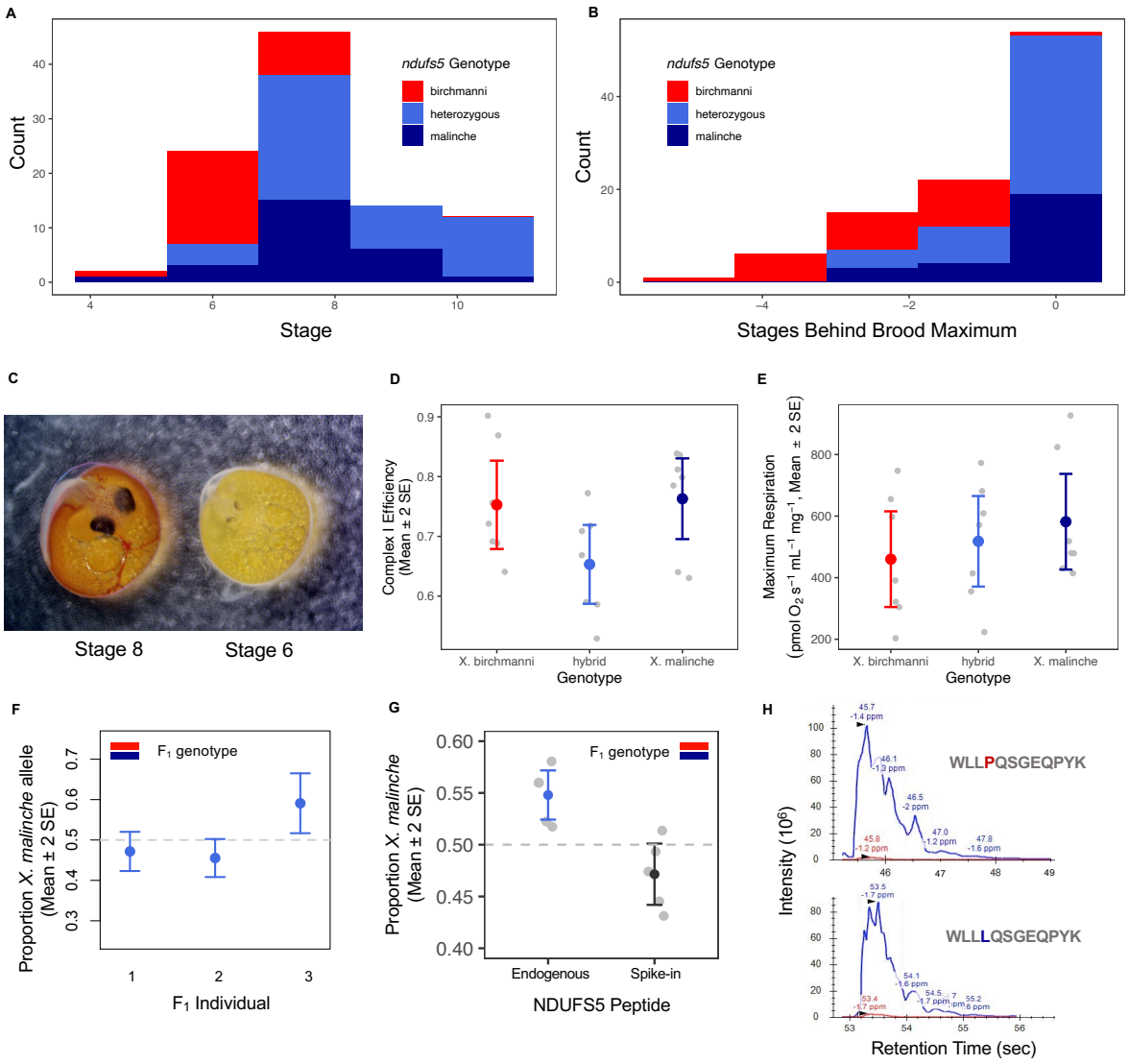

Figure 2. Developmental, physiological, and proteomic phenotypes associated with the hybrid incompatibility. (**A**) Developmental stage and *ndufs5* genotypes of hybrid embryos with *X. malinche* mitochondria. (**B**) Lag in development of hybrid embryos with *X. malinche* mitochondria compared to the most developed embryo in the brood as a function of *ndufs5* genotype. (**C**) Siblings from the admixture mapping population at different developmental stages. (**D**) Results of Oroboros O2K respirometer Complex I efficiency assay for adult *X. birchmanni*, *X. malinche*, and hybrid individuals with *X. malinche* mitochondria and heterozygous ancestry at *ndufs5* and *ndufa13* (n=7 per genotype). Colored points show the mean, whiskers show two standard errors, and gray points show individual data. (**E**) Despite differences in Complex I efficiency, maximum respiration rates do not differ between groups. (**F**) Allele-

specific expression of *ndufs5* in three adult F_1 hybrids. Points show the proportion of reads in F_1 hybrids supporting the *X. malinche* allele, bars show two standard errors. (**G**) Results of quantitative mass spectrometry analysis of *ndufs5* peptides in mitochondrial proteomes derived from five adult F_1 hybrids. Grey points show the proportion of area under the spectral curves contributed by the *X. malinche* allele in a given individual, colored points show the mean, and bars show two standard errors. The left column shows results for endogenous peptides present in F_1s, the right column shows results for the control where heavy-labeled standards of each peptide were spiked in. (**H**) Representative Skyline traces of the *X. malinche* and *X. birchmanni* versions of the *ndufs5* peptide from parallel reaction monitoring. Blue lines represent the spectral intensity from heavy-labeled peptides, red lines represent endogenous peptides. Inset shows the sequence of the *X. birchmanni* and *X. malinche* peptides being detected.

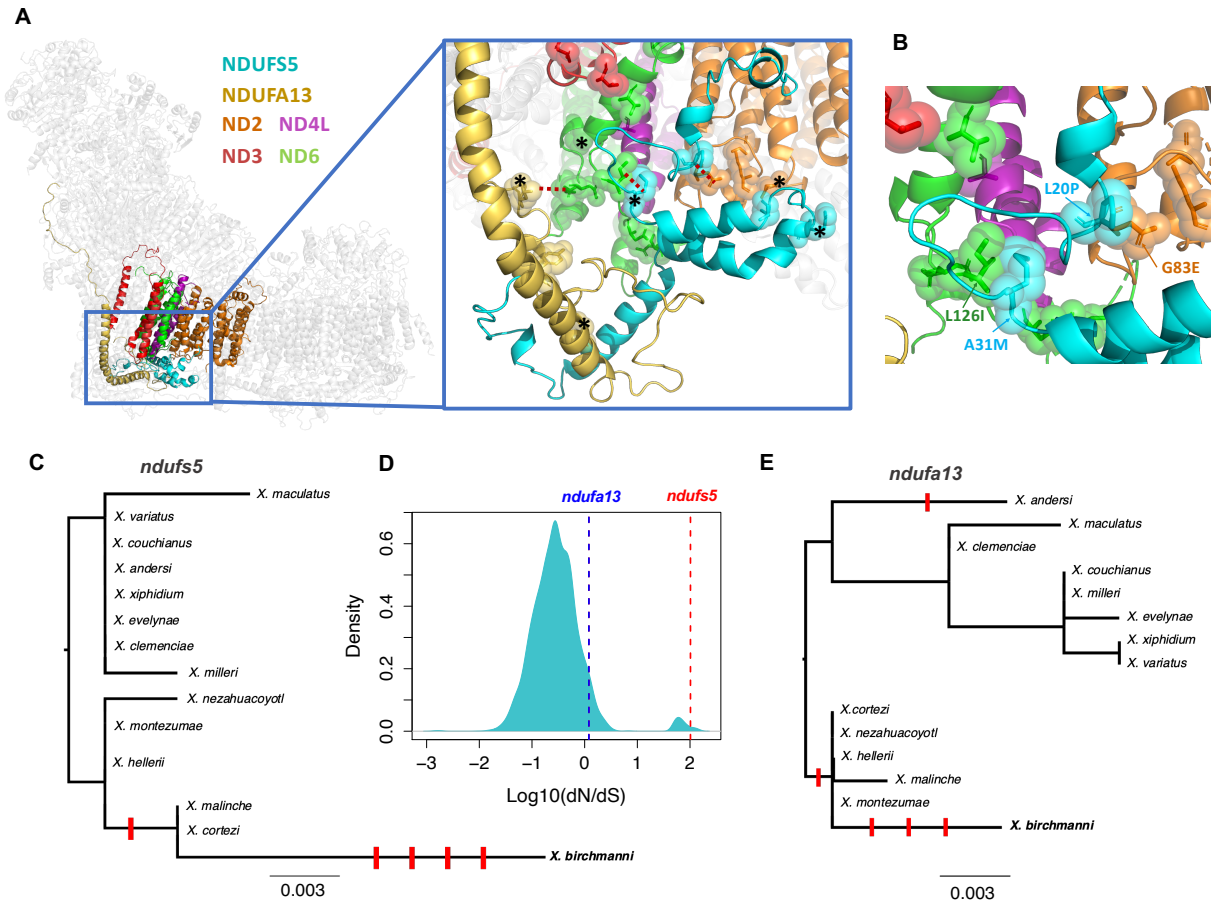

Figure 3. Predicted structures of *Xiphophorus* respiratory Complex I reveal interacting substitutions at protein interfaces. (**A**) *Xiphophorus* respiratory Complex I structures generated by RaptorX using alignment to a template mouse cryo-EM structure. Colored protein structures include *ndufs5, ndufa13*, and the four mitochondrially encoded *nd* genes in contact with *ndufs5* or *ndufa13*. Inset shows interaction surface between these genes. Asterisks denote residues with substitutions in *X. birchmanni* predicted to affect protein function (Table S4). Dotted red lines highlight substitutions in *ndufs5, ndufa13, nd2,* and *nd6* in close proximity. (**B**) Detailed view of interaction interface between *ndufs5*, *nd2*, and *nd6.* Arrows highlight substitutions in direct contact (represented as spheres), with letters denoting the *X. malinche* allele, the residue number, and the *X. birchmanni* allele, respectively. (**C**) Gene tree for *ndufs5* generated with RAxML, highlighting an excess of substitutions along the *X. birchmanni* branch. Scale bar represents number of nucleotide substitutions per site. Derived non-synonymous substitutions are indicated by red ticks along the phylogeny. Note that spacing between ticks

is arbitrary. (**D**) Distribution of Log_{10} dN/dS between *X. birchmanni* and *X. malinche* across all nuclear genes in the genome with values for *ndufs5* and *ndufa13* highlighted. (**E**) Gene tree for *ndufa13* generated with RAxML, highlighting an excess of substitutions along the *X. birchmanni* branch (as in C).

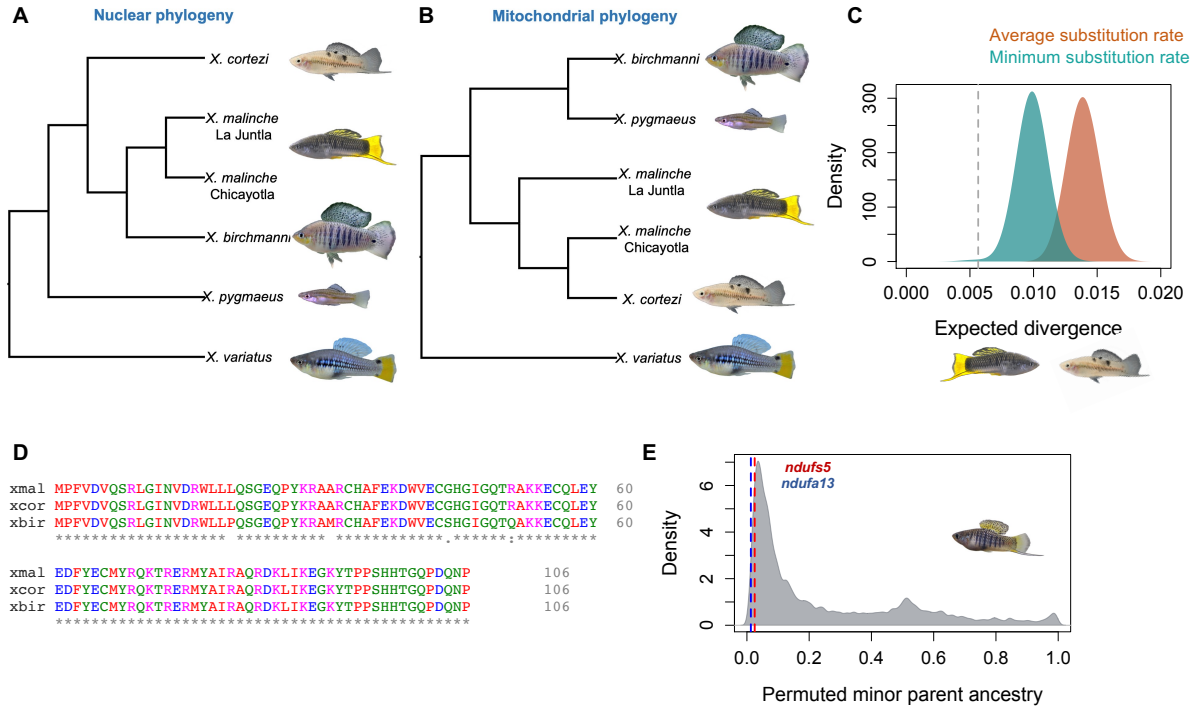

Figure 4. Phylogenetic analysis and ancestry mapping suggest that genes underlying the mitonuclear incompatibility have introgressed from *X. malinche* into *X. cortezi*. (**A**) Nuclear phylogeny of *Xiphophorus* species, showing that *X. birchmanni* and *X. malinche* are sister species (*49*). (**B**) Phylogeny constructed from whole mitochondrial genome sequences showing that *X. cortezi* mitochondria are nested within *X. malinche* mitochondrial diversity. (**C**) Results of simulations modeling expected mitochondrial divergence between *X. malinche* and *X. cortezi* in a scenario with no gene flow. Distributions represent pairwise sequence divergence in two sets of simulations. The first set used the average mitochondrial substitution rate observed between *Xiphophorus* species (red), and the second used the minimum mitochondrial substitution rate observed (blue). The dotted line shows observed divergence between mitochondrial haplotypes in *X. malinche* and *X. cortezi*, indicating that past mitochondrial introgression is more consistent with the observed data than incomplete lineage sorting (Supplementary Materials 1.5.3). (**D**) Clustal alignment of *ndufs5* sequences. *X. malinche* and *X. cortezi* have identical amino acid sequences at *ndufs5*, while *X. birchmanni* is separated from them by four substitutions. Colors indicate properties of the amino acid, asterisks indicate locations where the amino acid sequences are

identical. (**E**) Non-mitochondrial parent ancestry is lower than expected by chance in two natural *X. cortezi* × *X. birchmanni* hybrid populations fixed for the *X. cortezi* mitochondrial haplotype (Fig. S24) at *ndufs5* (red line) and *ndufa13* (blue line). Gray distribution shows permutations randomly drawing 0.1 centimorgan windows from the two *X. cortezi* × *X. birchmanni* hybrid populations. Inset shows a *X. cortezi* × *X. birchmanni* hybrid.

References

1. A. J. Dagilis, M. Kirkpatrick, D. I. Bolnick, The evolution of hybrid fitness during speciation. *PLOS Genetics*. **15**, e1008125 (2019).

2. Th. Dobzhansky, Genetic Nature of Species Differences. *The American Naturalist*. **71**, 404–420 (1937).

3. H. Müller, Isolating mechanisms, evolution, and temperature. *Biological Symposium*. **6**, 71–125 (1942).

4. S. Maheshwari, D. A. Barbash, The Genetics of Hybrid Incompatibilities. *Annual Review of Genetics*. **45**, 331–355 (2011).

5. D. C. Presgraves, The molecular evolutionary basis of species formation. *Nat Rev Genet*. **11**, 175–180 (2010).

6. K. B. S. Swamy, S. C. Schuyler, J.-Y. Leu, Protein Complexes Form a Basis for Complex Hybrid Incompatibility. *Front. Genet.* **12** (2021), doi:10.3389/fgene.2021.609766.

7. H. A. Orr, The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*. **139**, 1805–1813 (1995).

8. J. Boocock, M. J. Sadhu, A. Durvasula, J. S. Bloom, L. Kruglyak, Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast. *Science*. **371**, 415–419 (2021).

9. W.-H. Wei, G. Hemani, C. S. Haley, Detecting epistasis in human complex traits. *Nat Rev Genet*. **15**, 722–733 (2014).

10. W. J. Gauderman, Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol*. **155**, 478–484 (2002).

11. J. Boocock, M. J. Sadhu, A. Durvasula, J. S. Bloom, L. Kruglyak, Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast. *Science*. **371**, 415–419 (2021).

12. N. Phadnis, E. P. Baker, J. C. Cooper, K. A. Frizzell, E. Hsieh, A. F. A. de la Cruz, J. Shendure, J. O. Kitzman, H. S. Malik, An essential cell cycle regulation gene causes hybrid inviability in Drosophila. *Science*. **350**, 1552–1555 (2015).

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

14. J. C. Cooper, P. Guo, J. Bladen, N. Phadnis, A triple-hybrid cross reveals a new hybrid incompatibility locus between D. melanogaster and D. sechellia. *bioRxiv*, 590588 (2019).

15. N. A. Johnson, Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends in Genetics*. **26**, 317–325 (2010).

16. R. S. Burton, F. S. Barreto, A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Molecular Ecology*. **21**, 4942–4957 (2012).

17. G. E. Hill, Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecology and Evolution*. **6**, 5831–5842 (2016).

18. N. Lane, W. Martin, The energetics of genome complexity. *Nature*. **467**, 929–934 (2010).

19. C. M. Barr, M. Neiman, D. R. Taylor, Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist*. **168**, 39–50 (2005).

20. A. L. Case, F. R. Finseth, C. M. Barr, L. Fishman, Selfish evolution of cytonuclear hybrid incompatibility in Mimulus. *Proceedings of the Royal Society B: Biological Sciences*. **283**, 20161493 (2016).

21. C. D. Chase, Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions. *Trends in Genetics*. **23**, 81–90 (2007).

22. J. W. O. Ballard, M. C. Whitlock, The incomplete natural history of mitochondria. *Molecular Ecology*. **13**, 729–744 (2004).

23. C. Haag-Liautard, N. Coffey, D. Houle, M. Lynch, B. Charlesworth, P. D. Keightley, Direct Estimation of the Mitochondrial DNA Mutation Rate in Drosophila melanogaster. *PLOS Biology*. **6**, e204 (2008).

24. R. Allio, S. Donega, N. Galtier, B. Nabholz, Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker. *Molecular Biology and Evolution*. **34**, 2762–2772 (2017).

25. N. Osada, H. Akashi, Mitochondrial–Nuclear Interactions and Accelerated Compensatory Evolution: Evidence from the Primate Cytochrome c Oxidase Complex. *Mol Biol Evol*. **29**, 337–346 (2012).

26. F. S. Barreto, R. S. Burton, Evidence for Compensatory Evolution of Ribosomal Proteins in Response to Rapid Divergence of Mitochondrial rRNA. *Molecular Biology and Evolution*. **30**, 310–314 (2013).

27. D. I. Bolnick, M. Turelli, H. López-Fernández, P. C. Wainwright, T. J. Near, Accelerated Mitochondrial Evolution and "Darwin's Corollary": Asymmetric Viability of Reciprocal F1 Hybrids in Centrarchid Fishes. *Genetics*. **178**, 1037–1048 (2008).

28. M. Turelli, L. C. Moyle, Asymmetric Postmating Isolation: Darwin's Corollary to Haldane's Rule. *Genetics*. **176**, 1059–1088 (2007).

29. Y. Brandvain, G. B. Pauly, M. R. May, M. Turelli, Explaining Darwin's Corollary to Haldane's Rule: The Role of Mitonuclear Interactions in Asymmetric Postzygotic Isolation Among Toads. *Genetics*. **197**, 743–747 (2014).

30. P. Tiffin, S. Olson, L. C. Moyle, Asymmetrical crossing barriers in angiosperms. *Proc. R. Soc. Lond. B*. **268**, 861–867 (2001).

31. C. D. Meiklejohn, M. A. Holmbeck, M. A. Siddiq, D. N. Abt, D. M. Rand, K. L. Montooth, An Incompatibility between a Mitochondrial tRNA and Its Nuclear-Encoded tRNA Synthetase Compromises Development and Fitness in Drosophila. *PLOS Genetics*. **9**, e1003238 (2013).

32. D. Luo, H. Xu, Z. Liu, J. Guo, H. Li, L. Chen, C. Fang, Q. Zhang, M. Bai, N. Yao, H. Wu, H. Wu, C. Ji, H. Zheng, Y. Chen, S. Ye, X. Li, X. Zhao, R. Li, Y.-G. Liu, A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nature Genetics*. **45**, 573–577 (2013).

33. H.-Y. Lee, J.-Y. Chou, L. Cheong, N.-H. Chang, S.-Y. Yang, J.-Y. Leu, Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility between Two Yeast Species. *Cell*. **135**, 1065–1073 (2008).

34. M. R. Hanson, S. Bentolila, Interactions of Mitochondrial and Nuclear Genes That Affect Male Gametophyte Development. *The Plant Cell*. **16**, S154–S169 (2004).

35. Z. W. Culumber, H. S. Fisher, M. Tobler, M. Mateos, P. H. Barber, M. D. Sorenson, G. G. Rosenthal, Replicated hybrid zones of *Xiphophorus* swordtails along an elevational gradient. *Molecular Ecology*. **20**, 342–356 (2011).

36. H. S. Fisher, B. B. M. Wong, G. G. Rosenthal, Alteration of the chemical environment disrupts communication in a freshwater fish. *Proc. R. Soc. B*. **273**, 1187–1193 (2006).

37. M. Schumer, C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S. Sankararaman, P. Andolfatto, G. G. Rosenthal, M. Przeworski, Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*. **360**, 656–660 (2018).

38. M. Schumer, Y. Brandvain, Determining epistatic selection in admixed populations. *Mol Ecol*. **25**, 2577–2591 (2016).

39. M. Schumer, R. Cui, D. L. Powell, R. Dresner, G. G. Rosenthal, P. Andolfatto, High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *eLife*. **3**, e02535 (2014).

40. M. Schumer, D. L. Powell, R. Corbett-Detig, "Versatile simulations of admixture and accurate local ancestry inference with *mixnmatch* and *ancestryinfer*" (preprint, Genomics, 2019), , doi:10.1101/860924.

41. J. L. Haynes, Standardized Classification of Poeciliid Development for Life-History Studies. *Copeia*. **1995**, 147 (1995).

42. W. N. Tavolga, Embryonic development of the platyfish (Platypoecilus), the swordtail (Xiphophorus), and their hybrids. Bulletin of the AMNH ; v. 94, article 4. *Embryonic development in fish* (1949) (available at http://digitallibrary.amnh.org/handle/2246/1000).

43. A.-N. A. Agip, J. N. Blaza, H. R. Bridges, C. Viscomi, S. Rawson, S. P. Muench, J. Hirst, Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nature Structural & Molecular Biology*. **25**, 548–556 (2018).

44. D. Kampjut, L. A. Sazanov, The coupling mechanism of mammalian respiratory complex I. *Science* (2020), doi:10.1126/science.abc4209.

45. K. Fiedorczuk, J. A. Letts, G. Degliesposti, K. Kaszuba, M. Skehel, L. A. Sazanov, Atomic structure of the entire mammalian mitochondrial complex I. *Nature*. **538**, 406–410 (2016).

46. K. L. Mack, M. W. Nachman, Gene Regulation and Speciation. *Trends in Genetics*. **33**, 68–80 (2017).

47. M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, J. Xu, Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*. **7**, 1511–1522 (2012).

48. G. E. Hill, Mitonuclear Compensatory Coevolution. *Trends in Genetics* (2020), doi:10.1016/j.tig.2020.03.002.

49. R. Cui, M. Schumer, K. Kruesi, R. Walter, P. Andolfatto, G. G. Rosenthal, Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution*. **67**, 2166–2179 (2013).

50. D. B. Sloan, J. C. Havird, J. Sharbrough, The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology*. **26**, 2212–2236 (2017).

51. Q. K. Langdon, D. L. Powell, B. Kim, S. M. Banerjee, C. Payne, T. O. Dodge, B. Moran, P. Fascinetto-Zago, M. Schumer, *bioRxiv*, in press, doi:10.1101/2021.07.06.451368.

52. C. K. Ellison, R. S. Burton, Disruption of Mitochondrial Function in Interpopulation Hybrids of Tigriopus Californicus. *Evolution*. **60**, 1382–1391 (2006).

53. C. K. Ellison, O. Niehuis, J. Gadau, Hybrid breakdown and mitochondrial dysfunction in hybrids of Nasonia parasitoid wasps. *Journal of Evolutionary Biology*. **21**, 1844–1851 (2008).

54. J. R. Olson, S. J. Cooper, D. L. Swanson, M. J. Braun, J. B. Williams, The Relationship of Metabolic Performance and Distribution in Black-Capped and Carolina Chickadees. *Physiological and Biochemical Zoology*. **83**, 263–275 (2010).

55. D. L. Powell, M. García-Olazábal, M. Keegan, P. Reilly, K. Du, A. P. Díaz-Loyo, S. Banerjee, D. Blakkan, D. Reich, P. Andolfatto, G. G. Rosenthal, M. Schartl, M. Schumer, Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science*. **368**, 731–736 (2020).

56. Y. Lu, A. Sandoval, S. Voss, Z. Lai, S. Kneitz, W. Boswell, M. Boswell, M. Savage, C. Walter, W. Warren, M. Schartl, R. Walter, Oncogenic allelic interaction in Xiphophorus highlights hybrid incompatibility. *PNAS*. **117**, 29786–29794 (2020).

57. T. C. Nelson, A. M. Stathos, D. D. Vanderpool, F. R. Finseth, Y. Yuan, L. Fishman, Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (Mimulus section Erythranthe). *PLOS Genetics*. **17**, e1009095 (2021).

Acknowledgements

We thank Peter Andolfatto, Stepfanie Aguillon, Yaniv Brandvain, Jenn Coughlan, Hunter Fraser, Yuki Haba, Molly Przeworski, Ken Thompson and members of the Schumer lab for helpful discussion and/or feedback on earlier versions of this manuscript. We thank the Federal Government of Mexico for permission to collect fish. Stanford University and the Stanford Research Computing Center provided computational support for this project. We thank the Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University Mass Spectrometry (RRID:SCR_017801) for technical and experimental support. **Funding:** This work was supported by a Knight-Hennessy Scholars fellowship and NSF GRFP 2019273798 to B. Moran, a CEHG fellowship and NSF PRFB (2010950) to Q. Langdon, NIH P30 CA124435 in utilizing the Stanford Cancer Institute Proteomics/Mass Spectrometry Shared Resource, NIH grant 1R35GM142836 to J. Havird, and a Hanna H. Gray fellowship, Sloan Fellowship, and NIH grant 1R35GM133774 to M. Schumer. **Author Contributions:** B.M.M., D.L.P, and M. Schu, designed the project; B.M.M., C.Y.P., E.N.K.I., S.M.B, F.L., R.M., K.S., O.H.-P., J.C.H., and M.Schu. collected data; B.M.M., C.Y.P., Q.L.K., F.L., J.C.H., and M. Schu. performed analyses; D.L.P., T.R.G., R.D.L., R.C.-D., and M. Scha. provided expertise and technical support. **Competing Interests:** The authors declare no competing interests. **Data and materials availability:** All data are available through NCBI (see Supplementary Materials) and Dryad (pending). Code is available on Github at https://github.com/Schumerlab/mitonuc_DMI.

Supplementary Materials

Materials and Methods

Tables S1 – S11

Figs. S1 – S27