Sensitivity of genome-wide tests for mitonuclear genetic incompatibilities

Kuster, Shady A.[1,2], Schumer, Molly[3,4], Havird, Justin C.[5], Sloan, Daniel B.[2]

- 1: Cell and Molecular Biology, Colorado State University
- 2: Department of Biology, Colorado State University
- 3: Department of Biology, Stanford University
- 4: Freeman Hrabowski Fellow, Howard Hughes Medical Institute
- 5: Department of Integrative Biology, University of Texas at Austin

Author for correspondence: Shady A. Kuster, Shady.Kuster@colostate.edu

ABSTRACT

Mismatches between interacting mitochondrial and nuclear gene products in hybrids have been proposed to disproportionately contribute to the formation of early species boundaries. Under this model, genetic incompatibilities emerge when mitochondrial haplotypes are placed into a cellular context without their coevolved nuclear-encoded mitochondrial (n-mt) proteins. Although there is strong evidence that mitonuclear coevolution has contributed to reproductive isolation in some cases, it is less clear how far-reaching the effects of mitonuclear incompatibilities are in speciation. Does disrupting co-adapted mitonuclear genotypes have broad, genome-wide effects with numerous n-mt loci contributing to reproductive isolation? We leverage a system with several hybridizing species pairs (*Xiphophorus* fishes) that have known mitonuclear incompatibilities of large effect to ask whether a general signal of incompatibility is present when considering all n-mt genes. After dividing nuclear-encoded proteins into three classes based on level of interaction with mitochondrial gene products, we found only inconsistent statistical evidence for a difference between these classes in the degree of conserved mitonuclear ancestry. Our results imply that genome-wide scans focused on enrichment of broad functional gene classes may sometimes be insufficient for detecting a history of mitonuclear coevolution, even when strong selection is acting on mitonuclear incompatibilities at multiple loci.

31 **INTRODUCTION**

32

33 Since its endosymbiotic origin, the mitochondrion has evolved into an integral component of the
34 eukaryotic cell, functioning in such processes as oxidative phosphorylation (OXPHOS), Ca^{2+}
35 acquisition and signaling, and biosynthesis of heme and iron-sulfur clusters. Bilaterian metazoan
36 mitochondrial genomes (mitogenomes) have been reduced to encode only 13 core protein
37 subunits of the electron transport chain, 2 rRNAs, and 22 tRNAs, making mitochondria
38 dependent on the import of nuclear-encoded, but mitochondrially targeted, (n-mt) proteins for all
39 functions [1–4].

40 Mitogenomes evolve at a faster rate than nuclear genomes in most bilaterian metazoan
41 systems [5–8]. These elevated substitution rates may result from a higher underlying mutation
42 rate in combination with inefficient purging of deleterious mutations due to smaller effective
43 population size, lack of recombination, and uniparental inheritance [9–12], although explanations
44 based on reduced efficacy of selection in mitogenomes have been recently challenged [13–15].
45 Regardless of the cause, the consequences of high mitochondrial DNA (mtDNA) substitution
46 rates are often deleterious [16]. As such, substitutions in mitochondrially encoded proteins may
47 create selection for compensatory changes in interacting n-mt proteins [17–19], and signatures of
48 coevolution have indeed been found between the mitochondrial and nuclear genomes [20–23].
49 Other forms of mitonuclear coevolution (i.e., reciprocal selection for compatible genotypes) are
50 also possible [20,24–26]. Regardless of the exact process, the outcome is a set of interacting
51 proteins that are functionally dependent on sufficient "matching" of lineage-specific amino acid
52 substitutions in mitochondrial and n-mt proteins [24].

53 As a result, disrupting the association between a mitochondrial haplotype and its
54 coevolved nuclear counterparts (e.g., via secondary contact following speciation) is predicted to
55 uncover genetic incompatibilities that made early contributions to reproductive isolation [5,27].
56 Genes that evolve faster are theorized to be the first in generating incompatibilities between
57 diverging populations [28]. Given the rate at which mitonuclear genes coevolve and importance
58 of mitochondrial function to cellular metabolism, mitonuclear incompatibilities are proposed to
59 have a disproportionate effect on the accumulation of reproductive barriers during speciation
60 [24,29,30]. Although studies have found evidence of mitonuclear incompatibilities (e.g., [31,32])
61 in some systems, it is still unclear how far-reaching these effects are. In other words, when

62 selection is acting to maintain "matched" mitonuclear genotypes, are relatively few nuclear loci
63 of large effect involved or many loci of small effect? Moreover, do n-mt genes preferentially
64 contribute to mitonuclear genetic incompatibilities? We hypothesize that, relative to other
65 nuclear genes, n-mt genes are more strongly affected by selection acting to maintain associations
66 between matched lineage-specific mitonuclear genotypes. Therefore, examining signatures of
67 mtDNA "match" in n-mt genes vs. nuclear-encoded genes that lack mitochondrial interactions
68 (non-n-mt genes) should reveal a history of mitonuclear coevolution across the genome at
69 species boundaries.
70 Species of the swordtail fish genus *Xiphophorus* have recently begun hybridizing in
71 multiple river systems in Mexico, providing a model of repeated hybridization and speciation
72 dynamics [33–42]. Recent work on hybrid zones between sister species *X. birchmanni* and *X.*
73 *malinche* has identified severe mitonuclear incompatibilities involving the *ndufs5* and *ndufa13* n-
74 mt genes [43]. Mitonuclear mismatches at these loci have severe fitness effects, including death
75 during embryonic development. Subsequent investigations have identified that at least nine
76 nuclear genomic regions (seven of which contain a total of 10 n-mt genes) are involved in
77 mitonuclear incompatibilities in these hybrids [44]. Some of these incompatibilities, including
78 the *ndufs5* and *ndufa13* n-mt genes, were also found in *X. birchmanni* and *X. cortezi* hybrids
79 [42]. We therefore leveraged this system to investigate evidence for mitonuclear matching on a
80 genome-wide scale in hybridization and the ability to detect these interactions with commonly
81 used approaches to partition nuclear genes into classes based on their role in mitochondrial
82 function. We categorized nuclear genes into three groups based on whether the encoded protein
83 (1) interacts with mitogenomic gene products (interacting n-mt), (2) functions in mitochondria
84 but does not physically interact with mitogenomic gene products (non-interacting n-mt), or (3)
85 does not function in mitochondria (non-n-mt). This approach allows for an unbiased test of
86 whether selection for matched mitonuclear genotypes leaves genome-wide signatures based on
87 this *a priori* partitioning of genes into mitochondrial functional categories. We found
88 inconsistent statistical evidence for differences among these categories for metrics of selection
89 on mitonuclear interactions in *Xiphophorus* hybrids, indicating that the role of n-mt genes (either
90 interacting or non-interacting) is not always detectable on a genome-wide scale.
91
92

93 **METHODS**

94

95 **Genomic data from *Xiphophorus* hybrid populations**

96 Three datasets resulting from previous local ancestry inference in hybrid *Xiphophorus* fishes
97 were acquired from [37,40,43]. These datasets were originally analyzed using the *ancestryinfer*
98 pipeline and resulted in ancestry probabilities at between 628,881 and 689,982 sites genome-
99 wide (Table 1). Each dataset includes samples from a population of hybrid *Xiphophorus*
100 individuals inhabiting sections of rivers in Hidalgo Mexico or San Luis Potosí Mexico. The first
101 (Calnali Low, or CALL), is on the Calnali River and consists of *X. birchmanni* and *X. malinche*
102 hybrids. The second is at Chahuaco Falls (CHAF) on the Calnali River. It is separated from the
103 CALL population by a waterfall (~50m tall) [43] and consists of hybrids of the same species as
104 CALL. The third consists of two populations (Huextetitla and Santa Cruz, or HUEX-STAC)
105 downstream of the same fork on the Santa Cruz River and derived from a single initial
106 hybridization event between *X. birchmanni* and *X. cortezi* followed by independent evolution
107 [37,38].

108 Ancestry informative markers (AIMs) across the nuclear and mitochondrial genomes
109 were assigned a posterior probability of ancestry to each parental species, as described
110 previously [36,39,45]. Informative sites were defined as those with at least a 98% frequency
111 difference between the parental species, as previously described [33,44]. We used a published
112 annotation of the *X. birchmanni* reference genome [46] to identify genes containing each AIM.
113 To obtain average nuclear ancestry and mitochondrial haplotype frequency for each population,
114 we first filtered posterior probabilities of AIMs so they were either below 0.2 or above 0.8 to
115 ensure high quality calls. These filtered probabilities were then averaged for nuclear and
116 mitochondrial AIMs, respectively (Table 1).

117

118 **Identification of *Xiphophorus* n-mt genes**

119 The gffread [47] program was used to translate the *X. birchmanni* fasta file into protein
120 sequences, using the -V option to remove any coding sequence with an in-frame stop codon. To
121 annotate n-mt proteins, we compared the *Xiphophorus* translated proteome to UniProt human
122 (UP000005640, GCA_000001405.27) and mouse (UP000000589, GCA_000001635.8) protein
123 sequences. N-mt proteins have been empirically identified in humans and mice as part of the

124 MitoCarta 3.0 project [4]. Because the UniProt human and mouse datasets differed in
125 nomenclature from the MitoCarta dataset, we assigned protein names by searching for MitoCarta
126 sequences in the UniProt human sequences with BLASTP v2.15.0+. A custom Perl script was
127 then used to parse the BLASTP output, and identified proteins were manually confirmed to be
128 the correct protein if percent coverage and sequence identity were not 100%. OrthoFinder v2.5.4
129 [48] was then run with the primary transcript for each *Xiphophorus* sequence and the full sets of
130 human and mouse protein sequences.
131 The resulting orthogroups were then organized into our three gene classes: (1) interacting
132 n-mts, (2) non-interacting n-mts, and (3) non-n-mts. MitoCarta proteins differentiated non-n-mt
133 from n-mt proteins. Interacting and non-interacting proteins were distinguished based on a
134 previously published [49] list of interacting n-mt proteins that includes OXPHOS subunits,
135 ribosomal subunits, aminoacyl-tRNA synthetases, and DNA and RNA polymerases, with
136 updates to match the most recent empirical evidence [4] (Supplemental Table 1. List of human
137 interacting n-mt genes used in this study, modified from Supplemental Table 1). Comparison of
138 OrthoFinder output to these two gene lists identified *X. birchmanni* proteins that have human and
139 mouse orthologs.
140
141 **Linkage disequilibrium calculation**
142 We used mitonuclear linkage disequilibrium (LD) as a metric for the association between
143 mitochondrial and nuclear alleles in hybrid populations. However, the mitogenome is fixed for
144 the haplotype from just one of the parental species (*X. cortezi*) in HUEX-STAC [38], so
145 mitonuclear LD could only be calculated for the CALL and CHAF populations (see below for
146 HUEX-STAC analysis methods). Calculations were performed using the D' and r^2 statistics [50].
147 In the equations that follow, A represents an n-mt locus and B the mitogenome, and pA and pB
148 represent their respective allele frequencies. pAB is the frequency of the haplotype AB across the
149 two sites. The A and B alleles were assigned to be from the same parental population so that
150 positive and negative values of LD indicate a higher frequency of parental or recombinant allele
151 combinations, respectively.

$$D = p_{AB} - p_A p_B$$

$$D' = \frac{D}{D_{max}}$$

where

$$D_{max} = \begin{cases} \min[p_A(1-p_B),(1-p_A)p_B]\ if\ D > 0 \\ \min[p_A p_B,(1-p_A)(1-p_B)]\ if\ D < 0 \end{cases}$$

and

$$r^2 = \frac{D^2}{p_A * (1-\ p_A) *\ p_B * (1-\ p_B)}$$

Because all mitochondrial genes are linked on the same non-recombining mitogenome, a single AIM was used (CALL and CHAF's AIM was at position 45 of the reference genome, GenBank accession CM071424.1, in *COX2*, and HUEX-STAC's AIM was at position 846 in *ATP6*) to identify the mitochondrial haplotype for each fish and to represent the mitogenome in all LD analyses. This approach was appropriate for our data, as all mtDNA AIMs had approximately the same posterior probability for ancestry assignment. Posterior probabilities were summed across each nuclear AIM as a proxy for frequency of the two alleles. The two LD statistics were calculated for each protein class separately in a custom R script. LD values were visualized using ggplot2 v3.4.0 and ggridges v0.5.4 in RStudio v2023.03.1+446.

LD values were then averaged across AIMs within each nuclear gene to limit pseudo-replication. A one-way ANOVA from the R stats v4.2.2 package (commands *lm* and *anova*) was used to determine whether protein class had a statistically significant impact on LD values. Summary statistics were collected using commands *n*, *mean*, and *sd* from the dplyr v1.1.0, base v4.2.2, and stats v4.2.2 packages, respectively. We also identified the AIMs and genes with the top 1% of LD values with a custom R script.

N-mt genes known to be involved in mitonuclear incompatibilities were identified from [44]. We tracked the location of these genes in the distribution of D', r^2, and allele frequency values for the respective populations. To determine how these genes compared to the mean value, their statistic was standardized to the mean by a Z-score calculation.

HUEX-STAC allele frequency calculation

173 The HUEX-STAC populations are both fixed for the *X. cortezi* mitochondrial haplotype, so to
174 see how mtDNA background affects selection on n-mt genes in this system, ancestry frequency
175 was calculated at interacting n-mt, non-interacting n-mt, and non-n-mt loci. These values were
176 averaged within each gene. Allele frequencies were identified and visualized using a custom R
177 script. A one-way ANOVA and summary statistics were computed following the same process as
178 described in the previous section.
179 A permutation test was then performed to determine whether significant results were
180 robust to violations of normality assumptions in ANOVA. Allele frequency values were
181 permutated by randomly reassigning them across genes 9999 times using base R v4.2.2 functions
182 in a custom R script. Permuted datasets were then used to calculate null expectations for
183 differences in allele frequencies between interacting n-mt and non-n-mt gene classes. The
184 observed difference was compared to the resulting distribution to determine the probability of
185 our result having arisen by chance.
186
187 **Data and code availability**
188 All code used in this project has been deposited in GitHub at
189 https://github.com/sakuster/sensitivity-of-genome-wide-mtnuc. Data can be accessed on figshare
190 (https://doi.org/10.6084/m9.figshare.c.7902011.v1).
191
192
193 **RESULTS**
194
195 **Identification of *Xiphophorus* n-mt sequences**
196 Of the 24,269 *Xiphophorus* annotated genes, 86.5% (20,988) were assigned to an orthogroup by
197 OrthoFinder. We found that 75.5% (13,264) of the 17,566 identified orthogroups contained at
198 least one *Xiphophorus* gene, and 69.4% (12,198) included at least one gene from all three species
199 (*X. birchmanni*, mouse, and human). Gene duplications during the divergence of these three
200 species resulted in some orthogroups with multiple *X. birchmanni* genes (24.1% of orthogroups,
201 or 4,234, contained 2 or more).
202 These orthogroups were used to classify *Xiphophorus* genes into our three functional
203 categories for each population (Table 2). MitoCarta 3.0 consists of 1,136 human and 1,140

204 mouse n-mt genes. All but three of the mouse n-mt genes (*Gm4984*, *Htd2*, and *Nsun4*) were
205 identified, and only one human n-mt gene (*NDUFB1*) was not identified in *X. birchmanni*.
206 Overall, 159 orthogroups were found to contain interacting n-mts, 822 to contain non-interacting
207 n-mts, and 15,545 to contain non-n-mts. These orthogroups had 184, 1,042, and 23,043 *X.*
208 *birchmanni* genes, respectively. We then determined which of these genes overlapped with the
209 AIMs in our three hybrid datasets (Table 2).
210
211 **Mitonuclear linkage disequilibrium is high but does not differ among gene classes in hybrid**
212 **populations with segregating mitochondrial haplotypes**
213 We sought to determine whether n-mt genes maintained higher mitonuclear LD in *Xiphophorus*
214 hybrids due to selection against mitonuclear incompatibilities. On average, mitonuclear LD
215 values were high across the entire nuclear genome in the CALL and CHAF hybrid populations
216 (Figure 1), with nuclear ancestry matching the mitochondrial haplotype (positive D' values). This
217 result is also illustrated by differences in average nuclear ancestry between individuals with
218 mitochondrial haplotypes from the two possible parent species (Figure 2). These strong
219 associations are likely due to population structure in these two populations [33,42].
220 Under a model of mitonuclear coevolution, we predicted that the interacting n-mt protein
221 class would have higher mitonuclear LD values compared to the other protein classes. However,
222 we did not find a significant difference between the protein classes for CALL (D': $p = 0.936$; r^2:
223 $p = 0.446$) or CHAF (D': $p = 0.994$; r^2: $p = 0.341$; summary statistics in Supplemental Table 2).
224 Therefore, it does not appear that mitochondrial haplotypes have preferentially maintained
225 associations with n-mt loci at a genome-wide scale in these hybridizing populations.
226 One possible explanation for the lack of observed differences in mitonuclear LD among
227 gene classes in the CALL and CHAF populations is that selection on a small number of n-mt
228 genes may not lead to detectable shifts in genome-wide averages but could still result in
229 individual loci with the strongest association with the mitogenome being enriched for interacting
230 n-mt genes. To test this possibility, we took the top 1% of all genes based on our association
231 metrics (D' and r^2) and tested for overrepresented gene classes. However, none of these gene sets
232 showed significant enrichment for interacting n-mt genes (Supplemental Table 4).
233 Previous analyses of mitonuclear incompatibilities in *X. birchmanni* x *X. malinche*
234 hybrids in the CALL population described nine incompatible nuclear genomic regions containing

235 10 n-mt genes [43,44]. Eight of these 10 genes were identified in our CALL analysis, and five
236 (*MTERF4*, *NDUFA13*, *NDUFS5*, *RMDN3*, *SMIM8*) of those eight genes were outliers for
237 mitonuclear LD (with a Z-score of >2 above the mean D' value) (Supplemental Table 3).
238 Therefore, although the top 1% genome-wide LD values did not show an overrepresentation of
239 n-mt genes, we do find that some genes of known biological importance exhibit signatures of
240 elevated mitonuclear LD. Two other genes (*LYRM2* and *MMUT*) were also above the mean but
241 not outliers by this definition, and the final gene (*UQCRC2*) was very close to the mean.
242 Repeating these analyses with r^2 values instead of D' values resulted in slight changes, but the
243 results were generally similar (Supplemental Table 3). Overall, our results show that n-mt genes
244 previously found to be involved in mitonuclear incompatibilities show the expected trend
245 towards higher mitonuclear LD values, but using D' or r^2 alone would have been insufficient to
246 identify these genes as the most significant outliers in the genome (Supplemental Figure 1,
247 Supplemental Figure 2).
248 To see whether these genetic incompatibilities showed similar signatures across
249 populations with varying mtDNA frequencies and different parental source species, we also
250 investigated mitonuclear associations for CHAF and HUEX-STAC (Supplemental Table 3,
251 Supplemental Figure 3, Supplemental Figure 4, Supplemental Figure 5). The CALL and CHAF
252 populations, which derive from the same parental species, exhibit a significant correlation in
253 mitonuclear LD (D') across this set of incompatibility genes (Pearson correlation; $r = 0.89$; $p =$
254 0.0012). Comparisons with allele frequencies in HUEX-STAC, which originates from a different
255 pairing of parental species, also show a positive trend, but the relationships are weaker and non-
256 significant for both CALL ($r = 0.35$; $p = 0.39$) and CHAF ($r = 0.68$; $p = 0.06$). Interestingly, the
257 two genes that exhibit an incompatibility with *X. malinche* mtDNA (*MTERF4, NDUFA13*) are
258 found to have some of the strongest LD signatures in CHAF (Supplemental Table 3), which has
259 a high frequency of *X. malinche* mtDNA (Table 1).
260
261 **Interacting n-mt genes show enriched mitochondrial match in an older hybrid population**
262 **fixed for a mitochondrial haplotype**
263 The HUEX-STAC populations are fixed for the *X. cortezi* mitochondrial haplotype, so we
264 expected to see *X. cortezi* ancestry disproportionately represented at n-mt loci. HUEX-STAC
265 individuals generally have a large percentage of *X. cortezi* ancestry across their entire nuclear

266 genome with a mean allele frequency of 82.7% ± a standard error of 0.001% (Figure 3). Non-
267 interacting n-mt and non-n-mt genes showed similar mean *X. cortezi* allele frequencies at 81.8 ±
268 0.07% and 82.7 ± 0.01%, respectively. Interacting n-mt genes had a higher mean allele
269 frequency (87.4 ± 0.01%, Figure 3). Thus, unlike in the CALL and CHAF populations,
270 individuals collected from the HUEX-STAC population did demonstrate a difference in allele
271 frequencies among gene classes (one way ANOVA: $p = 0.005$) and supported the predicted
272 higher degree of matched mitonuclear ancestries for interacting n-mt genes (Figure 3). This
273 result was also supported by a non-parametric permutation test (Supplemental Figure 6).
274
275
276 **DISCUSSION**
277
278 Because loci that have faster rates of evolution are expected to cause genetic incompatibilities
279 between diverging populations, mtDNA, which evolves on average 19× faster than nuclear DNA
280 in vertebrates [8], is proposed to play an important role in the evolution of incompatibilities [30].
281 Higher rates of genome evolution in mtDNA can result in corresponding compensatory
282 mutations in n-mt genes [17,19–21,51]. Some investigators have taken to splitting up the nuclear
283 genome into n-mt and non-n-mt portions to investigate the effects of mitonuclear interactions
284 and test for a disproportionate role of n-mt genes (Supplemental Table 5). However, it is unclear
285 how sensitive this approach is, so we applied it to a system known to have severe effects of
286 mitonuclear incompatibilities involving multiple loci.
287 Here, we specifically tested the prediction that the interacting n-mt gene class would
288 show the strongest "matched" association with mitochondrial haplotype. Despite the known role
289 of n-mt genes in hybrid incompatibilities between *X. birchmanni* and *X. malinche*, the CALL and
290 CHAF populations did not support this prediction. This negative result likely speaks to the
291 statistical power issues inherent in identifying mitonuclear interactions. The sensitivity of
292 genome-wide tests for signatures of selection acting preferentially on n-mt genes will likely be
293 affected by the number of genes under selection and/or where the n-mt genes are distributed in
294 the genome – factors that are also important for detecting loci involved in nuclear-nuclear hybrid
295 incompatibilities [52].

296 N-mt genes make up a small portion of the genome, and only a subset of those genes may
297 be participating in mitonuclear epistasis. Incompatibilities can potentially involve a few loci with
298 a large effect size or many loci of small effect. The analyses performed by Robles et al. [44]
299 identified 10 n-mt genes among nine genomic regions contributing to mitonuclear
300 incompatibilities in *Xiphophorus* species hybrid populations, one of which was CALL. These
301 genes represent only 6.8% of all interacting n-mt genes and 1.0% of all n-mt genes. Selection on
302 these few genes would have to be exceptionally strong and have had enough time to act to detect
303 a significant shift in the average for the entire interacting n-mt gene class. Therefore, even in
304 cases where n-mt genes are contributing to negative mitonuclear epistasis, the number of loci
305 may be too small to detect genome-wide effects with comparisons such as those implemented
306 here, which are inherently designed to detect polygenic contributions to incompatibilities. In
307 contrast, systems such as *Tigriopus* copepods, in which many genes are involved in mitonuclear
308 incompatibilities, have proven to be very powerful in advancing the field of mitonuclear genetics
309 and genomics [21,31].
310 Background nuclear-nuclear LD and the distribution of n-mt genes across the genome
311 will also affect the ability to detect a difference in mtDNA ancestry matching between n-mt and
312 non-n-mt genes. Any signal of this difference arising from selection on an individual n-mt target
313 could be diluted by hitchhiking at linked non-n-mt loci [53]. This effect would be particularly
314 problematic for populations that are the result of recent admixture events, such as CALL and
315 CHAF (initial admixture ~46 generations ago [35]), and therefore, have not had many
316 generations for recombination to break up haplotype blocks in the nuclear genome [41,54]. In
317 some cases, clustering of n-mt genes in specific regions of the genome could create strong targets
318 of selection that overcome this dilution effect. For example, genomic regions with a high density
319 of n-mt genes have been identified in the long-tailed finch [55] and the eastern yellow robin [56],
320 both of which have found evidence that n-mt genes disproportionately share ancestry with
321 mtDNA.
322 Additionally, not all mitonuclear incompatibilities are necessarily mediated by n-mt
323 genes. Instead, some incompatibilities may involve nuclear-encoded proteins that are not
324 targeted to mitochondria but nonetheless create genetic interactions with mtDNA through co-
325 functionality in biochemical and signaling pathways that span multiple cellular compartments

326 [51]. Any contributions of non-n-mt genes to mitonuclear incompatibilities would directly
327 undermine search strategies based on partitioning the genome into n-mt and non-n-mt categories.
328 Despite these challenges, one of the three hybrid populations we investigated (HUEX-
329 STAC) supported our prediction of enrichment in matched ancestry for n-mt genes. Notably, this
330 hybrid population is ~217 generations older than CALL and CHAF [33,35], and it is the only one
331 of the populations we investigated that is fixed for a mitochondrial haplotype from just one of the
332 parental species. Its older age means there has been more time for incompatible genotype
333 combinations to be purged from the population and for recombination to have broken up
334 haplotype blocks in the nuclear genome, thereby limiting the effects of hitchhiking. We therefore
335 speculate that the age of HUEX-STAC and the fixation of a mitochondrial haplotype may have
336 made selection on n-mt loci stronger and/or more effective. In such populations, half of the
337 possible mitonuclear combinations have been removed, so selection for matched ancestry would
338 consistently favor the same n-mt allele. In contrast, when mitochondrial haplotypes are still
339 segregating, the fitness effects of an n-mt allele may change depending on an individual's
340 mtDNA. Because mitochondrial and nuclear loci are not physically linked, very strong fitness
341 effects would be needed to preserve mitonuclear LD in segregating populations [57].
342 Although there is an enormous body of literature investigating mitonuclear interactions,
343 the number of studies that specifically test for preferential effects on associations involving n-mt
344 genes in the context of hybridization is much more limited. We identified 16 previous studies
345 that specifically tested this prediction (Supplemental Table 5). In these 16 studies, 23 different
346 hybrid populations or species with an ancient history of admixture were used, and nine of those
347 23 found support for preferential effects on n-mt genes. However, we found examples that both
348 support and oppose the hypothesis that fixation of a mitochondrial haplotype strengthens
349 selection on n-mt genes. Expanding the availability of such genomic datasets from hybrid
350 systems and performing a formal meta-analysis would be important future directions to assess
351 relevant variables, such as whether mitochondrial haplotypes are segregating in the population,
352 the age of the hybrid population, the number of loci involved and their effect sizes, and the
353 genetic structure of the population. Reanalysis of publicly available genome-level data from
354 hybrid populations may also be a valuable approach to control for method of ancestry inference
355 and metric of association between mtDNA and nuclear loci.

356 Another variable that differentiates the HUEX-STAC population is the parental makeup
357 of the hybrids. Although geographically distinct, CHAF and CALL share parental species (*X.*
358 *birchmanni* and *X. malinche*). In contrast, HUEX-STAC was formed by hybridization between
359 *X. birchmanni* and *X. cortezi*. These three species demonstrate mitonuclear discordance, as *X.*
360 *malinche* and *X. cortezi* mitogenomes are more closely related to each other, but the *X. malinche*
361 nuclear genome is sister to *X. birchmanni* [58]. Therefore, hybrids in the HUEX-STAC
362 population have nuclear genomes that are more divergent, on average, from one another than in
363 the CALL and CHAF populations. Furthermore, recent work has shown that the *X. cortezi*, *X.*
364 *malinche*, and *X. birchmanni* species have a complex evolutionary history that involves the
365 introgression of n-mt genes *ndufs5* and *ndufa13* from *X. malinche* into *X. cortezi*, which
366 produces similar incompatibility phenotypes in hybrids between *X. cortezi* and *X. birchmanni*
367 [42]. HUEX-STAC hybrids are then faced with nuclear alleles that are farther diverged from one
368 another in combination with a prior history of introgression contributing to mitonuclear
369 incompatibilities. These two factors may cause stronger and more widespread incompatibilities
370 involving n-mt genes, resulting in stronger selection for matched mitonuclear ancestry. Given
371 these different demographic histories, each of the hybrid populations we analyze here is expected
372 to have numerous unique features, so caution is warranted in attributing any of our results to a
373 single feature (such as the fixed mitochondrial haplotype in HUEX-STAC). Nevertheless, we
374 view the potential role of fixed vs. segregating mitochondrial haplotypes to be an important area
375 of future research given the small number of studies conducted to date.
376 Overall, our findings suggest that splitting up the genome into n-mt and non-n-mt gene
377 classes may be an underpowered method to detect effects of mitonuclear epistasis in recently
378 admixed populations, even when there are clear biological effects of said epistasis. Although
379 such tests for genome-wide signatures may offer a relatively unbiased approach, they may be
380 susceptible to false negatives. Therefore, important contributions of mitonuclear interactions to
381 speciation may remain unrecognized in many systems.
382
383
384 **ACKNOWLEDGEMENTS**
385

SAK was supported by an NIH T32 fellowship (GM132057) and an NSF GRFP (006784). This work was also supported by NIH grants to DBS (R35GM148134) and JCH (R35GM142836) and an EDGE - NSF grant (IOS-2421661) to JCH and MS.

TABLES

Table 1. Summary of Xiphophorus hybrid populations used in this study.

Site Abbreviation	**CALL (Calnali Low)**	**CHAF (Chahuaco Falls)**	**HUEX-STAC (Huextetitla and Santa Cruz)**
Population Site	Calnali River	Calnali River	Santa Cruz River
Major Parent Species	*X. birchmanni*	*X. malinche*	*X. cortezi*
Minor Parent Species	*X. malinche*	*X. birchmanni*	*X. birchmanni*
Number of Individuals	281	250	255
Mitochondrial Haplotype Frequency[a]	63.7% *X. birchmanni*	93.6% *X. malinche*	99.6% *X. cortezi*
Average Major Parent Nuclear Ancestry	52.2%	68.2%	84.6%
Nuclear Ancestry Informative Sites	629,430	628,788	689,966
Mitochondrial Ancestry Informative Sites	93	93	16
Estimated Generations Since Initial Admixture	46 ± 1 [40]	46 ± 1 [40]	263 [33]

[a]Haplotype frequencies for each population were calculated following filtering of mitochondrial AIMs with low-confidence genotypes (posterior probability between 0.2 and 0.8). Filtering resulted in sample sizes slightly lower than the Number of Individuals row: CALL n = 278, CHAF n = 249, HUEX-STAC n = 244.

Table 2. AIMs and genes belonging to each protein class. The orthogroup row indicates how many orthogroups containing X. birchmanni *protein sequences were found for each gene class.*

Population	**Interacting n-mt**	**Non-interacting n-mt**	**Non-n-mt**
Orthogroups	159	822	15,545
CALL AIMs (unique genes)	954 (148)	8,580 (883)	287,834 (19,312)
CHAF AIMs (unique genes)	952 (148)	8,575 (883)	287,592 (19,307)
HUEX-STAC AIMs (unique genes)	997 (162)	9,399 (947)	319,679 (20,544)

FIGURES

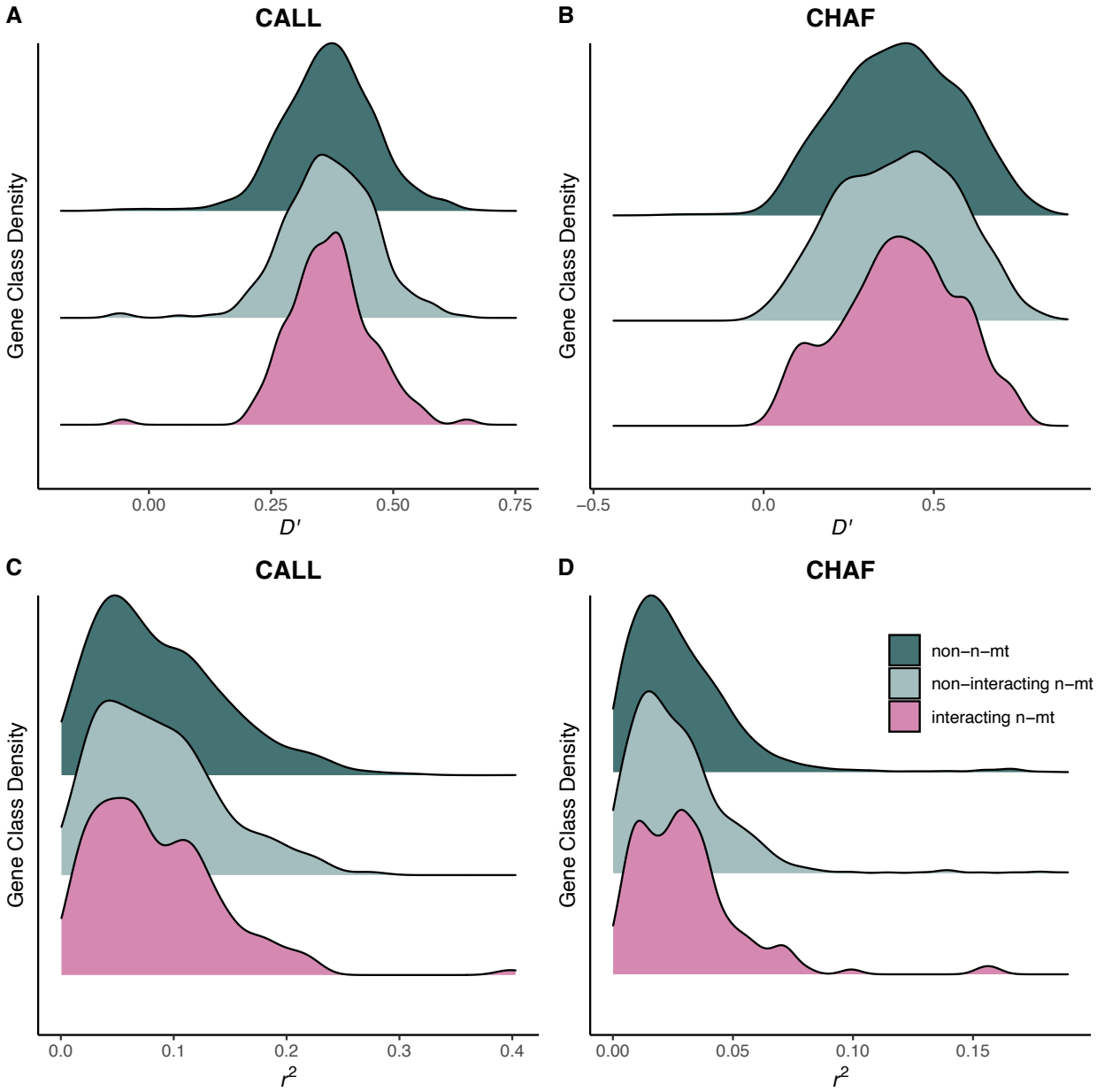


Figure 1. Distribution of mitonuclear LD values for each gene class in the CALL (A and B) and CHAF (C and D) populations. LD is calculated as D′ *in panels A and C and as* r^2 *in panels B and D. Note the change of scale on the x-axis across populations and LD statistics. P-value for a one-way ANOVA testing if gene class affects LD value: CALL* D′ *0.94 (A), CHAF* D′ *0.99 (B), CALL* r^2 *0.45 (C), CHAF* r^2 *0.34 (D). Summary statistics are reported in Supplemental Table 2.*

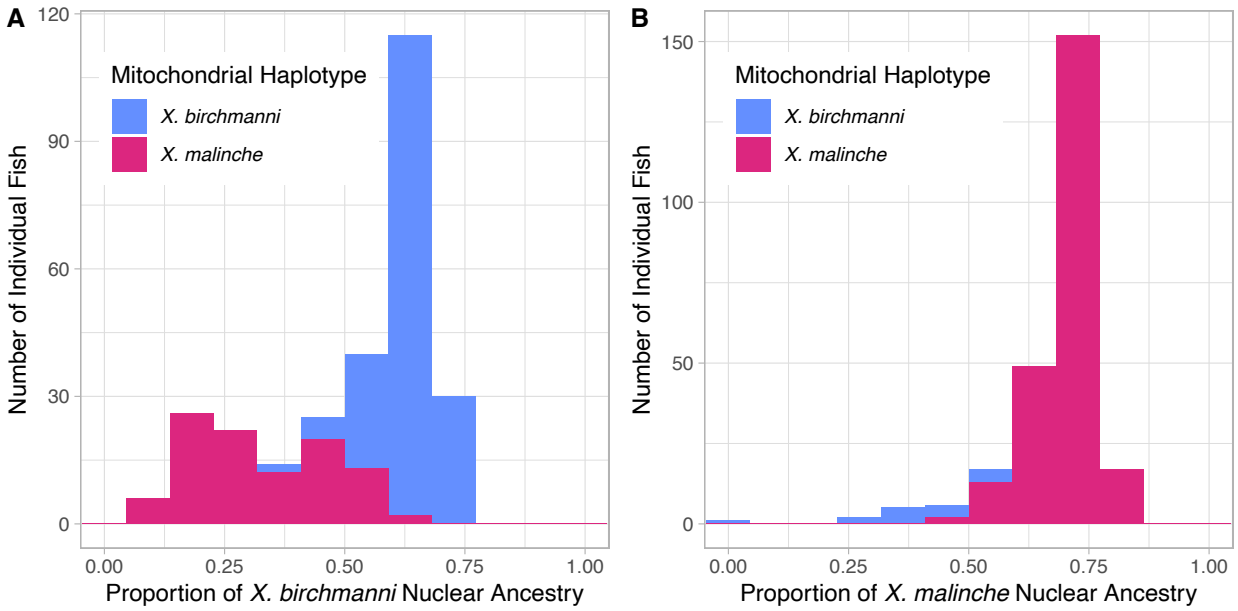


Figure 2. Stacked bar histogram for the proportion of the nuclear genome derived from the major parent for each fish in A) CALL and B) CHAF. Proportion is calculated as the sum of all major parent AIMs divided by number of AIMs. Each fish is colored by which of the two segregating mitochondrial haplotypes it has (blue = X. birchmanni*, fuchsia =* X. malinche*). Note that y-axis upper limits differ between the panels.*

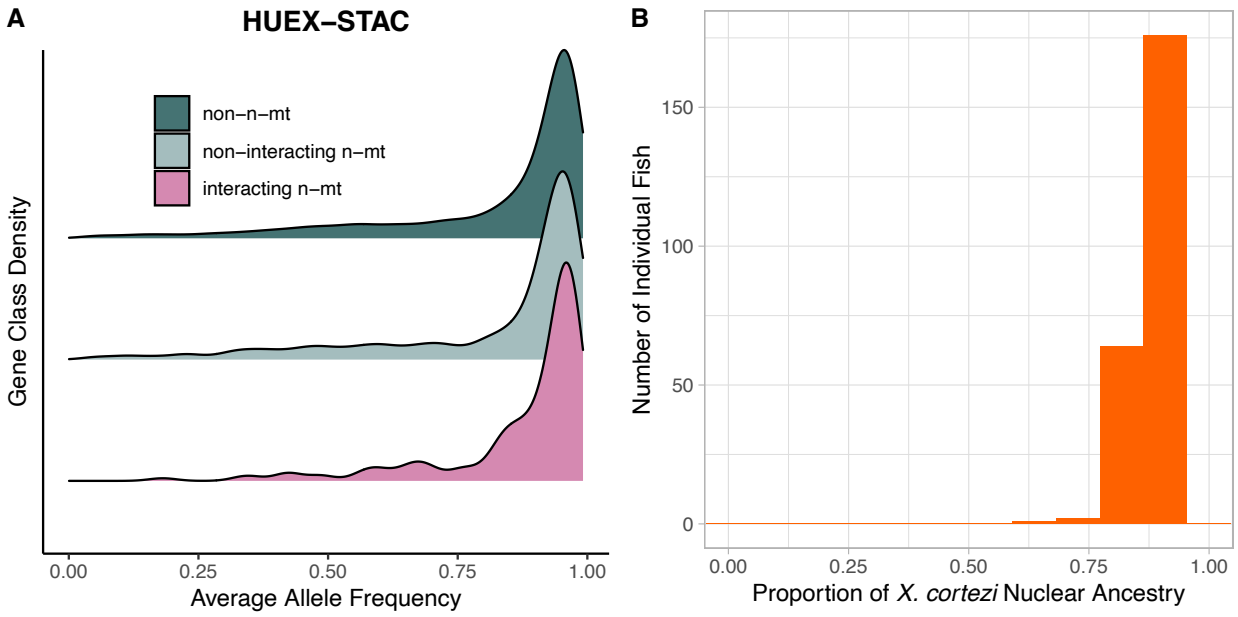
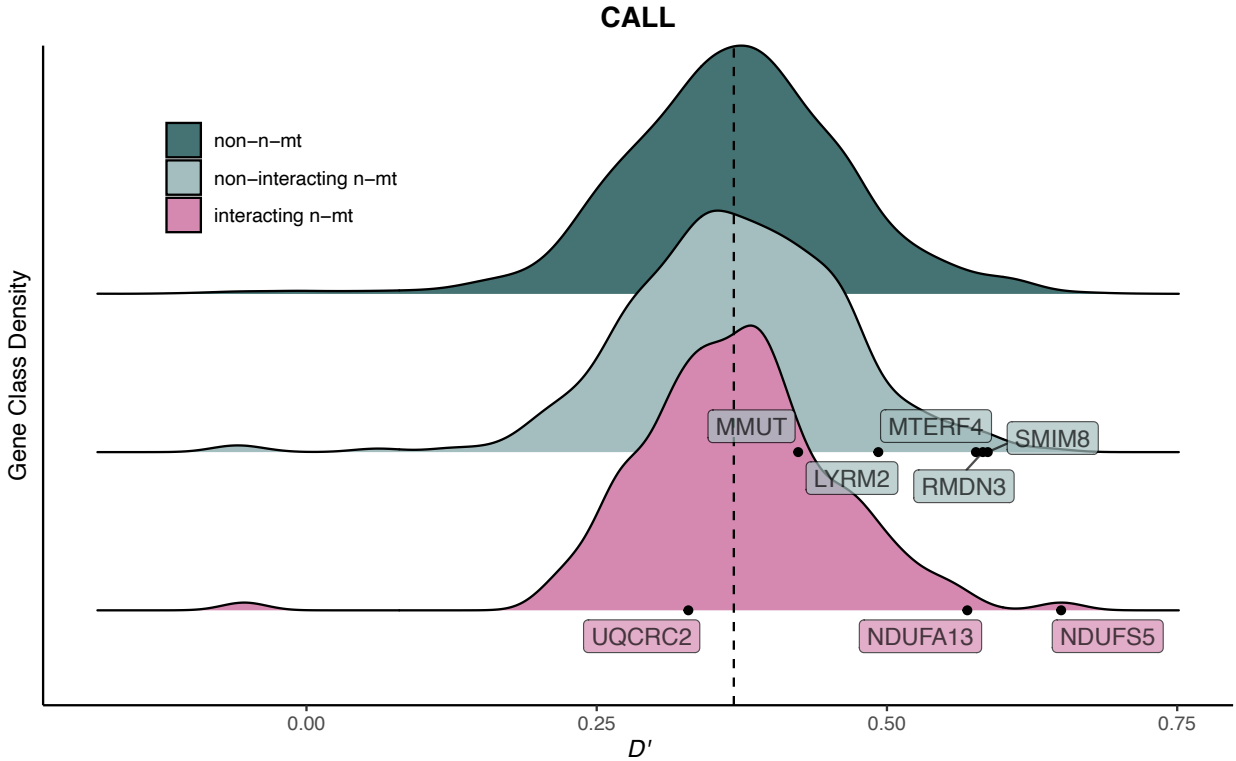
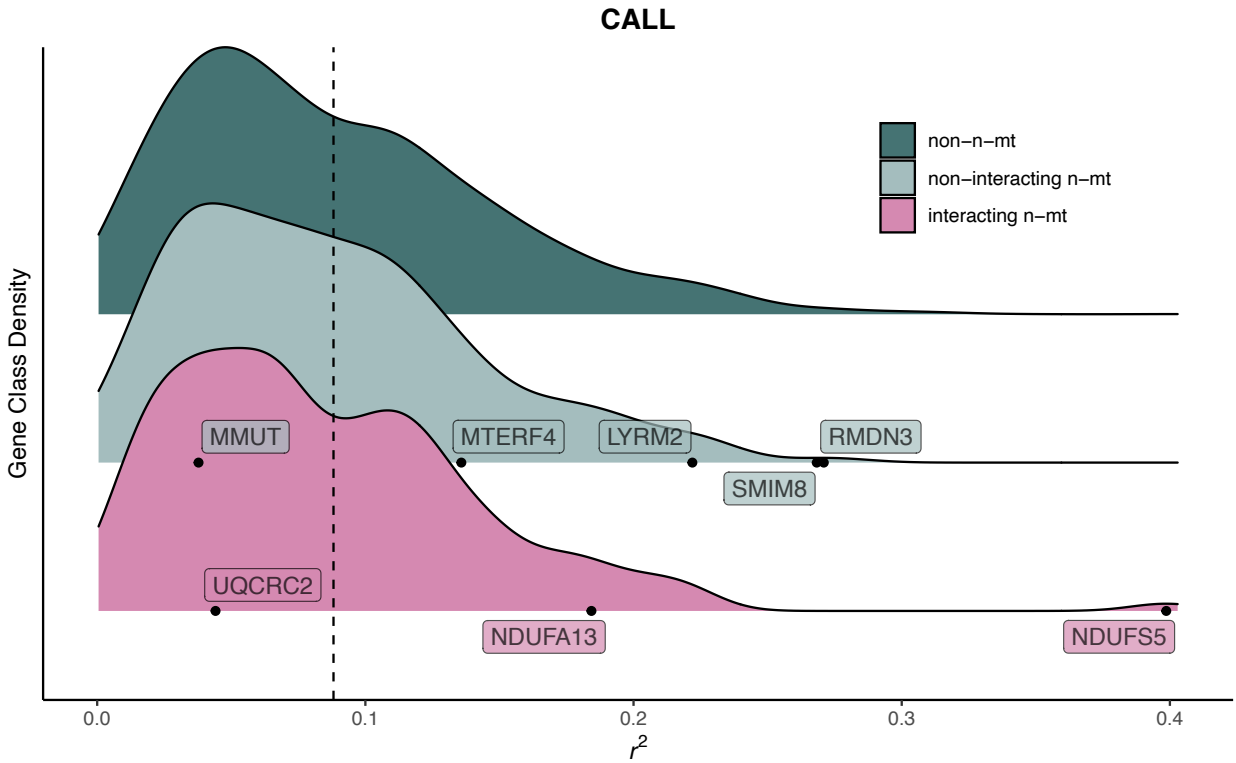


Figure 3. HUEX-STAC analyses for A) allele frequencies for each gene class and B) nuclear genome-wide ancestry for the HUEX-STAC population, a hybrid population fixed for the X. cortezi *mitogenome.*

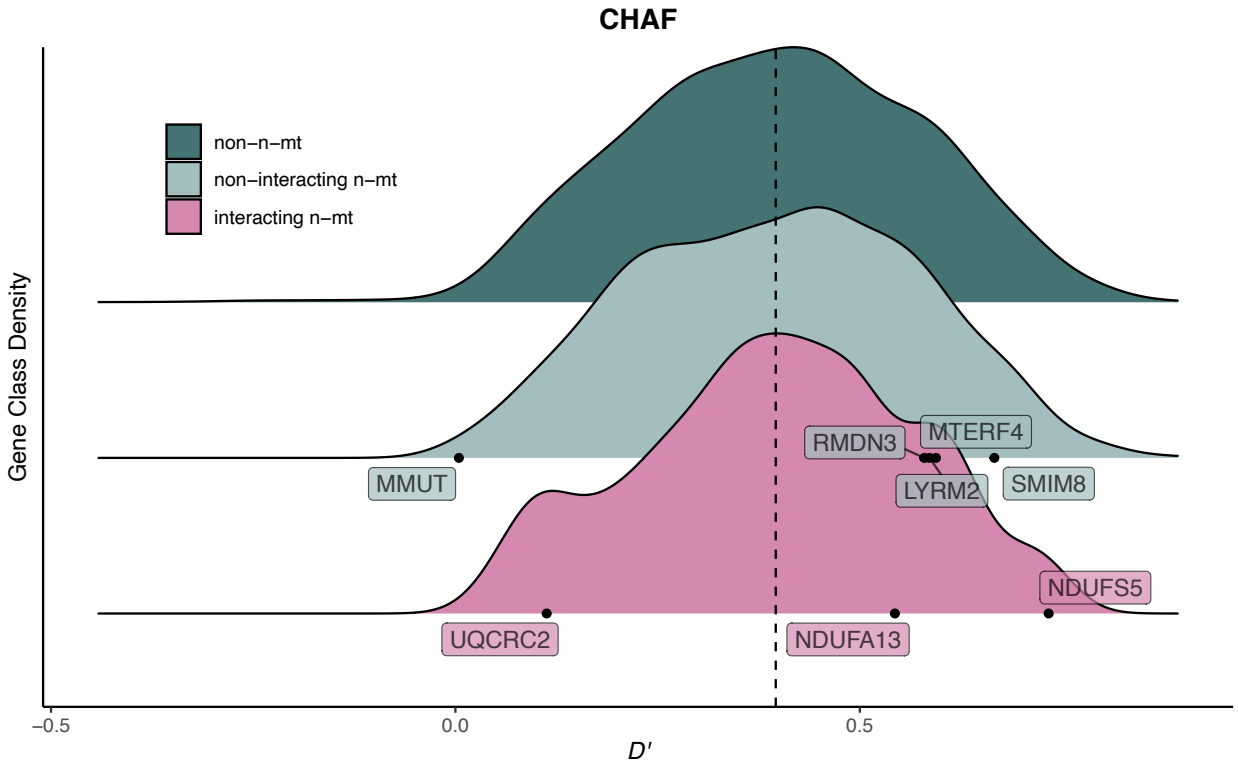
SUPPLEMENTAL FIGURES



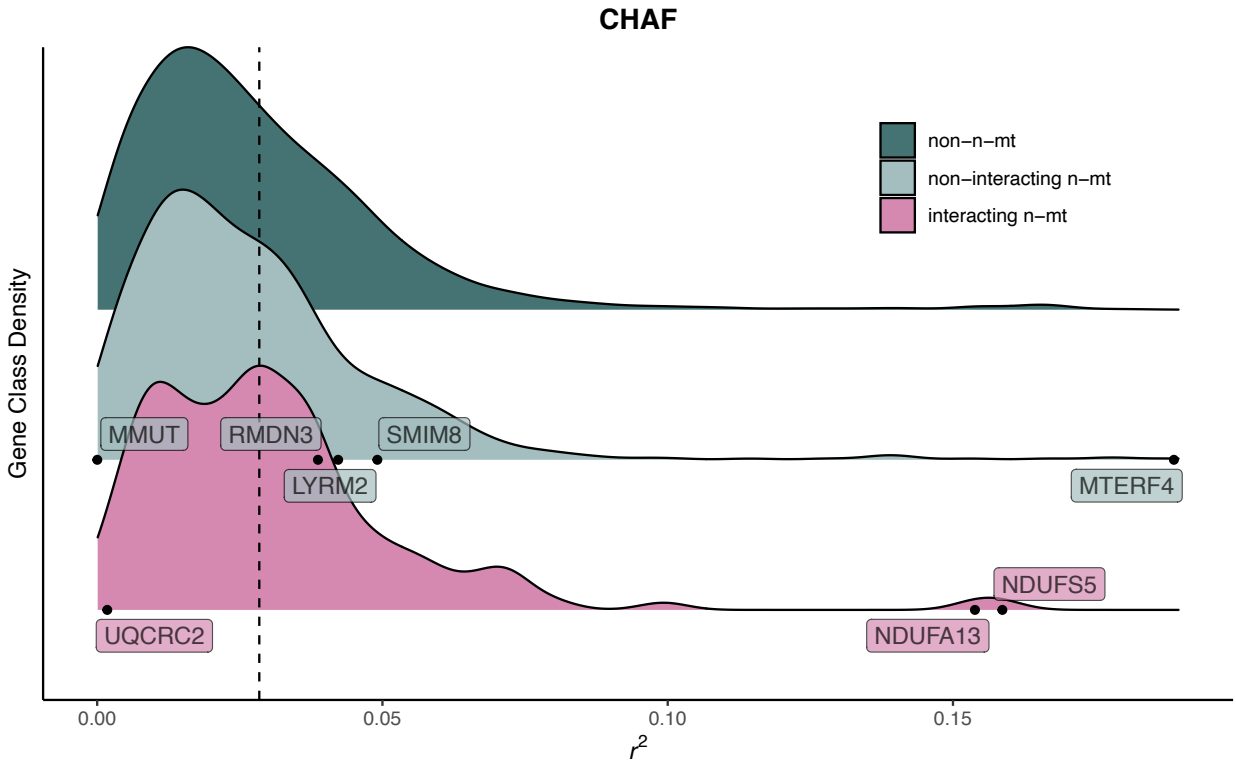
Supplemental Figure 1. Location of previously identified incompatibility genes [44] in the CALL population D' *values. Dashed line indicates genome-wide mean* D'.



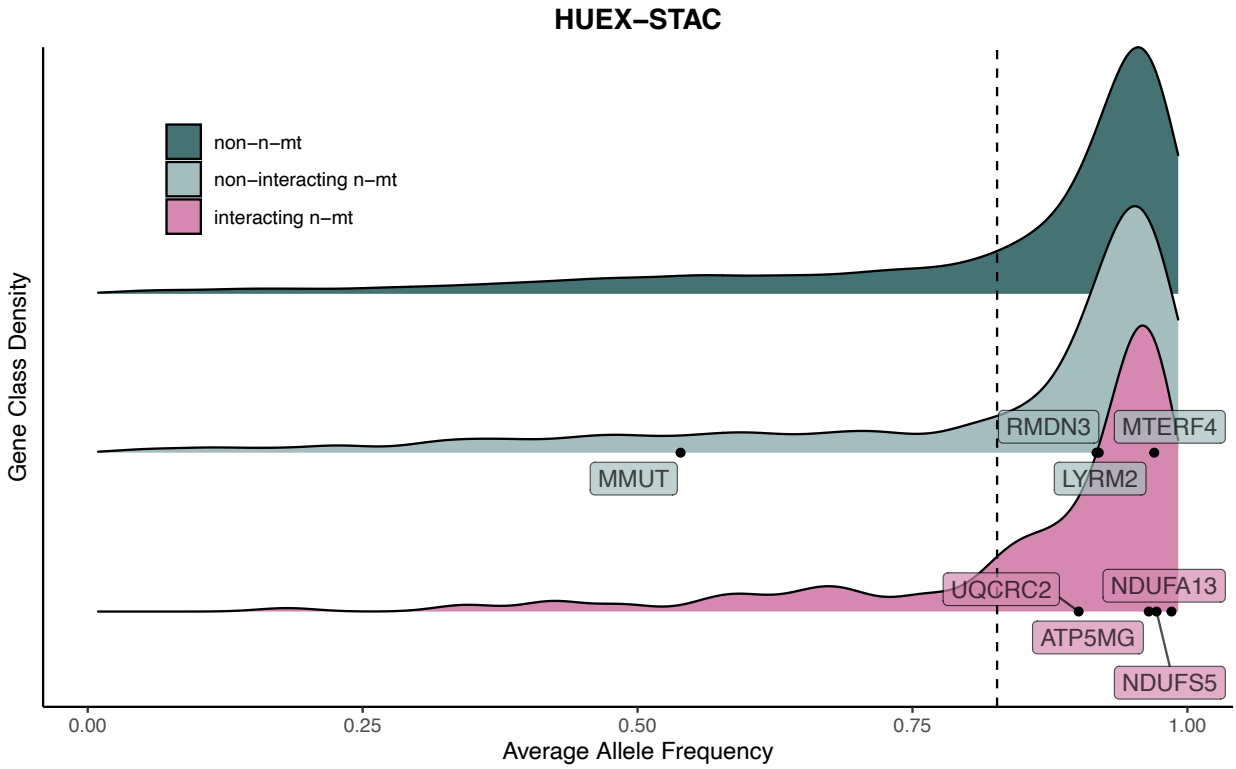
Supplemental Figure 2. Location of previously identified incompatibility genes [44] in the CALL population r^2 *values. Dashed line indicates genome-wide mean* r^2.



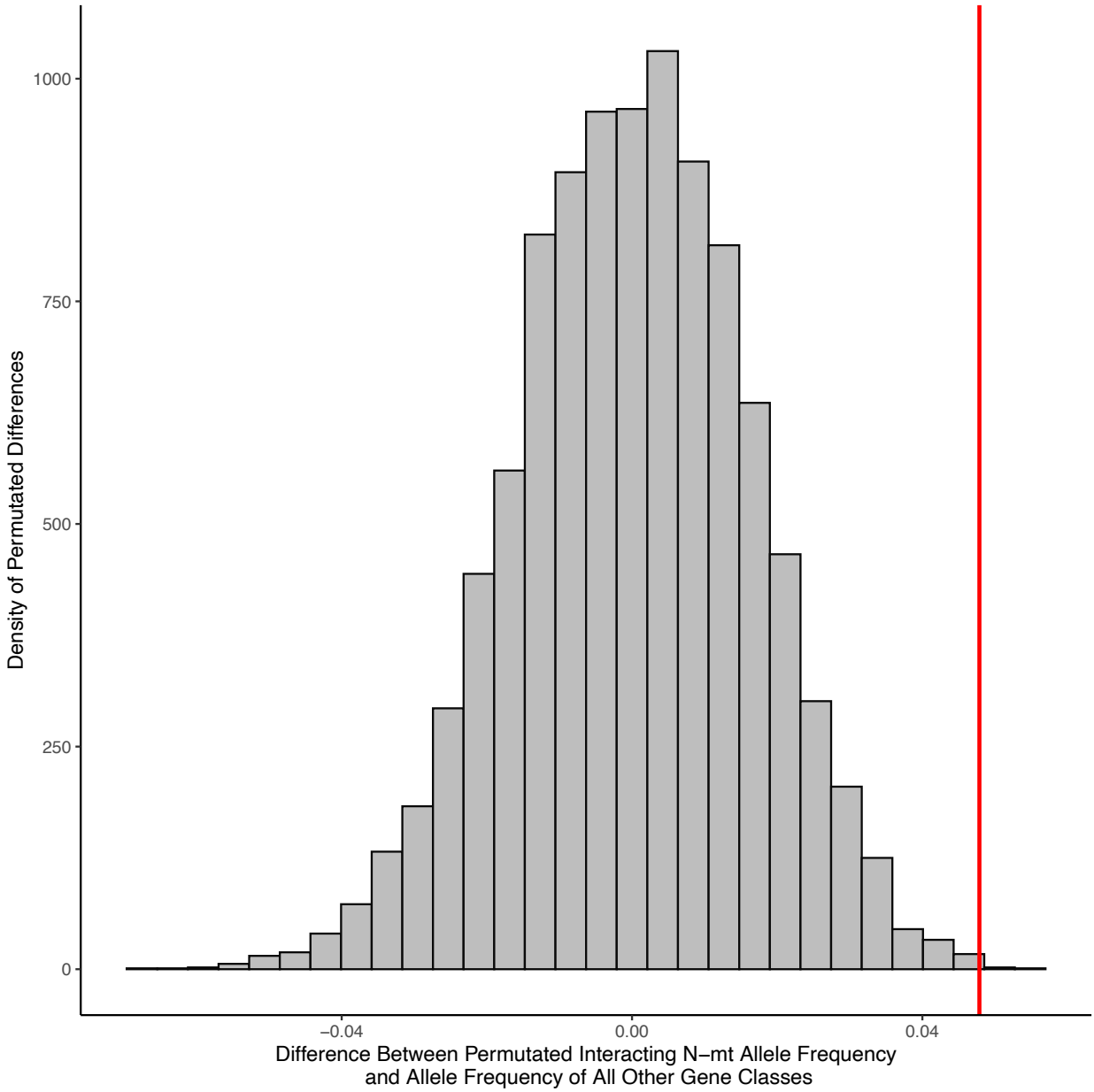
Supplemental Figure 3. Location of previously identified incompatibility genes [44] in the CHAF population D' *values. Dashed line indicates genome-wide mean* D'.



Supplemental Figure 4. Location of previously identified incompatibility genes [44]in the CHAF population r^2*values. Dashed line indicates genome-wide mean* r^2*.*



Supplemental Figure 5. Location of previously identified incompatibility genes [44] in the HUEX-STAC population X. cortezi *allele frequency values. Dashed line indicates genome-wide mean allele frequency.*



Supplemental Figure 6. Histogram demonstrating permuted differences in mean allele frequency between all interacting n-mts and the mean of all other genes. Red line represents the observed difference in interacting n-mt gene class and all other classes. In this permutation test, we found the interacting n-mt gene class to be different from the other two gene classes ($p = 5 \times 10^{-4}$*), supporting the parametric analysis (ANOVA) in this population.*

SUPPLEMENTAL TABLES

Supplemental Table 1. List of human interacting n-mt genes used in this study, modified from [49].

Uniprot ID	Entrez symbol	Gene symbol	Gene Name
P82663	64432	MRPS25	mitochondrial ribosomal protein S25
O15239	4694	NDUFA1	NADH:ubiquinone oxidoreductase subunit A1
Q9Y3B7	65003	MRPL11	mitochondrial ribosomal protein L11
O43678	4695	NDUFA2	NADH:ubiquinone oxidoreductase subunit A2
Q96EL2	64951	MRPS24	mitochondrial ribosomal protein S24
O95167	4696	NDUFA3	NADH:ubiquinone oxidoreductase subunit A3
Q9BYD2	65005	MRPL9	mitochondrial ribosomal protein L9
Q9BYD6	65008	MRPL1	mitochondrial ribosomal protein L1
Q9Y6G3	28977	MRPL42	mitochondrial ribosomal protein L42
P14854	1340	COX6B1	cytochrome c oxidase subunit 6B1
Q9Y2R9	51081	MRPS7	mitochondrial ribosomal protein S7
Q9NVS2	55168	MRPS18A	mitochondrial ribosomal protein S18A
P12074	1337	COX6A1	cytochrome c oxidase subunit 6A1
O00217	4728	NDUFS8	NADH:ubiquinone oxidoreductase core subunit S8
Q9BYN8	64949	MRPS26	mitochondrial ribosomal protein S26
P19404	4729	NDUFV2	NADH:ubiquinone oxidoreductase core subunit V2
Q86Y39	126328	NDUFA11	NADH:ubiquinone oxidoreductase subunit A11
Q9Y676	28973	MRPS18B	mitochondrial ribosomal protein S18B
O14949	27089	UQCRQ	ubiquinol-cytochrome c reductase complex III subunit VII
P30049	513	ATP5F1D	ATP synthase F1 subunit delta
P56381	514	ATP5F1E	ATP synthase F1 subunit epsilon
Q9H9J2	65080	MRPL44	mitochondrial ribosomal protein L44
P24539	515	ATP5PB	ATP synthase peripheral stalk-membrane subunit b
O75306	4720	NDUFS2	NADH:ubiquinone oxidoreductase core subunit S2
Q9UI09	55967	NDUFA12	NADH:ubiquinone oxidoreductase subunit A12
Q96GC5	51642	MRPL48	mitochondrial ribosomal protein L48
P05496	516	ATP5MC1	ATP synthase membrane subunit c locus 1
Q06055	517	ATP5MC2	ATP synthase membrane subunit c locus 2
O75489	4722	NDUFS3	NADH:ubiquinone oxidoreductase core subunit S3
P09001	11222	MRPL3	mitochondrial ribosomal protein L3
P48201	518	ATP5MC3	ATP synthase membrane subunit c locus 3
P49821	4723	NDUFV1	NADH:ubiquinone oxidoreductase core subunit V1
Q5JTZ9	57505	AARS2	alanyl-tRNA synthetase 2, mitochondrial
O43181	4724	NDUFS4	NADH:ubiquinone oxidoreductase subunit S4
O43920	4725	NDUFS5	NADH:ubiquinone oxidoreductase subunit S5
Q9Y3D9	51649	MRPS23	mitochondrial ribosomal protein S23
Q02221	1339	COX6A2	cytochrome c oxidase subunit 6A2
O75380	4726	NDUFS6	NADH:ubiquinone oxidoreductase subunit S6
P82914	64960	MRPS15	mitochondrial ribosomal protein S15
Q9NRX2	63875	MRPL17	mitochondrial ribosomal protein L17
P82912	64963	MRPS11	mitochondrial ribosomal protein S11
P15954	1350	COX7C	cytochrome c oxidase subunit 7C
P82933	64965	MRPS9	mitochondrial ribosomal protein S9
P82932	64968	MRPS6	mitochondrial ribosomal protein S6
Q9BYC9	55052	MRPL20	mitochondrial ribosomal protein L20
P82664	55173	MRPS10	mitochondrial ribosomal protein S10
P82675	64969	MRPS5	mitochondrial ribosomal protein S5
P24311	1349	COX7B	cytochrome c oxidase subunit 7B
O95363	10667	FARS2	phenylalanyl-tRNA synthetase 2, mitochondrial
P14406	1347	COX7A2	cytochrome c oxidase subunit 7A2
Q9Y2R5	51373	MRPS17	mitochondrial ribosomal protein S17
P24310	1346	COX7A1	cytochrome c oxidase subunit 7A1
P56385	521	ATP5ME	ATP synthase membrane subunit e

P09669	1345	COX6C	cytochrome c oxidase subunit 6C
Q9Y291	51650	MRPS33	mitochondrial ribosomal protein S33
P18859	522	ATP5PF	ATP synthase peripheral stalk subunit F6
Q9BZE1	51253	MRPL37	mitochondrial ribosomal protein L37
P82673	60488	MRPS35	mitochondrial ribosomal protein S35
Q9NSE4	55699	IARS2	isoleucyl-tRNA synthetase 2, mitochondrial
Q9UDW1	29796	UQCR10	ubiquinol-cytochrome c reductase, complex III subunit X
Q4U2R6	51258	MRPL51	mitochondrial ribosomal protein L51
O00483	4697	NDUFA4	NDUFA4 mitochondrial complex associated
Q9UHN1	11232	POLG2	DNA polymerase gamma 2, accessory subunit
Q16718	4698	NDUFA5	NADH:ubiquinone oxidoreductase subunit A5
P56181	4731	NDUFV3	NADH:ubiquinone oxidoreductase subunit V3
Q9BRJ2	84311	MRPL45	mitochondrial ribosomal protein L45
Q7Z2W9	219927	MRPL21	mitochondrial ribosomal protein L21
P54098	5428	POLG	DNA polymerase gamma, catalytic subunit
P82650	56945	MRPS22	mitochondrial ribosomal protein S22
O00411	5442	POLRMT	RNA polymerase mitochondrial
Q8IXM3	64975	MRPL41	mitochondrial ribosomal protein L41
Q9NQ50	64976	MRPL40	mitochondrial ribosomal protein L40
Q9P0J6	64979	MRPL36	mitochondrial ribosomal protein L36
Q96DV4	64978	MRPL38	mitochondrial ribosomal protein L38
Q9Y2Q9	28957	MRPS28	mitochondrial ribosomal protein S28
Q99766	27109	DMAC2L	distal membrane arm assembly component 2 like
Q9HD33	57129	MRPL47	mitochondrial ribosomal protein L47
Q8TCC3	51263	MRPL30	mitochondrial ribosomal protein L30
Q9Y3D3	51021	MRPS16	mitochondrial ribosomal protein S16
Q9P015	29088	MRPL15	mitochondrial ribosomal protein L15
P25705	498	ATP5F1A	ATP synthase F1 subunit alpha
Q8N5N7	54534	MRPL50	mitochondrial ribosomal protein L50
Q9Y3D5	51023	MRPS18C	mitochondrial ribosomal protein S18C
Q9P0M9	51264	MRPL27	mitochondrial ribosomal protein L27
Q9UGM6	10352	WARS2	tryptophanyl tRNA synthetase 2, mitochondrial
Q9BQC6	78988	MRPL57	mitochondrial ribosomal protein L57
P48047	539	ATP5PO	ATP synthase peripheral stalk subunit OSCP
O75251	374291	NDUFS7	NADH:ubiquinone oxidoreductase core subunit S7
Q9NP81	54938	SARS2	seryl-tRNA synthetase 2, mitochondrial
O75947	10476	ATP5PD	ATP synthase peripheral stalk subunit d
Q9NX14	54539	NDUFB11	NADH:ubiquinone oxidoreductase subunit B11
O75964	10632	ATP5MG	ATP synthase membrane subunit g
P49590	23438	HARS2	histidyl-tRNA synthetase 2, mitochondrial
P20674	9377	COX5A	cytochrome c oxidase subunit 5A
Q9BQ48	64981	MRPL34	mitochondrial ribosomal protein L34
Q92665	10240	MRPS31	mitochondrial ribosomal protein S31
Q9BYC8	64983	MRPL32	mitochondrial ribosomal protein L32
Q7L3T8	25973	PARS2	prolyl-tRNA synthetase 2, mitochondrial
O60783	63931	MRPS14	mitochondrial ribosomal protein S14
Q6YFQ2	125965	COX6B2	cytochrome c oxidase subunit 6B2
Q7Z7H8	124995	MRPL10	mitochondrial ribosomal protein L10
Q9NYK5	54148	MRPL39	mitochondrial ribosomal protein L39
Q15031	23395	LARS2	leucyl-tRNA synthetase 2, mitochondrial
Q7Z4L0	341947	COX8C	cytochrome c oxidase subunit 8C
Q9NWU5	29093	MRPL22	mitochondrial ribosomal protein L22
P49406	9801	MRPL19	mitochondrial ribosomal protein L19
Q9NP92	10884	MRPS30	mitochondrial ribosomal protein S30
P41250	2617	GARS1	glycyl-tRNA synthetase 1
Q9NX20	54948	MRPL16	mitochondrial ribosomal protein L16
Q9NZE8	51318	MRPL35	mitochondrial ribosomal protein L35
Q16540	6150	MRPL23	mitochondrial ribosomal protein L23
Q9Y2S7	26073	POLDIP2	DNA polymerase delta interacting protein 2

O14957	10975	UQCR11	ubiquinol-cytochrome c reductase, complex III subunit XI
Q96I59	79731	NARS2	asparaginyl-tRNA synthetase 2, mitochondrial
Q96GW9	92935	MARS2	methionyl-tRNA synthetase 2, mitochondrial
Q13084	10573	MRPL28	mitochondrial ribosomal protein L28
Q7Z7F7	128308	MRPL55	mitochondrial ribosomal protein L55
P08574	1537	CYC1	cytochrome c1
O75394	9553	MRPL33	mitochondrial ribosomal protein L33
P56134	9551	ATP5MF	ATP synthase membrane subunit f
Q5JPH6	124454	EARS2	glutamyl-tRNA synthetase 2, mitochondrial
Q5T160	57038	RARS2	arginyl-tRNA synthetase 2, mitochondrial
Q9H0U6	29074	MRPL18	mitochondrial ribosomal protein L18
P22695	7385	UQCRC2	ubiquinol-cytochrome c reductase core protein 2
P31930	7384	UQCRC1	ubiquinol-cytochrome c reductase core protein 1
P47985	7386	UQCRFS1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1
P07919	7388	UQCRH	ubiquinol-cytochrome c reductase hinge protein
Q9BYD1	28998	MRPL13	mitochondrial ribosomal protein L13
Q96A35	79590	MRPL24	mitochondrial ribosomal protein L24
Q92552	23107	MRPS27	mitochondrial ribosomal protein S27
O95299	4705	NDUFA10	NADH:ubiquinone oxidoreductase subunit A10
Q6P1L8	64928	MRPL14	mitochondrial ribosomal protein L14
Q9BW92	80222	TARS2	threonyl-tRNA synthetase 2, mitochondrial
O14561	4706	NDUFAB1	NADH:ubiquinone oxidoreductase subunit AB1
Q96KJ9	84701	COX4I2	cytochrome c oxidase subunit 4I2
O95178	4708	NDUFB2	NADH:ubiquinone oxidoreductase subunit B2
O43676	4709	NDUFB3	NADH:ubiquinone oxidoreductase subunit B3
Q9Y2Z4	51067	YARS2	tyrosyl-tRNA synthetase 2
Q5T653	51069	MRPL2	mitochondrial ribosomal protein L2
Q8N983	84545	MRPL43	mitochondrial ribosomal protein L43
Q9Y375	51103	NDUFAF1	NADH:ubiquinone oxidoreductase complex assembly factor 1
P56556	4700	NDUFA6	NADH:ubiquinone oxidoreductase subunit A6
O95182	4701	NDUFA7	NADH:ubiquinone oxidoreductase subunit A7
P51970	4702	NDUFA8	NADH:ubiquinone oxidoreductase subunit A8
Q15046	3735	KARS1	lysyl-tRNA synthetase 1
Q86TS9	122704	MRPL52	mitochondrial ribosomal protein L52
Q16795	4704	NDUFA9	NADH:ubiquinone oxidoreductase subunit A9
P82930	65993	MRPS34	mitochondrial ribosomal protein S34
Q5ST30	57176	VARS2	valyl-tRNA synthetase 2, mitochondrial
Q9H2W6	26589	MRPL46	mitochondrial ribosomal protein L46
P52815	6182	MRPL12	mitochondrial ribosomal protein L12
P82921	54460	MRPS21	mitochondrial ribosomal protein S21
O15235	6183	MRPS12	mitochondrial ribosomal protein S12
P13073	1327	COX4I1	cytochrome c oxidase subunit 4I1
O96000	4716	NDUFB10	NADH:ubiquinone oxidoreductase subunit B10
Q6PI48	55157	DARS2	aspartyl-tRNA synthetase 2, mitochondrial
O43677	4717	NDUFC1	NADH:ubiquinone oxidoreductase subunit C1
Q9BYD3	51073	MRPL4	mitochondrial ribosomal protein L4
Q96EL3	116540	MRPL53	mitochondrial ribosomal protein L53
O95298	4718	NDUFC2	NADH:ubiquinone oxidoreductase subunit C2
P28331	4719	NDUFS1	NADH:ubiquinone oxidoreductase core subunit S1
Q13405	740	MRPL49	mitochondrial ribosomal protein L49
Q6P161	116541	MRPL54	mitochondrial ribosomal protein L54
Q9P0J0	51079	NDUFA13	NADH:ubiquinone oxidoreductase subunit A13
O95168	4710	NDUFB4	NADH:ubiquinone oxidoreductase subunit B4
O43674	4711	NDUFB5	NADH:ubiquinone oxidoreductase subunit B5
Q9Y399	51116	MRPS2	mitochondrial ribosomal protein S2
P06576	506	ATP5F1B	ATP synthase F1 subunit beta
O95139	4712	NDUFB6	NADH:ubiquinone oxidoreductase subunit B6
P17568	4713	NDUFB7	NADH:ubiquinone oxidoreductase subunit B7
P10606	1329	COX5B	cytochrome c oxidase subunit 5B

O95169	4714	NDUFB8	NADH:ubiquinone oxidoreductase subunit B8
P36542	509	ATP5F1C	ATP synthase F1 subunit gamma
Q9Y6M9	4715	NDUFB9	NADH:ubiquinone oxidoreductase subunit B9

Supplemental Table 2. Summary statistics of mitonuclear association values for each hybrid population.

Population	**Statistic**	**Gene Class**	**Count**	**Mean**	**Standard Deviation**
CALL	*D′*	Interacting n-mt	148	0.370	0.086
		Non-interacting n-mt	883	0.369	0.094
		Non-n-mt	19,312	0.368	0.097
	r^2	Interacting n-mt	148	0.084	0.058
		Non-interacting n-mt	883	0.086	0.054
		Non-n-mt	19,312	0.088	0.058
CHAF	*D′*	Interacting n-mt	148	0.397	0.169
		Non-interacting n-mt	883	0.396	0.174
		Non-n-mt	19,307	0.396	0.177
	r^2	Interacting n-mt	148	0.030	0.024
		Non-interacting n-mt	883	0.028	0.022
		Non-n-mt	19.307	0.028	0.023
HUEX-STAC	Allele Frequency	Interacting n-mt	162	0.874	0.150
		Non-interacting n-mt	947	0.818	0.217
		Non-n-mt	20544	0.827	0.206

Supplemental Table 3. Measures of mitonuclear association for incompatible n-mt genes previously identified by Robles et al. [44].

		CALL[c]		CHAF[c]		HUEX-STAC[c]
Gene[a]	**Incompatible mtDNA[b]**	***D′***	***r²***	***D′***	***r²***	**Allele frequency**
NDUFA13 (g3279.t1)	*X. malinche*	0.57 (2.08)	0.18 (1.65)	0.54 (0.83)	0.15 (5.37)	0.99 (0.77)
MTERF4 (g3322.t1)	*X. malinche*	0.58 (2.16)	0.14 (0.82)	0.59 (1.12)	0.19 (6.86)	0.97 (0.69)
ATP5MG (g6088.t1)	*X. malinche*	ND	ND	ND	ND	0.96 (0.67)
NDUFS5 (g8054.t1)	bidirectional	0.65 (2.91)	0.40 (5.34)	0.73 (1.91)	0.16 (5.58)	0.97 (0.70)
MMUT (g15308.t1)	bidirectional	0.42 (0.57)	0.04 (-0.87)	0.00 (-2.21)	0.00 (-1.22)	0.54 (-1.40)
SMIM8 (g15060.t1)	*X. birchmanni*	0.59 (2.26)	0.27 (3.10)	0.67 (1.53)	0.05 (0.89)	ND
LYRM2 (g15051.t1)	*X. birchmanni*	0.49 (1.29)	0.22 (2.30)	0.59 (1.07)	0.04 (0.59)	0.92 (0.45)
RMDN3 (g15056.t1)	*X. birchmanni*	0.58 (2.22)	0.27 (3.15)	0.58 (1.04)	0.04 (0.44)	0.92 (0.44)
UQCRC2 (g11575.t1)	*X. birchmanni*	0.33 (-0.40)	0.04 (-0.76)	0.11 (-1.60)	0.00 (-1.14)	0.90 (0.36)

[a]Human gene names are given with the X. birchmanni annotation ID indicated in parentheses. UQCR11 has also been identified as a gene involved in mitonuclear incompatibilities [41], but it was not included in the annotation used in this analysis.
[b]Species listed indicate the mtDNA that is incompatible with the other species' n-mt allele. For example, X. malinche means the X. malinche mtDNA is incompatible with X. birchmanni ancestry at NDUFA13. Bidirectional means the incompatibility occurs for both mtDNAs..
[c]Values are reported for each metric of mitonuclear association (D', r^2, and allele frequency), and the corresponding Z-score for each value is indicated in parentheses, which was calculated by subtracting the mean value and dividing by the standard deviation.

Supplemental Table 4. Results of Fisher's exact test for enrichment of interacting n-mt genes in top 1% of genome-wide association statistics.

Population	Test	Result
CALL	One-sided Fisher's Exact Test for Count Data on D' values	$p = 0.42$
	One-sided Fisher's Exact Test for Count Data r^2 values	$p = 0.22$
CHAF	One-sided Fisher's Exact Test for Count Data on D' values	$p = 0.66$

Supplemental Table 5. Studies investigating differential effects of n-mt and non-n-mt gene groups or genomic windows.

Citation	**Admixture category**[a]	**Hybridizing groups**	**Segregating mtDNA?**[b]	**Disproportionate n-mt gene effect?**[c]
Mollusks				
Fields et al. 2024 [59]	Historical introgression	*Potamopyrgus estuarinus* and *P. kaitunuparaoa*	No	Yes
Fish				
Kato et al. 2023 [60]	Historical introgression	*Chaenogobius annularis* and a ghost lineage	No	No
Mammals				
Kwon et al. 2022 [61]	Interspecific hybridization	*Bos taurus* and *B. indicus*	No	Yes
Shi et al. 2025 [62]	Interspecific hybridization	*B. taurus* and *B. indicus*	No	No[d]
Jensen et al. 2023 [63]	Historical introgression	*Allochrocebus* and *Chlorocebus* lineages	No	Yes
Zhu & Evans 2023 [64]	Historical introgression	*Macaca arctoides* and *M. fascicularis aurea*	No	No
Evans et al. 2021 [65]	Interspecies individual hybrid	*M. tonkeana* and *M. maura*	This individual has *M. tonkeana* mtDNA	N/A[f]
	Interspecies individual hybrid	*M. tonkeana* and *M. hecki*	This individual has *M. hecki* mtDNA	N/A[f]
Bailey & Stevison 2021 [66]	Interspecific hybridization	*M. arctoides* and *M. sinica* or *M. fascicularis*	No	No
Sloan et al. 2015 [49]	Intraspecific admixture between populations	*Homo sapiens*	Yes	No
Sharbrough et al. 2017 [67]	Historical introgression	*H. sapiens* and *H. neanderthalensis*	No	Yes
	Historical introgression	*H. sapiens* and Denisovans	No	No

Zaidi & Makova 2019 [68]	Intraspecific admixture between populations	*H. sapiens* African American and European	Yes	Yes
	Intraspecific admixture between populations	*H. sapiens* Puerto Rican and European	Yes	Yes
	Intraspecific admixture between populations	*H. sapiens* Caribbean and European	Yes	No
	Intraspecific admixture between populations	*H. sapiens* Columbian and European	Yes	No
	Intraspecific admixture between populations	*H. sapiens* Mexican and European	Yes	No
	Intraspecific admixture between populations	*H. sapiens* Peruvian and European	Yes	No
Birds				
Lopez et al. 2021 [55]	Intraspecific admixture between populations	*Poephila acuticauda acuticauda* and *P. a. hecki*	Yes	Yes[e]
Morales et al. 2018 [69]	Intraspecific admixture between populations	northern and southern populations of *Eopsaltria australis*	Yes	Yes, but see [56]
Low et al. 2024 [56]	Intraspecific admixture between populations	inland and coastal mitochondrial lineages of *E. australis*	Yes	Yes[e]
Rancilhac et al. 2024 [70]	Interspecific hybridization	*Pogoniulus pusillus* and *P. chrysoconus*	Yes	No
Musher et al. 2024 [71]	Intraspecific admixture	*Thamnophilus aethiops*	Yes	No

These papers were collected, using two groups of search terms to identify published articles that included similar comparisons between n-mt and non-n-mt genes in the context of hybridization. The first search terms were "n-mt" AND ("reproductive isolation" OR "speciation") AND “hybrid” on Google Scholar in October 2024. Inclusion criteria for the resulting hits were that the work (1) categorized genes into n-mt and non-n-mt classes, (2) measured association between n-mt genetic variants and mitochondrial haplotypes, and (3) was focused on current or ancient admixture of genomes (via hybridization or introgression). Papers were excluded if the only aim was to use phylogenetic analyses to study mutation rates or phylogenetic discordance between mtDNA and non-n-mt loci. This search (conducted on 18 October 2024) identified 193 possible articles, but only 10 met our inclusion criteria. We then performed a separate search using the terms “mito*nuclear” AND (“reproductive isolation” OR “speciation.”), returning 57 articles, but none that met our inclusion criteria. We supplemented this list with six additional papers that were not returned as hits in the above searches but that we encountered as citations in other studies or during unrelated literature searches. The combination of search terms and narrow inclusion criteria likely excluded some relevant studies from this table, but the goal was not to capture all studies pertinent to genomic analysis of mitonuclear interactions.

[a]Hybridization systems were categorized as being either intraspecific admixture between populations, interspecific hybridization, or historical introgression to indicate timing of hybridization and how removed from initial admixture a population may be.

[b]Indication of whether the mtDNA from the hybrid source populations or species are still segregating in the hybrid system investigated.
[c]Investigations that found n-mt gene groups or genomic windows to share ancestry more than non-n-mt groups or windows were given a 'yes' to indicate whole-genome based methods successfully identified effects of n-mt selection.
[d]This study instead found evidence of enrichment for mismatch of n-mt and mtDNA ancestry.
[e]These studies find positive relationships between n-mt and mtDNA ancestry, but these relationships also include matching of sex chromosome ancestry or that n-mt genes are located in regions of a chromosome associated with sex determination.
[f]Because these studies only determined n-mt disproportionality with a single individual, they are omitted from our investigation into whether a fixing a mitochondrial haplotype affects the sensitivity of genome-wide tests of shared n-mt and mtDNA ancestry.

389 **REFERENCES**
390

391 1. Gray MW, Burger G, Lang BF. 1999 Mitochondrial Evolution. *Science (1979)* **283**, 1476–
392 1481.
393 2. Karlberg O, Canback B, Kurland CG, Andersson SGE. 2000 The dual origin of the yeast
394 mitochondrial proteome. *Yeast* **17**, 170–187.
395 3. Gray MW. 2015 Mosaic nature of the mitochondrial proteome: Implications for the origin
396 and evolution of mitochondria. *Proc Natl Acad Sci U S A* **112**, 10133–10138.
397 (doi:10.1073/pnas.1421379112)
398 4. Rath S *et al.* 2021 MitoCarta3.0: An updated mitochondrial proteome now with sub-
399 organelle localization and pathway annotations. *Nucleic Acids Res* **49**, D1541–D1547.
400 (doi:10.1093/nar/gkaa1011)
401 5. Burton RS. 2022 The role of mitonuclear incompatibilities in allopatric speciation.
402 *Cellular and Molecular Life Sciences* **79**. (doi:10.1007/s00018-021-04059-3)
403 6. Lynch M, Blanchard JL. 1998 Deleterious mutation accumulation in organelle genomes.
404 *Felsenstein* **102**, 29–39.
405 7. Nabholz B, Glémin S, Galtier N. 2008 Strong variations of mitochondrial mutation rate
406 across mammals - The longevity hypothesis. *Mol Biol Evol* **25**, 120–130.
407 (doi:10.1093/molbev/msm248)
408 8. Allio R, Donega S, Galtier N, Nabholz B. 2017 Large variation in the ratio of
409 mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity
410 and the use of mitochondrial DNA as a molecular marker. *Mol Biol Evol* **34**, 2762–2772.
411 (doi:10.1093/molbev/msx197)
412 9. Lynch M. 1996 Mutation Accumulation in Transfer RNAs: Molecular Evidence for
413 Muller's Ratchet in Mitochondrial Genomes. *Mol. Biol. Evol.* **13**, 209–220.
414 10. Melvin RG, Ballard JWO. 2017 Cellular and population level processes influence the rate,
415 accumulation and observed frequency of inherited and somatic mtDNA mutations.
416 *Mutagenesis* **32**, 323–334. (doi:10.1093/mutage/gex004)
417 11. Neiman M, Taylor DR. 2009 The causes of mutation accumulation in mitochondrial
418 genomes. *Proceedings of the Royal Society B: Biological Sciences* **276**, 1201–1209.
419 (doi:10.1098/rspb.2008.1758)
420 12. Sanchez-Contreras M *et al.* 2021 A replication-linked mutational gradient drives somatic
421 mutation accumulation and influences germline polymorphisms and genome composition
422 in mitochondrial DNA. *Nucleic Acids Res* **49**, 11103–11118. (doi:10.1093/nar/gkab901)
423 13. Radzvilavicius AL, Kokko H, Christie JR. 2017 Mitigating mitochondrial genome erosion
424 without recombination. *Genetics* **207**, 1079–1088. (doi:10.1534/genetics.117.300273)
425 14. Cooper BS, Burrus CR, Ji C, Hahn MW, Montooth KL. 2015 Similar efficacies of
426 selection shape mitochondrial and nuclear genes in both Drosophila melanogaster and
427 Homo sapiens. *G3: Genes, Genomes, Genetics* **5**, 2165–2176.
428 (doi:10.1534/g3.114.016493)
429 15. Edwards DM, Røyrvik EC, Chustecki JM, Giannakis K, Glastad RC, Radzvilavicius AL,
430 Johnston IG. 2021 Avoiding organelle mutational meltdown across eukaryotes with or
431 without a germline bottleneck. *PLoS Biol* **19**. (doi:10.1371/journal.pbio.3001153)
432 16. Gershoni M, Templeton AR, Mishmar D. 2009 Mitochondrial bioenergetics as a major
433 motive force of speciation. *BioEssays* **31**, 642–650. (doi:10.1002/bies.200800139)

17. Rand DM, Haney RA, Fry AJ. 2004 Cytonuclear coevolution: The genomics of cooperation. *Trends Ecol Evol* **19**, 645–653. (doi:10.1016/j.tree.2004.10.003)
18. Schmidt TR, Wu W, Goodman M, Grossman LI. 2001 Evolution of Nuclear-and Mitochondrial-Encoded Subunit Interaction in Cytochrome c Oxidase. *Mol. Biol. Evol* **18**, 563–569.
19. Hill GE. 2020 Mitonuclear Compensatory Coevolution. *Trends in Genetics* **36**, 403–414. (doi:10.1016/j.tig.2020.03.002)
20. Weaver RJ, Rabinowitz S, Thueson K, Havird JC. 2022 Genomic Signatures of Mitonuclear Coevolution in Mammals. *Mol Biol Evol* **39**. (doi:10.1093/molbev/msac233)
21. Barreto FS, Watson ET, Lima TG, Willett CS, Edmands S, Li W, Burton RS. 2018 Genomic signatures of mitonuclear coevolution across populations of Tigriopus californicus. *Nat Ecol Evol* **2**, 1250–1257. (doi:10.1038/s41559-018-0588-1)
22. Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015 Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution (N Y)* **69**, 3069–3081. (doi:10.1111/evo.12808)
23. Osada N, Akashi H. 2012 Mitochondrial-nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome c oxidase complex. *Mol Biol Evol* **29**, 337–346. (doi:10.1093/molbev/msr211)
24. Sloan DB, Havird JC, Sharbrough J. 2017 The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol Ecol* **26**, 2212–2236. (doi:10.1111/mec.13959)
25. Piccinini G, Iannello M, Puccio G, Plazzi F, Havird JC, Ghiselli F. 2021 Mitonuclear coevolution, but not nuclear compensation, drives evolution of OXPHOS complexes in bivalves. *Mol Biol Evol* **38**, 2597–2614. (doi:10.1093/molbev/msab054)
26. Zavala-Paez M, Sutara B, Keller S, Holliday J, Fitzpatrick MC, Hamilton J. 2025 The role of cytonuclear interactions to plant adaptation across a *Populus* hybrid zone. *bioRxiv* (doi:10.1101/2025.05.13.653687)
27. Ellison CK, Niehuis O, Gadau J. 2008 Hybrid breakdown and mitochondrial dysfunction in hybrids of Nasonia parasitoid wasps. *J Evol Biol* **21**, 1844–1851. (doi:10.1111/j.1420-9101.2008.01608.x)
28. Coyne JA. 1992 Genetics and speciation. *Nature* **355**.
29. Hill GE. 2016 Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol Evol* **6**, 5831–5842. (doi:10.1002/ece3.2338)
30. Burton RS, Barreto FS. 2012 A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol* **21**, 4942–4957. (doi:10.1111/mec.12006)
31. Lima TG, Burton RS, Willett CS. 2019 Genomic scans reveal multiple mito-nuclear incompatibilities in population crosses of the copepod Tigriopus californicus. *Evolution (N Y)* **73**, 609–620. (doi:10.1111/evo.13690)
32. Forsythe ES, Nelson ADL, Beilstein MA. 2020 Biased gene retention in the face of introgression obscures species relationships. *Genome Biol Evol* **12**, 1646–1663. (doi:10.1093/GBE/EVAA149)
33. Langdon QK *et al.* 2024 Swordtail fish hybrids reveal that genome evolution is surprisingly predictable after initial hybridization. *PLoS Biol* (doi:10.1101/2023.12.21.572897)

34. Schumer M, Cui R, Powell DL, Dresner R, Rosenthal GG, Andolfatto P. 2014 High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *Elife* **2014**. (doi:10.7554/eLife.02535)
35. Powell DL, Rose AD, Rosenthal GG. 2022 A widely-used pollutant causes reversal of conspecific mate preference in a freshwater fish. *bioRxiv* , 2022.09.07.507014. (doi:10.1101/2022.09.07.507014)
36. Powell DL, Payne C, Banerjee SM, Keegan M, Bashkirova E, Cui R, Andolfatto P, Rosenthal GG, Schumer M. 2021 The Genetic Architecture of Variation in the Sexually Selected Sword Ornament and Its Evolution in Hybrid Populations. *Current Biology* **31**, 923-935.e11. (doi:https://doi.org/10.1016/j.cub.2020.12.049)
37. Langdon QK, Powell DL, Kim B, Banerjee SM, Payne C, Dodge TO, Moran B, Fascinetto-Zago P, Schumer M. 2022 Predictability and parallelism in the contemporary evolution of hybrid genomes. *PLoS Genet* **18**. (doi:10.1371/journal.pgen.1009914)
38. Powell DL, Moran BM, Kim BY, Banerjee SM, Aguillon SM, Fascinetto-Zago P, Langdon QK, Schumer M. 2021 Two new hybrid populations expand the swordtail hybridization model system. *Evolution (N Y)* **75**, 2524–2539. (doi:10.1111/evo.14337)
39. Banerjee SM, Powell DL, Moran BM, Ramírez-Duarte WF, Langdon QK, Gunn TR, Vazquez G, Rochman C, Schumer M. 2023 Complex hybridization between deeply diverged fish species in a disturbed ecosystem. *Evolution (N Y)* **77**, 995–1005. (doi:10.1093/evolut/qpad019)
40. Powell DL *et al.* 2020 Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science (1979)* **368**, 731–736.
41. Schumer M *et al.* 2018 Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science (1979)* , 656–660.
42. Aguillon SM *et al.* 2025 Pervasive gene flow despite strong and varied reproductive barriers in swordtails. *Nat Ecol Evol* **9**, 867–878. (doi:https://doi.org/10.1038/s41559-025-02669-9)
43. Moran BM *et al.* 2024 A Lethal Genetic Incompatibility between Naturally Hybridizing Species in Mitochondrial Complex I. *Nature* (doi:https://doi.org/10.1038/s41586-023-06895-8)
44. Robles N V. *et al.* 2025 Admixture mapping reveals evidence for multiple mitonuclear incompatibilities in swordtail fish hybrids. *bioRxiv* (doi:10.1101/2025.01.30.635158)
45. Schumer M, Powell DL, Corbett-Detig R. 2020 Versatile simulations of admixture and accurate local ancestry inference with mixnmatch and ancestryinfer. *Mol Ecol Resour* **20**, 1141–1151. (doi:https://doi.org/10.1111/1755-0998.13175)
46. Dodge TO *et al.* 2024 Structural genomic variation and behavioral interactions underpin a balanced sexual mimicry polymorphism. *Current Biology* **34**, 4662-4676.e9. (doi:10.1016/j.cub.2024.08.053)
47. Pertea G, Pertea M. 2020 GFF Utilities: GffRead and GffCompare. *F1000Res* **9**, 304. (doi:10.12688/f1000research.23297.2)
48. Emms DM, Kelly S. 2019 OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**. (doi:10.1186/s13059-019-1832-y)
49. Sloan DB, Fields PD, Havird JC. 2015 Mitonuclear linkage disequilibrium in human populations. *Proceedings of the Royal Society B: Biological Sciences* **282**. (doi:10.1098/rspb.2015.1704)

50. Zhao H, Nettleton D, Soller M, Dekkers JCM. 2005 Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet Res* **86**, 77–87. (doi:10.1017/S001667230500769X)
51. Forsythe ES, Williams AM, Sloan DB. 2021 Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell* **33**, 980–997. (doi:10.1093/plcell/koab021)
52. Schumer M, Brandvain Y. 2016 Determining epistatic selection in admixed populations. *Mol Ecol* **25**, 2577–2591. (doi:10.1111/mec.13641)
53. Barton NH. 2000 Genetic hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355**, 1553–1562. (doi:10.1098/rstb.2000.0716)
54. Veller C, Edelman NB, Muralidhar P, Nowak MA. 2023 Recombination and selection against introgressed DNA. *Evolution (N Y)* **77**, 1131–1144. (doi:10.1093/evolut/qpad021)
55. Lopez KA, McDiarmid CS, Griffith SC, Lovette IJ, Hooper DM. 2021 Evaluating evidence of mitonuclear incompatibilities with the sex chromosomes in an avian hybrid zone. *Evolution (N Y)* **75**, 1395–1414. (doi:10.1111/evo.14243)
56. Low GW *et al.* 2024 Accelerated differentiation of neo-W nuclear-encoded mitochondrial genes between two climate-associated bird lineages signals potential co-evolution with mitogenomes. *Heredity (Edinb)* (doi:10.1038/s41437-024-00718-w)
57. Eyre-Walker A. 2017 Mitochondrial replacement therapy: Are mito-nuclear interactions likely to be a problem? *Genetics* **205**, 1365–1372. (doi:10.1534/genetics.116.196436)
58. Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal GG. 2013 Phylogenomics Reveals Extensive Reticulate Evolution In Xiphophorus Fishes. *Evolution (N Y)* **67**, 2166–2179. (doi:10.1111/evo.12099)
59. Fields PD *et al.* 2024 Genome Evolution and Introgression in the New Zealand mud Snails Potamopyrgus estuarinus and Potamopyrgus kaitunuparaoa. *Genome Biol Evol* **16**. (doi:10.1093/gbe/evae091)
60. Kato S, Arakaki S, Nagano AJ, Kikuchi K, Hirase S. 2023 Genomic landscape of introgression from the ghost lineage in a gobiid fish uncovers the generality of forces shaping hybrid genomes. *Mol Ecol* (doi:10.1111/mec.17216)
61. Kwon T, Kim K, Caetano-Anolles K, Sung S, Cho S, Jeong C, Hanotte O, Kim H. 2022 Mitonuclear incompatibility as a hidden driver behind the genome ancestry of African admixed cattle. *BMC Biol* **20**. (doi:10.1186/s12915-021-01206-x)
62. Shi X *et al.* 2025 Selection increases mitonuclear DNA discordance but reconciles incompatibility in African cattle. *Mol Biol Evol* (doi:10.1093/molbev/msaf039/8005707)
63. Jensen A *et al.* 2023 Complex Evolutionary History With Extensive Ancestral Gene Flow in an African Primate Radiation. *Mol Biol Evol* **40**. (doi:10.1093/molbev/msad247)
64. Zhu J, Evans BJ. 2023 Mitonuclear Interactions and the Origin of Macaque Societies. *Genome Biol Evol* **15**. (doi:10.1093/gbe/evad010)
65. Evans BJ, Peter BM, Melnick DJ, Andayani N, Supriatna J, Zhu J, Tosi AJ. 2021 Mitonuclear interactions and introgression genomics of macaque monkeys (Macaca) highlight the influence of behaviour on genome evolution. *Proceedings of the Royal Society B: Biological Sciences* **288**. (doi:10.1098/rspb.2021.1756)
66. Bailey NP, Stevison LS. 2021 Mitonuclear conflict in a macaque species exhibiting phylogenomic discordance. *J Evol Biol* **34**, 1568–1579. (doi:10.1111/jeb.13914)

67. Sharbrough J, Havird JC, Noe GR, Warren JM, Sloan DB. 2017 The mitonuclear dimension of Neanderthal and Denisovan ancestry in modern human genomes. *Genome Biol Evol* **9**, 1567–1581. (doi:10.1093/gbe/evx114)
68. Zaidi AA, Makova KD. 2019 Investigating mitonuclear interactions in human admixed populations. *Nat Ecol Evol* **3**, 213–222. (doi:10.1038/s41559-018-0766-1)
69. Morales HE, Pavlova A, Amos N, Major R, Kilian A, Greening C, Sunnucks P. 2018 Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates. *Nat Ecol Evol* **2**, 1258–1267. (doi:10.1038/s41559-018-0606-3)
70. Rancilhac L *et al.* 2024 Introgression across narrow contact zones shapes the genomic landscape of phylogenetic variation in an African bird clade. (doi:10.1101/2024.08.13.607717)
71. Musher LJ, Del-Rio G, Marcondes RS, Brumfield RT, Bravo GA, Thom G. 2024 Geogenomic Predictors of Genetree Heterogeneity Explain Phylogeographic and Introgression History: A Case Study in an Amazonian Bird (Thamnophilus aethiops). *Syst Biol* **73**, 36–52. (doi:10.1093/sysbio/syad061)