Mitonuclear Epistasis for Development Time and Its Modification by Diet in *Drosophila*

Jim A. Mossman,[1] Leann M. Biancani,[2] Chen-Tseh Zhu, and David M. Rand[1]
Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912

ABSTRACT Mitochondrial (mtDNA) and nuclear genes have to operate in a coordinated manner to maintain organismal function, and the regulation of this homeostasis presents a substantial source of potential epistatic ($G \times G$) interactions. How these interactions shape the fitness landscape is poorly understood. Here we developed a novel mitonuclear epistasis model, using selected strains of the *Drosophila* Genetic Reference Panel (DGRP) and mitochondrial genomes from within *Drosophila melanogaster* and *D. simulans* to test the hypothesis that mtDNA × nDNA interactions influence fitness. In total we built 72 genotypes (12 nuclear backgrounds × 6 mtDNA haplotypes, with 3 from each species) to dissect the relationship between genotype and phenotype. Each genotype was assayed on four food environments. We found considerable variation in several phenotypes, including development time and egg-to-adult viability, and this variation was partitioned into genetic (G), environmental (E), and higher-order ($G \times G$, $G \times E$, and $G \times G \times E$) components. Food type had a significant impact on development time and also modified mitonuclear epistases, evidencing a broad spectrum of $G \times G \times E$ across these genotypes. Nuclear background effects were substantial, followed by mtDNA effects and their $G \times G$ interaction. The species of mtDNA haplotype had negligible effects on phenotypic variation and there was no evidence that mtDNA variation has different effects on male and female fitness traits. Our results demonstrate that mitonuclear epistases are context dependent, suggesting the selective pressure acting on mitonuclear genotypes may vary with food environment in a genotype-specific manner.

KEYWORDS MtDNA; epistasis; DGRP; genotype-by-environment interaction; fitness; development time; mitonuclear

UNDERSTANDING the genetic architecture of quantitative traits requires knowledge of higher-order genetic effects (Phillips 2008), since genes rarely operate in isolation in time and space (Bateson 1909; Fisher 1930; Wright 1931; Chetverikov *et al.* 1961). Nonlinear interaction between alleles—epistasis—is widespread within species (Corbett-Detig *et al.* 2013; Mackay 2014) and is central to a number of genetic processes, including canalization (Waddington 1942) and speciation (Dobzhansky 1937; Muller 1940). Moreover, epistasis is understood to play a major role in the expression of phenotypic variation and has wide implications for population genetics, evolutionary biology (Wagner and Altenberg 1996), human disease genetics (Marchini *et al.* 2005; Mackay and Moore 2014), and agriculture (Li *et al.* 1997).

In its simplest form, epistasis captures the phenotypic effect of one SNP being conditional on the allelic state of another SNP. Interlocus interactions are common and have been demonstrated to occur both within (Kondrashov *et al.* 2002; Kern and Kondrashov 2004; Weinreich *et al.* 2006; Povolotskaya and Kondrashov 2010; Hinkley *et al.* 2011) and between genes (reviewed in Lehner 2011). Epistatic interactions can help to explain why some organisms express genetic variants and others do not, *e.g.*, in the same way that disease-causing mutations in humans do not affect all people equally (Carlborg and Haley 2004; Lehner 2013). As a result, epistasis is a complex phenomenon that is very challenging to characterize (Mackay 2014) because the epistatic interactions cannot be predicted (in magnitude or direction) by single SNPs in isolation.

One major goal in quantitative genetics is to accurately predict phenotype from genotype. Over the past

doi: 10.1534/genetics.116.187286
Manuscript received January 18, 2016; accepted for publication March 4, 2016; published Early Online March 9, 2016.
Available freely online through the author-supported open access option.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.187286/-/DC1.
[1]Corresponding authors: Department of Ecology and Evolutionary Biology, Box G, Brown University, Providence, RI 02912. E-mail: Jim_Mossman@brown.edu; David_Rand@brown.edu
[2]Present address: Department of Biology, University of Maryland, College Park, MD 20742.

30 years, quantitative trait locus (QTL) studies have helped pinpoint genes underlying quantitative variation in traits across many organisms. However, in a large number of cases only a small amount of the phenotypic variation can be explained by genetic variants, leaving a large proportion of "missing" heritability (Manolio 2009). One possible source of missing heritability is a complex genetic architecture (Yang *et al.* 2015) involving the interaction among genes and also with their environment. Gene-by-environment interactions ($G \times E$) are prevalent in model species studies (Hillenmeyer 2008; Gerke *et al.* 2010) but demonstrate little conservation between species (Tischler *et al.* 2008), making them difficult to isolate. To address this problem, here we used the model organism *Drosophila melanogaster* to test whether epistases ($G \times G$) between mitochondrial DNA (mtDNA) and nuclear backgrounds (nDNA) are context specific and dependent on, or are modified by, their environment. We asked whether there are $G \times G \times E$ effects on phenotypic variation in the most commonly used population genetic resource in *Drosophila*: the *Drosophila* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012).

Previous investigations have revealed that life history traits in the DGRP have a mainly polygenic architecture, with both additive and nonadditive (epistatic) components (Huang *et al.* 2012). There is extensive pleiotropy and transcript connectivity between common pathways underlying complex traits in the DGRP (Ayroles *et al.* 2009; Zhou *et al.* 2012), and this provides an excellent opportunity to disrupt pathways that are influenced by mtDNA–nDNA epistasis. Surprisingly, there has been no assessment of the relative contribution of mtDNA genetic variation or the interaction between different mtDNA haplotypes on life history traits in the DGRP. A main goal of this study, therefore, was to construct a panel of mitonuclear genotypes, to disentangle the effects of mtDNA and nDNA variation and their ($G \times G$ and $G \times G \times E$) interactions. In this context, the term "environment" can be genetic and abiotic, as either one can modify the effects of a given gene. Such "background" effects are clearly evident in the DGRP (Huang *et al.* 2012) and this is likely the result of wide variation in genome size (169.7–192.8 Mb) and genomic architecture (indels and inversions, haplotype structure, presence of *Wolbachia*) between the strains (Huang *et al.* 2014).

The genetic architecture of mitochondrial function provides strong motivation for a genetic introgression of mtDNAs onto DGRP nuclear backgrounds. First, mitochondria have been existing semiautonomously in eukaryotic cells for ~2 billion years and in that time have evolved to function in the environment of the cell that is maintained by both genomes (Rand *et al.* 2004). Second, this coexistence requires extensive crosstalk between genomes and interaction via signaling pathways and coordinated gene expression (Woodson and Chory 2008). By disrupting these coevolved mtDNA–nDNA complexes, we can ask whether greater sequence divergence is associated with greater phenotypic disruption. Previous studies have suggested that components of fitness are sensitive to mitonuclear disruption both within species (Clark and Lyckegaard 1988; MacRae and Anderson 1988; Fos *et al.* 1990; Kilpatrick and Rand 1995; Garcia-Martinez *et al.* 1998; Rand *et al.* 2001; James and Ballard 2003; Ballard and James 2004; Dowling *et al.* 2007b, 2008; Innocenti *et al.* 2011; Camus *et al.* 2012; Yee *et al.* 2013; Paliwal *et al.* 2014; Chang *et al.* 2016) and in heterospecific mitonuclear introgressions (Hutter and Rand 1995; Rand *et al.* 2006; Montooth *et al.* 2010; Meiklejohn *et al.* 2013; Villa-Cuesta *et al.* 2014; Zhu *et al.* 2014; Holmbeck *et al.* 2015). The motivation for introgressing across the species boundary is to increase the amount of sequence divergence outside that experienced in nature, where more distantly related populations (and their DNA molecules) show greater mitonuclear effects (Clark and Lyckegaard 1988; Burton and Barreto 2012).

How much $G \times G$ variation is modified by environment is a hotly debated topic, and there is some good evidence that mitonuclear interactions are context specific (Dowling *et al.* 2007a; Arnqvist *et al.* 2010; Hoekstra *et al.* 2013; Zhu *et al.* 2014). However, only recently have studies focused on context-specific traits in the DGRP (Durham *et al.* 2014). Studies in *Drosophila* have demonstrated large $G \times E$ effects on a suite of physiological phenotypes, life history traits, and gene expression profiles (Takano *et al.* 1987; Zhou *et al.* 2012; Reed *et al.* 2014), including gene × diet interactions that can confer as large a magnitude of effect as diet in the first order (Reed *et al.* 2010). Since diet has been shown to influence genetic effects in *Drosophila* physiological traits (Vieira *et al.* 2000; Reed *et al.* 2010, 2014) and can be precisely manipulated for constituents and caloric content, we used diets with variable protein:carbohydrate ratios, along with a standard cornmeal laboratory food that flies are routinely maintained on, to test for $G \times G \times E$ effects in the mitochondrial DGRP (mitoDGRP).

One additional motivation is the question of sex-specific fitness effects for mitochondrial genes. The Frank and Hurst hypothesis (Frank and Hurst 1996) posits that mtDNA haplotype effects should be more prevalent in males than in females, due to stronger purifying selection from maternal inheritance. The hypothesis has gained some support in recent studies and there is some evidence in *Drosophila* that males are sensitive to mtDNA variation in both sex-limited traits such as sperm competitive ability (Yee *et al.* 2013) (but see Friberg and Dowling 2008) and sex-independent traits such as aging (Camus *et al.* 2012), mitochondrial enzyme activity (Sackton *et al.* 2003), and gene expression (Innocenti *et al.* 2011). In sex-limited traits (*e.g.*, male-specific sperm competitiveness) it is not possible to determine whether the mtDNA effects are really affecting males more than females, since females do not express them and cannot be tested. An arguably more rigorous test of the hypothesis is to select a trait (such as development time or aging) that is expressed in both sexes, allowing sex-dependent effects to be delineated. Thus, we focus on development time because it is an important fitness trait, it is sexually dimorphic in *Drosophila* (Bainbridge and Bownes 1981), and males and females

share a common developmental environment. Under the Frank and Hurst hypothesis, we predict that the coefficient of variation for development time among mtDNAs in males should be greater than that for females if purifying selection on mtDNA is indeed stronger in females. Here, we test this prediction across a variety of nuclear genetic backgrounds in an effort to identify general properties of mitonuclear interaction effects ($G \times G$) for fitness. Finally, we ask whether the dietary environment modified any $G \times G$ effects to test the robustness of first-order mtDNA effects and nuclear $\times$ mtDNA epistatic effects across environments. While our primary motivation is understanding the fitness landscapes of mitonuclear interactions, these studies are relevant to the ongoing debate about the genetic consequences of mitochondrial replacement therapies in humans.

Materials and Methods

The 72 mitonuclear genotypes panel

Experiments were conducted on a panel of mitonuclear genotypes in which six divergent mtDNA haplotypes, three from the *D. simulans* clade, (i) *D. simulans siI*, (ii) *D. simulans sm21*, and (iii) *D. mauritiana mau12*, and three *D. melanogaster* haplotypes, (iv) *D. melanogaster OreR*, (v) *D. melanogaster AutW132*, and (vi) *D. melanogaster Zim53*, were placed on each of 12 DGRP nuclear backgrounds (*DGRP-304*, *DGRP-313*, *DGRP-315*, *DGRP-358*, *DGRP-375*, *DGRP-517*, *DGRP-707*, *DGRP-712*, *DGRP-714*, *DGRP-765*, *DGRP-786*, and *DGRP-820*). The *D. simulans* and *D. mauritiana* mtDNA haplotypes are from distinct clades, although the *D. mauritiana mau12* mtDNA differs from the *D. simulans siIII* haplotype by only one nucleotide in the exonic region (Ballard 2000). In statistical models of species effects, we grouped the *D. simulans* and *D. mauritiana* haplotypes together as in previous analyses (Montooth *et al.* 2010). Pairwise divergence estimates between haplogroups show there are up to 103 amino acid differences, 438 synonymous nucleotide substitutions, and 79 RNA nucleotide substitutions among these six mtDNAs (see Montooth *et al.* 2010 for all pairwise comparisons).

Nuclear genetic backgrounds for mitonuclear introgression in the DGRP were selected to span the range of phenotypic values for (i) starvation resistance, (ii) lifespan, (iii) chill coma recovery, (iv) locomotor reactivity, (v) fitness, and (vi) copulation latency as revealed in the "core 40" DGRP lines (Ayroles *et al.* 2009). We hypothesized that disruption of mitonuclear genomes would alter phenotypic expression but we did not want to bias our analyses by selecting genotypes that would be unlikely to be informative or were from only one part of the global phenotypic distribution. Mitochondrial DNAs were introgressed via precise balancer substitution, using the crossing scheme outlined in Zhu *et al.* (2014). After chromosome substitutions were completed, male parents from each original DGRP line were backcrossed to virgin females from each newly constructed mitoDGRP line for several generations to remove any residual nuclear genomic variation that might have been retained during the introgressions.

Eliminating Wolbachia

Prior to the development time assays (>6 months), larvae (and subsequent adults) of the 72 constructed mitonuclear genotypes were cultured on Instant Carolina Media with 0.03% tetracycline for two generations. Strains were then screened for *Wolbachia* infection after treatment, using two separate *Wolbachia*-specific primer pairs: (i) 1F, 5′-ttgtagcctgctatggataact-3′, 1R, 5′- gaataggtatgattttcatgt-3′ and (ii) 2F, 5′-tgtggtgccagagtacttgaa-3′, 2R, 5′-gctttataagcgcgttcagc-3′. *Wolbachia*-positive controls were run in the same PCRs and failure of samples to amplify either PCR product was evidence of *Wolbachia*-negative status. All the mitonuclear genotypes used in this study were *Wolbachia* negative.

Diet

To test whether the balance of protein and carbohydrate in the form of yeast and sugar, respectively, could modify $G \times G$ mitonuclear effects, we tested flies on three isocaloric diets, along with a "standard" diet the flies in our laboratory are routinely maintained on. The three experimental food treatments were previously published diets whose caloric content is essentially equal [high protein:carbohydrate (P:C) = 452 kJ/100 g, equal P:C = 456 kJ/100 g, and low P:C = 469 kJ/100 g] but with a protein and yeast imbalance (Matzkin *et al.* 2011). Using these diets, the development time of *D. melanogaster* was found to be significantly slower in the low P:C diet, compared to the high P:C diet (Matzkin *et al.* 2011). The dietary compositions of all four diets in the present study are shown in Table 1. Tegosept was used as an antifungal agent in replacement of methyl paraben used in Matzkin *et al.* (2011). To make the food, water was first boiled and the agar added and stirred until dissolved. The sucrose, SAF yeast, and yellow cornmeal were then added to the agar solution and cooked at high temperature (95°) until the food was of uniform consistency and the cornmeal cooked. The food was allowed to cool to ~55° and a mixture of the Tegosept dissolved in 95% ethanol was then added and well stirred. Ten milliliters of cooled food was pumped into 8-dram glass vials and cooled until solid.

Development time assays

The experiment was conducted in two discreet blocks. In the first block (April 2014), we assessed development time, along with the number of eggs produced from a known number of females in the laboratory food treatment. In the second block (December 2014), we conducted development time assays on the three isocaloric diets but did not count the number of eggs produced due to logistical limitations. Our data collected from both experimental blocks allowed the full four-diet $G \times G \times E$ analysis to be conducted across all phenotypes except number of eggs laid. Block 1 data were collinear with the laboratory food diet and any effect of block could not be delineated

Table 1 The dietary composition of the four food types

Ingredient	High P:C	Equal P:C	Low P:C	Laboratory food[a]
Agar (g)	1	1	1	1.8
SAF yeast (g)	32	20	8	5
Yellow cornmeal (g)	9	9	9	10.4
Sucrose (g)	8	20	32	22
Tegosept (g)	0.45	0.45	0.45	0.9
95% ethanol (ml)	4.5	4.5	4.5	4.5
Distilled H_2O (ml)	200	200	200	200

The three isocaloric nutritional geometry food types [high protein:carbohydrate (P:C) ratio, equal P:C, and low P:C], along with the standard laboratory food are shown.

[a] The standard laboratory food uses a 20% Tegosept mixture instead of 10% methyl paraben in the Matzkin *et al.* (2011) diets.

from diet in statistical analyses. As such, we pooled the data for both studies; the number of eggs per female could therefore be parsed only into $G \times G$ effects and not $G \times G \times E$ effects, since it was conducted only in one environment.

The experimental setup was the same for all food types. Prior to the development time assays, flies were reared on our standard laboratory food (see Table 1). Two generations of density control (25 ♂ and 25 ♀) in bottles preceded the egg-laying phase.

To control for larval density effects, development time was scored using a known number of eggs picked from egg-lay plates. For egg laying, ~50 mated males and females from each of the 72 genotypes were placed into two replicate egg-laying cups on a solid grape juice–agar laying medium. The laboratory food experiment used a known number of females in each egg-laying plate (mean = 16.86 ± 4.67, 1 SD, range = 15–30). Because of the large number of genotypes to be assayed, we conducted the egg lays over 3 consecutive days with each genotype having two replicate plates per day. Every genotype was sampled on each day. There was no significant effect of "day" on development time or egg-to-adult survival, and we therefore pooled the data across the 3 days.

Flies were introduced onto egg-lay plates without CO_2 anesthesia at 6 PM the day before egg picking and were removed from the plates at 6 AM the following morning, allowing them to lay eggs for 12 hr. While this was a long egg-lay period, the eggs were randomly picked across the plate and allocated to each of the four diet types. While a long egg lay may inflate the within-genotype variance in each food type, we assumed each genotype × food combination would experience the same degree of inflation, since eggs were randomly allocated across foods. As eggs from the same plate were allocated across food treatments, we fitted the egg-lay plate ID as a random effect variable in statistical models to account for this shared provenance and egg-laying environment.

The target number of eggs that were placed in vials was 30 across all genotype and food combinations. In some cases (630/2849 vials), genotypes produced low yields of eggs and insufficient numbers for 30 eggs per vial. In these cases, the largest possible number of eggs was picked to make a complete set of vial replicates (12 replicates in protein:carbohydrate isocaloric foods and 6 replicates in laboratory food). The numbers of eggs in the vials with <30 eggs were as follows: mean = 12.23 ± 6.87 (1 SD), range = 1–29. To model these differences in the data set, we conducted our model selection and fitted the number of eclosed offspring as a fixed-effect covariate. In no case was the model with the term fitted a significantly better fit of the data, based on lowest Akaike information criterion (AIC) values (see *Model selection and validation*).

Phenotyping

After egg picking, vials were placed into a controlled environment room at 25° on a 12-hr light:12-hr dark cycle. Vials were scored for eclosing offspring twice daily, at 9 AM and 5 PM. All eclosed flies were scored for time and sex. After eclosion, flies were removed from the vial and discarded. Development time scoring continued until all pupae had eclosed or all remaining pupae were visibly dead. Using these data we compiled vial estimates on (i) mean development time (across both sexes), (ii) mean male development time, (iii) mean female development time, (iv) egg-to-adult viability (survival), (v) coefficient of variation for development time in males, (vi) coefficient of variation for development time in females, (vii) the male–female difference in mean development time, (viii) vial sex ratio, and (ix) the male–female coefficient of variation difference in development time. Coefficient of variation (CV) was calculated as the standard deviation (σ) divided by the mean (μ): CV = σ/μ.

We wanted to ascertain whether our DGRP development time phenotypes were correlated with other published phenotype and gene expression data, utilizing the rich resources available to the DGRP community. There are no publicly available development time data for the complete set of DGRP genotypes; however, Ellis *et al.* (2014) sampled egg-to-pupae and eclosion development times in a subset of 50 DGRP lines of which 4 overlapped with the lines in the present study. We collected these data and correlated them to appropriate comparable development time data in the present study (taking into consideration temperature and food type in the previously published study).

The data that we correlated were (i) development time in 4/12 DGRP lines (Ellis *et al.* 2014), (ii) starvation stress resistance, (iii) chill coma recovery, and (iv) startle response (Mackay *et al.* 2012) (downloaded from http://dgrp2.gnets.ncsu.edu); (v) life span (Ivanov *et al.* 2015); and (vi) gene expression measured on Affymetrix *Drosophila* 2.0 arrays (Ayroles *et al.* 2009) (downloaded from http://dgrp2.gnets.ncsu.edu). For the *in silico* screen of gene expression correlations with development time phenotypes, we downloaded a list of 280 nuclear genes that are known to encode mitochondrially targeted proteins [MitoDrome (Sardiello *et al.* 2003; D'Elia *et al.* 2006)]. These are good candidates for genes that interact with mtDNAs, since components of the oxidative phosphorylation (OXPHOS) pathway of the electron transport chain are jointly encoded by both genomes and are expressed in the same regions of the mitochondrion.

Of the 280 protein-coding genes that are known, 179 show variable expression in the DGRP (Ayroles *et al.* 2009) and we screened these for putative expression quantitative trait loci (eQTL), whose expression segregated with our development time phenotypes.

As a supplementary analysis, we aligned available complete protein-coding regions of mtDNA in five of the six haplotypes used in this study (*AutW132* is not fully sequenced) to identify amino acid polymorphisms segregating between the haplotypes (Supplemental Material, Table S1). The aligned sequences were (i) *D. melanogaster* isolate Zim53 (AF200829.1), (ii) *D. melanogaster* isolate Oregon R (AF200828.1), (iii) *D. simulans* (*siI*) isolate TT01 (AF200834.1), (iv) *D. simulans* (*siII*) isolate DSR (AF200841.1), and (v) *D. mauritiana* isolate BG1 (AF200831.1).

Mutations in mtDNAs are in complete linkage disequilibrium within a haplotype and mapping disease-related mutations is challenging. To help understand whether any mutations could have a potential role in mitonuclear interactions at the protein function level, for each protein-coding gene we took the polypeptide sequence of a mtDNA haplotype (*e.g.*, *Oregon R*) and made *in silico* amino acid substitutions of mutations that are resident on an alternative haplotype (*e.g.*, *siI*). This was repeated for all mutations that were different between the two haplotypes. We conducted an *in silico* screen of all these polymorphisms, using PredictSNP software (Bendl *et al.* 2014). PredictSNP is a consensus classifier of disease-related mutations that are likely to affect protein function. The software considers the prediction output of best-predictor mutation evaluation tools that assess evolutionary, physico-chemical, and structural characteristics of the focal protein. We used the output from six best-predictor tools: (i) MAPP (Stone and Sidow 2005), (ii) PhD-SNP (Capriotti *et al.* 2006), (iii) PolyPhen-1 (Ramensky *et al.* 2002), (iv) PolyPhen-2 (Adzhubei *et al.* 2010), (v) SIFT (Sim *et al.* 2012), and (vi) SNAP (Bromberg and Rost 2007), along with the consensus caller (PredictSNP) to assess the putative effect of amino acid polymorphisms. The output is a consensus estimate of whether the mutation is likely to affect protein function or is likely to be neutral (Figure S1).

Statistical analyses

Our experimental design required that we conduct our investigation over 3 consecutive days. There was no significant effect of the day of the study on our estimates of development time and so we pooled these data together, along with the data from all four food types (see above). Development time data were first log-transformed prior to analysis. Development time, sex differences in development time, sex bias (number of males in brood–number of females), and sex differences in coefficients of variation in development time data were modeled as Gaussian distributions in linear mixed-effect (lmer) models implemented in the lme4 R package (Bates *et al.* 2014). Egg-to-adult viability data were modeled using generalized linear mixed-effect (glmer) models with binomial distribution and a logit link function.

We tested the hypotheses that DGRP nuclear variation, mtDNA variation, and DGRP nuclear × mtDNA variation were associated with development time. Also we included the diet type as a fixed and interaction effect to determine whether diet modified DGRP nuclear and mtDNA effects and their potential epistatic interactions. Since our data set was fully factorial (12 DGRP × 6 mtDNA haplotypes × 4 diets), we performed analysis of variance to partition the amount of variance in the dependent variable that was explained by the fixed effects. We modeled egg-pick plate ID as a random effect in mixed models (see above). Phenotypic variance was partitioned using the ANOVA model,

$$Y_{ijklm} = \mu + M_j + N_k + E_l + (MN)_{jk} + (ME)_{jl} + (NE)_{kl} + (MNE)_{jkl} + T_{ijklm} + Pl_i + \varepsilon_{ijklm},$$

where Y_{ijklm} is the phenotypic measurement taken from the mth individual with the jth mtDNA genotype (M) and the kth nuclear background (N) on the lth food type (E). $(MN)_{jk}$ is the mito × nuclear interaction (epistasis) term, $(ME)_{jl}$ is the mito × food type interaction, $(NE)_{kl}$ is the nuclear genotype × food interaction, and MNE_{jkl} is the mito × nuclear × food type ($G \times G \times E$) interaction term. The total number of offspring eclosing from the vial for each treatment was fitted as a model covariate (T_{ijklm}) and Pl_i is the random effect of plate ID.

Handling missing data

In the nutritional diet treatments, there were missing data for high P:C, equal P:C, and low P:C foods in one of the 72 mitonuclear genotypes [*sm21*; *DGRP-313*: that is, the *sm21* mtDNA haplotype on the DGRP 313 nuclear background (*DGRP-313*)]. From here on, we use this notation to describe mito;nuclear genotypes. These missing data represent 3/288 $G \times G \times E$ treatments (~1% of the total data). We wanted to confirm that these missing data did not influence the qualitative results in our analyses so we imputed development time data for those food types, using a *hot deck* procedure (Andridge and Little 2010). Briefly, a hot deck imputation involves replacing missing values of a variable (in this case development time, egg-to-adult survival, sex ratio) with values from a "similar" sample with respect to shared characteristics, *e.g.*, food type and genotype. In the first case, we filtered the total data set to include only samples that had the *sm21* mtDNA haplotype *and* the high P:C food type. We randomly sampled 10 rows of the data set that had these shared mtDNA and food type variables for each missing genotype × food treatment. In the statistical analyses we conducted the same analyses on the imputed (balanced) data sets and the nonimputed data sets. The results of the separate analyses were qualitatively identical (data not shown). In the *Results* section we report the nonimputed data.

Model selection and validation

Full models including the terms "nuclear genotype," "mtDNA genotype," and "food type" and all pairwise and three-way

interaction terms were constructed with total number of eclosed offspring in the vial fitted as a covariate. To conduct best-fit model selection for development time we used the *dredge* function implemented in the MuMIn R package (Barton 2015). This package provides automatic model selection based on the global model (above) and allows a subset of the terms to be retained during model selection (specified were first-order effects of nuclear genotype, mtDNA genotype, and food type). Models were fitted by maximum likelihood (ML), as recommended for comparing changing fixed-effect structures (Faraway 2005; Zuur *et al.* 2011), and then ranked based on AIC values, with $\Delta AIC > 2$ providing good support for a model term being removed (Table S3). We used these model evaluations to confirm our ANOVA results. For lmer models (lmers), we used the *anova* function in the lmerTest R package (Kuznetsova *et al.* 2015) with type III sums of squares and Satterthwaite's degrees-of-freedom approximation. Significant nuclear DGRP × mtDNA interaction terms in the ANOVA tables (lmer) or analysis of deviance tables (glmer) were considered evidence of $G \times G$ epistasis. Higher-order significant interactions with diet were considered evidence of $G \times G \times E$ effects. That is, the environment modified the $G \times G$ effect.

To test whether nuclear DGRP × mtDNA interactions (epistases) were of different magnitude across food types, we calculated the intraclass correlation (ICC) in each food type separately. The ICC in this context is an indicator of the amount of crossing of norms of reaction—in other words, we asked whether development time in a particular food type was associated with more crossing reaction norms or was more consistent between mtDNA types. To illustrate this variable we made two dummy data sets, one with a uniform value of development time for each mtDNA type in a given nuclear background and the second dummy data set with a random value (in the same range as the first dummy data set) for each mtDNA. The first dummy data set produced straight lines with zero crossing and the second dummy data set randomly crossed between mtDNA types. The ICC values were 1.0 for the first (uniform) data set and 0.02 for the random data set. There is therefore an index from 0 to 1 that corresponds to the amount of reaction norm crossing (inconsistency) between mtDNA types. For our real data, we calculated the ICC in two ways. First, we constructed a linear mixed-effect model with development time as the dependent variable and mtDNA type as the independent variable (fixed effect). The nuclear genotype was fitted as a random variable. The ICC was calculated as the (random effect variance)/(random effect variance + residual) for the model. The second method used the *ICCest* procedure in the ICC R package (Wolak *et al.* 2012). Both methods provided qualitatively identical results. We report values from the second (ICC) method.

Reciprocal crosses

Mitochondrial DNAs are maternally inherited in *Drosophila*. To experimentally validate the mtDNA effects we found in this study, we selected two mtDNA genotypes within seven DGRP backgrounds that demonstrated divergent phenotypes and performed reciprocal crosses between them. The rationale of this experiment was to validate that the phenotypic effects we observed would track the transmission of mtDNA in the cross and provide good evidence that the genetic entity responsible was of cytoplasmic origin. For each of the seven nuclear backgrounds we performed two reciprocal crosses between mtDNA genotypes, in addition to the within mitonuclear genotype crosses (two per nuclear background). We density controlled the flies for one generation prior to the reciprocal crosses. For each cross ~30 males and 30 females of the desired genotypes were mated for 2 days and transferred to an egg-lay plate. Egg lays occurred for 12 hr (as in the main experiment) and eggs were picked and placed into vials of the standard laboratory food (Table 1). Offspring were scored for development time as in the main experiment. We used *post hoc* Tukey's honest significant difference (HSD) tests to determine whether reciprocal crosses were similar to either of the parental genotypes.

Data availability

All statistical procedures were performed in R v3.1.3 (R Core Team 2015). The genotypes used in this investigation are available upon request. Data are available in File S1 and File S2, respectively, to perform reanalyses of development time phenotypes in the main experiment and in the reciprocal cross experiment. Female fecundity data are in File S3 and coefficients of development time variation data are in File S4.

Results

Development time

We assayed the egg-to-adult development time across 72 genotypes and four food types with different dietary composition. At the individual vial level, there was large variation in mean development time across all mitonuclear × food type combinations ranging from 192 to 288 hr (Figure 1). By far the largest source of this variation was the food type ($F = 3157.79$, $P < 2.2e\text{-}16$). The rank order of development from fastest to slowest was high P:C < equal P:C < low P:C < laboratory food, matching the rank order of decreasing protein (yeast) in each diet. There was a significant genetic component to development time phenotypes, with both nuclear variation ($F = 127.65$, $P < 2.2e\text{-}16$) and mtDNA variation ($F = 10.14$, $P < 3.23e\text{-}09$) accounting for much of the variation (Table 2A). There was no significant influence of the vial productivity on development time in the model ($P = 0.42$).

Second-order interaction effects revealed there were significant nuclear × mtDNA epistases for development time among the genotypes ($F = 7.3$, $P < 2.2e\text{-}16$). Nuclear genetic variation was also sensitive to food type and there was a strong interaction effect ($F = 24.63$, $P < 2.15e\text{-}16$) whereby the impact of nuclear variation changed in different food types. Figure 2, A–D, illustrates these mitonuclear $G \times G$

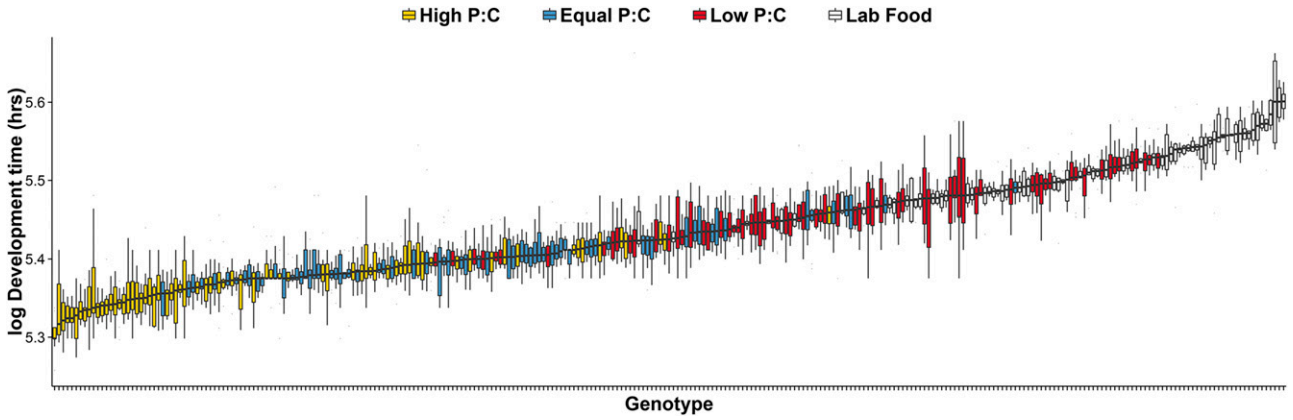


Figure 1 Development time variation across all $G \times G \times E$ combinations. Boxplots of each DGRP ID × mtDNA ID × food type are shown and are ranked by median value. Boxes in yellow, blue, red, and white correspond to the high P:C, equal P:C, low P:C, and laboratory food, respectively. Mitonuclear genotype names are omitted due to space constraints.

effects and the impact of diet on these genotypes (comparisons across Figure 2, A–D). Along with nuclear variation, mtDNA variation was sensitive to food type and there was a significant mtDNA × food type interaction ($F = 3.38$, $P = 1.24e\text{-}05$).

We further asked whether the $G \times G$ epistatic interactions were context dependent and whether their effect on development time changed in the different food types. We found a significant $G \times G \times E$ (nuclear × mtDNA × food type) interaction ($F = 2.99$, $P < 2.2e\text{-}16$) showing the epistatic effects were modified by the different food types (Table 2A).

The magnitude of mtDNA effect on development time was variable in different foods and in different nuclear backgrounds. Considering the whole panel of genotypes (Figure 3), some nuclear backgrounds, *e.g.*, *DGRP-517* and *DGRP-786*, showed highly canalized responses to both mtDNA type and food type (there was negligible mtDNA variation and the food types were tightly grouped). In contrast, some genotypes, *e.g.*, *DGRP-820*, demonstrated greater mtDNA variation. There are clear examples of food affecting development time in different ways in alternative nuclear backgrounds. *DGRP-358* shows a bimodal distribution of development times conditional on food type; the low protein foods (low P:C and laboratory food) were clearly separated from the higher protein concentration diets (high P:C and equal P:C). Likewise, development times among the isocaloric foods were tightly clustered in *DGRP-786*, but were more divergent in the laboratory food treatment. Development time in *mau12*, *siI*, and *sm21* haplotypes in *DGRP-304* increased monotonically with longer development times in the *sm21* haplotype. In contrast, the opposite pattern could be observed in the same haplotypes in the *DGRP-517* nuclear background and these patterns were not disrupted by the food type.

All models with mtDNA species effects on phenotypes are described in Table S2. We hypothesized that the *D. simulans* species' haplotypes would confer greater effects on development time than those originating in *D. melanogaster*. In spite of the greater number of SNPs that have accumulated over evolutionary time in the noncoevolved *D. simulans* mtDNA haplotypes, there was only a hint of a mtDNA "species" effect on development time (Table S2A), with *D. simulans* haplotypes developing marginally slower overall when compared to the model intercept. Importantly, mtDNA species effects could be observed in some nuclear backgrounds and food conditions but were absent in others. To test whether there was any significant phylogenetic signal in our development time data (Figure 3), we conducted separate analyses using a phylogeny of the six mtDNA haplotypes and quantified the amount of phylogenetically independent signal, using Moran's autocorrelation coefficient, I (see Table S4, Figure S7 and File S5 for full details and results). In a subset of the complete data (development time in laboratory food), we found clear examples of DGRP nuclear backgrounds that were sensitive to mtDNA phylogenetic signal (*e.g.*, *DGRP-714* and *DGRP-765*), while others showed no effect of phylogeny, consistent with the main result of no overall species effect (Table S2A).

When analyzed as individual haplotypes, the patterns were different. There were some clear examples of mitonuclear epistasis conferring developmental delay or acceleration (*e.g.*, *mau12*; *DGRP-712*, see Figure 3). These "line" effects were generally evident when one of the haplotypes within a species deviated from the other two. For example, *siI*; *DGRP-315* shows negative epistasis (slower development time) in all food types when compared to the *mau12* and *sm21* haplotypes. However, food type modifies $G \times G$ interactions and we can see the opposite pattern in *Aut*; *DGRP-315* in which a positive epistasis (faster development) is not evident across all food types and is restricted to the laboratory food environment. These interactions underpin the nature of the $G \times G \times E$ effects we observed in the complete data set and serve as good examples of how food type can modify $G \times G$ effects, even when they are absent in one or multiple environments.

To quantify the degree of reaction norm crossing between mtDNA types on alternative nuclear backgrounds, we

Table 2 *G* × *G* × *E* effects on fitness in the mitonuclear genotype panel

Phenotype	Model term	Num DF	Den DF	*F*-value	*P*-value	Random variable	σ^2
A. Development time	Nuclear	11	469.35	127.65	**0**	Plate ID	0.00017
	mtDNA	5	454.37	10.14	**3.23*e*-09**	Residual	0.00045
	Food	3	1231.58	3157.79	**0**		
	Vial productivity	1	2282.90	0.66	0.42		
	Nuclear × mtDNA	55	453.08	7.30	**0**		
	Nuclear × food	33	1135.37	24.63	**0**		
	mtDNA × food	15	1131.02	3.38	**1.24*e*-05**		
	Nuclear × mtDNA × food	162	1130.68	2.99	**0**		

Phenotype	Model term	Num DF	Den DF	Chisq	*P*-value
B. Egg-to-adult viability	Nuclear	11		202.77	**1.98*e*-37**
	mtDNA	5		14.85	**0.01**
	Food	3		26.95	**6.04*e*-06**
	Vial productivity	1		3556.21	**0**
	Nuclear × mtDNA	55		183.04	**1.15*e*-15**
	Nuclear × food	33		171.61	**1.19*e*-20**
	mtDNA × food	15		12.45	0.64
	Nuclear × mtDNA × food	162		110.95	0.99

Phenotype	Model term	Num DF	Den DF	*F*-value	*P*-value	Random variable	σ^2
C. Male–female sex bias	Nuclear	11	563.64	2.69	**<0.01**	Plate ID	0.22
	mtDNA	5	554.13	0.80	0.55	Residual	13.37
	Food	3	1550.83	2.05	0.11		
	Vial productivity	1	1213.61	0.17	0.68		
	Nuclear × mtDNA	55	550.10	0.84	0.78		
	Nuclear × food	33	1526.56	2.14	**<0.001**		
	MtDNA × food	15	1546.86	0.71	0.78		
	Nuclear × mtDNA × food	162	1523.02	1.10	0.20		

Phenotype	Model term	Num DF	Den DF	*F*-value	*P*-value	Random variable	σ^2
D. $DT_{MALE} - DT_{FEMALE}$	Nuclear	11	392.36	17.73	**0**	Plate ID	3.41
	mtDNA	5	403.16	1.03	0.40	Residual	48.18
	Food	3	1165.94	0.66	0.57		
	Vial productivity	1	1253.45	0.19	0.66		
	Nuclear × mtDNA	55	377.61	1.25	0.12		
	Nuclear × food	33	1109.79	3.00	**4.53*e*-08**		
	mtDNA × food	15	1116.24	1.15	0.30		
	Nuclear × mtDNA × food	160	1124.10	1.34	**0**		

Phenotype	Model term	Num DF	Den DF	*F*-value	*P*-value	Random variable	σ^2
E. $CV_{MALES} - CV_{FEMALES}$	Nuclear	11	2120	1.62	0.09	Plate ID	0
	mtDNA	5	2120	0.97	0.44	Residual	0.001
	Food	3	2120	0.32	0.81		
	Vial productivity	1	2120	0.58	0.45		
	Nuclear × mtDNA	55	2120	1.45	**0.02**		
	Nuclear × food	33	2120	2.58	**2.43*e*-06**		
	mtDNA × food	15	2120	1.46	0.11		
	Nuclear × mtDNA × food	155	2120	1.17	0.08		

Phenotype	Model term	Num DF	Sum Sq	Mean Sq	*F*-value	*P*-value
F. Eggs per female	Nuclear	11	7519.41	683.58	34.07	**3.58*e*-24**
	mtDNA	5	336.05	67.21	3.35	**<0.01**
	Nuclear × mtDNA	55	3242.86	58.96	2.94	**1.06*e*-05**
	Residuals	72	1444.52	20.06		

Summaries of linear mixed-effect models (lmer), generalized linear mixed-effect models (glmer), and linear models (lm) are shown. (A) Development time, (C) male–female sex bias, (D) development time differences between males and females ($DT_{MALE} - DT_{FEMALE}$), and (E) the coefficient of variation differences between males and females ($CV_{MALES} - CV_{FEMALE}$) were modeled as Gaussian distributions in lmers and fitted with restricted maximum likelihood (REML). (B) Egg-to-adult viability (survival) was modeled as a binomial error distribution with a logit-link function and fitted with maximum likelihood (ML) and is reported as an analysis of deviance table. (F) Eggs per female were modeled as a Gaussian distribution (lm) and an ANOVA is reported. *P*-values in boldface type are nominally significant at $\alpha = 0.05$.

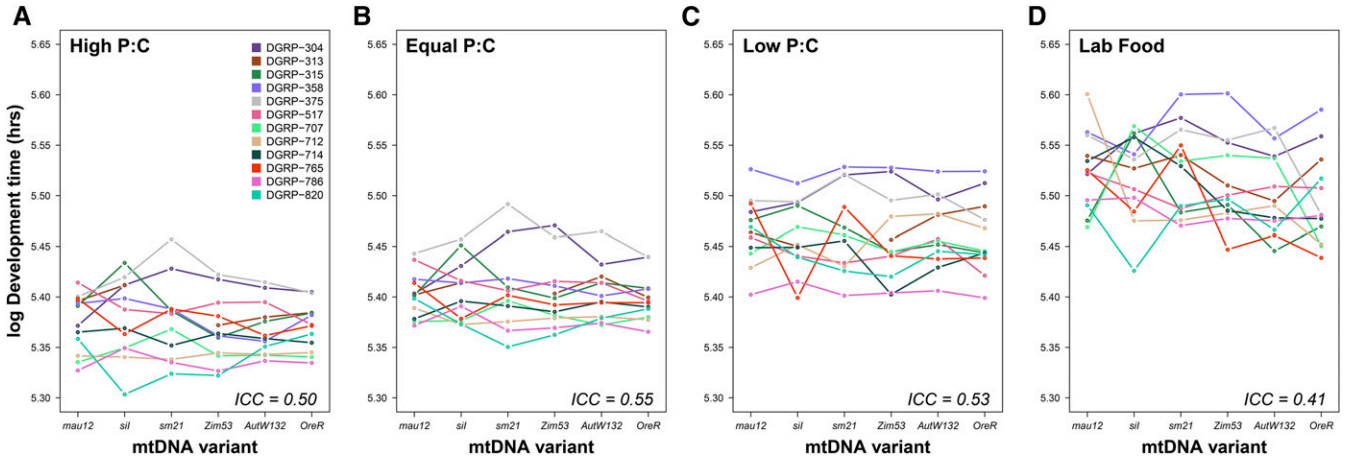


Figure 2 Mitonuclear epistases for development time across food types. Interaction plots show the mean values of each mitonuclear combination. DGRP IDs are color coded. Crossing norms of reaction are observable in the high P:C (A), equal P:C (B), low P:C (C), and laboratory food (D) types and are evidence of pervasive mito × nuclear epistases. Intraclass correlation coefficients (ICCs) are shown in each plot and are a measure of the magnitude of the change in rank order of DGRP mean across mtDNA genotypes. Line crossing was most evident in the laboratory food type (D), corresponding with the lowest ICC value.

calculated ICCs in the separate food types. Because of the unequal sampling (6 replicate vials for laboratory food and 12 replicate vials for the isocaloric foods), we randomly sampled 400 rows of data from each food type and repeated the ICC estimation 1000 times. The estimates from the real data are reported before the mean sampling estimates (in parentheses). The ICCs were as follows: (i) high P:C = 0.50 (0.498), (ii) equal P:C = 0.55 (0.550), (iii) low P:C = 0.53 (0.534), and (iv) laboratory food = 0.41 (0.407). The amount of reaction norm crossing was therefore greatest in the laboratory food type and lowest in the equal P:C food type, suggesting the mtDNA effects were dampened in the equal P:C diet. The high correlation between the estimates from the real data and the sampled subsets of the data suggests the ICC estimate on unbalanced sample sizes successfully captures the crossing differences between food types. Interestingly, there was no linear relationship between protein or carbohydrate concentration and the amount of crossing of reaction norms. The difference in crossing between the isocaloric food types was minor (range 0.50–0.55) with the laboratory food outside of this range (0.41).

Egg-to-adult viability (survival)

We assayed the proportion of the eggs that survived to adult stage and modeled this as a binomial distribution. The data are summarized in Figure 4 and Table 2. Nuclear variation is significantly associated with egg-to-adult survival (χ^2 = 202.77, P = 1.98*e*-37) in this panel of mitonuclear DGRP genotypes. The lowest overall survival was found in the *DGRP-313* nuclear background and the highest was in the *DGRP-712* background. There was a small, nominally significant effect of mtDNA type on survival (χ^2 = 14.85, P = 0.01) and a significant effect of food type (χ^2 = 26.95, P = 6.04*e*-06). Among the genotypes there were some mitonuclear combinations that showed significant deviation from other genotypes in the same nuclear background (*e.g.*, *Zm53;DGRP-313*) and some that showed significant line effects within species (*e.g.*, *siI;DGRP-714* and *siI;DGRP-765*). In the second examples the *siI* haplotype was associated with greater survival. Overall, mitonuclear epistatic effects were significant for egg-to-adult survival (χ^2 = 183.04, P = 1.15*e*-15). We found significant interaction effects between nuclear background and food type (χ^2 = 171.61, P = 1.19*e*-20), suggesting the DGRP backgrounds were conferring different effects in different food types. This can be observed in Figure 4, *e.g.*, in the *DGRP-707* and *DGRP-304* nuclear backgrounds, in which the rank order of egg-to-adult survival changes across the food types. Overall, there were no significant $G \times G \times E$ effects ($\chi^2 = 110.95, P = 0.99$). Notably, in *DGRP-304*, there is a monotonic relationship between food type and survival across all mtDNA backgrounds. In contrast, there are examples of food type having no consistent association across mtDNA types within a nuclear background (*e.g.*, *DGRP-707*).

Several nuclear backgrounds demonstrated canalized mtDNA variation for survival (Figure 4, *DGRP-358*, *DGRP-712*, and *DGRP-786*). Interestingly, food type had little influence on survival in these nuclear backgrounds, but had a large effect in others (*e.g.*, *DGRP-304*).

Sex ratio

Deviations in sex ratio can reflect sex-specific effects of genetic and environmental variation on organismal viability. The global sex ratio of all eclosed offspring in this study was 19,274 males:19,408 females—almost equality. Table 2C reports the analysis of male–female sex bias as a dependent variable. Male–female bias is a normally distributed trait and was modeled as a Gaussian distribution. There was a significant effect of nuclear background on the male–female bias across all treatments (Table 2C: F = 2.69, $P < 0.01$), but no significant main effects of mtDNA type (F = 0.8, P = 0.55) or

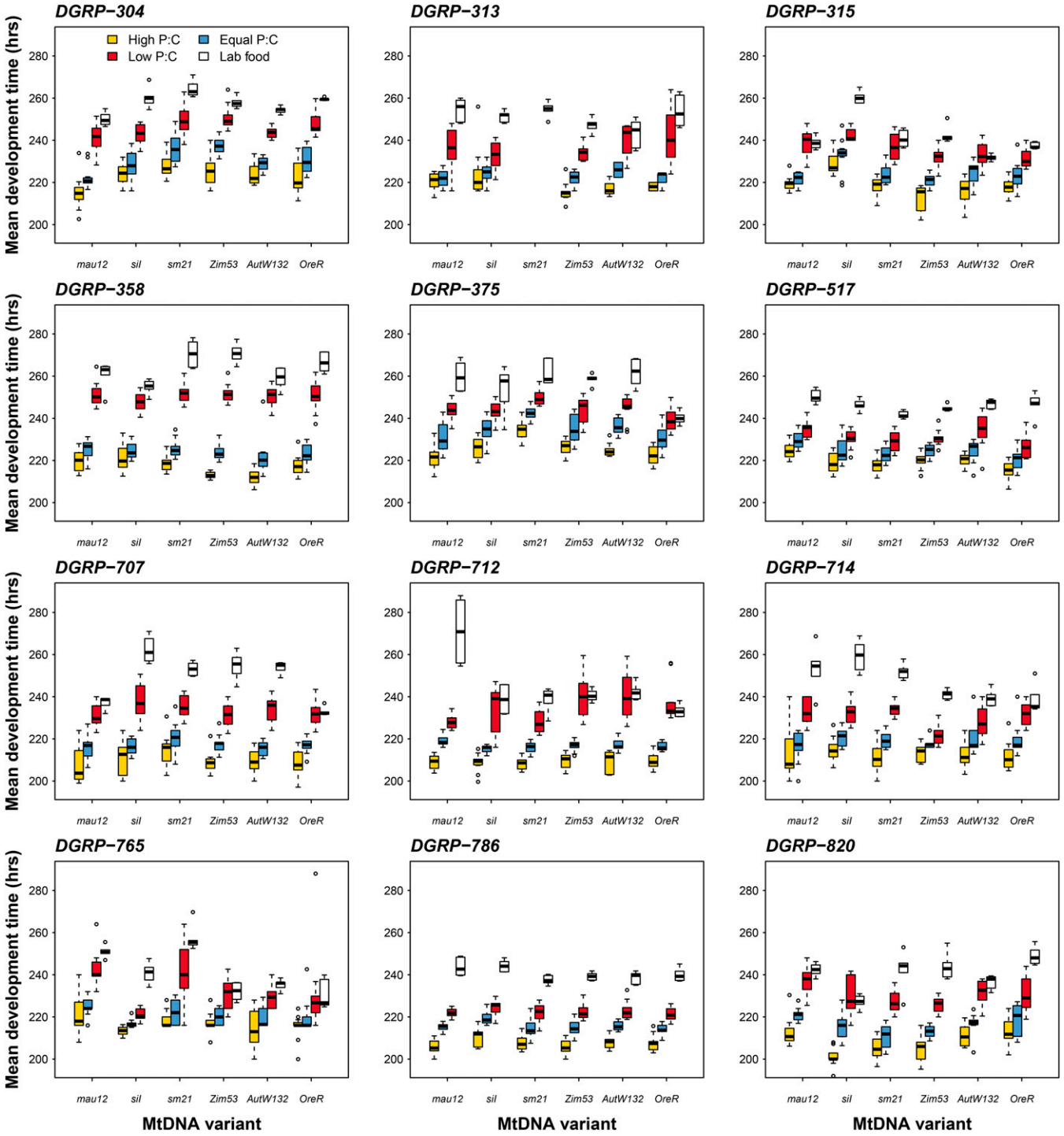


Figure 3 Phenotypic distributions of development time in each DGRP nuclear background in the mitonuclear genotype panel. MtDNA haplotypes are shown on the abscissa and development time is on the ordinal axis. Boxplots in yellow, blue, red, and white correspond to high P:C, equal P:C, low P:C, and laboratory food types, respectively. Each panel represents an individual DGRP nuclear background.

food type ($F = 2.05$, $P = 0.11$). Vial productivity was not associated with vial male–female bias ($F = 0.17$, $P = 0.68$), suggesting larger numbers of offspring were not associated with a larger proportion of males and vice versa. There was no evidence for $G \times G$ epistatic effects on sex ratio and the interaction term in the model was nonsignificant ($F = 0.84$, $P = 0.78$). Nuclear background × food type was, however, a significant term ($F = 2.14$, $P < 0.001$), resulting from some nuclear backgrounds that show larger variation between food types (*e.g.*, *DGRP-304*), whereas some nuclear backgrounds are invariant (*e.g.*, *DGRP-786*). There was no significant

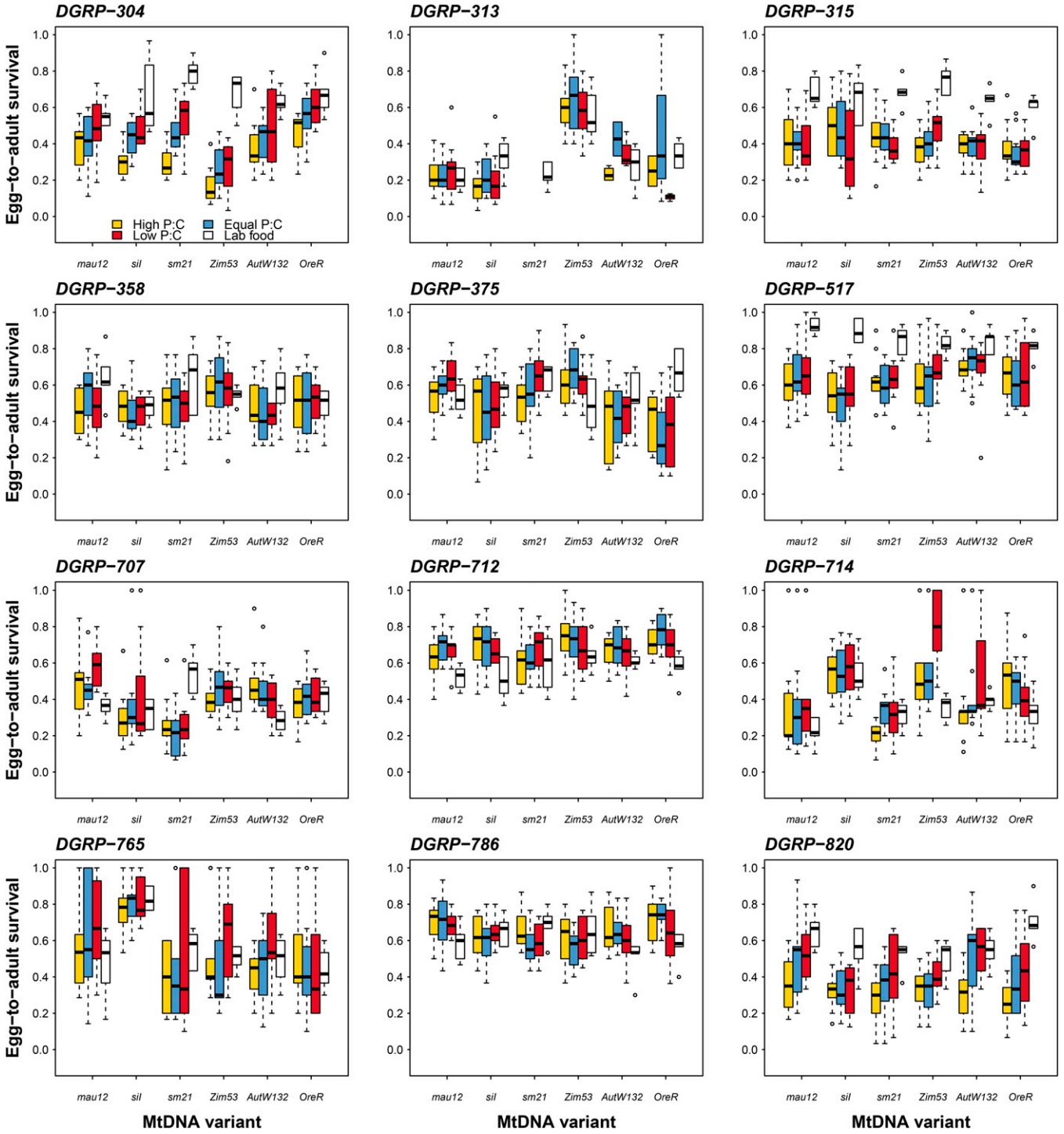


Figure 4 Phenotypic distributions of egg-to-adult viability (survival). The layout is the same as in Figure 3 and the proportions of eggs that survive to adult fly are shown on each ordinal axis.

mtDNA × food type interaction ($F = 0.78$, $P = 0.78$). The $G \times G \times E$ term was nonsignificant ($F = 1.10$, $P = 0.20$).

Male–female differences in development time ($DT_{MALE} - DT_{FEMALE}$)

We tested the Frank and Hurst hypothesis (Frank and Hurst 1996) that mtDNA effects would be more prevalent in males than in females. There is sexual dimorphism for development time in *D. melanogaster* (Bainbridge and Bownes 1981) and this is suggested to affect the expression of male-specific deleterious mtDNA mutation loads (Rand *et al.* 2001, 2004; Gemmell *et al.* 2004; Innocenti *et al.* 2011).

Drosophila females generally develop faster than males [*e.g.*, puparium formation to eclosion ~3.7–4% slower in male

D. melanogaster Oregon R strain (Powsner 1935; Bainbridge and Bownes 1981) and in *D. pseudoobscura* ~4% (Anderson 1966)], although there is overlap between the sexes in a given vial. Figure S2 shows interaction plots of male and female development times across all four food types. The lines are generally parallel and demonstrate that in almost all cases, the mean male development time was slower than the mean female development time. The magnitude of this difference was estimated for (i) all individuals across all vials and (ii) the fastest individual of each sex across all vials. For the global estimate (i) males developed on average 2.07% slower than females from egg to adult; for the fastest individual of each sex (ii) males developed 2.09% slower than females. We next tested whether there was a correlation between male and female development times across all mtDNA × nuclear background × food types. For this analysis we used the mean male and mean female development times per individual vial. This analysis is therefore restricted to vials that produced both male and female offspring. There is a significant positive correlation between male and female development times ($r = 0.90$, $t = 103.14$, d.f. = 2397, $P < 2.2e\text{-}16$; Figure 5A). In the majority of vials, males developed slower than females (data above the dotted line of equality; Figure 5A). We next asked whether the difference between males and females, $DT_{MALE} - DT_{FEMALE}$, was related to the mean value of the average (male + female) development times. This plot type (Bland and Altman 1986) demonstrates that there is no significant relationship between the mean and the difference ($r = 0.02$, $t = 0.85$, d.f. = 2397, $P = 0.40$; Figure 5B), suggesting the male–female dimorphism in development time scales isometrically (as the mean development time increases, females do not develop relatively faster than males).

There was a significant effect of nuclear background on the $DT_{MALE} - DT_{FEMALE}$ estimates (Table 2D: $F = 17.73$, $P < 2.2e\text{-}16$). The nuclear background with the greatest $DT_{MALE} - DT_{FEMALE}$ estimate across all mtDNA and food treatments was *DGRP-358*, while *DGRP-315* demonstrated the smallest difference. There was no main effect of mtDNA type ($F = 1.03$, $P = 0.40$), food type ($F = 0.66$, $P = 0.57$), or vial productivity ($F = 0.19$, $P = 0.66$) on the $DT_{MALE} - DT_{FEMALE}$ estimates. There was no significant mtDNA × nuclear background ($G \times G$) effect ($F = 1.25$, $P = 0.12$). On some backgrounds, *e.g.*, *DGRP-517*, there was very little mtDNA variation (Figure S3). In contrast, mtDNA variation conferred measureable differences in the *DGRP-765* nuclear background. Overall, the $DT_{MALE} - DT_{FEMALE}$ estimates were above the grand mean as often as they were below and there was no evidence that males were specifically expressing harmful mtDNA variation in the form of the $DT_{MALE} - DT_{FEMALE}$ estimate. We detected a significant nuclear background × food type effect ($F = 3.00$, $P = 4.53e\text{-}08$) and the $DT_{MALE} - DT_{FEMALE}$ estimate revealed that different food types conferred different effects in different nuclear backgrounds. There was no mtDNA × food type interaction ($F = 1.15$, $P = 0.30$). The higher-order $G \times G \times E$ interaction effect was significant ($F = 1.34$, $P < 0.01$) and this reflects the inconsistency of nuclear and mtDNA interactions across food types.

Coefficient of variation differences between males and females ($CV_{MALES} - CV_{FEMALES}$)

To test whether the amount of male–female variation in development times was different across nuclear, mtDNA, and food treatments, we calculated the $CV_{MALES} - CV_{FEMALES}$ of individual vials. This metric for mtDNAs provides the most stringent test of the Frank and Hurst hypothesis, with the prediction that CVmale > CVfemale. This analysis is restricted to vials that produced at least two males and at least two females. There is no relationship between mean development time of a vial and the $CV_{MALES} - CV_{FEMALES}$ ($r = 0.01$, $P = 0.54$; Figure 5C). There were no significant nuclear ($F = 1.62$, $P = 0.09$) or mtDNA effects ($F = 0.97$, $P = 0.44$) on $CV_{MALES} - CV_{FEMALES}$ (Table 2E). Neither food type ($F = 0.32$, $P = 0.81$) nor vial productivity ($F = 0.58$, $P = 0.45$) was associated with $CV_{MALES} - CV_{FEMALES}$. Nuclear × mtDNA effects were significant ($F = 1.45$, $P = 0.02$), along with nuclear × food type ($F = 2.58$, $P = 2.43e\text{-}06$) interactions. mtDNA × food type interactions were nonsignificant ($F = 1.46$, $P = 0.11$). The $G \times G \times E$ interaction term was marginally nonsignificant ($F = 1.17$, $P = 0.08$).

We further asked whether there were correlations between male and female coefficients of variation for development time for the terms in the model. In a stringent test of the Frank and Hurst hypothesis, we found significant positive correlations between CV_{MALES} and $CV_{FEMALES}$, when means were calculated based on nuclear background ($r = 0.97$, d.f. = 10, $t = 13.52$, $P = 9.43e\text{-}08$; Figure 6A), mtDNA variation ($r = 0.86$, d.f. = 4, $t = 3.35$, $P = 0.03$; Figure 6B), DGRP background × mtDNA ($r = 0.93$, d.f. = 70, $t = 20.44$, $P = <2.2e\text{-}16$; Figure 6D), DGRP background × food type ($r = 0.84$, d.f. = 46, $t = 10.46$, $P = 9.64e\text{-}14$; Figure 6E), mtDNA × food type ($r = 0.73$, d.f. = 22, $t = 5.05$, $P = 4.62e\text{-}05$; Figure 6F), and the three-way interaction ($r = 0.34$, d.f. = 270, $t = 5.97$, $P = 7.29e\text{-}09$; Figure 6G). The correlation between CV_{MALES} and $CV_{FEMALES}$ was borderline significant when samples were averaged across the four food types ($r = 0.93$, d.f. = 2, $t = 3.58$, $P = 0.07$; Figure 6C).The global comparison across individual vials is shown in Figure 6H ($r = 0.10$, d.f. = 2397, $t = 5.10$, $P = 3.64e\text{-}07$), evidencing a symmetrical pattern about the line of equality. In contrast to the predictions of the Frank and Hurst hypothesis, there were as many female biases in coefficients of variation as there were male biases. These data suggest that females are as sensitive to mtDNA variation as males and in some cases show more variability in development time (Figure 6B). The point estimates in Figure 6 are centroids of the data averaged over other terms in the model; *e.g.*, nuclear estimates in Figure 6A are averaged over all mtDNA types and food types. In the correlation between male and female coefficients of variation across food types (Figure 6C), there was a tendency for lower protein foods with slower development to have

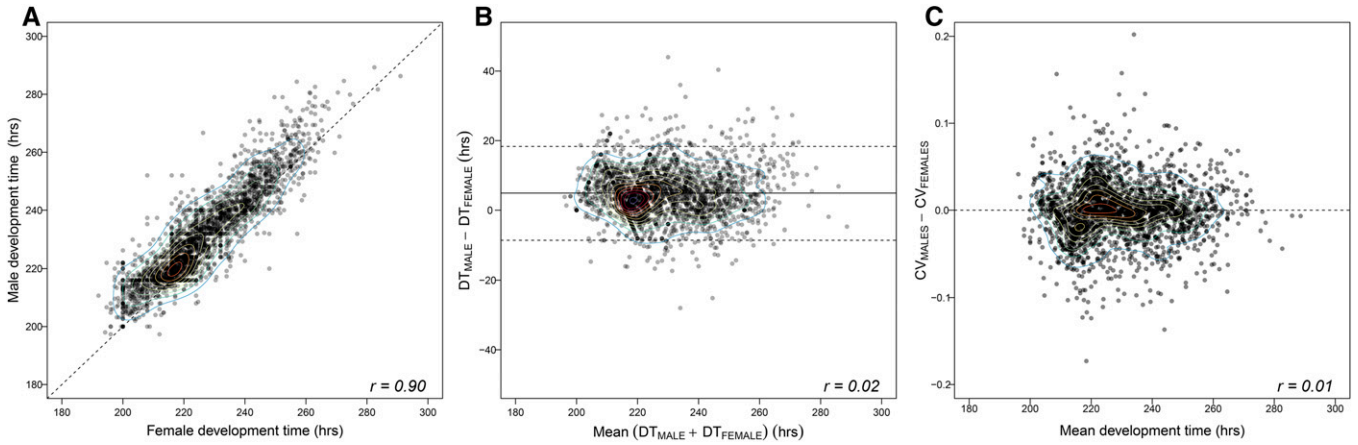


Figure 5 Sex differences in development time across all vials in the mitonuclear genotype panel. (A) A biplot of male development time against female development time for individual vials. There is a strong positive correlation between male and female development times ($r = 0.9$), and in most vials, males developed slower than females and are generally above the dashed line of equality. (B) A Bland and Altman plot shows that the difference between male and female development times (ordinal) is not correlated with the mean of males and females (abscissa) ($r = 0.02$). The dashed horizontal lines are the mean ± 1.96 SD, and the solid horizontal line is the mean. (C) A biplot of the coefficient of variation difference between males and females reveals no correlation ($r = 0.01$) between the mean vial development time (abscissa) and the coefficient of variation difference ($CV_{MALES} - CV_{FEMALES}$). Contours show the kernel density estimation of the data.

greater variation in males; the rank order of variation was as follows: (equal P:C < high P:C < low: P:C < laboratory food). Interestingly, there were two distinct clusters, with the high P:C and equal P:C clustering together and likewise for low P:C and the laboratory food diets. These data are consistent with a protein threshold effect on development time and its accompanying variation. There was a tendency for high protein foods to confer relatively high development time variation in females, yet the opposite effect was seen in low protein foods; males showed greater variation (Figure 6C).

There is an approximately twofold difference in the coefficient of variation for development time between *DGRP-517* (low) and *DGRP-358* (high) backgrounds (Figure 6A). The majority of the data show female-biased estimates suggesting females have, on average, greater variation in development time.

We tested whether there were differences between males and females in gene expression and the degree of phenotypic variation in development time [coefficient of variation (CV)], using Affymetrix expression data (Ayroles *et al.* 2009). We grouped the high CV DGRP backgrounds together (*DGRP-358*, *DGRP-707*, *DGRP-712*, *DGRP-714*) and the low CV DGRP backgrounds together (*DGRP-517*, *DGRP-315*, *DGRP-786*, *DGRP-375*) based on their distribution in Figure 6A. We readily acknowledge that the genotypes of the mtDNA-introgressed DGRP strains we have constructed are different from those reported in Ayroles *et al.* (2009), but sought to examine a relationship nonetheless. Using the Panther Classification System (http://pantherdb.org) (Thomas *et al.* 2003) and filtering for *D. melanogaster*, we downloaded Gene Ontology (GO) gene lists of biological processes, including "apoptotic process," "biological adhesion," "biological regulation," "cell killing," "cellular component organization or biogenesis," "cellular process," "developmental process," "immune system process," "localization," "locomotion," "metabolic process," "multicellular organismal process," "reproduction," and "response to stimulus." We also downloaded the gene set list for "mitochondrion" in the cellular component subset and a random selection of 200 genes in the Affymetrix array (Ayroles *et al.* 2009). For each GO category gene list we calculated the coefficient of variation in expression for each of the eight genotypes in both sexes. Each gene list had four gene expression CV estimates per phenotypic group per sex. We found a striking difference between the sexes in the degree of gene expression variation and phenotypic variation across the gene lists. In males, coefficients of variation in gene expression were consistently higher in the high phenotypic variation group than in the low variation group. In contrast, females in the low phenotypic variation group had consistently greater variation in gene expression (Figure S4A).

We conducted an ANOVA with sex and group (high or low) and their interaction as explanatory variables of CV in gene expression (dependent variable). Neither sex nor group was a significant first-order effect ($P > 0.05$); however, their interaction was significant ($F = 8.23$, $P = 0.005$), suggesting the effect of "group" was in significantly different directions in the different sexes (see Figure S4). We next calculated a mean point estimate of the four DGRPs in each GO category in each of the "high" and "low" phenotypic variance groups. In a biplot of the high and low gene expression variances, the sexual differences can be observed, with the female data generally lying above the line of equality and the male data generally below the line of equality (Figure 7A). Interestingly, a male datum lies above the line of equality, corresponding to a different direction of effect in the mitochondrion gene category (Figure 7B). In the "random" gene category, gene expression in females was

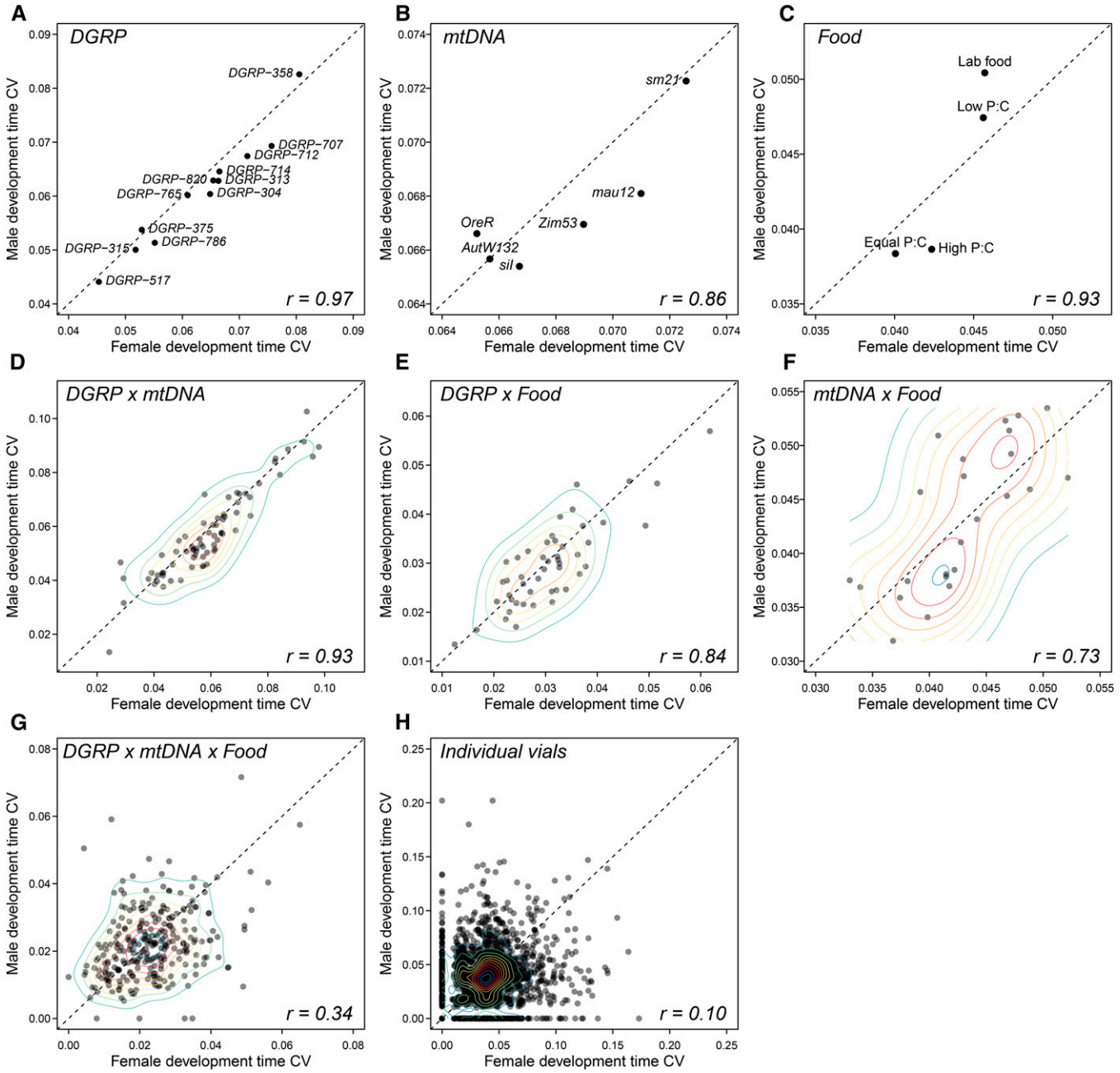


Figure 6 Phenotypic variance of development time for model terms. Coefficients of variation of males against females are shown for (A) DGRP background, (B) mtDNA haplotype, (C) food type, (D) DGRP ID × mtDNA ID, (E) DGRP ID × food type, (F) mtDNA haplotype × food type, and (G) DGRP ID × mtDNA ID × food type. CV_{MALES} against $CV_{FEMALES}$ for all individual vials is shown in H. In all panels the dashed line represents equality and deviations from equality show a sex bias in phenotypic coefficients of variation. The data in each plot represent centroids of the model variables (*e.g.*, the DGRP background centroids in A are averaged over all mtDNA types and all food types; mtDNA variants in B are averaged over all DGRP backgrounds and food types). Correlation coefficients are shown. A subset of DGRP backgrounds shown in A was selected based on high and low phenotypic variances (see main text for details). Contours show the kernel density estimation of the data.

no different between the high and low phenotypic variance groups.

DGRP phenotype and gene expression correlations

We investigated the relationship between development time in this investigation and that in previously published data sets, using DGRP lines. We have generally low power to detect correlations between nuclear backgrounds and other phenotypes since we used 12 DGRP backgrounds. In spite of this, we were first able to verify that development time in our flies corresponded well to the only available development time data (female pupation time at 25°: $r = 0.99$, $t = 19.52$, $P = 0.003$; Figure S5A) and female eclosion time ($r = 0.99$, $t = 8.23$, $P = 0.014$; Figure S5B) (Ellis *et al.* 2014). There was no

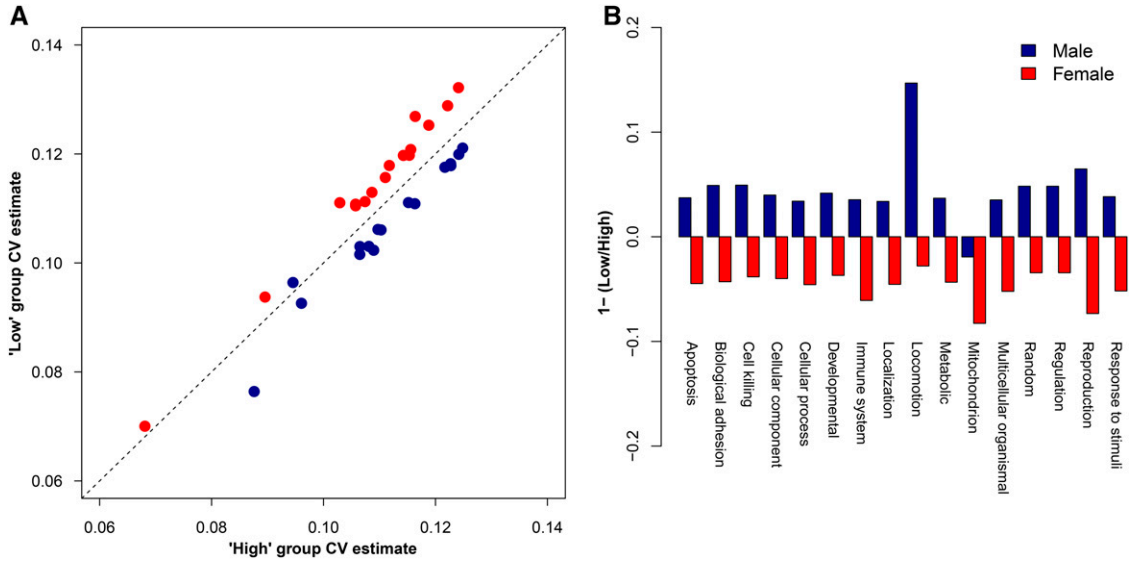


Figure 7 Phenotypic and gene expression coefficients of variation. DGRP backgrounds from the high and low phenotypic variance groups (identified in Figure 6A) were estimated for gene expression variance (CV) across 15 GO categories and a random set of 200 genes. (A) Plot of the high *vs.* the low CV for each category. Female (red) and male (blue) gene expression data are shown. Females are generally above the line of equality and males below, suggesting the relationship between phenotypic and gene expression variation is not uniform across the sexes. (B) The sex differences can be observed across the labeled GO categories (abscissa). The ordinal axis is 1 − the ratio of low:high gene expression CV. In only one GO term (mitochondrion) did the males show a qualitatively similar pattern to that of females.

correlation between development time and starvation resistance (females, $r = 0.35$, $t = 1.19$, $P = 0.26$; males, $r = -0.12$, $t = -0.39$, $P = 0.70$; Figure S5C), chill coma recovery (females, $r = -0.22$, $t = -0.703$, $P = 0.49$; males, $r = -0.08$, $t = -0.27$, $P = 0.80$; Figure S5D), startle response (females, $r = 0.16$, $t = 0.50$, $P = 0.63$; males, $r = 0.31$, $t = 1.04$, $P = 0.32$; Figure S5E), and life span ($r = 0.49$, $t = 1.71$, $P = 0.12$; Figure S5F).

Using Ayroles *et al.*'s (2009) microarray data we were able to identify a number of candidate genes whose expression profiles differ between nuclear backgrounds and correspond to divergent development time phenotypes (Figure S6). Of particular interest are the *DGRP-315* and *DGRP-820* lines, shown in red in Figure S6. When paired with a *siI* mtDNA variant, *DGRP-315* shows slow development compared to the other *D. simulans* mtDNAs (Figure 3). In contrast, *siI* shows accelerated development time when on the *DGRP-820* background. Of the 179 genes screened, ~20 demonstrate putative eQTL profiles (the development phenotype corresponds to the expression status of the gene or is inversely proportional). Our rationale for selecting probes with divergent expression patterns is because these are more likely to influence phenotypic variation (and are a bigger target for disruption by an incompatible mtDNA). However, we discuss why this may not be the case in the *Discussion* section, below.

Egg production

There were significant effects of nuclear background ($F = 34.07$, $P = 3.58e\text{-}24$), mtDNA ($F = 3.35$, $P < 0.01$), and their interaction ($F = 2.94$, $P = 1.06e\text{-}05$) on the number of eggs laid per female (Table 2F). This study was conducted only in the laboratory food treatment. Egg number was adjusted to the known number of females in the egg-lay plate (mean = 16.57, SD = 4.44, range 13–30).

Reciprocal crosses

We conducted reciprocal crosses to test whether development time segregating with mtDNA variation was consistent with maternal inheritance (Figure 8). In 12 of 14 reciprocal crosses, the phenotypic distribution of development times matched the maternal genotype, judged using Tukey's HSD tests. There were two DGRP backgrounds in which a reciprocal cross did not correspond with the maternal genotype (*DGRP-707* and *DGRP-712*). In the first case, one reciprocal cross showed an intermediate phenotype (within the range of the parental genotypes). In the second case, the parental genotypes did not recapitulate the phenotypic divergence in the main experiment, and one of the reciprocal crosses showed a small transgressive segregation pattern.

The landscape of G × G × E effects

We summarize the phenotypic data in column and row format as a tiled heat map. Different traits are arranged as columns, and rows with different colors represent different diets. Missing elements were imputed with the global mean of that particular assay × food combination (Figure 9). Figure 9 shows that the phenotypes are a mosaic of nuclear and mitochondrial interaction with few rows, columns, or cells showing a consistent ranking across the relative fitness scores.

Predicted mtDNA SNP effects

We identified several *in silico* SNP substitutions between *OregonR* and *siI* haplotypes that conferred putative deleterious

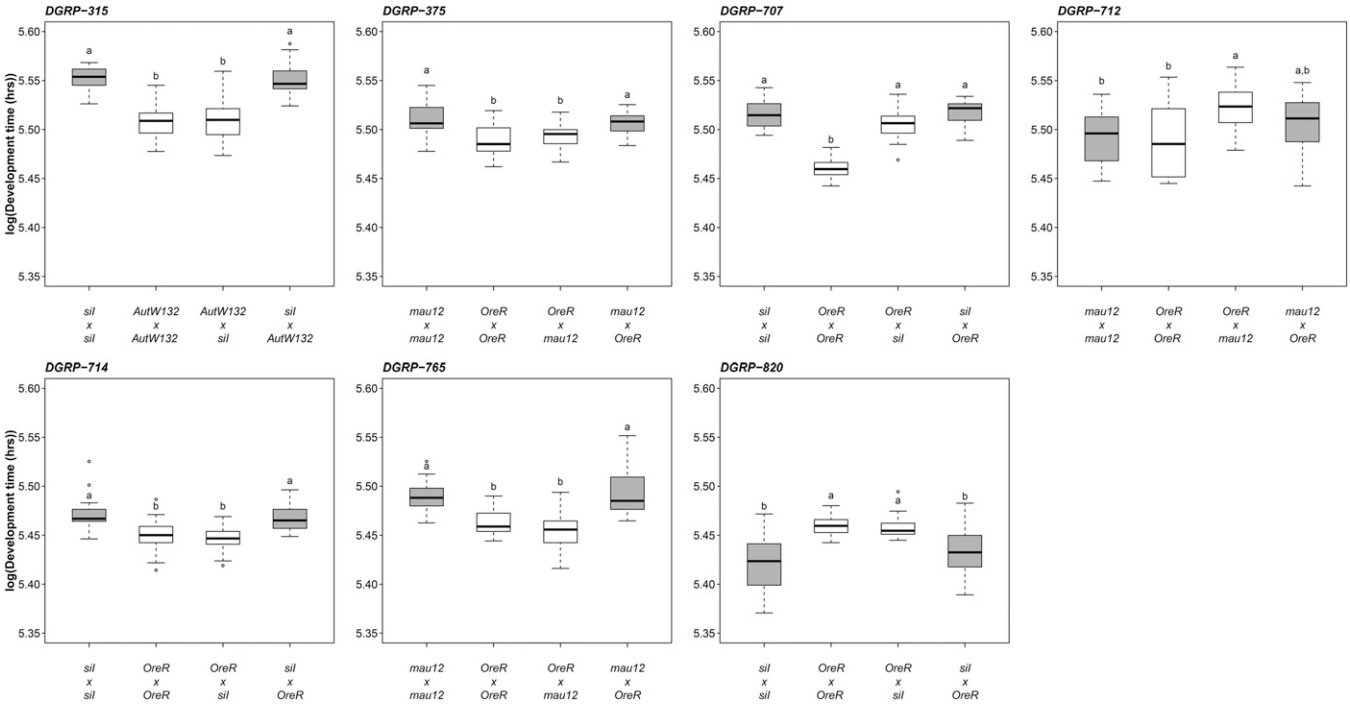


Figure 8 Offspring development time in reciprocal mtDNA crosses. Each panel corresponds to a separate DGRP background and the (female × male) cross is shown on the abscissa. The parental crosses are leftmost and the reciprocal crosses are rightmost. Boxplots that share a color have a shared maternal mtDNA haplotype. Results of Tukey's *post hoc* HSD tests are shown as letters. Boxes with the same letter are from the same statistical distribution.

effects on protein function in at least one SNP prediction software tool (Figure S1). The sites highlighted in blue in Figure S1 are nonsynonymous substitutions private to the *OregonR–siI* pairwise comparison and the remaining substitutions were shared in at least one other haplotype (see Table S1). Only two substitutions (*ND2* N313I and *ND5* S65A) conferred a consensus prediction of deleterious effect; however, neither of these were private to the *OregonR–siI* pair. SNPs identified as pathological in at least two prediction tools and private to the *OregonR–siI* pair include *ND2* I315N, *CO-II* N115S, *ATP6* L115M, *ND4* F29V, and *ND4L* V79I. The private *ND2* I315N mutation is separated by only two amino acids from a consensus deleterious substitution and therefore may interact locally, suggesting it may be an informative focal mutation for follow-up studies.

Discussion

This study found extensive evidence of mitonuclear epistatic interactions in *Drosophila*, when mtDNAs from distinct haplotypes were placed alongside controlled nuclear genetic backgrounds from the DGRP. Moreover, these $G \times G$ epistatic effects were influenced by food type, showing good support for pervasive $G \times G \times E$ interactions. Development time was a primary focus of this study and we also found similar higher-order interaction effects on egg-to-adult viability and development time differences between males and females. Importantly, there was no evidence that males suffered greater mutation burdens from mtDNA substitution than females, nor was there evidence for higher variance among mtDNA genotypes in males compared to females. Thus, our data provide no support for the Frank and Hurst (1996) hypothesis, which posits that maternal inheritance of mtDNA should allow stronger selection on female-limited phenotypes while permitting deleterious male-limited phenotypes to persist. We discuss our results in the context of epistasis, canalization, and the future avenues of our research program to identify the genes and mutations underpinning $G \times G \times E$ effects.

G × G and G × E effects

Without exception, all DGRP nuclear backgrounds show some influence of mtDNA or food interactions. However, the degree to which development time, for example, is affected varies considerably across nuclear genetic backgrounds. In some cases mtDNA variation has a large effect within a nuclear background, with different mtDNAs accelerating or delaying development (Figure 3, *e.g.*, *DGRP-765* and *DGRP-820*). Conversely, some nuclear backgrounds are canalized to mtDNA variation and appear to suppress the effect of alternative mtDNA (Figure 3, *e.g.*, *DGRP-517* and *DGRP-786*). The dietary environment had the most pronounced effect on development time in this $G \times G \times E$ experiment, with lower protein foods (laboratory food and low P:C) associated with slower development times than higher protein foods, consistent with a previous study that used the same nutritional

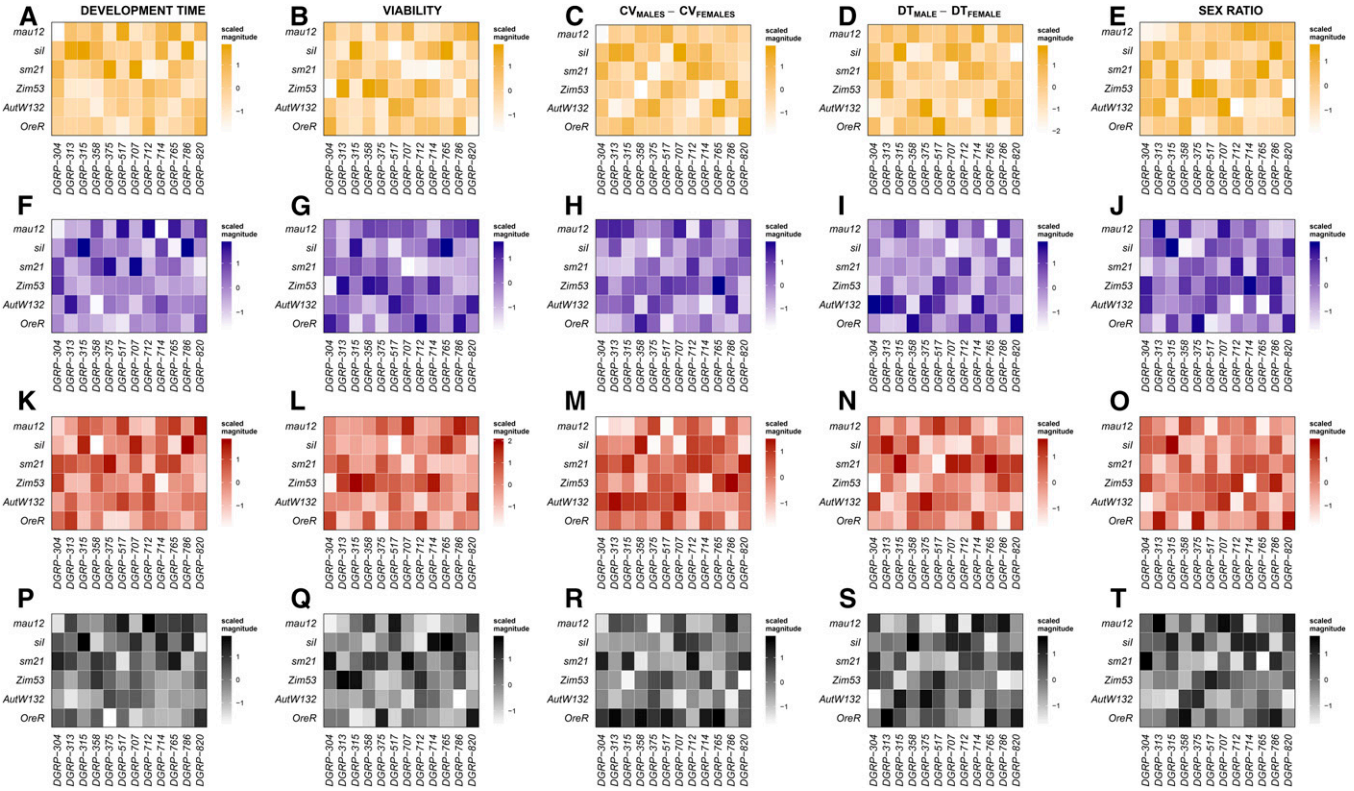


Figure 9 The $G \times G \times E$ fitness landscape. (A–T) Tiled heatmaps are scaled and show hot and cold tiles corresponding to high and low phenotypic values, respectively. Each column of panels represents a different phenotype in each of four food types (high P:C, yellow; equal P:C, blue; low P:C, red; and laboratory food, black). Missing elements were imputed as the global mean for the phenotype × food combination. Epistatic interactions are represented by hot and cold spots.

geometry foods (Matzkin *et al.* 2011). We also detected considerable conditional neutrality whereby mtDNA × nuclear interactions were more prevalent in the laboratory food than in the higher P:C food types. This form of cryptic genetic variation is hypothesized to be exposed by novel or stressful environments (Waddington 1957; Schlichting 2008), although differentiating between these scenarios is a contentious issue (Paaby and Rockman 2014). Waddington first proposed that organisms exposed to novel environments can reveal cryptic genetic variation because the "buffering" between phenotypic capacitors (genes that conceal the effects of mutations at other loci) breaks down (Waddington 1957). More recent work has pinpointed capacitors that are mediated by environment and genetic background in *Drosophila* [*e.g.*, temperature and the Hsp90 heat-shock protein (Rutherford and Lindquist 1998), genetic background effects on wing shape (Takahashi 2013), and sensory bristle traits (Takahashi 2015)]. Here, we have demonstrated that in some cases exposure of mtDNA and nuclear effects is context specific as evidenced by significant mtDNA × food type and DGRP ID × food type interactions. One interesting question arises: Are the mtDNAs or nuclear backgrounds buffering each other and to what extent could this coevolved mitonuclear system break down in low protein environments? Our results support a model of mitochondrial DNA as a capacitor that is sensitive to environment (in this case diet). Since our laboratory food is not isocaloric with the other diets, but segregates phenotypically with the low P:C food type, we favor the suggestion that protein concentration is the nutrient component most closely linked with development time and its sensitivity to mitonuclear interactions.

In line with previous studies (Montooth *et al.* 2010), there are at best only mild effects of mtDNA mutations that are fixed between *D. melanogaster* and *D. simulans* haplotypes (*e.g.*, when expressed in *DGRP-714* and *DGRP-765* in the laboratory food environment) in spite of considerable numbers of fixed variants between the species' mtDNAs (Table S1). Instead, we found greater numbers of larger effect epistases that were associated with specific haplotypes, suggesting the private SNPs in those haplotypes were driving the majority of mitonuclear epistases.

How can we hunt down "causative" SNPs? In an attempt to highlight some of these mutations for future fine mapping in the mtDNA, we focused on a haplotype pair (*Oregon R* and *siI*) that demonstrated divergent effects in the *DGRP-820* background. By substituting the *siI* amino acids into the *Oregon R* haplotype *in silico*, we identified a number of mutations that had putative significant effects on protein function (Figure S1). The majority of mutations had a suggested neutral effect. Could these mutations that change protein function act as positive epistatic factors in the *DGRP-820* background yet act as negative epistatic factors in the

DGRP-315 background? In the same analysis we identified two amino acid polymorphisms in mtDNAs that segregate with species (ND2 N313I and ND5 S65A) and show a consensus deleterious signal. We also identified one mutation (*ND2* I315N) that was private to the *OregonR–siI* pairwise comparison and only two amino acids downstream from a consensus deleterious substitution, providing opportunity for a within-protein interaction that is in complete linkage disequilibrium and specific to the *OregonR–siI* pair. Future work should aim to understand whether these fixed "pathological" mutations are compensated by nuclear variants in those DGRPs that do not highlight mtDNA species effects. For fine-scale mapping purposes, overcoming the complete linkage between alleles in mtDNA molecules is challenging as there is inevitable interaction between linked alleles. Novel techniques to successfully recombine mtDNAs in *Drosophila* have recently been developed (Ma and O'Farrell 2015) and, along with potential applications of the CRISPR-Cas9 genome editing system in mtDNA (Jo *et al.* 2015), these represent a significant forward step to making mtDNA mapping to fine-scale resolution a possibility in the near future.

G × G × E effects

This study provides a framework for understanding the $G \times G \times E$ fitness landscape (Figure 9) and under what environmental conditions we are more likely to successfully map mitonuclear interactions. There is pervasive $G \times G \times E$ for development time in the 72-genotype panel, consistent with known $G \times E$ (Vieira *et al.* 2000) and $G \times G \times E$ (Hoekstra *et al.* 2013; Zhu *et al.* 2014) effects in *Drosophila*. In a recent study (Zhou *et al.* 2012), environment (including yeast and sugar food content) has been shown to significantly expose phenotypic plasticity in the *Drosophila* DGRP transcriptome, with ~15% of the transcriptome exhibiting phenotypic plasticity and ~85% of transcripts being environmentally canalized. Other studies have found no main effect of diet on *Drosophila* transcript variation (*e.g.*, Reed *et al.* 2014) in spite of large genotype and genotype-by-diet interactions.

Is the study of epistasis a predictive discipline?

One complication of pervasive $G \times G \times E$ is the unpredictable nature of the effects on variance as well as mean values. Our working model predicts that more phylogenetically divergent mtDNA molecules should confer the greatest effects on phenotypes when paired with a standard *D. melanogaster* nuclear background. We base this assumption on the intimate relationship between mtDNA and nuclear DNA molecules that act in concert directly via proteins in the OXPHOS complexes of the electron transport chain and indirectly when a mitochondrion communicates with the cell's nucleus (because the mitochondrion itself is jointly encoded) (Woodson and Chory 2008). We specifically designed our orthogonal experiment to test for higher-order ($G \times G \times E$) effects, with food types that were likely to affect mitochondrial function, and therefore provide a form of metabolic stress. For development time, we found food type was a major modifier of mtDNA and nuclear DNA effects, as evidenced by significant mtDNA × food type and nuclear DNA × food type interactions, respectively. Higher-order genetic effects that are context specific are a major obstacle for personalized genomic medicine when the success of a treatment may depend on the genetic background, genetic interactions, and the environment. Moreover, in "natural" heterogeneous environments, epistasis effects may be an ephemeral phenomenon and this form of cryptic genetic variation may be concealed and ultimately maintained by stabilizing selection (Hermisson and Wagner 2004).

The search for genotype–phenotype associations in human genetics has been dominated by genome-wide association studies that do not capture the complete architecture of variation, but it is now apparent that $G \times G$ and $G \times E$ effects are likely sources of this missing heritability (Eichler *et al.* 2010; Yang *et al.* 2015). The future of personal genomics therefore rests on a finer-resolution understanding of epistatic interactions and a paradigm shift in hypothesis testing firmly placing nonlinear genetic effects at center stage (Moore and Williams 2009). Only then will accurate prediction of phenotype from genotype be recognized.

We found little support for a simple mitonuclear sequence divergence model affecting fitness, contrary to another mitonuclear study in *Drosophila* (Camus *et al.* 2012), which found increased nonsynonymous mtDNA divergence was associated with increased phenotypic divergence in longevity and senescence. It is possible that in the $\sim 2.3 \pm 0.3$ million years (Li *et al.* 1999) since the *D. melanogaster* and *D. simulans* divergence the mutations that are fixed between species do not, in combination, have a sufficient effect size to be detected on a *D. melanogaster* background. This could arise if there have been compensatory substitutions in the mtDNA during the divergence of the two species haplotypes, masking transient mtDNA mutations that may exhibit phenotypic effects within species. A previous mitonuclear investigation on fitness in two independent nuclear backgrounds (*Oregon R* and *AutW132*) (Montooth *et al.* 2010) also found no mtDNA sequence divergence effect. Five of six of the mtDNA haplotypes in the present study were also tested in Montooth *et al.*'s (2010) study.

We suggest that compensatory mutations in *D. melanogaster* nuclear genomes may help buffer against the deleterious effects of mtDNA mutations more generally, and any given nuclear genome will have some inherent buffering, or compensation, for pathological mtDNA mutations. Indeed, there is good evidence in mammals that mtDNA mutations within a transfer RNA (tRNA) stem structure drive second-site compensatory mutations in the same tRNA molecule and these occur to offset the deleterious effects of the primary mutation (Kern and Kondrashov 2004). Kern and Kondrashov (2004) estimate that between 10% and 50% of tRNA mtDNA substitutions act epistatically within a tRNA molecule, suggesting the evolution of tRNAs is tightly linked with compensatory and epistatically acting mutations. To further

dissect possible mitonuclear compensation mechanisms in *Drosophila*, future work could increase the degree of molecular divergence in the mitoDGRP substitution model to include more divergent haplotypes from other *Drosophila* species that may highlight any realized threshold of deleterious mutation load.

No support for the Frank and Hurst hypothesis

Contrary to other studies in *Drosophila*, we found no support for the Frank and Hurst hypothesis (Frank and Hurst 1996). We chose to focus on development time, because it is an important fitness trait, it is sexually dimorphic in *D. melanogaster*, but it is not a sex-limited trait, so a fair comparison can be made between sexes for the impact of mtDNA variation. Contrary to the predictions of the hypothesis, there was a hint that females demonstrate greater development time variation as a consequence of mtDNA substitution (Figure 6B). We used development time coefficient of variation differences between males and females as a proxy of variation that was comparable across the sexes. In fact, we found a robust absence of any evidence that males suffered from mtDNA substitution more severely than females. Development time is not a sex-limited trait and, unlike sperm traits, can be compared between the sexes. Across multiple taxa, tests of the Frank and Hurst hypothesis have provided equivocal results. One of the first articles to test the hypothesis (Ruiz-Pesini *et al.* 2000) showed there were associations between mtDNA haplotype and sperm motility in a Spanish population of men. Other studies in different geographical regions found no association between mtDNA haplotype and human sperm traits [Portugal (Pereira *et al.* 2007) and the United Kingdom (Mossman *et al.* 2012)]. Low repeatability of sex-limited (sperm-linked) mtDNA haplotype studies is also evident in *Drosophila* [significant effects (Yee *et al.* 2013) and nonsignificant effects (Friberg and Dowling 2008)]. While prior studies may have demonstrated the Frank and Hurst effect, the study reported here examined mtDNA effects in 12 different nuclear backgrounds on four different diet conditions, which may provide a closer approximation to the variable backgrounds in which mtDNAs find themselves in natural populations. Across phenotypes and genotypes in this study, there is no evidence that mtDNA haplotypes influence the direction or magnitude of phenotypic variation differently between the sexes, but taken together with other studies, this study does suggest some phenotypes may be more sensitive to mtDNA substitution.

This is the first study to assess mitonuclear genotypes in the DGRP. Using the rich genomic resources available we were able to make associations between phenotypic variation in development time and gene expression variation. We found striking differences between the sexes in the high and low phenotypic variance groups (based on coefficient of variation of development time, see Figure 6A) for variance in gene expression: males in the high phenotypic variation group had relatively wider variation in gene expression. In contrast, females in the low phenotypic variation group had relatively high variance in gene expression. This result suggests that the sexes differ in the relationship between variability in development time and variability in gene expression, with the exception of the mitochondrion GO category. What could explain this sex difference? Sex differences in *Drosophila* gene expression are a well-established phenomenon (Gibson *et al.* 2004) and extensively evident in the wild-type DGRP (Ayroles *et al.* 2009). Mitonuclear interactions have also been shown to modify gene expression in a largely sex-specific manner with male gene expression being affected more obviously than that of females when alternative mtDNAs are compared on a single nuclear background (Innocenti *et al.* 2011). This finding is consistent with a sex-specific selective sieve, in which male-harming mutations can accumulate in mtDNAs due to the lack of opportunity for selection in males (Frank and Hurst 1996). Our data are consistent with the notion that males and females differ in expression patterns in nuclear encoded mitochondrial genes. Given the nature of the data (gene expression data were from the original DGRP lines and the male–female contrast was based on variation in development time in the current experiment) it is difficult to determine whether the low variance across genes in males is due to a limited response of gene expression or whether it is a regulated homeostatic state that differs between the sexes. Resolving this pattern will require additional experiments in gene expression in the appropriate mitoDGRP genotypes. While the results in the present study do not provide support for the Frank and Hurst hypothesis, there is intriguing evidence that studies of mitonuclear variation can reveal important, unrecognized differences in sex-specific effects of mtDNA haplotypes. Whether these findings are robust across multiple nuclear backgrounds is unknown and such studies are now underway with the mitoDGRP panel.

In summary, we have developed a mitonuclear epistasis model, utilizing the DGRP resource and specifically tested for three-way ($G \times G \times E$) interaction effects. We found evidence for pervasive mitonuclear effects and three-way interactions for development time in the panel and suggest that mitochondrial DNAs (and the genes they contain) act as capacitors to release cryptic genetic variation in alternative environments. We have uncovered interesting genotypes that will be the focus of future quantitative genetic mapping and identified a sexual dichotomy in the relationship between phenotypic and genotypic expression. A striking outcome of these studies has been the unpredictable fitness consequences of different mitonuclear combinations. Efforts to test the vigor of mitonuclear combinations in human cell lines from diverse backgrounds could provide important information about genetic risk factors associated with mitochondrial replacement therapies in humans.

Acknowledgments

We thank a number of graduate students and undergraduates for logistical help during the experiment: M. Holmbeck,

A. Spierer, J. Santiago, C. Hale-Phillips, D. Yoon, J. Donner, J. Dewey, R. Mabeza, and A. Pascal. F. Lemieux prepared food for the reciprocal cross experiment. We are grateful to M. Simons and S. Ramachandran for statistical advice. F. Zapata gave valuable advice on phylogenetic analyses. The manuscript was improved with comments from D. Presgraves and two anonymous reviewers. This project was funded by National Institutes of Health grants R01-GM067862 and AG027849 (to D.M.R.). The authors declare no conflict of interest.

Author contributions: J.A.M. and D.M.R. designed the study with logistical input from L.M.B. L.M.B. prepared the cultures and organized experimental setup. Reciprocal crosses were performed by J.A.M. J.A.M. analyzed the data. J.A.M. and D.M.R. wrote the article.

Literature Cited

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova *et al.*, 2010 A method and server for predicting damaging missense mutations. Nat. Methods 7: 248–249.

Anderson, W. W., 1966 Genetic divergence in M. Vetukhiv's experimental populations of Drosophila pseudoobscura 3. Divergence in body size. Genet. Res. 7: 255–266.

Andridge, R. R., and R. J. A. Little, 2010 A review of hot deck imputation for survey non-response. Int. Stat. Rev. 78: 40–64.

Arnqvist, G., D. K. Dowling, P. Eady, L. Gay, T. Tregenza *et al.*, 2010 Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect. Evolution 64: 3354–3363.

Ayroles, J. F., M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman *et al.*, 2009 Systems genetics of complex traits in *Drosophila melanogaster*. Nat. Genet. 41: 299–307.

Bainbridge, S. P., and M. Bownes, 1981 Staging the metamorphosis of *Drosophila melanogaster*. J. Embryol. Exp. Morphol. 66: 57–80.

Ballard, J. W. O., 2000 Comparative genomics of mitochondrial DNA in Drosophila simulans. J. Mol. Evol. 51: 64–75.

Ballard, J. W. O., and A. C. James, 2004 Differential fitness of mitochondrial DNA in perturbation cage studies correlates with global abundance and population history in Drosophila simulans. Proc. Biol. Sci. 271: 1197–1201.

Barton, K., 2015 *MuMIn: Multi-Model Inference* (R Package Version 1.15.1). Available at: http://CRAN.R-project.org/package=MuMIn.

Bates, D., M. Maechler, B. Bolker, and S. Walker, 2014 Fitting Linear Mixed-Effects Models using lme4 (R Package Version 1.1-7). Available at: http://arxiv.org/abs/1406.5823v1.

Bateson, W., 1909 *Mendel's Principles of Heredity*. Cambridge University Press, Cambridge, England.

Bendl, J., J. Stourac, O. Salanda, A. Pavelka, E. D. Wieben *et al.*, 2014 PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput. Biol. 10: e1003440.

Bland, J. M., and D. G. Altman, 1986 Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1: 307–310.

Bromberg, Y., and B. Rost, 2007 SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res. 35: 3823–3835.

Burton, R. S., and F. S. Barreto, 2012 A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Mol. Ecol. 21: 4942–4957.

Camus, M. F., D. J. Clancy, and D. K. Dowling, 2012 Mitochondria, maternal inheritance, and male aging. Curr. Biol. 22: 1717–1721.

Capriotti, E., R. Calabrese, and R. Casadio, 2006 Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 22: 2729–2734.

Carlborg, O., and C. S. Haley, 2004 Epistasis: too often neglected in complex trait studies? Nat. Rev. Genet. 5: 618–625.

Chang, C.-C., J. Rodriguez, and J. Ross, 2016 Mitochondrial–nuclear epistasis impacts fitness and mitochondrial physiology of interpopulation *Caenorhabditis briggsae* hybrids. G3 6: 209–219.

Chetverikov, S. S., M. Barker, and I. M. Lerner, 1961 On certain aspects of the evolutionary process from the standpoint of modern genetics. Proc. Am. Philos. Soc. 105: 167–195.

Clark, A. G., and E. M. S. Lyckegaard, 1988 Natural-selection with nuclear and cytoplasmic transmission. 3. Joint analysis of segregation and mtDNA in *Drosophila melanogaster*. Genetics 118: 471–481.

Corbett-Detig, R. B., J. Zhou, A. G. Clark, D. L. Hartl, and J. F. Ayroles, 2013 Genetic incompatibilities are widespread within species. Nature 504: 135–137.

D'Elia, D., D. Catalano, F. Licciulli, A. Turi, G. Tripoli *et al.*, 2006 The MitoDrome database annotates and compares the OXPHOS nuclear genes of Drosophila melanogaster, Drosophila pseudoobscura and Anopheles gambiae. Mitochondrion 6: 252–257.

Dobzhansky, T., 1937 *Genetics and the Origin of Species*. Columbia University Press, New York, NY.

Dowling, D. K., K. C. Abiega, and G. Arnqvist, 2007a Temperature-specific outcomes of cytoplasmic-nuclear interactions on egg-to-adult development time in seed beetles. Evolution 61: 194–201.

Dowling, D. K., U. Friberg, F. Hailer, and G. Arnqvist, 2007b Intergenomic epistasis for fitness: within-population interactions between cytoplasmic and nuclear genes in *Drosophila melanogaster*. Genetics 175: 235–244.

Dowling, D. K., U. Friberg, and J. Lindell, 2008 Evolutionary implications of non-neutral mitochondrial genetic variation. Trends Ecol. Evol. 23: 546–554.

Durham, M. F., M. M. Magwire, E. A. Stone, and J. Leips, 2014 Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. Nat. Commun. 5: 4338.

Eichler, E. E., J. Flint, G. Gibson, A. Kong, S. M. Leal *et al.*, 2010 Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet. 11: 446–450.

Ellis, L. L., W. Huang, A. M. Quinn, A. Ahuja, B. Alfrejd *et al.*, 2014 Intrapopulation genome size variation in *D. melanogaster* reflects life history variation and plasticity. PLoS Genet. 10: e1004522.

Faraway, J. J., 2005 *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. CRC Press, Cleveland, OH.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, England.

Fos, M., M. A. Dominguez, A. Latorre, and A. Moya, 1990 Mitochondrial DNA evolution in experimental populations of Drosophila subobscura. Proc. Natl. Acad. Sci. USA 87: 4198–4201.

Frank, S. A., and L. D. Hurst, 1996 Mitochondria and male disease. Nature 383: 224.

Friberg, U., and D. K. Dowling, 2008 No evidence of mitochondrial genetic variation for sperm competition within a population of *Drosophila melanogaster*. J. Evol. Biol. 21: 1798–1807.

Garcia-Martinez, J., J. A. Castro, M. Ramon, A. Latorre, and A. Moya, 1998 Mitochondrial DNA haplotype frequencies in natural and experimental populations of *Drosophila subobscura*. Genetics 149: 1377–1382.

Gemmell, N. J., V. J. Metcalf, and F. W. Allendorf, 2004 Mother's curse: the effect of mtDNA on individual fitness and population viability. Trends Ecol. Evol. 19: 238–244.
Gerke, J., K. Lorenz, S. Ramnarine, and B. Cohen, 2010 Gene-environment interactions at nucleotide resolution. PLoS Genet. 6: e1001144.
Gibson, G., R. Riley-Berger, L. Harshman, A. Kopp, S. Vacha *et al.*, 2004 Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. Genetics 167: 1791–1799.
Hermisson, J., and G. P. Wagner, 2004 The population genetic theory of hidden variation and genetic robustness. Genetics 168: 2271–2284.
Hillenmeyer, M. E., 2008 The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320: 362–365.
Hinkley, T., J. Martins, C. Chappey, M. Haddad, E. Stawiski *et al.*, 2011 A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. Nat. Genet. 43: 487–489.
Hoekstra, L. A., M. A. Siddiq, and K. L. Montooth, 2013 Pleiotropic effects of a mitochondrial-nuclear incompatibility depend upon the accelerating effect of temperature in *Drosophila*. Genetics 195: 1129–1139.
Holmbeck, M. A., J. R. Donner, E. Villa-Cuesta, and D. M. Rand, 2015 A *Drosophila* model for mito-nuclear diseases generated by an incompatible interaction between tRNA and tRNA synthetase. Dis. Model. Mech. 8: 843–854.
Huang, W., S. Richards, M. A. Carbone, D. H. Zhu, R. R. H. Anholt *et al.*, 2012 Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. Proc. Natl. Acad. Sci. USA 109: 15553–15559.
Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Ramia *et al.*, 2014 Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. Genome Res. 24: 1193–1208.
Hutter, C. M., and D. M. Rand, 1995 Competition between mitochondrial haplotypes in distinct nuclear genetic environments – *Drosophila pseudoobscura vs. D. persimilis*. Genetics 140: 537–548.
Innocenti, P., E. H. Morrow, and D. K. Dowling, 2011 Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. Science 332: 845–848.
Ivanov, D. K., V. Escott-Price, M. Ziehm, M. M. Magwire, T. F. C. Mackay *et al.*, 2015 Longevity GWAS using the *Drosophila* Genetic Reference Panel. J. Gerontol. A Biol. Sci. Med. Sci. 70: 1470–1478.
James, A. C., and J. W. O. Ballard, 2003 Mitochondrial genotype affects fitness in *Drosophila simulans*. Genetics 164: 187–194.
Jo, A., S. Ham, G. H. Lee, Y.-I. Lee, S. Kim *et al.*, 2015 Efficient mitochondrial genome editing by CRISPR/Cas9. BioMed Res. Int. 2015: 10.
Kern, A. D., and F. A. Kondrashov, 2004 Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. Nat. Genet. 36: 1207–1212.
Kilpatrick, S. T., and D. M. Rand, 1995 Conditional hitchhiking of mitochondrial DNA: frequency shifts of *Drosophila melanogaster* mtDNA variants depend on nuclear genetic background. Genetics 141: 1113–1124.
Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov, 2002 Dobzhansky-Muller incompatibilities in protein evolution. Proc. Natl. Acad. Sci. USA 99: 14878–14883.
Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen, 2015 *lmerTest: Tests in Linear Mixed Effects Models* (R Package Version 2.0-29). Available at: http://CRAN.R-project.org/package=lmerTest.
Lehner, B., 2011 Molecular mechanisms of epistasis within and between genes. Trends Genet. 27: 323–331.
Lehner, B., 2013 Genotype to phenotype: lessons from model organisms for human genetics. Nat. Rev. Genet. 14: 168–178.
Li, Y. J., Y. Satta, and N. Takahata, 1999 Paleo-demography of the *Drosophila melanogaster* subgroup: application of the maximum likelihood method. Genes Genet. Syst. 74: 117–127.
Li, Z. K., S. R. M. Pinson, W. D. Park, A. H. Paterson, and J. W. Stansel, 1997 Epistasis for three grain yield components in rice (*Oryza sativa* L). Genetics 145: 453–465.
Ma, H., and P. H. O'Farrell, 2015 Selections that isolate recombinant mitochondrial genomes in animals. eLife 4: 07247.
Mackay, T. F. C., 2014 Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat. Rev. Genet. 15: 22–33.
Mackay, T. F. C., and J. H. Moore, 2014 Why epistasis is important for tackling complex human disease genetics. Genome Med. 6: 124.
Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* Genetic Reference Panel. Nature 482: 173–178.
MacRae, A. F., and W. W. Anderson, 1988 Evidence for non-neutrality of mitochondrial-DNA haplotypes in *Drosophila pseudoobscura*. Genetics 120: 485–494.
Manolio, T. A., 2009 Finding the missing heritability of complex diseases. Nature 461: 747–753.
Marchini, J., P. Donnelly, and L. R. Cardon, 2005 Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat. Genet. 37: 413–417.
Matzkin, L. M., S. Johnson, C. Paight, G. Bozinovic, and T. A. Markow, 2011 Dietary protein and sugar differentially affect development and metabolic pools in ecologically diverse *Drosophila*. J. Nutr. 141: 1127–1133.
Meiklejohn, C. D., M. A. Holmbeck, M. A. Siddiq, D. N. Abt, D. M. Rand *et al.*, 2013 An incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in *Drosophila*. PLoS Genet. 9: e1003238.
Montooth, K. L., C. D. Meiklejohn, D. N. Abt, and D. M. Rand, 2010 Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of *Drosophila*. Evolution 64: 3364–3379.
Moore, J. H., and S. M. Williams, 2009 Epistasis and its implications for personal genetics. Am. J. Hum. Genet. 85: 309–320.
Mossman, J. A., J. Slate, T. R. Birkhead, H. D. Moore, and A. A. Pacey, 2012 Mitochondrial haplotype does not influence sperm motility in a UK population of men. Hum. Reprod. 27: 641–651.
Muller, H. J., 1940 *The New Systematics*, pp. 185–268. Clarendon Press, Oxford, England.
Paaby, A. B., and M. V. Rockman, 2014 Cryptic genetic variation: evolution's hidden substrate. Nat. Rev. Genet. 15: 247–258.
Paliwal, S., A. C. Fiumera, and H. L. Fiumera, 2014 Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. Genetics 198: 1251–1265.
Pereira, L., J. Goncalves, R. Franco-Duarte, J. Silva, T. Rocha *et al.*, 2007 No evidence for an mtDNA role in sperm motility: data from complete sequencing of asthenozoospermic males. Mol. Biol. Evol. 24: 868–874.
Phillips, P. C., 2008 Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. 9: 855–867.
Povolotskaya, I. S., and F. A. Kondrashov, 2010 Sequence space and the ongoing expansion of the protein universe. Nature 465: 922–926.
Powsner, L., 1935 The effects of temperature on the durations of the developmental stages of Drosophila melanogaster. Physiol. Zool. 8: 474–520.
R Core Team, 2015 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Ramensky, V., P. Bork, and S. Sunyaev, 2002 Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 30: 3894–3900.
Rand, D. M., A. G. Clark, and L. M. Kann, 2001 Sexually antagonistic cytonuclear fitness interactions in *Drosophila melanogaster*. Genetics 159: 173–187.
Rand, D. M., R. A. Haney, and A. J. Fry, 2004 Cytonuclear coevolution: the genomics of cooperation. Trends Ecol. Evol. 19: 645–653.
Rand, D. M., A. Fry, and L. Sheldahl, 2006 Nuclear-mitochondrial epistasis and *Drosophila* aging: introgression of *Drosophila simulans* mtDNA modifies longevity in *D. melanogaster* nuclear backgrounds. Genetics 172: 329–341.
Reed, L. K., S. Williams, M. Springston, J. Brown, K. Freeman *et al.*, 2010 Genotype-by-diet interactions drive metabolic phenotype variation in *Drosophila melanogaster*. Genetics 185: 1009–1019.
Reed, L. K., K. Lee, Z. Zhang, L. Rashid, A. Poe *et al.*, 2014 Systems genomics of metabolic phenotypes in wild-type *Drosophila melanogaster*. Genetics 197: 781–793.
Ruiz-Pesini, E., A. C. Lapena, C. Diez-Sanchez, A. Perez-Martos, J. Montoya *et al.*, 2000 Human mtDNA haplogroups associated with high or reduced spermatozoa motility. Am. J. Hum. Genet. 67: 682–696.
Rutherford, S. L., and S. Lindquist, 1998 Hsp90 as a capacitor for morphological evolution. Nature 396: 336–342.
Sackton, T. B., R. A. Haney, and D. M. Rand, 2003 Cytonuclear coadaptation in Drosophila: disruption of cytochrome c oxidase activity in backcross genotypes. Evolution 57: 2315–2325.
Sardiello, M., F. Licciulli, D. Catalano, M. Attimonelli, and C. Caggese, 2003 MitoDrome: a database of *Drosophila melanogaster* nuclear genes encoding proteins targeted to the mitochondrion. Nucleic Acids Res. 31: 322–324.
Schlichting, C. D., 2008 Hidden reaction norms, cryptic genetic variation, and evolvability. Ann. N. Y. Acad. Sci. 1133: 187–203.
Sim, N.-L., P. Kumar, J. Hu, S. Henikoff, G. Schneider *et al.*, 2012 SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 40: 452–457.
Stone, E. A., and A. Sidow, 2005 Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res. 15: 978–986.
Takahashi, K. H., 2013 Multiple capacitors for natural genetic variation in *Drosophila melanogaster*. Mol. Ecol. 22: 1356–1365.
Takahashi, K. H., 2015 Novel genetic capacitors and potentiators for the natural genetic variation of sensory bristles and their trait specificity in *Drosophila melanogaster*. Mol. Ecol. 24: 5561–5572.
Takano, T., S. Kusakabe, and T. Mukai, 1987 The genetic-structure of natural populations of *Drosophila melanogaster*. XX. Comparison of genotype-environment interaction in viability between a northern and a southern population. Genetics 117: 245–254.
Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Y. Mi, B. Karlak *et al.*, 2003 PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13: 2129–2141.
Tischler, J., B. Lehner, and A. G. Fraser, 2008 Evolutionary plasticity of genetic interaction networks. Nat. Genet. 40: 390–391.
Vieira, C., E. G. Pasyukova, Z. B. Zeng, J. B. Hackett, R. F. Lyman *et al.*, 2000 Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. Genetics 154: 213–227.
Villa-Cuesta, E., M. A. Holmbeck, and D. M. Rand, 2014 Rapamycin increases mitochondrial efficiency by mtDNA-dependent reprogramming of mitochondrial metabolism in *Drosophila*. J. Cell Sci. 127: 2282–2290.
Waddington, C. H., 1942 Canalization of development and the inheritance of acquired characters. Nature 150: 563–565.
Waddington, C. H., 1957 *The Strategy of the Genes*. George, Allen, and Unwin, London, England.
Wagner, G. P., and L. Altenberg, 1996 Perspective: complex adaptations and the evolution of evolvability. Evolution 50: 967–976.
Weinreich, D. M., N. F. Delaney, M. A. Depristo, and D. L. Hartl, 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312: 111–114.
Wolak, M. E., D. J. Fairbairn, and Y. R. Paulsen, 2012 Guidelines for estimating repeatability. Methods Ecol. Evol. 3: 129–137.
Woodson, J. D., and J. Chory, 2008 Coordination of gene expression between organellar and nuclear genomes. Nat. Rev. Genet. 9: 383–395.
Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97–159.
Yang, J., A. Bakshi, Z. Zhu, G. Hemani, A. A. E. Vinkhuyzen *et al.*, 2015 Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. 47: 1114–1120.
Yee, W. K. W., K. L. Sutton, and D. K. Dowling, 2013 *In vivo* male fertility is affected by naturally occurring mitochondrial haplotypes. Curr. Biol. 23: 55–56.
Zhou, S., T. G. Campbell, E. A. Stone, T. F. C. Mackay, and R. R. H. Anholt, 2012 Phenotypic plasticity of the Drosophila transcriptome. PLoS Genet. 8: e1002593.
Zhu, C.-T., P. Ingelmo, and D. M. Rand, 2014 G×G×E for lifespan in *Drosophila*: mitochondrial, nuclear, and dietary interactions that modify longevity. PLoS Genet. 10: e1004354.
Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith, 2011 *Mixed Effects Models and Extensions in Ecology with R*. Springer-Verlag, Berlin/Heidelberg, Germany.

Communicating editor: D. C. Presgraves

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.187286/-/DC1

Mitonuclear Epistasis for Development Time and Its Modification by Diet in *Drosophila*

Jim A. Mossman, Leann M. Biancani, Chen-Tseh Zhu, and David M. Rand

DOI: 10.1534/genetics.116.187286

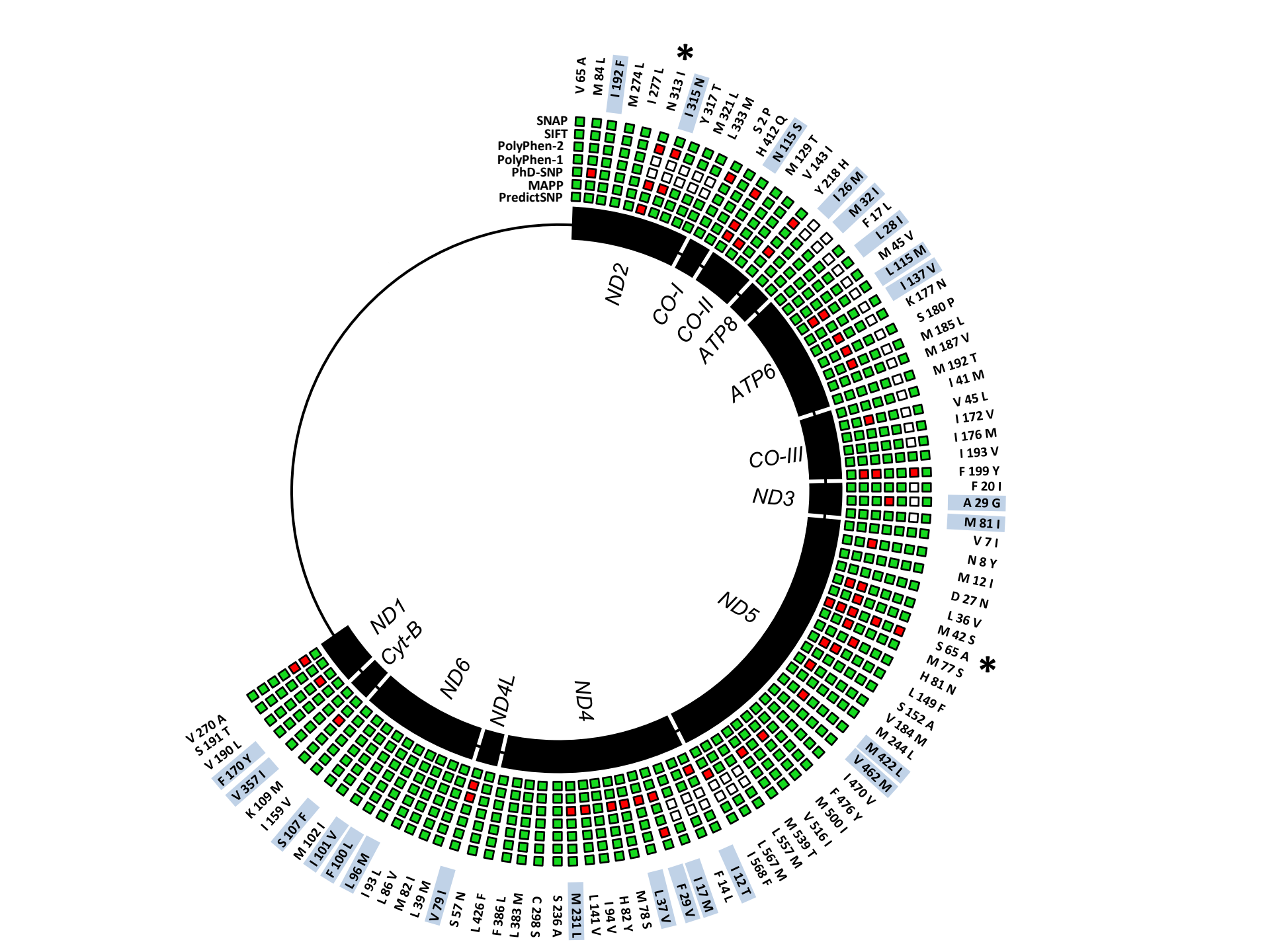

SNAP
SIFT
PolyPhen-2
PolyPhen-1
PhD-SNP
MAPP
PredictSNP
ND2
CO-I
CO-II
ATP8
ATP6
CO-III
ND3
ND5
ND4
ND4L
ND6
Cyt-B
ND1
V 65 A
M 84 L
I 192 F
M 274 L
I 277 L
N 313 I
I 315 N
Y 317 T
M 321 L
L 333 M
S 2 P
H 412 Q
N 115 S
M 129 T
V 143 I
Y 218 H
I 26 M
M 32 I
F 17 L
L 28 I
M 45 V
L 115 M
I 137 V
K 177 N
S 180 P
M 185 L
M 187 V
M 192 T
I 41 M
V 45 L
I 172 V
I 176 M
I 193 V
F 199 Y
F 20 I
A 29 G
M 81 I
V 7 I
N 8 Y
M 12 I
D 27 N
L 36 V
M 42 S
S 65 A
M 77 S
H 81 N
L 149 F
S 152 A
V 184 M
M 244 L
M 422 L
V 462 M
I 470 V
F 476 Y
M 500 I
V 516 I
M 539 T
L 557 M
L 567 M
I 568 F
I 12 T
F 14 L
I 17 M
F 29 V
L 37 V
M 78 S
I 94 V
H 82 Y
L 141 V
M 231 L
S 236 A
C 298 S
M 383 T
F 386 L
L 426 F
S 57 N
V 79 I
L 39 M
M 82 I
L 86 V
I 93 L
L 96 M
F 100 L
I 101 V
M 102 I
S 107 F
I 159 V
K 109 M
V 357 I
F 170 Y
V 190 L
S 191 T
V 270 A

Figure S1. In silico mtDNA amino acid substitutions between siI and OregonR haplotypes. Each gene is labeled on the molecule (not drawn to scale). Concentric arcs of boxes show different prediction software results of amino acid substitutions. Boxes in red show substitutions with putative deleterious effects, boxes in green are putatively neutral. Asterisks show two mutations with a consensus deleterious effect across calling softwares. Amino acid positions and substitutions are shown. Substitutions in blue boxes are private to the siI haplotype within the suite of D. simulans species mutations.

DGRP−304

DGRP−313

DGRP−315

DGRP−358

DGRP−375

DGRP−517

DGRP−707

DGRP−712

DGRP−714

DGRP−765

DGRP−786

DGRP−820

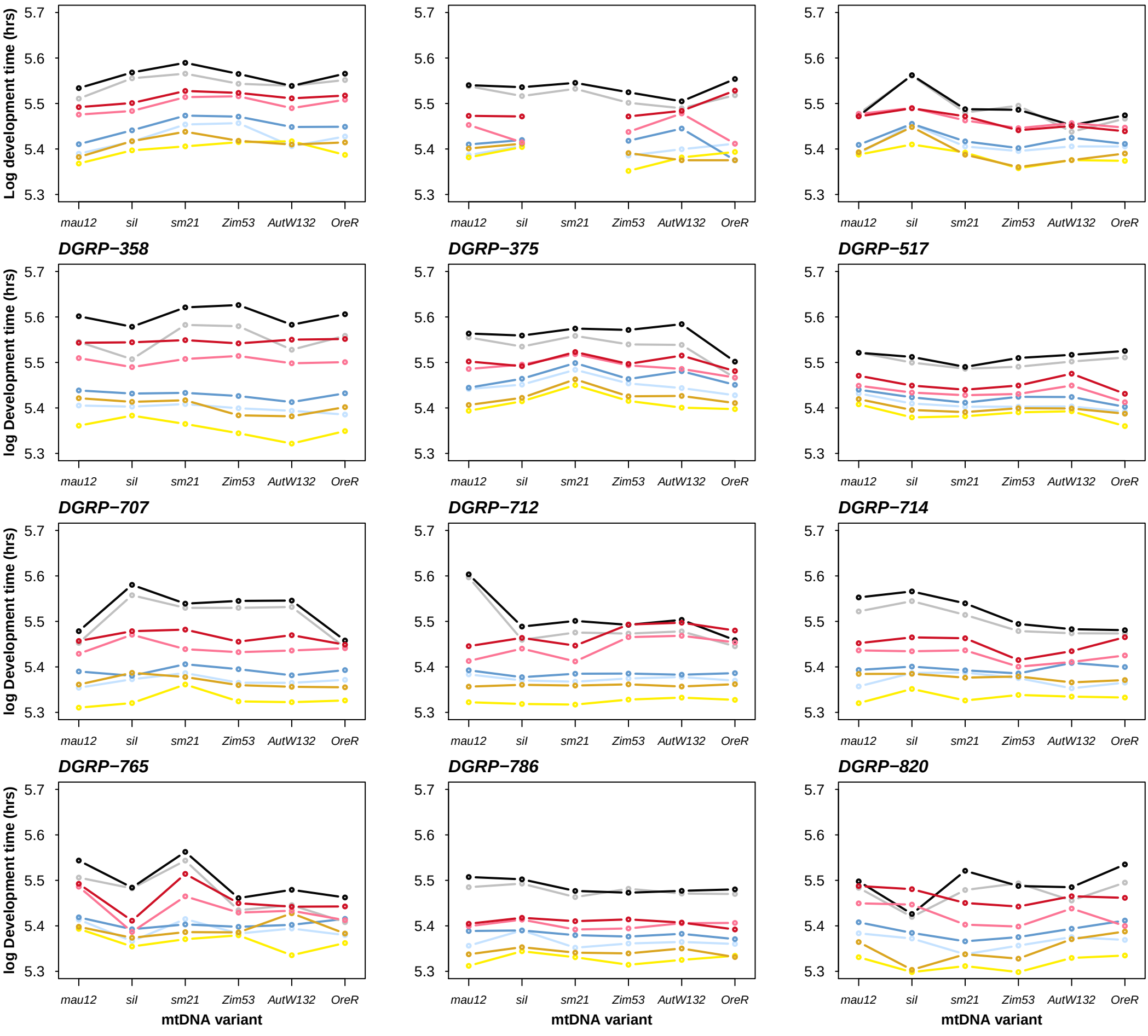


Figure S2. Male and female development time differences across DGRP backgrounds. Interaction plots are of mean estimates in each mtDNA x food combination. Lighter shades correspond to Females and darker shades, males. Parallelism between the sexes shows males and females respond to mtDNA variation in a similar way across nuclear backgrounds and food types.

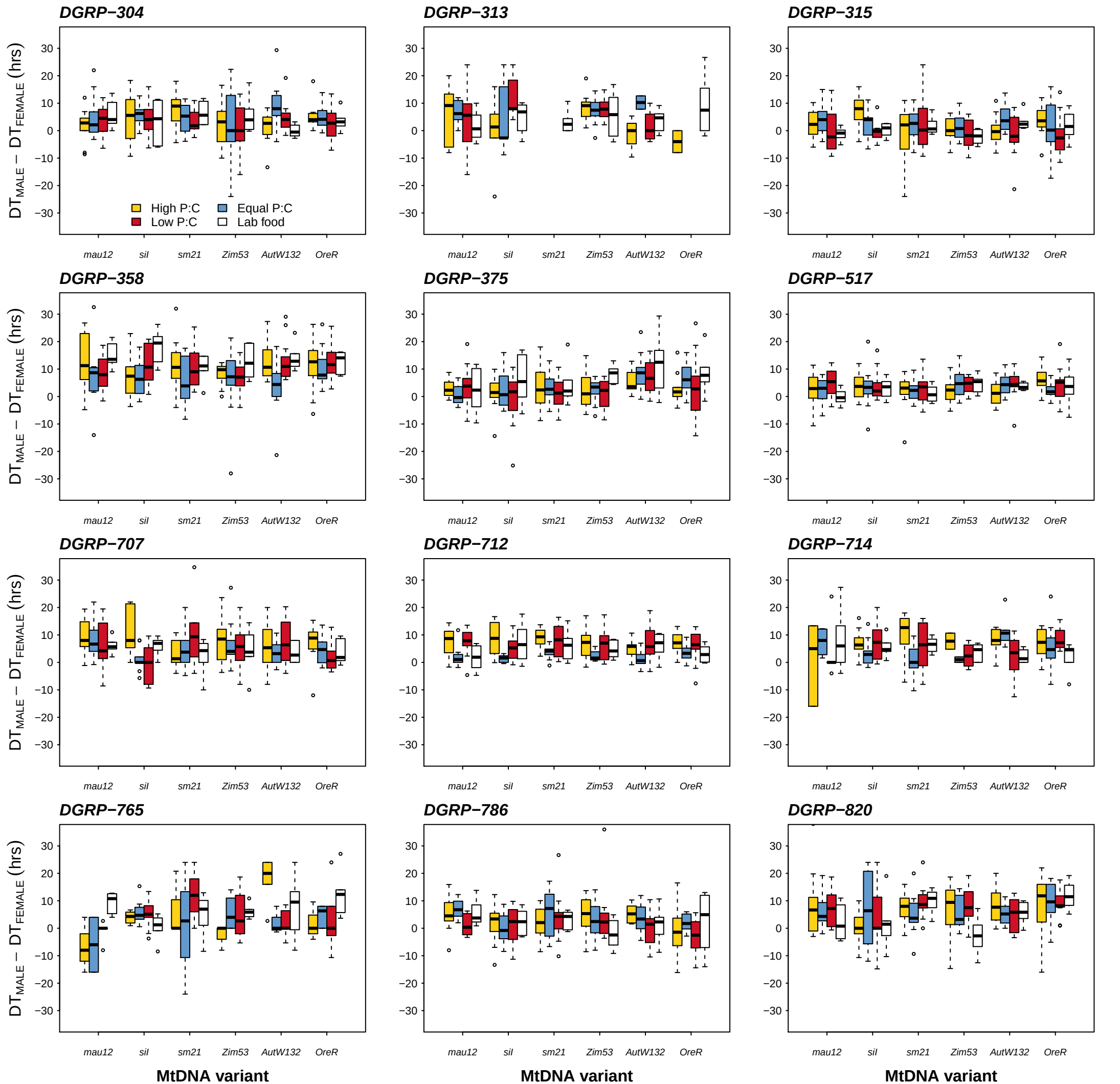

DGRP−304
DGRP−313
DGRP−315
DGRP−358
DGRP−375
DGRP−517
DGRP−707
DGRP−712
DGRP−714
DGRP−765
DGRP−786
DGRP−820
DT_MALE − DT_FEMALE (hrs)
30
20
10
0
−10
−20
−30
High P:C
Low P:C
Equal P:C
Lab food
mau12
sil
sm21
Zim53
AutW132
OreR
MtDNA variant

Figure S3. DTMALE - DTFEMALE estimates across DGRP backgrounds. MtDNA haplotypes are shown on the abscissa and the mean difference in development time between males and females is on the ordinal axis. Boxplots in yellow, blue, red and white correspond to High P:C, Equal P:C, Low P:C and Lab food types, respectively. Each panel represents an individual DGRP nuclear background.

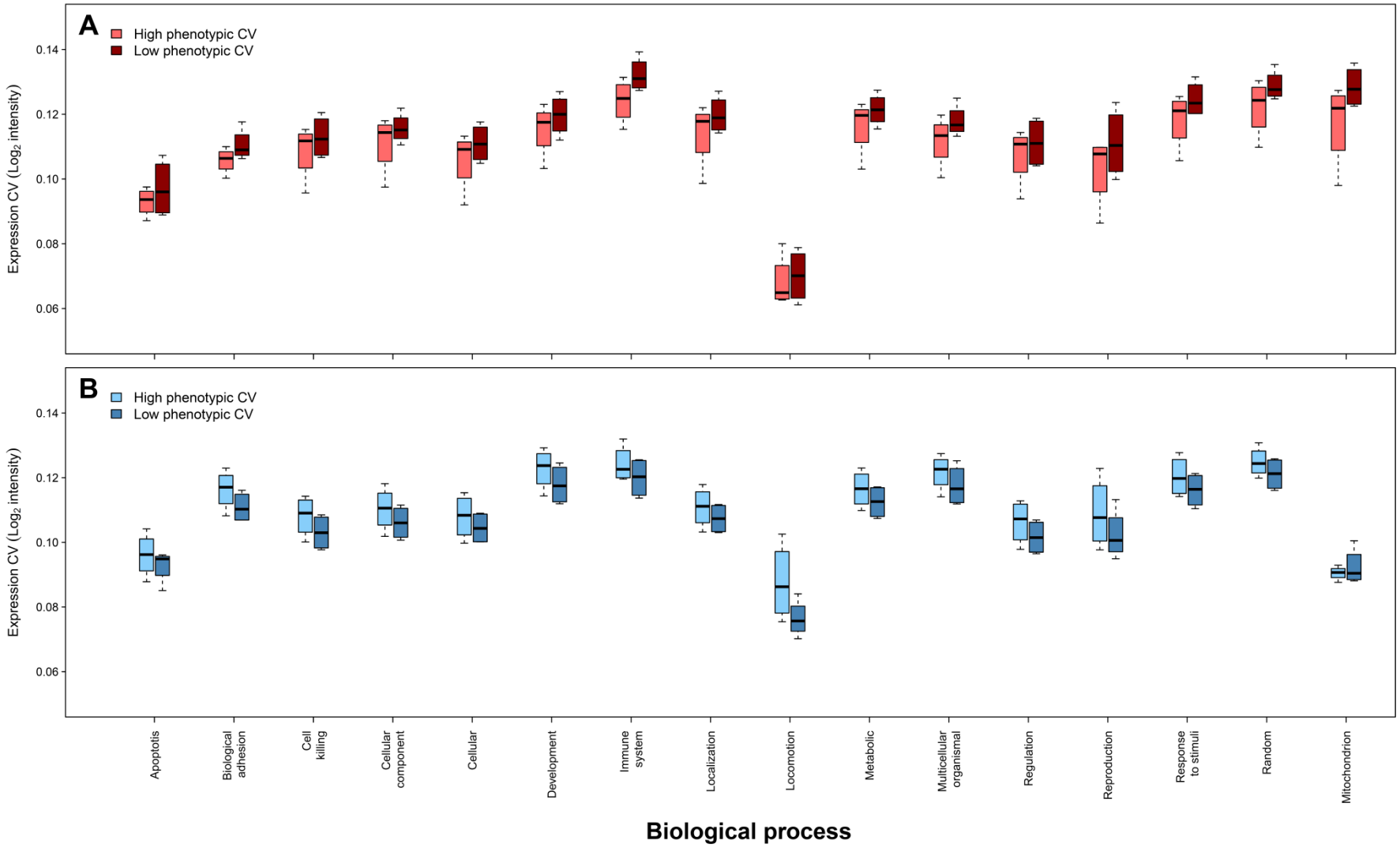
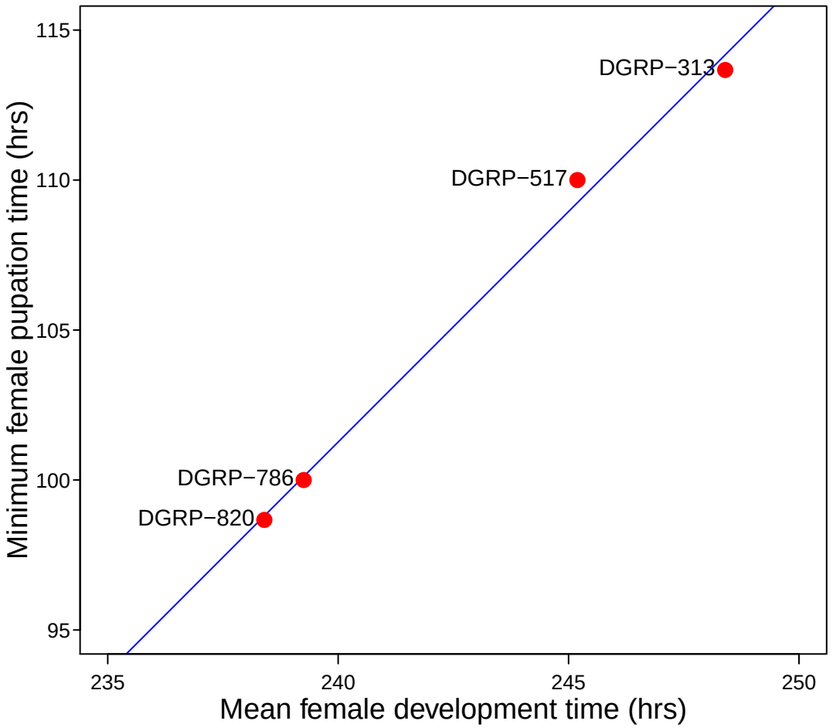
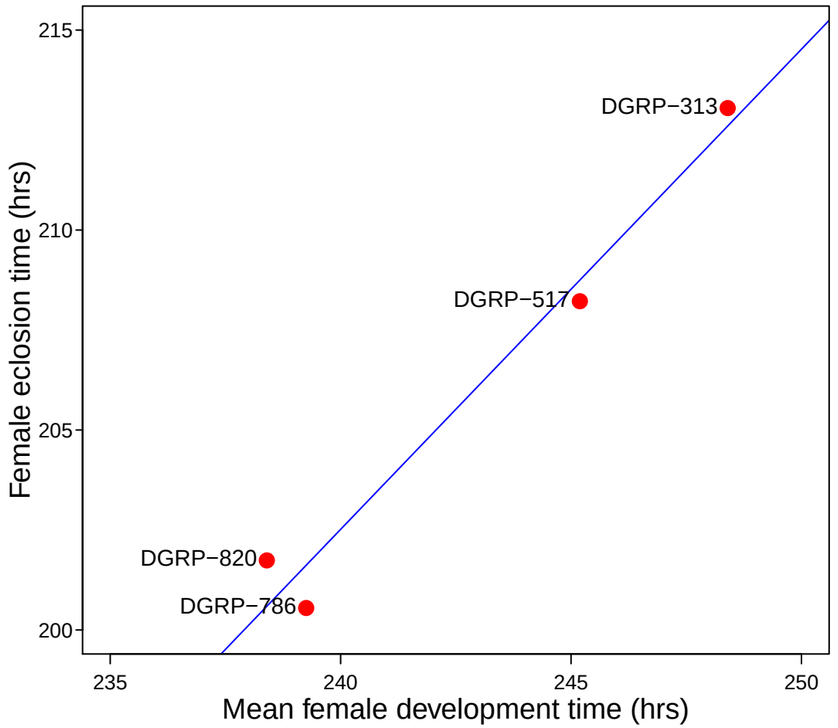

A
High phenotypic CV
Low phenotypic CV
Expression CV (Log2 intensity)
0.14
0.12
0.10
0.08
0.06
B
High phenotypic CV
Low phenotypic CV
Expression CV (Log2 intensity)
0.14
0.12
0.10
0.08
0.06
Apoptotis
Biological adhesion
Cell killing
Cellular component
Cellular
Development
Immune system
Localization
Locomotion
Metabolic
Multicellular organismal
Regulation
Reproduction
Response to stimuli
Random
Mitochondrion
Biological process

Figure S4. Male and female gene expression CV in 'high' and 'low' phenotypic variance groups. Female gene expression CV (A) and male gene expression CV (B) are shown. Lighter shades are high phenotypic variance groups and dark shades are low phenotypic variance groups. Each box represents four data, across each of 15 GO categories plus a random set of 200 genes. Males and females show opposite patterns of gene expression variation between the high and low phenotypic variance groups.

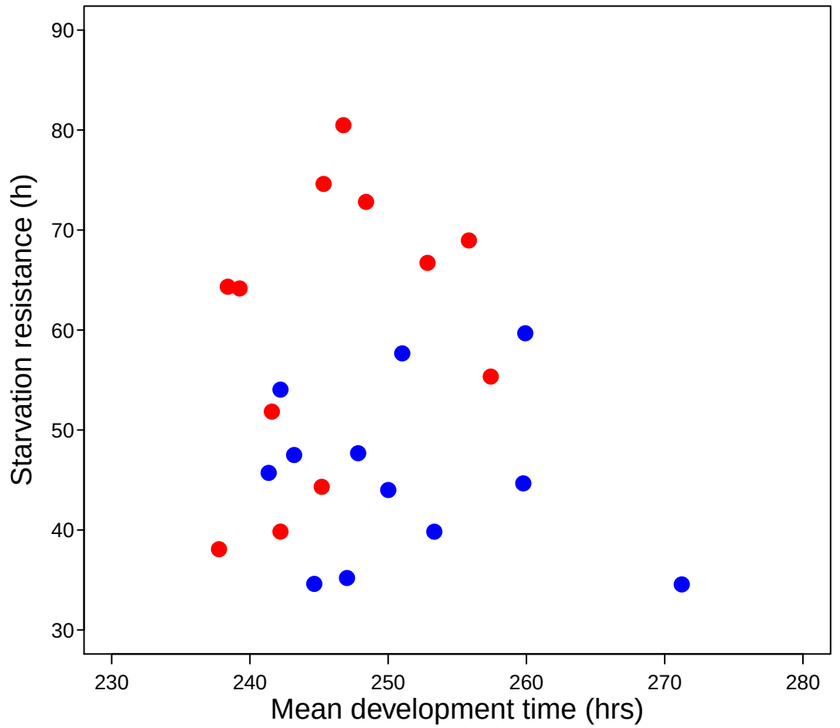
A


Minimum female pupation time (hrs)
115
110
105
100
95
DGRP−313
DGRP−517
DGRP−786
DGRP−820
235
240
245
250
Mean female development time (hrs)

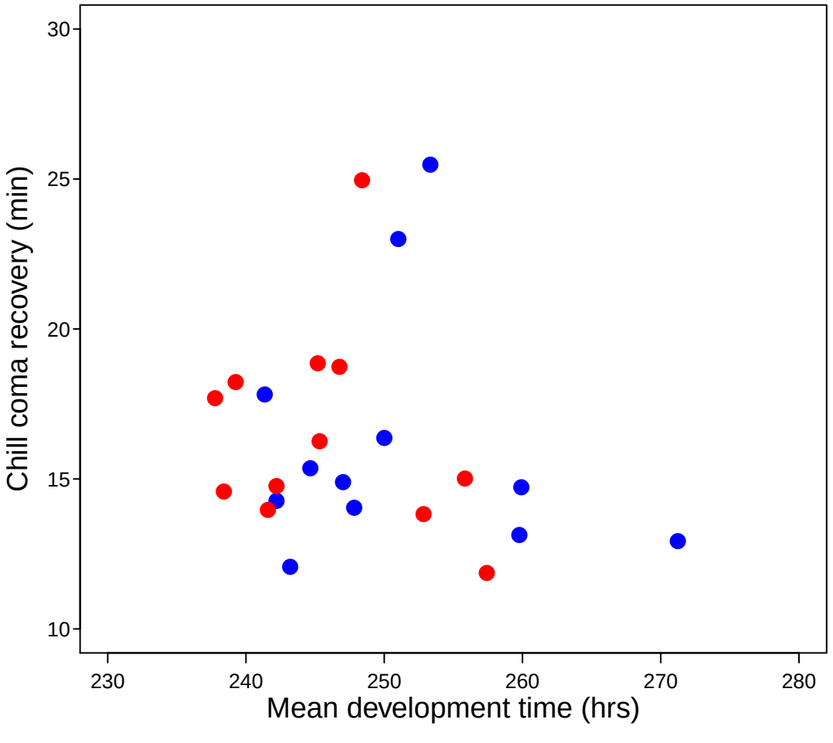
B


Female eclosion time (hrs)
215
210
205
200
DGRP−313
DGRP−517
DGRP−820
DGRP−786
235
240
245
250
Mean female development time (hrs)

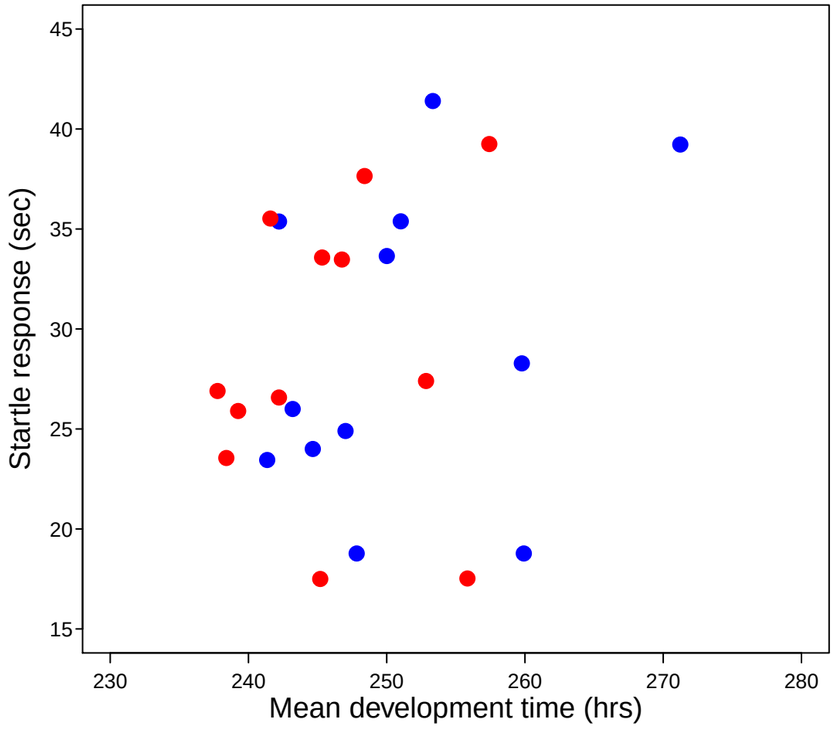
C


Starvation resistance (h)
90
80
70
60
50
40
30
230
240
250
260
270
280
Mean development time (hrs)

D


Chill coma recovery (min)
30
25
20
15
10
230
240
250
260
270
280
Mean development time (hrs)

E


Startle response (sec)
45
40
35
30
25
20
15
230
240
250
260
270
280
Mean development time (hrs)

F

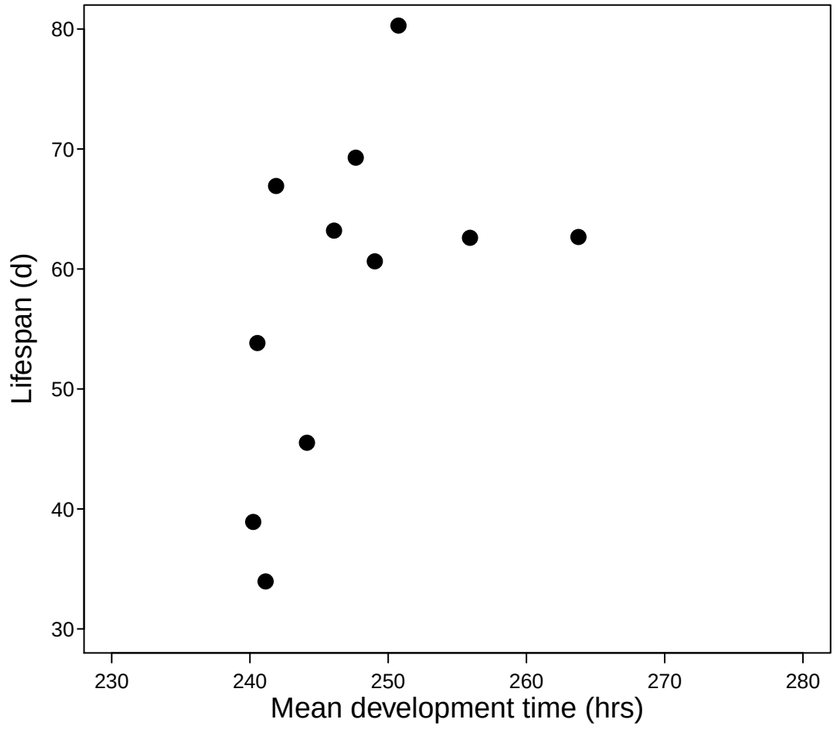
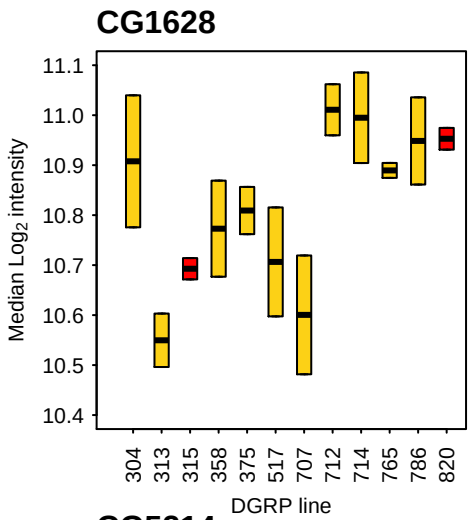
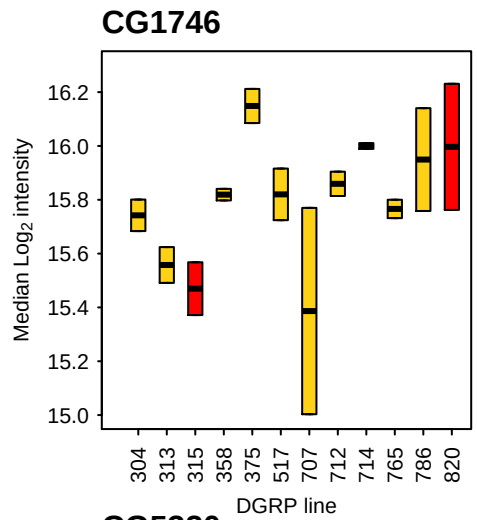
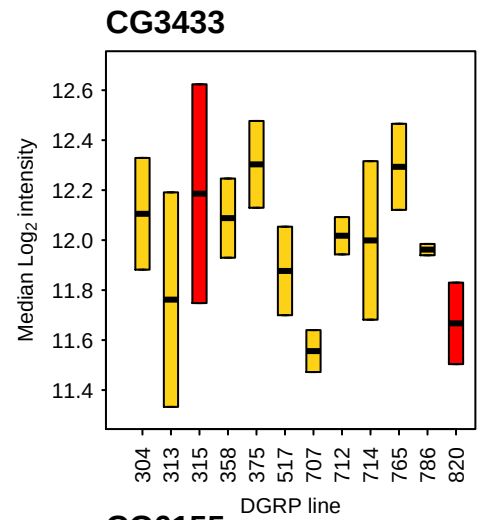
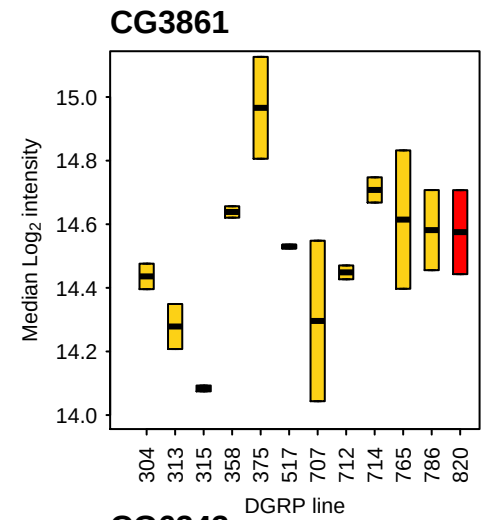

Lifespan (d)
80
70
60
50
40
30
230
240
250
260
270
280
Mean development time (hrs)

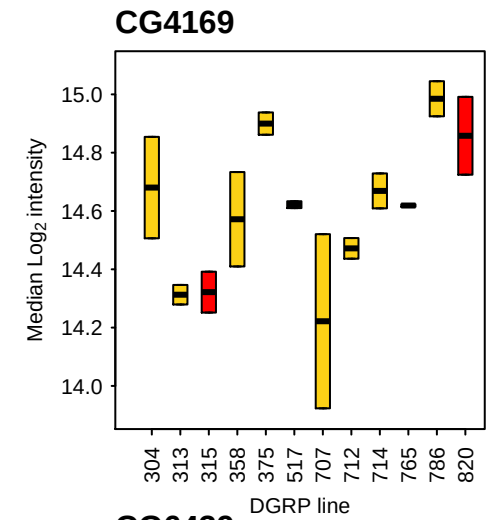
Figure S5.Development time correlations and other published studies. Egg to puparium (A) and egg to eclosion correlations (B) between the present study (abscissa) and Ellis et al's (ELLIS et al. 2014) (ordinal). Strong correlations between laboratories reveal a robust development time variation between DGRP backgrounds. Data from the present study are point estimates in comparable food and temperature treatments. Blue and red data are comparable males and females, respectively. Starvation resistance (C), chill coma recovery (D) and startle response (E) and lifespan (F) showed no correlations with development time in the present study.

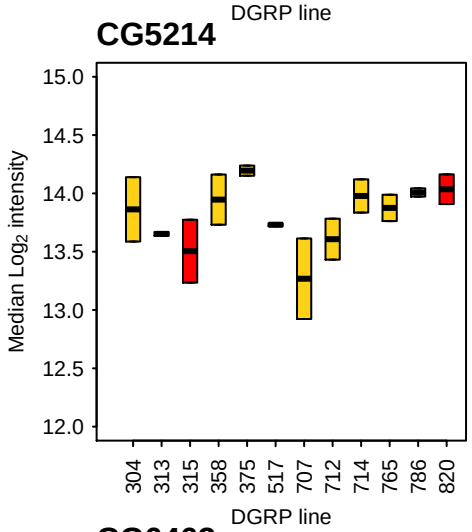

CG1628
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

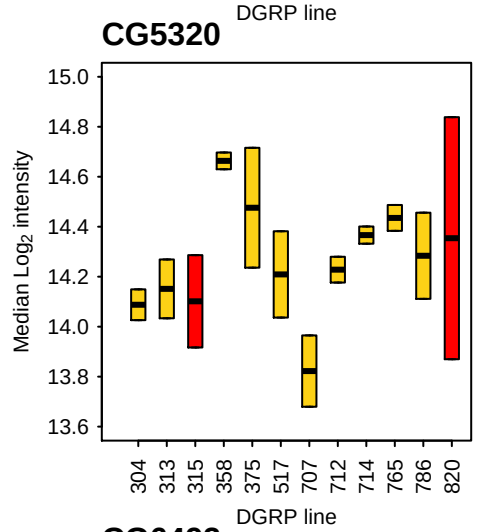

CG1746
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

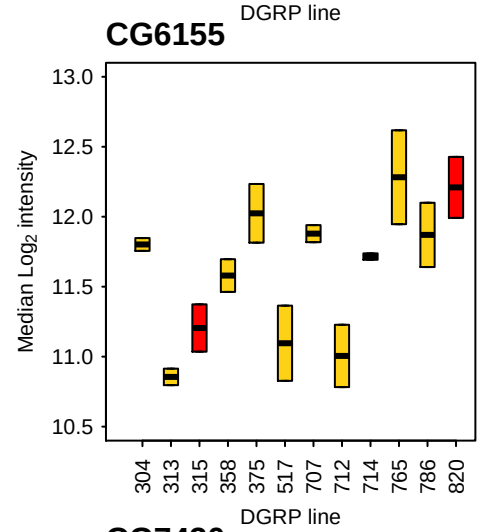

CG3433
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

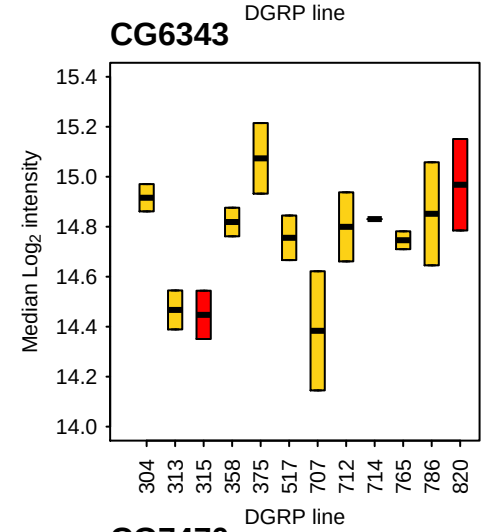

CG3861
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

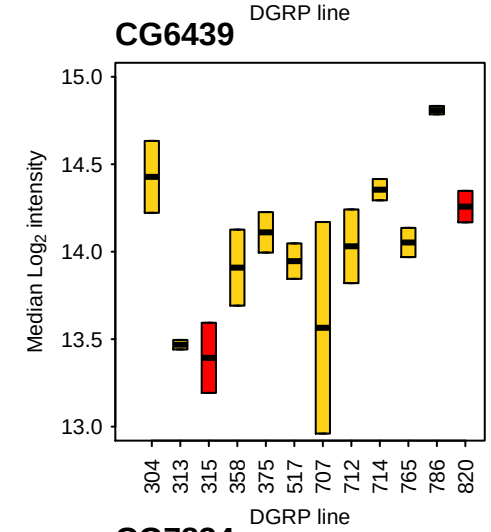

CG4169
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

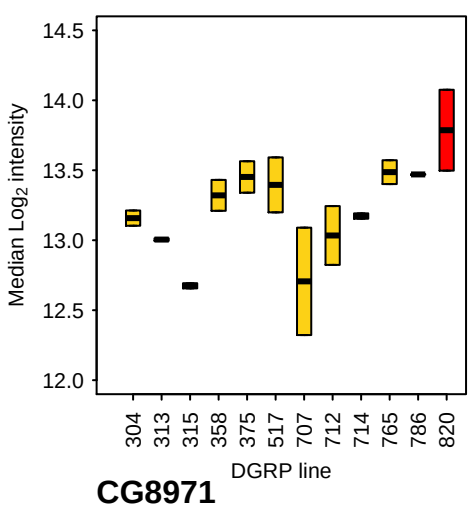

CG5214
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

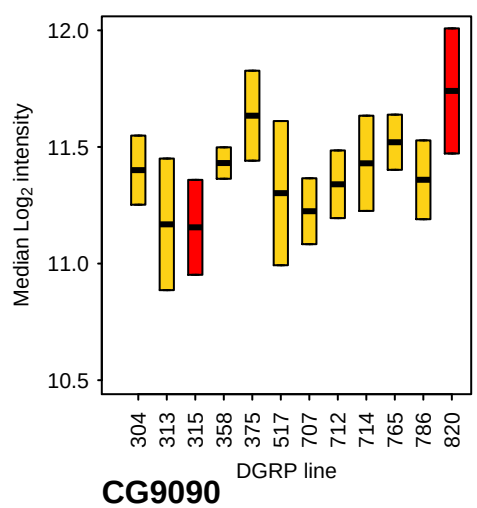

CG5320
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

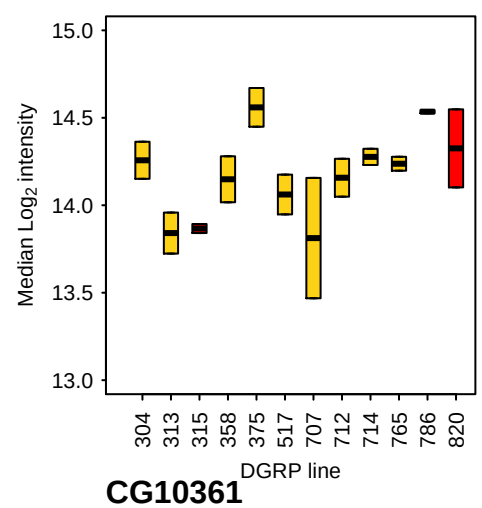

CG6155
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

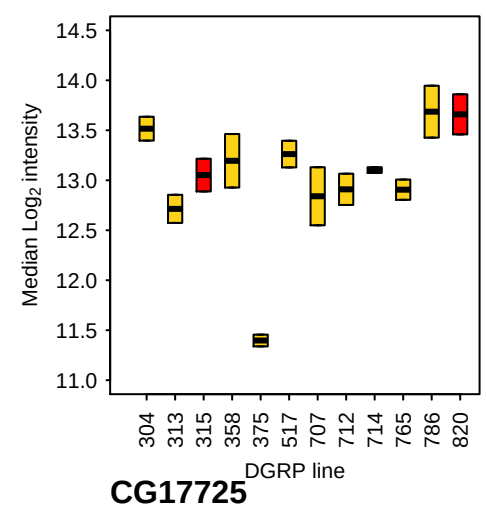

CG6343
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

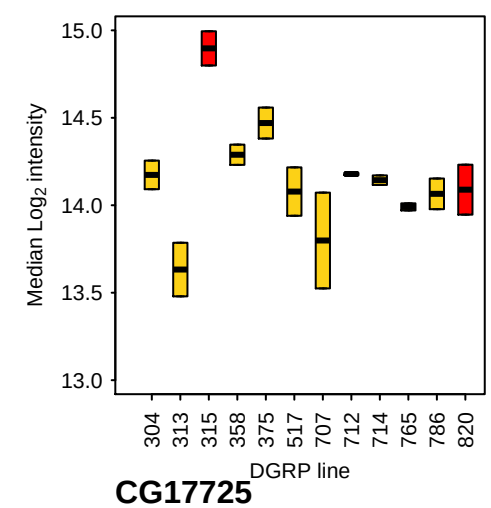

CG6439
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

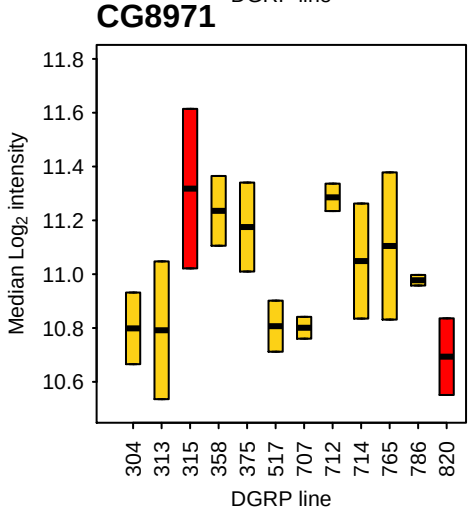

CG6463
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

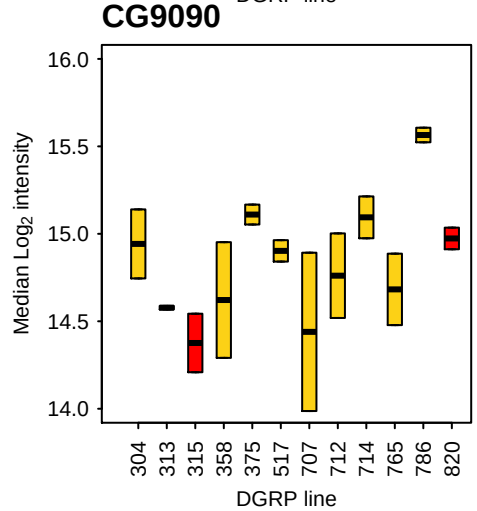

CG6492
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

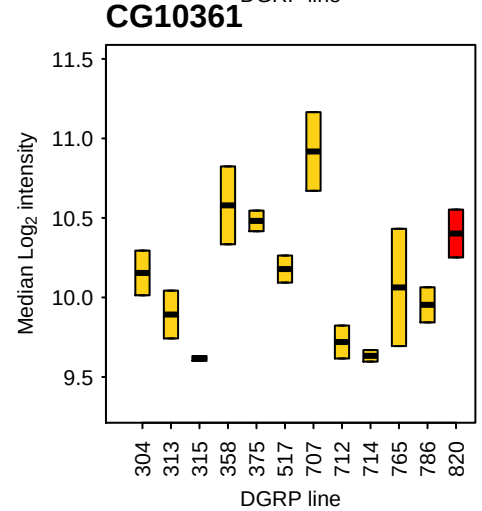

CG7430
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

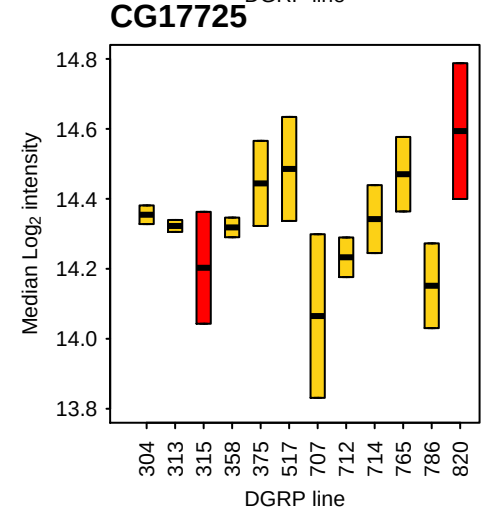

CG7470
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820


CG7834
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820


CG8971
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820


CG9090
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820


CG10361
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820


CG17725
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

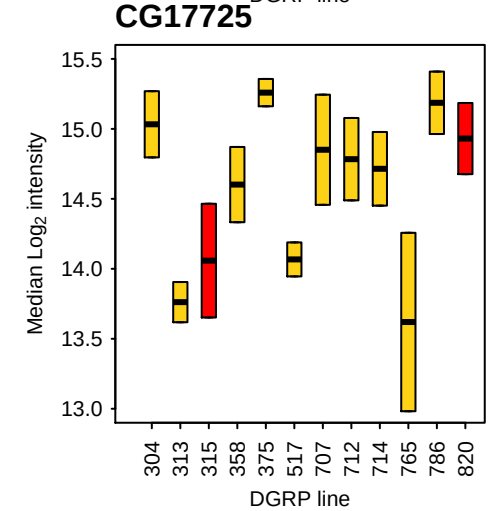

CG17725
Median Log2 intensity
DGRP line
304 313 315 358 375 517 707 712 714 765 786 820

Figure S6. Gene expression in a sub-set of 20 nuclear-mitochondrial genes in the DGRP. The 12 DGRP backgrounds are those in the present study, but in their native mitonuclear configuration (e.g. DGRP-304; DGRP-304). Each panel represents pooled gene expression variation in males (n=1) and females (n=1). Boxes in red are focal DGRP backgrounds DGRP-315 and DGRP-820 whose phenotypic variation in development time is altered by siI and OregonR mtDNA variants.

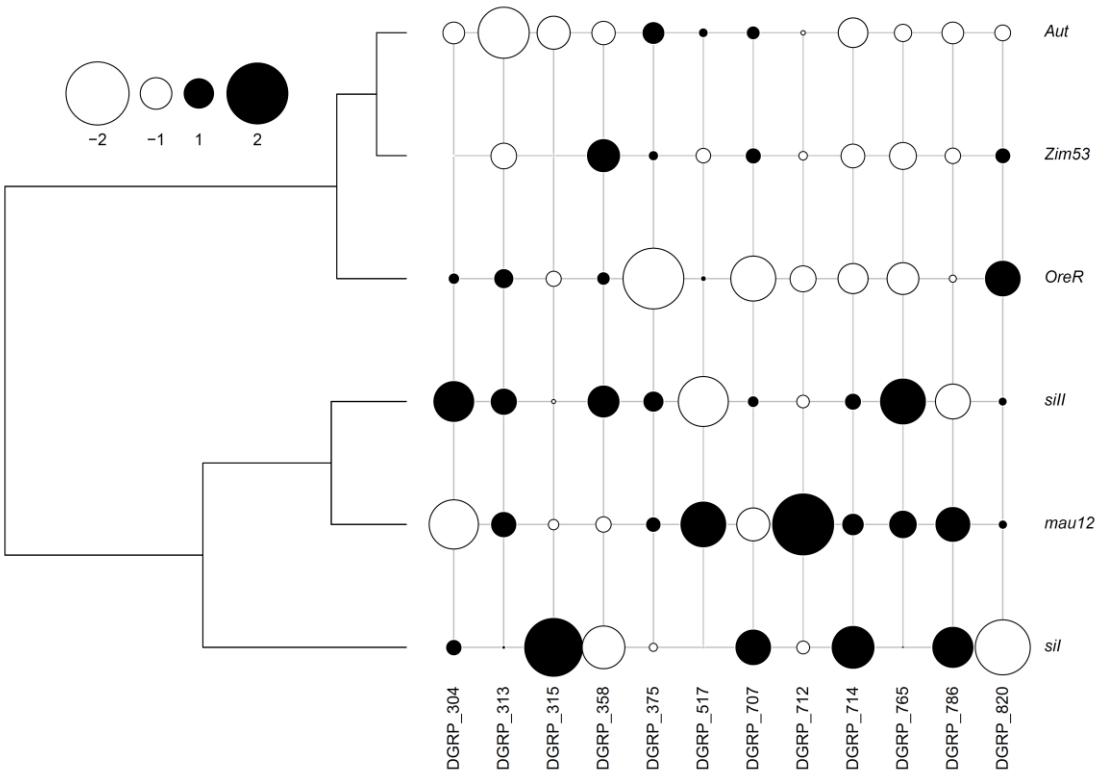


Figure S7. mtDNA phylogenetic relationships and development time phenotypic variation in Lab food for the 72 genotype mito-DGRP panel. Circles representing the magnitude of phenotype are size and color scaled, with black circles representing longer development times. A scale is included in the top left corner. Development times in Lab food are represented and each vertical column corresponds with a DGRP nuclear background. Horizontal rows represent mtDNA haplotypes and their IDs are at the rightmost edge.

ND2

Position	65	84	128	142	148	192	196	198	274	276	277	313	315	317	321	333
Zim53	V	M	M	S	F	I	I	L	M	T	L	N	I	Y	M	L
OreR	.	.	.	.	Y	.	.	.	.	.	I	.	.	.	.	.
sil	A	L	.	.	Y	F	.	.	L	.	.	I	N	T	L	M
silI	A	L	.	T	Y	.	V	F	L	L	M	I	S	T	L	M
mau	A	L	V	.	Y	.	.	.	L	L	.	I	S	T	L	M

CO-I

Position	2	128	412	451	469
Zim53	S	F	H	I	F
OreR	.	Y	.	.	.
sil	P	Y	Q	.	.
silI	P	Y	Q	V	Y
mau	P	Y	Q	V	Y

CO-II

Position	115	129	130	143	165	218
Zim53	N	M	T	V	V	Y
OreR	.	.	.	.	.	.
sil	S	T	.	I	.	H
silI	.	S	I	I	I	H
mau	.	T	.	I	.	H

ATP8

Position	26	32	34
Zim53	I	M	N
OreR	.	.	.
sil	M	I	.
silI	.	.	D
mau	.	.	.

ATP6

Position	17	28	45	115	137	177	180	185	187	192
Zim53	F	L	M	L	I	N	P	I	V	M
OreR	.	.	.	.	.	K	S	M	M	.
sil	L	I	V	M	V	.	.	L	.	T
silI	L	.	V	.	.	.	.	L	.	T
mau	L	.	V	.	.	.	.	L	.	T

CO-III

Position	41	45	170	172	176	193	199
Zim53	I	V	M	I	I	I	F
OreR	.	.	L	.	.	.	.
sil	M	L	L	V	M	V	Y
silI	M	L	L	V	M	V	Y
mau	M	L	L	V	M	V	Y

ND3

Position	20	29	81
Zim53	F	A	M
OreR	.	.	.
sil	I	G	I
silI	I	.	.
mau	I	.	.

ND5

Position	7	8	12	27	36	42	65	77	80	81	149	152	179	184	244	289	411	416	422	462	466	469	470	476	487	500	516	539	557	567	568	569
Zim53	V	N	M	D	L	M	S	M	N	H	L	S	M	V	M	L	N	I	M	V	F	L	I	F	L	I	V	M	L	L	-	I
OreR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	-	.
sil	I	Y	I	N	V	S	A	S	.	N	F	A	.	M	L	.	.	.	L	M	.	.	V	Y	.	.	I	T	M	M	N	F
silI	I	Y	I	N	V	S	A	S	.	N	F	A	V	M	L	F	.	.	.	.	S	.	.	Y	.	.	I	T	M	M	N	F
mau	I	Y	I	N	V	S	A	S	Y	N	F	A	V	M	L	F	S	V	.	.	.	M	V	Y	M	.	I	T	M	M	N	F

ND4

Position	9	12	14	17	29	37	78	82	94	141	165	231	236	243	258	298	383	386	426
Zim53	L	I	F	I	F	L	M	H	I	L	I	M	S	L	V	C	L	F	L
OreR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sil	.	T	L	M	V	V	S	Y	V	V	.	L	A	.	.	S	M	L	F
silI	V	S	L	.	.	.	S	Y	V	V	.	.	A	.	.	S	M	L	F
mau	V	S	L	.	.	.	S	Y	V	V	T	.	A	M	M	S	M	L	F

ND4L

Position	57	79
Zim53	S	V
OreR	.	.
sil	N	I
silI	N	.
mau	N	.

ND6

Position	39	82	86	90	93	94	96	100	101	102	107	159
Zim53	L	M	L	L	I	F	L	F	I	M	S	I
OreR	.	.	.	.	.	.	.	.	.	.	.	.
sil	M	I	V	.	L	.	M	L	V	I	F	V
silI	M	I	V	F	.	.	.	.	.	I	.	V
mau	M	I	V	F	L	L	.	.	.	I	.	V

CYT-B

Position	109	235	357
Zim53	K	V	V
OreR	.	.	.
sil	M	.	I
silI	M	.	.
mau	M	M	.

ND1

Position	170	186	190	191	258	270
Zim53	F	A	M	S	M	V
OreR	.	.	V	.	.	.
sil	Y	.	L	T	.	A
silI	.	G	L	T	.	I
mau	.	G	L	T	V	A

Table S1. Amino acid polymorphisms across 5 out of 6 mtDNA haplotypes used in this study. Each amino acid polymorphism position is shown relative to the reference genome (*Zim53*) across all 13 protein coding genes. A dot (.) denotes no difference to the reference and a dash (-) denotes an amino acid deletion. The IUPAC amino acid code was used.

Table S2. MtDNA 'species' effects on fitness in the mito-nuclear genotype panel. Mean point estimates of each mtDNA x nuclear background x food type were evaluated and mitochondrial genetic variation was collapsed into each species (e.g. *mau12*, *siI* and *sm21* were '*D. simulans*' haplotypes and *Zim53*, *AutW132* and *Oregon R* were *D. melanogaster* types). We could not model a random term (data were averaged and consolidated) and all models were fit as linear models and reported as ANOVA tables. Egg-to-adult viability was modeled as a Gaussian error distribution on the proportion of eggs that survived.

Phenotype	Model term	Num DF	Sum Sq	Mean Sq	F value	P-value
(a) *Development time*	Nuclear	11	10081.58	916.51	7.69	**7.58e-11**
	Mt Species	1	390.93	390.93	3.28	0.07
	Food	3	19182.33	6394.11	53.62	**6.97e-25**
	Nuclear x Mt Species	11	524.09	47.64	0.40	0.95
	Nuclear x Food	33	8429.76	255.45	2.14	**<0.0001**
	Mt Species x Food	3	152.02	50.67	0.42	0.74
	Nuclear x Mt Species x Food	33	4102.23	124.31	1.04	0.41
	Residuals	182	21703.85	119.25		
(b) *Egg-to-adult viability*	Nuclear	11	2.63	0.24	27.51	**3.52e-33**
	Mt Species	1	0.00	0.00	0.02	0.88
	Food	3	0.12	0.04	4.68	**<0.01**
	Nuclear x Mt Species	11	0.35	0.03	3.65	**<0.001**
	Nuclear x Food	33	0.62	0.02	2.15	**<0.001**
	Mt Species x Food	3	0.02	0.01	0.69	0.56
	Nuclear x Mt Species x Food	33	0.21	0.01	0.72	0.87
	Residuals	182	1.58	0.01		
(c) *Male – female sex bias*	Nuclear	11	38.72	3.52	1.82	0.053
	Mt Species	1	5.92	5.92	3.06	0.08
	Food	3	13.44	4.48	2.32	0.08
	Nuclear x Mt Species	11	32.86	2.99	1.54	0.12
	Nuclear x Food	33	85.83	2.60	1.34	0.11
	Mt Species x Food	3	3.40	1.13	0.59	0.62
	Nuclear x Mt Species x Food	33	56.83	1.72	0.89	0.64
	Residuals	182	352.09	1.93		
(d) DT_{MALE} - DT_{FEMALE}	Nuclear	11	1238.84	112.62	11.86	**1.30e-16**
	Mt Species	1	3.29	3.29	0.35	0.56
	Food	3	158.67	52.89	5.57	**<0.01**
	Nuclear x Mt Species	11	115.11	10.46	1.10	0.36
	Nuclear x Food	33	706.01	21.39	2.25	**<0.001**
	Mt Species x Food	3	21.86	7.29	0.77	0.51
	Nuclear x Mt Species x Food	33	314.07	9.52	1.00	0.47
	Residuals	182	1727.81	9.49		
(e) CV_{MALES}-$CV_{FEMALES}$	Nuclear	11	0.003	0.000	1.41	0.17
	Mt Species	1	0.000	0.000	0.03	0.87
	Food	3	0.001	0.000	2.07	0.11
	Nuclear x Mt Species	11	0.002	0.000	1.05	0.40
	Nuclear x Food	33	0.013	0.000	1.86	**0.01**
	Mt Species x Food	3	0.000	0.000	0.52	0.67
	Nuclear x Mt Species x Food	33	0.004	0.000	0.60	0.96
	Residuals	182	0.037	0.000		
(f) *Eggs per female*	Nuclear	11	3759.70	341.79	11.51	**4.40e-10**
	Mt Species	1	35.26	35.26	1.19	0.28
	Nuclear x Mt Species	11	328.92	29.90	1.01	0.45
	Residuals	48	1425.28	29.69		

Table S3. Development time best fit model assessment using the *dredge* function in the MuMIn R package. A positive sign (+) denotes the model term was included and a negative sign (-) denotes a model term was absent. All models were constrained to include DGRP ID, mtDNA ID and food type as first order terms. The removal of total offspring per vial did not significantly improve the model fit (ΔAIC<2) and we retained this covariate in all lmer and glmer models (model highlighted in blue). The models are ranked and confirm the development time ANOVA test results. Model terms, degrees-of-freedom, log liklihoods (LogLik), Akaike Information Criterion (AIC) values and the change in AIC (ΔAIC) are shown for each model. Models were fit using Maximum likelihood (ML) as recommended for changing fixed effect structures (FARAWAY 2006; ZUUR *et al.* 2009).

mtDNA	DGRP ID	Food type	Total offspring	Food type x mtDNA	Food type x DGRP ID	mtDNA x DGRP ID	Food type x mtDNA x DGRP ID	df	logLik	AIC	ΔAIC
+	+	+	-	+	+	+	+	287	6787.332131	-13000.66426	0
+	+	+	-0.000124884	+	+	+	+	288	6787.908887	-12999.81777	0.846486687
+	+	+	-0.000209748	+	+	+	-	126	6537.172383	-12822.34477	178.3194958
+	+	+	-	+	+	+	-	125	6535.723676	-12821.44735	179.2169099
+	+	+	-0.000194768	-	+	+	-	111	6513.677561	-12805.35512	195.3091387
+	+	+	-	-	+	+	-	110	6512.444718	-12804.88944	195.7748246
+	+	+	-0.000176871	+	+	-	-	71	6402.811778	-12663.62356	337.0407052
+	+	+	-	+	+	-	-	70	6401.726836	-12663.45367	337.2105889
+	+	+	-	-	+	-	-	55	6382.908279	-12655.81656	344.8477028
+	+	+	-0.000167657	-	+	-	-	56	6383.872151	-12655.7443	344.9199591
+	+	+	-0.00040398	+	-	+	-	93	6197.954979	-12209.90996	790.7543028
+	+	+	-	+	-	+	-	92	6193.268533	-12202.53707	798.1271959
+	+	+	-0.000385298	-	-	+	-	78	6176.630544	-12197.26109	803.4031733
+	+	+	-	-	-	+	-	77	6172.419051	-12190.8381	809.8261603
+	+	+	-0.000317219	+	-	-	-	38	6080.563891	-12085.12778	915.5364787
+	+	+	-	+	-	-	-	37	6077.507811	-12081.01562	919.6486402
+	+	+	-0.000305596	-	-	-	-	23	6062.829042	-12079.65808	921.0061765
+	+	+	-	-	-	-	-	22	6060.023703	-12076.04741	924.6168562

Table S4. Moran's autocorrelation coefficients, *I*. Estimates of *I* were conducted in each of 12 DGRP nuclear backgrounds. Observed and expected coefficients are shown, along with the standard deviations and associated p-values. Values in bold are significant at α=0.05. Accounting for multiple testing, a more conservative Bonferroni-corrected p-value threshold of 0.05/12 = 0.004 was not exceeded in any DGRP background.

DGRP line	*I* (observed)	*I* (expected)	SD	p-value
DGRP-304	-0.476	-0.2	0.213	0.196
DGRP-313	0.188	-0.2	0.224	0.084
DGRP-315	-0.116	-0.2	0.144	0.561
DGRP-358	-0.374	-0.2	0.246	0.478
DGRP-375	-0.193	-0.2	0.145	0.963
DGRP-517	-0.520	-0.2	0.190	0.092
DGRP-707	-0.289	-0.2	0.236	0.707
DGRP-712	-0.211	-0.2	0.100	0.912
DGRP-714	**0.372**	**-0.2**	**0.249**	**0.022**
DGRP-765	**0.391**	**-0.2**	**0.237**	**0.012**
DGRP-786	-0.214	-0.2	0.243	0.953
DGRP-820	-0.199	-0.2	0.190	0.995

File S1: Development time data. (.xls, 676 KB)

Available for download as a .xls file at:

http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.187286/-/DC1/FileS1.xls

File S2: Reciprocal cross data. (.xls, 82 KB)

Available for download as a .xls file at:

http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.187286/-/DC1/FileS2.xls

File S3: Female fecundity. (.xls, 37 KB)

Available for download as a .xls file at:

http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.187286/-/DC1/FileS3.xls

File S4: CV of development times. (.xls, 469 KB)

Available for download as a .xls file at:

http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.187286/-/DC1/FileS4.xls

File S5: Supporting Information

Mitonuclear epistasis for development time and its modification by diet in *Drosophila*

Jim A. Mossman, Leann M. Biancani, David M. Rand

This file contains:

1) **A description of the supporting phylogenetic analysis**
2) **Supporting references**

Phylogenetic analysis - Moran's autocorrelation coefficient, I

Moran's autocorrelation coefficient, I, is a statistic used to describe whether the distribution of a trait among a set of species is affected or not by their phylogenetic relationship (PARADIS 2014). In the present context, we used Moran's *I* statistic to determine whether the differences in development time we observed between haplotypes was influenced by the phylogenetic relationships between them. In other words, could we capture lineage affects, for example the 'species' effect by using a phylogenetic tree instead of using linear mixed effect statistical models. As an example, we have focused on a single trait here, development time, and in a single development environment (the Lab food treatment). In this treatment, the intra-class correlation coefficient (ICC) was lowest, suggesting a large degree of crossing of norms of reaction and hence increased likely mitonuclear epistasis (Figure 2, main text). Given the small number of haplotypes in this study (x6) and a balanced species representation (3x haplotypes per species), our analysis is more powerful in testing the impact of deeper bifurcations in the phylogeny; those that delineate species.

We built a phylogeny of the six mtDNA haplotypes based on concatenated amino acid sequences of previously published ND2, COI, ATPase6, ND5, ND4, CytB, and ND1 genes, in that order. The accession numbers of genomes containing the amino acid sequences were: (i) *D. melanogaster* isolate Zim53 (AF200829.1), (ii) *D. melanogaster* isolate Oregon R (AF200828.1), (iii) *D. simulans (sil)* isolate TT01 (AF200834.1), (iv) *D. simulans (silI)* isolate DSR (AF200841.1), and (v) *D. mauritiana* isolate BG1 (AF200831.1). In addition, we downloaded amino acid sequences from a representative Austrian haplotype (KABE11:(NUNES *et al.* 2013)) with the accession numbers ND2 (JX475716), COI (JX458192), ATPase6 (JX467115), ND5 (JX475888), ND4 (JX475802), CytB (JX475547), ND1 (JX475630). All concatenated sequences were aligned and all amino acid sequences were of equal length with no gaps.

Phenotype data were mean development times of each mtDNA haplotype (point estimate) in each of the 12 DGRP nuclear backgrounds in the Lab food.

The phylogeny was estimated using the Kitsch-Fitch-Margoliash distance matrix method (FELSENSTEIN 1989) with molecular clock as implemented in BioEdit (HALL 1999). Moran's *I* statistics were calculated using the method described in (GITTLEMAN and KOT 1990), as implemented in the ape R package (PARADIS *et al.* 2004). We estimated an autocorrelation statistic for the haplotype tree in each DGRP nuclear genetic background. The autocorrelation estimates, along with associated p-values are shown in Table S4.

Figure S7 shows the phylogeny and a matrix of phenotypic values mapped to the tree for each of the 12 DGRP backgrounds. The mtDNA haplotype tips of the tree are shown, along with circles of varying sizes representing the phenotype data. The phenotype data were scaled as in Figure 9 of the main text and mapped to the phylogeny using the table.phylo4d function in the adephylo R package (JOMBART *et al.* 2010). The two DGRP backgrounds with significant phylogenetic signal for development time (DGRP-714 and DGRP-765) show opposing colored circles in the *D. simulans* and *D. melanogaster* clades, corresponding to appreciable and consistent differences in development time. In these cases, development time is sensitive to the species of mtDNA.

Supporting References

FELSENSTEIN, J., 1989 PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics **5:** 164-166.

GITTLEMAN, J. L., and M. KOT, 1990 Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects. Systematic Zoology **39:** 227-241.

HALL, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, pp. 95-98 in *Nucleic acids symposium series*.

JOMBART, T., F. BALLOUX and S. DRAY, 2010 Adephylo: new tools for investigating the phylogenetic signal in biological traits. Bioinformatics **26:** 1907-1909.

NUNES, M. D. S., M. DOLEZAL and C. SCHLÖTTERER, 2013 Extensive paternal mtDNA leakage in natural populations of Drosophila melanogaster. Molecular Ecology **22:** 2106-2117.

PARADIS, E., 2014 Moran's Autocorrelation Coefficient in Comparative Methods. R Vignette.

PARADIS, E., J. CLAUDE and K. STRIMMER, 2004 APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **20:** 289-290.