Article type: Article

Experimental evidence that thermal selection shapes mitochondrial genome evolution

Zdeněk Lajbner 1*, Reuven Pnini 1, M. Florencia Camus 2,3, Jonathan Miller 1, Damian K. Dowling 2

1 Physics and Biology Unit, Okinawa Institute of Science and Technology Graduate University (OIST), 1919-1 Tancha, Onna-son, Okinawa 904-0945, Japan,
2 School of Biological Sciences, Monash University, Clayton, Victoria 3800, Australia,
3 Department of Genetics, Evolution & Environment, University College London, London WC1E 6BT, UK.

*Correspondence: Zdeněk Lajbner, Physics and Biology Unit, Okinawa Institute of Science and Technology Graduate University (OIST), 1919-1 Tancha, Onna-son, Okinawa 904-0945, Japan.

Email: z.lajbner@seznam.cz

Running head: Thermal selection on mitogenomes.

Keywords
mtDNA; hybridization; secondary contact; thermal selection; experimental evolution,
Drosophila melanogaster
Mitochondria are essential organelles found within eukaryotic cells, which contain their own DNA. Mitochondrial DNA (mtDNA) is frequently used in population genetic and biogeographic studies as a maternally-inherited and evolutionary-neutral genetic marker, despite increasing evidence that polymorphisms within the mtDNA sequence are sensitive to thermal selection. Currently, however, all published evidence for this “mitochondrial climatic adaptation” hypothesis is correlational. Here, we use laboratory-based experimental evolution in the fruit fly, Drosophila melanogaster, to test whether thermal selection can shift population frequencies of two mtDNA haplotypes, whose natural frequencies exhibit clinal associations with latitude along the Australian east-coast. We present experimental evidence the haplotypes changed in frequency, across generations, when subjected to different thermal regimes. Our results thus contradict the widely-accepted paradigm that intra-specific mtDNA variants are selectively neutral; suggesting spatial distributions of mtDNA haplotypes reflect adaptation to climatic environments rather than within-population coalescence and diffusion of selectively-neutral haplotypes across populations.

Mitochondrial DNA (mtDNA) is usually maternally inherited (Greiner et al. 2015), and has long been considered a neutral evolutionary marker (Moritz et al. 1987). Accordingly, the mtDNA has been routinely harnessed as the quintessential tool used in phylogenetics, population genetic studies, and phylogeographic reconstructions (Avise et al. 1987). However, non-neutral evolution of DNA can compromise historical inferences in population and evolutionary biology (Rand et al. 1994). New evidence published over the past decade has suggested that a sizeable amount of the genetic variation that exists within the mitochondrial genome is sensitive to natural selection, and exerts strong effects on the phenotype (Dowling et al. 2008, Dowling 2014, Wallace 2016). Furthermore, emerging data
indicate that not all mitochondrial haplotypes perform equally well under the same thermal conditions – some perform best when it is warmer, others when it is colder (Matsuura et al. 1993, Doi et al. 1999, Dowling et al. 2007, Arnqvist et al. 2010, Wolff et al. 2016).

Correlative molecular data in humans are also consistent with the idea that certain mitochondrial mutations might represent adaptations to cold climates (Mishmar et al. 2003, Ruiz-Pesini et al. 2004, Balloux et al. 2009, Luo et al. 2011), and thus support is growing for a “mitochondrial climatic adaptation” hypothesis, which suggests that polymorphisms that accumulate across mtDNA haplotypes found in different spatial locations have been shaped by selection to the prevailing climate.

However, these contentions remain debated primarily because the conclusions of previous studies are based on correlations between mutational patterns in the mtDNA sequence and climatic region, which have proven difficult to replicate in other or larger datasets (Kivisild et al. 2006, Sun et al. 2007). We therefore decided to apply an experimental evolution approach to test the mitochondrial climatic adaptation hypothesis, by determining whether multigenerational exposure of replicated populations of fruit flies to different thermal conditions leads to consistent changes in the population frequencies of naturally-occurring mtDNA haplotypes.

In the wild, different locally-adapted populations can routinely come into secondary contact and hybridize. This enables selection of novel mito-nuclear genotypes that might be better suited to a new or changing environment (Cannestrelli et al. 2016). This evolutionary scenario is common in the Anthropocene, when humans have rapidly and unprecedentedly changed both climatic conditions and levels of habitat connectivity (Lewis & Maslin 2015). We reproduced such a hybridization event under controlled laboratory conditions, by interbreeding two subpopulations of D. melanogaster, which had adapted to different thermal environments, at different ends of an established and well-studied...
It is thought that the species was introduced into Australia during the past one to two hundred years, probably via recurrent introductions of flies from both African and European origins (David & Capy 1988, Bergland et al. 2016). The species has been studied extensively in the context of thermal adaptation along latitudinal clines, both within Australia, and other replicated clines in other continents (Hoffmann & Weeks 2007, Adrion et al. 2015, Bergland et al. 2016). This research has shown that numerous phenotypes related to thermal tolerance exhibit linear associations with latitude, and that these patterns are underscored by linear associations of key candidate nuclear genes (Hoffmann & Weeks 2007). Yet, no research had focused on the quantitative spatial distribution of mtDNA variants (Adrion et al. 2015), until Camus et al. (2017) reported that similar clinal patterns are found for two phylogenetic groups of mtDNA haplotypes along the eastern coast of Australia. Furthermore, Camus et al. (2017) were able to map these clinal patterns of mtDNA variation to the phenotype, showing that the mtDNA haplotype that predominates at subtropical latitudes confers superior resistance to extreme heat exposure, but inferior resistance to cold exposure than its temperate-predominant counterparts.
Results

We collected 20 mated-females from the Townsville subpopulation (latitude -19.26) and 20 from Melbourne (latitude -37.99). These females were used to found isofemale lineages. Genotyping of these lines revealed two deeply-divergent mtDNA haplotypes that coexist in both of the wild subpopulations we sampled, but at different frequencies (Fig. 1). The haplotypes correspond with the haplogroups of Camus et al. (2017). The A haplotype is found to predominate in the low-latitude, hot, tropical subpopulation from Townsville (H), whilst the B haplotype predominates in the temperate, cooler Melbourne subpopulation (C).

Wild fruit flies are often hosts of intracellular parasites, such as Wolbachia and associated maternally-transmitted microbiomes that are known to manipulate host phenotypes (Fry et al. 2004, Hurst & Jiggins 2005, Koukou et al. 2006). In order to assess the effects of thermal selection on the standing mitochondrial variation in our experiment, both in the presence and the absence of these maternally-inherited microbiota that co-transmit with the mtDNA, we treated a full copy of our isofemale lineages with the antibiotic tetracycline hydrochloride, such that we maintained a full copy with putative Wolbachia and unperturbed microbiomes, and one copy without Wolbachia and with perturbed microbiomes.

After multigenerational acclimatisation to the laboratory, we combined the isofemale lineages, via an admixture procedure, to form 15 replicated experimental populations, seven of which were derived from tetracycline-treated lineages, and eight derived from untreated lineages (Fig. 2). Starting haplotype frequencies in our experimental populations reflect the composition of haplotypes in the wild populations. On average, 45% of flies at the outset of the experiment possessed the A haplotype and 55% the B haplotype. These frequencies were confirmed by individual genotyping of virtually all flies in all 15 experimental populations, at this starting generation of the experimental evolution (Supplementary Table 1). Within each antibiotic treatment (ancestors tetracycline-treated
versus untreated), each of the experimental populations were then split into quadruplicates and each experimental subpopulation was then maintained at one of four different thermal conditions (Fig. 2). These were a constant 19°C, constant 25°C, fluctuating around a mean of 17.4°C, and fluctuating around a mean of 26.4°C (see Methods for details). Selection was applied for the subsequent three generations for the two colder treatments, and for seven generations for the two warmer treatments (~3 months of experimental evolution). Following this, the haplotype frequencies of each experimental subpopulation were estimated, and changes in frequencies calculated. In total, 4,410 fruit fly individuals were genotyped across the experiment (Supplementary Table 1).

We divided the dataset into four groups for analysis: females from populations treated with antibiotics (denoted FA), females from populations left untreated (FN), males from populations treated with antibiotics (MA), and males from populations left untreated (MN). We did this, since our main terms of interest centred on the level of the three-way interaction (sex \times antibiotic treatment \times thermal regime), which was significant in a mixed model analysis using maximum likelihood estimation (Extended Table 1). We found a statistically significant effect of the thermal regime on changes in haplotype frequencies sampled from females derived from antibiotic-treated lineages (Group FA, \( P = 0.0152 \) and power \( \equiv 1 - \beta = 78\% \), Fig. 3, Table 1). In this group, we found that the frequency of the B haplotype, which is naturally predominant in the temperate south of Australia, had decreased in both of the warmer treatments, but increased in the cooler treatments; in concordance with patterns found along the Australian cline. That is, in the FA group, the A haplotype increased under positive selection in each of the warmer experimental conditions; these conditions reflect those experienced in our low latitude H subpopulation (estimated selection coefficient of the A haplotype for fluctuating warm conditions \( s_A = \) ...
0.082±0.026; estimated selection coefficient of the B haplotype for fluctuating cold conditions $s_B = 0.085±0.050$).

There might also have been an effect of the thermal regime on changes in haplotype frequencies, sampled from males derived from antibiotic-untreated lineages (Group MN, $P = 0.0702$, power = 23%). However, the pattern of frequency change across the four thermal conditions was opposite to that observed in Group FA, with the frequency of the A haplotype increasing under colder temperatures, and the B haplotype under warmer temperatures.
Discussion

We have provided direct evidence that population-frequencies of naturally-occurring mtDNA haplotypes, sampled from a continuous distribution of *D. melanogaster* in east coast Australia, are shaped by thermal selection. However, our support for the mitochondrial climatic hypothesis was limited to the group of females whose ancestors had had their coevolved microbiomes, including *Wolbachia* infection, disrupted by antibiotic treatment (FA). While the patterns observed in males derived from antibiotic-untreated lineages were opposite in their direction, we note that selection on mtDNA in males cannot directly contribute to shaping patterns of mtDNA variation between generations, because males virtually never transmit their mtDNA haplotypes to their offspring. As such, mitochondrial genomes are predicted to evolve under a sex-specific selective sieve (Innocenti et al. 2011), in which mutations in the mtDNA sequence that confer harm to males can nonetheless accumulate in wild populations, as long as these same mutations are neutral or beneficial for females (Frank & Hurst 1996, Gemmell et al. 2004, Camus et al. 2015). In the absence of inter-sexual positive pleiotropy, such male-expression specific mtDNA mutations could in theory shape patterns of haplotype frequencies within a generation, if they affect male-specific patterns of juvenile or adult survival, but would not be passed on to the next generation, and would thus not shape haplotype frequencies across generations. That said, we feel it is unlikely that such male-harming mutations could explain the patterns detected in males here, and indeed the A and B haplotypes are probably largely sex-general in their effects, at least on thermal tolerance phenotypes studied (Camus et al. 2017).

On the other hand, the haplotype frequencies sampled from male offspring in the antibiotic-free treatment might have been affected by *Wolbachia*-induced cytonuclear incompatibilities. In the wild, there is a latitudinal cline in *Wolbachia* presence (Hoffmann et al. 1998), indicating that *Wolbachia* prevalence itself might be shaped by thermal selection.
The low-latitude Australian sub-tropical populations exhibit higher levels of *Wolbachia* infection than higher latitude temperate populations (Hoffmann et al. 1998). The complicated host-parasite dynamics make predictions for future changes in mito-genomic compositions of wild fruit flies populations difficult (see Kriesner et al. 2016, Corbin et al. 2017). *Wolbachia* clades also exhibit habitat-specific fitness dynamics (Versace et al. 2014), and it is possible that different *Wolbachia*, or other microsymbiotic, strains are linked to the two different mtDNA haplotypes studied here, given that each co-transmit with the mtDNA in perfect association along the maternal lineage, and that the mtDNA frequencies in the antibiotic-free treatments hitchhiked on frequency changes involving these microsymbiotic assemblages, as is expected by theory, and has been observed previously (Rasgon et al. 2006, Schuler et al. 2016).

Mitochondrial genetic markers remain an important tool for population genetics, despite growing experimental evidence that mitochondrial genetic variation is affected by thermal (Camus et al. 2017), and other kinds of selection (Kazancıoğlu & Arnqvist 2014). The evolutionary trajectories of distinct mitochondrial haplotypes might, furthermore, be selected together with functionally-linked nuclear gene complexes (Wolff et al. 2014, Hill 2015). This reinforces the point that phylogenetic, population-genetic, and biogeographic studies involving mtDNA should incorporate statistical tests to investigate the forces shaping sequence variation and evolution (Ballard & Kreitman 1995), and examine variation at multiple genetic loci (Galtier et al. 2009). Moreover, to date, researchers have focused mainly on the effects of nonsynonymous mutations in the evolutionary dynamics of mitochondrial genomes (James et al. 2016). However, the evidence is growing that mitochondrial molecular function is also affected by single nucleotides in synonymous and non-protein coding positions on mtDNA (Camus et al. 2017); a contention that is further supported by the current
study given that there are no non-synonymous SNPs separating the A and B haplotypes in this study (Camus et al. 2017).

Our study advances our understanding of DNA polymorphism by providing experimental evidence that thermal selection acts upon standing variation in the mtDNA sequence. Further research is, however, needed to resolve the dynamics of this thermal evolution; for instance, by determining whether thermal selection acts on the mtDNA sequence directly, or on epistatic combinations of mitochondrial-nuclear genotype; and whether thermal selection is the main driver of adaptive variation that we see within the mitochondrial genome or whether other environmental variables, such as the nutritional environment (Mossman et al. 2016), are salient. Furthermore, it remains unclear how much of the pool of non-neutral genetic variation that delineates distinct mitochondrial haplotypes has actually been shaped by adaptive relative to non-adaptive processes. Finally, almost all experimental work investigating the adaptive capacity of the mitochondrial genome has been conducted on just a few model invertebrate species (Dowling et al. 2010, Barreto & Burton 2013, Kazanciöglu & Arnqvist 2014, Camus et al. 2015), with few exceptions (Fontanillas et al. 2005, Boratyński et al. 2016), and this is due simply to the intractability of applying experimental evolutionary approaches to vertebrate species. Future studies should involve a combination of ecological and experimental evolutionary approaches with high resolution transcriptomics and proteomics applied more generally across eukaryotes, and also the development of tests enabling us to reliably uncover the footprint of thermal selection in wild populations (Sunnucks et al. 2017).


Acknowledgements

We thank Vanessa Kellerman and Winston Yee for assistance with wild sample collection, and Mary Ann Price, Carla Sgrò, Ritsuko Suyama, Garth Illsley, Richard Lee, Nicholas Luscombe, Pavel Munclinger, Takeshi Noda, and Oleg Simakov for helpful advices. We thank Yuan Liu for help with artwork design. This work was supported by the Physics and Biology Unit of the Okinawa Institute of Science and Technology Graduate University (J.M.) and JSPS P12751 + 24 2751 to Z.L. and J.M, the Hermon-Slade Foundation (HSF 15/2) and the Australian Research Council (FT160100022) to D.K.D. Initial stages of the study were funded by Go8EURFA11 2011003556 to Z.L. and D.K.D.

Author Contributions

Z.L. and D.K.D. designed the experiment. Z.L. performed the experiment. Z.L. and M.F.C. provided mitogenomic sequences. R.P. performed the major part of data analyses. Z.L., D.K.D., M.F.C., and J.M. contributed to the data analyses. Z.L., D.K.D., R.P., M.F.C., and J.M. wrote the manuscript.

Competing Financial Interests

The authors declare no competing financial interests.
Figures

Figure 1: Relationship of mtDNA haplotypes A and B.

The circle area for each haplotype is proportional to its frequency in the wild sample (A=18 females, B=22 females). Colours indicate the sampling region: Townsville (red, 20 females) and Melbourne (blue, 20 females). Small grey circles represent genotyped-SNP divergence.
Figure 2: Scheme of experimental evolution by hybridization of differentially thermally-adapted subpopulations of fruit fly.

Prior to the application of thermal selection, we created a series of replicated experimental populations, by combining flies of isofemale lineages collected from the Melbourne (putatively cool-adapted, or “C”) subpopulation, denoted in blue, and the Townsville (putatively hot-adapted, “H”) subpopulation (red). This was achieved over two generations, via a process of admixture of the individual isofemale lineages. In the Admixture 1 step, we pooled 5 virgin females (♀) from each of 18 of the H isofemale lineages, with 5 virgin males (♂) from each of 18 C isofemale lineages into one bottle, denoted by $HC = 18 \times 5 (♀H) + 18 \times 5 (♂C)$. In parallel in Admixture 1, we performed the reciprocal cross wherein H $\leftrightarrow$
C above, denoted by $CH = 18 \times 5(\varphi C) + 18 \times 5(\varphi H)$. Each bottle contained 90 males and 90 females (180 flies). In the following generation, at Admixture step 2, we combined 25 virgin females and 25 virgin males from HC bottles together with 25 virgin females and 25 virgin males from CH bottles, $25(\varphi CH) + 25(\varphi HC) + 25(\varphi HC) + 25(\varphi HC)$, across 15 biological replicates (7 of which were descendants of flies treated by antibiotics, 8 of which were descendants of untreated flies). At this stage, all flies had been maintained in standard laboratory conditions ($25^\circ C$) for 16 generations (14 generations as isofemale lineages, 2 during the admixture process). We then divided each of these 15 biological replicates into 4 subpopulations, subjecting each subpopulation to one of four thermal treatments ($19^\circ C, 25^\circ C, \text{fluctuating cold, and fluctuating warm}$), with each experimental subpopulation containing around 500 individuals. On the left side of the figure, yellow text denotes sample sizes associated with each stage of the admixture process for flies whose ancestors had been exposed to antibiotic treatment (ATB), while grey text on the right corresponds with untreated flies (UTR).
Figure 3: Mean change of mtDNA haplotype B frequency per thermal environment.

Interaction plots depict changing frequencies (final generation - initial generation) in stable 19°C, 25°C, fluctuating cold, and fluctuating warm environments for female and male descendants of flies treated by antibiotics (FA, MA; 7 replicates) and untreated (FN, MN; 8 replicates; in which Wolbachia and associated maternally transmitted microbiomes present).

The error-bars are estimated using Eqs. (3-5).
Table 1: The Wald statistics and the one-way ANOVA statistics for 4 population groups (FA, FN, MA, MN). (I) linear mixed model (II) fix effects only (III) the results of lme4 and (IV) one-way ANOVA. The $P$-values associated with Wald test (I, II) are calculated twice: (P) for finite-size samples and ($P_\infty$) for large samples, assuming the validity of the asymptotic $\chi^2$ distribution. The yellow background indicates statistical significance.

<table>
<thead>
<tr>
<th>group</th>
<th>(I) mixed model</th>
<th>(II) fixed effects ($\rho_g^2 \equiv 0$)</th>
<th>(III) lme4</th>
<th>(IV) ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
<td>$P_\infty$</td>
<td>$P$</td>
<td>W</td>
</tr>
<tr>
<td>FA</td>
<td>16.5667</td>
<td>0.0009</td>
<td>0.0147</td>
<td>15.8185</td>
</tr>
<tr>
<td>FN</td>
<td>1.7616</td>
<td>0.6233</td>
<td>0.6341</td>
<td>1.7296</td>
</tr>
<tr>
<td>MA</td>
<td>2.0993</td>
<td>0.5521</td>
<td>0.5715</td>
<td>2.1414</td>
</tr>
<tr>
<td>MN</td>
<td>10.0702</td>
<td>0.0180</td>
<td>0.0494</td>
<td>10.0724</td>
</tr>
</tbody>
</table>
Methods (3000 words)

Experimental procedures

Wild subpopulations of *D. melanogaster* were sampled in Australia. We sampled a hot adapted subpopulation (H; Townsville: -19.26, 146.79) in the north-east, and a cool adapted subpopulation (C; Melbourne: -37.99, 145.27) in the south of the continent. We collected fertilised females and established 20 isofemale lineages from each wild population. After 3 generations of acclimatisation to laboratory conditions, we split each isofemale lineage into two replicates, and treated one replicate of each lineage with 0.164 mg mL\(^{-1}\) tetracycline in food for 3 generations to remove any intracellular and cytoplasmically-inherited bacteria, such as *Wolbachia* (Clancy & Hoffman 1998). We then propagated these lineages for a further 10 generations to mitigate any effects of the antibiotic treatment. Flies were reared at 25°C, on a 12:12 hour light:dark cycle, in 10 dram plastic vials, on a potato-dextrose-agar medium, with *ad libitum* live yeast added to each vial. All isofemale lineages were then transferred from our laboratories in Australia to those in Japan, and their food medium changed to a corn flour-glucose-agar medium (Supplementary Table 2), with *ad libitum* live yeast added to each vial. They acclimatized for a further 3 generations at 25°C, before entering an admixture process described below, in order to set up a series of replicated experimental populations.

We pooled 5 virgin females (♀) from each of 18 of the H isofemale lineages, mentioned above, with 5 virgin males (♂) from each of 18 of the C lineages in one bottle (HC = 18 x 5 ♀H + 18 x 5 ♂C), and 5 virgin females from each of the 18 C isofemale lineages with 5 virgin males from each of the 18 H isofemale lineages in another bottle (CH = 18 x 5 ♀C + 18 x 5 ♂H), such that each bottle contained 90 males and 90 females. This step was performed separately for flies sourced from the tetracycline-treated isofemale lineages and flies sourced from untreated isofemale lineages, separately (Fig. 2; Supplementary Table
We then allowed the flies to lay eggs over 8 consecutive days, and transferred them to fresh bottles as indicated in Supplementary Table 3. We reared all experimental populations in 250 ml bottles on a corn flour-glucose-agar medium. In the next step, we mixed F1 offspring (25 virgin males and 25 virgin females) from the HC bottles with corresponding F1 offspring (25 virgin males and 25 virgin females) from the CH bottles (25 ♂CH + 25 ♀HC + 25 ♀CH + 25 ♂HC). We established 7 experimental populations from the tetracycline-treated isofemale lineages and 8 experimental populations from the untreated lineages. We allowed flies of these populations to mate and lay eggs at 25°C, and then we transferred bottles with approximately 500 eggs into four thermal regimes, represented by cool versus warm temperatures, on either a constant or fluctuating temperature cycle. Bottles maintained in the cool and constant temperature were kept at a constant 19°C, and those in the warm and constant temperature at 25°C temperature. We used Environmental Chambers (MIR-154, Sanyo) to generate fluctuating thermal conditions that are common in areas of origin of our experimental populations (The Australian Government, Bureau of Meteorology), Melbourne: 8:00(22°C); 11:30(28°C); 16:00(20°C); 20:00(17°C); 22:00(14°C); 8:00(15°C); 11:30(20°C); 16:00(16°C); 20:00(15°C); 22:00(14°C), and Townsville: 8:00(27°C); 10:30(28°C); 20:00(27°C); 22:30(26°C); 0:00(25°C). The temperature in all conditions was continually monitored and in fluctuating conditions recorded by Thermo-hydro SD Data Loggers (AD-5696; A&D Ltd). We propagated all replicate populations for three months (3 or 7 successive generations depending on the thermal condition; Supplementary Table 4). We regulated the size of each population by trimming egg numbers per generation to approximately 500 eggs. At the end of the experimental evolution period of three months, adult flies were collected and fixed in 95% ethanol.
Data collection

Total genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen). We sequenced total DNA of H and C population samples quantitatively, using an Illumina platform at Micromon (Monash University, Australia). Length of reads was set to 70bp and we reached a maximum coverage 500x on coding parts of mitogenomes (Camus et al. 2017).

We mapped all reads on published mitogenomic sequence NC 001709 in Geneious R6 (http://www.geneious.com, Kearse et al. 2012). We observed overall mitogenomic variability and picked 14 mtDNA polymorphic sites (SNPs), which are not unique to the H or C populations, which delineate all flies into one of two corresponding mtDNA haplotypes, which we denote A and B, and which are described in the Results (Fig. 1, Supplementary Table 5), using multiplexPCR and MALDI/Tof (mass spectrometry; Geneworks, Australia). We genotyped virtually all flies in the starting generation (formed by 50 males and 50 females per each of the 15 replicates), and then more than 24 males and 25 females per bottle in final generation of the experiment, upon completion of experimental evolution. The minimal number of sequenced samples was estimated assuming the relative thermal effect of 10% at power of $1 - \beta = 70\%$ and the F-distribution. In total, we genotyped 4410 individuals (Supplementary Table 1).

Data analysis

A multilevel linear model showed a significant three-way interaction between thermal regime, sex, and antibiotic treatment on the change in frequency of haplotype B (Extended Table 1). Accordingly, we divided the total of 60 experimental populations into 4 groups: FA=females whose ancestors had been treated with antibiotics, FN=females untreated, MA=males whose ancestors had been treated with antibiotics, and MN=males untreated. Flies in each of these groups are propagated under 1 of 4 thermal treatments: 19°C,
27°C, fluctuating cold, fluctuating warm. The measured frequencies of haplotype B are thus denoted as $f_{gi}^n$, where $g = 1, 2, 3, 4 = G$ stands for the group, $i = 1, 2, 3, 4 = T$ is the thermal regime, and $n = 1, 2, ..., N$ is the biological replicate’s number. The frequencies $f_{gi}^n$ are then subtracted from the corresponding frequencies of the initial female population (regardless of the sex we are examining), $f_0^n$ and the frequency changes, $y_{gi}^n = f_{gi}^n - f_0^n$ were fit to a linear model (Kutner et al. 2013)

$$y_{gi}^n = \theta_{gi} + \epsilon_{gi}^n \quad (i = 1, 2, 3, 4 = T; n = 1, 2, ..., N)$$

(1)

where $\theta_{gi}$ denotes the mean value of the frequency difference, obtained for thermal regime $i$ in group $g$ (averaged over $N$ samples) and $\epsilon_{gi}^n$ is the measurement noise associated with sample $b$. The sample sizes in Eq. (1) are $N = 7$ for groups FA and MA, whose ancestors had been treated with antibiotics, and $N = 8$ for untreated groups FN and MN. The statistical properties of the Gaussian noise in (1) are defined by having zero mean, $\langle \epsilon_{gi}^n \rangle = 0$, and a positive definite correlation matrix

$$\langle \epsilon_{gi}^n \epsilon_{g'i}^{n'} \rangle = \delta_{ggi'} \sigma_{gi} \sigma_{g'i} [\delta_{ij}(1 - \rho_{g}^n) + \rho_{g}^n]$$

(2)

Here, angular brackets $\langle \cdots \rangle$ denote ensemble averaging over the noise, so that $\sigma_{gi}^2 = \langle (\epsilon_{gi}^n)^2 \rangle$ is the variance, and $-\frac{1}{3} < \rho_{g}^n < 1$ is the correlation-coefficient associated with replicate number $n$. The constraint imposed on $\rho_{g}^n$ is required to ensure the positive-definiteness of the correlation. Eq. (2) allows one to consider two distinctive random effects: (i) different noise levels for each thermal regime, and (ii) possible dependencies between replicates which belong to different regimes, and yet originated from the same parental generation.

The parameters in Eq. (2) can be estimated by a standard maximum-likelihood (ML) calculation. Replacing ensemble averages by sampled means one obtains

$$\hat{\theta}_{gi} = N^{-1} \Sigma_n y_{gi}^n \quad , \quad \hat{\sigma}_{gi}^2 = N^{-1} \Sigma_n (y_{gi}^n - \hat{\theta}_{gi})^2$$

(3)
where $\hat{\theta}$ denotes an estimator of a random variable $x$. Note, that for sequel convenience, $\sigma_{\hat{\theta}i}^2$ is normalized by the number of samples (rather than by $N - 1$). Having $\hat{\theta}_{gi}$ and $\sigma_{\hat{\theta}i}^2$ as written in Eqs. (3), the ML estimator of $\rho_{gi}^n$ is found by solving the following quadratic equation

$$12\rho_{gi}^n = -(1 + 3\rho_{gi}^n)(1 - \rho_{gi}^n)(\delta y_{gi}^n)^T[\Gamma(\rho_{gi}^n)]\Omega[\Gamma(\rho_{gi}^n)](\delta y_{gi}^n)$$  \hspace{1cm} (4a)$$

where $(\delta y_{gi}^n) \equiv y_{gi}^n - \hat{\theta}_{gi}$ is the 4-component (column) vector of fluctuations, $(\delta y_{gi}^n)^T$ is the corresponding transposed (row) vector, $\Gamma(\rho_{gi}^n)$ is the inverse of the correlation-matrix given by

$$\Gamma_{ij} = (1 + 2\rho_{gi}^n)/[\sigma_{\hat{\theta}i}^2(3\rho_{gi}^n + 1)(1 - \rho_{gi}^n)], \Gamma_{i\neq j} = -\rho_{gi}^n/[(\sigma_{\hat{\theta}i}^2\sigma_{\hat{\theta}j}^2(3\rho_{gi}^n + 1)(1 - \rho_{gi}^n)] \hspace{1cm} (4b)$$

and $\Omega$ is a traceless matrix of rank-1, such that $\Omega_{ij} = \sigma_{\hat{\theta}i}\sigma_{\hat{\theta}j}(1 - \delta_{ij})$. After solving Eq. (4a) for all values of $b$ while keeping $g$ fixed - the estimated errors of $\hat{\theta}_{gi}$ [i.e., the fix-effect errors presented in (1)] are found by inverting the Fisher information matrix (Cover & Thomas 2006) $\partial \log P / (\partial \theta_i \partial \theta_j)$, where $P$ is the joint probability distribution of $y_{gi}^n$. This leads to an error-matrix of the form:

$$(V_g)_{ij} \equiv \langle \delta \hat{\theta}_{gi}\delta \hat{\theta}_{gj} \rangle = \langle (\hat{\theta}_{gi} - \langle \hat{\theta}_{gi} \rangle)(\hat{\theta}_{gj} - \langle \hat{\theta}_{gj} \rangle) \rangle = -(H_g)_{ij}^{-1} = [\sum_b \Gamma_{ij}(\rho_{gi}^n)]^{-1} \hspace{1cm} (5)$$

with $H_g$ being the Fisher matrix evaluated at the ML solution. Note, that when $\rho_{gi}^n = 0$, Eq. (5) is reduced to the expected diagonal form: $(V_g)_{ij} = \delta_{ij}\sigma_{\hat{\theta}i}^2/N$.

Fig. 2 suggests that group FA (and to a lesser extent group MN), is showing some degree of thermal selectivity. In order to quantify this assertion, $\hat{\theta}_{gi}$ are re-parametrized as

$$\hat{\theta}_{gi} = \mu_g + \alpha_{gi}, \text{ where } \mu \text{ is the average over } T = 4 \text{ levels, } \mu_g \equiv \sum_i \hat{\theta}_{gi} / T, \text{ and}$$

$$\sum_{i=1}^T \alpha_{gi} = 0. \text{ The F-statistic as defined by the standard 1-factor ANOVA, is then: } F_g =$$

$$[(\text{mean-square-error between levels})/(\text{mean-square-error within levels})]. \text{ Namely, for each group separately,}$$

$$F_g = [\sum_{i=1}^T \alpha_{gi}^2 / (T - 1)] / [\sum_{i=1}^T \alpha_{gi}^2 / (T(N - 1))] \hspace{1cm} (g = 1,2,\ldots,4 = G) \hspace{1cm} (6)$$
with \( \{\hat{\theta}_{gi}, \bar{\delta}_{gi}\} \) given in Eq. (3). The F-values for all groups, together with the corresponding P-values, are shown in Table 1 (IV). Indeed, group FA (females on antibiotics) exhibits a significant thermal effect with sufficiently high power \((P = 0.0152\) and power \(\geq 78\%\)).

The P-value and power of \(F_g\) are given by: \(P = 1 - \Phi_{v_1,v_2}(F_g, \lambda = 0)\) and power \(= 1 - \beta = 1 - \Phi_{v_1,v_2}(F_g, \lambda = TF_g)\), where \(\Phi_{v_1,v_2}(f, \lambda)\) is the cumulative non-central F-distribution with degrees-of-freedom \(v_1 = T - 1, v_2 = T(N - 1)\), and non-centrality parameter \(\lambda\).

In addition to \(F_g\), we also studied the Wald-statistic of the thermal effects which is defined as \(W_g = \alpha_g^\top(\delta \alpha_g \delta \alpha_g^\top)^{-1} \alpha_g\), where \(\alpha_g\) (with g kept fixed) is the vector of thermal deviations, \(\alpha_{gi} = \hat{\theta}_{gi} - \mu_g\) \((i = 1, 2, 3 = T - 1)\), and \(\delta \alpha\) the fluctuation \(\delta \alpha = \alpha - \langle \alpha \rangle\).

Introducing \(\vec{\alpha}_g \equiv (\theta_{g1}, ..., \theta_{g4})\), \(\vec{\alpha}_g \equiv (\mu_g, \alpha_{g1}, \alpha_{g2}, \alpha_{g3})\) and changing variables, \(\vec{\alpha}_g = R\hat{\alpha}_g\), one finds that

\[
W_g = (R\hat{\alpha}_g)^\top(R V_g R^\top)^{-1} (R\hat{\alpha}_g) \quad (g = 1, 2, ..., A = G) \tag{7}
\]

where \(R\) is a \(3 \times 4\) transformation matrix \((R_{ii} = 3/4, R_{i\neq j} = -1/4)\), and \(V_g\) is the correlation matrix previously given in Eq. (5). We have applied Eq. (7) both to a mixed model [in which \(\rho_g^n\) are obtained using Eq. (4a)], and to a fixed-effects model [in which all \(\rho_g^n\) vanish identically and \(V_g\) is a diagonal matrix]. The results are shown in Table 1 (I-II). For finite-size samples, \(W_g(T-1)\) is distributed according to the F-distribution (Engle 1984, Parker 2016).

Therefore, the P-values and power associated with \(W_g\) take the form: \(P = 1 - \Phi_{v_1,v_2}\left[\frac{W_g}{T-1}, \lambda = 0\right]\), power \(= 1 - \beta = 1 - \Phi_{v_1,v_2}\left[\frac{W_g}{T-1}, \lambda = \lambda_w\right]\), where \(v_1 = T - 1\) and \(v_2 = N(T - 2) - (T - 1)\). In the limit of large samples, \(N \gg 1\) [or, equivalently, when the correlation \(V_g\) in Eq. (7) is given a-priori rather than being estimated], \(W_g\) is distributed as \(\chi^2\) with \(T-1\) degrees of freedom. As a result, the corresponding P-values decrease significantly (see Table 1).
While the $\chi^2$ distribution may be appropriate in large-sample studies, it is inapplicable in our case where $N = (7, 8)$. Focusing on the statistically significant FA group ($g = 1$), we’ve verified the cross-over of $W_g$ from F to the $\chi^2$ distribution and the consistency of expressions (6-7) by performing MC simulations. The starting point of the simulations is the set of ‘actual parameters’ as measured by the experiment:

$$\begin{align*}
\theta &= [+0.0409, -0.0506, +0.0601, -0.1477], \\
\sigma &= [ +0.0408, +0.0427, +0.0917, +0.2002, -0.0771, +0.0758, -0.1162] \\
\rho &= [+0.0409, -0.0506, +0.0601, -0.1477] \\
\end{align*}$$

(8)

Using these empirical values, we generate an ensemble of 20,000 realizations of random frequencies – all sampled out of a multivariate Gaussian noise as defined by Eqs. (1-2). For each of these replicas we then infer a set of $15 = 2T + N$ parameters $\{\hat{\theta}_{gi}, \hat{\sigma}_{gi}, \hat{\rho}_{g}\}$ according to Eqs. (3-5). Finally, using Eqs. (6-7), we generate the statistics and compute histograms for $F_g$ and $W_g$. As shown in Extended Fig. 2, the simulated histograms are in good agreement with the theoretically expected non-central $F$-distributions (as well as the asymptotic non-central $\chi^2$).

When the correlation matrix $V_g$ in (7) is replaced by expression (2) with fixed parameters given by (8) - the Wald statistic crosses over from F to the $\chi^2$ distribution. Consequently, the $P$-value of the experiment decreases: $P = 0.0147 \rightarrow 9 \times 10^{-4}$ and the power decreases from 73% to about 60% (see Extended Fig. 1).

We compared the standard maximum likelihood estimation described in detail here with lme4 v. 1.1-10 package (Bates et al. 2015) of R version 3.2.2. (R Development Core Team 2015) in RStudio server v. 0.99.465 (RStudio Team 2015). The average frequency differences for all groups, $\hat{\theta}_{gi} (g = 1, 2, 3, 4)$, together with their estimated errors, $\delta \hat{\theta}_{gi}$, are shown in Extended Table 2 (see also Fig. 3). The error estimates of lme4 are level (i.e.,
thermal) independent and given by the RMS value, $\bar{\theta}_g^2 = \sum_i \theta_{gi}^2 / (NT)$. This feature is presumably a result of the internal constraints that are imposed on the random model of lme4. Referring to Eq. (2), the constraints are: $\sigma_{gi} = \sigma_g \forall i$, $\sum_b \rho_g^n = 0$. The differences in the error estimates between our computations [rows (I) and (II) in Extended Table 2] and the results of lme4 [row (III)] are clear. The relative differences due to random effects seem to be rather small, $|\delta \bar{\theta}_{gi}(I) - \delta \bar{\theta}_{gi}(II)| / |\bar{\theta}_{gi}| \leq 5\%$.

Note, that our computation produces higher $P$-values i.e., “less significant” as compared to those of lme4. The reason being that, in calculating $P$-values out of the Wald statistics lme4 implicitly assumes an infinite number of samples. In this limiting case, the Wald statistic $W_g$ is distributed according to the $\chi^2$ distribution and, as a result, the $P$-value decreases by an order of magnitude to give a $P = 6 \times 10^{-4} - 9 \times 10^{-4}$ and power $\approx 60\%$. Yet, the differences between $W_g(I)$ and $W_g(III)$, shown in Table 1, are not sufficiently large to alter any of the conclusions related to statistical significance.

MitDNA is transmitted mainly maternally, but paternal mitDNA leakage is documented in our model organism (Nunes et al. 2013). For the reason, we verified using lme4 that our results remain unchanged when one takes in account frequencies of both sexes in the starting generation (Supplementary Table 6). That is, the results remain qualitatively similar, and their interpretations identical, if we use the frequencies of males and females combined in the starting generation, rather than females only, in deriving the change in frequencies across the experiment.

References


Cover T. M. & Thomas J. A. Elements of information theory, 2nd Ed. (John Willey & Sons, 2006).


Extended data

Extended Figure 1 MC simulations for the statistically significant FA group.

Each histogram consists of 20,000 independent realizations that are generated according to Eqs. (1-2). Asterisk marks the empirical value.

(a) Histogram of the ANOVA $F$ statistic (Eq. 6) compared to the non-central $F$-distribution ($H_1$) with parameters $\nu_1 = T - 1 = 3$, $\nu_2 = T(N - 1) = 24$, $\lambda_F = \bar{m}_F \nu_1 (\nu_2 - 2) / 2 - \nu_1 = 15$. The probability distribution of the null hypothesis ($H_0$) is obtained by setting $\lambda_F = 0$. 
(b) The non-central $\chi^2$ distribution ($H_1$) describing the Wald statistic $W_g$ in the limit of large samples ($N \gg 1$). Here, $\nu_1 = T - 1 = 3$, $\lambda_w = \bar{m}_w - \nu_1 = 16.6$. For the null hypothesis $(H_0)$, $\lambda_w = 0$.

(c) Histogram of $W_g$ for finite-size samples (Eq. 7), compared to the scaled non-central $F$-distribution ($H_1$) for the random variable $f = W_g / (T - 1)$. The parameters of ($H_1$) are $\nu_1 = T - 1 = 3$, $\nu_2 = N (T - 2) - (T - 1) = 11$. $\lambda_w = \bar{m}_w \nu_1 (\nu_2 - 2) / [2(T - 1)] - \nu_1 = 21$.

For $(H_0)$, $\lambda_w = 0$. The asymptotic $\chi^2$ distribution (b) is shown as a dashed line.

(d) The cumulative distribution functions (CDF) of the Wald statistics, comparing the empirical distribution with the expected $\chi^2$ and scaled-$F$ distributions.

Note that the only free (fitting) parameters in (a-d) are the sampled means $\{\bar{m}_F, \bar{m}_W\}$ which are obtained by averaging over all realizations of $\{F_g, W_g\}$, respectively.
Extended Table 1: Multilevel model examining the effect of sex, antibiotic treatment, and thermal regime on B haplotype frequency change, as a response variable.

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$</th>
<th>d.f.</th>
<th>$P (&gt;\chi^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.1518</td>
<td>1</td>
<td>0.696824</td>
</tr>
<tr>
<td>sex</td>
<td>0.8269</td>
<td>1</td>
<td>0.363163</td>
</tr>
<tr>
<td>antibiotic treatment</td>
<td>0.0850</td>
<td>1</td>
<td>0.770666</td>
</tr>
<tr>
<td>thermal regime</td>
<td>1.9834</td>
<td>3</td>
<td>0.575866</td>
</tr>
<tr>
<td>antibiotic treatment: thermal regime</td>
<td>8.9186</td>
<td>3</td>
<td>0.030393</td>
</tr>
<tr>
<td>sex: antibiotic treatment</td>
<td>0.0326</td>
<td>1</td>
<td>0.856811</td>
</tr>
<tr>
<td>sex: thermal regime</td>
<td>15.4599</td>
<td>3</td>
<td>0.001463</td>
</tr>
<tr>
<td>sex: antibiotic treatment: thermal regime</td>
<td>10.7419</td>
<td>3</td>
<td>0.013207</td>
</tr>
</tbody>
</table>

$\sigma$

| ID                       | 0.119988 |
| Population               | 0.009593 |

Sex, antibiotic treatment, and thermal regime were modelled as fixed effects. ID and Population were modelled as random effects. Population indicates the biological replicate i.e., group of 4 bottles descending from a single starting bottle. ID pairs males and females which are sharing the same bottle. The yellow background indicates statistical significance.
Extended Table 2: The fixed parameters, $\bar{\theta}_g$, and their errors, $\delta \bar{\theta}_g$, for 4 population groups (FA, FN, MA, MN), evaluated by three schemes of estimation: (I) linear mixed model (II) fix effects only (III) lme4.

<table>
<thead>
<tr>
<th>model of error(†)</th>
<th>group(‡)</th>
<th>19°C</th>
<th>25°C</th>
<th>fluctuating cold</th>
<th>fluctuating warm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\bar{\theta}_1$</td>
<td>$\pm \delta \bar{\theta}_1$</td>
<td>$\bar{\theta}_2$</td>
<td>$\pm \delta \bar{\theta}_2$</td>
</tr>
<tr>
<td>(I)</td>
<td>FA(7)</td>
<td>+0.0409</td>
<td>0.0348</td>
<td>0.0354</td>
<td>0.0428</td>
</tr>
<tr>
<td>(II)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(III)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I)</td>
<td>FN(8)</td>
<td>+0.0195</td>
<td>0.0584</td>
<td>0.0589</td>
<td>0.0554</td>
</tr>
<tr>
<td>(II)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(III)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I)</td>
<td>MA(7)</td>
<td>-0.0029</td>
<td>0.0438</td>
<td>0.0442</td>
<td>0.0489</td>
</tr>
<tr>
<td>(II)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(III)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I)</td>
<td>MN(8)</td>
<td>-0.0144</td>
<td>0.0409</td>
<td>0.0418</td>
<td>0.0555</td>
</tr>
<tr>
<td>(II)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(III)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† (I) $\delta \bar{\theta}_g$ according to Eq. (5). (II) $\delta \bar{\theta}_g = \sigma_g / \sqrt{N}$ (i.e., assuming that $\rho_g^n = 0$). ‡ the numbers in parenthesis indicate the sample-size ($N = 7$, 8 replicates for each thermal level).
Supplementary Information Legends

Supplementary Table 1: List of samples
In experimental populations ATB1-ATB7, the ancestors had been exposed to antibiotic treatment, while experimental populations UTR1-UTR8 correspond with untreated flies.

Supplementary Table 2: Fly food composition

Supplementary Table 3: Starting dates of experimental populations.
Foundation date marks the date at which virgin flies were combined in a bottle as outlined in Admixture Step 2 (Fig. 2) to form the Starting generation. In Admixture Step 1, we allowed their parents to lay eggs for about 1 day, and transferred them to a new bottle. This process was repeated across nine days. We call the process by which we transfer the flies to a new bottle a “tip”. Virgin flies of each sex were sourced from several tips, in order to ensure we had an adequate supply of flies to initiate the experimental populations. We show the dates of maternal ovipositioning and virgin collection in a separate column (date). Number means the number of virgin flies sourced from the tip.

Supplementary Table 4: Propagation of experimental populations in dates.
Foundation date corresponds with Supplementary Table 3. We imposed the four different thermal regimes on multiple generations starting by eggs laid by the Starting generation. Within each generation, the flies of each experimental bottle were transferred to new bottles, and the column “tip number” reflects the “tip” that was used to propagate the next generation per experimental population. Although the tip we used to initiate each experimental population varied in Generation 1, in subsequent generations, we propagated experimental populations mostly from the first tip.

Supplementary Table 5: Selected SNPs characteristics and individual haplotypes.
Supplementary Table 6: The Wald statistics of the mixed model obtained with lme4, when the frequencies $f^n_{gi}$ are subtracted from the corresponding frequencies in the starting generation (males and females combined).