G3 · Genes | Genomes | Genetics

Similar efficacies of selection shape mitochondrial and nuclear genes in both *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
[†] School of Informatics and Computing, Indiana University, Bloomington, IN 47405

[1] Current address: Department of Evolution and Ecology, University of California, Davis, CA 95616
[2] Corresponding author, current address: School of Biological Sciences, University of Nebraska, Lincoln, NE 68588. E-mail: kmontooth2@unl.edu

DOI: 10.1534/g3.114.016493

File S1

Supporting Data: FASTA file containing the 38 *Drosophila melanogaster* assembled mtDNA genomes used in this study aligned with *D. yakuba* (NC_001322)

Available for download as a .fasta file at www.g3journal.org/lookup/suppl/doi:10.1534/g3.114.016493/-/DC1

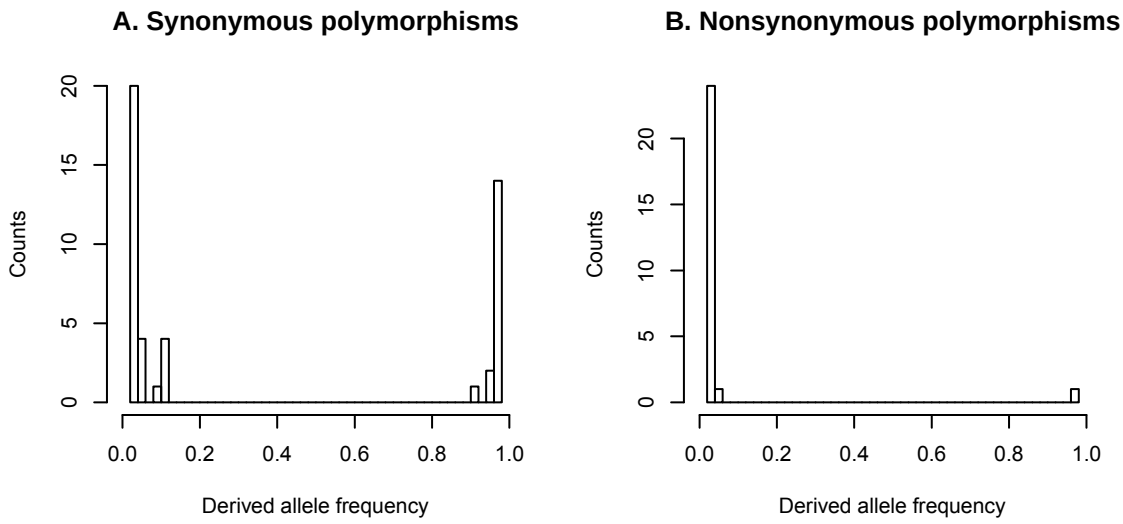

Figure S1 Unfolded site-frequency spectra of synonymous (A) and nonsynonymous (B) polymorphisms in the *D. melanogaster* mtDNA, using *D. simulans* as a close outgroup and *D. yakuba* as a distant outgroup to polarize mutations. 37% of derived synonymous polymorphisms were present at frequencies greater than 0.92, while only a single derived nonsynonymous polymorphism was at high frequency. Distributions contain 46 (A) and 26 (B) segregating sites.

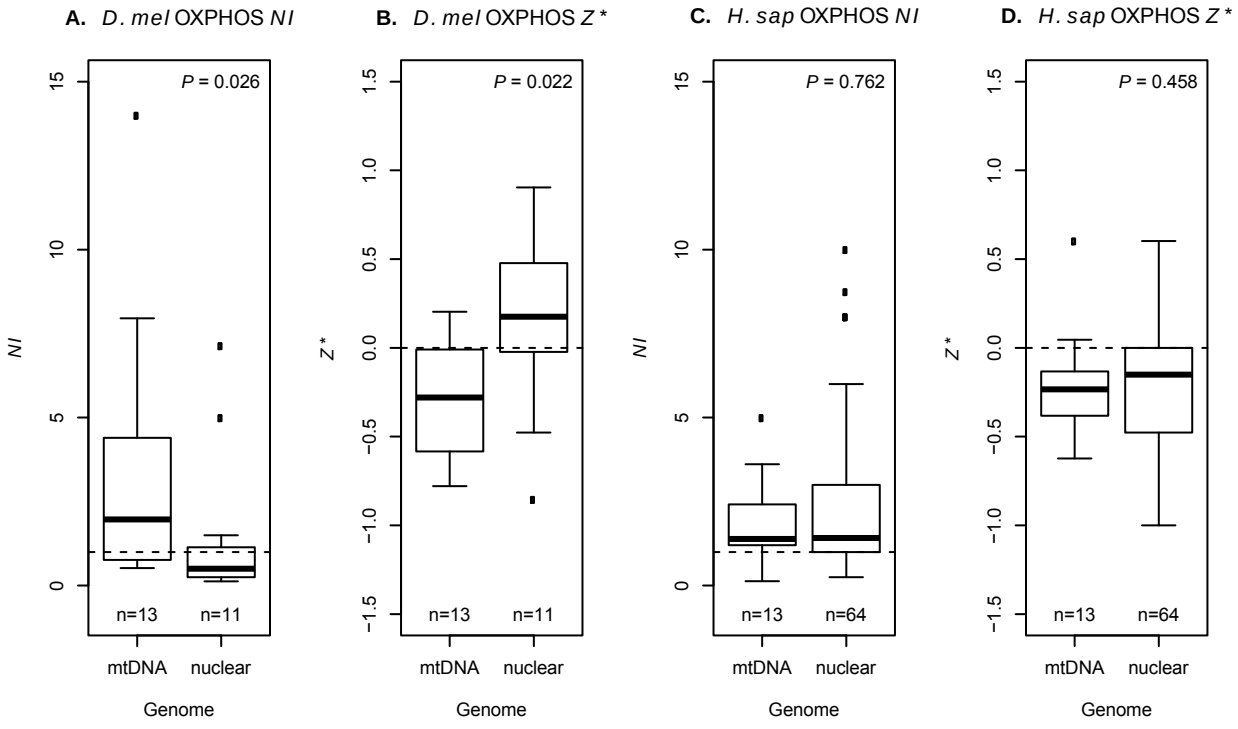

Figure S2 Contrasts of polymorphism and divergence summary statistics between mitochondrial genes and nuclear genes in our data set that are annotated to have OXPHOS function. *P*-values are from Mann-Whitney *U* tests. Dashed lines represent the neutral expectation for each statistic and the number of genes in each set (n) is provided. Mitochondrial *NI* and *Z** are calculated from the "more inclusive" method using *D. yakuba* (A, B) and chimpanzee (C,D) as outgroups.

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 ***D. melanogaster*** **MK tables calculated using different methods**

A. Changes along the ***D. melanogaster*** **branch** [a]

Gene	P_N	P_S	D_N	D_S	NI [b]	Z^* [c]	P_{FET} [d]
ATPase6	5	1	2	10	25.000	-1.041	0.013
ATPase8	1	0	0	5	12.000	-1.079	0.167
COI	0	12	3	23	0.462	0.336	0.538
COII	0	3	2	11	1.000	0.000	1.000
COIII	1	6	3	13	0.722	0.000	1.000
Cyt-b	2	2	1	18	18.000	-0.978	0.067
ND1	1	3	1	13	4.333	-0.544	0.405
ND2	3	3	3	8	2.667	-0.352	0.600
ND3	1	0	0	6	14.000	-1.146	0.143
ND4	6	5	4	11	3.300	-0.447	0.228
ND4L	2	1	1	4	8.000	-0.574	0.464
ND5	4	8	9	32	1.778	-0.263	0.459
ND6	0	2	3	3	0.333	0.477	0.464
Sum	26	46	32	157	2.773	-0.439	0.001
NI_{TG} [e]	2.57 (1.27, 5.08)						

	mt v. nuc [f]	mt v. X [f]	mt v. auto [f]
NI	0.045	0.011	0.054
*Z**	0.037	0.01	0.043

B. ***D. melanogaster-D. simulans*** **(less inclusive)** [g]

Gene	P_N	P_S	D_N	D_S	NI [b]	Z^* [c]	P_{FET} [d]
ATPase6	5	3	4	18	7.500	-0.756	0.032
ATPase8	1	0	1	9	10.000	-1.000	0.182
COI	0	13	4	78	1.130	-0.053	1.000
COII	0	5	6	23	0.570	0.243	0.558
COIII	1	7	6	33	0.786	-0.084	1.000
Cyt-b	2	4	1	48	24.000	-1.167	0.029
ND1	3	3	4	31	7.750	-0.806	0.050
ND2	3	4	11	29	1.977	-0.301	0.657
ND3	1	0	1	16	17.000	-1.230	0.111
ND4	6	5	10	38	4.560	-0.617	0.054
ND4L	2	1	1	9	18.000	-0.875	0.108
ND5	4	10	19	59	1.242	-0.135	0.744
ND6	0	3	5	8	0.375	0.426	0.509
Sum	28	58	73	399	2.640	-0.424	0.0004
NI_{TG} [e]	2.39 (1.30, 4.39)						

	mt v. nuc [f]	mt v. X [f]	mt v. auto [f]
NI	0.039	0.009	0.05
*Z**	0.02	0.005	0.02

C. ***D. melanogaster-D. simulans*** **(more inclusive)** [g]

Gene	P_N	P_S	D_N	D_S	NI [b]	Z^* [c]	P_{FET} [d]
ATPase6	5	2	5	17	8.500	-0.778	0.030
ATPase8	1	0	1	9	10.000	-1.000	0.154
COI	0	16	6	77	0.655	0.183	1.000
COII	1	5	6.5	22.5	0.692	-0.019	1.000
COIII	1	9	6.5	34.5	0.590	0.024	1.000
Cyt-b	2	5	2	47	9.400	-0.903	0.072
ND1	3	3	5	33	6.600	-0.753	0.063
ND2	6	5	15	29	2.320	-0.340	0.300
ND3	1	1	1	17	17.000	-0.954	0.195
ND4	7	7	13	40	3.077	-0.467	0.099
ND4L	2	1	1	9	18.000	-0.875	0.108
ND5	4	10	26.5	59.5	0.898	0.000	1.000
ND6	2	4	7	9	0.643	0.125	1.000
Sum	35	68	95.5	403.5	2.170	-0.340	0.002
NI_{TG} [e]	2.02 (1.22,3.58)						

	mt v. nuc [f]	mt v. X [f]	mt v. auto [f]
NI	0.075	0.018	0.089
*Z**	0.028	0.009	0.033

D. *D. melanogaster-D. yakuba* (less inclusive) [g]								**E. *D. melanogaster-D. yakuba* (more inclusive)** [g]						
Gene	P_N	P_S	D_N	D_S	NI [b]	Z^* [c]	P_{FET} [d]	P_N	P_S	D_N	D_S	NI [b]	Z^* [c]	P_{FET} [d]
ATPase6	5	3	6	40	11.111	-0.944	0.006	5	2	11	35	7.955	-0.778	0.021
ATPase8	1	0	2	8	6.000	-0.778	0.273	1	0	2	8	6.000	-0.778	0.273
COI	0	13	5	98	1.180	-0.071	1.000	0	16	8	101	0.667	0.176	0.595
COII	0	5	5	39	1.110	-0.046	1.000	1	5	6	39	1.300	-0.280	1.000
COIII	1	7	5	50	1.429	-0.327	0.573	1	9	8.5	47.5	0.621	-0.009	1.000
Cyt-b	2	4	12	73	3.042	-0.533	0.229	2	5	17.5	67.5	1.543	-0.267	0.641
ND1	3	3	9	45	5.000	-0.663	0.088	3	3	11	45	4.091	-0.584	0.122
ND2	3	4	15	45	2.250	-0.362	0.375	6	5	25	41	1.968	-0.275	0.334
ND3	1	0	4	21	8.800	-0.944	0.192	1	1	5	22	4.400	-0.584	0.377
ND4	6	5	15	65	5.200	-0.682	0.016	7	7	24	63	2.625	-0.408	0.120
ND4L	2	1	1	7	14.000	-0.778	0.152	2	1	1	7	14.000	-0.778	0.152
ND5	4	10	45	109	0.969	-0.036	1.000	4	10	55.83	107.17	0.768	0.063	0.775
ND6	0	3	14	21	0.367	0.436	0.283	2	4	21.5	22.5	0.523	0.203	0.669
Sum	28	58	138	621	2.172	-0.342	0.004	35	68	196.3	605.67	1.588	-0.205	0.041
NI_{TG} [e]	2.29 (1.26, 4.71)							1.67 (1.05, 2.86)						

	mt v. nuc [f]	mt v. X [f]	mt v. auto [f]	mt v. nuc [f]	mt v. X [f]	mt v. auto [f]
NI	0.038	0.007	0.047	0.277	0.065	0.325
*Z**	0.01	0.002	0.012	0.087	0.019	0.104

[a] Changes were assigned to the *D. melanogaster* branch using *D. simulans* as the close outgroup and *D. yakuba* as the distant outgroup. Codons with missing data and mitochondrial genome SRX022291 were omitted from the analysis.

[b] A count of one was added to each cell in order to calculate $NI = \frac{D_S P_N}{D_N P_S}$ when any single cell had a count of zero.

[c] $Z^* = Log_{10}\left(\frac{(D_N+1)(P_S+1)}{(D_S+1)(P_N+1)}\right)$, as in Presgraves (2005).

[d] *P*-value from Fisher's Exact Test of the MK table.

[e] $NI_{TG} = \frac{\sum D_{Si} P_{Ni}/(P_{Si}+D_{Si})}{\sum P_{Si} D_{Ni}/(P_{Si}+D_{Si})}$ with confidence intervals from 5000 bootstrap samples (Stoletzki and Eyre-Walker 2011).

[f] *P*-values from Mann Whitney *U* tests contrasting statistics between the mitochondrial gene set and either the entire nuclear (nuc), X chromosome (X), or autosomal (A) gene sets.

[g] Divergence was estimated between *D. melanogaster* and *D. simulans* or between *D. melanogaster* and *D. yakuba*. The less inclusive analysis omitted codons with missing data in any genome, did not include mitochondrial genome SRX022291, and used a conservative criterion that minimized D_N for codons with multiple substitutions. The more inclusive analysis did not omit codons with missing data, included mitochondrial genome SRX022291, and averaged across all possible mutational paths between codons with multiple substitutions.

Table S3 Summary statistics of the MK table for mitochondrially encoded OXPHOS complexes in *D. melanogaster*

Species	Gene set [a]	*NI* [b]	NI_{TG} [c]	*Z* [b]	P_{FET} [d]
D. melanogaster – D. yakuba	Complex I	1.731	1.59 (0.94, 3.17)	-0.238	0.070
(more inclusive)	Complex IV	0.556	0.55 (0, 1.30)	0.255	0.750
	Complex V	9.923	9.64 (undefined)	-0.997	0.007
D. melanogaster – D. yakuba	Complex I	2.221	2.06 (0.98, 5.38)	-0.346	0.019
(less inclusive)	Complex IV	0.499	0.50 (0, 1.43)	0.302	1.000
	Complex V	12.000	13.5 (undefined)	-1.079	0.002
D. melanogaster – D. simulans	Complex I	2.313	2.00 (1.17, 4.04)	-0.364	0.006
(more inclusive)	Complex IV	0.470	0.45 (0, 0.69)	0.328	0.539
	Complex V	13.000	10.4 (undefined)	-1.114	0.005
D. melanogaster – D. simulans	Complex I	2.722	2.39 (1.18, 5.25)	-0.435	0.004
(less inclusive)	Complex IV	0.335	0.31 (0, 0.79)	0.475	0.474
	Complex V	10.800	9.25 (undefined)	-1.033	0.006

[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, *Cyt-b.*

[b] $NI = \frac{D_S P_N}{D_N P_S}$ and $Z = Log_{10}\left(\frac{D_N P_S}{D_S P_N}\right)$ were calculated using counts of P_N, P_S, D_N and D_S summed across genes within gene sets.

[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).

[d] *P*-value for Fisher's Exact Test of the MK table containing counts of P_N, P_S, D_N and D_S summed across genes within gene sets.

Table S4 Counts of polymorphic (*P*) and divergent (*D*) nonsynonymous (*N*) and synonymous (*S*) sites along with summary statistics of the MK table using 36 of the 38 mitochondrial haplotypes in our sample that were independently sequenced and assembled by Richardson *et al*. (2012). MK counts are from the "more inclusive" method.

Gene	P_N	P_S	D_N	D_S	NI [a]	Z^* [b]	P_{FET} [c]
ATPase6	5	2	11	35	7.95	-0.78	0.021
ATPase8	1	0	2	8	6.00	-0.78	0.273
COI	1	16	8	101	0.79	-0.12	1.000
COII	0	5	6	39	0.95	0.02	1.000
COIII	1	9	8.5	47.5	0.62	-0.01	1.000
Cyt-b	2	5	17.5	67.5	1.54	-0.27	0.641
ND1	2	3	11	45	2.73	-0.46	0.287
ND2	1	5	23	39	0.34	0.26	0.413
ND3	0	1	5	22	1.92	-0.28	1.000
ND4	1	6	23	63	0.46	0.12	0.673
ND4L	1	1	1	7	7.00	-0.60	0.378
ND5	2	10	61.83	107.17	0.35	0.33	0.218
ND6	0	4	21.5	22.5	0.21	0.68	0.114

[a] A count of 1 was added to each cell when calculating $NI = \frac{D_S P_N}{D_N P_S}$ for any locus with a zero count in any cell.

[b] $Z^* = Log_{10}\left(\frac{(D_N+1)(P_S+1)}{(D_S+1)(P_N+1)}\right)$, as in (Presgraves 2005).

[c] *P*-value from Fisher's Exact Test of the MK table.

Table S5 Effects of sampling on *NI*

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

Gene	P_N	P_S	D_N	D_S	Pseudo[a]	*NI*	*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
Cyt-b	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	P_N	P_S	D_N	D_S	Pseudo[a]	*NI*	*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
Cyt-b	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

Gene	P_N	P_S	D_N	D_S	Pseudo[a]	*NI*	*Z**
ATPase	5	6	17	58	no	2.84	-0.45
COI	2	4	6	124	no	**10.33**	**-1.03**
COII	1	3	6	59	yes	3.28	-0.52
COIII	1	6	7	67	no	1.60	-0.39
Cyt-b	4	4	28	102	no	3.64	-0.55
ND1	2	3	18	71	no	2.63	-0.45
ND2	4	4	11	85	no	**7.73**	**-0.86**
ND3	1	2	6	34	no	2.83	-0.52
ND4	1	9	23	107	no	0.52	0.05
ND4L	1	4	2	21	yes	2.63	-0.42
ND5	3	4	49	146	no	2.23	-0.37
ND6	1	3	8	43	yes	1.79	-0.25
All	23	49	178	914	no	**2.41***	**-0.39***

D. 19 African-American/20 European-American								E. 30 African-American/30 European-American						
Gene	P_N	P_S	D_N	D_S	Pseudo[a]	*NI*	*Z**	P_N	P_S	D_N	D_S	Pseudo[a]	*NI*	*Z**
ATPase	7	14	17	55	no	1.62	-0.22	9	17	17	53	no	1.65	-0.22
COI	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
Cyt-b	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 has been added to all cells in the MK table when indicated by yes.

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 S6 Accession numbers for mtDNAs used in Table S5

Accessions used in Table S5B

Population	Accession number
African	D38112
European	NC_012920
Japanese	AP008310

Accessions used in Table S5C

Population	Accession number
African	D38112
European	NC_012920
Japanese	AF346989

Accessions used in Table S5D

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 S5E

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	

Table S7 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

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.