Similar efficacies of selection shape mitochondrial and nuclear genes in *Drosophila melanogaster* and *Homo sapiens*

Brandon S. Cooper[*,1], Chad R. Burrus[§], Chao Ji[§], Matthew W. Hahn[*,§], and Kristi L. Montooth[*,2]

[*] Department of Biology, Indiana University, Bloomington, IN 47405, USA

[§] School of Informatics and Computing, Indiana University, Bloomington, IN 47405, USA

[1]Current address: Department of Evolution and Ecology, University of California, Davis, CA 95616, USA

[2]Current address: School of Biological Sciences, University of Nebraska, Lincoln, NE 68588

Short running title: Mitochondrial and nuclear *NI*

Keywords: cytoplasmic sweep, mtDNA, neutrality index, tests of selection

Author for correspondence: Kristi Montooth, School of Biological Sciences, 1104 T Street, Lincoln, NE 68588-0118, kmontooth2@unl.edu

ABSTRACT

Deleterious mutations contribute to polymorphism even when selection effectively prevents their fixation. The efficacy of selection in removing deleterious mitochondrial mutations from populations depends on the effective population size (N_e) of the mtDNA, and the degree to which a lack of recombination magnifies the effects of linked selection. Using complete mitochondrial genomes from *Drosophila melanogaster* and nuclear data available from the same samples, we re-examine the hypothesis that non-recombining animal mtDNA harbor an excess of deleterious polymorphisms relative to the nuclear genome. We find no evidence of recombination in the mitochondrial genome, and the much-reduced level of mitochondrial synonymous polymorphism relative to nuclear genes is consistent with a reduction in N_e. Nevertheless, we find that the neutrality index (*NI*), a measure of the excess on nonsynonymous polymorphism relative to the neutral expectation, is not significantly different between mitochondrial and nuclear loci. Reanalysis of published data from *Homo sapiens* reveals the same lack of a difference between the two genomes, though small samples in previous studies had suggested a strong difference in both species. Thus, despite a smaller N_e, mitochondrial loci of both flies and humans appear to experience similar efficacies of selection as do loci in the recombining nuclear genome.

20 INTRODUCTION

The effective size of a population (N_e) impacts how effectively selection removes deleterious mutations and fixes advantageous mutations. The unique genetics of the mitochondrial genome (mtDNA) are thought to reduce its N_e relative to the nuclear genome, via haploid, uniparental inheritance, the mitochondrial bottleneck in the maternal germline, and a lack of recombination
25 that decreases N_e via selection on linked sites (HILL and ROBERTSON 1966; MAYNARD SMITH and HAIGH 1974; GILLESPIE 2000; MEIKLEJOHN *et al.* 2007; WHITE *et al.* 2008; CHARLESWORTH 2012). In addition, cytoplasmic transmission can link the mtDNA to selfish cytoplasmic elements (e.g., *Wolbachia* in insects) that may sweep through populations, further decreasing mitochondrial N_e and possibly increasing mitochondrial substitution rates via the fixation of
30 slightly deleterious mutations (SHOEMAKER *et al.* 2004). For these reasons it has been widely hypothesized that selection is less effective in mitochondrial genomes than in their nuclear counterparts, and that mitochondrial genomes may accumulate greater numbers of deleterious substitutions (LYNCH 1996; LYNCH 1997). Analyses of sequence data in *Drosophila* and mammals have appeared to largely support these conclusions (BALLARD and KREITMAN 1994;
35 RAND and KANN 1996; NACHMAN 1998; RAND and KANN 1998; WEINREICH and RAND 2000).

N_e is not the only evolutionary parameter that distinguishes mitochondrial and nuclear genomes. The distinct functional landscape of the mitochondrial genome likely affects the distribution of selective effects (s) of mutations that arise in this genome. Animal mitochondrial genomes typically encode regulatory information for replication and transcription nested within a
40 hypervariable region (also known as the D-loop, control, or A+T-rich region), 22 tRNAs, 2 ribosomal components, and 13 protein-coding genes—all core components of oxidative phosphorylation (OXPHOS). Outside of the control region, there is little noncoding DNA in

animal mtDNAs. In Drosophilids, over 98% of the genome outside of the control region encodes DNA and RNA genes with highly conserved sequences that function in mitochondrial protein synthesis and aerobic respiration (CLARY and WOLSTENHOLME 1985; WOLSTENHOLME and CLARY 1985; BALLARD 2000; MONTOOTH *et al.* 2009). This suggests that the distribution of selective effects in the mtDNA may be shifted towards larger negative effects on fitness.

The mutational landscape of the mtDNA also differs from the nuclear genome. In most animal taxa the mitochondrial mutation rate greatly exceeds that of the nuclear genome (LYNCH *et al.* 2008), and the mitochondrial mutational process is also highly biased (MONTOOTH and RAND 2008). For example, nearly all mitochondrial mutations in *D. melanogaster* change a G:C base pair to an A:T (HAAG-LIAUTARD *et al.* 2008). When combined with the strong A+T-bias in this mitochondrial genome, where 95% of third codon positions are an A or a T (MONTOOTH *et al.* 2009), this indicates that the most commonly occurring mutations in protein-coding loci of the *Drosophila* mtDNA will change an amino acid. Relative to the nuclear genome, animal mitochondrial genomes thus experience a greater mutational pressure that can also be biased in some taxa towards nonsynonymous mutations; these are likely to have deleterious effects in a molecule that encodes such highly conserved functions.

Some of the strongest population genetic patterns in support of distinct selective pressures acting on mitochondrial and nuclear genomes come from analyses of the neutrality index (*NI*) (RAND and KANN 1996; NACHMAN 1998; RAND and KANN 1998; WEINREICH and RAND 2000). *NI* is a summary statistic of the deviation from the neutral expectation in the McDonald-Kreitman test (MCDONALD and KREITMAN 1991; RAND and KANN 1996), and is calculated from counts of synonymous and nonsynonymous polymorphic and fixed sites within and between related species. Weakly deleterious nonsynonymous mutations that segregate in the

population but that will not contribute to divergence lead to a value of *NI* greater than 1. When the efficacy of selection is decreased, the expectation is that the number of segregating weakly deleterious polymorphisms will increase; this is the pattern that has been observed in mtDNA. Meta-analyses of McDonald-Kreitman tables and their associated *NI* values for mitochondrial
70 and nuclear loci in animals have concluded that *NI* is predominantly greater than the neutral expectation of 1 for the mtDNA (RAND and KANN 1996; NACHMAN 1998; RAND and KANN 1998; BETANCOURT *et al.* 2012) and exceeds the average *NI* of the nuclear genome (WEINREICH and RAND 2000). Although the relative sparseness of the data was recognized early on (NACHMAN 1998), these patterns have been taken as evidence that selection is largely ineffective
75 in the mtDNA because of its reduced N_e, and that mitochondrial genomes harbor more deleterious polymorphisms than do their nuclear counterparts.

Here we revisit this pattern using new, complete mitochondrial genomes from *D. melanogaster* that we compare with published nuclear data from the same samples (LANGLEY *et al.* 2012) and with available human data from both nuclear and mitochondrial genomes
80 (BUSTAMANTE *et al.* 2005; JUST *et al.* 2008; RUBINO *et al.* 2012). We find little evidence that the effects of selection differ on average between mitochondrial and nuclear genomes of flies or humans, despite evidence that there is much reduced N_e due to a lack of recombination and linkage with the cytoplasm. We discuss reasons why *NI* is, on average, similar between mitochondrial and nuclear loci, despite the distinct population genetic properties of these two
85 genomes.

MATERIALS AND METHODS

D. melanogaster mtDNA assembly, annotation, and estimates of sequence diversity

Raw sequence read files from 38 genetic lines of *D. melanogaster* from Raleigh, NC, USA
90 (MACKAY *et al.* 2012) sequenced by the 50 Genomes subproject of the *Drosophila* Population Genomics Project (DPGP) (LANGLEY *et al.* 2012) were downloaded from the NCBI Sequence Read Archive. We used the Burrows-Wheeler Aligner (BWA), and specifically the fast and accurate short read alignment with Burrows-Wheeler Transform (LI and DURBIN 2009), to map sequence reads to the *D. melanogaster* mitochondrial reference genome (NC_001709). We
95 allowed up to 5 gaps, 5 gap extensions, and 5 mismatches per aligned read, but few reads needed such flexibility and most were filtered out in later steps. Using SAMtools, we post-processed the alignments to filter out low-quality alignments and to detect single nucleotide polymorphisms (SNPs) (LI *et al.* 2009). SNPs with a quality score greater than 20 and indels with a quality score greater than 50 were kept for further analyses. We then generated a consensus sequence for each
100 of the *D. melanogaster* mtDNAs listed in Table S1. Due to high variance in coverage across the A+T-rich D-loop, we did not include this region in our final assemblies or analyses.

We annotated the consensus sequence for each mtDNA using the GenBank annotation of the *D. melanogaster* reference sequence (NC_001709). Using ClustalW (LARKIN *et al.* 2007), we performed a whole genome alignment, as well as gene-specific alignments, of each consensus
105 sequence to the reference sequence and to the outgroup species *D. yakuba* (NC_001322). There are very few indels in the coding regions of *Drosophila* mtDNA (MONTOOTH *et al.* 2009), making alignment straightforward. From these alignments we calculated expected heterozygosity (π), the number of segregating sites (S), and Watterson's θ_W (WATTERSON 1975) as measures of sequence diversity. The mitochondrial haplotype network was inferred from 80 segregating sites
110 in the coding region of the mtDNA for which there were no missing or ambiguous data using TCS version 1.21 (CLEMENT *et al.* 2000).

Tests for recombination in the *D. melanogaster* mtDNA

We estimated linkage disequilibrium (LD) between all pairs of mitochondrial SNPs using the
115 statistic *D'* (LEWONTIN 1964), where *D*' = 0 indicates no LD and |*D*'| = 1 indicates perfect LD. Because recombination erodes LD as a function of distance, a negative correlation between |*D*'| and genetic distance between pairs of SNPs has been used as evidence for recombination (AWADALLA *et al.* 1999). To test this prediction we looked for significant negative correlations between |*D*'| and genetic distance. We also conducted these same tests using another statistical
120 measure of association, r^2 (HILL and ROBERTSON 1966), which is more robust to variation in mutation rates (AWADALLA *et al.* 1999; MEUNIER and EYRE-WALKER 2001; INNAN and NORDBORG 2002). We calculated these correlations using a variety of minor allele frequency cutoffs (Table S2). We also tested for the presence of all four genotypes at pairs of SNPs (the "four-gamete test"; HUDSON and KAPLAN 1985) using DNAsp version 5 (ROZAS *et al.* 2003).
125

Neutrality tests

Using π and θ_W, we calculated Tajima's *D* (TAJIMA 1989), which is expected to be 0 for a neutrally evolving locus. Demographic effects will skew the site-frequency spectrum of both synonymous and nonsynonymous polymorphisms at a locus. Contrasting Tajima's *D* between
130 nonsynonymous and synonymous polymorphisms therefore tests whether nonsynonymous alleles experience a greater skew in frequency relative to putatively neutral synonymous alleles, indicative of selection (RAND and KANN 1996). We implemented this analysis using the heterogeneity test (HAHN *et al.* 2002), which simulates 10,000 genealogies with no recombination using the values of synonymous and nonsynonymous *S* calculated from the data

135 and compares the estimated difference in Tajima's *D* to the random distribution of differences between synonymous and nonsynonymous polymorphisms. We calculated several other summaries of the site-frequency spectrum, including Fu and Li's *D*, which characterizes the proportion of mutations on external and internal branches of a genealogy (FU and LI 1993) and Fay and Wu's *H*, which tests for an excess of high-frequency derived alleles in a sample relative
140 to the neutral expectation (FAY and WU 2000). These latter statistics were calculated using a set of 80 segregating sites in the coding region of the mtDNA for which there were no missing or ambiguous data. Significance was determined using 10,000 coalescent simulations as implemented in DNAsp version 5 (ROZAS *et al.* 2003).

We conducted McDonald-Kreitman (MK) tests (MCDONALD and KREITMAN 1991) using
145 Fisher's exact tests of the two-by-two contingency tables of counts of nonsynonymous and synonymous polymorphisms (P_N and P_S) within *D. melanogaster* and nonsynonymous and synonymous fixed differences (D_N and D_S) between *D. melanogaster* and *D. yakuba*. Polymorphic sites within *D. melanogaster* only contributed to fixed differences if the allele in *D. yakuba* was not present in *D. melanogaster.* For codons with more than one change, we
150 calculated the number of nonsynonymous and synonymous differences as the average over all possible mutational pathways between codons (NEI and GOJOBORI 1986). For any gene with a count of 0 in any cell of the MK table, we added a count of 1 to all cells (SHELDAHL *et al.* 2003; PRESGRAVES 2005). 23% of thirteen *D. melanogaster* mitochondrial genes, 9.5% of 6,113 *D. melanogaster* nuclear genes, 0% of thirteen *H. sapiens* mitochondrial genes, and 73% of 11,624
155 *H. sapiens* nuclear genes required these additional counts. Fisher's exact tests were performed using R version 2.15.1 (R CORE TEAM 2012).

We calculated *NI*—the ratio of P_N/P_S to D_N/D_S—as a summary statistic of the MK table (RAND and KANN 1996). Assuming that selection is constant, the neutral expectation is that D_N/D_s will equal P_N/P_S (KIMURA 1983; MCDONALD and KREITMAN 1991), and *NI* is expected to be 1. When the MK test is significant, an *NI* less than 1 indicates a significant excess of nonsynonymous fixed differences and an *NI* greater than 1 indicates a significant excess of nonsynonymous polymorphisms. We calculated several modified statistics similar to *NI*, including $Z^* = Log\left(\frac{(D_N+1)(P_S+1)}{(D_S+1)(P_N+1)}\right)$ (PRESGRAVES 2005) and the direction of selection $DoS = \frac{D_N}{D_N+D_S} - \frac{P_N}{P_N+P_S}$, which is an unbiased summary statistic of the two-by-two MK table (STOLETZKI and EYRE-WALKER 2011). The sign of these latter two statistics is more intuitive; negative values are consistent with weakly deleterious (negatively selected) polymorphisms and positive vales are consistent with advantageous (positively selected) substitutions.

The short length and reduced polymorphism of mitochondrial genes decreases the power of the MK test and upwardly biases *NI* (WEINREICH and RAND 2000). Because of this, we also performed MK tests and calculated *NI* using the summed counts of polymorphic and divergent sites for each of the OXPHOS complexes: Complex I (NADH dehydrogenase, ND), Complex IV (cytochrome *c* oxidase, CO) and Complex V (ATP synthetase, ATPase). Cytochrome B (CytB) is the only Complex III gene encoded by the mtDNA. Stoletzki and Eyre-Walker (2011) emphasize that contingency data should not generally be summed, particularly when there is heterogeneity among contingency tables, and they provide an unbiased estimator of overall *NI* for combining counts, $NI_{TG} = \frac{\sum D_{Si}P_{Ni}/(P_{Si}+D_{Si})}{\sum P_{Si}D_{Ni}/(P_{Si}+D_{Si})}$. We calculated this statistic and used the DoFE software package (STOLETZKI and EYRE-WALKER 2011) to calculate bootstrap confidence intervals and to conduct Woolf's tests of homogeneity (WOOLF 1955). The only data set with significant heterogeneity was the *D. melanogaster* nuclear gene set (P<0.0001). The same

180 approaches were used to analyze polymorphism and divergence in the human data sets (see below), as well as in a subset of the mitochondrial haplotypes reported in our study that were independently sequenced and assembled by Richardson et al. (2012) (Table S3).

Comparisons of mitochondrial and nuclear *NI* in flies and humans

185 To compare patterns of polymorphism and divergence between mitochondrial and nuclear genomes, we obtained existing data for nuclear genes in *D. melanogaster* and for nuclear and mitochondrial genes in *Homo sapiens.* Counts of polymorphic and divergent sites for *D. melanogaster* nuclear genes were obtained from the DPGP analysis of the same 38 genomes from Raleigh, NC, USA, using *D. yakuba* as the outgroup species (LANGLEY *et al.* 2012). The 190 human nuclear data from Bustamante et al. (2005) included counts of polymorphic and divergent sites from 19 African Americans and 20 European Americans, using the chimpanzee *Pan troglodytes* as an outgroup species. We calculated the number of polymorphic and divergent sites for human mitochondrial genes using mtDNA sequences from 19 African Americans (JUST *et al.* 2008), 20 European Americans (RUBINO *et al.* 2012), and the chimpanzee mitochondrial 195 reference genome D38113.1 (HORAI *et al.* 1995). We also analyzed subsets of the human mitochondrial sequence data to characterize the sensitivity of *NI* to sampling (Table S4). Table S5 provides accession numbers for the human mtDNA sequences used. Comparisons of the distributions of *NI* between data sets were performed using Mann-Whitney *U* tests in R version 2.15.1 (R CORE TEAM 2012).

200

RESULTS

An excess of low- and high-frequency derived mitochondrial polymorphisms

We assembled 14,916 bp of sequence containing the transcribed regions of the mtDNA with a median coverage of 32x for 38 genetic lines sampled from a single population of *D. melanogaster* in Raleigh, NC (LANGLEY *et al.* 2012; MACKAY *et al.* 2012)(Table S1). Over 98% of these nucleotides encode the 13 protein-coding genes, 22 tRNAs and two ribosomal RNAs. The per-site expected heterozygosity in this region (π) of the mtDNA was 0.0008. We identified 137 segregating sites in this population sample, 103 of which were in protein-coding genes. Median heterozygosity in protein coding genes was 0.0023 per synonymous site and 0.0002 per nonsynonymous site. Silent site heterozygosity was significantly lower in mitochondrial genes relative to nuclear genes (Mann-Whitney U, mtDNA v. X chromosome, $P_{MWU} = 0.00002$; mtDNA v. autosomes, $P_{MWU} < 0.00001$) and was only 0.16 times that of the autosomes (Figure 1A), lower than what is expected if the mtDNA has an effective population size that is one-quarter that of the autosomes.

In addition to there being very few segregating sites in the *D. melanogaster* mtDNA, polymorphisms at these sites were skewed towards low frequencies (Figure 2A,B), as evidenced by consistently negative values of Tajima's D (Table 1). Tajima's D across the mtDNA was -2.607 and differed significantly from the neutral expectation of 0 (S=80 for coding sites with no missing data, $P<0.0001$), as did Fu and Li's D ($D = -2.67$, $P<0.05$). The minor allele frequency for unpolarized synonymous polymorphisms was always less than 11%, and all but one of the nonsynonymous polymorphisms was a singleton (Figure 2A,B). Using *D. yakuba* as an outgroup revealed that the derived allele was nearly fixed for 44% of segregating synonymous sites, while there was only a single derived nonsynonymous polymorphism at high frequency (Figure 2C,D). Thus, the mitochondrial genome was essentially devoid of intermediate frequency polymorphisms, with derived synonymous mutations at either very high (greater than 89%) or

very low (less than 11%) frequencies and nearly all derived nonsynonymous polymorphisms at frequencies less than 5%. This skew towards high-frequency derived alleles resulted in a significant negative value of Fay and Wu's *H* statistic ($H = -41.2$, $P = 0.005$).

230 **A partial selective sweep in the *D. melanogaster* mtDNA**

The large fraction of derived alleles at high frequencies is a consequence of the haplotype structure of this sample (Figure 3). Nearly 30% of individuals in this population shared an identical mitochondrial haplotype and an additional 66% of individuals differed from this haplotype by only one to five mutations. The two remaining haplotypes (RAL-639 and RAL-
235 335) were highly divergent from this common haplotype group, contributing nearly half of the segregating sites to the population sample. These two haplotypes shared the ancestral state with *D. yakuba* at 17 of the 23 derived high-frequency synonymous polymorphisms (i.e., they have the low-frequency ancestral allele; Figure 2C). When these two haplotypes were removed from the analysis, there remained a strong skew toward rare alleles (Tajima's $D = -2.31$, $P < 0.01$; Fu
240 and Li's $D = -3.14$, $P < 0.02$), but Fay and Wu's *H*, which is sensitive to the number of high-frequency derived alleles, was only weakly significant ($H = -10.34$, $P = 0.043$). The remaining six derived, high-frequency synonymous polymorphisms, as well as the single derived, high-frequency nonsynonymous polymorphism, were the result of single mtDNAs within the common haplotype group having the same allelic state as *D. yakuba*. Given the lack of recombination in
245 the mtDNA (see below), these are likely new, rather than ancestral, mutations. Six of these seven mutations would have changed a C or G to a T or A, consistent with the mutation bias in the *D. melanogaster* mtDNA (HAAG-LIAUTARD *et al.* 2008).

No evidence for recombination in the *D. melanogaster* mtDNA

250 We tested for a negative correlation between linkage disequilibrium (LD) and the distance between each pair of polymorphic sites in the *D. melanogaster* mitochondrial genome, as a signature of the decay of LD over distance via recombination (AWADALLA *et al.* 1999). There was no evidence for a decrease in LD with increasing distance between sites, regardless of the measure of LD or the minor allele cutoff used (Table S2). There were no pairs of polymorphic 255 sites for which all four gametes were present (HUDSON and KAPLAN 1985; BRUEN *et al.* 2006), further supporting an absence of effective recombination.

Weakly deleterious polymorphism in the *D. melanogaster* mtDNA

The skew in the site-frequency spectrum toward rare alleles (Figure 2) resulted in negative 260 values of Tajima's *D* across the entire mtDNA (Table 1). There was no evidence that the skew toward rare alleles differed between synonymous and nonsynonymous polymorphisms (Figure 2A,B): heterogeneity tests (HAHN *et al.* 2002) of Tajima's *D* between synonymous and nonsynonymous sites were never significant ($P > 0.35$ for all genes and complexes). However, unfolding the frequency spectra revealed that the large number of high-frequency derived sites 265 were nearly all synonymous (Figure 2C,D). This would indicate that the haplotype that has increased in frequency carried many more synonymous than nonsynonymous polymorphisms. Given that the mutation bias in the *D. melanogaster* mtDNA greatly favors nonsynonymous mutations (HAAG-LIAUTARD *et al.* 2008), this pattern suggests a history of effective purifying selection removing mitochondrial haplotypes that contain nonsynonymous polymorphisms. 270 Furthermore, all nonsynonymous polymorphisms that have arisen on the common mitochondrial haplotype are present at very low frequencies.

The current distribution of polymorphisms relative to divergence showed little evidence for an excess of segregating deleterious polymorphisms. No single gene and only one OXPHOS complex significantly departed from the neutral expectation after Bonferroni correction (Table 2 and Table 3); the MK test was significant for Complex V (ATPase), with an excess of nonsynonymous polymorphism (Fisher's exact test, $P_{FET} = 0.007$). For the entire set of protein-coding mitochondrial genes, there was a slight excess of nonsynonymous polymorphism relative to the neutral expectation that was not significant after Bonferroni correction ($P_{FET} = 0.041$). Using the unbiased estimator NI_{TG} (STOLETZKI and EYRE-WALKER 2011) did not change this inference: NI_{TG} for mitochondrial-encoded proteins was 1.67, but the confidence intervals on NI_{TG} nearly included the neutral expectation of 1 (Table 3).

These patterns were verified using 36 of the 38 mitochondrial haplotypes in our sample that were independently sequenced and assembled by Richardson et al. (2012). No single mitochondrial gene rejected the neutral expectation and the MK table counts for Complex V (ATPase) were exactly the same in the two data sets (Table S3). When counts of polymorphism and divergence differed between our datasets, they typically differed by only a single count. The only exception was in several Complex I (ND) genes, for which our assembled mtDNAs had a small number of additional nonsynonymous polymorphisms relative to the Richardson et al. (2012) data set (ND genes, $df = 6$, $P_{MWU,\ paired} = 0.021$; all other genes, $df = 5$, $P_{MWU,\ paired} = 1$) that resulted in slightly higher values of NI (ND included, $df = 12$, $P_{MWU,\ paired} = 0.016$; all other genes, $df = 5$, $P_{MWU,\ paired} = 1$). This was not due to exclusion of two mitochondrial haplotypes in the Richardson et al. (2012) data set, and the additional polymorphisms in our data were not clustered on any single haplotype. The reduced number of nonsynonymous polymorphisms in the Richardson et al. (2012) data provided even less support for an excess of nonsynonymous

295 segregating variation in the mitochondrial genome. Patterns of polymorphism and divergence for the entire set of mitochondrial-encoded proteins in this dataset did not deviate from the neutral expectation (P_{FET} = 0.423), and the confidence intervals on NI_{TG} for mitochondrially encoded proteins contained the neutral expectation of 1 (NI_{TG} = 0.821, 95% CI = 0.386 to 1.90).

300 **On average, *NI* is the same for mitochondrial and nuclear genes in flies**

The distribution of *D. melanogaster* mitochondrial gene *NI* values was contained within that of the 6,151 nuclear genes for which we had MK table counts, with many nuclear genes having both more positive and more negative values of *Z** and *DoS* (Figure 1 and Figure 4). The median value of *NI* was 1.97 (NI_{TG} = 1.67) for the 13 mitochondrial genes and 1.48 (NI_{TG} = 1.35) for

305 6,113 nuclear genes, and did not differ significantly between genomes (P_{MWU} = 0.278). Confidence intervals for NI_{TG} were overlapping for the two genomes (Table 3). Median values of *Z** and *DoS* were negative for both genomes (Figure 1 and Figure 4), consistent with the presence of a slight excess of deleterious polymorphisms in both genomes. *Z** for *D. melanogaster* mitochondrial genes did not differ significantly from nuclear genes, independent of

310 nuclear location (mtDNA vs. X chromosome, P_{MWU} = 0.065; mtDNA vs. autosomes, P_{MWU} = 0.325) (Figure 1C). *DoS* was modestly significantly different between mitochondrial and X-chromosome genes (mtDNA vs. X, P_{MWU} = 0.044), but not between the mtDNA and autosomal genes (mtDNA vs. autosomes, P_{MWU} = 0.126). Nevertheless, there was a trend for mitochondrial genes to differ more from X-linked loci than they did from autosomal loci for both *Z** and *DoS*

315 (Figure 1C,D). Analysis of assembled mtDNAs from Richardson et al. (2012) (see above) yielded qualitatively similar results; *NI* did not differ between mitochondrial and nuclear genes (P_{MWU} = 0.500), and neither *Z** or *DoS* differed significantly between the mitochondria and any

nuclear location (Z^*, $P_{MWU} > 0.471$ for both comparisons; *DoS*, $P_{MWU} > 0.726$ for both comparisons).

320

_NI_ does not differ between mitochondrial and nuclear genes in humans

Summary statistics of the MK table also did not differ between mitochondrial and nuclear genes in *H. sapiens* (*NI*, $P_{MWU} = 0.657$; Z^*, $P_{MWU} = 0.243$; *DoS*, $P_{MWU} = 0.700$)(Figure 4). Values of Z^* and *DoS* for mitochondrial genes in humans were well within the distribution of the 11,624

325 nuclear genes examined, and the confidence intervals of NI_{TG} were overlapping for the mitochondrial and nuclear genomes (Table 3). The distributions of mitochondrial gene MK summary statistics were also largely overlapping and did not differ significantly between *D. melanogaster* and *H. sapiens* (*NI*, $P_{MWU} = 0.545$; Z^*, $P_{MWU} = 0.441$; *DoS*, $P_{MWU} = 0.310$)(Figure 4).

330 We also found that *NI* is very sensitive to sampling (Table 4). When only a few individuals are sampled, the choice of genomes can lead to high variability and extreme values in *NI*, potentially as a result of single haplotypes that may carry multiple polymorphisms, as appears to be the case for human *ND6* (Table 4). For example, depending on which Japanese individual we included in our analyses, *NI* for *ND6* takes on values of 30.71 (MK, $P_{FET} = 0.001$),

335 5.50 (MK, $P_{FET} = 0.31$), or 1.79 (MK, $P_{FET} = 0.16$) when sampling only 3 mtDNAs. As more mtDNAs are sampled, point estimates of *NI* and Z^* for each mitochondrial gene become more similar to the neutral expectation (Table 4) and average values of *NI* and NI_{TG} no longer show strong evidence for deviations from the neutral expectation.

340 DISCUSSION

Using a large sample of whole-genome sequence data, we have tested a number of hypotheses about mtDNA evolution, and about differences in the efficacy of selection on mitochondrial versus nuclear genes. Our data confirm that mtDNA do not recombine and have lower silent site
345 diversity than do nuclear genes, which supports the prediction that the mitochondrial genome has a lower N_e than does the nuclear genome. We also show a skew in the site-frequency spectrum toward rare alleles that likely has two sources: 1) the accumulation of new mutations on what appears to be a mtDNA haplotype that has swept to high frequency in the recent past, and 2) the ancestral polymorphisms contained on migrant or remnant haplotypes (RAL-639 and RAL-335)
350 that are now rare in this population. Despite the apparent reduction in N_e for mtDNA, our findings indicate that selection is equally effective at purging deleterious polymorphisms from the mitochondrial and nuclear genomes of *D. melanogaster* and *H. sapiens*.

Given its uniparental and haploid transmission, the expectation under neutrality is that the mtDNA has one-quarter the population size of the autosomes. This reduced value of N (and
355 subsequently N_e) matches that expected for the Y (or W) chromosome, and, like the Y, the mtDNA has little to no recombination. However, very much unlike the Y chromosomes that have been sequenced (e.g. CHARLESWORTH and CHARLESWORTH 2000; CARVALHO *et al.* 2009; CARVALHO and CLARK 2013; BELLOTT *et al.* 2014), animal mtDNA genomes do not show an accumulation of deleterious mutations. Animal mtDNA does not show an accumulation of
360 transposable elements, and the gene content of the mitochondrial genome is remarkably stable, with few gene losses and even fewer pseudogenes (BALLARD and RAND 2005). Across mammals, d_N/d_S is two to fifteen times lower for mitochondrial genes than for nuclear genes (POPADIN *et al.* 2012), and in *Drosophila* average values of d_N/d_S for mitochondrial genes are well under 0.1, and are on average only 13% that of nuclear genes (BAZIN *et al.* 2006;

365 MONTOOTH *et al.* 2009). This pattern is particularly striking given that the mutation rate in the *D. melanogaster* mtDNA is an order of magnitude greater than the per-site mutation rate in the nuclear genome (HAAG-LIAUTARD *et al.* 2007; HAAG-LIAUTARD *et al.* 2008). While heteromorphic Y chromosomes do show signatures of less effective purifying selection, such as proliferation of satellite repeats and reduced codon bias (BACHTROG 2013; SINGH *et al.* 2014),
370 the single copy, X-degenerate genes that have remained on the human Y chromosome experience effective purifying selection (BELLOTT *et al.* 2014) (Bellott et al. 2014; Rozen et al. 2009), as do the protein sequences of Drosophila Y-linked genes (SINGH *et al.* 2014). Thus, despite early loss of many genes when heteromorphic Y chromosomes and mtDNA formed, both these non-recombining chromosomes have maintained a stable set of genes that experience effective
375 purifying selection despite their reduced N_e.

Many researchers have cited the early work on *NI* in Drosophila and mammals in support of the idea that mtDNA has accumulated deleterious mutations (MEIKLEJOHN *et al.* 2007; GREEN *et al.* 2008; NEIMAN and TAYLOR 2009; AKASHI *et al.* 2012). In fact, this idea has become so engrained that it is regularly cited in reviews of mitochondrial gene evolution (e.g. BALLARD and
380 WHITLOCK 2004; LYNCH 2007, p. 338). What is surprising about this conversion of a small set of intriguing initial studies into dogma is that the early studies themselves were quite circumspect about the implications of their results. For instance, Nachman (1998), in noting that very few nuclear loci were available for comparison, stated "It is also unclear whether the patterns reported here are unique to mitochondrial DNA." Data from the few nuclear genes that had been
385 sequenced raised "the possibility that the patterns reported here for mtDNA may also be found at some nuclear loci" (NACHMAN 1998). Even the relatively more recent studies that did have access to additional nuclear datasets were only able to calculate *NI* for 36 nuclear loci

(WEINREICH and RAND 2000), and the *NI* values that were available often did not deviate significantly from neutrality (NACHMAN 1998; WEINREICH and RAND 2000). Those that did
390 reject neutrality tended to do so weakly, perhaps due to the small number of polymorphisms in mitochondrial samples even when the number of individuals sampled is high (e.g. ND3, NACHMAN *et al.* 1996). Nevertheless, there are mitochondrial genes that do strongly reject neutrality and some of these had *NI* values that greatly exceeded *NI* for the sampled nuclear loci. Based on these and similar comparisons, many authors have reached the conclusion that mtDNA
395 evolves in a manner distinct from the nuclear genome. Our results using the whole genomes of flies and humans suggest that they are not evolving differently.

Reductions in N_e—due either to reductions in census population size or to the increased effect of linked selected variants in regions of low recombination—are expected to result in a reduction in the efficacy of selection. Indeed, comparisons of McDonald-Kreitman test results
400 across a range of species with different values of N_e has revealed this expected relationship (e.g. LI *et al.* 2008; WRIGHT and ANDOLFATTO 2008; GOSSMANN *et al.* 2010), as have comparisons of *NI* across regions of the *D. melanogaster* genome with different recombination rates (PRESGRAVES 2005; LANGLEY *et al.* 2012). Therefore, all things being equal, mitochondrial loci would be expected to harbor an excess of replacement polymorphisms relative to nuclear loci
405 due to reduced N_e. Our results suggest that all things are not equal between these two cellular compartments, and that there may be features of the mitochondrion that make it less likely to accumulate deleterious mutations. One such feature is the "bottleneck" that occurs in the number of mtDNAs that are passed from mother to offspring—this event makes it possible for selection to act within hosts, possibly increasing the power of selection to remove deleterious mutations
410 (BERGSTROM and PRITCHARD 1998; RAND 2011). The additional layers of selection imposed by

mitochondrial inheritance, combined with stronger negative selective effects of amino acid changing mutations in mitochondrial genes, may therefore allow the mtDNA to escape the accumulation of deleterious mutation, resulting in equal values of *NI* between nucleus and mitochondria.

415 Our results come with several caveats. First, we have only studied two organisms—it may be that a more comprehensive review of *NI* in mtDNA and nuclear loci across many species will reveal a difference in the average efficacy of selection. The early meta-analyses of *NI* contained loci from a wide range of animals (NACHMAN 1998; WEINREICH and RAND 2000), possibly suggesting that using only Drosophila and humans results in a view that is too limited.
420 Nevertheless, these are two model organisms for evolutionary biology that span a large range of mtDNA:nuclear substitution rates, and studies of these species have set the bar for much of modern population genetics. Second, it is clear from our analysis of the *D. melanogaster* mtDNA that it is not at equilibrium. Much of the theory used to predict *NI* values from N_e and s assumes mutation-selection-drift balance (see, e.g. NACHMAN 1998), and deviations from this equilibrium
425 can result in more complex relationships between N_e, s, and *NI* (MESSER and PETROV 2013). While non-equilibrium histories may mean that mtDNA *NI* values are not at equilibrium, it is equally likely that nuclear genes from *D. melanogaster* are not at mutation-selection-drift equilibrium (HAHN 2008; LANGLEY *et al.* 2012). Whether or not the mtDNA is at equilibrium, and whether or not the *NI* values calculated from this snapshot of two species represent
430 equilibrium values, our results still imply that there is no difference between nuclear and mitochondrial measures of the efficacy of selection.

Despite the mitochondrial genome experiencing a distinct population genetic environment relative to the nuclear genome, our whole-genome analyses uncovered no evidence

for an excess accumulation of slightly deleterious mutations. In fact, the only strong evidence for
435 a reduced efficacy of selection in mtDNA, relative to nuclear genomes, comes from comparative studies of nuclear and mitochondrial tRNAs (LYNCH 1996; LYNCH 1997). As discussed above, in the absence of a pattern in *NI*, there are essentially no other patterns of molecular evolution in mtDNA indicative of deleterious mutation accumulation. This pattern is in stark contrast to the patterns found in analogous nuclear regions with reduced N_e and low recombination, like the Y
440 chromosome. Determining whether mtDNA accumulates deleterious substitutions more readily than nuclear DNA in a larger sample of species—and what type of loci may be affected—will be a particularly fruitful avenue for future studies.

445 ACKNOWLEDGEMENTS

We thank Colin Meiklejohn and members of the Montooth and Hahn labs for constructive feedback. B.S.C. was supported on the Indiana University Genetics, Molecular and Cellular Sciences Training Grant T32-GM007757 funded by the National Institutes of Health and a
450 Doctoral Dissertation Improvement Grant funded by the National Science Foundation. This research was supported by funding from Indiana University, award DBI-0845494 from the National Science Foundation to M.W.H., and a National Science Foundation CAREER award IOS-1149178 to K.L.M.

455

LITERATURE CITED

AKASHI, H., N. OSADA and T. OHTA, 2012 Weak Selection and Protein Evolution. Genetics **192:** 15-U43.

AWADALLA, P., A. EYRE-WALKER and J. M. SMITH, 1999 Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science **286:** 2524-2525.

BACHTROG, D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nature Reviews Genetics **14:** 113-124.

BALLARD, J. W., 2000 Comparative genomics of mitochondrial DNA in members of the Drosophila melanogaster subgroup. J Mol Evol **51:** 48-63.

BALLARD, J. W., and M. KREITMAN, 1994 Unraveling selection in the mitochondrial genome of Drosophila. Genetics **138:** 757-772.

BALLARD, J. W. O., and D. M. RAND, 2005 The population biology of mitochondrial DNA and its phylogenetic implications, pp. 621-642 in *Annu Rev Ecol Evol S*.

BALLARD, J. W. O., and M. C. WHITLOCK, 2004 The incomplete natural history of mitochondria. Molecular Ecology **13:** 729-744.

BAZIN, E., S. GLEMIN and N. GALTIER, 2006 Population size does not influence mitochondrial genetic diversity in animals. Science **312:** 570-572.

BELLOTT, D. W., J. F. HUGHES, H. SKALETSKY, L. G. BROWN, T. PYNTIKOVA *et al*., 2014 Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature **508:** 494-+.

BERGSTROM, C. T., and J. PRITCHARD, 1998 Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics **149:** 2135-2146.

BETANCOURT, A. J., B. BLANCO-MARTIN and B. CHARLESWORTH, 2012 The relation between the neutrality index for mitochondrial genes and the distribution of mutational effects on fitness. Evolution **66:** 2427-2438.

BRUEN, T. C., H. PHILIPPE and D. BRYANT, 2006 A simple and robust statistical test for detecting the presence of recombination. Genetics **172:** 2665-2681.

BUSTAMANTE, C. D., A. FLEDEL-ALON, S. WILLIAMSON, R. NIELSEN, M. T. HUBISZ *et al*., 2005 Natural selection on protein-coding genes in the human genome. Nature **437:** 1153-1157.

CARVALHO, A. B., and A. G. CLARK, 2013 Efficient identification of Y chromosome sequences in the human and Drosophila genomes. Genome Research **23:** 1894-1907.

CARVALHO, A. B., L. B. KOERICH and A. G. CLARK, 2009 Origin and evolution of Y chromosomes: Drosophila tales. Trends in Genetics **25:** 270-277.

CHARLESWORTH, B., 2012 The effects of deleterious mutations on evolution at linked sites. Genetics **190:** 5-22.

CHARLESWORTH, B., and D. CHARLESWORTH, 2000 The degeneration of Y chromosomes. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **355:** 1563-1572.

CLARY, D. O., and D. R. WOLSTENHOLME, 1985 The mitochondrial DNA molecular of Drosophila yakuba: nucleotide sequence, gene organization, and genetic code. J Mol Evol **22:** 252-271.

CLEMENT, M., D. POSADA and K. A. CRANDALL, 2000 TCS: a computer program to estimate gene genealogies. Mol Ecol **9:** 1657-1659.

FAY, J. C., and C. I. WU, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405-1413.

FU, Y. X., and W. H. LI, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693-709.

GILLESPIE, J. H., 2000 Genetic drift in an infinite population: The pseudohitchhiking model. Genetics **155:** 909-919.

GOSSMANN, T. I., B.-H. SONG, A. J. WINDSOR, T. MITCHELL-OLDS, C. J. DIXON *et al*., 2010 Genome Wide Analyses Reveal Little Evidence for Adaptive Evolution in Many Plant Species. Molecular Biology and Evolution **27:** 1822-1832.

GREEN, R. E., A.-S. MALASPINAS, J. KRAUSE, A. W. BRIGGS, P. L. F. JOHNSON *et al*., 2008 A complete neandertal mitochondrial genome sequence determined by high-throughput Sequencing. Cell **134:** 416-426.

HAAG-LIAUTARD, C., N. COFFEY, D. HOULE, M. LYNCH, B. CHARLESWORTH *et al*., 2008 Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS Biol **6:** e204.

HAAG-LIAUTARD, C., M. DORRIS, X. MASIDE, S. MACASKILL, D. L. HALLIGAN *et al*., 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature **445:** 82-85.

HAHN, M. W., 2008 Toward a selection theory of molecular evolution. Evolution **62:** 255-265.

HAHN, M. W., M. D. RAUSHER and C. W. CUNNINGHAM, 2002 Distinguishing between selection and population expansion in an experimental lineage of bacteriophage t7. Genetics **161:** 11-20.

HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet Res **8:** 269-294.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE and N. TAKAHATA, 1995 Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci U S A **92:** 532-536.

HUDSON, R. R., and N. L. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147-164.

INNAN, H., and M. NORDBORG, 2002 Recombination or mutational hot spots in human mtDNA? Mol Biol Evol **19:** 1122-1127.

JUST, R. S., T. M. DIEGOLI, J. L. SAUNIER, J. A. IRWIN and T. J. PARSONS, 2008 Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. Forensic Sci Int Genet **2:** e45-48.

KIMURA, M., 1983 *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.

LANGLEY, C., K. STEVENS, C. CARDENO, Y. LEE, D. SCHRIDER *et al*., 2012 Genomic Variation in Natural Populations of Drosophila melanogaster. Genetics **192:** 533-598.

LARKIN, M. A., G. BLACKSHIELDS, N. P. BROWN, R. CHENNA, P. A. MCGETTIGAN *et al*., 2007 Clustal W and clustal X version 2.0. Bioinformatics **23:** 2947-2948.

LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49-67.

LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:** 1754-1760.

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN *et al*., 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079.

LI, Y. F., J. C. COSTELLO, A. K. HOLLOWAY and M. W. HAHN, 2008 "Reverse ecology" and the power of population genomics. Evolution **62:** 2984-2994.

LYNCH, M., 1996 Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol Biol Evol **13:** 209-220.

LYNCH, M., 1997 Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. Mol Biol Evol **14:** 914-925.

LYNCH, M., 2007 *Origins of Genome Architecture*.

LYNCH, M., W. SUNG, K. MORRIS, N. COFFEY, C. R. LANDRY *et al*., 2008 A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA.

MACKAY, T., S. RICHARDS, E. STONE, A. BARBADILLA, J. AYROLES *et al*., 2012 The Drosophila melanogaster Genetic Reference Panel. Nature **482:** 173-178.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genetical Research **23:** 23-35.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature **351:** 652-654.

MEIKLEJOHN, C. D., K. L. MONTOOTH and D. M. RAND, 2007 Positive and negative selection on the mitochondrial genome. Trends Genet **23:** 259-263.

MESSER, P. W., and D. A. PETROV, 2013 Frequent adaptation and the McDonald–Kreitman test. Proceedings of the National Academy of Sciences **110:** 8615-8620.

MEUNIER, J., and A. EYRE-WALKER, 2001 The correlation between linkage disequilibrium and distance: implications for recombination in hominid mitochondria. Mol Biol Evol **18:** 2132-2135.

MONTOOTH, K. L., D. N. ABT, J. HOFMANN and D. M. RAND, 2009 Comparative genomics of Drosophila mtDNA: Novel features of conservation and change across functional domains and lineages. J Mol Evol **69:** 94-114.

MONTOOTH, K. L., and D. M. RAND, 2008 The spectrum of mitochondrial mutation differs across species. PLoS Biol **6:** e213.

NACHMAN, M. W., 1998 Deleterious mutations in animal mitochondrial DNA. Genetica **102-103:** 61-69.

NACHMAN, M. W., W. M. BROWN, M. STONEKING and C. F. AQUADRO, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953-963.

NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol **3:** 418-426.

NEIMAN, M., and D. R. TAYLOR, 2009 The causes of mutation accumulation in mitochondrial genomes. Proc Biol Sci **276:** 1201-1209.

POPADIN, K. Y., S. I. NIKOLAEV, T. JUNIER, M. BARANOVA and S. E. ANTONARAKIS, 2012 Purifying Selection in Mammalian Mitochondrial Protein-Coding Genes Is Highly Effective and Congruent with Evolution of Nuclear Genes. Mol Biol Evol.

PRESGRAVES, D., 2005 Recombination enhances protein adaptation in Drosophila melanogaster. Curr Biol **15:** 1651-1656.

R CORE TEAM, 2012 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria **ISBN 3-900051-07-0:** URL http://www.R-project.org/.

RAND, D. M., 2011 Population genetics of the cytoplasm and the units of selection on mitochondrial DNA in Drosophila melanogaster. Genetica **139:** 685-697.

RAND, D. M., and L. M. KANN, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol Biol Evol **13:** 735-748.

RAND, D. M., and L. M. KANN, 1998 Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica **102-103:** 393-407.

RICHARDSON, M. F., L. A. WEINERT, J. J. WELCH, R. S. LINHEIRO, M. M. MAGWIRE *et al*., 2012 Population Genomics of the Wolbachia Endosymbiont in Drosophila melanogaster. PLoS Genet **8:** e1003129.

ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496-2497.

RUBINO, F., R. PIREDDA, F. M. CALABRESE, D. SIMONE, M. LANG *et al*., 2012 HmtDB, a genomic resource for mitochondrion-based human variability studies. Nucleic Acids Res **40:** D1150-1159.

SHELDAHL, L. A., D. M. WEINREICH and D. M. RAND, 2003 Recombination, dominance and selection on amino acid polymorphism in the Drosophila genome: contrasting patterns on the X and fourth chromosomes. Genetics **165:** 1195-1208.

SHOEMAKER, D. D., K. A. DYER, M. AHRENS, K. MCABEE and J. JAENIKE, 2004 Decreased diversity but increased substitution rate in host mtDNA as a consequence of Wolbachia endosymbiont infection. Genetics **168:** 2049-2058.

SINGH, N. D., L. B. KOERICH, A. B. CARVALHO and A. G. CLARK, 2014 Positive and purifying selection on the Drosophila y chromosome. Mol Biol Evol **31:** 2612-2623.

STOLETZKI, N., and A. EYRE-WALKER, 2011 Estimation of the neutrality index. Mol Biol Evol **28:** 63-70.

TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585-595.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor Popul Biol **7:** 256-276.

620 WEINREICH, D. M., and D. M. RAND, 2000 Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. Genetics **156:** 385-399.

WHITE, D. J., J. N. WOLFF, M. PIERSON and N. J. GEMMELL, 2008 Revealing the hidden complexities of mtDNA inheritance. Mol Ecol **17:** 4925-4942.

WOLSTENHOLME, D. R., and D. O. CLARY, 1985 Sequence evolution of Drosophila mitochondrial

625 DNA. Genetics **109:** 725-744.

WOOLF, B., 1955 On Estimating the Relation between Blood Group and Disease. Annals of Human Genetics **19:** 251-253.

WRIGHT, S. I., and P. ANDOLFATTO, 2008 The Impact of Natural Selection on the Genome: Emerging Patterns in Drosophila and Arabidopsis, pp. 193-213 in *Annual Review of Ecology*

630 *Evolution and Systematics*.

Supporting Information

Table S1 mtDNAs assembled in this study along with the average coverage

Table S2 Correlations between linkage disequilibrium (LD) and distance (bp) between pairs of SNPs in the *D. melanogaster* mtDNA

Table S3 Counts of polymorphic (*P*) and divergent (*D*) nonsynonymous (*N*) and synonymous (*S*) sites along with summary statistics of the McDonald-Kreitman table using 36 of the 38 mitochondrial haplotypes in our sample that were independently sequenced and assembled by Richardson et al. (2012).

Table S4 Effects of sampling on *NI*

Table S5 Accession numbers for mtDNAs used in Table S4

File S1 FASTA file with the 38 assembled genomes used in this study

Table 1 Synonymous and nonsynonymous variation in *D. melanogaster* mitochondrial genes and OXPHOS complexes

	Synonymous sites[b]					Nonsynonymous sites				
Gene/Complex[a]	n (bp)	*S*	π	Θ_W	*D*	n (bp)	*S*	π	Θ_W	*D*
COI	347	16	1.57	3.81	-1.91	1183	0	0	0	undef
COII	142	5	0.31	1.19	-1.90	539	1	0.05	0.24	-1.13
COIII	169.33	9	0.72	2.14	-1.96	613.67	1	0.05	0.24	-1.13
NDI	194	3	0.16	0.71	-1.72	739	3	0.16	0.71	-1.72
NDII	198	5	0.41	1.19	-1.69	822	6	0.34	1.43	-2.07
NDIII	68	1	0.05	0.24	-1.12	280	1	0.05	0.24	-1.12
NDIV	273.33	7	0.51	1.67	-1.95	1061.67	7	0.38	1.67	-2.17
NDIVL	56	1	0.05	0.24	-1.13	229	2	0.11	0.48	-1.49
NDV	351	10	0.68	2.38	-2.16	1365	4	0.21	0.95	-1.88
NDVI	103.33	4	0.26	0.95	-1.75	415.67	2	0.11	0.48	-1.49
Complex I	1243.66	31	2.12	7.38	-2.48	4912.34	25	1.35	5.95	-2.64
Complex III	239	5	0.55	1.19	-1.38	892	2	0.16	0.48	-1.29
Complex IV	658.33	30	2.60	7.14	-2.21	2335.67	2	0.11	0.48	-1.49
Complex V	174.67	2	0.16	0.71	-1.72	650.33	6	0.32	1.43	-2.10
Total	2315.66	68	5.43	16.42	-2.44	8790.34	35	1.93	8.33	-2.70

[a] Complex I, ND, 7 loci; Complex III, Cyt-b, 1 locus; Complex IV, CO, 3 loci; Complex V, ATPase, 2 overlapping loci

[b] For synonymous and nonsynonymous sites we calculated the number of segregating sites (*S*), heterozygosity (π), Watterson's Θ_W, and Tajima's *D*. The heterogeneity test for differences between synonymous and nonsynonymous *D* was never significant ($P > 0.35$ for all comparisons).

Table 2 Counts of polymorphic (*P*) and divergent (*D*) nonsynonymous (*N*) and synonymous (*S*) sites along with summary statistics of the McDonald-Kreitman table for *D. melanogaster* mitochondrial genes

Gene	P_N	P_S	D_N	D_S	NI[a]	Z^* [b]	DoS[c]
ATPase6	5	2	11	35	7.955	-0.778	-0.475
ATPase8	1	0	2	8	6	-0.778	undef
COI	0	16	8	101	0.667	0.176	undef
COII	1	5	6	39	1.3	-0.280	-0.033
COIII	1	9	8.5	47.5	0.621	-0.009	0.052
CytB	2	5	17.5	67.5	1.543	-0.267	-0.080
NDI	3	3	11	45	4.091	-0.584	-0.304
NDII	6	5	25	41	1.968	-0.275	-0.167
NDIII	1	1	5	22	4.4	-0.584	-0.315
NDIV	7	7	24	63	2.625	-0.408	-0.224
NDIVL	2	1	1	7	14	-0.778	-0.542
NDV	4	10	55.833	107.167	0.768	0.063	0.057
NDVI	2	4	21.5	22.5	0.523	0.203	0.155

[a] For *ATPase8* and *COI*, a count of 1 was added to each cell when calculating $NI = \frac{D_S P_N}{D_N P_S}$. For no gene was the neutral expectation rejected via a Fisher's exact test of the MK table at the table-wise Bonferroni corrected *P*=0.004.

[b] $Z^* = Log\left(\frac{(D_N+1)(P_S+1)}{(D_S+1)(P_N+1)}\right)$, as in (PRESGRAVES 2005).

[c] $DoS = \frac{D_N}{D_N+D_S} - \frac{P_N}{P_N+P_S}$ was not calculated for loci with a count of zero (STOLETZKI and EYRE-WALKER 2011).

Table 3 Summary statistics of the McDonald-Kreitman table for mitochondrial OXPHOS complexes and nuclear genes

Species	Genome	Gene set[a]	*NI* [b]	NI_{TG}[c]	*Z* [b]	*DoS* [b]
D. melanogaster	mtDNA	Complex I	1.73	1.59 (0.94, 3.17)	-0.238	-0.129
	mtDNA	Complex IV	0.56	0.55 (0, 1.30)	0.255	0.045
	mtDNA	**Complex V**	**9.92**	**9.64 (undef)**	**-0.997**	**-0.518**
	mtDNA	All coding	1.59	1.67 (1.03, 2.86)	-0.201	-0.095
	nuclear	All coding	1.13	1.35 (1.31, 1.38)	-0.053	-0.023
H. sapiens	mtDNA	Complex I	1.19	1.20 (0.56, 2.40)	-0.074	-0.027
	mtDNA	Complex IV	1.98	2.00 (0.86, 3.61)	-0.296	-0.060
	mtDNA	Complex V	1.62	1.66 (0.79, 2.42)	-0.209	-0.097
	mtDNA	All coding	1.46	1.48 (0.90, 2.24)	-0.165	-0.059
	nuclear	All coding	1.51	1.57 (1.51, 1.63)	-0.180	-0.101

[a] Complex I, ND, seven loci; Complex IV, CO, three loci; Complex V, ATPase, two overlapping loci. Complex II is nuclear encoded and Complex III has only a single mitochondrial locus, *CytB.*

[b] *NI*, $Z = Log\left(\frac{D_N P_S}{D_S P_N}\right)$, and *DoS* were calculated using counts of P_N, P_S, D_N and D_S summed across genes within gene sets. **Bold** indicates *P* = 0.007 for Fisher's exact test of the MK table, which is significant after sample- but not table-wise Bonferroni correction.

[c] $NI_{TG} = \frac{\sum D_{Si} P_{Ni}/(P_{Si}+D_{Si})}{\sum P_{Si} D_{Ni}/(P_{Si}+D_{Si})}$ with confidence intervals from 5000 bootstrap samples (STOLETZKI and EYRE-WALKER 2011). Only the *D. melanogaster* nuclear dataset had significant heterogeneity as assessed by Woolf's test of homogeneity.

Table 4 The sensitivity of *NI* to sampling

Sample	1 African 1 European 1 Japanese [a]		1 African 1 European 1 Japanese [b]		1 African 1 European 1 Japanese [b]		19 African-American 20 European-American[c]		30 African-American 30 European-American[c]	
	NI [d]	Z*	*NI*	Z*	*NI*	Z*	*NI*	Z*	*NI*	Z*
ATPase	2.27	-0.37	2.78	-0.44	2.84	-0.45	1.62	-0.22	1.65	-0.22
COI	**12.50**	**-1.08**	6.89	-0.88	**10.33**	**-1.03**	**3.61**	**-0.56**	**3.68**	**-0.56**
COII	3.33	-0.52	5.80	-0.82	3.28	-0.52	0.86	-0.13	0.64	-0.01
COIII	2.36	-0.53	1.38	-0.14	1.60	-0.39	1.38	-0.23	0.95	-0.08
Cyt-b	**3.78**	**-0.57**	3.64	-0.55	3.64	-0.55	2.29	-0.36	**3.23***	**-0.50**
NDI	1.58	-0.28	2.51	-0.43	2.63	-0.45	1.49	-0.20	1.49	-0.19
NDII	**5.86**	**-0.76**	**5.86**	**-0.76**	**7.73**	**-0.86**	3.59	-0.55	**2.95**	**-0.46**
NDIII	2.83	-0.52	6.80	-0.77	2.83	-0.52	1.33	-0.28	1.52	-0.25
NDIV	0.93	-0.09	0.52	0.05	0.52	0.05	**0.13**	**0.60**	**0.11**	**0.70**
NDIVL	2.20	-0.34	2.63	-0.42	2.63	-0.42	5.00	-0.62	4.00	-0.54
NDV	2.00	-0.32	2.38	-0.39	2.23	-0.37	1.20	-0.09	1.44	-0.16
NDVI	**30.71***	**-1.22**	5.50	-0.74	1.79	-0.25	1.30	-0.20	1.17	-0.16
All coding	**2.88***	**-0.46**	**2.50***	**-0.40**	**2.41***	**-0.39**	**1.46**	**-0.17**	**1.55**	**-0.19**
NI_{TG} (CI)[e]	**2.85** (1.83,4.99)		**2.60** (1.65,3.91)		**2.56** (1.59,3.81)		1.48 (0.92,2.25)		1.59 (0.93,2.36)	

[a] MK table counts from Nachman et al. (1996)

[b] MK table counts as above, but substituting two different, randomly chosen Japanese samples

[c] MK table counts from African-American and European-American sequences sampled from (JUST *et al.* 2008) and (RUBINO *et al.* 2012) with the chimpanzee mitochondrial reference genome as an outgroup (HORAI *et al.* 1995).

[d] A count of 1 was added to each cell when calculating NI for an locus with a zero count in any cell. **Bold** indicates $P \leq 0.05$; ***** indicates significant a sample-wise Bonferroni-corrected *P*-value of less than 0.004 for Fisher's exact test of the MK table.

[e] Calculated as in Table 3. No sample rejected Woolf's test of homogeneity ($P > 0.19$ for all samples).

Bold indicates that the confidence intervals do not overlap the neutral expectation of 1.

FIGURE LEGENDS

Figure 1 Effect of genomic location on silent-site heterozygosity and on summary statistics of polymorphism and divergence in *D. melanogaster.* (A) Genomic location has a significant effect on silent-site heterozygosity ($P_{MWU} < 0.001$ for all pairwise contrasts), consistent with predicted differences in N_e between these chromosomes. The ratio of median mitochondrial to autosomal silent site heterozygosity was 0.157, less than predicted for neutral sites if mitochondrial N_e is one quarter that of the autosomes. mtDNA, X chromosome, and autosome data sets contained 12, 1255, and 8073 genes, respectively. (B-D) Distributions of neutrality indices *NI*, *Z** and *DoS* are similar between mitochondrial and nuclear genes despite differences in N_e. Dashed lines represent the neutral expectations for these statistics. Three nuclear loci for which *NI* exceeded 50 were excluded from (B) to improve visualization. See main text for description of statistics, number of genes in each dataset and statistical results.

Figure 2 Site-frequency spectra of synonymous and nonsynonymous polymorphisms in the *D. melanogaster* mtDNA. (A, B) Folded site-frequency spectra for synonymous and nonsynonymous segregating sites across the mitochondrial coding region reveal that mitochondrial polymorphisms are skewed to low frequencies and are more so for nonsynonymous sites. (C, D) Unfolded site-frequency spectra reveal that derived, synonymous polymorphisms are almost equally likely to be at low frequency (56% of 59 sites at frequencies less than 0.11) or nearly fixed (44% of 59 sites at frequencies greater than 0.89), while derived, nonsynonymous polymorphisms are nearly always present as singletons (94% of 32 sites). There are essentially no mitochondrial polymorphisms at intermediate frequencies. Sites were omitted from the unfolded site frequency spectra if neither allelic state was shared with *D. yakuba*. The number of sites included in each distribution is 67 (A), 35 (B), 59 (C), and 32 (D).

Figure 3 Haplotype network for 38 *D. melanogaster* mtDNAs sampled from Raleigh, NC. The network, inferred from 80 coding region SNPs with no missing information, reveals that nearly 30% of individuals sampled (11/38) share the same common haplotype (red) and an additional 65% of individuals carry a

haplotype only a few mutations away from this haplotype. This common set of mitochondrial haplotypes is highly diverged from the two other mtDNAs sampled in the population; lines RAL-639 and RAL-335 differ from the common haplotype at 14 and 34 SNPs, respectively. At least one of these two haplotypes carries the ancestral state (shared with *D. yakuba*) at 38% of these SNPs. Numbers represent the Raleigh (DGRP) line carrying the haplotype. Red, yellow, blue and white nodes were present in 11, 3, 2 and 1 lines, respectively.

Figure 4 Distributions of polymorphism and divergence summary statistics for mitochondrial and nuclear genes in flies and in humans. (A) Distributions of *NI* are similar between genomes within species and across species. A value of 1 was added to each cell if any cell contained a count of zero for any given gene. Values of *NI* greater than 50 were removed to improve visualization (3 nuclear genes for flies and 2 nuclear genes for humans). Distributions of (B) *Z** and (C) *DoS* for mitochondrial genes are also similar to those for nuclear genes in both flies and humans (see text for statistical contrasts). Positive values of these summary statistics of the MK contingency table are consistent with positive selection and negative values with negative selection. *DoS* was not calculated for genes with zero values in any cell of the MK table. Dashed lines represent the neutral expectation for each statistic, and the numbers of genes in each set are indicated in parentheses.

Figure 1

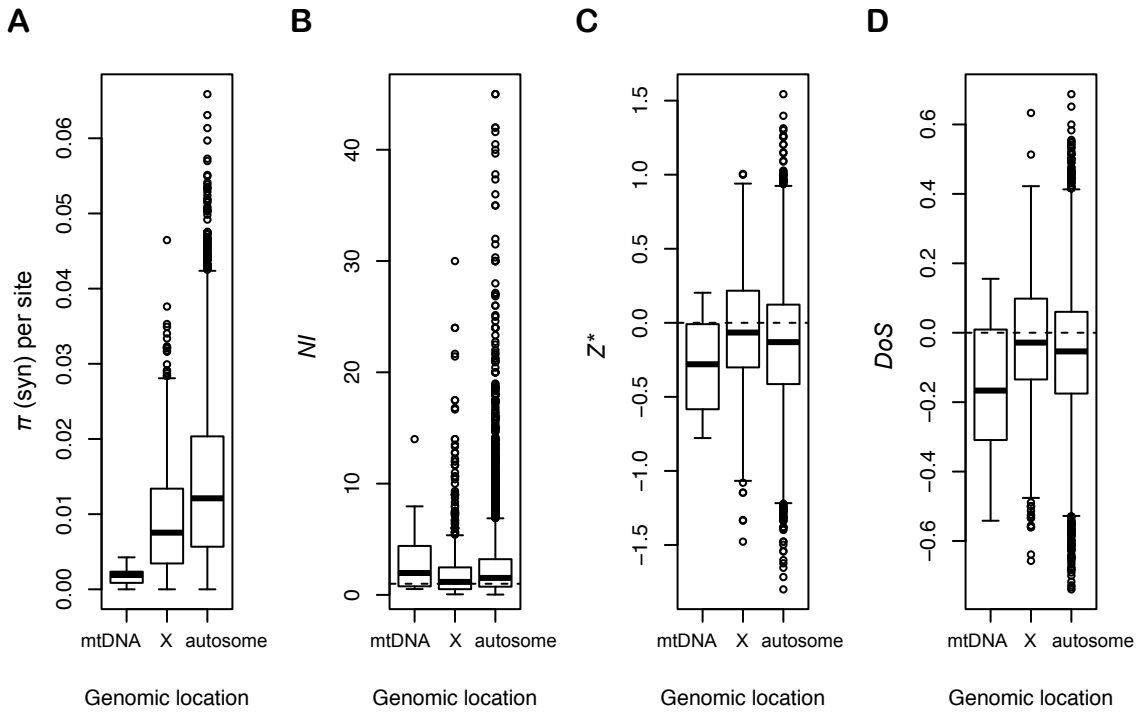

A
B
C
D
π (syn) per site
NI
Z*
DoS
mtDNA
X
autosome
Genomic location

Figure 2

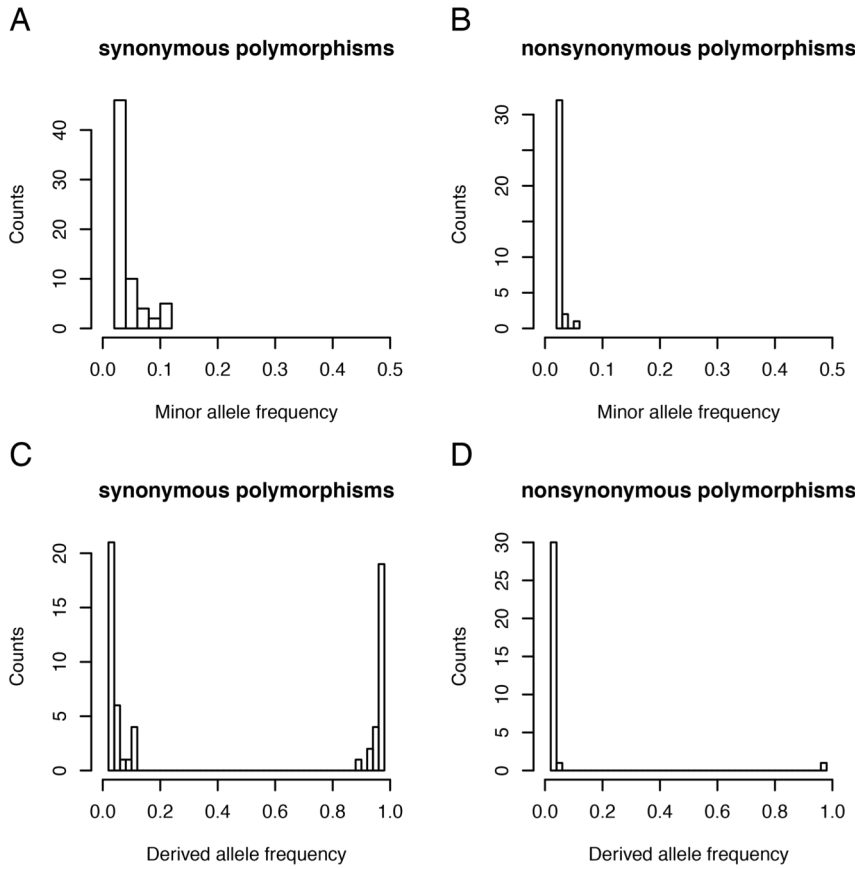

A
synonymous polymorphisms
Counts
0
10
20
30
40
0.0
0.1
0.2
0.3
0.4
0.5
Minor allele frequency
B
nonsynonymous polymorphisms
Counts
0
5
10
20
30
0.0
0.1
0.2
0.3
0.4
0.5
Minor allele frequency
C
synonymous polymorphisms
Counts
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0
Derived allele frequency
D
nonsynonymous polymorphisms
Counts
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
Derived allele frequency

Figure 3

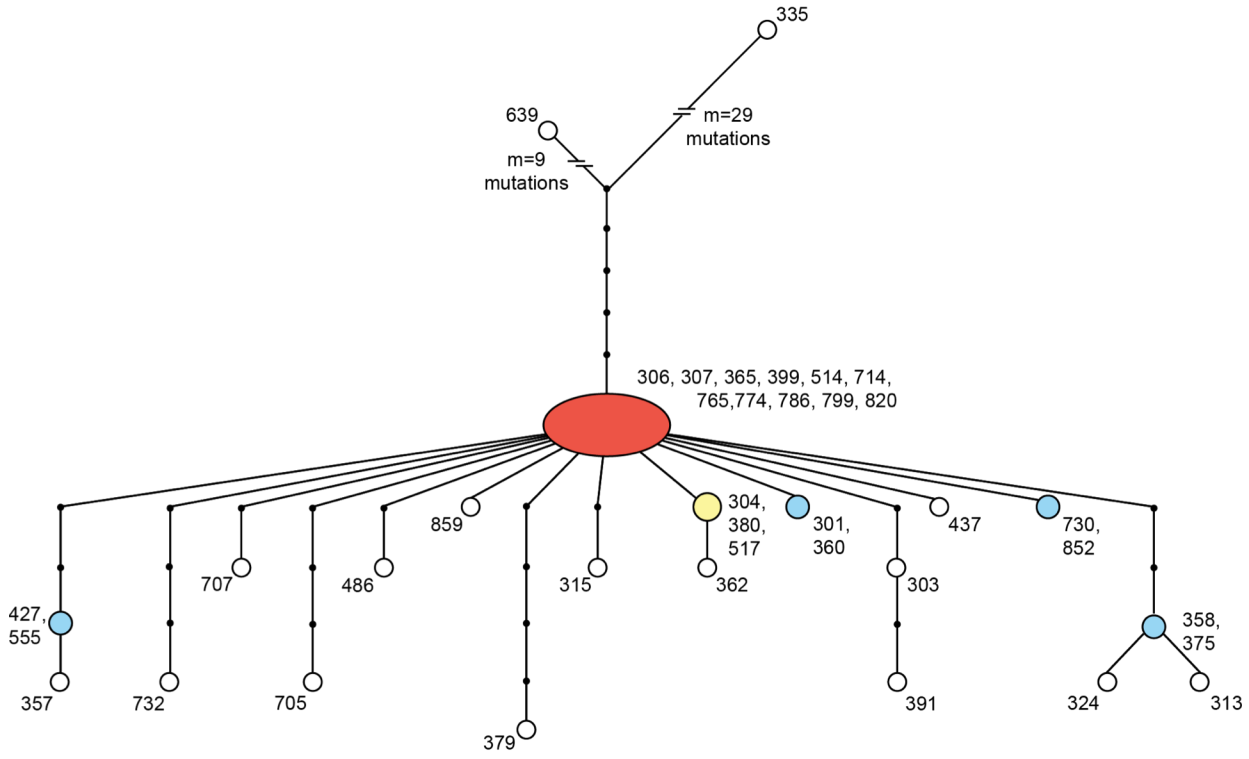
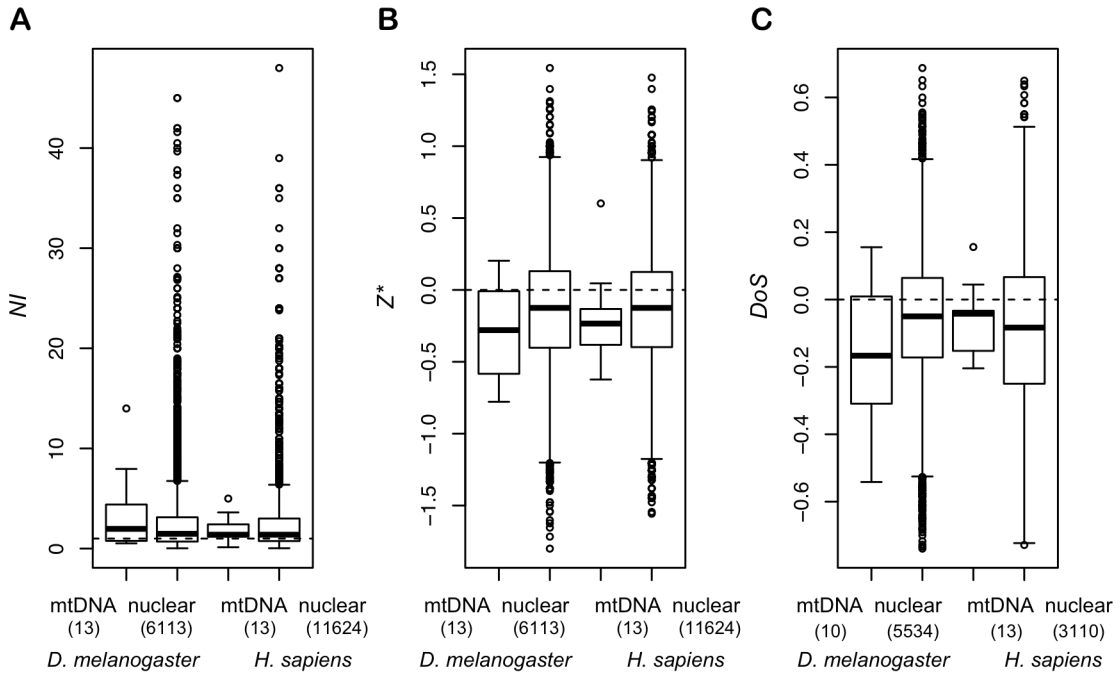

335
639
m=29
mutations
m=9
mutations
306, 307, 365, 399, 514, 714,
765,774, 786, 799, 820
859
304,
380,
517
301,
360
437
730,
852
707
486
315
362
303
427,
555
358,
375
357
732
705
379
391
324
313

Figure 4


A
B
C
NI
Z*
DoS
mtDNA
nuclear
(13)
(6113)
(11624)
(10)
(5534)
(3110)
D. melanogaster
H. sapiens

Supporting Information

Table S1 mtDNAs assembled in this study along with the average coverage

DGRP line	SRA Accession	Average coverage
RAL-301	SRX000530	103
RAL-303	SRX000529	38
RAL-304	SRX000531	19
RAL-306	SRX000532	14
RAL-307	SRX000533	26
RAL-313	SRX022270	57
RAL-315	SRX000535	23
RAL-324	SRX010933	10
RAL-335	SRX022273	32
RAL-357	SRX022274	28
RAL-358	SRX000536	42
RAL-360	SRX000534	28
RAL-362	SRX022277	107
RAL-365	SRX000537	18
RAL-375	SRX000538	32
RAL-379	SRX000539	44
RAL-380	SRX000556	65
RAL-391	SRX000557	23
RAL-399	SRX000558	38
RAL-427	SRX000528	146
RAL-437	SRX010938	55
RAL-486	SRX022286	21
RAL-514	SRX022287	131
RAL-517	SRX022288	142
RAL-555	SRX022289	22
RAL-639	SRX022290	13
RAL-705	SRX022291	6
RAL-707	SRX022292	12
RAL-714	SRX022294	27
RAL-730	SRX022295	54
RAL-732	SRX022296	20
RAL-765	SRX022297	14
RAL-774	SRX022298	119
RAL-786	SRX022299	11
RAL-799	SRX022300	192

RAL-820	SRX022301	158
RAL-852	SRX022302	159
RAL-859	SRX010956	36

Table S2 Correlations between linkage disequilibrium (LD) and distance (bp) between pairs of SNPs in the *D. melanogaster* mtDNA

Measure of LD	MAF cutoff[a]	df	Pearson's *r*	*P*-value
D'	none	11552	0.007	0.451
R^2	none	10821	-0.007	0.457
D'	no singletons	1822	0.014	0.560
R^2	no singletons	1393	0.035	0.194
D'	0.10	43	-0.080	0.599
R^2	0.10	39	-0.100	0.535

[a] Minor allele frequency cutoff below which SNPs were not included

Table S3 Counts of polymorphic (*P*) and divergent (*D*) nonsynonymous (*N*) and synonymous (*S*) sites along with summary statistics of the McDonald-Kreitman table using 36 of the 38 mitochondrial haplotypes in our sample that were independently sequenced and assembled by Richardson et al. (2012).

Gene	P_N	P_S	D_N	D_S	NI^a	$Z^{*\,b}$	DoS^c
ATPase6	5	2	11	35	7.95	-0.78	-0.48
ATPase8	1	0	2	8	6.00	-0.78	undef
COI	1	16	8	101	0.79	-0.12	0.01
COII	0	5	6	39	0.95	0.02	undef
COIII	1	9	8.5	47.5	0.62	-0.01	0.05
CytB	2	5	17.5	67.5	1.54	-0.27	-0.08
NDI	2	3	11	45	2.73	-0.46	-0.20
NDII	1	5	23	39	0.34	0.26	0.20
NDIII	0	1	5	22	1.92	-0.28	undef
NDIV	1	6	23	63	0.46	0.12	0.12
NDIVL	1	1	1	7	7.00	-0.60	-0.38
NDV	2	10	61.83	107.17	0.35	0.33	0.20
NDVI	0	4	21.5	22.5	0.21	0.68	undef

[a] For *ATPase8*, *COII, NDIII* and *NDVI*, a count of 1 was added to each cell when calculating $NI = \frac{D_S P_N}{D_N P_S}$. For no gene was the neutral expectation rejected via a Fisher's exact test of the MK table at the table-wise Bonferroni corrected *P*=0.004.

[b] $Z^* = Log\left(\frac{(D_N+1)(P_S+1)}{(D_S+1)(P_N+1)}\right)$, as in (Presgraves 2005).

[c] $DoS = \frac{D_N}{D_N+D_S} - \frac{P_N}{P_N+P_S}$ was not calculated for loci with a count of zero (Stoletzki and Eyre-Walker 2011).

Table S4 Effects of sampling on *NI*

A. 1 African, 1 European, 1 Japanese (Nachman *et al.* 1996)

Gene	*Pn*	*Ps*	*Dn*	*Ds*	Pseudo[a]	*NI*[b]	*Z**
ATPase	4	6	17	58	no	2.27	-0.37
COI	3	5	6	125	no	**12.5**	**-1.08**
COII	1	3	6	60	yes	3.33	-0.52
COIII	1	4	7	66	no	2.36	-0.53
Cytb	5	5	27	102	no	**3.78**	**-0.57**
ND1	2	5	18	71	no	1.58	-0.28
ND2	3	4	11	86	no	**5.86**	**-0.76**
ND3	1	2	6	34	no	2.83	-0.52
ND4	2	10	23	107	no	0.93	-0.09
ND4L	1	5	2	22	yes	2.20	-0.34
ND5	5	7	51	143	no	2.00	-0.32
ND6	5	1	7	43	no	**30.7***	**-1.22***
All	31	55	179	915	no	**2.88***	**-0.46***

B. 1 African, 1 European, 1 Japanese

Gene	*Pn*	*Ps*	*Dn*	*Ds*	Pseudo[a]	*NI*[b]	*Z**
ATPase	4	5	17	59	no	2.78	-0.44
COI	2	6	6	124	no	6.89	-0.88
COII	1	2	5	58	no	5.80	-0.82
COIII	1	6	8	66	yes	1.38	-0.14
Cytb	4	4	28	102	no	3.64	-0.55
ND1	3	5	17	71	no	2.51	-0.43
ND2	3	4	11	86	no	**5.86**	**-0.76**
ND3	2	2	5	34	no	6.80	-0.77
ND4	1	9	23	107	no	0.52	0.05
ND4L	1	4	2	21	yes	2.63	-0.42
ND5	4	5	49	146	no	2.38	-0.39
ND6	1	1	8	44	yes	5.50	-0.74
All	24	50	176	915	no	**2.50***	**-0.40***

C. Change Japanese sequence

Pn	*Ps*	*Dn*	*Ds*	Pseudo[a]	*NI*[b]	*Z**
5	6	17	58	no	2.84	-0.45
2	4	6	124	no	**10.33**	**-1.03**
1	3	6	59	yes	3.28	-0.52
1	6	7	67	no	1.60	-0.39
4	4	28	102	no	3.64	-0.55
2	3	18	71	no	2.63	-0.45
4	4	11	85	no	**7.73**	**-0.86**
1	2	6	34	no	2.83	-0.52
1	9	23	107	no	0.52	0.05
1	4	2	21	yes	2.63	-0.42
3	4	49	146	no	2.23	-0.37
1	3	8	43	yes	1.79	-0.25
23	49	178	914	no	**2.41***	**-0.39***

	D. 19 African-American/20 European-American							E. 30 African-American/30 European-American						
Gene	*Pn*	*Ps*	*Dn*	*Ds*	Pseudo[a]	*NI*[b]	*Z**	*Pn*	*Ps*	*Dn*	*Ds*	Pseudo[a]	*NI*[b]	*Z**
ATPase	7	14	17	55	no	1.62	-0.22	9	17	17	53	no	1.65	-0.22
CO!	5	27	6	117	no	**3.61**	**-0.56**	6	31	6	114	no	**3.68**	**-0.56**
COII	1	13	5	56	no	0.86	-0.13	1	17	5	54	no	0.64	-0.01
COIII	2	13	7	63	no	1.38	-0.23	2	18	7	60	no	0.95	-0.08
Cytb	12	17	28	91	no	2.29	-0.36	18	19	27	92	no	**3.23***	**-0.50***
ND1	5	15	15	67	no	1.49	-0.20	6	18	15	67	no	1.49	-0.19
ND2	8	17	11	84	no	**3.59**	**-0.55**	10	25	11	81	no	**2.95**	**-0.46**
ND3	1	4	6	32	no	1.33	-0.28	2	7	6	32	no	1.52	-0.25
ND4	1	33	23	101	no	**0.13**	**0.60**	1	41	23	100	no	**0.11**	**0.70**
ND4L	1	4	1	20	no	5.00	-0.62	1	5	1	20	no	4.00	-0.54
ND5	12	30	46	138	no	1.20	-0.09	18	39	43	134	no	1.44	-0.16
ND6	2	9	7	41	no	1.30	-0.20	2	10	7	41	no	1.17	-0.16
All	57	196	172	865	no	**1.46**	**-0.17**	76	247	168	848	no	**1.55**	**-0.19**

[a] A count of one was added to all cells in the MK table when indicated by yes.
[b] **Bold** indicates $P \leq 0.05$; ***** indicates significant a sample-wise Bonferroni-corrected *P*-value of less than 0.004 for Fisher's exact test of the MK table

Table S5 Accession numbers for mtDNAs used in Table S4

Accessions used in Table S3B

Population	Accession number
African	D38112
European	NC_012920
Japanese	AP008310

Accessions used in Table S3C

Population	Accession number
African	D38112
European	NC_012920
Japanese	AF346989

Accessions used in Table S3D

African-American	European-American
DQ304898	EU670874
DQ304902	EU714298
DQ304906	EU714300
DQ304916	EU747355
DQ304918	EU862197
DQ304921	EU882063
DQ304924	EU914954
DQ304928	EU919746
DQ304933	FJ156761
DQ304938	FJ190383
DQ304940	FJ664616
DQ304944	FJ705809
DQ304945	FJ842500
DQ304957	FJ858802
DQ304958	FJ866786
DQ304963	FJ984932
DQ304964	FJ985851
DQ304965	GQ175058
DQ304966	GQ200588
	GQ249257

Additional accessions used in Table S3E

African-American	European-American
DQ304971	GQ332765
DQ304992	GU045487
DQ304993	GU147938
DQ305000	GU198193
DQ305001	GU252762
DQ305002	GU295665
DQ305004	GU361772
DQ305009	GU371909
DQ305014	GU390312
DQ305017	GU433215
DQ305028	

References

Nachman, M. W., W. M. Brown, M. Stoneking, and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142**:** 953-963.

Presgraves, D., 2005 Recombination enhances protein adaptation in Drosophila melanogaster. Curr Biol 15**:** 1651-1656.

Richardson, M. F., L. A. Weinert, J. J. Welch, R. S. Linheiro, M. M. Magwire *et al.*, 2012 Population Genomics of the Wolbachia Endosymbiont in Drosophila melanogaster. PLoS Genet 8**:** e1003129.

Stoletzki, N., and A. Eyre-Walker, 2011 Estimation of the neutrality index. Mol Biol Evol 28**:** 63-70.