**Supplementary Materials**

**Mitochondrial DNA fitness depends on nuclear genetic background in *Drosophila***

Mossman J.A.\*, §, Ge Y.\*, Navarro F.\* and Rand D.M.\* ,§

\*Department of Ecology and Evolutionary Biology, 80 Waterman Street, Box G, Brown University, Providence, Rhode Island 02912, USA

§ Corresponding authors: ([jim\_mossman@brown.edu](mailto:jim_mossman@brown.edu)) ([David\_Rand@Brown.edu](mailto:David_Rand@Brown.edu))

**Corresponding authors’ address:**

Department of Ecology and Evolutionary Biology, 80 Waterman Street, Box G, Brown University, Providence, Rhode Island 02912, USA

Contents

[(i) SNP variation in our laboratory stocks 3](#_Toc360596)

[(ii) Summary of population sizes in cages 7](#_Toc360597)

[(iii) Supporting References 9](#_Toc360598)

## (i) SNP variation in our laboratory stocks

We estimated the number of SNPs in four out of ten genotypes used in this study, following a previous method ([Mossman et al. 2019](#_ENREF_3" \o "Mossman, 2019 #5240)). These genotypes were: *OreR;OreR*, *siI;OreR*, *OreR;Aut*, *siI;Aut* and the sequencing data were obtained from RNA-seq dataset from a previous experiment ([Mossman et al. 2016](#_ENREF_5); [Mossman et al. 2017](#_ENREF_4)). The rationale for using RNA-seq data to screen for SNPs is that we can estimate the number of variants in the samples transcriptome-wide with much higher resolution than microsatellite or targeted SNP genotyping.

For each genotype we selected 3 x RNA-seq libraries (technical replicates), each corresponding to a pool of 30 male flies that were 5 days old at the time of the experiment. The 30 males were snap-frozen in liquid nitrogen prior to RNA extraction. mRNA was extracted using Direct Dynabeads (Invitrogen) following manufacturer’s instructions, followed by RNA Fragmentation (RNA Fragmentation Kit: Ambion). cDNA first and second strand synthesis followed mRNA extraction using random hexamer primer (Invitrogen) and Super Script III Reverse transcriptase Kit (Invitrogen). For Illumina RNA-seq we performed end-repair using the End-It DNA End Repair Kit (Epicentre), following manufacturer’s instructions, then addition of A-bases to 3’ ends using Klenow fragment (3’->5’ exo-). We then adapter ligated and PCR-enriched our libraries. Prior to pre-sequencing quality control, we size selected using LabChip XT DNA 750 Chip (Perkin Elmer) then performed 50bp single end RNA-seq on the Illumina HiSeq platform at Brown University’s Genomics Core Facility.

After sequencing we performed quality control on the sequence data. Fastq files were first filtered of reads with more than 80% of bases scoring less than Q20 (phred scaled 99% accuracy) using FastQC v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Adapters were clipped using FastX-Toolkit v2.6 (http://hannonlab.cshl.edu/fastx\_toolkit/) and sequences were aligned to the reference *Drosophila melanogaster* (dm3) genome along with splice junction mapping using Tophat v2.0.8b ([Trapnell et al. 2012](#_ENREF_6)). Finally, reads were counted using HTseq-count ([Anders et al. 2015](#_ENREF_1)).

*Pair and trio analyses*

We next estimated the numbers of SNPs that were segregating in our stock populations using pair and trioauto analyses, implemented in bcftools (samtools v.1.1.19) ([Li et al. 2009](#_ENREF_2)). Pair analyses detect differences in the allele calls across the genome between two datasets (in this case, different mtDNA haplotypes on the same nuclear background). For the pairs analyses three replicate libraries (.bam files) were merged to form one large library representing that genotype (Figure S1). The rationale for trio analysis is to highlight variants that are not consistent with Mendelian inheritance and likely to be *de novo* mutations. In other words, the offspring genotypes are not consistent with parent genotypes. In the trio analyses, the replicates remained separate and took the position of either a parent or the offspring in the trio.

The trio analysis outputs a list of SNPs that are incompatibly inherited between putative parents and offspring- in this case, the three RNA-seq libraries of each genotype. In the example shown, Libraries #2 and #3 were parents and Library #1 was the putative offspring (see Figure S1). For clarity, no experimental crosses were performed and here we report a modeling exercise based on RNA-seq data. The assignment of ‘parent’ or ‘offspring’ does not change the results of the trio analysis.



**Figure S1. Pairs and Trios experimental scheme.**

In the trio analyses, phred log ratios of genotype likelihoods with and without the trio/pair constraint (CLRs) are calculated. Higher CLR values correspond with higher likelihood of a mismatch between parent and offspring.

*SNP estimates*

In the main text we have plotted the samples to show regions of the genome that show signatures of enrichment for high CLR scores. These are most evident in the between-*Aut* haplotype contrast (Figure 3B) and most likely driven by SNPs that are found specifically in the *siI;Aut* genotype (Figure 4D).

Figure S2 shows the distribution of CLR values within each haplotype pair on the same nuclear background (see Figure 3 of main text). Figure S3 shows the distribution of CLR values based on the trio analyses (see Figure 4 of main text).



**Figure S2. Distribution of putative *pair* mismatches and their corresponding CLR values.** The contrast between *siI;OreR* and *OreR;OreR* is shown in (A), and the contrast between *siI;Aut* and *OreR;Aut* is shown in (B).

****

**Figure S3. Distribution of putative *trio* mismatches and their corresponding CLR values.** The contrast within *OreR;OreR* (A), *OreR;Aut* (B), *siI;OreR* (C), and *siI;Aut* are shown.

There are no exact CLR threshold values that correspond with confident SNP calls. Clearly values of 255 on a 0->255 scale indicate high confidence scores and these were achieved in very low numbers in the *OreR* nuclear pairs analysis (Figure S2A; Table S1) and a much higher number in the *Aut* pair analysis. For the trio analyses, only the *siI;Aut* genotype demonstrated high confidence scores. Polymorphism is more likely to be detected in the pairs analysis, since the merged .bam files contain deeper reads and therefore genotype differences can be more confidently called. Table S1 describes the numbers of putative SNPs that are segregating in the populations at threshold values ranging from 100 to 255 in increments of 5.

**Table S1. Numbers of putative SNPs in each analysis type as a function of greater than or equal to (>=) CLR.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLR(>=) | *OreR pair* | *Aut pair* | *OreR;Ore trio* | *siI;OreR trio* | *OreR;Aut trio* | *siI;Aut trio* |
| 100 | 86 | 3275 | 3 | 0 | 1 | 177 |
| 105 | 78 | 3199 | 3 | 0 | 1 | 173 |
| 110 | 72 | 3082 | 3 | 0 | 1 | 169 |
| 115 | 68 | 2992 | 1 | 0 | 1 | 164 |
| 120 | 65 | 2919 | 1 | 0 | 1 | 160 |
| 125 | 57 | 2850 | 0 | 0 | 0 | 155 |
| 130 | 56 | 2764 | 0 | 0 | 0 | 152 |
| 135 | 55 | 2685 | 0 | 0 | 0 | 150 |
| 140 | 54 | 2605 | 0 | 0 | 0 | 149 |
| 145 | 53 | 2516 | 0 | 0 | 0 | 144 |
| 150 | 50 | 2454 | 0 | 0 | 0 | 142 |
| 155 | 46 | 2386 | 0 | 0 | 0 | 139 |
| 160 | 42 | 2326 | 0 | 0 | 0 | 135 |
| 165 | 42 | 2268 | 0 | 0 | 0 | 135 |
| 170 | 42 | 2198 | 0 | 0 | 0 | 134 |
| 175 | 42 | 2140 | 0 | 0 | 0 | 132 |
| 180 | 41 | 2074 | 0 | 0 | 0 | 129 |
| 185 | 39 | 2008 | 0 | 0 | 0 | 128 |
| 190 | 38 | 1959 | 0 | 0 | 0 | 128 |
| 195 | 38 | 1910 | 0 | 0 | 0 | 127 |
| 200 | 36 | 1875 | 0 | 0 | 0 | 124 |
| 205 | 35 | 1826 | 0 | 0 | 0 | 121 |
| 210 | 35 | 1789 | 0 | 0 | 0 | 119 |
| 215 | 34 | 1741 | 0 | 0 | 0 | 119 |
| 220 | 33 | 1699 | 0 | 0 | 0 | 119 |
| 225 | 33 | 1648 | 0 | 0 | 0 | 118 |
| 230 | 30 | 1610 | 0 | 0 | 0 | 114 |
| 235 | 28 | 1566 | 0 | 0 | 0 | 113 |
| 240 | 28 | 1531 | 0 | 0 | 0 | 112 |
| 245 | 28 | 1494 | 0 | 0 | 0 | 111 |
| 250 | 27 | 1463 | 0 | 0 | 0 | 110 |
| 255 | 26 | 1434 | 0 | 0 | 0 | 109 |

## (ii) Summary of population sizes in cages

**Table S2 A summary of the numbers of flies in each cage at the sampled generations.** Generations 0 and 10 were scored as the numbers of flies that seeded the pre-perturbation and post-perturbation phases (700 for *D. melanogaster* type haplotype cages, 699 for *D. simulans* haplotype cages). The figures for the remaining generations are underestimates since flies that remained stuck in the food were not frozen for genotyping and subsequently counted. The flies that were frozen were counted and added to the number of genotyped flies (92 or 93 in each cage at each timepoint).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Generation** | | | | | | | | |
| **Cage** | **nDNA** | **mtDNA** | **Replicate** | **0** | **1** | **3** | **5** | **7** | **9** | **10** | **16** | **23** |
| 1 | *OreR* | *sim* | 1 | **699** | 453 | 536 | 723 | 495 | 724 | **699** | 675 | 793 |
| 2 | *OreR* | *sim* | 2 | **699** | 674 | 809 | 656 | 726 | 796 | **699** | 883 | 747 |
| 3 | *OreR* | *sim* | 3 | **699** | 572 | 604 | 763 | 593 | 759 | **699** | 540 | 613 |
| 4 | *OreR* | *sim* | 4 | **699** | 525 | 944 | 797 | 719 | 816 | **699** | 836 | 754 |
| 5 | *Aut* | *sim* | 1 | **699** | 791 | 1097 | 870 | 783 | 860 | **699** | 986 | 929 |
| 6 | *Aut* | *sim* | 2 | **699** | 929 | 743 | 672 | 467 | 573 | **699** | 811 | 866 |
| 7 | *Aut* | *sim* | 3 | **699** | 953 | 779 | 822 | 552 | 591 | **699** | 1010 | 833 |
| 8 | *Aut* | *sim* | 4 | **699** | 1067 | 1016 | 951 | 703 | 764 | **699** | 875 | 773 |
| 9 | *OreR* | *mel* | 1 | **700** | 607 | 719 | 810 | 570 | 650 | **700** | 779 | 689 |
| 10 | *OreR* | *mel* | 2 | **700** | 605 | 803 | 917 | 711 | 768 | **700** | 914 | 811 |
| 11 | *OreR* | *mel* | 3 | **700** | 713 | 734 | 791 | 854 | 719 | **700** | 711 | 780 |
| 12 | *OreR* | *mel* | 4 | **700** | 686 | 933 | 956 | 727 | 793 | **700** | 903 | 858 |
| 13 | *Aut* | *mel* | 1 | **700** | 690 | 864 | 804 | 663 | 704 | **700** | 742 | 803 |
| 14 | *Aut* | *mel* | 2 | **700** | 976 | 1100 | 875 | 599 | 780 | **700** | 608 | 775 |
| 15 | *Aut* | *mel* | 3 | **700** | 572 | 1021 | 899 | 733 | 688 | **700** | 719 | 832 |
| 16 | *Aut* | *mel* | 4 | **700** | 839 | 818 | 774 | 657 | 720 | **700** | 863 | 904 |



**Figure S4. Population sizes at the sampled generations across the four replicate populations.** The generational time is shown along the abscissa and the numbers of flies on the ordinal axis. The *D. melanogaster* mtDNA haplotype cages are shown in (A) and (C), and the *D. simulans* mtDNA haplotype cages are shown in (B) and (D). *OreR* nuclear backgrounds are shown in (A) and (B), *Aut* nuclear backgrounds are shown in (C) and (D). Populations fluctuated over time and the grand mean across all cages = 753.52 flies (range= 453-1100, SD=122.5). As noted above, the grand mean is a slight underestimate since flies often became stuck in the food and were not retrieved for genotyping and subsequent counting.

## (iii) Supporting References

**Anders, S., P.T. Pyl, and W. Huber, 2015 HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2):166-169.**

**Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16):2078-2079.**

**Mossman, J.A., R.M.S. Mabeza, E. Blake, N. Mehta, and D.M. Rand, 2019 Age of both parents influences reproduction and egg dumping behavior in Drosophila melanogaster. *Journal of Heredity*.**

**Mossman, J.A., J.G. Tross, N.A. Jourjine, N. Li, Z. Wu *et al.*, 2017 Mitonuclear Interactions Mediate Transcriptional Responses to Hypoxia in Drosophila. *Molecular Biology and Evolution* 34 (2):447-466.**

**Mossman, J.A., J.G. Tross, N. Li, Z. Wu, and D.M. Rand, 2016 Mitochondrial-Nuclear Interactions Mediate Sex-Specific Transcriptional Profiles in Drosophila. *Genetics* 204 (2):613-630.**

**Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim *et al.*, 2012 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols* 7 (3):562-578.**