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

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], Angel Madero[1], Theresa R. Gunn[1], Quinn K. Langdon[1], Fang Liu[4], Rowan Matney[4], Kratika Singhal[4], Ryan D. Leib[4], Osvaldo Hernandez-Perez[2], Russell Corbett-Detig[5,6], Judith Frydman[1,7], Manfred Schartl[8,9], Justin C. Havird[3], Molly Schumer[1,2,10]*

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]Department of Genetics, Stanford University, Stanford, CA, USA

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

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

[10]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 the first step in 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 hybrid incompatibilities involving multiple genes should be common and that rapidly evolving genes will be more likely to cause incompatibilities, but there has been sparse empirical data to evaluate these predictions. Here, we describe a mitonuclear incompatibility involving three genes within respiratory Complex I in naturally hybridizing swordtail fish species. 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 document the evolutionary history of the genes involved and localize the protein-protein interactions most likely to contribute to the incompatibility. This work thus provides the first glimpse into the genetic architecture, physiological impacts, and evolutionary origin of a multi-gene incompatibility impacting naturally hybridizing species.

Introduction

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 dozen 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].

This knowledge gap leaves key predictions about the genetic architecture of hybrid incompatibilities and the evolutionary processes that drive their emergence 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[8,11–13]). Another largely untested prediction is that rapid molecular evolution will increase the rate at which incompatibilities accumulate between species[4,5,14]. 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[5,14].

Another open question is the degree to which the genes that become involved in hybrid incompatibilities are predictable from their molecular or evolutionary properties. The mitochondrial genome, in particular, has been proposed as a hotspot for the accumulation of genetic incompatibilities[15,16]. Mitochondria are essential for energy production in nearly all eukaryotic organisms[17]. 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 some fungi[18], creating the potential for sexually antagonistic selection[19,20]. 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 ~25× higher mitochondrial substitution rates in some species[21–23]. 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[24,25]. 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[15].

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[26–29]. 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[30–33] and none of those identified to date have been studied in species that naturally hybridize.

As we begin to identify 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[34] after premating barriers were disrupted by habitat disturbance[35], and are a powerful system to study the emergence of hybrid incompatibilities in young species. Despite their recent divergence[36] (~250,000 generations; 0.5% divergence per basepair), some hybrids between *X. birchmanni* and *X. malinche* experience strong selection against incompatibilities[36,37]. One incompatibility causing hybrid melanoma has been previously mapped in this system and population genetic patterns suggest that dozens may be segregating in natural hybrid populations[36–39]. Moreover, the ability to generate controlled crosses[40,41] and the development of high-quality genomic resources[39,42] makes this system particularly tractable for identifying hybrid incompatibilities that impact natural populations and characterizing their evolution. Leveraging data from controlled laboratory crosses and natural hybrid zones, we pinpoint two *X. birchmanni* genes that are lethal when mismatched with the *X. malinche* mitochondria in hybrids, explore the developmental and physiological effects of this incompatibility, 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 an F_2 laboratory cross and 359 wild-caught hybrid adults, and applied a hidden Markov model to data at 629,661 ancestry-informative sites along the genome in order to infer local ancestry (~1 informative site per kb[38,43]; Methods, Supplementary Information 1.1.1-1.1.4). Using these results, we found evidence for a previously unknown incompatibility between the nuclear genome of *X. birchmanni* and the mitochondrial genome of *X. malinche* (Supplementary Information 1.1.5-1.1.11). Our first direct evidence for this incompatibility came from controlled laboratory crosses (Methods, Supplementary Information 1.1.1). 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 six segregation distorters genome-wide[41], with the most extreme signals falling along a 6.5 Mb block of chromosome 13 and an 4.9 Mb block of chromosome 6 (Fig. 1A; Fig. 1D).

Closer examination of genotypes in the chromosome 13 region showed that almost none of the surviving individuals harbored homozygous *X. birchmanni* ancestry in a 3.75 Mb subregion (Fig. 1C; Fig. S2; 0.1% observed vs 25% expected). This pattern is unexpected even in the case of a lethal nuclear-nuclear incompatibility (where simulations indicate that we should

recover homozygous *X. birchmanni* ancestry in ~10% of surviving individuals; Supplementary Information 1.1.1), 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 inferred that 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. 1B; Fig. S3; Methods; Supplementary Information 1.2.1-1.2.2).

The degree of segregation distortion observed in F_2 individuals on chromosome 6 is also surprising (Fig. 1D). Only 3% of individuals harbor homozygous *X. birchmanni* ancestry in this region (compared to 0.1% on chromosome 13 and 25% on average at other loci across the genome; Fig 1F). The frequency of homozygous *X. birchmanni* ancestry at the center of chromosome 6 is lower than expected in the absence of selection and lower than expected for a nuclear-nuclear hybrid incompatibility (Supplementary Information 1.1.1). ABC approaches indicate that selection on homozygous *X. birchmanni* ancestry on chromosome 6 is also severe (maximum a posteriori estimate $s = 0.91$, 95% credible interval 0.87-0.94; $h = 0.09$, 95% credible interval 0.01-0.21; Fig. 1E, Fig. S3, Supplementary Information 1.2.2). Thus, our F_2 data indicate that homozygous *X. birchmanni* ancestry at either chromosome 13 or chromosome 6 is nearly lethal in hybrids with *X. malinche* mitochondria (Fig. 1H).

To formally test for the presence of a mitonuclear incompatibility involving chromosome 13 and chromosome 6, or elsewhere in the genome, 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 Information 1.1.2, 1.1.6).

156 However, a few segregate for the mitochondrial genomes of both parental types, and we focused
157 on one such population (the "Calnali Low" population, hereafter the admixture mapping
158 population). Admixture mapping for associations between nuclear genotype and mitochondrial
159 ancestry (after adjusting for expected covariance due to genome-wide ancestry[37]) revealed two
160 genome-wide significant peaks and one peak that approached genome-wide significance (Fig.
161 1G, Table S1-S3). The strongest peak of association spanned approximately 77 kb and fell within
162 the region of chromosome 13 identified using F_2 crosses (Fig. 1G). This peak was also replicated
163 in another hybrid population (Fig. S4; Methods, Supplementary Information 1.1.5) and contains
164 only three genes: the NADH dehydrogenase ubiquinone iron-sulfur protein 5 (*ndufs5*), E3
165 ubiquitin-protein ligase, and microtubule-actin cross-linking factor 1. Of these three genes,
166 *ndufs5* forms a protein complex with mitochondrially encoded proteins, which along with other
167 evidence implicates it as one of the nuclear components of the mitonuclear incompatibility (Fig.
168 1C; see Supplementary Information 1.1.8 for analysis of other genes).
169 Analysis of three natural hybrid populations that had fixed the mitochondrial haplotype of
170 one of the parental species (Fig. S5) confirmed that this region on chromosome 13 is under
171 selection in natural hybrid populations, with the strongest signal of selection localizing precisely
172 to the same three genes found under the admixture mapping peak (Fig. S6A; Supplementary
173 Information 1.1.6). Moreover, comparing genotypes and phenotypes in siblings allowed us to
174 exclude maternal effects as a driver of the chromosome 13 signal (Supplementary Information
175 1.1.7), and we ruled out the possibility that other confounding factors could generate the
176 observed patterns (Supplementary Information 1.1.9).
177 We also identified a peak on chromosome 6 that approached genome-wide significance
178 (Fig. 1G; Table S2; Supplementary Information 1.1.10) and fell precisely within the segregation

distortion region previously mapped in F_2 hybrids (Fig. 1D; Supplementary Information 1.1.1). This peak contained 20 genes including the mitochondrial Complex I gene *ndufa13* (Fig. S7-S8; Methods, Supplementary Information 1.1.10). Depletion of non-mitochondrial parent ancestry at *ndufa13* was unidirectional (Fig. 1F), consistent with selection acting only against the combination of the *X. malinche* mitochondria with homozygous *X. birchmanni* ancestry at *ndufa13* (see Supplementary Information 1.2.3-1.2.4). Genomic analyses in natural hybrid populations reflect this asymmetry, with ancestry at *ndufa13* fixed in populations with the *X. malinche* mitochondrial haplotype and segregating in populations with the *X. birchmanni* mitochondrial haplotype (Fig. S6B).

Together, these results indicate that at least two *X. birchmanni* nuclear genes are incompatible with the *X. malinche* mitochondria (Fig. 1H), although we discuss uncertainty about the exact architecture of the interaction in Supplementary Information 1.1.11. These genes, *ndufs5* and *ndufa13*, belong to a group of proteins and assembly factors that form respiratory Complex I[44] (see Table S1 for locations of the 51 annotated Complex I genes in the *Xiphophorus* genome). Complex I is the first component of the mitochondrial electron transport chain that ultimately allows the cell to generate ATP. Both nuclear proteins interface with several mitochondrially derived proteins at the core of the Complex I structure, pointing to the possibility that physical interactions underlie this multi-gene mitonuclear incompatibility.

Interactions with the *X. birchmanni* mitochondria

Admixture mapping analysis also identified a strong peak of mitonuclear association on chromosome 15, which we briefly discuss here and in Supplementary Information 1.1.10 and

1.2.1. This peak was associated with *X. birchmanni* mitochondrial ancestry (Fig. S9), indicating that it has a distinct genetic architecture from the incompatibility involving the *X. malinche* mitochondria and *X. birchmanni ndufs5* and *ndufa13*. Specifically, analysis of genotypes at the admixture mapping peak indicates that the *X. birchmanni* mitochondria is incompatible with homozygous *X. malinche* ancestry on chromosome 15 (Fig. S9). This region did not contain any members of Complex I, but dozens of genes in this interval interact with known mitonuclear genes (see Table S3; Supplementary Information 1.1.10), presenting an exciting direction for future work. The fact that we detect incompatible interactions with both the *X. malinche* and *X. birchmanni* mitochondria in our admixture mapping results underscores the importance of mitonuclear interactions as "hotspots" for the evolution of hybrid incompatibilities[15].

Lethal Effect of Incompatibility in Early Development

The incompatibility involving the *X. malinche* mitochondria appears to be lethal by the time individuals reach adulthood. To investigate the developmental timing of the incompatibility, we genotyped pregnant females from the admixture mapping population and recorded the developmental stages of their embryos[45] (swordtails are livebearing fish; Methods). We focused on the 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. S10-13; Supplementary Information 1.3.1). While developmental asynchrony is typically on the scale of 0-2 days in pure species[46] (Supplementary Information 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 Information 1.3.1; Fig. 2A-B). Genotyping results revealed that embryos with homozygous *X. birchmanni* ancestry at *ndufs5* and *X. malinche* mitochondria are present 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 21-28 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-D).

In contrast to other species, in *Xiphophorus* this developmental lag could itself cause mortality, since embryos that do not complete embryonic development within the mother fail to survive more than a few days after birth (Supplementary Information 1.3.1). Moreover, *Xiphophorus* fry appear to be more sensitive to Complex I inhibition than zebrafish. Pharmacological inhibition of Complex I in newborn *Xiphophorus* fry caused nearly 100% lethality over 24 hours at concentrations that are not lethal to zebrafish fry on the same timescale[47,48] (Supplementary Information 1.3.2). Notably, zebrafish larvae with Complex I inhibition also exhibit delayed or arrested development[47,48].

Mitochondrial Biology in Viable Hybrids Heterozygous for the Incompatibility

Our analysis of developing embryos indicates that individuals with the mitonuclear incompatibility exhibit delayed or arrested embryonic development. While this developmental delay phenotype may itself be sufficient to cause lethality, since premature birth is almost always lethal in both parental *Xiphophorus* species and in hybrids (Table S4; Supplementary Information 1.3.1), we were curious to investigate physiological impacts of the mitonuclear

incompatibility further. However, because individuals homozygous for the incompatibility involving the *X. malinche* mitochondria generally do not complete embryonic development, it is difficult to evaluate this question directly.

To begin to explore effects of the hybrid incompatibility on Complex I function *in vivo*, albeit in a form that does not impact viability, we turned to F_1 hybrids between *X. birchmanni* and *X. malinche* (Fig. 3A). Since F_1 hybrids that derive their mitochondria from *X. malinche* and are heterozygous for ancestry at *ndufs5* and *ndufa13* are viable, we asked whether there was evidence for compensatory nuclear or mitochondrial regulation that might be protective in F_1 hybrids. We found no evidence for significant differences in expression of *ndufs5* or *ndufa13* (Supplementary Information 1.3.3; Fig. 3F, Fig. S14-15) or in mitochondrial copy number (Supplementary Information 1.3.4; Fig. S16) in F_1 hybrids.

With no clear indication of a compensatory regulatory response, we reasoned that we might be able to detect reduced mitochondrial Complex I function in hybrids heterozygous for the incompatibility. To examine mitochondrial function in *X. birchmanni*, *X. malinche*, and hybrids harboring the *X. malinche* mitochondria and heterozygous ancestry at *ndufs5* and *ndufa13,* we quantified respiratory phenotypes in isolated mitochondria using a multiple substrate, uncoupler, and inhibitor titration protocol with the Oroboros O2K respirometer (Fig. S17; Methods, Supplementary Information 1.3.5). We found that Complex I efficiency was somewhat lower in hybrids compared to the two parental species (Fig. 3B, Fig. S18, orthogonal contrast $t = -2.53$, $P = 0.023$, n = 7 per genotype), although overall levels of mitochondrial respiration were unchanged (Fig. 3C, orthogonal contrast $t = 0.078$, $P = 0.94$, n = 7 per genotype; Supplementary Information 1.3.5). While Complex I efficiency can also be affected by the integrity of the mitochondrial membrane, neither measurement of LEAK state respiration (Fig.

S18, t = -1.213, P = 0.24, n = 7; Supplementary Information 1.3.5) nor flow cytometry assays (Fig. S19; Supplementary Information 1.3.6) showed differences in mitochondrial membrane integrity between genotypes, pointing to reduced function of Complex I in hybrids. More dramatic than differences in absolute Complex I efficiency was the time required for hybrids to reach maximum Complex I-driven respiration, which was substantially longer than in either of the parental species (orthogonal contrast t = 4.303, $P < 0.001$; Fig. 3D; Fig. S20). In contrast, time to peak respiration after activation of Complex II was similar across genotypes (orthogonal contrast t = -0.705, P = 0.49; Fig. 3E). Together, these data point to reduced function of Complex I in heterozygous individuals, as well as possible physiological compensation by other components in the respiratory pathway.

Given physiological evidence for some reduction in Complex I function in hybrids heterozygous at *ndufs5* and *ndufa13*, we predicted that there might be an altered frequency of protein complexes incorporating both *X. malinche* mitochondrial proteins and *X. birchmanni* proteins at *ndufs5* and *ndufa13* in F_1 hybrids. To test this prediction, we took a mass spectrometry based quantitative proteomics approach. We used stable isotope-labeled peptides to distinguish between the *X. birchmanni* and *X. malinche ndufs5* and *ndufa13* peptides in mitochondrial proteomes extracted from F_1 hybrids (see Methods, Supplementary Information 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. 3G; Fig. S21; Supplementary Information 1.4.5). Since we did not observe allele-specific expression of *ndufs5* (Fig 3F; Supplementary Information 1.3.3), this result is consistent with disproportionate degradation of *X. birchmanni*-derived *ndufs5* peptides

in the mitochondrial proteome or differences in translation of *ndufs5* transcripts derived from the two species.

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

While we can leverage high resolution admixture mapping 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[49–51]. Although these structures are only available for distant relatives of swordtails, the presence of the same set of supernumerary Complex I subunits and high sequence similarity suggest that using these structures is appropriate (Table S5-S6; Fig. S22-S24; Supplementary Information 1.4.6).

Barring a hybrid incompatibility generated by regulatory divergence (see Supplementary Information 1.3.3), we expect hybrid incompatibilities to be driven by amino acid changes in interacting proteins[52]. We used the program RaptorX[53] 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[49] (Fig. 4A; Fig. S22-S24; Methods). Using these structures, we visualized amino acid substitutions between *X. birchmanni* and *X. malinche* at the interfaces of *ndufs5, ndufa13* and mitochondrially encoded genes (Fig. S25-S27). While there are dozens of substitutions in the four mitochondrially encoded genes that are in close physical proximity to *ndufs5* or *ndufa13* (Fig. S22; *nd2, nd3, nd4l,* and *nd6*), there are only five cases where amino acid substitutions in either nuclear gene are predicted to be close

317 enough to contact substitutions in any mitochondrial gene (Fig. 4A; see Fig. S26 for pairwise
318 visualizations of interacting proteins). One *ndufs5* substitution directly contacts a substitution in
319 *nd2* (Fig. 4B, Fig. S26), whereas another *ndufs5* substitution contacts a cluster of four *nd6*
320 substitutions (Fig. 4B, Fig. S26, Supplementary Information 1.4.6, Table S7). The same *nd6*
321 cluster is potentially in contact with one *ndufa13* substitution, and two other pairs of
322 substitutions, in *ndufa13* and *nd6* and *ndufs5* and *nd6* respectively, may also be close enough to
323 physically interact (Fig 4A, Fig. S26-27). While there is some uncertainty in the structure of one
324 region of *nd6* (Supplementary Information 1.4.6, Table S7), we found that the contact points
325 between *nd2*, *nd6, ndufs5*, and *ndufa13* were otherwise robust to a number of technical variables
326 (Supplementary Information 1.4.6). Based on these results, we predict that *nd2* and *nd6* are the
327 genes most likely to be involved in the mitochondrial component of the hybrid incompatibility
328 (Fig. 4A, 4B; Fig. S27-S30).
329
330 **Rapid evolution of Complex I proteins**
331
332 Theory predicts that hybrid incompatibilities are more likely to arise in rapidly evolving
333 genes[4,5,14]. Consistent with this hypothesis, *ndufs5* is among the most rapidly evolving genes
334 genome-wide between *X. birchmanni* and *X. malinche* (Fig. 4C, 4D). Aligning the *ndufs5* coding
335 sequences of *X. birchmanni, X. malinche*, and twelve other swordtail species revealed that all
336 four amino acid substitutions that differentiate *X. birchmanni* and *X. malinche* at *ndufs5* were
337 derived on the *X. birchmanni* branch (Fig. 4C). Phylogenetic tests indicate that there has been
338 accelerated evolution of *ndufs5* on this branch (dN/dS > 99, N = 4, S = 0, codeml branch test P =
339 0.005, Fig. 4C). Similar patterns of rapid evolution are observed at *ndufa13*, which also showed

evidence for accelerated evolution in *X. birchmanni* (Fig. 4E; dN/dS = 1.2, N = 3, S = 1, codeml branch test $P = 0.002$). While explicit tests for adaptive evolution at *ndufs5* and *ndufa13* could not exclude a scenario of relaxed selection (Supplementary Information 1.5.1, 1.5.2), our comparisons across phylogenetic scales highlight strong conservation in some regions of the proteins and rapid turnover in others, complicating our interpretation of this test (Fig. S31).

Rapid evolution of *ndufs5* and *ndufa13* could be driven by coevolution with mitochondrial substitutions, a mechanism that has been proposed to explain the outsized role of the mitochondria in hybrid incompatibilities[15,54]. Indeed, there is an excess of derived substitutions in the *X. birchmanni* mitochondrial protein *nd6*, one of the proteins that physically contacts *ndufs5* and *ndufa13* (Table S8; Fig. S28; codeml branch test $P = 0.005$). Moreover, a number of the substitutions observed in both mitochondrial and nuclear genes are predicted to have functional consequences (based on SIFT analysis, Supplementary Information 1.5.1; Table S9), including ones likely to be in contact between *ndufs5, ndufa13, nd2*, and *nd6* (Fig. 4A, 4B; Fig. S27). Such colocalization of substitutions predicted to affect protein function 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[55]. 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[55] (Fig. 5A; 5B). As we show, all *X. cortezi* mitochondria sequenced to

363 date are nested within *X. malinche* mitochondrial diversity (Fig. 5B; Supplementary Information
364 1.5.3-1.5.4), including the likely mitochondrial partners *nd2* and *nd6* (Fig. S32). Simulations
365 indicate that gene flow, rather than incomplete lineage sorting, drove replacement of the *X.*
366 *cortezi* mitochondria with the *X. malinche* sequence ($P < 0.002$ by simulation; Fig. 5C;
367 Supplementary Information 1.5.4).

368 The introgression of the mitochondrial genome from *X. malinche* into *X. cortezi* raises the
369 possibility that other Complex I genes may have co-introgressed[56]. Indeed, the nucleotide
370 sequence for *ndufs5* is identical between *X. malinche* and *X. cortezi*, and the sequence of *ndufa13*
371 differs by a single synonymous mutation (although conservation of both genes is high throughout
372 *Xiphophorus*; Fig. S33-S34). Identical amino acid sequences at the genes underlying the hybrid
373 incompatibility between *X. malinche* and *X. birchmanni* suggest that *X. cortezi* and *X.*
374 *birchmanni* are likely to harbor the same mitonuclear incompatibility, as a result of ancient
375 introgression between *X. malinche* and *X. cortezi* (Fig. 5D; Supplementary Information 1.5.3-
376 1.5.5). This inference is supported by analysis of ancestry in two contemporary *X. birchmanni* ×
377 *X. cortezi* hybrid populations[41], which reveals a significant depletion of non-mitochondrial parent
378 ancestry at *ndufs5* and *ndufa13* (Fig. 5E; Fig. S35; Supplementary Information 1.5.6). These
379 results are consistent with the mitonuclear incompatibility observed in *X. birchmanni* × *X.*
380 *malinche* being active in *X. birchmanni* × *X. cortezi* populations (see also [41]). This exciting
381 finding hints that genes underlying hybrid incompatibilities can introgress together, transferring
382 incompatibilities between related species.
383
384

Discussion

What genetic and evolutionary forces drive the emergence of hybrid incompatibilities, especially between closely related species? Theory predicts that hybrid incompatibilities involving multiple genes should be common[6,7], but with few exceptions[8,11–13], they remain virtually uncharacterized at the genic level[6]. Here, we identify a mitonuclear incompatibility that involves at least three genes and causes hybrid lethality in lab and wild populations. The mitonuclear incompatibility we map in naturally hybridizing species echoes predictions from theory and studies in lab models[8,11–13] that protein complexes may be a critical site of hybrid breakdown.

Researchers have proposed mitonuclear interactions as "hotspots" for the emergence of hybrid incompatibilities, given that mitochondrial genomes often experience higher substitution rates between species[21–23], yet must intimately interact with nuclear proteins to perform essential cellular functions[24,25]. Our findings support this prediction, identifying incompatible interactions with both the *X. malinche* and *X. birchmanni* mitochondria. We also show that there has been exceptionally rapid evolution in both mitochondrial and interacting nuclear genes in *X. birchmanni*, which may have introduced mutations that are incompatible in hybrids (Fig. 4). Whether driven by adaptation or some other mechanism, our findings support the hypothesis that the coevolution of mitochondrial and nuclear genes could drive the overrepresentation of mitonuclear interactions in hybrid incompatibilities[24,25,54]. More broadly, our results are consistent with predictions that rapidly evolving proteins are more likely to become involved in hybrid incompatibilities than their slowly evolving counterparts[4,5,14].

Characterizing the incompatibility across multiple scales of organization allowed us to begin to explore the mechanisms through which it acts[57–59]. Our results suggest that hybrid lethality is mediated in part through developmental delay of individuals with *X. malinche* mitochondria and homozygous *X. birchmanni* ancestry at *ndufs5*. Among heterozygous individuals, we detect physiological impacts on Complex I function, even though these individuals escape the lethal effects of the incompatibility.

Finally, this mitonuclear incompatibility provides a new case in which the same genes are involved in incompatibilities across multiple species[39,60,61]. However, tracing the evolutionary history of the genes that underlie it adds further complexity to this prediction: we found that introgression has resulted in the transfer of genes underlying the incompatibility from *X. malinche* to *X. cortezi*, and evidence from *X. birchmanni* × *X. cortezi* hybrid populations indicates that the incompatibility is likely under selection in these populations as well. The possibility that hybridization could transfer incompatibilities between species has not been previously recognized, perhaps due to an underappreciation of the frequency of hybridization. The importance of past hybridization in the structure of present reproductive barriers is a promising area for future inquiry.

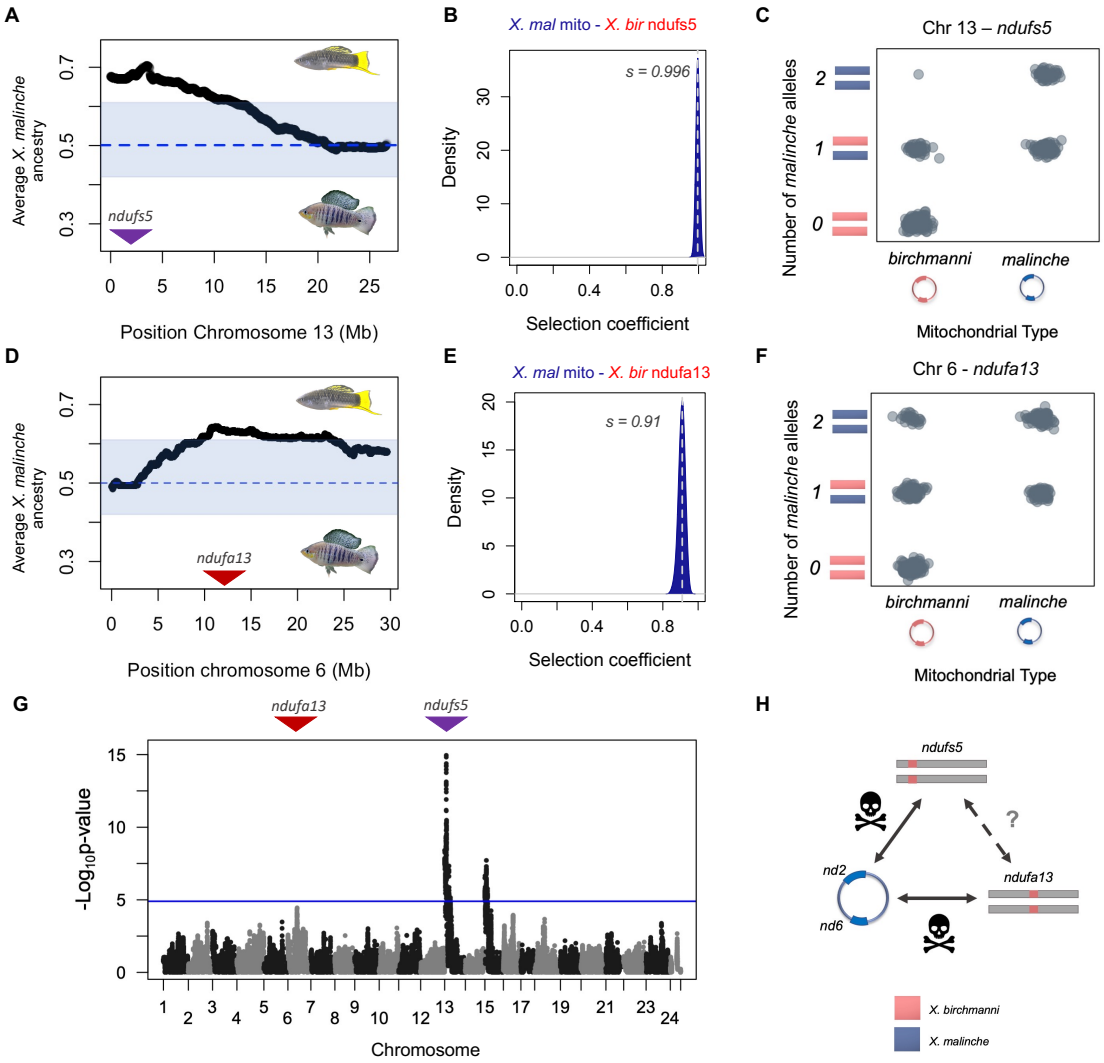

Fig. 1 | Admixture mapping pinpoints a mitonuclear incompatibility in *Xiphophorus*. (**A**) Average ancestry of F_2s on chromosome 13 reveals segregation distortion towards *X. malinche* ancestry across a large region of the chromosome. Blue envelope shows the 99% quantiles of *X. malinche* ancestry at all ancestry informative sites genome wide. Dashed line represents the expected *X. malinche* ancestry for this cross. Purple arrow points to the position of *ndufs5*. (**B**) Results of approximate Bayesian computation (ABC) simulations estimating the strength of selection on the *X. malinche* mitochondria when combined with *X. birchmanni* ancestry at *ndufs5*. Shown here is the posterior distribution from accepted simulations; the vertical line and inset indicates the maximum a posteriori estimate for the selection coefficient. (**C**) Observed genotype frequencies of different genotype combinations of *ndufs5*

(chromosome 13) and mitochondrial haplotypes in the admixture mapping population. (**D**) Average ancestry of F_2s on chromosome 6, reveals segregation distortion towards *X. malinche* ancestry across a large region of the chromosome. Blue envelope and dashed line indicate 99% ancestry quantiles and expected ancestry in the cross as in (**A**), red arrow points to the position of *ndufa13*. (**E**) Results of approximate Bayesian computation (ABC) simulations estimating the strength of selection on the combination of *X. malinche* mitochondria with *X. birchmanni ndufa13*, as in (**B**). (**F**) Observed genotype frequencies of different genotype combinations of *ndufa13* (chromosome 6) and mitochondrial haplotypes in the admixture mapping population. (**G**) Admixture mapping results for associations between nuclear ancestry and mitochondrial haplotype in natural hybrids, controlling for genome-wide ancestry. Blue line indicates the 10% false positive rate genome-wide significance threshold determined by simulations. The peak visible on chromosome 15 is driven by interactions with the *X. birchmanni* mitochondria and an unknown nuclear gene, and is discussed in Supplementary Information 1.1.10 and 1.2.1. (**H**) Schematic of identified interactions with the *X. malinche* mitochondrial genome from our mapping data. We discuss evidence for possible interactions between *ndufs5* and *ndufa13* (indicated by the dashed line) in Supplementary Information 1.1.11.

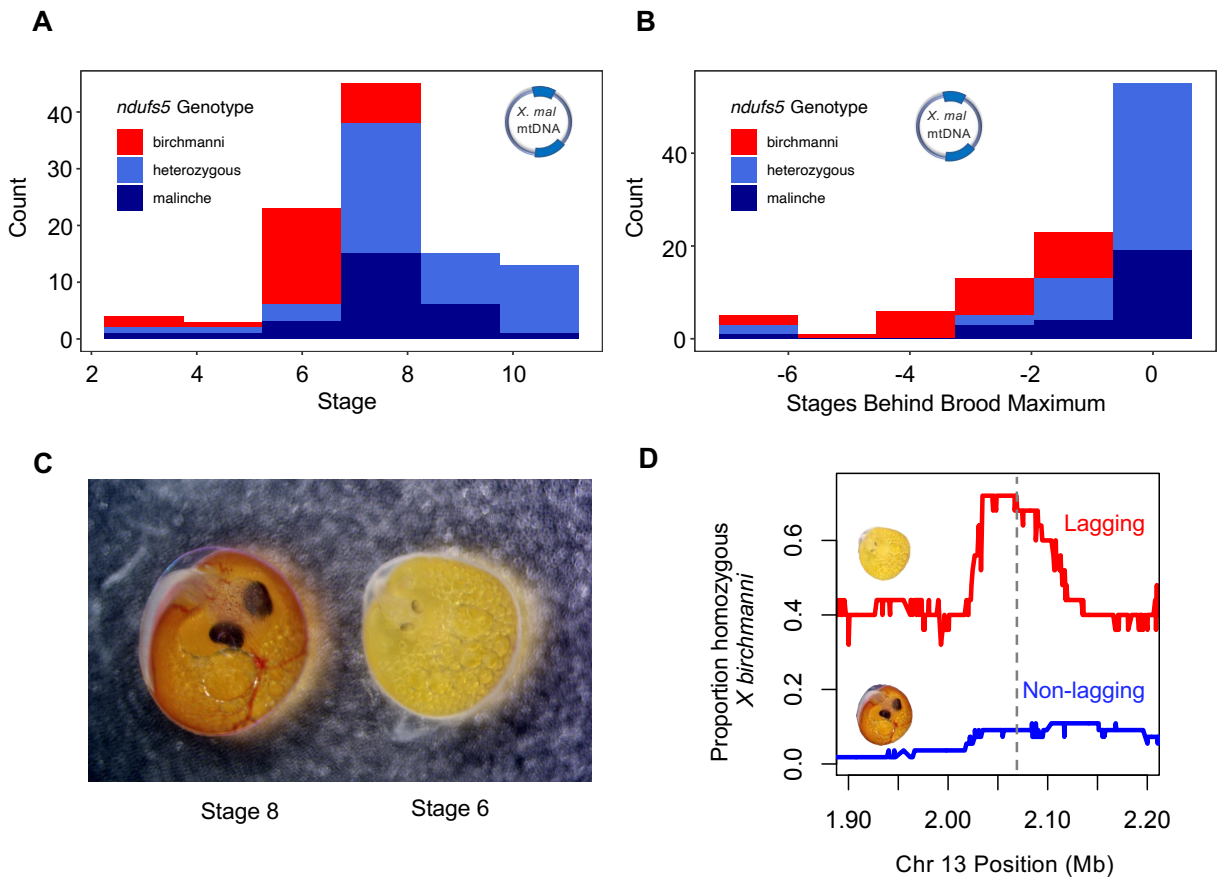

Fig. 2 | Impact of the hybrid incompatibility on *Xiphophorus* hybrid embryos. (**A**) Developmental stage and *ndufs5* genotypes of hybrid embryos with *X. malinche* mitochondria. **(B)** Lag in development of hybrid embryos with *X. malinche* mitochondria compared to the most developed embryo in their brood as a function of *ndufs5* genotype. (**C**) Siblings from the admixture mapping population at different developmental stages. (**D**) Frequency of homozygous *X. birchmanni* ancestry along chromosome 13 in embryos with *X. malinche* mitochondria that lagged their siblings in developmental stage by ≥1 developmental stage (red) versus the frequency of homozygous *X. birchmanni* ancestry in embryos that did not exhibit developmental lag (blue, see Supplementary Information 1.1.7). Dashed line indicates the location of *ndufs5*. Note that only 69% of embryos with developmental lag have homozygous *X. birchmanni* in this region, indicating that there are other causes of this phenotype, either environmentally or elsewhere in the genome. For the same analysis of chromosomes 6 and 15, where we see no clear difference in average ancestry as a function of lag status, see Fig. S11-13.

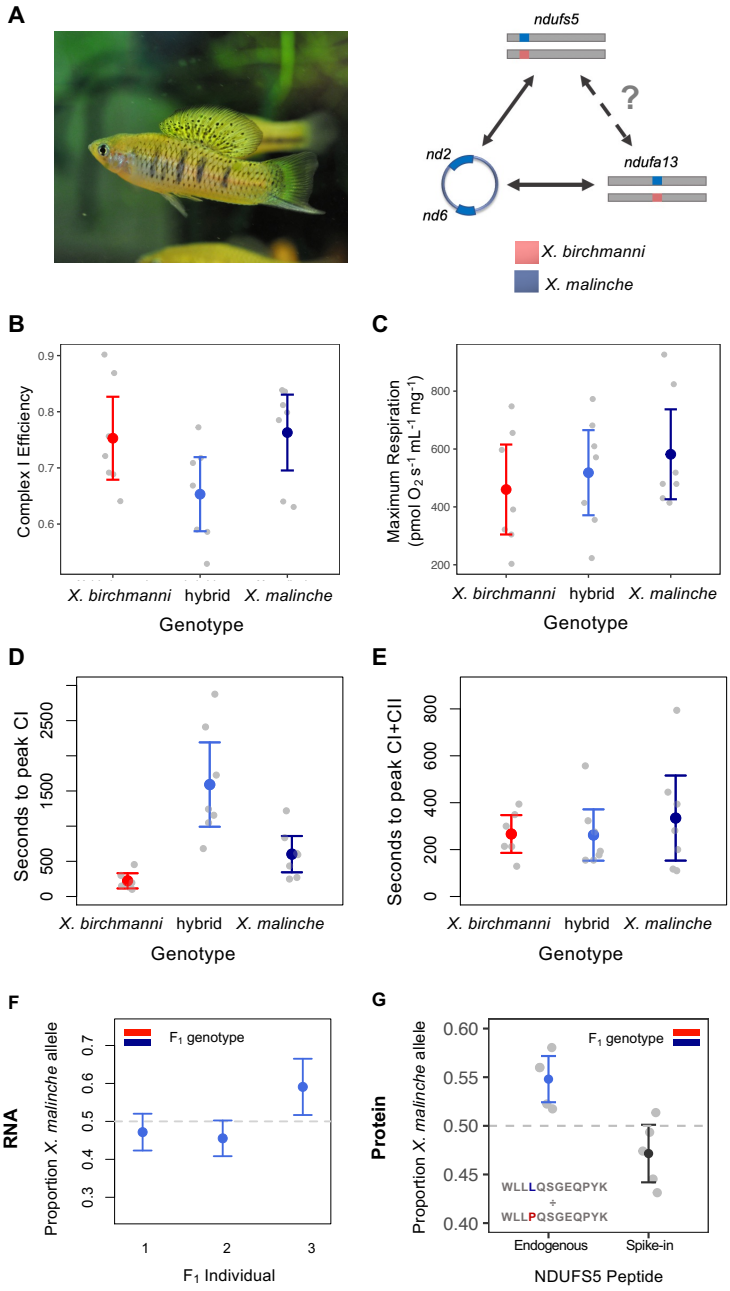

Fig. 3 | Physiological and proteomic phenotypes of viable heterozygotes harboring the hybrid incompatibility. In all panels, colored points and whiskers show the mean ± 2 standard errors, and gray points show individual data. (**A**) Representative image and schematic of ancestry at loci involved in the incompatibility in an *X. birchmanni* × *malinche* F_1 hybrid. Heterozygous hybrids avoid the lethal effects of the interaction between the *X. malinche* mitochondria and *X. birchmanni* ancestry at *ndufs5* and *ndufa13* (Supplementary Information 1.2.2). (**B**) Results of Oroboros O2K respirometer 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) point to lower Complex I efficiency in hybrids. (**C**) Maximum respiration rates during full O2K protocol as a function of genotype did not differ between groups despite significant differences in Complex I efficiency. (**D**) Time to reach the maximum rate of Complex I-driven respiration after the addition of ADP differed between hybrids and parental species. Complex I-driven respiration begins with the addition of ADP, as the flow of electrons is previously limited by the inability of Complex V to relieve the proton gradient in the absence of its substrate (see Fig. S20 for example time-to-peak curves). (**E**) Time to reach the peak in Complex I- and Complex II-driven respiration after the addition of succinate did not differ across genotypes. Complex II-driven respiration begins with the addition of succinate, which is the electron donating substrate of Complex II. (**F**) Allele-specific expression of *ndufs5* in three adult F_1 hybrids. (**G**) Results of quantitative mass spectrometry analysis of *ndufs5* peptides in mitochondrial proteomes derived from five adult F_1 hybrids. Data points show the proportion of area under the spectral curves contributed by the *X. malinche* allele in a given individual. 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. Inset shows the identities of heavy-labeled peptides for each species.

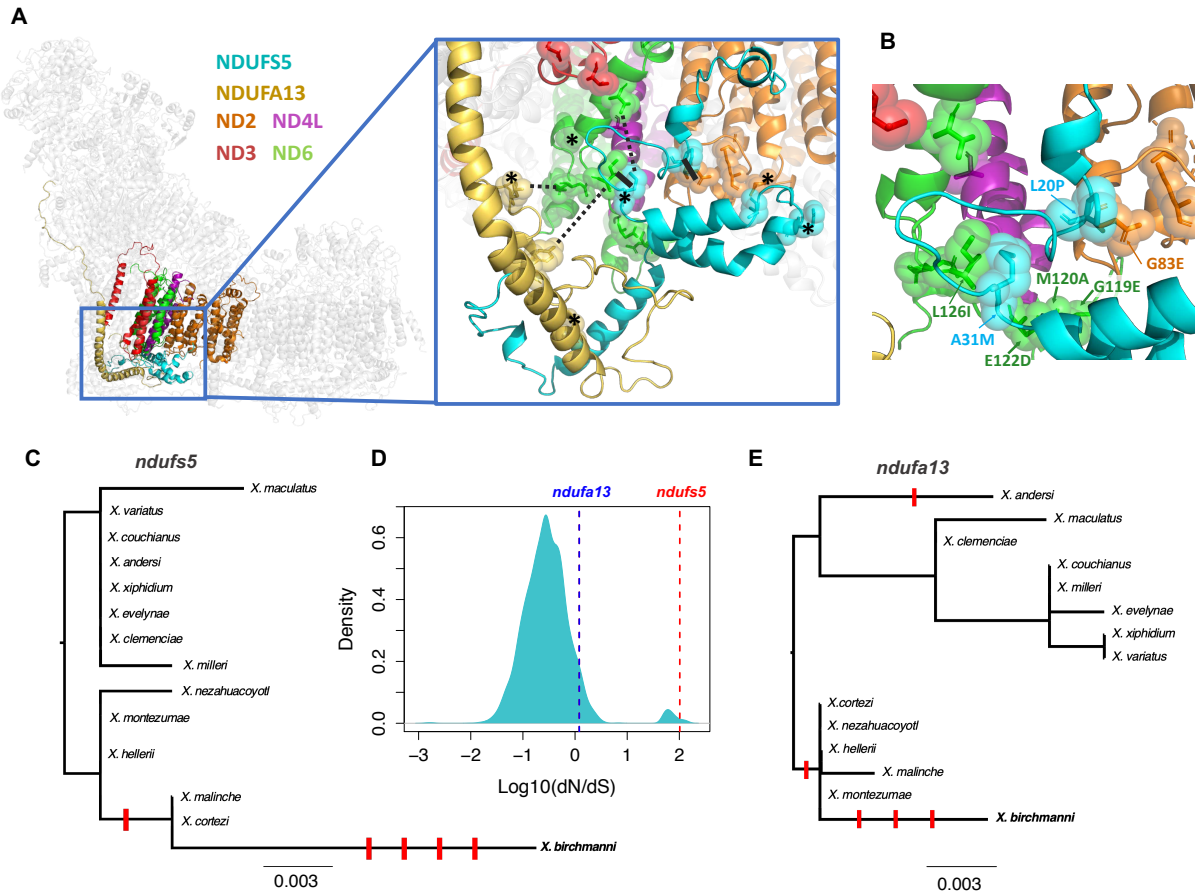

Fig. 4 | Predicted structures of *Xiphophorus* respiratory Complex I reveal interacting substitutions at protein interfaces. (**A**) *Xiphophorus* respiratory Complex I structures generated by RaptorX using alignment to a template mouse cryo-EM structure. Colored protein structures include *ndufs5, ndufa13*, and the four mitochondrially encoded *nd* genes in contact with *ndufs5* or *ndufa13*. Inset shows the surface of contact between these genes. Solid black lines highlight two areas of close contact between interspecific substitutions (alpha carbon distance ≤ 10 Angstrom for all models), while dashed lines show three additional areas in which there was weaker evidence for contact (side chain distance ≤ 12 Angstrom in at least one model). Asterisks denote residues with substitutions in *X. birchmanni* predicted to affect protein function (Table S9). (**B**) Detailed view of interaction interface between *ndufs5*, *nd2*, and *nd6.* Spheres highlight substitutions between *X. birchmanni* and *X. malinche*. For substitutions in close proximity, the residues are labeled with letters denoting the *X. malinche* allele, the residue number, and

500 the *X. birchmanni* allele, respectively (see Supplementary Information 1.4.6 and Table S9 for details). (**C**)
501 Gene tree for *ndufs5* generated with RAxML, highlighting an excess of substitutions along the *X.*
502 *birchmanni* branch. Scale bar represents number of nucleotide substitutions per site. Derived non-
503 synonymous substitutions are indicated by red ticks along the phylogeny. Note that spacing between ticks
504 is arbitrary. (**D**) Distribution of Log_{10} dN/dS between *X. birchmanni* and *X. malinche* across all nuclear
505 genes in the genome with values for *ndufs5* and *ndufa13* highlighted. (**E**) Gene tree for *ndufa13* generated
506 with RAxML, highlighting an excess of substitutions along the *X. birchmanni* branch (as in C).
507
508

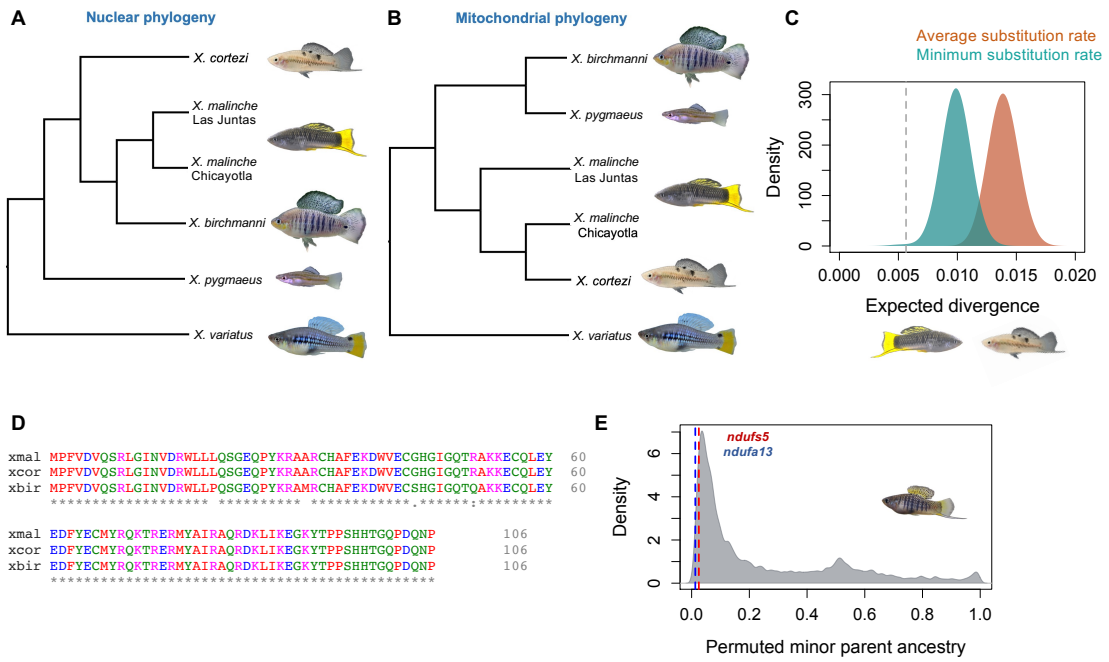

Fig. 5 | Phylogenetic analysis and ancestry mapping suggest that genes underlying the mitonuclear incompatibility have introgressed from *X. malinche* into *X. cortezi*. (**A**) Nuclear phylogeny of *Xiphophorus* species, showing that *X. birchmanni* and *X. malinche* are sister species[55]. (**B**) 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 Information 1.5.4). (**D**) Clustal alignment of *ndufs5* sequences shows that *X. malinche* and *X. cortezi* have identical amino acid sequences at *ndufs5*, hinting at possible introgression of this nuclear gene, while *X. birchmanni* is separated from them by four substitutions. Similar patterns are observed for *ndufa13*. 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. S35) 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.

Methods

Biological Materials

Wild parental and hybrid individuals used in this study were collected from natural populations in Hidalgo, Mexico (Permit No. PPF/DGOPA-002/19). Artificial F_1 and F_2 hybrids were generated using large mesocosm tanks at the Centro de Investigaciones Científicas de las Huastecas "Aguazarca", as described previously[40]. Caudal fin clips were used as the source for all DNA isolation and for flow cytometry, and liver tissue for RNAseq, respirometry, and proteomic assays were collected following Stanford APLAC protocol #33071.

Genotyping and local ancestry calling

Genomic DNA was extracted from fin clips collected from natural and artificial hybrids and individually barcoded tagmentation based libraries were generated for each individual (Supplementary Information 1.1.3). Hybrids were genotyped with low-coverage whole genome sequencing followed by local ancestry inference across the 24 *Xiphophorus* chromosomes and the mitochondrial genome using the *ancestryinfer* pipeline[39,40,43,62] (Supplementary Information 1.1.3-1.1.4). We converted posterior probabilities for each ancestry state (homozygous *X. birchmanni*, heterozygous, and homozygous *X. malinche*) to hard-calls for downstream analysis, using a posterior probability threshold of 0.9, and analyzed ancestry variation across the genome.

QTL and admixture mapping

The region interacting with the mitochondrial genome was first identified based on analysis of segregation distortion in 943 F_2 hybrids generated from F_1 crosses between *X. malinche* females and *X. birchmanni* males (Supplementary Information 1.1.1 and Langdon et

554 al[41]). Since all hybrids in this artificial cross harbored the *X. malinche* mitochondria, we scanned
555 for regions of exceptionally high *X. malinche* ancestry along the genome (>60% *X. malinche*
556 ancestry), identifying one such region on chromosome 13 and one on chromosome 6 (Fig. 1; see
557 also[41]). Evidence for interactions between these regions and the mitochondrial genome were
558 confirmed using admixture mapping. Selection against incompatible genotype combinations
559 generates "missing" two-locus genotypes, and induces unexpectedly high correlations in ancestry
560 between physically unlinked loci. To search for these associations between mitochondrial and
561 nuclear genotypes, we took advantage of two hybrid populations that segregated for the
562 mitochondrial haplotype of both species (Supplementary Information 1.1.2): the Calnali Low
563 hybrid population (N = 359) and the Chahuaco falls hybrid population (N = 244). Briefly, we
564 used a partial correlation analysis to identify regions of the genome strongly associated with
565 mitochondrial ancestry, after regressing out genome-wide ancestry to account for covariance in
566 ancestry due to population structure (see [37] and Supplementary Information 1.1.5, 1.1.9).
567 Significance thresholds for admixture mapping analyses were determined using simulations
568 (Supplementary Information 1.1.5).
569 We also took advantage of data from three natural hybrid populations that had fixed
570 either the *X. birchmanni* (Acuapa: N=117 and Aguazarca: N=126) or *X. malinche* mitochondrial
571 haplotype (Tlatemaco: N=126) to evaluate evidence of selection on regions identified through
572 admixture mapping (Supplementary Information 1.1.6). Natural hybrids from these populations
573 were genotyped using low-coverage whole genome sequencing and by applying the
574 *ancestryinfer* pipeline described above[39,40,43,62].
575
576

Estimates of selection on the mitonuclear incompatibility

We used an ABC approach to estimate the strength of selection against the incompatible interaction between the *X. malinche* mitochondrial haplotype and *X. birchmanni* ancestry at the two nuclear genes involved in the hybrid incompatibility: *ndufs5* and *ndufa13* (Supplementary Information 1.2.2). For these simulations, we took advantage of data from F_2 hybrids, where the known cross design simplifies the parameter space we must explore. Specifically, in F_2 hybrids, the expectation is that across individuals 50% of alleles will be derived from *X. malinche* and 50% of alleles will be derived from *X. birchmanni* in the nuclear genome. Since we observed substantial deviations from this expectation at *ndufs5* and *ndufa13*, we asked what selection coefficients (0-1) and dominance coefficients (0-1) could generate the observed genotypes in F_2 hybrids at *ndufs5* and *ndufa13* after two generations of selection. We performed 500,000 simulations for each interaction and accepted or rejected simulations based on comparisons to the real data using a 5% tolerance threshold (Supplementary Information 1.2.2).

We also evaluated evidence for incompatible interactions with the *X. birchmanni* mitochondrial haplotype and inferred the strength of selection on this direction of the *ndufs5* interaction (Supplementary Information 1.2.1-1.2.2). We again used an ABC approach but due to differences in the type of data available we implemented these simulations using the population simulator SELAM. See Supplementary Information 1.2.1-1.2.4 for more details.

Developmental staging and genotyping of embryos

To pinpoint when in development the incompatibility between the *X. malinche* mitochondria and *X. birchmanni* nuclear genotypes causes lethality, we collected a dataset on the developmental stages of embryos with different genotype combinations. We focused our

sampling efforts on pregnant females from the Calnali Low hybrid population, where incompatible genotypes are more common. Poeciliid embryos must complete all stages of embryonic development in the maternal environment to survive after birth (Supplementary Information 1.3.1).

Whole ovaries were removed from pregnant females and embryos were individually dissected. Each embryo was assigned a developmental stage ranging from 1-11 based on established protocols for poeciliid embryos[45]. Unfertilized eggs were excluded from analysis. Following staging, individual embryos (N = 296) were genotyped as described above and in Supplementary Information 1.3.1. We tested for significant differences in developmental stage between siblings with compatible and incompatible genotype combinations using a two-sided two-sample t-test (Supplementary Information 1.3.1) and examined differences in ancestry between large groups of siblings that varied in their developmental stages (Supplementary Information 1.1.7). We also collected data on embryonic stage and variability between siblings in embryonic stage from both pure parental species for comparison to the hybrid data (Supplementary Information 1.3.1).

Mitochondrial respirometry

Our results indicate that the hybrid incompatibility between the *X. malinche* mitochondria and *X. birchmanni* nuclear genes is at least partially recessive. Gene expression, allele specific expression, and analysis of mitochondrial copy number indicated that F_1 hybrids heterozygous for the mitonuclear incompatibility are not compensating for the incompatibility through altered expression or changes in mitochondrial copy number (Supplementary Information 1.3.3-1.3.4).

To further evaluate mitochondrial function in individuals heterozygous for the mitonuclear incompatibility, we conducted respirometry assays on *X. birchmanni*, *X. malinche*, and hybrid individuals that had the *X. malinche* mitochondria and were heterozygous for the nuclear components of the hybrid incompatibility (N=7 of each genotype). Mitochondria were isolated from whole liver tissue from each of these individuals (Supplementary Information 1.3.5). Mitochondrial respiration was quantified using the Oroboros O2K respirometry system fitted with small volume modules. Using a standardized concentration of mitochondrial isolate (0.15 mg of protein), we followed a multi-substrate, inhibitor, and uncoupler titration protocol to quantify respiration in eight distinct states (Fig. S17) based on a protocol adapted from killifish[63]. A step-by-step description of this protocol and methods used to calculate respiratory flux control factors is outlined in Supplementary Information 1.3.5. To test for effects of genotype on Complex I efficiency and maximum respiration rate, we constructed an orthogonal contrast between parental and hybrid values of these parameters, using test date as a covariate in the linear model. Although we focus our comparisons in the main text on measures of Complex I efficiency and maximum respiration, we report analyses for all respiratory flux control factors in Supplementary Information 1.3.5. We complemented the results of these respirometry experiments with measures of mitochondrial membrane potential in hybrids and parental species using a flow cytometry-based approach (Supplementary Information 1.3.6).

Parallel reaction monitoring proteomics

For Parallel Reaction Monitoring (PRM) with mass spectrometry, we used a similar approach to that used for respirometry to isolate whole mitochondria from five F_1 hybrids (which harbored *X. malinche* mitochondria). This approach is described in detail in Supplementary

Information 1.4.1. Briefly, we designed heavy labeled peptides to distinguish between the *X. birchmanni* and *X. malinche* copies of *ndufs5* and *ndufa13* (Supplementary Information 1.4.2). These peptides were designed to mimic the products of trypsin digestion of the native peptides from both *X. birchmanni* and *X. malinche*, but had ^{13}C- and ^{15}N-labeled arginine or lysine at the C terminus, and cysteine residues carbamoylated with iodoacetamide. This approach facilitates quantification of the peptides of interest in the mitochondrial proteome (Supplementary Information 1.4.2).

Mitochondrial isolates were prepared for mass spectrometry and combined with heavy labeled peptides in known quantities (see Supplementary Information 1.4.3). Mass spectrometry experiments were performed on a Q Exactive HF-X Hybrid Quadrupole - Orbitrap mass spectrometer with liquid chromatography using a Nanoacquity UPLC, and a parallel reaction monitoring method was used for ion selection. The protocol for mass spectrometry and PRM is described in detail in Supplementary Information 1.4.4.

To analyze the results, raw data were imported into the Skyline program. We called the focal peptide's spectral peak so that the window captured the signal from the heavy labeled spike-in peptide and applied the same retention time interval to detect the endogenous peptide. We focused analysis on the *ndufs5* peptide WLL[L/P]QSGEQPYK since other endogenous peptides were below the expected sensitivity limits of our PRM protocol (Supplementary Information 1.4.5). Given known quantities of the heavy-labeled spike-in peptide, we normalized intensities of the endogenous peptides and asked what proportion of endogenous *ndufs5* peptides in the mitochondrial proteome of each F_1 individual were derived from *X. malinche* versus *X. birchmanni* (see Supplementary Information 1.4.5 for more details). We asked whether these

667 ratios significantly deviated from the 50-50 expectation for F_1 hybrids using a two-sided one-
668 sample *t*-test.
669
670 **Complex I protein modeling**
671 Mapping results allowed us to identify *ndufs5* and *ndufa13* as the *X. birchmanni* genes
672 that interact negatively with *X. malinche* mitochondrial genes. We used a protein-modeling based
673 approach with RaptorX (http://raptorx.uchicago.edu) to identify the mitochondrial genes most
674 likely to interact with *ndufs5* and *ndufa13* (see Supplementary Information 1.4.6). Using the
675 mouse Cryo-EM structure (PDB ID 6G2J) of Complex I, we identified proteins in contact with
676 *ndufs5* and *ndufa13*, which included several mitochondrial (*nd2, nd3, nd4l, nd6*) and nuclear
677 (*ndufa1*, *ndufa8*, *ndufb5*, *ndufc2*) genes. We then used RaptorX to predict structures for both the
678 *X. birchmanni* and *X. malinche* versions of the proteins. We were especially interested in cases
679 where species-specific substitutions were predicted to be in physical contact between the
680 mitochondrial and nuclear proteins in Complex I. In addition, we evaluated the robustness of
681 these predictions to choice of Cryo-EM template; see Supplementary Information 1.4.6 for
682 results and discussion.
683
684 **Analysis of evolutionary rates**
685 Comparison of predicted protein sequences from *ndufs5, ndufa13*, and mitochondrial
686 genes of interest (*nd2* and *nd6*) revealed a large number of substitutions between *X. birchmanni*
687 and *X. malinche*. Using the program PAML, we calculated dN/dS between *X. birchmanni* and *X.*
688 *malinche* for all annotated protein coding genes throughout the genome and found that both
689 *ndufs5* and *ndufa13* have unexpectedly high rates of protein evolution (Fig. 4D; Supplementary

690 Information 1.5.1). Examining mutations that distinguished species in a phylogenetic context
691 revealed that a large number of substitutions in *ndufs5*, *nudfa13*, and *nd6* were derived in *X.*
692 *birchmanni*. We implemented a branch test using the codeml function in PAML to test for
693 significant differences in evolutionary rates of *ndufs5*, *nudfa13*, and *nd6* on the *X. birchmanni*
694 lineage (species included: *X. birchmanni*, *X. malinche*, *X. cortezi*, *X. pygmaeus*, *X.*
695 *nezahualcotoyl*, *X. montezumae*, *X. hellerii*, *X. couchianus*, *X. variatus*, and *X. maculatus*). We
696 also evaluated the predicted functional impacts of individual substitutions using protein
697 alignments for mitochondrial and nuclear proteins of interest from across bony fish and the
698 program SIFT[64]. See Supplementary Information 1.5.1 for more information on both PAML and
699 SIFT analyses.
700
701 **Tests for ancient introgression**
702 Previous work had indicated that the mitochondrial phylogeny in *Xiphophorus* is
703 discordant with the whole-genome species tree[55]. Specifically, although *X. birchmanni* and *X.*
704 *malinche* are sister species based on the nuclear genome, *X. malinche* and *X. cortezi* are sister
705 species based on the mitochondrial genome. We used a combination of PacBio amplicon
706 sequencing of 10 individuals (2 or more per species, Supplementary Information 1.5.3) and
707 newly available whole-genome resequencing data to confirm this result and polarize the direction
708 of the discordance by constructing maximum likelihood mitochondrial phylogenies with the
709 program RAxML[65]. We performed similar phylogenetic analyses of the nuclear genes that
710 interact with the *X. malinche* mitochondria (*ndufs5* and *ndufa13*; Supplementary Information
711 1.5.3).

Combined with phylogenetic results, simulation results suggest that gene flow from *X. malinche* into *X. cortezi* is the most likely cause of the discordance we observe between the mitochondrial and nuclear phylogenies (Supplementary Information 1.5.3-1.5.4). Since *X. malinche* and *X. cortezi* are not currently sympatric, this suggests ancient gene flow between them. Using whole genome alignments of *X. birchmanni*, *X. malinche*, and *X. cortezi* individuals and the admixtools package[66] we tested for genome-wide evidence of admixture between *X. malinche* and *X. cortezi* using the qpDstat function with a block-jackknife block size of 5 Mb, (Supplementary Information 1.5.5).

Contemporary hybridization between *X. birchmanni* and *X. cortezi*

Comparison of *X. cortezi* and *X. malinche* sequences at *ndufs5*, *ndufa13*, and their mitochondrial interactors (*nd2* and *nd6*) indicate that *X. cortezi* and *X. malinche* have identical or nearly identical amino acid sequences at these Complex I genes, and that both species differ substantially from *X. birchmanni* (Supplementary Information 1.5.6). To investigate the possibility that hybrids between *X. birchmanni* and *X. cortezi* share the same mitonuclear incompatibility as observed in hybrids between *X. birchmanni* and *X. malinche*, we took advantage of genomic data from recently discovered hybrid populations between *X. birchmanni* and *X. cortezi*[67]. Using a permutation-based approach, we asked whether ancestry at *ndufs5* and *ndufa13* showed lower mismatch with mitochondrial ancestry than expected given the genome-wide ancestry distribution, across two *X. birchmanni* × *X. cortezi* hybrid populations. These permutations are described in detail in Supplementary Information 1.5.6.

Animal Care and Use

All methods were performed in compliance with Stanford Administrative Panel on Laboratory Animal Care protocol #33071.

Data Availability

Raw sequencing reads used in this project are available under NCBI SRA Bioprojects PRJNA744894, PRJNA746324, PRJNA610049, and PRJNA745218. Mass spectrometry data are available on PRIDE (accession pending), and all other datasets necessary to recreate the results of the publication are available on Dryad (accessions pending).

Code Availability

All new customs scripts used to generate results will be made available on Github at https://github.com/Schumerlab/mitonuc_DMI and https://github.com/Schumerlab/Lab_shared_scripts.

References

1. Dagilis, A. J., Kirkpatrick, M. & Bolnick, D. I. The evolution of hybrid fitness during speciation. *PLOS Genetics* **15**, e1008125 (2019).
2. Dobzhansky, Th. Genetic Nature of Species Differences. *The American Naturalist* **71**, 404–420 (1937).
3. Müller, H. Isolating mechanisms, evolution, and temperature. *Biological Symposium* **6**, 71–125 (1942).
4. Maheshwari, S. & Barbash, D. A. The Genetics of Hybrid Incompatibilities. *Annu Rev Genet* **45**, 331–355 (2011).
5. Presgraves, D. C. The molecular evolutionary basis of species formation. *Nat Rev Genet* **11**, 175–180 (2010).
6. Swamy, K. B. S., Schuyler, S. C. & Leu, J.-Y. Protein Complexes Form a Basis for Complex Hybrid Incompatibility. *Front. Genet.* **12**, (2021).
7. Orr, H. A. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**, 1805–1813 (1995).
8. Boocock, J., Sadhu, M. J., Durvasula, A., Bloom, J. S. & Kruglyak, L. Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast. *Science* **371**, 415–419 (2021).
9. Wei, W.-H., Hemani, G. & Haley, C. S. Detecting epistasis in human complex traits. *Nat Rev Genet* **15**, 722–733 (2014).
10. Gauderman, W. J. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* **155**, 478–484 (2002).
11. Phadnis, N. *et al.* An essential cell cycle regulation gene causes hybrid inviability in Drosophila. *Science* **350**, 1552–1555 (2015).
12. Tang, S. & Presgraves, D. C. Evolution of the Drosophila Nuclear Pore Complex Results in Multiple Hybrid Incompatibilities. *Science* **323**, 779–782 (2009).

13. Cooper, J. C., Guo, P., Bladen, J. & Phadnis, N. A triple-hybrid cross reveals a new hybrid incompatibility locus between D. melanogaster and D. sechellia. *bioRxiv* 590588 (2019) doi:10.1101/590588.
14. Johnson, N. A. Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends Genet* **26**, 317–325 (2010).
15. Burton, R. S. & Barreto, F. S. A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Mol Ecol* **21**, 4942–4957 (2012).
16. Hill, G. E. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol Evol* **6**, 5831–5842 (2016).
17. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
18. Barr, C. M., Neiman, M. & Taylor, D. R. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168**, 39–50 (2005).
19. Chase, C. D. Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions. *Trends Genet* **23**, 81–90 (2007).
20. Case, A. L., Finseth, F. R., Barr, C. M. & Fishman, L. Selfish evolution of cytonuclear hybrid incompatibility in Mimulus. *Proc. R. Soc. Lond. B* **283**, 20161493 (2016).
21. Ballard, J. W. O. & Whitlock, M. C. The incomplete natural history of mitochondria. *Mol Ecol* **13**, 729–744 (2004).
22. Haag-Liautard, C. *et al.* Direct Estimation of the Mitochondrial DNA Mutation Rate in Drosophila melanogaster. *PLOS Biology* **6**, e204 (2008).
23. Allio, R., Donega, S., Galtier, N. & Nabholz, B. 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. *Mol Biol Evol* **34**, 2762–2772 (2017).
24. Osada, N. & Akashi, H. Mitochondrial–Nuclear Interactions and Accelerated Compensatory Evolution: Evidence from the Primate Cytochrome c Oxidase Complex. *Mol Biol Evol* **29**, 337–346 (2012).

801 25. Barreto, F. S. & Burton, R. S. Evidence for Compensatory Evolution of Ribosomal Proteins in
802 Response to Rapid Divergence of Mitochondrial rRNA. *Mol Biol Evol* **30**, 310–314 (2013).
803 26. Bolnick, D. I., Turelli, M., López-Fernández, H., Wainwright, P. C. & Near, T. J. Accelerated
804 Mitochondrial Evolution and "Darwin's Corollary": Asymmetric Viability of Reciprocal F1 Hybrids
805 in Centrarchid Fishes. *Genetics* **178**, 1037–1048 (2008).
806 27. Turelli, M. & Moyle, L. C. Asymmetric Postmating Isolation: Darwin's Corollary to Haldane's Rule.
807 *Genetics* **176**, 1059–1088 (2007).
808 28. Brandvain, Y., Pauly, G. B., May, M. R. & Turelli, M. Explaining Darwin's Corollary to Haldane's
809 Rule: The Role of Mitonuclear Interactions in Asymmetric Postzygotic Isolation Among Toads.
810 *Genetics* **197**, 743–747 (2014).
811 29. Tiffin, P., Olson, S. & Moyle, L. C. Asymmetrical crossing barriers in angiosperms. *Proc. R. Soc.*
812 *Lond. B* **268**, 861–867 (2001).
813 30. Meiklejohn, C. D. *et al.* An Incompatibility between a Mitochondrial tRNA and Its Nuclear-Encoded
814 tRNA Synthetase Compromises Development and Fitness in Drosophila. *PLOS Genetics* **9**, e1003238
815 (2013).
816 31. Luo, D. *et al.* A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in
817 rice. *Nat Genet* **45**, 573–577 (2013).
818 32. Lee, H.-Y. *et al.* Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility
819 between Two Yeast Species. *Cell* **135**, 1065–1073 (2008).
820 33. Hanson, M. R. & Bentolila, S. Interactions of Mitochondrial and Nuclear Genes That Affect Male
821 Gametophyte Development. *The Plant Cell* **16**, S154–S169 (2004).
822 34. Culumber, Z. W. *et al.* Replicated hybrid zones of *Xiphophorus* swordtails along an elevational
823 gradient. *Mol Ecol* **20**, 342–356 (2011).
824 35. Fisher, H. S., Wong, B. B. M. & Rosenthal, G. G. Alteration of the chemical environment disrupts
825 communication in a freshwater fish. *Proc. R. Soc. B* **273**, 1187–1193 (2006).

826 36. Schumer, M. *et al.* Natural selection interacts with recombination to shape the evolution of hybrid
827 genomes. *Science* **360**, 656–660 (2018).
828 37. Schumer, M. & Brandvain, Y. Determining epistatic selection in admixed populations. *Mol Ecol* **25**,
829 2577–2591 (2016).
830 38. Schumer, M. *et al.* High-resolution mapping reveals hundreds of genetic incompatibilities in
831 hybridizing fish species. *eLife* **3**, e02535 (2014).
832 39. Powell, D. L. *et al.* Natural hybridization reveals incompatible alleles that cause melanoma in
833 swordtail fish. *Science* **368**, 731–736 (2020).
834 40. Powell, D. L. *et al.* The Genetic Architecture of Variation in the Sexually Selected Sword Ornament
835 and Its Evolution in Hybrid Populations. *Current Biology* (2021) doi:10.1016/j.cub.2020.12.049.
836 41. Langdon, Q. K. *et al.* Predictability and parallelism in the contemporary evolution of hybrid
837 genomes. *PLOS Genetics* **18**, e1009914 (2022).
838 42. Schartl, M. *et al.* The genome of the platyfish, Xiphophorus maculatus, provides insights into
839 evolutionary adaptation and several complex traits. *Nature Genetics* **45**, 567 (2013).
840 43. Schumer, M., Powell, D. L. & Corbett-Detig, R. Versatile simulations of admixture and accurate
841 local ancestry inference with mixnmatch and ancestryinfer. *Mol Ecol Resour* **20**, 1141–1151 (2020).
842 44. Sharma, L. K., Lu, J. & Bai, Y. Mitochondrial Respiratory Complex I: Structure, Function and
843 Implication in Human Diseases. *Curr Med Chem* **16**, 1266–1277 (2009).
844 45. Haynes, J. L. Standardized Classification of Poeciliid Development for Life-History Studies. *Copeia*
845 **1995**, 147 (1995).
846 46. Tavolga, W. N. Embryonic development of the platyfish (Platypoecilus), the swordtail
847 (Xiphophorus), and their hybrids. Bulletin of the AMNH ; v. 94, article 4. *Embryonic development in*
848 *fish* (1949).
849 47. Byrnes, J. *et al.* Pharmacologic modeling of primary mitochondrial respiratory chain dysfunction in
850 zebrafish. *Neurochem Int* **117**, 23–34 (2018).

48. Pinho, B. R. *et al.* How mitochondrial dysfunction affects zebrafish development and cardiovascular function: an in vivo model for testing mitochondria-targeted drugs. *Br J Pharmacol* **169**, 1072–1090 (2013).
49. Agip, A.-N. A. *et al.* Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat Struct Mol Biol* **25**, 548–556 (2018).
50. Kampjut, D. & Sazanov, L. A. The coupling mechanism of mammalian respiratory complex I. *Science* (2020) doi:10.1126/science.abc4209.
51. Fiedorczuk, K. *et al.* Atomic structure of the entire mammalian mitochondrial complex I. *Nature* **538**, 406–410 (2016).
52. Mack, K. L. & Nachman, M. W. Gene Regulation and Speciation. *Trends Genet* **33**, 68–80 (2017).
53. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nature Protocols* **7**, 1511–1522 (2012).
54. Hill, G. E. Mitonuclear Compensatory Coevolution. *Trends Genet* (2020) doi:10.1016/j.tig.2020.03.002.
55. Cui, R. *et al.* Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution* **67**, 2166–2179 (2013).
56. Sloan, D. B., Havird, J. C. & Sharbrough, J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol Ecol* **26**, 2212–2236 (2017).
57. Ellison, C. K. & Burton, R. S. Disruption of Mitochondrial Function in Interpopulation Hybrids of Tigriopus Californicus. *Evolution* **60**, 1382–1391 (2006).
58. Ellison, C. K., Niehuis, O. & Gadau, J. Hybrid breakdown and mitochondrial dysfunction in hybrids of Nasonia parasitoid wasps. *J Evol Biol* **21**, 1844–1851 (2008).
59. Olson, J. R., Cooper, S. J., Swanson, D. L., Braun, M. J. & Williams, J. B. The Relationship of Metabolic Performance and Distribution in Black-Capped and Carolina Chickadees. *Physiol Biochem Zool* **83**, 263–275 (2010).

60. Lu, Y. *et al.* Oncogenic allelic interaction in Xiphophorus highlights hybrid incompatibility. *PNAS* **117**, 29786–29794 (2020).
61. Nelson, T. C. *et al.* 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).
62. Corbett-Detig, R. & Nielsen, R. A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLOS Genetics* **13**, e1006529 (2017).
63. Chung, D. J., Bryant, H. J. & Schulte, P. M. Thermal acclimation and subspecies-specific effects on heart and brain mitochondrial performance in a eurythermal teleost (Fundulus heteroclitus). *J Exp Biol* **220**, 1459–1471 (2017).
64. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812–3814 (2003).
65. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
66. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065–1093 (2012).
67. Powell, D. L. *et al.* Two new hybrid populations expand the swordtail hybridization model system. *Evolution* **75**, 2524–2539 (2021).

Acknowledgements

We thank Peter Andolfatto, Stepfanie Aguillon, Yaniv Brandvain, Jenn Coughlan, Hunter Fraser, Yuki Haba, Nitin Phadnis, Molly Przeworski, Ken Thompson and members of the Schumer lab for helpful discussion and/or feedback on earlier versions of this manuscript. We thank Alexa Pollock for help performing rotenone trials. 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., A. M. and M. Schu. collected data; B.M.M., C.Y.P., Q.L.K., F.L., J.C.H., A.M., and M. Schu. performed analyses; D.L.P., T.R.G., R.D.L., R.C.-D., J.F. and M. Scha. provided expertise and technical support.

Competing Interests: the authors declare no competing interests.

Materials & Correspondence: correspondence and requests for materials should be directed to B.M.M. (benmoran@stanford.edu) or M. Schu (schumer@stanford.edu).