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 early species boundaries. Under this model, genetic incompatibilities emerge when mitochondrial haplotypes are in a cellular context without their coevolved nuclear-encoded mitochondrial (n-mt) proteins. Some case studies have shown that such disruptions in mitonuclear coevolution can contribute to reproductive isolation, but whether mitonuclear incompatibilities generate selection that impacts multiple n-mt loci and/or causes broad, genome-wide contributions to speciation is unclear. Here, we leverage a system with several hybridizing species pairs (*Xiphophorus* fishes) that have known mitonuclear incompatibilities of large effect. We divided nuclear-encoded genes into three classes based on level of interaction with mitochondrial gene products. We found only inconsistent statistical support for a difference between these classes in the degree of positive covariation in mitonuclear ancestry. We discuss evidence that these analyses are sensitive to the amount of non-synonymous divergence between parent species in interacting n-mt genes or the age of the hybridization event. Overall, our results imply that genome-wide scans focused on enrichment of broad functional gene classes may often 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
34 component of the eukaryotic cell, functioning in such processes as oxidative phosphorylation
35 (OXPHOS), Ca^{2+} acquisition and signaling, and biosynthesis of heme and iron-sulfur clusters.
36 Bilaterian metazoan mitochondrial genomes (mitogenomes) have been reduced to encode only
37 13 core protein subunits of the electron transport chain, 2 rRNAs, and 22 tRNAs, making
38 mitochondria dependent on the import of nuclear-encoded, but mitochondrially targeted, (n-mt)
39 proteins for all functions [1–4]. Due to the importance of mitochondrial function to fitness,
40 maintaining these intimate interactions among coevolved partners may be essential.
41 Mitogenomes evolve at a faster rate than nuclear genomes in most bilaterian metazoan
42 systems, with an average rate of 19× faster in vertebrates [5–8]. These elevated substitution rates
43 may result from a higher underlying mutation rate in combination with inefficient purging of
44 deleterious mutations due to smaller effective population size, lack of recombination, and
45 uniparental inheritance [9–12], although explanations based on reduced efficacy of selection in
46 mitogenomes have been recently challenged [13–15]. Regardless of the cause, the consequences
47 of high mitochondrial DNA (mtDNA) substitution rates are often deleterious [6]. As such,
48 substitutions in mitochondrially encoded proteins may create selection for compensatory changes
49 in interacting n-mt proteins [16–18], and signatures of coevolution have indeed been found
50 between the mitochondrial and nuclear genomes [19–22]. Other forms of mitonuclear
51 coevolution (i.e., reciprocal selection for compatible genotypes) are also possible [19,23–25].
52 Regardless of the exact process, the outcome is a set of interacting proteins that are functionally
53 dependent on sufficient "matching" of lineage-specific substitutions in mitochondrial and n-mt
54 proteins [23].
55 As a result, disrupting the association between a mitochondrial haplotype and its
56 coevolved nuclear counterparts (e.g., via secondary contact following speciation) is predicted to
57 uncover genetic incompatibilities that made early contributions to reproductive isolation [5,26].
58 Given the rate at which mitonuclear genes coevolve and the importance of mitochondrial
59 function to cellular metabolism, mitonuclear incompatibilities are proposed to have a
60 disproportionate effect on the accumulation of reproductive barriers during lineage divergence
61 [23,27–30]. These incompatibilities could potentially be driven by only a few genes with major

62 effects on fitness or alternatively by many genes of small effect. Although studies have found
63 evidence of mitonuclear incompatibilities (e.g., [31,32]) in some systems, it is still unclear how
64 far-reaching these effects are and how many genes are involved. For example, the immediate
65 effects of selection on mitonuclear incompatibilities could be restricted to a subset of n-mt
66 proteins involved in direct physical interactions with mitochondrial gene products. On the other
67 hand, because mitochondria are hubs of various biological processes, selection may be more
68 widespread and affect nuclear genes that have no direct connection to mitochondrial function.
69 We hypothesize that, relative to other nuclear genes, n-mt genes are more strongly affected by
70 selection acting to maintain associations between matched lineage-specific mitonuclear
71 genotypes. Therefore, examining signatures of mtDNA "match" in n-mt genes vs. nuclear-
72 encoded genes that lack mitochondrial interactions should reveal a history of mitonuclear
73 coevolution across the genome at species boundaries. While partitioning genomic data based on
74 n-mt gene function has been used to investigate mitonuclear incompatibilities in many systems
75 (Table S1), the effectiveness of this approach has not been tested in a system with previously
76 known mitonuclear incompatibilities.
77 Species of the swordtail fish genus *Xiphophorus* have recently begun hybridizing in
78 multiple river systems in Mexico, providing a model of repeated hybridization and speciation
79 dynamics [33–42]. Recent work on hybrid zones between sister species *X. birchmanni* and *X.*
80 *malinche* has identified severe mitonuclear incompatibilities involving the *ndufs5* and *ndufa13* n-
81 mt genes [43]. Mitonuclear ancestry mismatches at these loci have severe fitness effects,
82 including death during embryonic development. Subsequent investigations have identified that at
83 least nine nuclear genomic regions (which contain a total of 10 known n-mt genes across seven
84 loci) are involved in mitonuclear incompatibilities in these hybrids [44]. Some of these
85 incompatibilities, including the *ndufs5* and *ndufa13* n-mt genes, were also found in *X.*
86 *birchmanni* and *X. cortezi* hybrids [42]. We therefore leveraged this system to investigate
87 evidence for mitonuclear matching during hybridization on a genome-wide scale and the ability
88 to detect these interactions with commonly used approaches that partition nuclear genes into
89 classes based on their role in mitochondrial function. We categorized nuclear genes into three
90 groups based on whether the encoded protein (1) interacts with mitogenomic gene products
91 (interacting n-mt), (2) functions in mitochondria but does not physically interact with
92 mitogenomic gene products (non-interacting n-mt), or (3) does not function in mitochondria

93 (non-n-mt). This *a priori* partitioning allows for an unbiased test of whether selection for
94 matched mitonuclear genotypes leaves genome-wide signatures. Overall, we found inconsistent
95 statistical evidence for differences among these categories based on analyses of mitonuclear
96 interactions in *Xiphophorus* hybrids. However, we found that interacting n-mt genes with larger
97 amounts of non-synonymous substitutions may have more severe consequences for mitonuclear
98 incompatibilities. Therefore, the role of n-mt genes may not be generally detectable on a
99 genome-wide scale, and the sensitivity of these tests should be taken into consideration when
100 interpreting results.
101
102
103 **METHODS**
104
105 **Genomic data from *Xiphophorus* hybrid populations**
106 Three datasets resulting from previous local ancestry inference in hybrid *Xiphophorus* fishes
107 were acquired as described previously [37,40,43]. Each dataset includes samples from a
108 population of hybrid *Xiphophorus* individuals inhabiting sections of rivers in the states of
109 Hidalgo or San Luis Potosí, Mexico. The first (Calnali Low, or CALL) is on the Calnali River
110 and consists of *X. birchmanni* and *X. malinche* hybrids. The second is at Chahuaco Falls (CHAF)
111 on the Calnali River. It is separated from the CALL population by a waterfall (~50m tall) [43]
112 and consists of hybrids of the same pair of species as CALL. The third represents two
113 populations (Huextetitla and Santa Cruz, or HUEX-STAC) downstream of the same fork on the
114 Santa Cruz River and derived from a shared hybridization event between *X. birchmanni* and *X.*
115 *cortezi* followed by independent evolution [37,38].
116 These datasets were originally analyzed using the *ancestryinfer* pipeline [45] and resulted
117 in posterior probabilities of species ancestry at between 628,881 and 689,982 ancestry
118 informative markers (AIMs) across the nuclear and mitochondrial genomes (Table 1), as
119 described previously [36,39,45]. Informative sites were defined as those with at least a 98%
120 frequency difference between the parental species, as previously described [33,44]. We used a
121 published annotation of the *X. birchmanni* reference genome [46] to identify genes containing
122 each AIM. We then transformed AIM probabilities into categorical genotype calls based on a
123 posterior probability threshold. If a probability for a parent species was greater than or equal to

124 0.9 and the other parent species' probability of ancestry was less than or equal to 0.1, ancestry
125 was assigned to the major probability. If instead the sum of both parents' ancestries was less than
126 or equal to 0.2, the AIM was assigned as heterozygous. AIMs that did not meet these criteria
127 were filtered out of the dataset. Allele counts were then averaged across all AIMs that passed the
128 prior filtering to calculate an average major parent ancestry proportion per individual. These
129 values were then averaged for nuclear and mitochondrial AIMs to obtain population-level
130 averages (Table 1).
131
132 **Identification of *Xiphophorus* n-mt genes**
133 The gffread [47] program was used to obtain predicted protein sequences, using the -V option to
134 remove any coding sequence with an in-frame stop codon. To annotate n-mt proteins, we
135 compared the *Xiphophorus* translated proteome to UniProt human (UP000005640,
136 GCA_000001405.27) and mouse (UP000000589, GCA_000001635.8) protein sequences. N-mt
137 proteins have been empirically identified in humans and mice as part of the MitoCarta 3.0 project
138 [4]. Because the UniProt human and mouse datasets differed in nomenclature from the MitoCarta
139 dataset, we assigned protein names by searching for MitoCarta sequences in the UniProt human
140 sequences with BLASTP v2.15.0+. A custom Perl script was then used to parse the BLASTP
141 output, and identified proteins were manually confirmed to be the correct protein if percent
142 coverage and sequence identity were not 100%. OrthoFinder v2.5.4 [48] was then run with the
143 primary translated transcript for each *Xiphophorus* sequence and the full sets of human and
144 mouse protein sequences.
145 MitoCarta protein classifications were used to differentiate non-n-mt from n-mt
146 orthogroups. Interacting and non-interacting orthogroups were distinguished based on a
147 previously published [49] list, updated here, of interacting n-mt proteins that includes OXPHOS
148 subunits, ribosomal subunits, aminoacyl-tRNA synthetases, and DNA/RNA polymerases (Table
149 S2).
150
151 **Linkage disequilibrium calculation**
152 We used mitonuclear linkage disequilibrium (LD) as a metric for the association between
153 mitochondrial and nuclear alleles in the CALL and CHAF hybrid populations (see below for
154 analysis methods for the HUEX-STAC population, which differed because their mitogenome is

155 fixed for the haplotype from the *X. cortezi* parental species [38]). Calculations were performed
156 using the D' and r^2 statistics [50]. In the equations that follow, A and B represent nuclear and
157 mitochondrial loci, respectively, and pA and pB represent the corresponding allele frequencies.
158 pAB is the frequency of the haplotype AB across the two sites. The A and B alleles were assigned
159 to be from the same parental population so that positive and negative values of LD indicate a
160 higher frequency of parental or recombinant allele combinations (i.e., matched and mismatched),
161 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

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

162 Because all mitochondrial genes are linked on the same non-recombining mitogenome, a
163 single AIM was used (CALL and CHAF's AIM was in *COX2* and HUEX-STAC's AIM was in
164 *ATP6*, positions 45 and 846 of GenBank accession CM071424.1, respectively) to identify the
165 mitochondrial haplotype for each fish and to represent the mitogenome in all LD analyses. This
166 approach was appropriate for our data, as all mtDNA AIMs within an individual had
167 approximately the same posterior probability for ancestry assignment. Posterior probabilities
168 were summed across each nuclear AIM as a proxy for frequency of the two alleles. The two LD
169 statistics were calculated for each gene class separately in a custom R script. LD values were
170 visualized using ggplot2 v3.4.0 and ggridges v0.5.4 in RStudio v2023.03.1+446.
171 LD values were then averaged across AIMs within each nuclear gene to limit pseudo-
172 replication. A one-way ANOVA from the R stats v4.2.2 package (commands *lm* and *anova*) was
173 used to determine whether gene class had a statistically significant impact on LD values.

174 Summary statistics were collected using commands *n*, *mean*, and *sd* from the dplyr v1.1.0, base
175 v4.2.2, and stats v4.2.2 packages, respectively.
176 In the case of selection on a small number of n-mt genes, shifts in genome-wide averages
177 may not be detectable but could still result in individual loci with the strongest association with
178 the mitogenome being enriched for interacting n-mt genes. We therefore also identified the AIMs
179 and genes with the top 1% of LD values with a custom R script.
180
181 **Partial correlation calculation**
182 High LD values can result from population structure (i.e., proportion of parent species ancestry
183 in an individual genome), and the populations used here vary in the proportion of background
184 ancestry (Table 1). We therefore performed a partial correlation (pcor) analysis, which calculated
185 the correlation between mitochondrial and nuclear ancestry while accounting for the correlation
186 expected given the variation in genome-wide ancestry among individuals. AIMs were filtered as
187 described above, and individual ancestry genotype calls for mitochondrial and nuclear AIMs
188 were used to estimate a correlation coefficient with the *pcor.test* function from the R ppcor v1.1
189 package. This estimate was then averaged across all AIMs in a gene, and a one-way ANOVA
190 was run to determine the effect of gene class on the correlation estimate.
191
192 **HUEX-STAC allele frequency calculation**
193 The HUEX-STAC populations are both fixed for the *X. cortezi* mitochondrial haplotype, so we
194 expected *X. cortezi* ancestry to be disproportionately represented at n-mt loci. However, due to
195 the lack of mtDNA variation, the average allele frequency for each nuclear gene was calculated
196 and visualized (with a custom R script) instead of mitonuclear LD to infer how mtDNA
197 background affects selection on n-mt genes in this system. A one-way ANOVA and summary
198 statistics were computed as described above.
199 A permutation test was then performed to determine whether significant results were
200 robust to violations of normality assumptions in ANOVA. Allele frequency values were
201 permutated by randomly reassigning them across genes 9999 times using base R v4.2.2 functions
202 in a custom R script. Permuted datasets were then used to calculate null expectations for
203 differences in allele frequencies between interacting n-mt and non-n-mt gene classes. The

204 observed difference was compared to the permutations to determine the probability of the pattern
205 arising by chance.
206
207 **Non-synonymous substitution analysis**
208 We further filtered the three datasets to include only genes that had substitutions in the coding
209 sequence. We quantified the number of non-synonymous substitutions normalized to the length
210 of the coding sequence. This filtered dataset was analyzed using a linear model with mitonuclear
211 LD as the response variable and an interaction between gene class and our normalized count of
212 non-synonymous substitutions as the explanatory variables. A type III ANOVA was then run
213 (via *Anova* from the R car v3.1-1 package) on the model to determine statistical significance of
214 our predictor variables followed by use of the *summary* function (from base R) with non-n-mts
215 as the reference group. An influential point analysis was performed on the interacting n-mt gene
216 class using DFFIT with the plotting command *ols_plot_dffits* from the R olsrr v0.6.1 package.
217 Identified points were then individually removed from the model with all gene classes to
218 determine their effects on the relationship between mitonuclear LD and non-synonymous
219 substitutions.
220
221 **Data and code availability**
222 All code has been deposited in GitHub at https://github.com/sakuster/sensitivity-of-genome-
223 wide-mtnuc. Data can be accessed on figshare (https://doi.org/10.6084/m9.figshare.c.7902011).
224
225
226 **RESULTS**
227
228 **Identification of *Xiphophorus* n-mt sequences**
229 Of the 24,269 *Xiphophorus* annotated genes, 86.5% (20,988) were assigned to an orthogroup by
230 OrthoFinder. We found that 75.5% (13,264) of the 17,566 identified orthogroups contained at
231 least one *Xiphophorus* gene, and 69.4% (12,198) included at least one gene from all three species
232 (*X. birchmanni*, mouse, and human). Gene and/or genome duplications during the evolutionary
233 history of these three species resulted in some orthogroups with multiple *X. birchmanni* genes
234 (24.1% of orthogroups, or 4,234, contained 2 or more).

These orthogroups were used to classify *Xiphophorus* genes into our three functional categories for each population (Table S3). MitoCarta 3.0 consists of 1,136 human and 1,140 mouse n-mt genes. All but three of the mouse n-mt genes (*Gm4984*, *Htd2*, and *Nsun4*) were identified, and only one human n-mt gene (*NDUFB1*) was not identified in *X. birchmanni*. Overall, 159 orthogroups were found to contain interacting n-mts, 822 to contain non-interacting n-mts, and 15,545 to contain non-n-mts. These orthogroups had 184, 1,042, and 23,043 *X. birchmanni* genes, respectively.

Mitonuclear linkage disequilibrium does not differ among gene classes in hybrid populations with segregating mitochondrial haplotypes

We sought to determine whether n-mt genes maintained higher mitonuclear LD in *Xiphophorus* hybrids due to selection against mitonuclear incompatibilities. On average, mitonuclear LD values were high across the entire nuclear genome in the CALL and CHAF hybrid populations (Figure 1, Supplemental Figure 1), meaning nuclear ancestry matched the mitochondrial haplotype (positive D' values). This result is also illustrated by differences in average nuclear ancestry between individuals with mitochondrial haplotypes from the two possible parent species (Supplemental Figure 2). These strong associations are likely driven by population structure in these two populations [33,42]. Accordingly, a partial correlation (pcor) test that controls for population structure resulted in lower mitonuclear association values (Supplemental Figure 1).

Under a model of mitonuclear coevolution, we predicted that the interacting n-mt gene class would have higher mitonuclear LD values compared to the other gene classes. However, we did not find a consistent significant difference between the gene classes for CALL (D': p = 0.44; r^2: p = 0.44, pcor: p = 0.04) or CHAF (D': p = 0.86; r^2: p = 0.29, pcor: p = 0.14; Figure 1, Supplementary Figure 1, summary statistics in Table S4). In fact, the pcor values in CALL instead showed higher mitonuclear LD in the non-n-mt gene class than in either of the two n-mt gene classes. Therefore, it does not appear that mitochondrial haplotypes have preferentially maintained associations with n-mt loci of the same ancestry at a genome-wide scale in these hybridizing populations.

We additionally tested whether interacting n-mt genes were overrepresented in the top 1% of all genes based on our association metrics (D', r^2, pcor). However, again, significant enrichment for interacting n-mt genes was inconsistent (Table S5).

266

267 **Known mitonuclear incompatibility genes show positive associations with mitochondrial**
268 **haplotype**

269 Nine of the 10 previously identified mitonuclear incompatibility genes [44] were
270 included in our CALL analysis, and only two (*NDUFS5* and *SMIM8*) were outliers for
271 mitonuclear LD (with a Z-score of >2 above the mean *D′* value) (Table S6, Figure S3).
272 Therefore, although the top 1% genome-wide LD values did not show an overrepresentation of
273 n-mt genes, we do find that some genes of known biological importance exhibit signatures of
274 elevated mitonuclear LD. Five other genes (*NDUFA13*, *MTERF4*, *ATP5MG*, *LYRM2* and
275 *RMDN3*) were also above the mean but not outliers by this definition, and the final two genes
276 (*MMUT* and *UQCRC2*) were very close to the mean. Repeating these analyses with r^2 and pcor
277 values instead of *D′* values resulted in slight changes, but the results were generally similar
278 (Table S6, Figures S4-5).

279 To see whether these genetic incompatibilities showed similar signatures across
280 populations with varying mtDNA frequencies and different parental source species, we also
281 investigated mitonuclear associations for CHAF and HUEX-STAC (Table S6, Figures S6-9. The
282 CALL and CHAF populations, which are hybrids between the same parental species, exhibit a
283 significant correlation in mitonuclear LD (*D′*) across this set of incompatibility genes (Pearson
284 correlation; $r = 0.87$; $p = 0.0051$). Comparisons with allele frequencies in HUEX-STAC, which
285 originates from a different pair of parental species, also show a positive trend, but the
286 relationships are weaker and/or non-significant with both CALL ($r = 0.34$; $p = 0.41$) and CHAF
287 ($r = 0.70$; $p = 0.051$).

288

289 **Interacting n-mt genes show enriched mitochondrial match in an older hybrid population**
290 **fixed for a mitochondrial haplotype**

291 The HUEX-STAC populations are fixed for the *X. cortezi* mitochondrial haplotype, so we
292 expected to see *X. cortezi* ancestry disproportionately represented at n-mt loci. HUEX-STAC
293 individuals generally have a large percentage of *X. cortezi* ancestry across their entire nuclear
294 genome with a mean allele frequency of 88.7% ± a standard error of 0.011% (Figure 2B). Non-
295 interacting n-mt and non-n-mt genes showed similar mean *X. cortezi* allele frequencies at 82.2 ±
296 0.02% and 83.1 ± 0.001%, respectively. Interacting n-mt genes had a higher mean allele

297 frequency (88.0 ± 0.097%, Figure 2A, Table S4). Thus, unlike in the CALL and CHAF
298 populations, individuals collected from the HUEX-STAC population did demonstrate a
299 difference in allele frequencies among gene classes (one way ANOVA: $p = 0.005$) and supported
300 the predicted higher degree of matched mitonuclear ancestries for interacting n-mt genes (Figure
301 2A). This result was also supported by a non-parametric permutation test (Figure S10).
302
303 **Interacting n-mt genes with more non-synonymous substitutions may show higher**
304 **mitonuclear LD**
305 Because many nuclear-encoded proteins may show few or no fixed differences in amino acid
306 sequence between these closely related species, we asked whether the presence of non-
307 synonymous substitutions between species affected how gene class related to mitonuclear LD.
308 There were inconsistent effects of gene class, non-synonymous substitutions, and the interaction
309 between these two across metrics and populations (Tables S7-8). There was, however, some
310 evidence that interacting n-mt genes with more non-synonymous substitutions have higher
311 mitonuclear LD compared to non-n-mt genes (Figure 3, Figures S11-12, Table S9). The
312 significant interactions in CALL and CHAF were found to be the result of a few influential genes
313 (Table S10). For example, *NDUFS5*, a gene previously shown to underly a lethal mitonuclear
314 incompatibility, consistently appeared as influential across association metrics and populations
315 where a significant interaction was found. When *NDUFS5* was removed from the dataset, the
316 difference between the interacting n-mt and non-n-mt gene classes was eliminated. Other known
317 incompatibility genes were not consistently influential, although *NDUFA13* and *ATP5MG* were
318 detected in some analyses in CHAF.
319
320
321 **DISCUSSION**
322
323 We tested the prediction that the interacting n-mt gene class would show the strongest "matched"
324 association with mitochondrial haplotype. Despite the known role of n-mt genes in hybrid
325 incompatibilities between *X. birchmanni* and *X. malinche*, the CALL and CHAF populations did
326 not support this prediction. This negative result likely speaks to the statistical power issues
327 inherent in identifying mitonuclear interactions. The sensitivity of genome-wide tests for

328 signatures of selection acting preferentially on n-mt genes will likely be affected by the structure
329 of the incompatibility and the genomic background of the genes under selection, factors that are
330 also important for detecting loci involved in nuclear-nuclear hybrid incompatibilities [51].
331 Specifically, the number of genes under selection, where those genes are distributed in the
332 genome, and/or whether only n-mt genes are involved could affect the sensitivity of these tests.
333 N-mt genes make up a small portion of the genome, and only a subset of those genes may
334 be participating in mitonuclear epistasis. Incompatibilities can potentially involve a few loci with
335 large effect size or many loci of small effect (or some combination of these extremes). The
336 analyses performed by Robles et al. [44] identified 10 n-mt genes among seven genomic regions
337 contributing to mitonuclear incompatibilities in *Xiphophorus* species hybrid populations, one of
338 which was CALL. These genes represent only 6.8% of all interacting n-mt genes and 1.0% of all
339 n-mt genes. Selection on these few genes would have to be exceptionally strong and have had
340 enough time to act to detect a significant shift in the average for the entire interacting n-mt gene
341 class. In contrast, systems such as *Tigriopus* copepods appear to have many genes involved in
342 mitonuclear incompatibilities and have proven to be powerful in advancing the field of
343 mitonuclear genetics and genomics [20,31]. Therefore, even in cases where n-mt genes are
344 contributing to negative mitonuclear epistasis, the number of loci may be too small to detect
345 genome-wide effects with the approach implemented here, which is inherently designed to detect
346 polygenic contributions to incompatibilities.
347 Elevations in background LD (e.g., due to population structure) and the distribution of n-
348 mt genes across the genome will also affect the ability to detect a difference in mtDNA ancestry
349 matching between n-mt and non-n-mt genes. Any signal of this difference arising from selection
350 on an individual n-mt target could be diluted by hitchhiking at linked non-n-mt loci [52]. This
351 effect would be particularly problematic for populations that are the result of recent admixture
352 events and, therefore, have not had many generations for recombination to break up haplotype
353 blocks in the nuclear genome [41,53], such as in CALL and CHAF (whose initial admixture was
354 estimated to be ~46 generations ago [35]). In some cases, clustering of n-mt genes in specific
355 regions of the genome could create strong targets of selection that overcome this dilution effect.
356 For example, genomic regions with a high density of n-mt genes have been identified in the
357 long-tailed finch [54] and the eastern yellow robin [55], both of which have found evidence that
358 n-mt genes disproportionately share ancestry with mtDNA.

359 Additionally, not all mitonuclear incompatibilities are necessarily mediated by n-mt
360 genes. Instead, some incompatibilities may involve nuclear-encoded proteins that are not
361 targeted to mitochondria but nonetheless create genetic interactions with mtDNA through co-
362 functionality in biochemical and signaling pathways that span multiple cellular compartments
363 [56]. For example, our definition of an n-mt gene is a gene whose product is targeted for import
364 into mitochondria. However, mitochondria are a part of a complex cellular network, and any
365 number of cellular processes could affect mitochondrial function. Genes involved in these
366 processes are, by our definition, non-n-mt genes. Therefore, any contributions of non-n-mt genes
367 to mitonuclear incompatibilities would directly undermine search strategies based on partitioning
368 the genome into n-mt and non-n-mt categories.
369 However, these challenges were somewhat mitigated in our analyses of CALL and
370 CHAF by including the number of non-synonymous substitutions for each gene in the model. By
371 looking only at genes with substitutions between species, we uncovered a pattern in which
372 interacting n-mt genes tended to show a stronger relationship between mitonuclear association
373 and the number of non-synonymous substitutions than non-n-mt genes (Figure 3, Figures S11-
374 12). This stronger, positive relationship could imply that putative functional divergence results in
375 stronger selection to match ancestry to the mitochondrial haplotype if the divergence occurs in an
376 interacting n-mt gene. However, this relationship was found to be disproportionately driven by a
377 few genes, namely *NDUFS5*, a known incompatibility gene [43]. Nevertheless, this result
378 suggests that including protein sequence divergence could increase the power to detect the
379 effects of mitonuclear incompatibilities, even when using genome-wide tests in a system with
380 few genes of large effect. We speculate that these effects will be easiest to detect in interacting n-
381 mt genes but that this result does not rule out the possibility of other genes contributing to
382 mitonuclear incompatibilities.
383 Despite the challenges outlined above, one of the three hybrid populations we
384 investigated (HUEX-STAC) supported our prediction of enrichment in matched ancestry for n-
385 mt genes. Notably, this hybrid population is older than CALL and CHAF, estimated to have
386 formed ~250 generations ago [33,35], and it is the only population we investigated that is fixed
387 for a mitochondrial haplotype from just one of the parental species. Its older age suggests there
388 has been more time for incompatible genotype combinations to be purged from the population
389 and for recombination to have broken up haplotype blocks in the nuclear genome, thereby

390 limiting the effects of hitchhiking. We therefore speculate that the age of HUEX-STAC and the
391 fixation of a mitochondrial haplotype may have made selection on n-mt loci stronger and/or
392 more effective. In such populations, half of the possible mitonuclear combinations have been
393 removed, so selection for matched ancestry would consistently favor the same n-mt allele. In
394 contrast, when mitochondrial haplotypes are still segregating, the fitness effects of an n-mt allele
395 may change depending on an individual's mtDNA. Because mitochondrial and nuclear loci are
396 not physically linked, very strong selection would be needed to preserve mitonuclear LD in
397 segregating populations [57].
398 Despite the potential importance of the age and fixed mitochondrial haplotype of the
399 HUEX-STAC population, previous studies have found examples of hybrid populations with both
400 fixed and segregating mitochondrial haplotypes that show a disproportionate effect of selection
401 on n-mt genes during hybridization (Table S1). Therefore, expanding the availability of genomic
402 datasets from hybrid systems and performing a formal meta-analysis would be important future
403 directions to assess relevant variables, such as whether mitochondrial haplotypes are segregating
404 in the population, the age of the hybrid population, the number of loci involved and their effect
405 sizes, and the genetic structure of the population. Reanalysis of publicly available genome-level
406 data from hybrid populations may also be a valuable approach to control for method of ancestry
407 inference and metric of association between mtDNA and nuclear loci.
408 Another variable that differentiates the HUEX-STAC population is the parental makeup
409 of the hybrids. Although geographically distinct, CHAF and CALL share parental species (*X.*
410 *birchmanni* and *X. malinche*). In contrast, HUEX-STAC was formed by hybridization between
411 *X. birchmanni* and *X. cortezi*. These three species demonstrate mitonuclear discordance [58] and
412 a history of introgression [42], both of which may impact mitonuclear incompatibilities in these
413 hybrids, potentially resulting in stronger selection for matched mitonuclear ancestry. Given these
414 different demographic histories, each of the hybrid populations we analyze here is expected to
415 have numerous unique features, so caution is warranted in attributing any of our results to a
416 single feature (such as the fixed mitochondrial haplotype in HUEX-STAC). Nevertheless, we
417 view the potential role of fixed vs. segregating mitochondrial haplotypes to be an important area
418 of future research given the small number of studies conducted to date.
419 Overall, our findings suggest that splitting up the genome into n-mt and non-n-mt gene
420 classes may be an underpowered method to detect effects of mitonuclear epistasis in recently

admixed populations, even when there are clear biological effects of said epistasis. Although such tests for genome-wide signatures may offer a relatively unbiased approach, they may be susceptible to false negatives. Therefore, important contributions of mitonuclear interactions to speciation may remain unrecognized in many systems.

ACKNOWLEDGEMENTS

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. We thank three anonymous reviewers for their insightful comments.

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. malinche*	*X. malinche*	*X. cortezi*
Minor Parent Species	*X. birchmanni*	*X. birchmanni*	*X. birchmanni*
Number of Individuals	281	250	255
Mitochondrial Haplotype Frequency[a]	63.7% *X. birchmanni*	93.5% *X. malinche*	97.8% *X. cortezi*
Average Major Parent Nuclear Ancestry	52.0%	71.3%	88.7%
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 = 243.

FIGURES

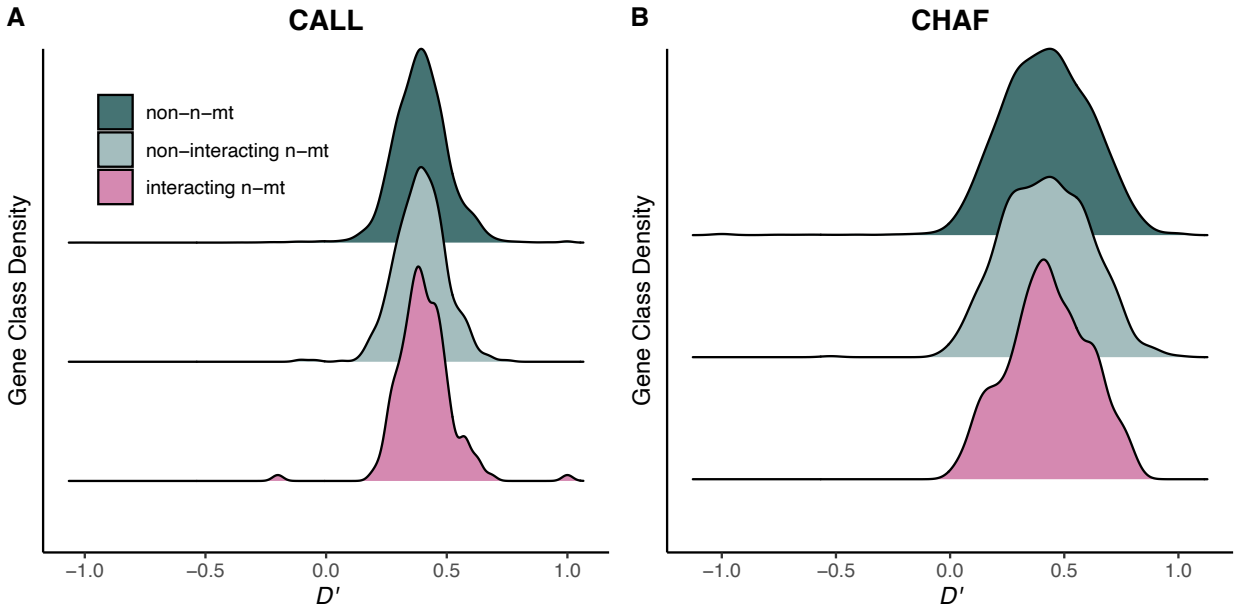


Figure 1. Distribution of mitonuclear D′ *values for each gene class in the CALL (A) and CHAF (B) populations. One-way ANOVA testing if D′ value differs among gene classes: CALL p = 0.44 (A); CHAF p = 0.86 (B). Summary statistics are reported in Table S4, and other mitonuclear LD metrics are shown in Figure S1.*

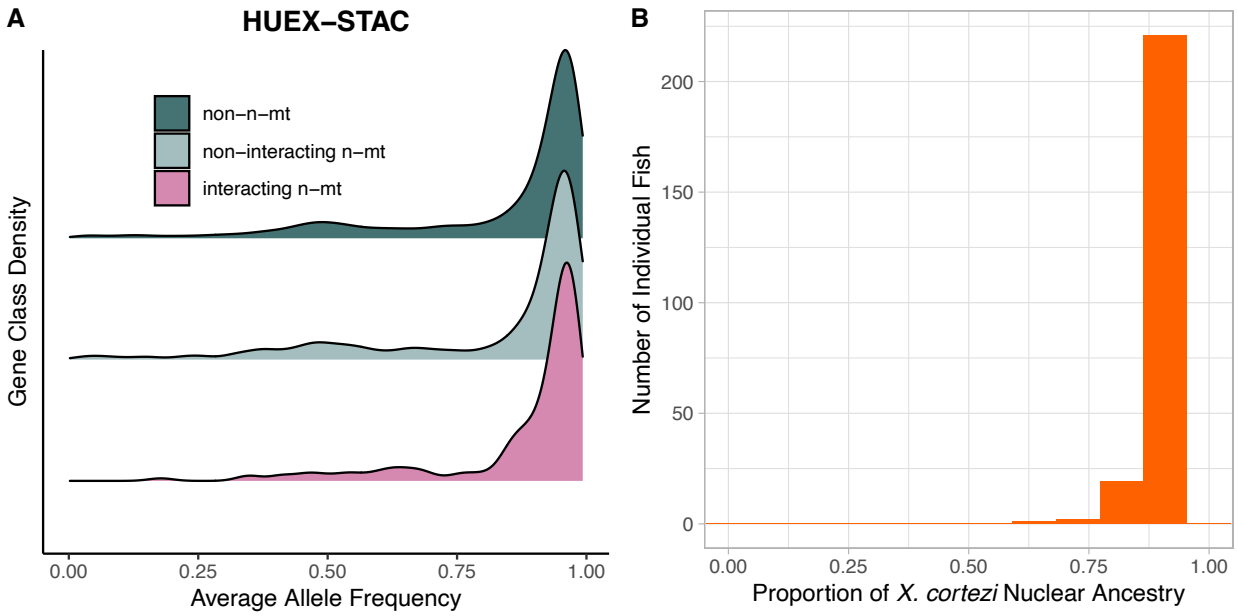


Figure 2. HUEX-STAC analyses for A) allele frequencies for each gene class and B) nuclear genome-wide ancestry for the HUEX-STAC population, a hybrid population fixed for the X. cortezi *mitogenome. A one-way ANOVA found gene class to significantly affect allele frequency (p = 0.006); see Table S4 for summary statistics showing differences in means Figure S10 for permutation test demonstrating the robustness of this result. The stacked bar histogram in B was made with 12 bins.*

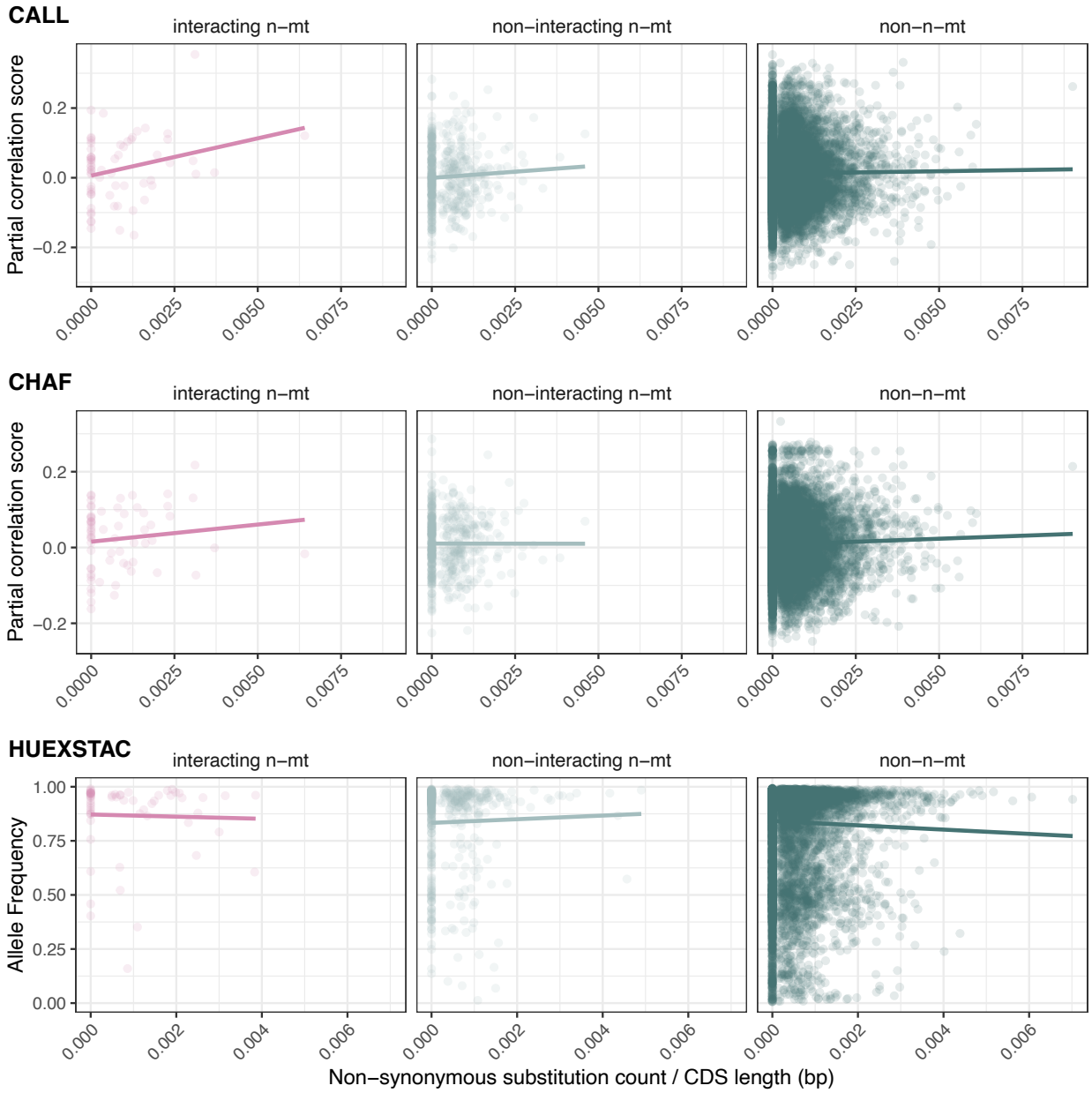


Figure 3. Regression analysis of mitonuclear association metric by non-synonymous substitutions. Partial correlation values for CALL do not show a significant interaction (type III ANOVA p = 0.073) but interacting n-mts have a stronger linear relationship than the non-n-mt relationship (p = 0.034). Partial correlation values for CHAF does not have a significant interaction (p = 0.67) or difference in slopes (p = 0.50). X. cortezi *allele frequencies in HUEX-STAC also do not show a significant interaction (p = 0.43) or a difference in slopes (p = 0.85). Other metrics of mitonuclear LD can be found in Figures S10-11. Statistics can be found in Tables S7-10.*

REFERENCES

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

17. 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.
18. Hill GE. 2020 Mitonuclear Compensatory Coevolution. *Trends in Genetics* **36**, 403–414. (doi:10.1016/j.tig.2020.03.002)
19. 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)
20. 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)
21. 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)
22. 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)
23. 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)
24. 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)
25. 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)
26. 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)
27. 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)
28. Burton RS, Barreto FS. 2012 A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol* **21**, 4942–4957. (doi:10.1111/mec.12006)
29. Coyne JA. 1992 Genetics and speciation. *Nature* **355**.
30. Gershoni M, Templeton AR, Mishmar D. 2009 Mitochondrial bioenergetics as a major motive force of speciation. *BioEssays* **31**, 642–650. (doi:10.1002/bies.200800139)
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. *Mol Ecol* **34**. (doi:https://doi.org/10.1111/mec.70106)
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. Schumer M, Brandvain Y. 2016 Determining epistatic selection in admixed populations. *Mol Ecol* **25**, 2577–2591. (doi:10.1111/mec.13641)
52. Barton NH. 2000 Genetic hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355**, 1553–1562. (doi:10.1098/rstb.2000.0716)
53. 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)
54. 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)
55. 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)
56. 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)
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)

Supplementary Materials for the manuscript:

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

SUPPLEMENTAL FIGURES

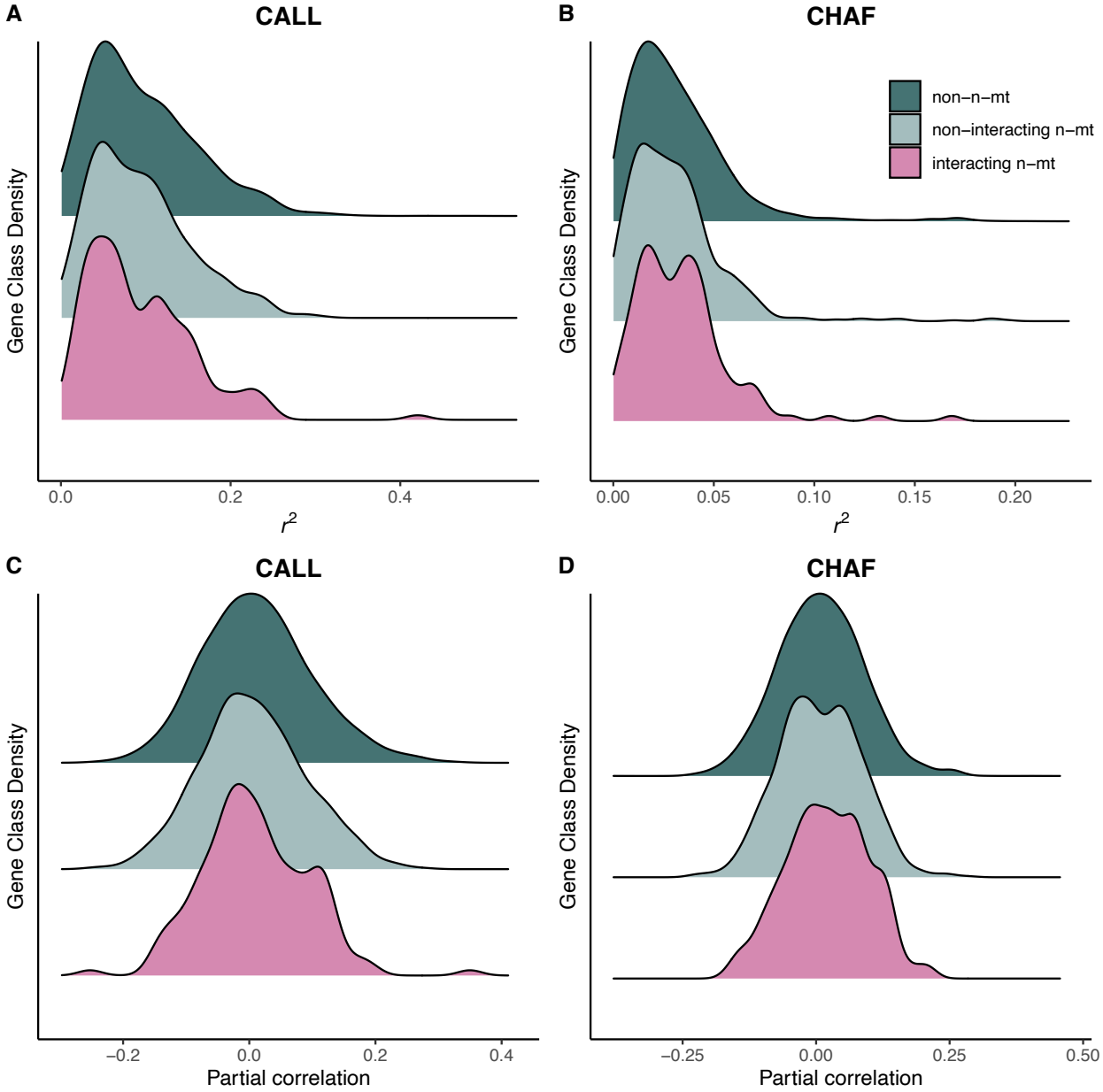


Figure S1. Other metrics of mitonuclear association values for each gene class in the CALL (A and C) and CHAF (B and D) populations. Association is calculated as r^2 *in panels A and B and as a partial correlation that accounts for genome-wide ancestry in panels C and D. Note the change of scale on the x-axis across populations and LD statistics. P-value for a one-way ANOVA testing if gene class affects LD value: CALL* r^2 *0.44 (A), CHAF* r^2 *0.29 (B), CALL partial correlation 0.04 (C), CHAF partial correlation 0.14 (D). Summary statistics are reported in Table S4.*

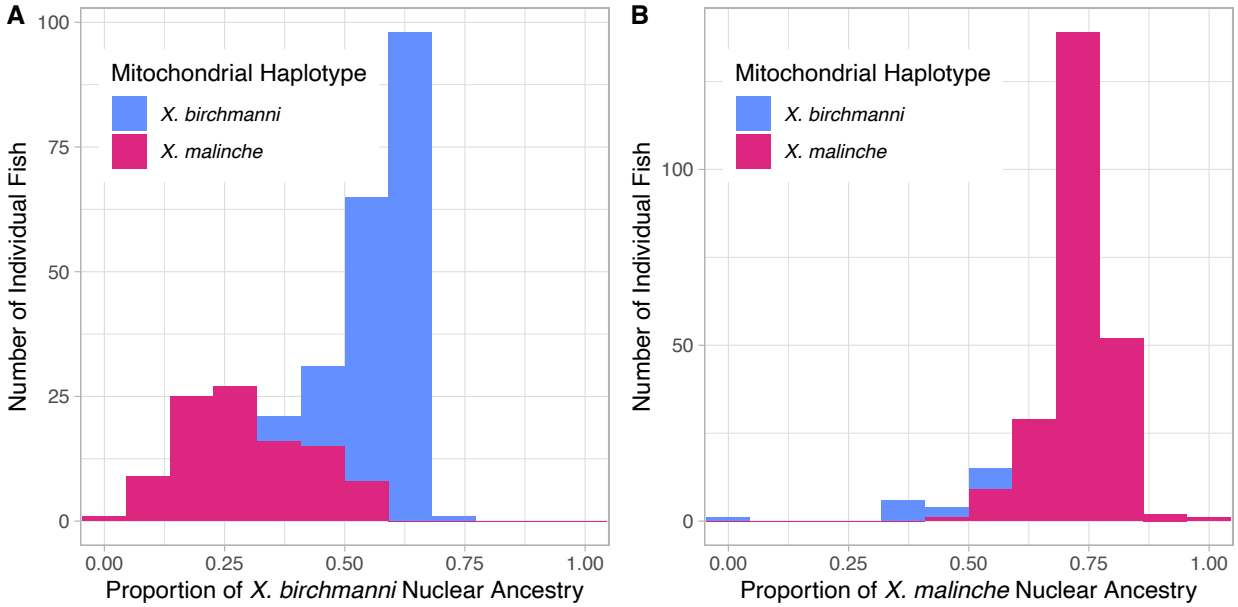


Figure S2. Stacked bar histogram for the proportion of the nuclear genome derived from the major parent for each fish in A) CALL and B) CHAF. Proportion is calculated as the sum of all major parent AIMs divided by number of AIMs. Each fish is colored by which of the two segregating mitochondrial haplotypes it has (blue = X. birchmanni*, fuchsia =* X. malinche*). Both stacked bar histograms were made with 12 bins. Note that y-axis upper limits differ between the panels.*

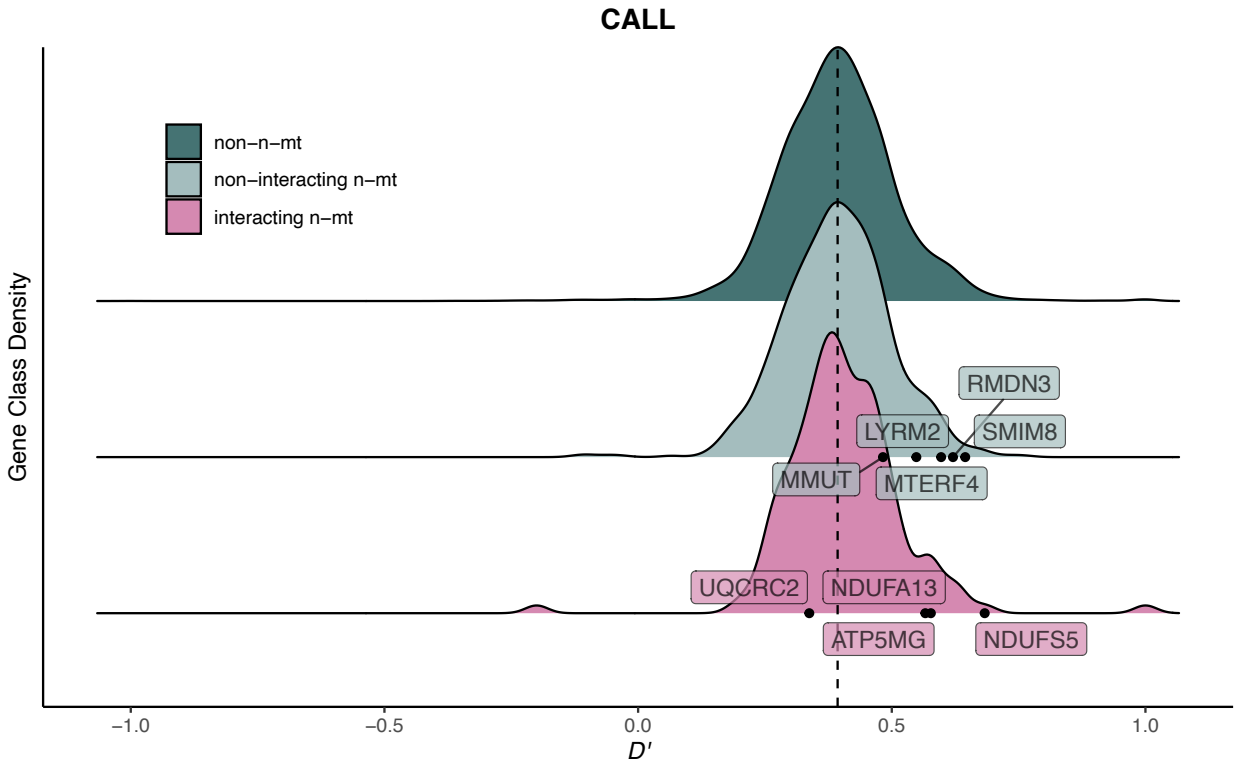


Figure S3. Location of previously identified incompatibility genes [1] in the CALL population D' *values. Dashed line indicates genome-wide mean* D'.

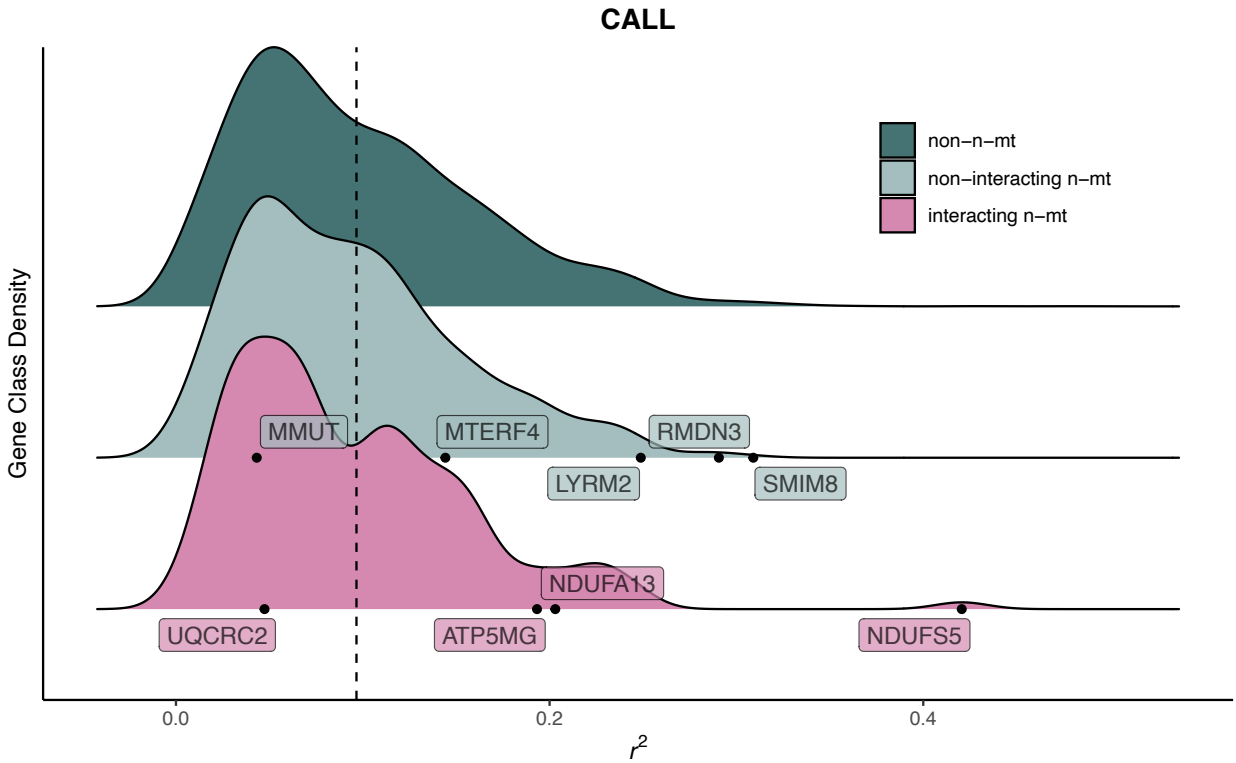


Figure S4. Location of previously identified incompatibility genes [1] in the CALL population r^2 *values. Dashed line indicates genome-wide mean* r^2.

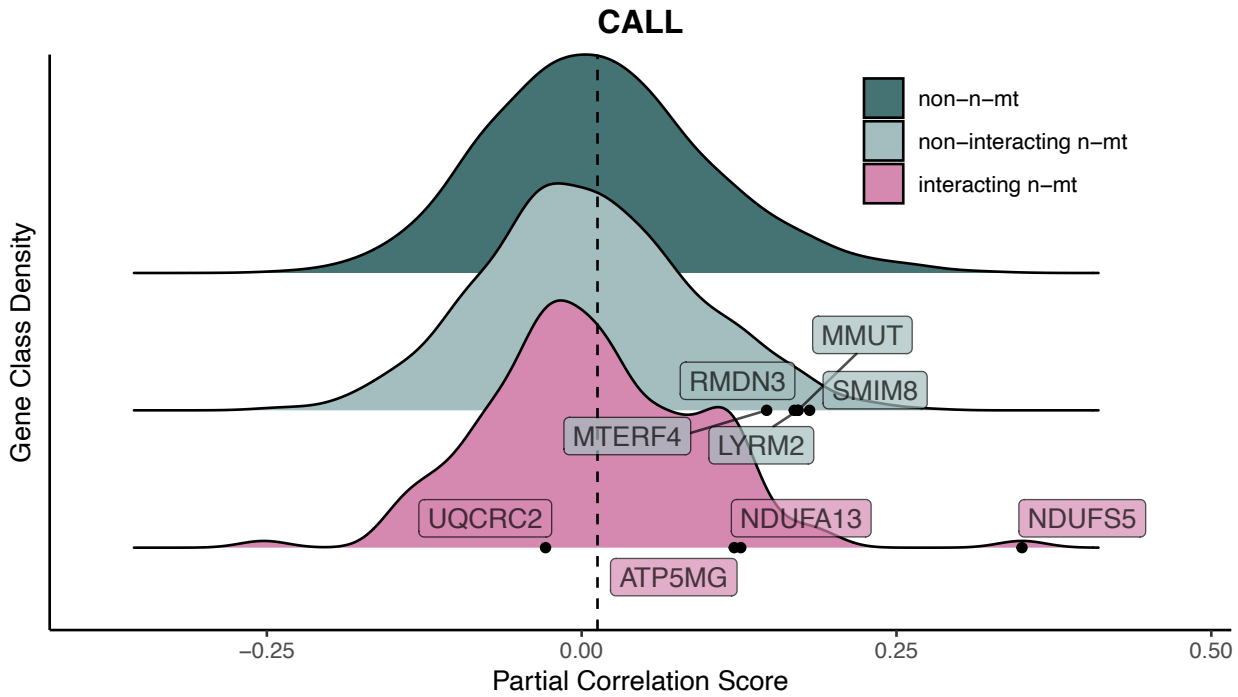


Figure S5. Location of previously identified incompatibility genes [1]in the CALL population partial correlation values. Dashed line indicates genome-wide mean partial correlation score.

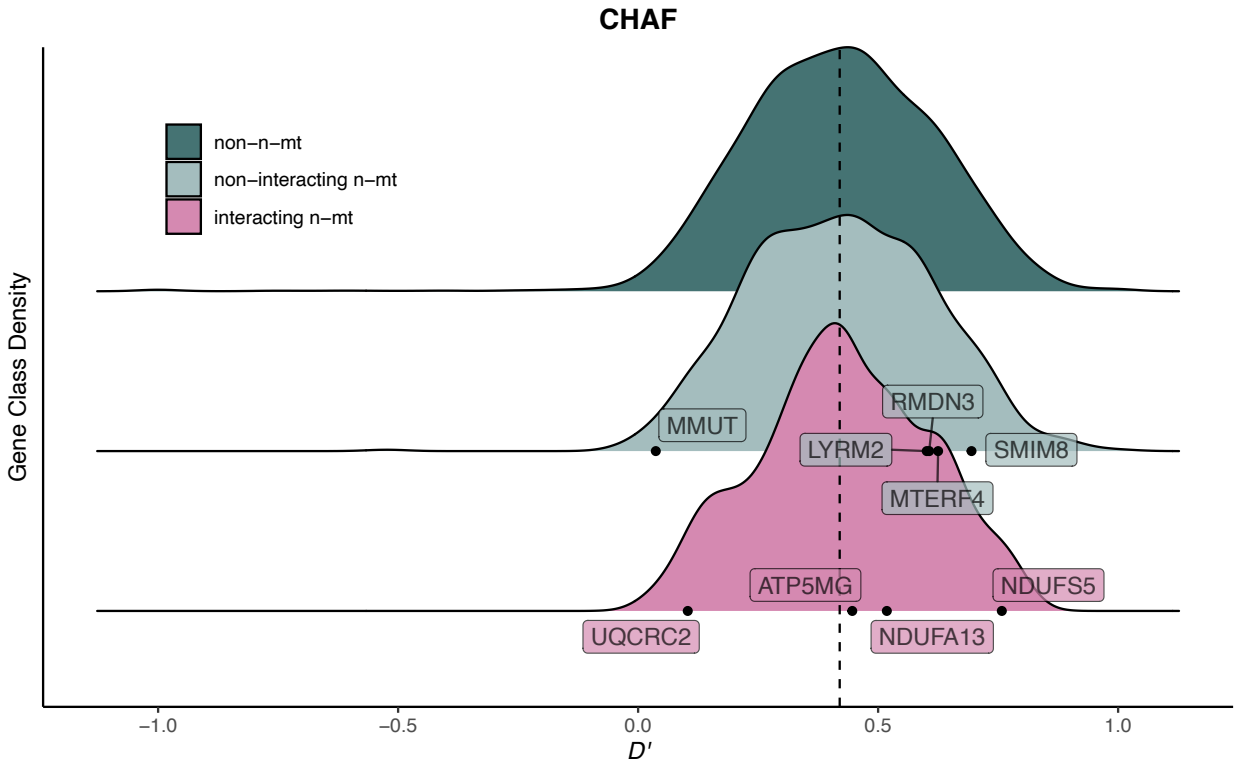


Figure S6. Location of previously identified incompatibility genes [1] in the CHAF population D' *values. Dashed line indicates genome-wide mean* D'.

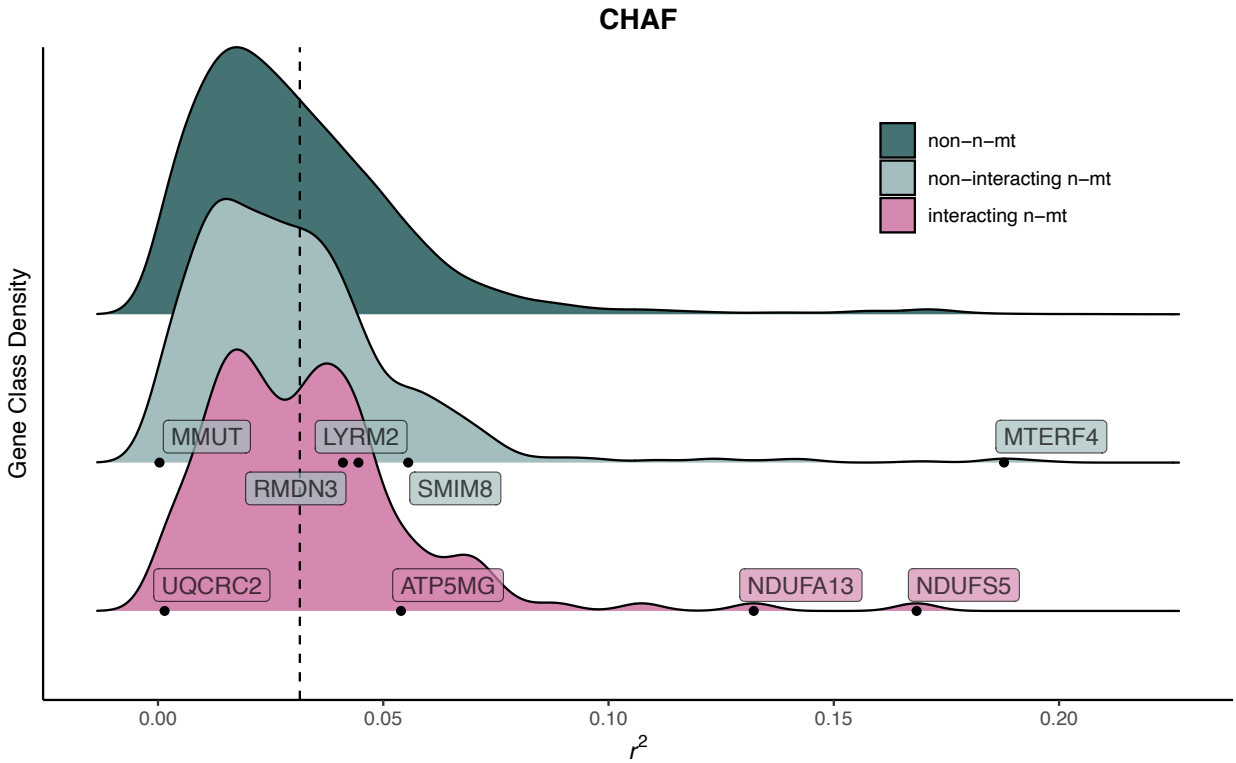


Figure S7. Location of previously identified incompatibility genes [1]in the CHAF population r^2*values. Dashed line indicates genome-wide mean* r^2.

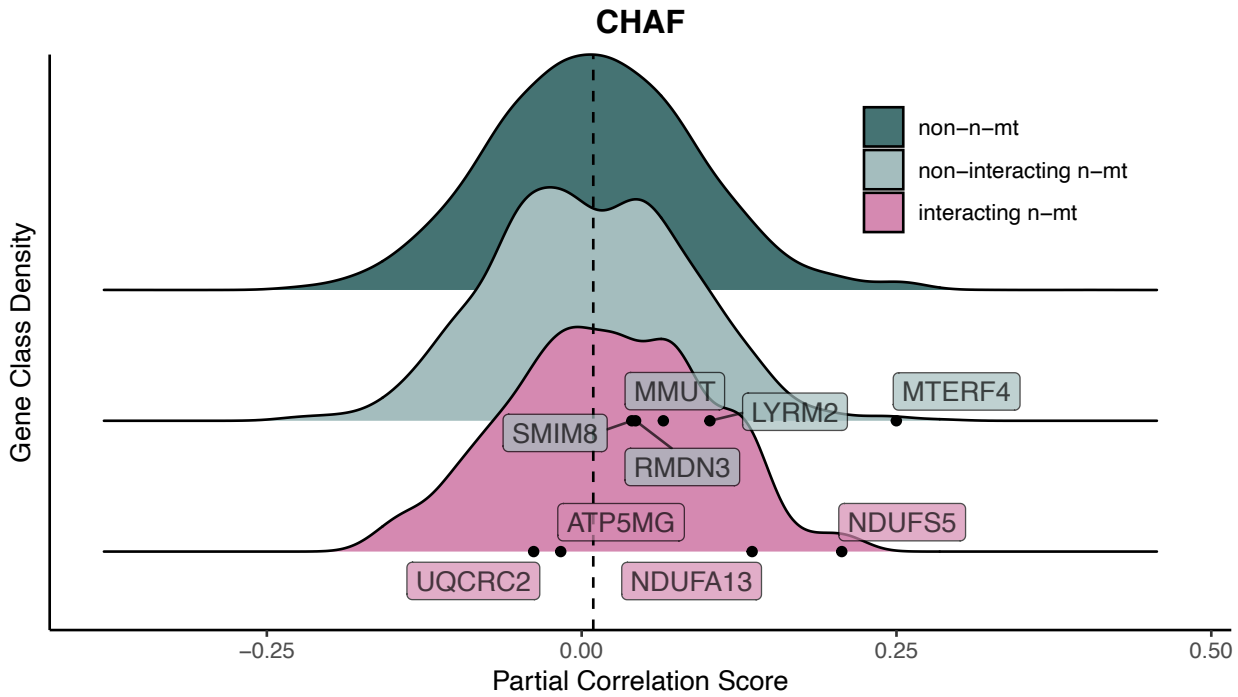


Figure S8. Location of previously identified incompatibility genes [1]in the CHAF population partial correlation values. Dashed line indicates genome-wide mean partial correlation score.

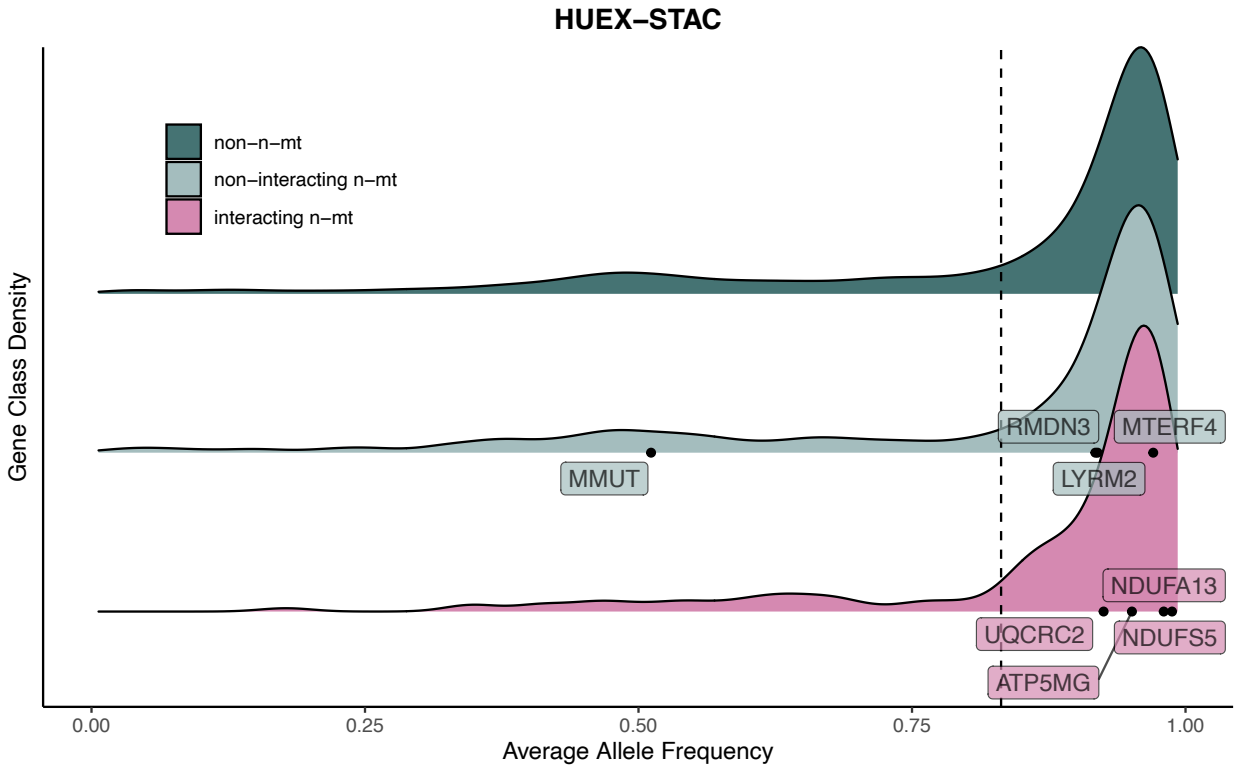


Figure S9. Location of previously identified incompatibility genes [1] in the HUEX-STAC population X. cortezi *allele frequency values. Dashed line indicates genome-wide mean allele frequency.*

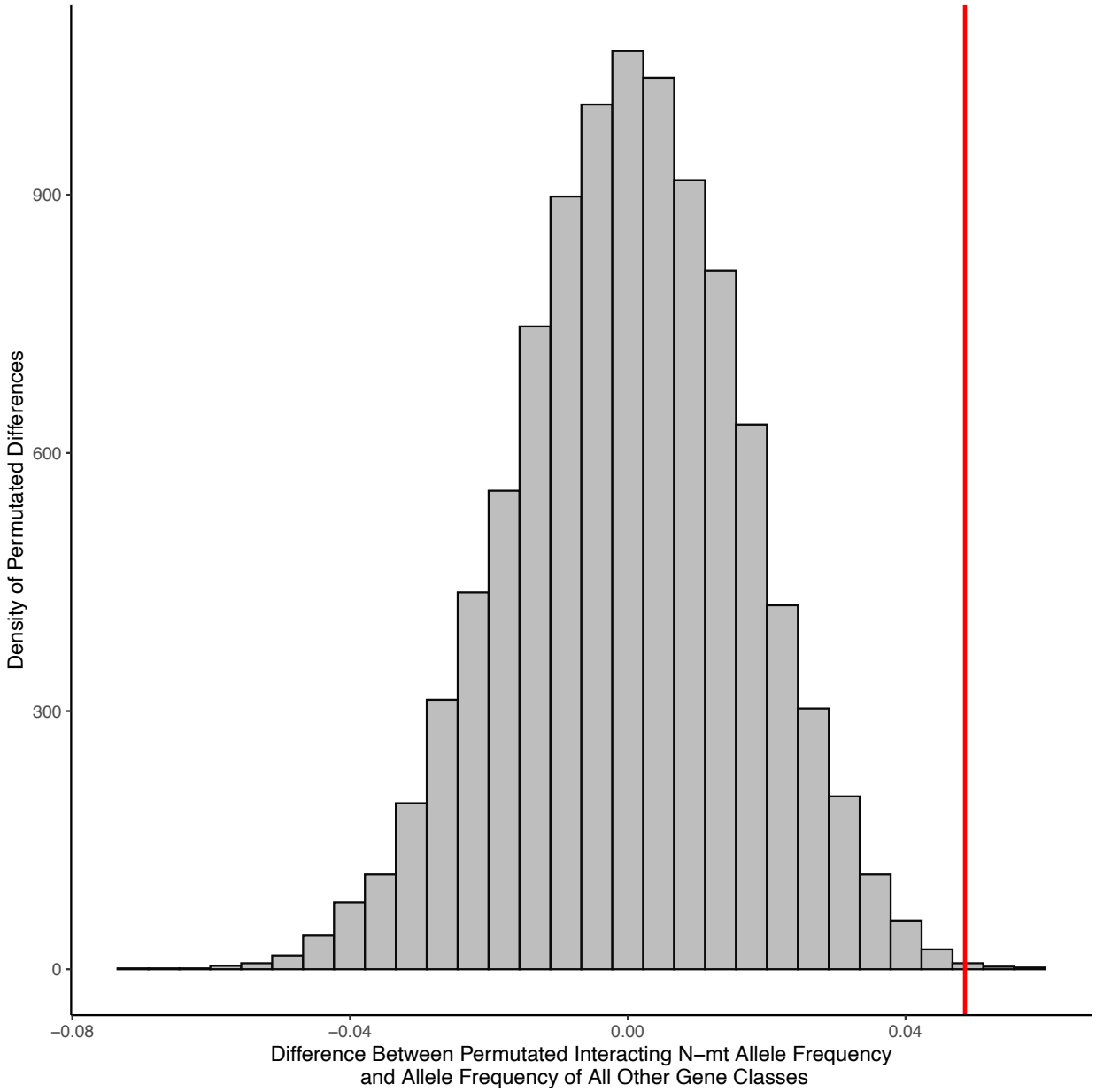


Figure S10. Histogram demonstrating permuted differences in mean allele frequency between all interacting n-mt genes and the mean of all other genes. Red line represents the observed difference in interacting n-mt gene class and all other classes. In this permutation test, we found the interacting n-mt gene class to be different from the other two gene classes ($p = 9 \times 10^{-4}$*), supporting the parametric analysis (ANOVA) in this population.*

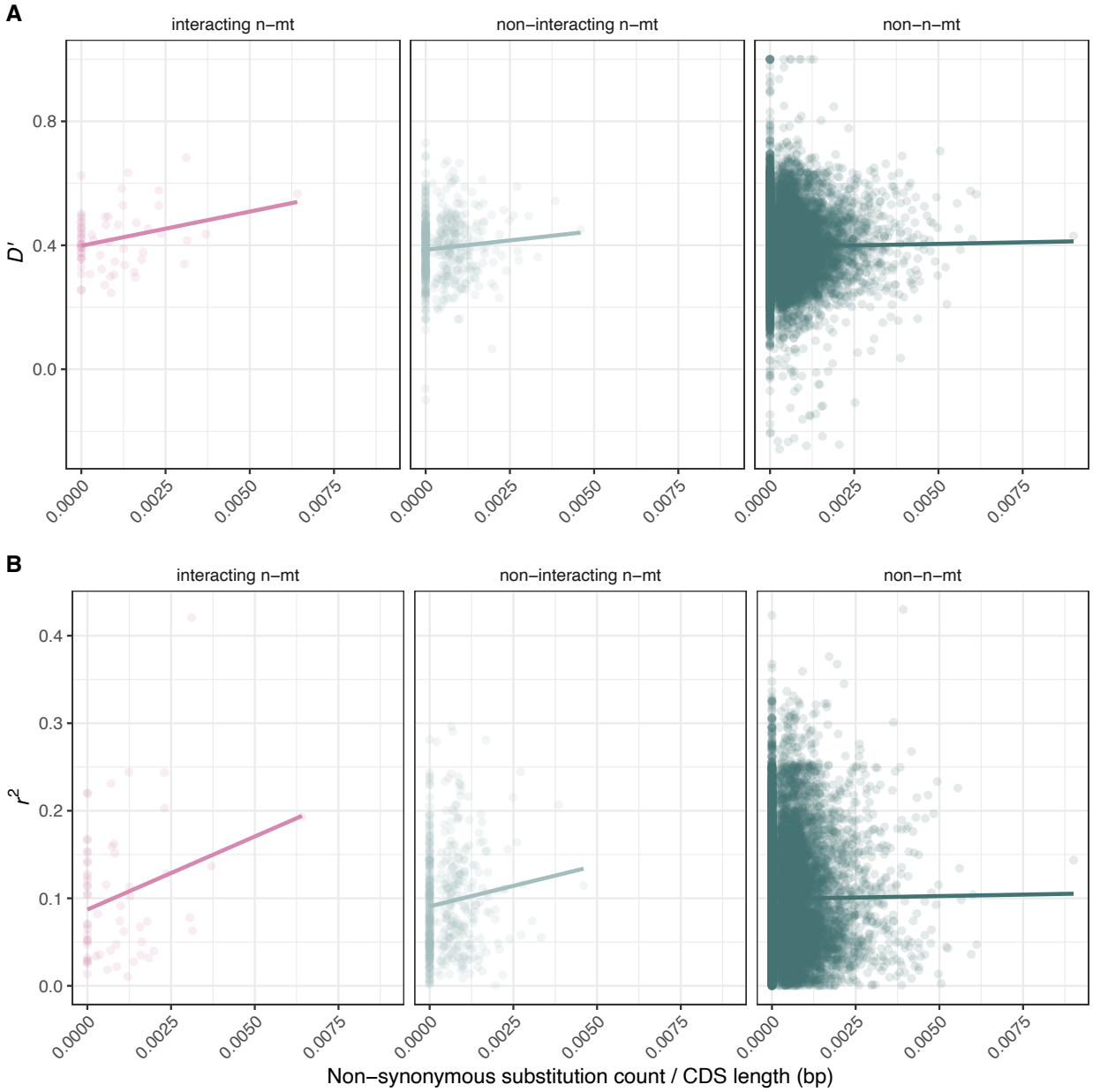


Figure S11. Non-synonymous substitution model for the CALL population for the A) D' *and B)* r^2 *metrics. In a model that accounts for an interaction between gene class and non-synonymous substitutions, interacting n-mt genes have a stronger relationship to the* r^2 *(p = 0.019) but not the* D' *(p = 0.092) metric (Table S9), but see points driving this relationship in Table S10 and summary statistics in Table S7.*

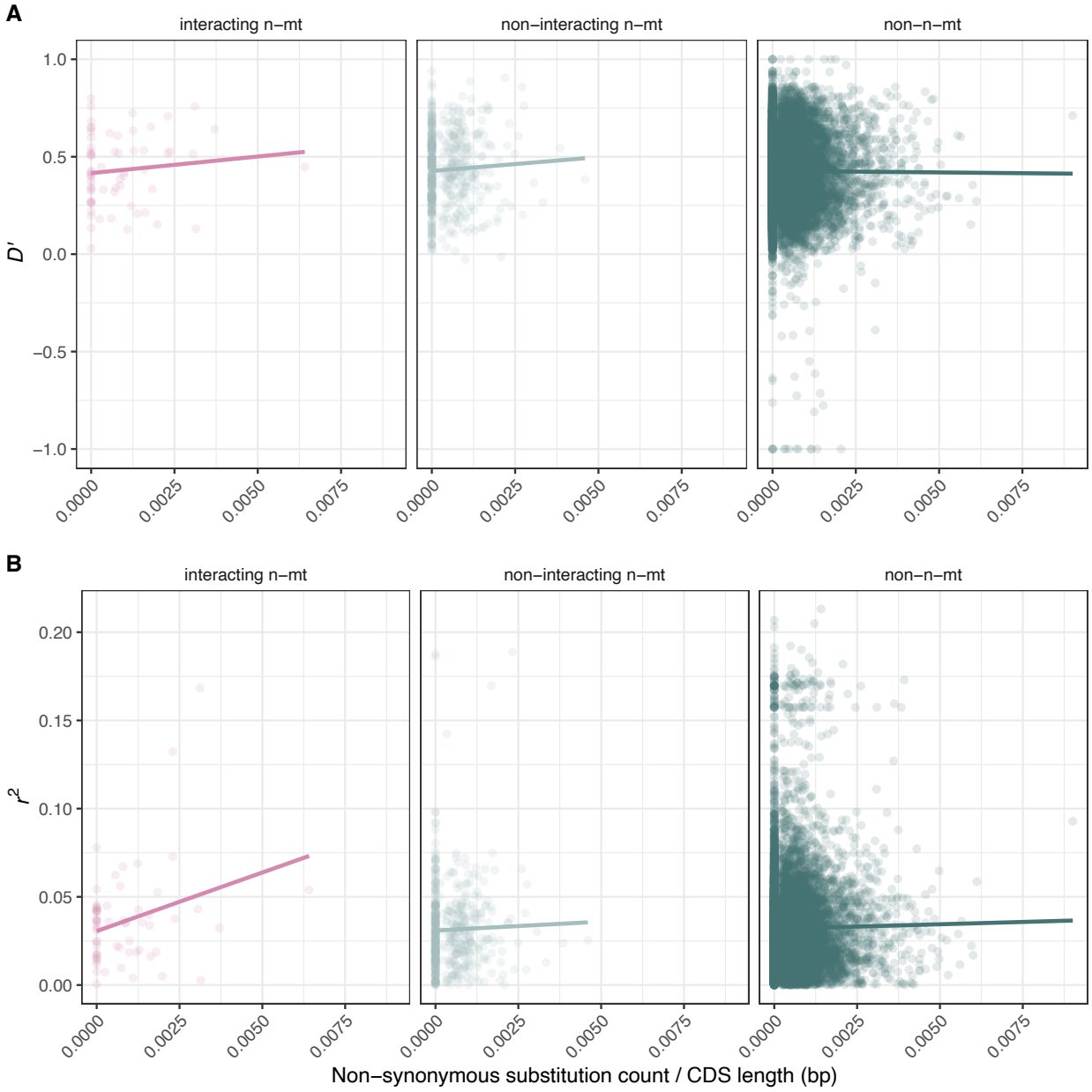


Figure S12. Non-synonymous substitution model for the CHAF. population for the A) D' *and B)* r^2 *metrics. In a model that accounts for an interaction between gene class and non-synonymous substitutions, interacting n-mt genes have a stronger relationship to the* r^2 *(p = 0.027) but not the* D' *(p = 0.37) metric (Table S9), but see points driving this relationship in Table S10 and summary statistics in Table S7.*

SUPPLEMENTAL TABLES

Table S1. 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 [2]	Historical introgression	*Potamopyrgus estuarinus* and *P. kaitunuparaoa*	No	Yes
Fish				
Kato et al. 2023 [3]	Historical introgression	*Chaenogobius annularis* and a ghost lineage	No	No
Mammals				
Kwon et al. 2022 [4]	Interspecific hybridization	*Bos taurus* and *B. indicus*	No	Yes
Shi et al. 2025 [5]	Interspecific hybridization	*B. taurus* and *B. indicus*	No	No[d]
Jensen et al. 2023 [6]	Historical introgression	*Allochrocebus* and *Chlorocebus* lineages	No	Yes
Zhu & Evans 2023 [7]	Historical introgression	*Macaca arctoides* and *M. fascicularis aurea*	No	No
Evans et al. 2021 [8]	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 [9]	Interspecific hybridization	*M. arctoides* and *M. sinica* or *M. fascicularis*	No	No
Sloan et al. 2015 [10]	Intraspecific admixture between populations	*Homo sapiens*	Yes	No
Sharbrough et al. 2017 [11]	Historical introgression	*H. sapiens* and *H. neanderthalensis*	No	Yes

	Historical introgression	*H. sapiens* and Denisovans	No	No
Zaidi & Makova 2019 [12]	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 [13]	Intraspecific admixture between populations	*Poephila acuticauda acuticauda* and *P. a. hecki*	Yes	Yes[e]
Morales et al. 2018 [14]	Intraspecific admixture between populations	northern and southern populations of *Eopsaltria australis*	Yes	Yes, but see [15]
Low et al. 2024 [15]	Intraspecific admixture between populations	inland and coastal mitochondrial lineages of *E. australis*	Yes	Yes[e]
Rancilhac et al. 2024 [16]	Interspecific hybridization	*Pogoniulus pusillus* and *P. chrysoconus*	Yes	No
Musher et al. 2024 [17]	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 within a single individual, they are omitted from our investigation into whether fixing a mitochondrial haplotype affects the sensitivity of genome-wide tests of shared n-mt and mtDNA ancestry.

Table S2. List of human interacting n-mt genes used in this study, modified from [10].

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

Table S3. Pre-filtered AIMs and genes (in parentheses) belonging to each gene 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)

Table S4. Summary statistics of mitonuclear association values for each hybrid population.

Population	Statistic	Gene Class	Count	Mean	Standard Deviation
CALL	*D′*	Interacting n-mt	148	0.405	0.116
		Non-interacting n-mt	883	0.392	0.105
		Non-n-mt	19,312	0.393	0.114
	r^2	Interacting n-mt	148	0.094	0.064
		Non-interacting n-mt	883	0.094	0.059
		Non-n-mt	19,312	0.097	0.063
	Partial correlation	Interacting n-mt	148	0.008	0.084
		Non-interacting n-mt	881	0.005	0.084
		Non-n-mt	19,293	0.013	0.087
CHAF	*D′*	Interacting n-mt	147	0.428	0.176
		Non-interacting n-mt	883	0.421	0.189
		Non-n-mt	19,296	0.420	0.198
	r^2	Interacting n-mt	147	0.033	0.024
		Non-interacting n-mt	883	0.030	0.024
		Non-n-mt	19,296	0.0315	0.025
	Partial correlation	Interacting n-mt	148	0.021	0.078
		Non-interacting n-mt	882	0.007	0.077
		Non-n-mt	19,297	0.009	0.080
HUEX-STAC	Allele Frequency	Interacting n-mt	162	0.880	0.156
		Non-interacting n-mt	947	0.822	0.222
		Non-n-mt	20543	0.831	0.211

Table S5. 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	$p = 0.33$
	One-sided Fisher's Exact Test for Count Data r^2 values	$p = 0.76$
	One-sided Fisher's Exact Test for Count Data partial correlation metric	$p = 0.03$
CHAF	One-sided Fisher's Exact Test for Count Data on D' values	$p = 0.31$
	One-sided Fisher's Exact Test for Count Data r^2 values	$p = 0.42$
	One-sided Fisher's Exact Test for Count Data partial correlation metric	$p = 0.25$

Table S6. Measures of mitonuclear association for incompatible n-mt genes previously identified by Robles et al. [1].

		CALL[c]			CHAF[c]			HUEX-STAC[c]
Gene[a]	**Incompatible mtDNA[b]**	***D′***	**r^2**	**Partial correlation**	***D′***	**r^2**	**Partial correlation**	**Allele frequency**
NDUFA13 (g3279.t1)	*X. malinche*	0.58 (1.61)	0.20 (1.70)	0.13 (1.30)	0.52 (0.50)	0.13 (4.02)	0.14 (1.57)	0.99 (0.74)
MTERF4 (g3322.t1)	*X. malinche*	0.60 (1.79)	0.14 (0.76)	0.15 (1.54)	0.63 (1.04)	0.19 (6.24)	0.25 (3.00)	0.97 (0.66)
ATP5MG (g3595.t1)	*X. malinche*	0.57 (1.52)	0.19 (1.54)	0.12 (1.24)	0.45 (0.13)	0.05 (0.90)	-0.02 (-0.32)	0.95 (0.57)
NDUFS5 (g8054.t1)	bidirectional	0.68 (2.54)	0.42 (5.17)	0.35 (3.87)	0.76 (1.71)	0.17 (5.47)	0.21 (2.46)	0.98 (0.70)
MMUT (g15308.t1)	bidirectional	0.48 (0.78)	0.04 (-0.85)	0.17 (1.82)	0.04 (-1.94)	0.00 (-1.24)	0.06 (0.69)	0.51 (-1.51)
SMIM8 (g15060.t1)	*X. birchmanni*	0.64 (2.21)	0.31 (3.39)	0.18 (1.93)	0.69 (1.39)	0.06 (0.96)	0.04 (0.38)	ND
LYRM2 (g15051.t1)	*X. birchmanni*	0.55 (1.36)	0.25 (2.43)	0.17 (1.83)	0.60 (0.92)	0.04 (0.52)	0.10 (1.16)	0.92 (0.42)
RMDN3 (g15056.t1)	*X. birchmanni*	0.62 (1.99)	0.29 (3.09)	0.17 (1.79)	0.61 (0.94)	0.04 (0.38)	0.04 (0.42)	0.92 (0.41)
UQCRC2 (g11575.t1)	*X. birchmanni*	0.34 (-0.49)	0.05 (-0.78)	-0.03 (-0.47)	0.20 (-1.60)	0.00 (-1.20)	-0.04 (-0.59)	0.93 (0.44)

[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, partial correlation, 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.

Table S7. Summary statistics for the non-synonymous substitution model. NS is the count of non-synonymous substitutions per coding sequence divided by the coding sequence length.

	Gene class	Number of genes	Mean NS ± standard deviation
CALL	Interacting n-mt	63	0.00085 ± 0.0012
	Non-interacting n-mt	473	0.00051 ± 0.0007
	Non-n-mt	11330	0.00053 ± 0.0007
CHAF	Interacting n-mt	63	0.00085 ± 0.0012
	Non-interacting n-mt	473	0.00051 ± 0.0007
	Non-n-mt	11324	0.00053 ± 0.0007
HUEX-STAC	Interacting n-mt	60	0.00089 ± 0.0011
	Non-interacting n-mt	461	0.00046 ± 0.0007
	Non-n-mt	10587	0.00041 ± 0.0006

Table S8. Type III ANOVA results for non-synonymous substitution models

	Predictor variables	p-value
D′		
CALL	**Intercept**	**< 2e-16**
	Gene class	0.40
	Non-synonymous substitutions / CDS (NS)	0.16
	Gene class * NS	0.11
CHAF	**Intercept**	**< 2e-16**
	Gene class	0.93
	Non-synonymous substitutions / CDS (NS)	0.54
	Gene class * NS	0.35
r^2		
CALL	**Intercept**	**< 2e-16**
	Gene class	**0.042**
	Non-synonymous substitutions / CDS (NS)	0.42
	Gene class * NS	**0.011**
CHAF	**Intercept**	**< 2e-16**
	Gene class	0.80
	Non-synonymous substitutions / CDS (NS)	0.12
	Gene class * NS	0.085
Partial correlation		
CALL	**Intercept**	**< 2e-16**
	Gene class	0.062
	Non-synonymous substitutions / CDS (NS)	0.23
	Gene class * NS	0.073
CHAF	**Intercept**	**< 2e-16**
	Gene class	0.67
	Non-synonymous substitutions / CDS (NS)	**0.0028**
	Gene class * NS	0.67
Allele Frequency		
HUEX-STAC	**Intercept**	**< 2e-16**
	Gene class	0.53
	Non-synonymous substitutions / CDS (NS)	**0.0034**
	Gene class * NS	0.83

Table S9. Results of summary() *function testing the difference between interacting n-mt genes' relationship between non-synonymous substitutions and mitonuclear LD and the non-n-mt genes' relationship. Please see attached excel.*

Table S10. Influential points of non-synonymous substitution analysis

Mitonuclear LD metric	Influential genes	Mitonuclear LD value	Normalized non-synonymous substitution count	Removed significance of interaction?[a]	Removed significance of interacting n-mt interaction?[b]
CALL					
r^2	*NDUFS5*	0.42	0.0031	Y	Y
Partial correlation	*NDUFS5*	0.35	0.0031	Y	Y
CHAF					
r^2	*NDUFS5*	0.17	0.0031	Y	Y
	NDUFA13	0.13	0.0023	Y	Y
	ATP5MG	0.45	0.0064	Y	N
	LOC102221627[c]	0.003	0.0031	N	N

The above points were identified as influential (i.e., significantly shaping the slope of the regression) in a model of mitonuclear LD vs an interaction between gene class and non-synonymous substitutions per coding sequence length.

[a]When this gene was removed from the model, did the interaction between gene class and normalized non-synonymous substitution count become non-significant?

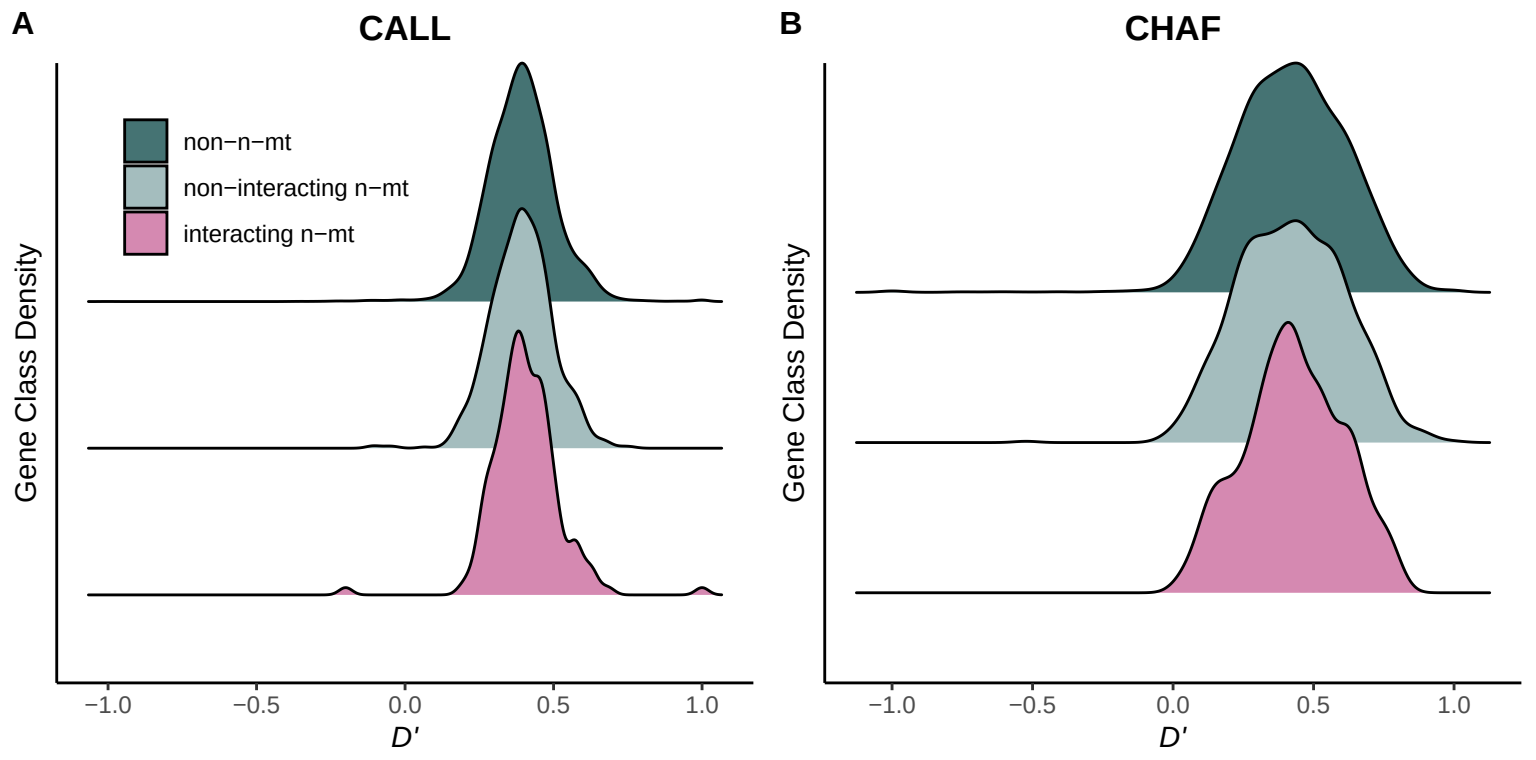
[b]When this gene was removed from the model, did the interacting n-mt gene class no longer have a significantly sharper slope than the non-n-mt gene class?

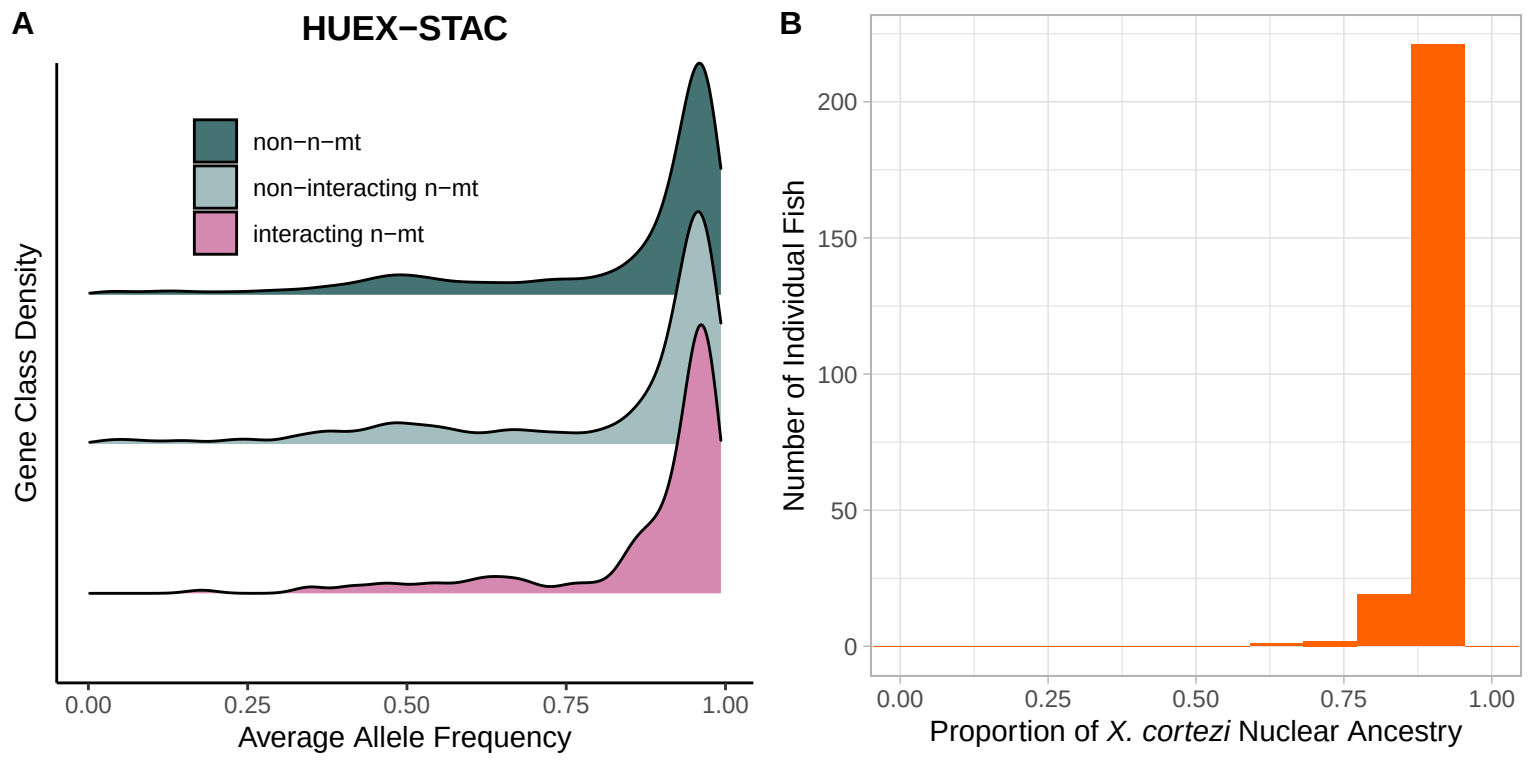
[c]LOC102221627 is annotated as “cytochrome b-c1 complex subunit 8”

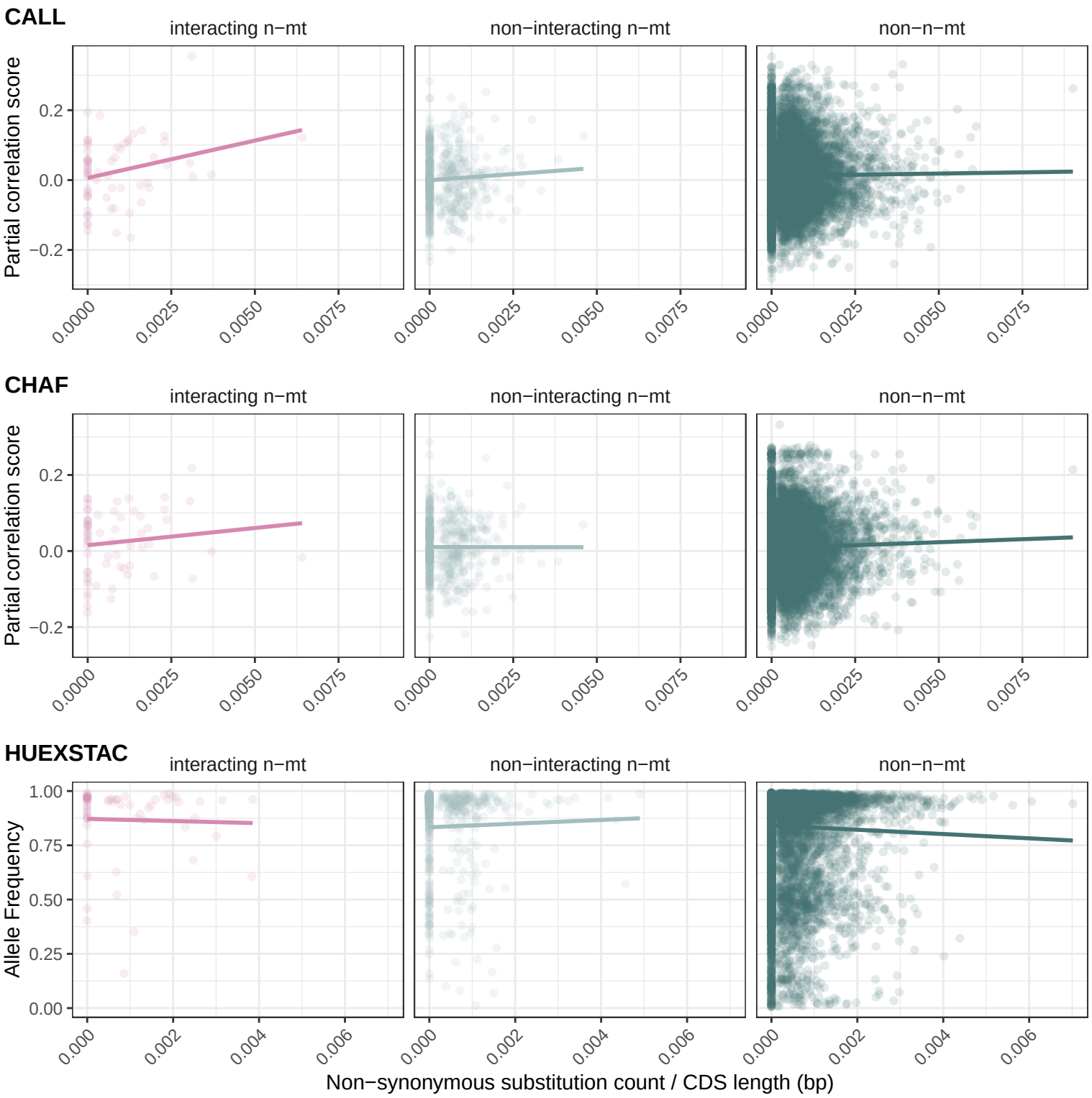
REFERENCES

1. Robles N V. *et al.* 2025 Admixture mapping reveals evidence for multiple mitonuclear incompatibilities in swordtail fish hybrids. *Mol Ecol* **34**. (doi:https://doi.org/10.1111/mec.70106)
2. 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)
3. 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)
4. 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)
5. 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)
6. 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)
7. Zhu J, Evans BJ. 2023 Mitonuclear Interactions and the Origin of Macaque Societies. *Genome Biol Evol* **15**. (doi:10.1093/gbe/evad010)
8. 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)
9. 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)
10. 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)
11. 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)
12. 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)
13. 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)
14. 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)
15. 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)
16. 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)
17. 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)


A
CALL
non−n−mt
non−interacting n−mt
interacting n−mt
Gene Class Density
−1.0
−0.5
0.0
0.5
1.0
D′
B
CHAF
Gene Class Density
−1.0
−0.5
0.0
0.5
1.0
D′


A
HUEX−STAC
non−n−mt
non−interacting n−mt
interacting n−mt
Gene Class Density
0.00
0.25
0.50
0.75
1.00
Average Allele Frequency
B
Number of Individual Fish
200
150
100
50
0
0.00
0.25
0.50
0.75
1.00
Proportion of X. cortezi Nuclear Ancestry


CALL
interacting n−mt
non−interacting n−mt
non−n−mt
Partial correlation score
0.2
0.0
−0.2
0.0000
0.0025
0.0050
0.0075
CHAF
interacting n−mt
non−interacting n−mt
non−n−mt
Partial correlation score
0.2
0.0
−0.2
0.0000
0.0025
0.0050
0.0075
HUEXSTAC
interacting n−mt
non−interacting n−mt
non−n−mt
Allele Frequency
1.00
0.75
0.50
0.25
0.00
0.000
0.002
0.004
0.006
Non−synonymous substitution count / CDS length (bp)