Genetic and environmental interactions outweigh mitonuclear coevolution for complex traits in *Drosophila*

Leah J. Darwin[1,2], Faye A. Lemieux[2], Rebecca Z. Bachtel[2], Jack H. Blocker[2], Camille P. Brown[2], Jacob D. Lerman[2], Olivia C. Maule[2], Yevgeniy Raynes[1,2], and David M. Rand[1,2]

[1]Center for Computational Molecular Biology, Brown University, Providence, RI
[2]Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, RI

Abstract

The interdependent relationship between mitochondrial and nuclear genomes is a powerful model for understanding how epistasis shapes the architecture and evolution of complex traits. Once considered a neutral marker, mitochondrial DNA variation is now recognized as critical to phenotypic evolution because of its epistatic interactions and history of coevolution with the nuclear genome. A central challenge in evolutionary genetics is to quantify the relative importance of stabilizing and directional selection shaping complex trait distributions within and among species. Both can act on interacting and/or co-evolving genes contributing to quantitative traits, but resolving their relative roles is complicated by the complex architecture of most traits. Here, we use a panel of 90 *Drosophila* mitonuclear genotypes to quantify the relative contributions of mitochondrial, nuclear, and environmental variation and their interactions to four metabolically demanding complex traits. We sample both within-species and between-species mitochondrial variation and observe stronger interaction effects attributable to within-species variation, consistent with stabilizing selection maintaining mitonuclear function. Additionally, culturing the flies on a mitochondrial Complex I inhibitor, rotenone, reveals significant genotype x environment (GxE and GxGxE) interaction effects, providing insight into how genetic variation can be maintained across changing environments. Our results have broader implications in medicine, where mitochondrial DNA donors with longer purifying selection histories may be safer for mitochondrial replacement therapies.

Introduction

The observation that many traits vary continuously among individuals raises fundamental questions about the genetic basis and maintenance of phenotypic diversity. These quantitative traits are often challenging to study because they arise from the combined segregation of many polymorphic loci and are sensitive to different environmental contexts [1]. Interactions between genes (G×G, epistasis) and between genes and the environment (G×E) can further complicate trait expression, leading to unpredictable outcomes. The extent to which quantitative traits are shaped by additive versus interactive effects is a matter of considerable debate. Some argue that although gene interactions contribute to variation in most biological systems, population-level genetic variance is mostly additive [2]. Others contend that epistasis is pervasive and that additivity can arise as an emergent property of complex epistatic networks [3]. Interaction effects could help to explain “missing” heritability common in genome-wide association studies [4,5]. They may also account for how genetic variation can be maintained under stabilizing selection [6,7], which appears to be a major force acting on quantitative traits [8,9].

Fully comprehensive tests for epistatic effects on traits are both experimentally and computationally intractable because the number of variant combinations– and thus statistical model terms or experimental treatments– increases exponentially as more loci are considered. One effective strategy to address this challenge is to reduce the number of testable interactions by considering sets of genes rather than individual gene combinations. A natural example of such sets is the mitochondrial (mtDNA) and nuclear genomes, which jointly encode oxidative phosphorylation (OXPHOS) complexes responsible for energy synthesis [10]. Their interdependence and the functional importance of OXPHOS to most traits makes the mitonuclear system a powerful model for studying adaptive evolution [11,12]. Much work has focused on the adaptive divergence of mitochondrial genes

and the response of the nuclear genome [13–15], with the expectation of positive selection on nuclear genes compensating for deleterious mutations in mtDNA. This model predicts that mismatched mitochondrial and nuclear genomes would reduce hybrid fitness, observed in systems such as marine copepods [16, 17], yeast [18], fruit flies [19, 20], nematodes [21], and swordtail fishes [22, 23]. However, analyses of the sites of amino acid contact between mtDNA- and nuclear-encoded proteins in bivalves [24] and mammals [25] have not revealed consistently significant dN/dS ratios, which would be expected under positive selection. Some instances of elevated dN/dS for these proteins are attributed to relaxed rather than positive selection [26] suggesting that such incompatibilities may be the exception rather than the rule [27]. Conflicting empirical findings highlight the need for alternative models of mitonuclear coevolution and its impact on quantitative traits.

In this study, we constructed a panel of mismatched mitonuclear genotypes in *Drosophila* to elucidate the contributions of genetic and environmental interaction effects to four metabolically demanding traits: body weight, climbing velocity, flight performance, and development time. We introduced 10 mtDNAs from each of two diverged natural populations of *D. melanogaster* and two additional species (*D. simulans* and *D. yakuba*) onto two common *D. melanogaster* nuclear backgrounds. These mtDNA replacements were built in duplicate generating 80 intraspecific and eight interspecific mitonuclear genotypes, each of which was exposed to the OXPHOS Complex I inhibitor rotenone to assess how genetic interactions are influenced by environmental perturbation. Through quantifying how G×G and G×E interactions shape phenotypic variation within populations, our results show how mildly deleterious mtDNA mutations that persist under stabilizing selection influence trait variation. Understanding the impact of mildly deleterious mutations in different genomic and environmental contexts provides a framework for anticipating their consequences in clinical settings such as mitochondrial replacement therapy [28, 29]. Our results further demonstrate that purifying selection acting on mtDNA prevents the accumulation of deleterious alleles, challenging compensatory models of mitonuclear coevolution.

Results

We constructed a panel of 22 mitochondrial haplotypes, including within-species variation from Beijing and Zimbabwe *D. melanogaster* (*D. mel*) populations and from two related species, *D. yakuba* (*D. yak*) and *D. simulans* (*D. sim*). Virgin females of each mitochondrial haplotype were backcrossed for 10 generations to males of the two North American *D.mel* nuclear backgrounds, OreR and DGRP375 (hereafter labeled as mtDNA;nucDNA, e.g., Bei05;OreR or ZW144;375). This process of backcrossing was done in duplicate, creating two "builds" of each mitonuclear line to control for potential nuclear variation (theoretically $1/2^{10}$, or 0.0976% of the nuclear genome after 10 generations of backcrosses) that could confound the effect of mtDNA genotype on phenotype. We report significant correlations between builds for all traits (Figure S1). In total, 90 lines were assayed over the course of this study, including the two lines retaining their "native" mitochondrial haplotypes, OreR;OreR and DGRP375;DGRP375 (Figure 1a).

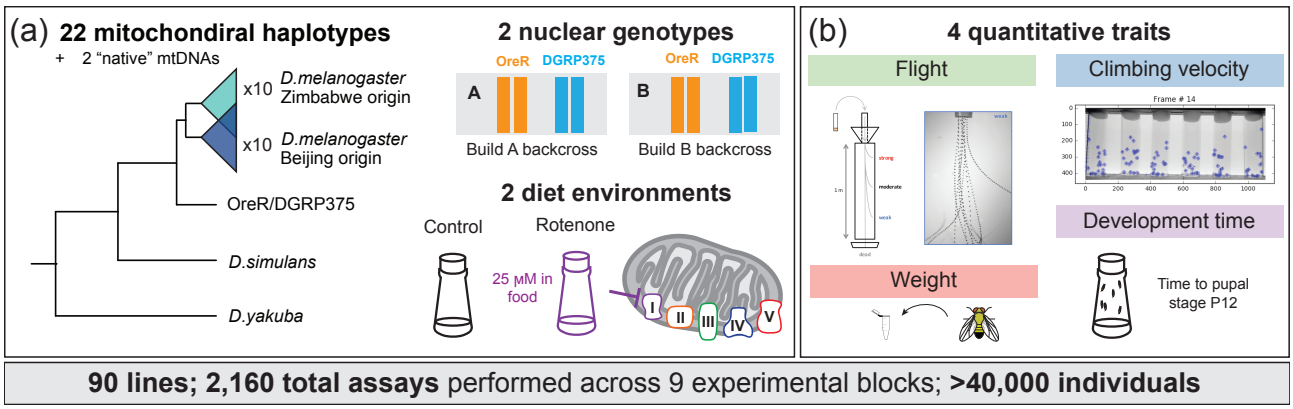

Figure 1: Experimental design and genotypes sampled. (a) Mitochondrial phylogeny shows 10 haplotypes from a *D. melanogaster* population in Zimbabwe, 10 from Beijing, 2 from *D. simulans* and *D. yakuba*, and the two "native" mtDNAs corresponding to the two nuclear genotypes used. Each mtDNA was backcrossed in duplicate builds A and B to the OreR and DGRP375 North American *D. melanogaster* nuclear backgrounds and exposed to control and rotenone diets. (b) Four quantitative traits were measured: flight performance (image source: [30], climbing velocity (image source: [31]), body weight, and development time. In total, 90 total lines were quantified across nine experimental blocks (2,160 assays).

We measured the quantitative traits development time, climbing velocity, flight performance, and body

weight in two environments: a control environment and one containing rotenone, a naturally occurring OXPHOS Complex I inhibitor (Figure 1b). For the weight, climbing, and flight traits, males and females were sorted and assayed separately. There were no significant correlation coefficients between any combination of two phenotypes for both females (Figure S2a, Table S1) and males (Figure S2b, Table S2). We assessed the reproducibility of our trait measurements by retesting a subset of lines. For all retested phenotypes, there were significant rank correlations between the original measurements and the repeated ones ($p < 0.01$, Table S3, Figure S4).

Nuclear, environmental, and mitochondrial main effects and interactions

Using a linear model, we detected significant main effects of nuclear genotype and rotenone treatment environment across all four traits (see Figure 2a-c, Extended Data Table 1, and Equation 2 in Methods for details). Nuclear genome had the largest effect on variance, with the principal component analysis of all phenotypes showing that PC1 (reflecting nuclear background effects) explained 51.11% of the total variance (Figure S3). We focused our results on a sex-stratified model and used a semi-partial R^2 (R^2_β) to quantify relative effect sizes. Across all phenotypes, the nuclear main effect was significant ($p < 0.0125$) in both sexes (Figure 2a), except for male weight which was only significant at $p < 0.025$. In both males and females, the nuclear genome had the largest effect for climbing ($R^2_{\beta^*} = 0.374$ for females; $R^2_{\beta^*} = 0.183$ for males), flight ($R^2_{\beta^*} = 0.752$ for females; $R^2_{\beta^*} = 0.335$ for males), and development time ($R^2_{\beta^*} = 0.217$, not stratified by sex) (Extended Data Table 1). We found that animals with the DGRP375 nuclear background developed slower than those with the OreR background but performed better in flight and climbing, which may be related smaller body weight (Figure 2a). The effect of rotenone was significant for the climbing trait ($p < 0.001$, $R^2_{\beta^*} = 0.034$ for females; $p < 0.001$, $R^2_{\beta^*} = 0.024$), development time ($p < 0.001$, $R^2_{\beta^*} = 0.024$), and flight in males ($p < 0.001$, $R^2_{\beta^*} = 0.128$) (Figure 2b, Extended Data Table 1). Treatment with rotenone negatively affected most traits, reducing flight and climbing performance and and slowing development time (Figure 2b).

The mitochondrial genome had significant main effects ($p < 0.00125$) for weight (both sexes) and for female climbing (Extended Data Table 1, Figure 2c). For the weight phenotype, the mitochondrial effect size was particularly notable as it was greater than the nuclear effect in both females ($R^2_{\beta^*} = 0.188$), and males ($R^2_{\beta^*} = 0.141$) (Extended Data Table 1). The significant mtDNA effects were not driven by similar effects from within either the Beijing or Zimbabwe *D. mel* populations, or by extreme phenotypes from the two outgroup species *D. sim* and *D. yak*. Rather, when the average mitochondrial effects were ranked by size, the Beijing and Zimbabwe mtDNAs are not distinctly clustered and the *D. sim* and *D. yak* mtDNAs were in the middle of this range (Figure 2c). To further investigate the relationship between the mitochondrial phylogeny and phenotypes across the four traits within nuclear background, sex, and treatment groups, we measured phylogenetic signal. Specifically, we used Blomberg's K, which tests for a relationship between a phylogeny and a trait. Under a mitonuclear coadaptation model, divergent mtDNAs (*D. yak*, *D. sim*) in a *D. mel* nuclear genomic background should disrupt mitonuclear communication and have larger phenotypic effects than conspecific mtDNAs. Conversely, variation within *D. mel* populations (Beijing vs. Zimbabwe) should show greater similarity within clades if traits have been shaped by shared selective pressures. Our tests revealed no Blomberg's K values for all combinations of sex, treatment, nuclear background, and phenotype (Table S4). Instead, the outgroups had intermediate trait measurements relative to the Beijing and Zimbabwe clades for all traits (Extended Data Figure 2).

Like the main effects, second-order interaction variance also differed across traits and sexes, indicating that each trait had a distinct architecture. The mitonuclear genetic epistasis (G$\times$G) was significant only for male weight ($p < 0.001$) (Figure 3a, Extended Data Table 1). In the climbing trait, GxG was significant at the less conservative threshold $p < 0.025$ in males ($p = 0.015$) and marginally insignificant in females ($p = 0.026$) (Extended Data Table 1). We find the mitochondrial-by-treatment GxE to be more common across traits. The mtDNA GxE effect was significant in the development time phenotype ($p = 0.006$) and in both females ($p < 0.001$) and males ($p < 0.001$) in the climbing phenotype (Figure 3b, Extended Data Table 1). In the development time trait, the mitochondrial-by-treatment effect size ($R^2_{\beta^*} = 0.047$) was larger than the treatment main effect ($R^2_{\beta^*} = 0.024$) (Extended Data Table 1). For all significant second order interaction effects, the slopes of the reaction norms were not more similar within mtDNA group. Furthermore, the two outgroups *D. sim* and *D. yak* appeared to be less reactive than many of the *D. mel* mtDNAs.

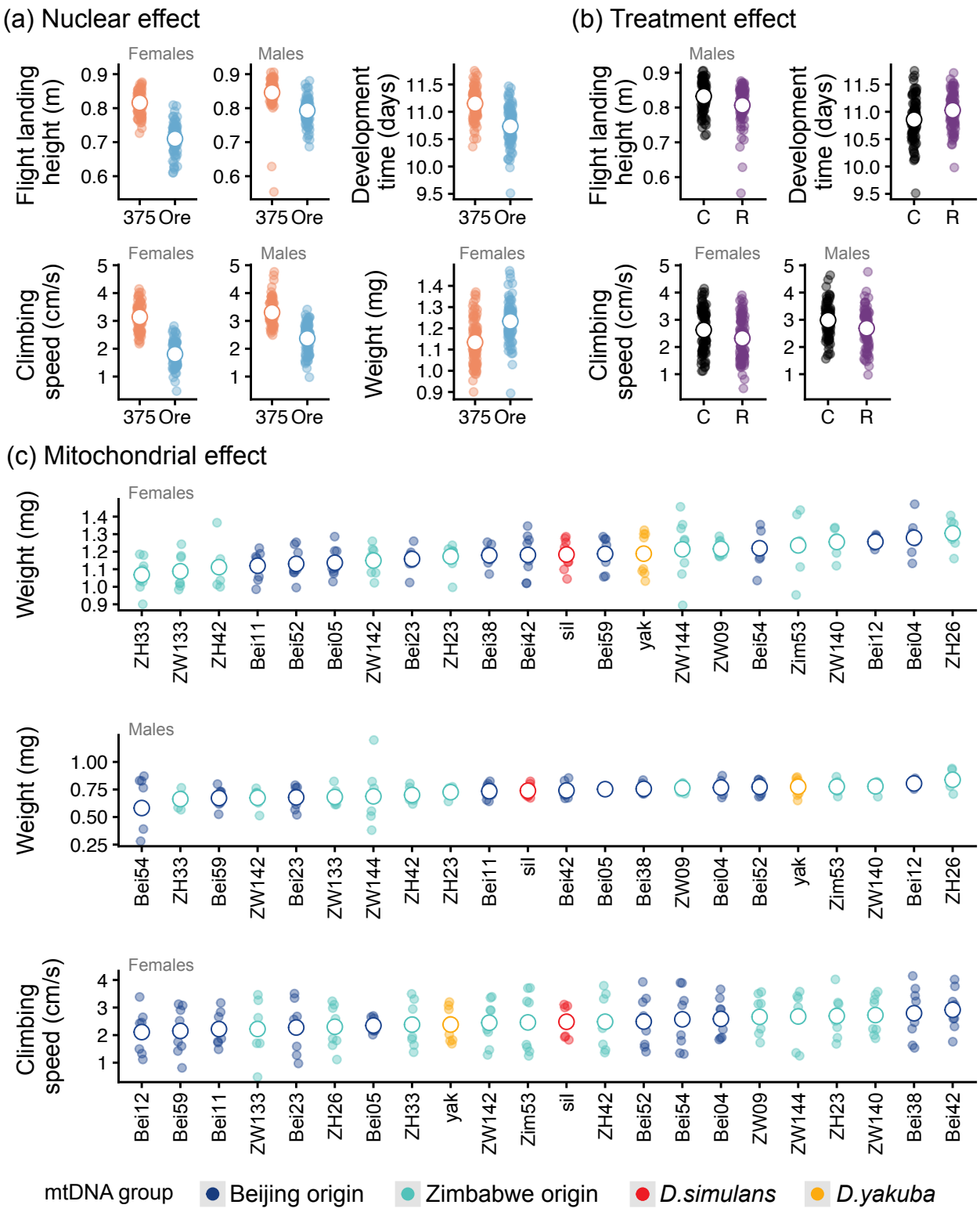

Figure 2: Significant main effects at a Bonferroni corrected significance threshold $\alpha = 0.05/4$ traits $= 0.00125$ for trait/sex combinations. Data point clouds show the average adjusted trait values for all 88 mitonuclear genotypes. (a) Nuclear effects across the DGRP375 (375) and OreR (Ore) genotypes. The large white circle is the estimated marginal mean of each nuclear background. (b) Treatment effects across the control (C) and rotenone (R) environments. The large white circle is the estimated marginal mean of each treatment. (c) Mitochondrial effects across the Beijing, Zimbabwe, *D. sim* and *D. yak* groups. The large white circle is the estimated marginal mean of each mtDNA. Note trait/sex combinations that are not shown were quantified but not statistically significant ($p > 0.00125$).

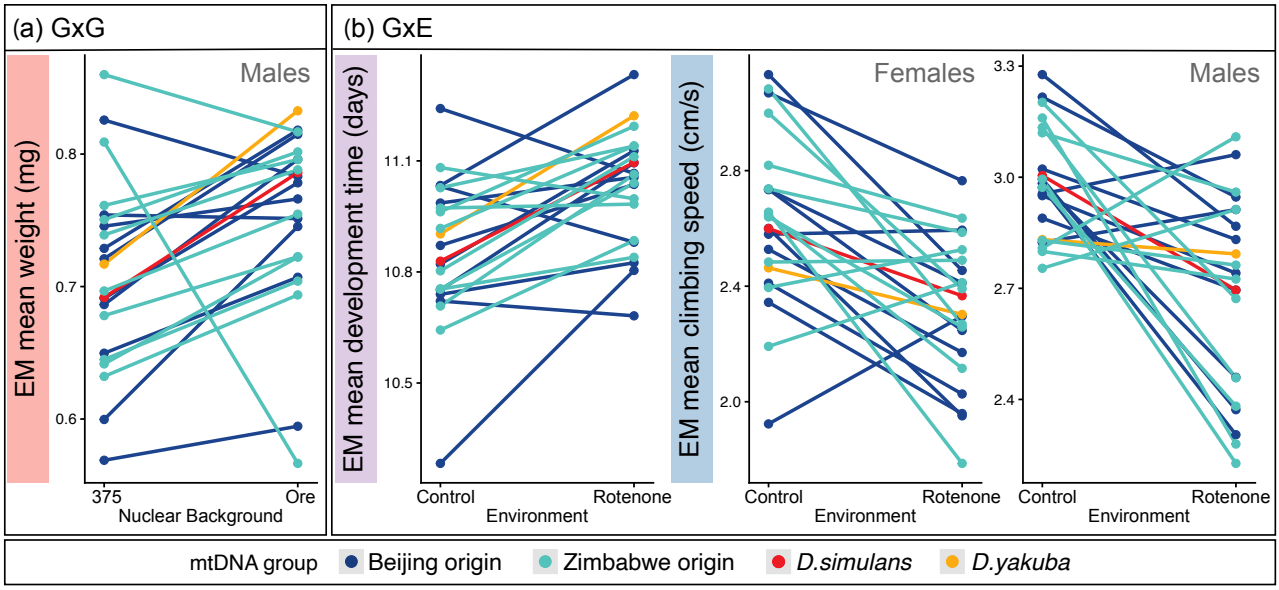

Figure 3: MtDNA modification of nuclear and environmental effects on complex traits. Significant second order interaction effects at a Bonferroni corrected significance threshold $\alpha = 0.05/4$ traits $= 0.00125$. Lines crossing indicate an interaction. (a) Mitonuclear GxG effects shown by the estimated marginal (EM) mean for the male weight trait across the two nuclear background for each mtDNA. (b) Mito-rotenone GxE effects for development time and female and male climbing. EM means for trait are given across the control and rotenone environments. Note trait/sex combinations that are not shown were quantified but not statistically significant (p>0.00125).

GxGxE effects are more pronounced within than between species

We identified a significant third order GxGxE interaction effect in the climbing phenotype in both females ($p < 0.001$) and males ($p < 0.001$). To further explore how mitochondrial genotypes differ phenotypically across nuclear genotype and treatment combinations, we used an additive main effects and multiplicative interaction (AMMI) model. This method applies principal component analysis (PCA) to decompose interaction effects after centering on additive effects, visualized with a biplot of genotypes and environments. We combined the nuclear background and treatment into a single environmental context for the mitochondrial term to capture third-order interaction effects (Mito:Nuc:Treatment), but in doing so the biplot also implicitly captures second-order interaction effects (Mito:Nuc and Mito:Treatment). Points (here, mtDNAs) near the interaction PCA (IPCA) axes origin are more stable, whereas those farther away show stronger interactions [32]. Acute angles between genotype points and environment vectors indicate positive interactions (e.g. ZW144 and O:C in Figure 4a inset), and obtuse angles (e.g. ZW144 and O:R in Figure 4a inset) indicate negative ones. Figure 4 shows the AMMI biplot of the first two IPCA axes for the climbing trait. IPC1 was significant in females (47.4%, $p < 0.0001$) and males (60.6%, $p < 0.0001$), IPC2 was also significant (36%, $p = 0.004$ for females; 34.9%, $p = 0.004$ for males), IPCA3 was not significant in either sex and is excluded.

Using the AMMI model, we observed that OreR nuclear background causes stronger climbing performance interactions among mitochondrial genotypes than the DGRP375 background. In both females (Figure 4a) and males (Figure 4b), the OreR nuclear background environment on both control (Ore:Control arrow) and rotenone (Ore:Rotenone arrow) food are much further from the origin than the DGRP375 nuclear backgrounds (375:Control and 375:Rotenone arrows) indicating stronger interaction effects.

There is no trend that suggests that certain mitochondrial groups (i.e. haplotypes from within Beijing or Zimbabwe) performed similarly better or worse on any given environment. However, there were some haplotypes from within the Beijing and Zimbabwe groups that stood out as having large interaction effects. In both males and females, ZW144 displayed a strong interaction effect indicated by its distance from the origin (Figure 4, insets) and its acute (<$90°$) angles with the Ore:Control arrows. We also observed that in both sexes, the "native" mitochondrial haplotypes 375 and Ore were close to the origin indicating these were stable across the treatment environments. These haplotypes were not reciprocally placed onto the two nuclear backgrounds so consequently cannot deviate along the horizontal plane.

Surprisingly, the interaction effects for both *D. sim* and *D. yak* were not stronger than those observed within either *D. mel* group despite having diverged as species ~2.5M and ~10M years ago, respectively. This is apparent from the placement of the *D. sim* and *D. yak* points inside the ellipse of points defined by either

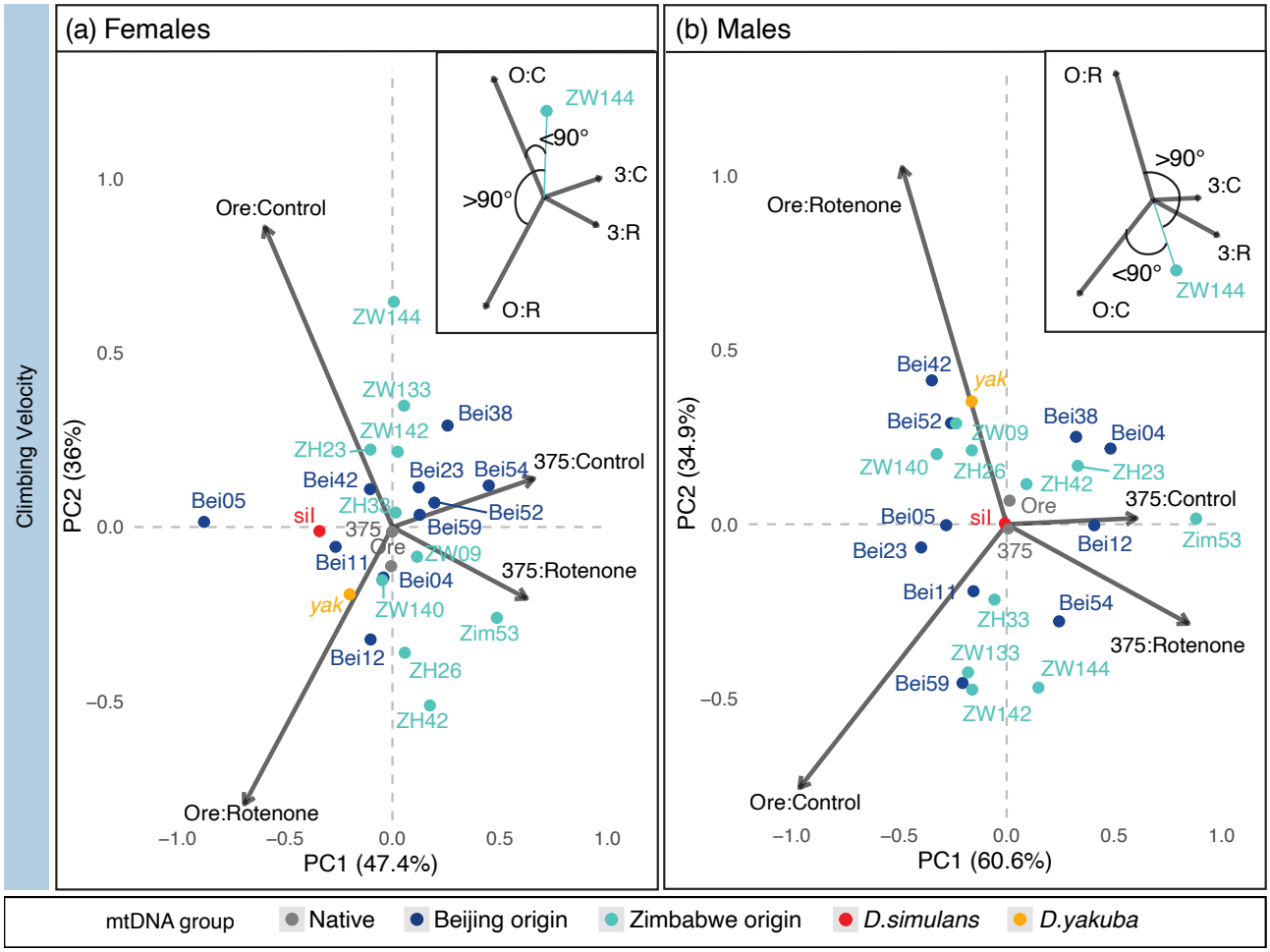

Figure 4: AMMI biplot of first two interaction PCs for the climbing trait. Nuclear:Treatment combinations are represented by arrows. Mitochondrial genotypes are represented by points and colored by general grouping (Beijing origin, Zimbabwe origin, *D. simulans*, *D. yakuba*, and the native mitochondria 375/Ore.) Mitochondrial genotypes closer to the origin show smaller interaction effects and are generally stable across different Nuclear:Treatment combinations. Genotypes further from the origin show greater interaction effects and align closer (acute angle, ZW144 and O:C in inset A) to environments where there is a positive interaction and are more distant (obtuse angle, ZW144 and O:R in inset B) where there is a negative interaction. Data from female individuals is shown in panel (a) and males in panel (b).

the Beijing or Zimbabwe mtDNAs from *D. mel*. Interestingly, the strength of interaction effects appears to be unrelated to amino acid (AA) divergence between mtDNA haplotypes: relative to the *D. mel* mtDNA reference genome, there are 142 changes between *D. mel* and *D. yak*, 92 between *D. mel* and the *D. sim* sil, 12 across all *D. mel* Beijing mtDNAs, and 11 across all *D. mel* Zimbabwe mtDNAs (Table S5). We then considered whether mitonuclear interactions could be attributable to AA changes at physical protein-protein contact sites between mtDNA and nucDNA residues, since OXPHOS Complex I (CI) inhibited by rotenone comprises both mitochondiral and nuclear encoded proteins. We used a Cryo-EM structure of CI [33] to map the missense changes to contact and non-contact positions and found that 62.5% of AA mutations among the 20 *D. mel* mtDNAs were in CI nuclear contact sites compared to 31.3% and 33.6% of the AA mutations fixed between *D. mel* and *D. sim* or *D. yak*, respectively (Figure 5, Extended Data Table 2). To assess the impact of amino acid changes, we classified them as conservative or radical, with radical changes involving changes between AA classes (nonpolar or uncharged polar) [34]. A higher proportion of radical substitutions occurred in *D. mel* mtDNAs (8/20, 40%) than in divergence to *D. sim* (24/92, 26%) or *D. yak* (43/142, 30%) (Table S5). While not significant by G-tests of heterogeneity, the trend is towards more deleterious conditions within than between species.

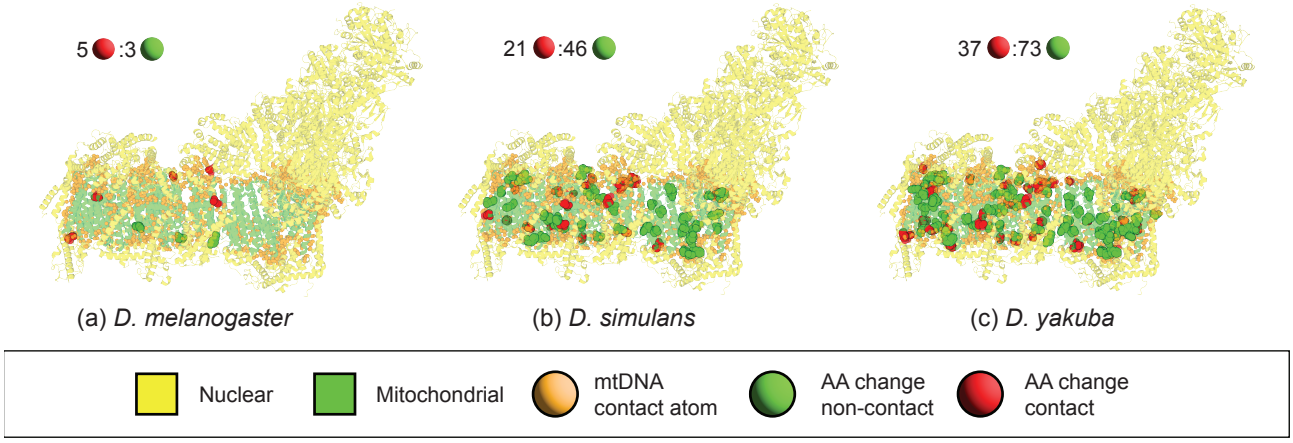

Figure 5: Complex I (CI) protein structure (from [33]) showing both mitochondrial and nuclear encoded proteins. Atoms from mtDNA encoded proteins <4Å from nuclear residues are shown as orange spheres (contact sites). The larger green and red spheres show non-contact and contact sites respectively for each CI missense mutation in the (a) *D. mel* haplotypes, (b) *D. sim* haplotype, and (c) *D. yak* haplotype. Ratios of contact to non-contact AA changes for each species are given. Note that the mutations shown are not strictly private to each group and many overlap.

Sex differences influence traits, but Mother's Curse is not evident

Using an ANOVA model that included sex as both a main and interaction effect (Equation 1), we found sex to be a significant main effect in all traits ($p < 0.0001$, Table S6), with the largest effect on weight ($R^2_{\beta*} = 0.7993$) reflecting the known sexual dimorphism of body size. We also found a number of significant genotype-by-sex, treatment-by-sex and third order interactions between mitochondrial genotype, sex, and nuclear genotype (Table S6).

We also examined how maternal inheritance has shaped mitochondrial haplotype variation. Maternal inheritance allows selection on female traits but can let male-harming mutations accumulate, a hypothesis known as Mother's Curse [35,36]. The weak form of this hypothesis predicts greater phenotypic variation among mtDNA genotypes in males than in females which we tested for by comparing coefficient of variation (CV) among mtDNAs in the two sexes. The strong form predicts variation in both sexes with negatively correlated male and female performance [37]. Across all 12 trait, environment, and nuclear genotype combinations there was no evidence of either weak or strong Mother's Curse, as no CV comparisons were significant (Extended Data Figure 1, Table S7), and none of the significantly correlated traits showed a negative correlation (Table S7).

Discussion

In this study, we leveraged a genetically diverse panel of mitonuclear genotypes of *Drosophila* to evaluate the impact of mitonuclear coevolution, epistasis, and GxE on four complex traits: climbing velocity, flight performance, body weight, and development time. Unlike many systems where a full-factorial test of epistasis is infeasible, our design allowed for a comprehensive partitioning of phenotypic variation that we used to explicitly test for mitonuclear epistasis between two nuclear genomes and 22 mtDNAs. We modified the environment with rotenone, a natural inhibitor of mitochondrial Complex I– the largest OXPHOS complex containing seven mitochondrial and about 36 nuclear-encoded genes [38]. This allowed us to challenge both the nuclear and mitochondrial genomes in our experiment and further partition phenotypic variation into environmental effects, GxE, and GxGxE interactions. The design was strengthened by duplicating the build of each mitonuclear combination and reproducibility was confirmed across experimental blocks. Significant main effects of the nuclear genome, mtDNA, and rotenone treatment were detected across all four traits, confirming that variation in both genomes and environmental stressors jointly shaped phenotypic outcomes.

We found significant second order mitochondria-by-environment (GxE) interaction effects that influenced climbing velocity and development time. Indeed, for development time, the effect size of the mitochondrial GxE interaction was even greater than the environmental effect alone. It has been predicted that when exposed to novel and/or stressful environments (such as rotenone inhibiting OXPHOS through Complex I), GxE effects may be amplified because genotypes will respond unpredictably relative to their reactions to familiar en-

vironments [39]. The mitochondrial by rotenone interaction observed here may indicate cryptic mitochondrial variation, particularly for development time and male climbing velocity where interaction effects occur without mitochondrial main effects.

Across four traits measured in both sexes, we only found a single instance of a significant mitochondria-by-nuclear DNA interaction (i.e., mitonuclear epistasis or GxG)– male body weight. Body weight is strongly linked to metabolic function and the modulation of this trait by mitonuclear effects could be due to differences in OXPHOS efficiency. Beyond male weight, we found evidence of mitonuclear epistatic effects in male climbing at a less conservative significance threshold ($p < 0.025$), consistent with the high metabolic demands of this behavior [40]. Previous studies have used covariation in evolutionary rates of mitochondrial and nuclear OXPHOS genes as strong evidence for mitonuclear coevolution in mammals [25], insects [41], and bivalves [24]. Therefore, disrupting this coevolved relationship may only be impactful for traits with high metabolic and OXPHOS demand, reflecting the idiosyncratic nature of mitonuclear interactions found in both our study and others [42, 43].

Consistent with previous reports of context-dependent mitonuclear epistasis [44–48], our results indicate that GxG interactions vary across environments and sexes. In particular, we detected significant third-order interaction effects with the environment (GxGxE) and with sex (GxGxSex) in the climbing trait. Classical theory predicts that haploid, uniparentally inherited mtDNA is unlikely to maintain polymorphism unless fitness effects are finely balanced [49, 50]. However, context-dependent selection arising from sex, environment, or nuclear background may promote the maintenance of mtDNA variation through asymmetric fitness effects [51–54]. Our finding of both GxGxE and GxGxSex interactions may provide a mechanism for prolonging the polymorphic sojourn time of non-neutral mtDNA variants. Shifting selective advantages across contexts could buffer these variants against fixation compared to mtDNA evolving in a homogeneous environment.

Remarkably, across all significant mtDNA main effects, second- and third-order interactions, we found that the two outgroup species, *D. sim* and *D. yak*, consistently showed intermediate values relative to the two *D. mel* groups despite being paired with mismatched *D. mel* nuclear genomes. This result is surprising as it contrasts sharply with predictions from the nuclear compensation model, which suggests that nuclear alleles evolve to counteract deleterious mitochondrial mutations [13]—a form of coevolution expected to break down in interspecific mismatches. Furthermore, analyses of 53 nuclear-encoded OXPHOS genes in *Drosophila* showed that most are under purifying selection, with only three Complex I genes exhibiting signs of positive selection (Table S8) [55, 56]. Our work bridges these molecular findings with phenotypic evidence, demonstrating that despite extensive amino acid divergence between species, mitonuclear communications have been maintained, limiting phenotypic differentiation across species boundaries. These results extend earlier work showing stronger within-species mitonuclear fitness effects [57], revealing that even across roughly 10 million years of divergence between *D. mel* and *D. yak*, nuclear compensation has not impacted these traits.

We propose that this observation of "excess polymorphism" relative to divergence is consistent with a model of stabilizing selection acting on mitonuclear communication, a pattern that is evident in other complex traits [9]. We found that the more divergent mtDNAs exhibited intermediate trait values as well as weaker interaction effects across environments for both GxE and GxGxE effects relative to the interspecific genotypes. These results are consistent with stabilizing selection for an optimal phenotype constraining divergence and promoting greater phenotypic stability across environmental gradients. To achieve this stability, purifying selection has likely prevented the fixation of harmful mutations between *D. mel*, *D. sim*, and *D. yak*. When we compared physical locations of amino acid changes in Complex I across *D. mel*, *D. sim* and *D. yak*, we found that within-species variants were enriched at protein-protein contact sites with the nuclear genome (Figure 5). Within-species variants also showed a higher proportion of radical AA changes (Table S5), suggesting that most mtDNA variation within populations is mildly deleterious and transient and large-effect AA changes are purged before they can be fixed. This contrasts theory that suggests that purifying selection is weaker in the mitochondrial genome due to its reduced effective population size [58]. Our findings, together with evidence from *D. melanogaster* showing comparable mitochondrial and nuclear purifying selection [59] and consistently low mtDNA dN/dS values across the Genus *Drosophila* [60], support efficient purifying selection acting on mtDNA in this system.

Finally, our study addresses the Mother's Curse hypothesis, which posits that maternal inheritance of mtDNA allows male-harming mutations to accumulate because selection acts only through female fitness [35, 36]. Cosmides & Tooby [61] presented early evidence for this selective asymmetry, showing that some mitochondrial diseases were more pronounced in males. By placing different mitochondrial haplotypes on a common nuclear background and measuring complex traits in both sexes, our design provides an effective empirical test of this hypothesis. Across the three traits we measured, we found no evidence for Mother's Curse in either its strong or weak form, despite a study design that specifically addressed previous concerns about mtDNA sampling and reproducibility [37]. The absence of Mother's Curse in our traits, combined with evidence of strong purifying or stabilizing selection on mtDNA, suggests that selection acting through females may indirectly constrain the accumulation of male-harming mutations.

Collectively, our results demonstrate that mtDNA and its interactions with nuclear genetic variation and external environments are significant factors in the distribution of phenotypic variation in natural populations. Moreover, the very limited influence of divergent mtDNAs on these fitness related traits in *Drosophila* support a model of stabilizing selection acting on mitonuclear interactions that offer a distinct contrast to models that implicate mitonuclear interactions in adaptive evolution and speciation. The lack of major mitonuclear incompatibilities in *Drosophila* may reflect its relatively large effective populations size (N_e) [62], which prevents the fixation of deleterious mtDNA mutations through efficient purifying selection. In contrast, smaller N_e systems such as *Tigriopus* copepods [63] or *Xiphophorus* swordtails [22] may allow for more harmful variants to accumulate, increasing the risk of incompatibilities. This "variable purging hypothesis" of mitonuclear incompatibilities is relevant to both evolutionary and medical applications. Our findings that mtDNA polymorphisms can be more impactful that divergent mtDNAs has implications for the choice of donor mtDNAs used in mitochondrial replacement therapy. If segregating mtDNA variants within various human population are more likely to harbor deleterious mutations than mtDNAs from more divergent populations, the latter may be more likely to have purged disease-causing mutations. Alternatively, widespread mtDNA haplogroups may have larger historical effective population sizes and may be less likely to carry such mutations. Future analyses should evaluate the important role of effective population size on the tempo and mode of mitonuclear coevolution.

Methods

Mitonuclear Lines

Polymorphism in mtDNA was sampled from 20 female lines of *D. melanogaster* (*D. mel)*, 10 from Beijing, China and 10 from Zimbabwe. Nineteen of these lines were originally characterized as the "Global Diversity Lines" (GDLs) [64], while the Zim53 line from Zimbabwe was originally characterized in Ballard 2000 [65]. Also included in the mtDNA panel were two genotypes from other species: *D. simulans* (*D. sim*) Hawaii (mitochondrial haplotype *sil* [66], mitonuclear construction described in [57]) and *D. yakuba* (*D. yak*) (mitonuclear construction described in [67]). Together, we used a total of 22 mitochondrial haplotypes incorporating intraspecific polymorphism and deep interspecific mitochondrial divergence from reproductively isolated outgroups.

To construct the mitonuclear lines, we first removed *Wolbachia* from each female line by treating them with tetracycline for two generations. PCR assays post treatment and again 2 years later confirmed the absence of *Wolbachia*. Following this, we created mitonuclear genotypes using 10 generations of recurrent male backcrossing from each of the nuclear lines to virgin females from the mitochondrial lines. We will describe the mitonuclear genotype of these lines using the notation *mtDNA;nucDNA*. We used two common nuclear backgrounds: *D. mel* Oregon R (*OreR*) and *D. mel* DGRP375 (*375*). For each mitonuclear combination, the construction was replicated by maintaining two independent lines ("builds"). Virgin females from each build were collected and independently crossed to males from either *OreR* or *375* nuclear backgrounds, providing an additional source of biological replication and controlling for any random effects of the backcrossing process. Also included in the panel were two additional "unaltered" lines for each nuclear background retaining their own native mitochondria (*OreR;OreR* and *375;375*). In total, 22 mitochondrial haplotypes were backcrossed in two independent builds to two nuclear backgrounds, plus the two genetically unaltered controls, resulting in $22 \times 2 \times 2 + 2 = 90$ lines that were assayed. A complete list of all genotypes is provided in Table S9.

Fly rearing

Backcross lines were maintained on standard *Drosophila* rearing food. The food contains 1.8 g agar, 5 g SAF yeast, 10.4 g yellow cornmeal, 22 g sucrose, 0.9 tegosept in 4.5 ml 95% ethanol, cooked in 200 mL of deionized water and larger batches were scaled up with the same ratios. The rotenone diet was the standard control food containing 25 μM of rotenone, a concentration selected to yield enough surviving adults with similar ages to controls. Prior to the phenotype assays, each line was expanded multiple culture bottles before sorting 20 mated females from each bottle onto either control or rotenone food. The 20 females were allowed to lay eggs on either the control or rotenone food for 48 hours before being removed. The offspring of these individuals were then sorted by sex in 6 replicate vials of each sex and scored for each trait while being maintained on their respective diets.

Phenotypic Assays

We measured body weight, climbing velocity, and flight performance in single-sex groups of adults. Prior to sorting adults by sex, we measured development time of mixed-sex groups. Development time was scored for three replicate vials following the 48-hour egg lay. Each day after the parental females were removed, we counted the number of pupae that have entered pupal stage P12 or above (identified by visible darkened wings [68]) for a total of 11 days post egg lay. We consider the average time since egg lay to be overall development time score for each vial. Following the development time assay, adult animals from the replicate vials were sorted by sex and flipped to new food of the same diet they were grown on. A total of six replicate vials of 20 males and six vials of 20 females were used for the remainder of the phenotype assays.

The climbing velocity or "negative geotaxis" assay was scored for all six vials in three technical replicate drops of the climbing rig to initiate climbing. The automated software FreeClimber was used to quantify the climbing performance [31]. FreeClimber quantifies the average velocity of each vial through a local linear regression on the average y-position vs. time and returns the steepest (i.e., fastest), linear two-second section of the curve as the final velocity. We consider the average of the three technical replicates to be the final score for each vial providing 3×6=18 estimates of climbing speed for each genotype, sex, and treatment combination.

To measure flight performance, we used three of the six sorted vials for each sex since this assay is sacrificial and the remaining vials needed to be saved for the body weight assay. The flight assay measures the response to an abrupt drop into a flight column where higher landing height is indicative of a stronger flight genotype. The details of the assay have been previously described in [30]. Due to the nature of the assay, all vials were scored simultaneously and we do not have access to individual trait scores for each vial. The trait value for each build of a genotype was then the average landing height for all flies across the three replicate vials.

Animal body weight was scored by combining the remaining surviving animals from three replicate vials of males and females and dividing the pool evenly between three 1.5 mL microcentrifuge tubes. Each microcentrifuge tube was pre-weighed with a 0.001 mg (1 μg) and then reweighed with a sample of approximately 20 animals. The exact number of animals per tube was recorded since it varied among samples. To get a final body weight, the initial weight of the tube was subtracted from the final weight of the tube with the flies and divided by the number of flies in each tube to estimate an average weight per fly. Since the animals from the remaining vials were pooled to maintain a more uniform distribution of animals per tube, we considered the average of the score for the replicate tubes to be the genotype score.

The concept of "build" introduced by the two independent backcrosses for each line provided an additional level of biological replication for each phenotype. That is, where we combine biological replicates to achieve an aggregate phenotype score in the flight and weight assays, there is still replication at the level of the build as the same genotype is tested multiple times.

Experimental Blocks and Reproducibility of Phenotype

Due to the size of the mitonuclear panel and the number of the traits measured, the lines were tested across seven initial experimental blocks. The first three blocks were larger and contained 12 mitonuclear genotypes with both builds for a total of 24 lines tested. The remainder of the experimental blocks were scaled down to four mitonuclear genotypes per block with both builds for a total of eight lines per block. Lines that were unsuccessful (i.e. low yield) in their original block were retested in a different block and removed from the original block. Six "reference" genotypes were tested in every block and were used to adjust each trait measurement for block effects. This included one of the builds of the lines *sil;OreR*, *sil;375*, *yak;OreR*, *yak;375* and the two unaltered lines *OreR;OreR* and *375;375*. For completeness, both builds of *sil;OreR*, *sil;375*, *yak;OreR*, *yak;375* were tested in the larger blocks 1-3.

To test for the reproducibility of the development time, the climbing, and the body weight phenotypes across experimental block environments, a subset of the lines tested in blocks 1 and 2 were retested in a final block (block 9). A total of 16 mitonuclear genotypes with both builds were retested in this final block for a total of 32 retested lines. We then compared the measures obtained from the second observation to those from the original testing block environment. The block assignments for each line are given in Table S9.

ANOVA Model of Phenotypic Response

To test for the effects of rotenone treatment and genotype-phenotype associations for both the mitochondrial and nuclear genotypes, we used an ANOVA mixed model for each phenotype independently. To correct for

variation between blocks, we subtracted the mean of the six "reference" genotypes (*OreR;OreR*, *375;375*), and four "build A" lines of *yak;OreR*, *yak;375*, *sil;OreR*, *sil;375*) by block (Bl), treatment (T), and sex (S) from each raw data point and added the treatment and sex mean across all blocks [$(Y_{i(Bl;S;T)} - \bar{Y}_{(Bl;S;T)}) + \bar{Y}_{(S;T)}$]. Including the interspecific (*D. yak* and *D. sim*) lines in the block correction helped prevent bias toward *D. mel*-specific responses, ensuring that block effects were removed based on the overall experimental panel rather solely on *D. mel* performance.

The full model is given by:

$$Y_{ijklmn} = \mu + M_i + N_j + T_k + S_l + (MN)_{ij} + (MT)_{ik} + (NT)_{jk} + (TS)_{kl} + (MS)_{il} + (NS)_{jl} + (MNT)_{ijk} + (MNS)_{ijl} + B_{m(ij)} + V_{n(ijkm)} + \varepsilon_{n(ijklm)} \quad (1)$$

The model included the mitochondrial genotype (M), the nuclear genotype (N), treatment (*T*), and sex (*S*) as fixed effects as well as their interactions. Build (B), nested in the combination of the mitochondrial and nuclear genomes is included as a random effect. Build is a form a biological replication for each combination of mitochondrial and nuclear genotype. Within each build, there is a second level of replication that we model with random effects. For the climbing phenotypes there is an additional level of biological replication that comes from a shared environment of a vial (V) that is a random effect nested within build and treatment. Due to the way the treatment was applied (through diet), two treatments cannot be given to the same vial. Multiple measurements of each vial's climbing velocity allows us to estimate the variance of this random effect. For the remainder of the phenotypes, we were only able to estimate the variance at the level of build (B). There is additionally an error term given by ε.

We also ran reduced models for each sex separately:

$$Y_{ijkmn} = \mu + M_i + N_j + T_k + (MN)_{ij} + (MT)_{ik} + (NT)_{jk} + (MNT)_{ijk} + B_{m(ij)} + V_{n(ijkm)} + \varepsilon_{n(ijkm)} \quad (2)$$

The development time phenotype does not have a sex term so only the reduced model was used. Additionally, since development time is affected by larval density, which was controlled for at the level of parental density rather than total egg count, there is an additional larval density term in the model which is approximated by the total pupal count in each vial divided by the total across all sets by treatment [$L = \text{total}_i/\text{total}_{(T)}$]. For the weight and flight phenotypes, averages at the level of build were used. We performed Type III ANOVA with Satterthwaite's approximation for degrees of freedom. The type III ANOVA uses a partial sum of squares (SS) approach which adjusts the main effects to exclude the overlap with interaction terms. This means that the SS do not add to the total SS and can not be interpreted to partition the total variance of the model. To measure the relative importance of each fixed effect (effect size), we instead used a semi-partial R^2 namely, $R^2_{\beta*}$, described in Jaeger et al., 2016 [69]. ANOVA models were implemented using the R packages lme4 [70] and lmerTest [71]; and $R^2_{\beta*}$ was calculated using the package r2glmm [69]. In Figure 2 we use the package emmeans [72] to estimate the marginal means (or model-adjusted mean) for each factor of interest. We do this to obtain a more accurate portrayal of the effect being captured by the model.

Note: When fitting the linear model for male weight, the random effect of build was estimated as having a near-zero variance, indicating that our data did not support among-build variation beyond residual error. This likely reflected limited replication and/or noisy measurements (males are ~30% smaller than females).

AMMI Model of Phenotypic Response

Where we identified significant third-order interactions using the ANOVA model, we used an Additive Main effects and Multiplicative Interaction model (AMMI) to further explore patterns in the data. AMMI is a frequently used technique for understanding complex genotype by environment interactions in agricultural yield trials and is particularly useful as a visualization technique [32, 73]. This method first applies analysis of variance to partition the variation in the phenotype of interest in to genotype effects (G) , environmental effects (E), and genotype-environment (GE) interaction effects, and then it uses principal components analysis to decompose the GE effect. Since AMMI models explicitly decompose the interaction effect by first subtracting G and E additive effects, we can use the resultant decomposition to visualize the interactions present in the data. The AMMIN model equation is:

$$Y_{ijm} = \mu + M_i + E_j + \sum_{n=1}^{N} \lambda_n \gamma_{in} \delta_{jn} + \rho_{ij} + \varepsilon_{ijm}, \quad (3)$$

where each response Y is indexed by the mitochondrial genotype i, the environment j, and the build m. For these analyses, we used the adjusted Y values. Like in the ANOVA model, the mitochondrial main effect is

given by M. Since the AMMI model is built to decompose GE interactions, we formed a single composite variable E for the "environment" that has factor levels for every combination of nuclear background and treatment (Ore:Control, Ore:Rotenone, 375:Control, and 375:Rotenone). This simplification is appropriate for visualizing patterns and aligns with prior AMMI applications incorporating additional environmental factors [32, 74]. The multiplicative part of the AMMI model uses PCA to decompose the remaining residual matrix after removing the M and E main effects into PCA axes 1 to N. The multiplicative parameters are: λ_n is the singular value for PCA axis n, γ_{in} is the mitochondrial genotype eigenvector for axis n, and δ_{jn} is the environment eigenvector for axis n. The AMMI multiplicative parameters $\lambda^{0.5}\gamma_i$ and $\lambda^{0.5}\delta_j$ (for genotype and environment, respectively) are often referred to as "interaction PCA scores" (IPCA) to specify the distinction between typical PCA [32]. There are at most min$(M-1, E-1)$ IPCA axes, but we are typically only interested in the first couple of axes ($n < N$). If not all axes are used, the remaining residual is included in ρ_{me}. Additionally, we considered error at the replication level of build with the term ε_{ijm}. Models were implemented using the R package agricolae [75].

Mother's Curse tests

We consider two forms of this hypothesis. The weak form predicts that (1) males show greater variation in traits among the mitochondrial haplotypes than females and (2) there is no clear evidence of sexual antagonism (no linear relationship between male and female performance). The strong form predicts that (1) males and females both show variation across mitochondrial haplotypes and (2) the mutations between the haplotypes are sexually antagonistic in nature (negative linear relationship between female and male performance) [37].

To test predictions related to variation in trait performance, we average across technical replicates and biological replicates (build, vial) and calculate the variation due to mitochondrial genetic variation alone. We then compare the sex difference in this variation on the two nuclear backgrounds, OreR and DGRP375 and independently on both environments, control and rotenone. We sampled 1000 bootstraps (only 500 are shown in Extended Data Figure 1) of the coefficient of variation (CV) for each sex and calculated the number of times the CV for males was less than or equal to the CV for females over the total number of bootstraps to obtain a p-value. Note that the Bei54 (375/Ore on rotenone) and ZW144 (Ore on rotenone) haplotypes were excluded from the analysis of the weight trait due to low sample sizes in males that artificially increased variance in males. To test predictions related to sexual antagonism between males and females, we calculated the correlation coefficient, r.

mtDNA variants and quantification of mitochondrial phylogenetic signal

To call mtDNA variants, we retrieved DNA short reads from NCBI (PRJNA268111, [64] for GDL mtDNAs) and from lines sequenced by our research group (Rand and Raynes unpublished for OreR and DGRP375). Reads were aligned to the FlyBase *D.melanogaster* 6.32 reference using bwa v. 0.7.17 [76]. Variants were called using a standard GATK pipeline (excluding the AT-rich region of the mitochondrial genome), with base recalibration performed using known variants from the DGRP freeze 2 data, and final variant filtering applied according to GATK's recommended hard-filtering parameters [77]. Variants were annotated with SNPeff v. 4.3t [78] to identify amino acid changes. The remaining mitochondrial genomes were downloaded from NCBI: Zim53 (KP843854.1, [79]), *D. yak* (NC_001322.1), and *D. sim* sil (AF200836.1). Changes to the amino acid sequence were obtained by aligning protein sequences with mafft v. 7.505 auto alignment and reporting changes in the sequence with custom python script. We retrieved the *D. mel* complex I protein structure from PDB (8B9Z, [33]) and used PyMOL v. 3.1.6.1 to identify protein-protein contact sites. Contact sites are defined as atoms from mtDNA-encoded proteins that are within 4Å of atoms from nucDNA-encoded proteins.

To evaluate the relationship between genetic relatedness and the trait measurements, we quantified phylogenetic signal. The major source of phylogenetic variation in our mitonuclear system was from the mitochondrial haplotypes, so we specifically quantified mitochondrial phylogenetic signal across an inferred mitochondrial phylogenetic tree. The filtered variants were used to create mitochondrial FASTA files for each sample using GATK's 'FastaAlternateReferenceMaker'. Multiple sequence alignment was performed using the mafft v. 7.505 'mafft-linsi' algorithm [80] and mitochondrial tree was inferred in RAxML-NG v. 1.2.2 using a majority-rule consensus tree derived from 1000 bootstrap replicates. We used the inferred phylogenetic tree to calculate Blomberg's K, which estimates the strength of phylogenetic signal under a Brownian motion model of evolution [81]. To test if K is significantly different from zero, we sampled 1000 randomizations of the trait data across the phylogeny and calculated the number of times the randomized K is larger than the observed K. Dividing the proportion by 1000 yielded a p-value, which we then calculated for every combination of sex, treatment, nuclear background, and phenotype using the adjusted phenotype values. The R package phytools [82] was used to calculate both K and p-values.

Data Availability

Code for data analyses and raw data files are available at the GitHub site (https://github.com/leahdarwin/MitoGDL). The authors affirm that all analyzed data necessary for confirming the conclusions of the article are present within the article, figures, tables, supplementary information and GitHub site. Drosophila strains are available upon request from David_Rand@brown.edu

Funding

The work was supported by grants from NIGMS: MIRA R35GM139607 to DMR, LJD was supported by NIH T32 GM128596 and an NSF Graduate Fellowship (2023-2025), CPB and OCM were supported by NSF REU Division of Biological Infrastructure award 2150328, JHB was supported by Brown University Undergraduate Teaching and Research Award. DMR acknowledges support of COBRE award P20GM109035.

References

[1] Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* **101**, 5–22 (2017). URL https://www.cell.com/ajhg/abstract/S0002-9297(17)30240-9. Publisher: Elsevier.

[2] Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS genetics* **4**, e1000008 (2008).

[3] Mackay, T. F. C. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nature Reviews Genetics* **15**, 22–33 (2014). URL https://www.nature.com/articles/nrg3627. Publisher: Nature Publishing Group.

[4] Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009). URL https://www.nature.com/articles/nature08494. Publisher: Nature Publishing Group.

[5] Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* **109**, 1193–1198 (2012). URL https://www.pnas.org/doi/abs/10.1073/pnas.1119675109. Publisher: Proceedings of the National Academy of Sciences.

[6] Gimelfarb, A. Genotypic Variation for a Quantitative Character Maintained under Stabilizing Selection without Mutations: Epistasis. *Genetics* **123**, 217–227 (1989). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1203785/.

[7] Barton, N. H. & Turelli, M. Effects of genetic drift on variance components under a general model of epistasis. *Evolution; International Journal of Organic Evolution* **58**, 2111–2132 (2004).

[8] Mackay, T. F. C. The Genetic Architecture of Quantitative Traits. *Annual Review of Genetics* **35**, 303–339 (2001). URL https://www.annualreviews.org/content/journals/10.1146/annurev.genet.35.102401.090633. Publisher: Annual Reviews.

[9] Mackay, T. F. C. Mutations and quantitative genetic variation: lessons from Drosophila. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **365**, 1229–1239 (2010).

[10] Sloan, D. B. *et al.* Cytonuclear integration and co-evolution. *Nature Reviews Genetics* **19**, 635–648 (2018). URL https://www.nature.com/articles/s41576-018-0035-9. Publisher: Nature Publishing Group.

[11] Wolff, J. N., Ladoukakis, E. D., Enríquez, J. A. & Dowling, D. K. Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130443 (2014). URL https://royalsocietypublishing.org/doi/full/10.1098/rstb.2013.0443. Publisher: Royal Society.

[12] Rand, D. M. & Mossman, J. A. Mitonuclear conflict and cooperation govern the integration of genotypes, phenotypes and environments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **375**, 20190188 (2020).

[13] Rand, D. M., Haney, R. A. & Fry, A. J. Cytonuclear coevolution: the genomics of cooperation. *Trends in Ecology & Evolution* **19**, 645–653 (2004). URL https://linkinghub.elsevier.com/retrieve/pii/S0169534704003052.

[14] Hill, G. E. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecology and Evolution* **6**, 5831–5842 (2016). URL https://onlinelibrary.wiley.com/doi/10.1002/ece3.2338.

[15] Hill, G. E. Mitonuclear Compensatory Coevolution. *Trends in Genetics* **36**, 403–414 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S0168952520300548.

[16] Lima, T. G., Burton, R. S. & Willett, C. S. Genomic scans reveal multiple mito-nuclear incompatibilities in population crosses of the copepod Tigriopus californicus. *Evolution* **73**, 609–620 (2019). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.13690. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/evo.13690.

[17] Pereira, R. J., Lima, T. G., Pierce-Ward, N. T., Chao, L. & Burton, R. S. Recovery from hybrid breakdown reveals a complex genetic architecture of mitonuclear incompatibilities. *Molecular Ecology* **30**, 6403–6416 (2021). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15985. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mec.15985.

[18] Nguyen, T. H. M. *et al.* Mapping mitonuclear epistasis using a novel recombinant yeast population. *PLOS Genetics* **19**, e1010401 (2023). URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010401. Publisher: Public Library of Science.

[19] Meiklejohn, C. D. *et al.* An Incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in Drosophila. *PLoS genetics* **9**, e1003238 (2013).

[20] Bettinazzi, S. *et al.* Assessing the role of mitonuclear interactions on mitochondrial function and organismal fitness in natural Drosophila populations. *Evolution Letters* **8**, 916–926 (2024). URL https://doi.org/10.1093/evlett/qrae043.

[21] Chang, C.-C., Rodriguez, J. & Ross, J. Mitochondrial–Nuclear Epistasis Impacts Fitness and Mitochondrial Physiology of Interpopulation Caenorhabditis briggsae Hybrids. *G3 Genes|Genomes|Genetics* **6**, 209–219 (2016). URL https://doi.org/10.1534/g3.115.022970.

[22] Moran, B. M. *et al.* A lethal mitonuclear incompatibility in complex I of natural hybrids. *Nature* **626**, 119–127 (2024). URL https://www.nature.com/articles/s41586-023-06895-8. Publisher: Nature Publishing Group.

[23] Robles, N. V. *et al.* Admixture mapping reveals evidence for multiple mitonuclear incompatibilities in swordtail fish hybrids (2025). URL https://www.biorxiv.org/content/10.1101/2025.01.30.635158v2. Pages: 2025.01.30.635158 Section: New Results.

[24] Piccinini, G. *et al.* Mitonuclear Coevolution, but not Nuclear Compensation, Drives Evolution of OXPHOS Complexes in Bivalves. *Molecular Biology and Evolution* **38**, 2597–2614 (2021). URL https://academic.oup.com/mbe/article/38/6/2597/6146411.

[25] Weaver, R. J., Rabinowitz, S., Thueson, K. & Havird, J. C. Genomic Signatures of Mitonuclear Coevolution in Mammals. *Molecular Biology and Evolution* **39**, msac233 (2022). URL https://doi.org/10.1093/molbev/msac233.

[26] Zwonitzer, K. D. *et al.* Genome copy number predicts extreme evolutionary rate variation in plant mitochondrial DNA. *Proceedings of the National Academy of Sciences* **121**, e2317240121 (2024). URL https://www.pnas.org/doi/abs/10.1073/pnas.2317240121. Company: National Academy of Sciences Distributor: National Academy of Sciences Institution: National Academy of Sciences Label: National Academy of Sciences Publisher: Proceedings of the National Academy of Sciences.

[27] Sloan, D. B., Havird, J. C. & Sharbrough, J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology* **26**, 2212–2236 (2017). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13959. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mec.13959.

[28] Eyre-Walker, A. Mitochondrial Replacement Therapy: Are Mito-nuclear Interactions Likely To Be a Problem? *Genetics* **205**, 1365–1372 (2017). URL https://academic.oup.com/genetics/article/205/4/1365/6066392.

[29] McFarland, R. *et al.* Mitochondrial Donation in a Reproductive Care Pathway for mtDNA Disease. *New England Journal of Medicine* **393**, 461–468 (2025). URL https://www.nejm.org/doi/full/10.1056/NEJMoa2503658. Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa2503658.

[30] Spierer, A. N. *et al.* Natural variation in the regulation of neurodevelopmental genes modifies flight performance in Drosophila. *PLOS Genetics* **17**, e1008887 (2021). URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008887. Publisher: Public Library of Science.

[31] Spierer, A. N., Yoon, D., Zhu, C.-T. & Rand, D. M. FreeClimber: automated quantification of climbing performance in Drosophila. *Journal of Experimental Biology* **224**, jeb229377 (2021). URL https://doi.org/10.1242/jeb.229377.

[32] Gauch (Jr.), H. G. *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs* (Elsevier Science, 1992). Google-Books-ID: LaW1ksodou0C.

[33] Agip, A.-N. A., Chung, I., Sanchez-Martinez, A., Whitworth, A. J. & Hirst, J. Cryo-EM structures of mitochondrial respiratory complex I from Drosophila melanogaster. *eLife* **12**, e84424 (2023). URL https://doi.org/10.7554/eLife.84424. Publisher: eLife Sciences Publications, Ltd.

[34] Dagan, T., Talmor, Y. & Graur, D. Ratios of Radical to Conservative Amino Acid Replacement are Affected by Mutational and Compositional Factors and May Not Be Indicative of Positive Darwinian Selection. *Molecular Biology and Evolution* **19**, 1022–1025 (2002). URL https://doi.org/10.1093/oxfordjournals.molbev.a004161.

[35] Frank, S. A. & Hurst, L. D. Mitochondria and male disease. *Nature* **383**, 224 (1996).

[36] Gemmell, N. J., Metcalf, V. J. & Allendorf, F. W. Mother's curse: the effect of mtDNA on individual fitness and population viability. *Trends in Ecology & Evolution* **19**, 238–244 (2004).

[37] Dowling, D. K. & Adrian, R. E. Challenges and Prospects for Testing the Mother's Curse Hypothesis. *Integrative and Comparative Biology* **59**, 875–889 (2019). URL https://doi.org/10.1093/icb/icz110.

[38] Sharma, L. K., Lu, J. & Bai, Y. Mitochondrial Respiratory Complex I: Structure, Function and Implication in Human Diseases. *Current medicinal chemistry* **16**, 1266–1277 (2009). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4706149/.

[39] Saltz, J. B. *et al.* Why does the magnitude of genotype-by-environment interaction vary? *Ecology and Evolution* **8**, 6342–6353 (2018). URL https://onlinelibrary.wiley.com/doi/10.1002/ece3.4128. Publisher: John Wiley & Sons, Ltd.

[40] Sujkowski, A. *et al.* Mito-nuclear interactions modify *Drosophila* exercise performance. *Mitochondrion* **47**, 188–205 (2019). URL https://www.sciencedirect.com/science/article/pii/S1567724918300618.

[41] Yan, Z., Ye, G. & Werren, J. H. Evolutionary Rate Correlation between Mitochondrial-Encoded and Mitochondria-Associated Nuclear-Encoded Proteins in Insects. *Molecular Biology and Evolution* **36**, 1022–1036 (2019). URL https://doi.org/10.1093/molbev/msz036.

[42] Edmands, S. Mother's Curse effects on lifespan and aging. *Frontiers in Aging* **5** (2024). URL https://www.frontiersin.org/journals/aging/articles/10.3389/fragi.2024.1361396/full. Publisher: Frontiers.

[43] Garlovsky, M. D. *et al.* Testing for age- and sex-specific mitonuclear epistasis in Drosophila. *Evolution* **79**, 1568–1582 (2025). URL https://doi.org/10.1093/evolut/qpaf096.

[44] Dowling, D. K. Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1840**, 1393–1403 (2014). URL https://www.sciencedirect.com/science/article/pii/S0304416513004960.

[45] Arnqvist, G. *et al.* Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect. *Evolution; International Journal of Organic Evolution* **64**, 3354–3363 (2010).

[46] Hoekstra, L. A., Siddiq, M. A. & Montooth, K. L. Pleiotropic Effects of a Mitochondrial–Nuclear Incompatibility Depend upon the Accelerating Effect of Temperature in Drosophila. *Genetics* **195**, 1129–1139 (2013). URL https://doi.org/10.1534/genetics.113.154914.

[47] Zhu, C.-T., Ingelmo, P. & Rand, D. M. G×G×E for Lifespan in Drosophila: Mitochondrial, Nuclear, and Dietary Interactions that Modify Longevity. *PLoS Genetics* **10**, e1004354 (2014). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4022469/.

[48] Mossman, J. A., Biancani, L. M., Zhu, C.-T. & Rand, D. M. Mitonuclear Epistasis for Development Time and Its Modification by Diet in Drosophila. *Genetics* **203**, 463–484 (2016). URL https://doi.org/10.1534/genetics.116.187286.

[49] Hedrick, P. W. GENETIC POLYMORPHISM IN HETEROGENEOUS ENVIRONMENTS: A DECADE LATER. *Annual Review of Ecology, Evolution, and Systematics* **17**, 535–566 (1986). URL https://www.annualreviews.org/content/journals/10.1146/annurev.es.17.110186.002535. Publisher: Annual Reviews.

[50] Clark, A. G. NATURAL SELECTION WITH NUCLEAR AND CYTOPLASMIC TRANSMISSION. I. A DETERMINISTIC MODEL. *Genetics* **107**, 679–701 (1984). URL https://doi.org/10.1093/genetics/107.4.679.

[51] Babcock, C. S. & Asmussen, M. A. Effects of Differential Selection in the Sexes on Cytonuclear Polymorphism and Disequilibria. *Genetics* **144**, 839–853 (1996). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1207574/.

[52] Babcock, C. S. & Asmussen, M. A. Effects of differential selection in the sexes on cytonuclear dynamics. Life stages with sex differences. *Genetics* **149**, 2063–2077 (1998). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1460269/.

[53] Rand, D. M., Clark, A. G. & Kann, L. M. Sexually Antagonistic Cytonuclear Fitness Interactions in Drosophila melanogaster. *Genetics* **159**, 173–187 (2001). URL https://doi.org/10.1093/genetics/159.1.173.

[54] Turelli, M. & Barton, N. H. Polygenic Variation Maintained by Balancing Selection: Pleiotropy, Sex-Dependent Allelic Effects and G × E Interactions. *Genetics* **166**, 1053–1079 (2004). URL https://doi.org/10.1093/genetics/166.2.1053.

[55] Clark, A. G. *et al.* Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**, 203–218 (2007). URL https://www.nature.com/articles/nature06341. Publisher: Nature Publishing Group.

[56] Larracuente, A. M. *et al.* Evolution of protein-coding genes in Drosophila. *Trends in genetics: TIG* **24**, 114–123 (2008).

[57] Montooth, K. L., Meiklejohn, C. D., Abt, D. N. & Rand, D. M. Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of Drosophila. *Evolution; International Journal of Organic Evolution* **64**, 3364–3379 (2010).

[58] Lynch, M. & Blanchard, J. L. Deleterious mutation accumulation in organelle genomes. *Genetica* **102-103**, 29–39 (1998).

[59] Cooper, B. S., Burrus, C. R., Ji, C., Hahn, M. W. & Montooth, K. L. Similar Efficacies of Selection Shape Mitochondrial and Nuclear Genes in Both Drosophila melanogaster and Homo sapiens. *G3: Genes|Genomes|Genetics* **5**, 2165–2176 (2015). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4592998/.

[60] Montooth, K. L., Abt, D. N., Hofmann, J. W. & Rand, D. M. Comparative genomics of Drosophila mtDNA: Novel features of conservation and change across functional domains and lineages. *Journal of Molecular Evolution* **69**, 94–114 (2009).

[61] Cosmides, L. M. & Tooby, J. Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology* **89**, 83–129 (1981).

[62] Sprengelmeyer, Q. D. *et al.* Recurrent Collection of Drosophila melanogaster from Wild African Environments and Genomic Insights into Species History. *Molecular Biology and Evolution* **37**, 627–638 (2020). URL https://doi.org/10.1093/molbev/msz271.

[63] Burton, R. S. The role of mitonuclear incompatibilities in allopatric speciation. *Cellular and Molecular Life Sciences* **79**, 103 (2022). URL https://doi.org/10.1007/s00018-021-04059-3.

[64] Grenier, J. K. *et al.* Global Diversity Lines–A Five-Continent Reference Panel of Sequenced Drosophila melanogaster Strains. *G3 Genes|Genomes|Genetics* **5**, 593–603 (2015). URL https://doi.org/10.1534/g3.114.015883.

[65] Ballard, J. W. O. Comparative Genomics of Mitochondrial DNA in Members of the Drosophila melanogaster Subgroup. *Journal of Molecular Evolution* **51**, 48–63 (2000). URL https://doi.org/10.1007/s002390010066.

[66] Ballard, J. W. O. Comparative Genomics of Mitochondrial DNA in Drosophila simulans. *Journal of Molecular Evolution* **51**, 64–75 (2000). URL https://doi.org/10.1007/s002390010067.

[67] Ma, L., Clark, A. G. & Keinan, A. Gene-Based Testing of Interactions in Association Studies of Quantitative Traits. *PLOS Genetics* **9**, e1003321 (2013). URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003321. Publisher: Public Library of Science.

[68] Chyb, S. & Gompel, N. Wild-type morphology. In Chyb, S. & Gompel, N. (eds.) *Atlas of Drosophila Morphology*, 1–23 (Academic Press, San Diego, 2013). URL https://www.sciencedirect.com/science/article/pii/B9780123846884000018.

[69] Jaeger, B. C., , E., Lloyd J., , D., Kalyan, & Sen, P. K. An R2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics* **44**, 1086–1105 (2017). URL https://doi.org/10.1080/02664763.2016.1193725. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02664763.2016.1193725.

[70] Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48 (2015). URL https://doi.org/10.18637/jss.v067.i01.

[71] Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**, 1–26 (2017). URL https://doi.org/10.18637/jss.v082.i13.

[72] Lenth, R. V. *et al.* emmeans: Estimated Marginal Means, aka Least-Squares Means (2025). URL https://cran.r-project.org/web/packages/emmeans/index.html.

[73] Gauch, H. G. Statistical Analysis of Yield Trials by AMMI and GGE. *Crop Science* **46**, 1488–1500 (2006). URL https://acsess.onlinelibrary.wiley.com/doi/10.2135/cropsci2005.07-0193.

[74] Kona, P. *et al.* AMMI and GGE biplot analysis of genotype by environment interaction for yield and yield contributing traits in confectionery groundnut. *Scientific Reports* **14**, 2943 (2024). URL https://www.nature.com/articles/s41598-024-52938-z. Publisher: Nature Publishing Group.

[75] Mendiburu, F. d. agricolae: Statistical Procedures for Agricultural Research (2023). URL https://cran.r-project.org/web/packages/agricolae/index.html.

[76] Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). URL https://doi.org/10.1093/bioinformatics/btp324.

[77] Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* **43**, 11.10.1–11.10.33 (2013).

[78] Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w^{1118} ; iso-2; iso-3. *Fly* **6**, 80–92 (2012). URL http://www.tandfonline.com/doi/abs/10.4161/fly.19695.

[79] Wolff, J. N. *et al.* Evolutionary implications of mitochondrial genetic variation: mitochondrial genetic effects on OXPHOS respiration and mitochondrial quantity change with age and sex in fruit flies. *Journal of Evolutionary Biology* **29**, 736–747 (2016). URL https://doi.org/10.1111/jeb.12822.

[80] Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3603318/.

[81] Blomberg, S. P., Garland JR., T. & Ives, A. R. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution* **57**, 717–745 (2003). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0014-3820.2003.tb00285.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0014-3820.2003.tb00285.x.

[82] Revell, L. J. phytools: Phylogenetic Tools for Comparative Biology (and Other Things) (2024). URL https://cran.r-project.org/web/packages/phytools/index.html.

[83] Early, A. M. & Clark, A. G. Monophyly of Wolbachia pipientis genomes within Drosophila melanogaster: geographic structuring, titre variation and host effects across five populations. *Molecular Ecology* **22**, 5765–5778 (2013). URL https://onlinelibrary.wiley.com/doi/10.1111/mec.12530. Publisher: John Wiley & Sons, Ltd.

Extended Data

Table 1: Analysis of variance for all four phenotypes measured stratified on sex groups. Emboldened/grey highlighted terms and p-values indicate significance at Bonferroni corrected significance threshold $\alpha = \frac{0.05}{4 \text{ traits}} = 0.0125$. Terms significant at a less conservative threshold $\alpha = \frac{0.1}{4 \text{ traits}} = 0.025$ are only emboldened. Note that the larval density effect is quantified in the development time trait.

	1st Order			2nd Order			3rd Order		
	Fixed Effect	*P*	$R^2_{\beta*}$	Fixed Effect	*P*	$R^2_{\beta*}$	Fixed Effect	*P*	$R^2_{\beta*}$
	Weight								
♀	**Mito**	**0.002**	**0.188**	Mito:Nuc	0.238	0.104	Mt:Nuc:Treat	0.036	0.049
	Nuc	**<0.001**	**0.145**	Nuc:Treat	0.763	<0.001			
	Treatment	0.753	<0.001	Mito:Treat	0.566	0.027			
♂	**Mito**	**<0.001**	**0.141**	**Mito:Nuc**	**<0.001**	**0.089**	Mt:Nuc:Treat	0.590	0.031
	Nuc	**0.016**	**0.009**	Nuc:Treat	0.322	0.001			
	Treatment	0.101	0.004	Mito:Treat	0.260	0.041			
	Climbing velocity								
♀	**Mito**	**0.005**	**0.050**	Mito:Nuc	0.026	0.040	**Mt:Nuc:Treat**	**<0.001**	**0.021**
	Nuc	**<0.001**	**0.374**	Nuc:Treat	0.987	<0.001			
	Treatment	**<0.001**	**0.034**	**Mito:Treat**	**<0.001**	**0.027**			
♂	Mito	0.683	0.020	**Mito:Nuc**	**0.015**	**0.049**	**Mt:Nuc:Treat**	**<0.001**	**0.032**
	Nuc	**<0.001**	**0.183**	Nuc:Treat	0.067	<0.001			
	Treatment	**<0.001**	**0.024**	**Mito:Treat**	**<0.001**	**0.024**			
	Flight performance								
♀	Mito	0.039	0.245	Mito:Nuc	0.204	0.189	Mt:Nuc:Treat	0.115	0.128
	Nuc	**<0.001**	**0.752**	Nuc:Treat	0.062	0.016			
	Treatment	0.625	0.001	**Mito:Treat**	**0.022**	**0.166**			
♂	Mito	0.529	0.118	Mito:Nuc	0.175	0.164	Mt:Nuc:Treat	0.895	0.066
	Nuc	**<0.001**	**0.335**	Nuc:Treat	0.480	0.002			
	Treatment	**<0.001**	**0.128**	**Mito:Treat**	**0.021**	**0.196**			
	Development time								
	Mito	0.439	0.082	Mito:Nuc	0.096	0.125	Mt:Nuc:Treat	0.712	0.020
	Nuc	**<0.001**	**0.217**	**Nuc:Treat**	**0.017**	**0.006**			
	Treatment	**<0.001**	**0.024**	**Mito:Treat**	**0.006**	**0.047**			
	Larval d	**<0.001**	**0.015**						

Table 2: Amino acid (AA) changes for mtDNA genes for *D. mel* haplotypes. Differences shown are between each sampled haplotype and the *D. mel* reference genome (REF). Radical changes between AA classes (non-polar, uncharged polar, negatively charged, and positively charged) are highlighted in gray. Complex I sites that contact nuclear-encoded proteins are highlighted in red and emboldened. AA changes private to *D. sim* and/or *D. yak* are given in Table S5 and not shown here.

	Complex I								Complex III			Complex IV							Complex V	
	ND2		ND4			ND5			Cyt-b			CoI	CoIII						ATPase6	
	148	**311**	**84**	161	368	**450**	**486**	**551**	71	264	358	106	77	110	115	143	170	249	120	185
REF	Y	N	N	L	Y	I	N	S	C	V	V	A	T	A	A	V	L	V	N	I
Bei04	.	.	.	V	S	.	S	.	.	.	M	.	.	.	.	I	.	I	.	M
Bei05	.	.	S	V	S	.	S	.	.	.	M	.	.	.	.	I	.	I	.	M
Bei11	.	.	.	V	S	.	S	.	.	.	M	.	.	.	.	I	.	I	.	M
Bei12	.	.	.	V	S	.	S	.	.	.	M	.	.	.	.	I	.	I	.	M
Bei23	.	.	.	V	S	.	S	.	.	.	M	.	.	.	.	I	.	I	.	M
Bei38	.	.	.	V	S	.	S	.	.	.	M	.	.	.	.	I	.	I	.	M
Bei42	.	.	.	V	S	.	.	.	.	M	.	.	.	.	.	.	.	.	.	M
Bei52	.	.	.	V	S	.	S	.	.	.	M	T	.	.	.	I	.	I	.	M
Bei54	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bei59	.	.	.	V	S	.	S	.	.	.	M	.	I	.	T	.	.	I	.	M
ZH23	.	.	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZH26	.	.	.	V	S	T	.	F	Y	.	.	.	.	.	.	.	.	.	.	.
ZH33	.	.	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M
ZH42	.	.	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZW09	.	.	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZW133	.	.	.	V	S	.	.	.	.	.	.	.	.	S	.	.	.	.	K	.
ZW140	.	S	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M
ZW142	.	.	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M
ZW144	.	.	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim53	F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.
D. sim sil	.	.	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. yak	.	.	.	V	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

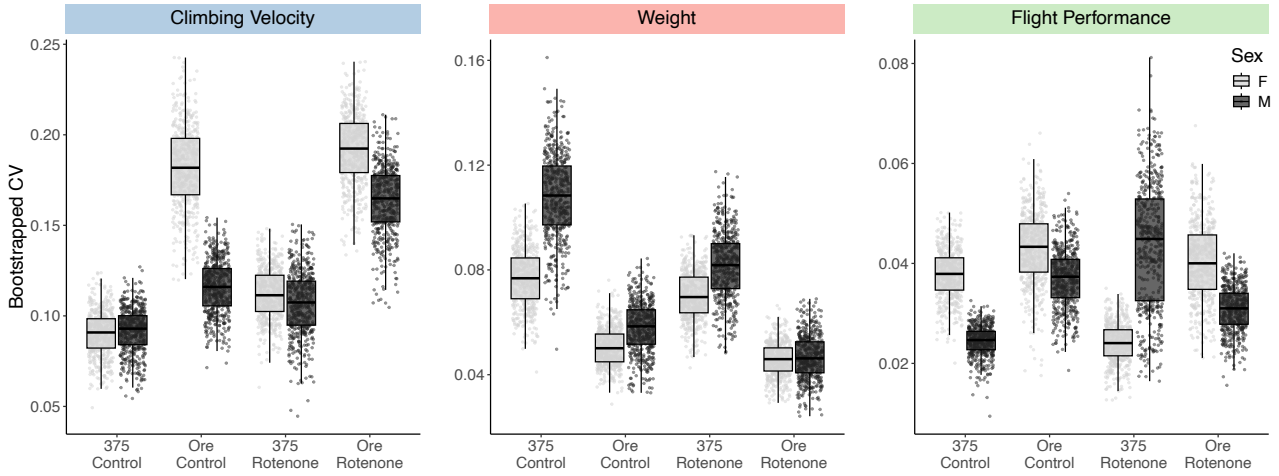

Figure 1: Bootstrapped (n=500) distributions of the coefficient of variation (CV) for each trait. Comparisons between females (F) and males (M) are given for each combination of nuclear background (OreR, DGRP375) and treatment environment (control, rotenone).

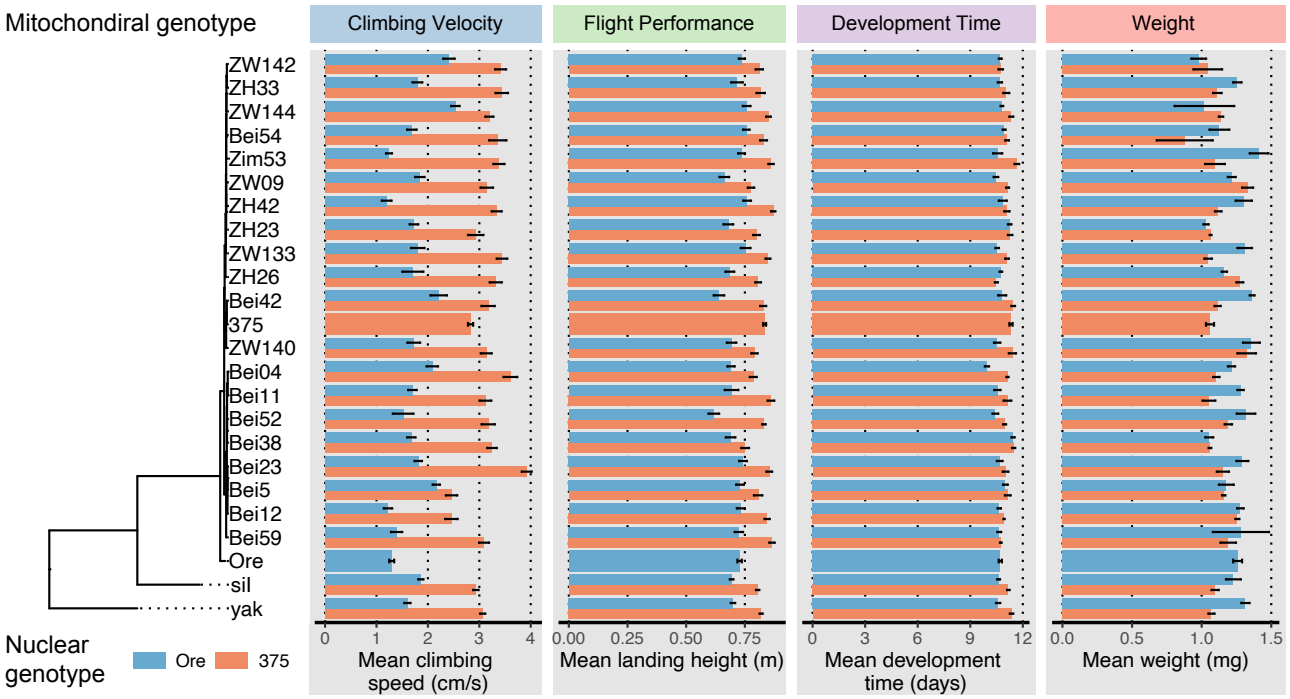

Figure 2: Relationship between mitochondrial phylogeny and all phenotypes for both OreR and DGRP375 nuclear genotypes. Only females on control food are shown.