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

51 indicate that not all mitochondrial haplotypes perform equally well under the same thermal
52 conditions – some perform best when it is warmer, others when it is colder (Matsuura et al.
53 1993, Doi et al. 1999, Dowling et al. 2007, Arnqvist et al. 2010, Wolff et al. 2016).
54 Correlative molecular data in humans are also consistent with the idea that certain
55 mitochondrial mutations might represent adaptations to cold climates (Mishmar et al. 2003,
56 Ruiz-Pesini et al. 2004, Balloux et al. 2009, Luo et al. 2011), and thus support is growing for
57 a "mitochondrial climatic adaptation" hypothesis, which suggests that polymorphisms that
58 accumulate across mtDNA haplotypes found in different spatial locations have been shaped
59 by selection to the prevailing climate.

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

68 In the wild, different locally-adapted populations can routinely come into
69 secondary contact and hybridize. This enables selection of novel mito-nuclear genotypes that
70 might be better suited to a new or changing environment (Cannestrelli et al. 2016). This
71 evolutionary scenario is common in the Anthropocene, when humans have rapidly and
72 unprecedentedly changed both climatic conditions and levels of habitat connectivity (Lewis
73 & Maslin 2015). We reproduced such a hybridization event under controlled laboratory
74 conditions, by interbreeding two subpopulations of *D. melanogaster,* which had adapted to
75 different thermal environments, at different ends of an established and well-studied

76 latitudinal cline (Hoffmann & Weeks 2007, Bergland et al. 2016). It is thought that the
77 species was introduced into Australia during the past one to two hundred years, probably via
78 recurrent introductions of flies from both African and European origins (David & Capy 1988,
79 Bergland et al. 2016). The species has been studied extensively in the context of thermal
80 adaptation along latitudinal clines, both within Australia, and other replicated clines in other
81 continents (Hoffmann & Weeks 2007, Adrion et al. 2015, Bergland et al. 2016). This
82 research has shown that numerous phenotypes related to thermal tolerance exhibit linear
83 associations with latitude, and that these patterns are underscored by linear associations of
84 key candidate nuclear genes (Hoffmann & Weeks 2007). Yet, no research had focused on the
85 quantitative spatial distribution of mtDNA variants (Adrion et al. 2015), until Camus et al.
86 (2017) reported that similar clinal patterns are found for two phylogenetic groups of mtDNA
87 haplotypes along the eastern coast of Australia. Furthermore, Camus et al. (2017) were able
88 to map these clinal patterns of mtDNA variation to the phenotype, showing that the mtDNA
89 haplotype that predominates at subtropical latitudes confers superior resistance to extreme
90 heat exposure, but inferior resistance to cold exposure than its temperate-predominant
91 counterparts.

92 **Results**

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

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

108 After multigenerational acclimatisation to the laboratory, we combined the
109 isofemale lineages, via an admixture procedure, to form 15 replicated experimental
110 populations, seven of which were derived from tetracycline-treated lineages, and eight
111 derived from untreated lineages (Fig. 2). Starting haplotype frequencies in our experimental
112 populations reflect the composition of haplotypes in the wild populations. On average, 45%
113 of flies at the outset of the experiment possessed the A haplotype and 55% the B haplotype.
114 These frequencies were confirmed by individual genotyping of virtually all flies in all 15
115 experimental populations, at this starting generation of the experimental evolution
116 (Supplementary Table 1). Within each antibiotic treatment (ancestors tetracycline-treated

117 versus untreated), each of the experimental populations were then split into quadruplicates
118 and each experimental subpopulation was then maintained at one of four different thermal
119 conditions (Fig. 2). These were a constant 19°C, constant 25°C, fluctuating around a mean of
120 17.4°C, and fluctuating around a mean of 26.4°C (see Methods for details). Selection was
121 applied for the subsequent three generations for the two colder treatments, and for seven
122 generations for the two warmer treatments (~3 months of experimental evolution). Following
123 this, the haplotype frequencies of each experimental subpopulation were estimated, and
124 changes in frequencies calculated. In total, 4410 fruit fly individuals were genotyped across
125 the experiment (Supplementary Table 1).

126 We divided the dataset into four groups for analysis: females from populations
127 treated with antibiotics (denoted FA), females from populations left untreated (FN), males
128 from populations treated with antibiotics (MA), and males from populations left untreated
129 (MN). We did this, since our main terms of interest centred on the level of the three-way
130 interaction (sex × antibiotic treatment × thermal regime), which was significant in a mixed
131 model analysis using maximum likelihood estimation (Extended Table 1). We found a
132 statistically significant effect of the thermal regime on changes in haplotype frequencies
133 sampled from females derived from antibiotic-treated lineages (Group FA, $P =$
134 0.0152 and power $\equiv 1 - \beta = 78\%$, Fig. 3, Table 1). In this group, we found that the
135 frequency of the B haplotype, which is naturally predominant in the temperate south of
136 Australia, had decreased in both of the warmer treatments, but increased in the cooler
137 treatments; in concordance with patterns found along the Australian cline. That is, in the FA
138 group, the A haplotype increased under positive selection in each of the warmer experimental
139 conditions; these conditions reflect those experienced in our low latitude H subpopulation
140 (estimated selection coefficient of the A haplotype for fluctuating warm conditions s_A =

141 0.082±0.026; estimated selection coefficient of the B haplotype for fluctuating cold

142 conditions s_B = 0.085±0.050).

143 There might also have been an effect of the thermal regime on changes in

144 haplotype frequencies, sampled from males derived from antibiotic-untreated lineages

145 (Group MN, P = 0.0702, power = 23%). However, the pattern of frequency change across the

146 four thermal conditions was opposite to that observed in Group FA, with the frequency of the

147 A haplotype increasing under colder temperatures, and the B haplotype under warmer

148 temperatures.

149 **Discussion**

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

170 On the other hand, the haplotype frequencies sampled from male offspring in the
171 antibiotic-free treatment might have been affected by *Wolbachia*-induced cytonuclear
172 incompatibilities. In the wild, there is a latitudinal cline in *Wolbachia* presence (Hoffmann et
173 al. 1998), indicating that *Wolbachia* prevalence itself might be shaped by thermal selection.

174 The low-latitude Australian sub-tropical populations exhibit higher levels of *Wolbachia*
175 infection than higher latitude temperate populations (Hoffmann et al. 1998). The complicated
176 host-parasite dynamics make predictions for future changes in mito-genomic compositions of
177 wild fruit flies populations difficult (see Kriesner et al. 2016, Corbin et al. 2017). *Wolbachia*
178 clades also exhibit habitat-specific fitness dynamics (Versace et al. 2014), and it is possible
179 that different *Wolbachia*, or other microsymbiont, strains are linked to the two different
180 mtDNA haplotypes studied here, given that each co-transmit with the mtDNA in perfect
181 association along the maternal lineage, and that the mtDNA frequencies in the antibiotic-free
182 treatments hitchhiked on frequency changes involving these microsymbiotic assemblages, as
183 is expected by theory, and has been observed previously (Rasgon et al. 2006, Schuler et al.
184 2016).

185 Mitochondrial genetic markers remain an important tool for population genetics,
186 despite growing experimental evidence that mitochondrial genetic variation is affected by
187 thermal (Camus et al. 2017), and other kinds of selection (Kazancıoğlu & Arnqvist 2014).
188 The evolutionary trajectories of distinct mitochondrial haplotypes might, furthermore, be
189 selected together with functionally-linked nuclear gene complexes (Wolff et al. 1014, Hill
190 2015). This reinforces the point that phylogenetic, population-genetic, and biogeographic
191 studies involving mtDNA should incorporate statistical tests to investigate the forces shaping
192 sequence variation and evolution (Ballard & Kreitman 1995), and examine variation at
193 multiple genetic loci (Galtier et al. 2009). Moreover, to date, researchers have focused mainly
194 on the effects of nonsynonymous mutations in the evolutionary dynamics of mitochondrial
195 genomes (James et al. 2016). However, the evidence is growing that mitochondrial molecular
196 function is also affected by single nucleotides in synonymous and non-protein coding
197 positions on mtDNA (Camus et al. 2017); a contention that is further supported by the current

198 study given that there are no non-synonymous SNPs separating the A and B haplotypes in
199 this study (Camus et al. 2017).
200 Our study advances our understanding of DNA polymorphism by providing
201 experimental evidence that thermal selection acts upon standing variation in the mtDNA
202 sequence. Further research is, however, needed to resolve the dynamics of this thermal
203 evolution; for instance, by determining whether thermal selection acts on the mtDNA
204 sequence directly, or on epistatic combinations of mitochondrial-nuclear genotype; and
205 whether thermal selection is the main driver of adaptive variation that we see within the
206 mitochondrial genome or whether other environmental variables, such as the nutritional
207 environment (Mossman et al. 2016), are salient. Furthermore, it remains unclear how much of
208 the pool of non-neutral genetic variation that delineates distinct mitochondrial haplotypes has
209 actually been shaped by adaptive relative to non-adaptive processes. Finally, almost all
210 experimental work investigating the adaptive capacity of the mitochondrial genome has been
211 conducted on just a few model invertebrate species (Dowling et al. 2010, Barreto & Burton
212 2013, Kazancıoğlu & Arnqvist 2014, Camus et al. 2015), with few exceptions (Fontanillas et
213 al. 2005, Boratyński et al. 2016), and this is due simply to the intractability of applying
214 experimental evolutionary approaches to vertebrate species. Future studies should involve a
215 combination of ecological and experimental evolutionary approaches with high resolution
216 transcriptomics and proteomics applied more generally across eukaryotes, and also the
217 development of tests enabling us to reliably uncover the footprint of thermal selection in wild
218 populations (Sunnucks et al. 2017).

219 **References**

220 Adrion, J. R., Hahn, M. W. & Cooper, B. S. Revisiting classic clines in *Drosophila*
221 *melanogaster* in the age of genomics. *Trends Genet.* **31**, 434–444 (2015).
222
223 Arnqvist, G. *et al*. Genetic architecture of metabolic rate: environment specific epistasis
224 between mitochondrial and nuclear genes in an insect. *Evolution* **64**, 3354-3363 (2010).
225
226 Avise, J. C. *et al*. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge
227 between population genetics and systematics. *Annu. Rev. Ecol. Evol. Syst.* **18**, 489-522
228 (1987).
229
230 Ballard, J. W. O. & Kreitman, M. Is mitochondrial DNA a strictly neutral marker? *Trends*
231 *Ecol. Evolut.* **10**, 485-488 (1995).
232
233 Balloux, F., Handley, L. J. L., Jombart, T., Liu, H., & Manica, A. Climate shaped the
234 worldwide distribution of human mitochondrial DNA sequence variation. *Proc. R. Soc. Lond.*
235 *[Biol]* **276**, 3447-3455 (2009).
236
237
238 Barreto, F. S. & Burton, R. S. Evidence for compensatory evolution of ribosomal proteins in
239 response to rapid divergence of mitochondrial rRNA. *Mol. Biol. Evol.* **30**, 310-314 (2013).
240
241 Bergland, A. O., Tobler, R., González, J., Schmidt, P. & Petrov, D. 2016. Secondary contact
242 and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila*
243 *melanogaster*. *Mol. Ecol.* **25**, 1157-1174 (2016).

Boratyński, Z., Ketola, T., Koskela, E. & Mappes, T. The sex specific genetic variation of energetics in bank voles, consequences of introgression? *Evol. Biol.* **43**, 37-47 (2016).

Camus, M. F., Wolf, J. B., Morrow, E. H. & Dowling, D. K. Single nucleotides in the mtDNA sequence modify mitochondrial molecular function and are associated with sex-specific effects on fertility and aging. *Curr. Biol.* **25**, 2717-2722 (2015).

Camus, M. F., Wolff, J. N., Sgrò, C. M. & Dowling, D. K. Experimental evidence that thermal selection has shaped the latitudinal distribution of mitochondrial haplotypes in Australian fruit flies. Preprint at http://biorxiv.org/content/early/2017/04/26/103606 (2017).

Canestrelli, D. *et al*. The tangled evolutionary legacies of range expansion and hybridization. *Trends Ecol. Evolut.* **31**, 677-688 (2016).

Corbin, C., Heyworth, E. R., Ferrari, J. & Hurst, G. D. Heritable symbionts in a world of varying temperature. *Heredity* **118**, 10-20 (2017).

David, J. R. & Capy, P. Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* **4**, 106-111 (1988).

Doi, A., Suzuki, H. & Matsuura, E. T. 1999. Genetic analysis of temperature-dependent transmission of mitochondrial DNA in *Drosophila*. *Heredity* **82**, 555-560 (1999).

268 Dowling, D. K., Abiega, K. C. & Arnqvist, G. Temperature-specific outcomes of
269 cytoplasmic-nuclear interactions on egg-to-adult development time in seed beetles. *Evolution*
270 **61**, 194-201 (2007).
271
272 Dowling, D. K., Friberg, U. & Lindell, J. Evolutionary implications of nonneutral
273 mitochondrial genetic variation. *Trends Ecol. Evolut.* **23**, 546-554 (2008).
274
275 Dowling, D. K., Meerupati, T. & Arnqvist, G. Cytonuclear interactions and the economics of
276 mating in seed beetles. *Am. Nat.* **176**, 131-140 (2010).
277
278 Dowling, D. K. Evolutionary perspectives on the links between mitochondrial genotype and
279 disease phenotype. *BBA-Gen. Subjects* **1840**, 1393-1403 (2014).
280
281 Fontanillas, P., Depraz, A., Giorgi, M. S. & Perrin, N. Nonshivering thermogenesis capacity
282 associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew,
283 *Crocidura russula*. *Mol. Ecol.* **14**, 661-670 (2005).
284
285 Frank, S. A. & Hurst, L. D. Mitochondria and male disease. *Nature* **383**, 224 (1996).
286
287 Fry, A. J., Palmer M. R. & Rand, D. M. Variable fitness effects of *Wolbachia* infection in
288 *Drosophila melanogaster*. *Heredity* **93**, 379–389 (2004).
289
290 Galtier, N., Nabholz, B., Glemin, S. & Hurst, G. D. D. Mitochondrial DNA as a marker of
291 molecular diversity: a reappraisal. *Mol. Ecol.* **18**, 4541-4550 (2009).
292

293 Gemmell, N. J., Metcalfe, V. J. & Allendorf, F. W. Mother's curse: the effect of mtDNA on
294 individual fitness and population viability. *Trends Ecol. Evolut.* **19**, 238-244 (2004).
295
296 Greiner, S., Sobanski, J. & Bock, R. Why are most organelle genomes transmitted
297 maternally? *BioEssays* **37**, 80-94 (2015).
298
299 Hill, G. E. Mitonuclear ecology. *Mol. Biol. Evol*. **32**, 1917-1927 (2015).
300
301 Hoffmann, A. A., Hercus, M. & Dagher, H. Population dynamics of the *Wolbachia* infection
302 causing cytoplasmic incompatibility in *Drosophila melanogaster*. *Genetics* **148**, 221-231
303 (1998).
304
305 Hoffmann A. A. & Weeks A. R. Climatic selection on genes and traits after a 100 year-old
306 invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from
307 eastern Australia. *Genetica* **129**, 133-147 (2007).
308
309 Hurst, G. D. & Jiggins, F. M. Problems with mitochondrial DNA as a marker in population,
310 phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. R. Soc.*
311 *Lond. [Biol]* **272**, 1525-1534 (2005).
312
313 Innocenti, P., Morrow, E. H. & Dowling, D. K. Experimental evidence supports a sex-
314 specific selective sieve in mitochondrial genome evolution. *Science* **332**, 845-848 (2011).
315
316 James, J. E., Piganeau, G. & Eyre-Walker, A. The rate of adaptive evolution in animal
317 mitochondria. *Mol. Ecol.* **25**, 67-78 (2016).

Kazancıoğlu, E., & Arnqvist, G. The maintenance of mitochondrial genetic variation by negative frequency-dependent selection. *Ecol. Lett.* **17**, 22-27 (2014).

Kivisild, T. *et al*. The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**, 373–387 (2006).

Koukou, K. *et al*. Influence of antibiotic treatment and *Wolbachia* curing on sexual isolation among *Drosophila melanogaster* cage populations. *Evolution* **60**, 87-96 (2006).

Kriesner, P., Conner, W. R., Weeks, A. R., Turelli, M. & Hoffmann, A. A. (2016). Persistence of a *Wolbachia* infection frequency cline in *Drosophila melanogaster* and the possible role of reproductive dormancy. *Evolution* **70**, 979-997 (2016).

Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **519**, 171-180 (2015).

Luo, Y., Gao, W., Liu, F. & Gao, Y. Mitochondrial nt3010G-nt3970C haplotype is implicated in high-altitude adaptation of Tibetans. *Mitochondrial DNA* **22**, 181-190 (2011).

Matsuura, E. T., Niki, Y. & Chigusa, S. I. Temperature-dependent selection in the transmission of mitochondrial-DNA in *Drosophila*. *Jpn. J. Genet.* **68**, 127-135 (1993).

Mishmar, D. *et al*. Natural selection shaped regional mtDNA variation in humans. *P. Natl. Acad. Sci. USA* **100**, 171-176 (2003).

343 Moritz, C., Dowling, T. E. & Brown, W. M. Evolution of animal mitochondrial DNA:
344 relevance for population biology and systematics. *Annu. Rev. Ecol. Evol. Syst.* **18**, 269-292
345 (1987).
346
347 Mossman, J. A., Biancani, L. M. & Rand, D. M. Mitonuclear epistasis for development time
348 and its modification by diet in *Drosophila*. *Genetics* **203**, 463-484 (2016).
349
350 Rand, D. M., Dorfsman, M. & Kann, L. M. Neutral and non-neutral evolution of *Drosophila*
351 mitochondrial DNA. *Genetics* **138**, 741-756 (1994).
352
353 Rasgon, J. L., Cornel, A. J. & Scott, T. W., 2006. Evolutionary history of a mosquito
354 endosymbiont revealed through mitochondrial hitchhiking. *Proc. R. Soc. Lond. [Biol]* **273**,
355 1603-1611 (2006).
356
357 Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V. & Wallace, D. C. Effects of
358 purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**, 223-
359 226 (2004).
360
361 Schuler, H. *et al*. The hitchhiker's guide to Europe: the infection dynamics of an ongoing
362 *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. *Mol. Ecol.* **25**,
363 1595-1609 (2016).
364
365 Sun, C., Kong, Q.-P. & Zhang, Y.-P. The role of climate in human mitochondrial DNA
366 evolution: A reappraisal. *Genomics* **89**, 338-342 (2007).
367

368 Sunnucks, P., Morales, H. E., Lamb, A. M., Pavlova, A. & Greening, C. Integrative
369 approaches for studying mitochondrial and nuclear genome co-evolution in oxidative
370 phosphorylation. *Front. Genet.* **8**, 25 (2017).
371
372 Versace, E., Nolte, V., Pandey, R. V., Tobler, R. & Schlötterer, C. Experimental evolution
373 reveals habitat-specific fitness dynamics among *Wolbachia* clades in *Drosophila*
374 *melanogaster*. *Mol. Ecol.* **23**, 802-814 (2014).
375
376 Wallace, D. C. Genetics: Mitochondrial DNA in evolution and disease. *Nature* **535**, 498-500
377 (2016).
378
379 Wolff, J. N., Ladoukakis, E. D., Enríquez, J. A. & Dowling, D. K. Mitonuclear interactions:
380 evolutionary consequences over multiple biological scales. *Phil. Trans. R. Soc. B* **369**,
381 20130443 **(**2014).
382
383 Wolff, J. N., Tompkins, D. M., Gemmell, N. J. & Dowling, D. K. Mitonuclear interactions,
384 mtDNA-mediated thermal plasticity, and implications for the Trojan Female Technique for
385 pest control. *Sci. Rep.* **6**, 30016 (2016).

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.

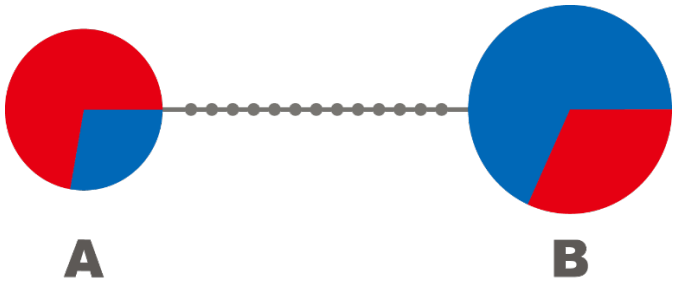
Author Contributions

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.

406 **Figures**



407

408 **Figure 1: Relationship of mtDNA haplotypes A and B.**

409 The circle area for each haplotype is proportional to its frequency in the wild sample (A=18

410 females, B=22 females). Colours indicate the sampling region: Townsville (red, 20 females)

411 and Melbourne (blue, 20 females). Small grey circles represent genotyped-SNP divergence.

412

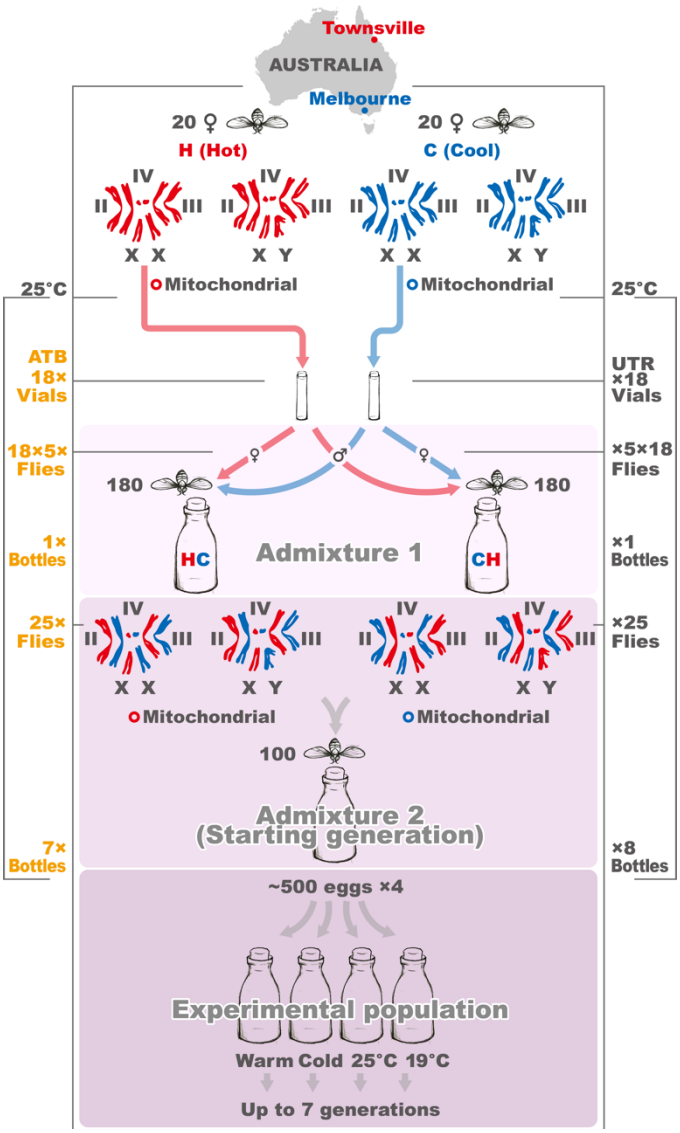


Figure 2: Scheme of experimental evolution by hybridization of differentially thermally-adapted subpopulations of fruit fly.

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 = 18×5(♀H) + 18×5(♂C). In parallel in Admixture 1, we performed the reciprocal cross wherein H <=>

424 C above, denoted by $\mathrm{CH} = 18\times5(♀\mathrm{C}) + 18\times5(♂\mathrm{H})$. Each bottle contained 90 males and
425 90 females (180 flies). In the following generation, at Admixture step 2, we combined 25
426 virgin females and 25 virgin males from HC bottles together with 25 virgin females and 25
427 virgin males from CH bottles, $25\ (♂\mathrm{CH}) + 25(♀\mathrm{HC}) + 25(♀\mathrm{CH}) + 25(♂\mathrm{HC})$, across 15
428 biological replicates (7 of which were descendants of flies treated by antibiotics, 8 of which
429 were descendants of untreated flies). At this stage, all flies had been maintained in standard
430 laboratory conditions (25°C) for 16 generations (14 generations as isofemale lineages, 2
431 during the admixture process). We then divided each of these 15 biological replicates into 4
432 subpopulations, subjecting each subpopulation to one of four thermal treatments (19°C, 25°C,
433 fluctuating cold, and fluctuating warm), with each experimental subpopulation containing
434 around 500 individuals. On the left side of the figure, yellow text denotes sample sizes
435 associated with each stage of the admixture process for flies whose ancestors had been
436 exposed to antibiotic treatment (ATB), while grey text on the right corresponds with
437 untreated flies (UTR).

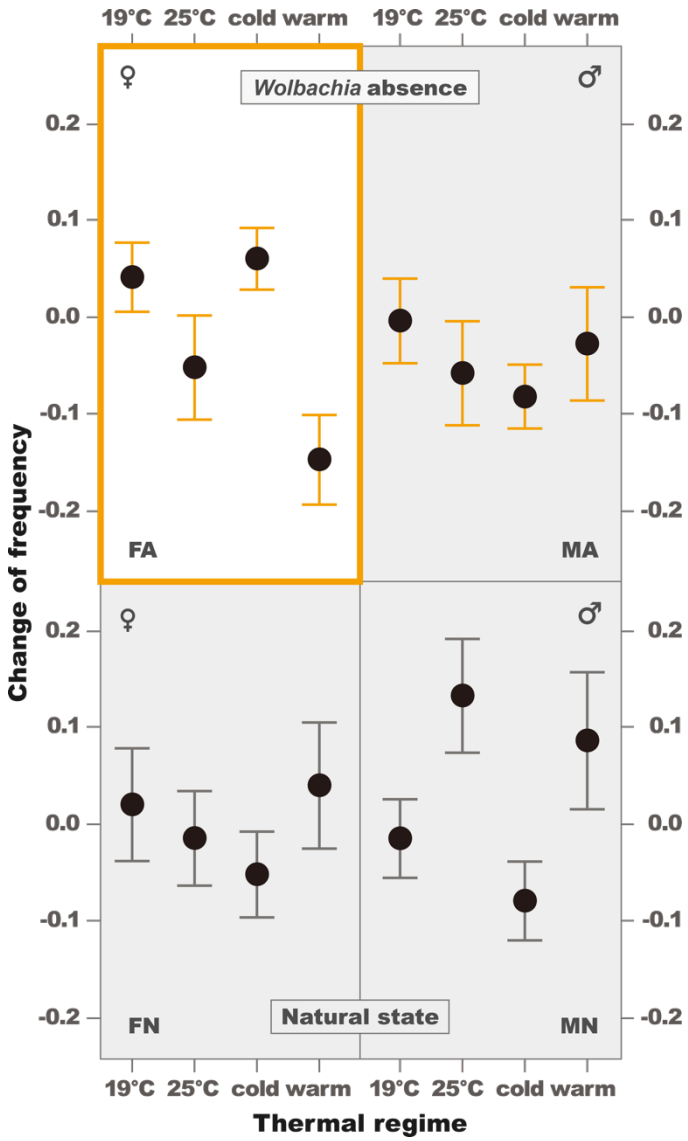


Figure 3: Mean change of mtDNA haplotype B frequency per thermal environment. Interaction plots depict changing frequencies (final generation -initial generation) in stable 19°C, 25°C, fluctuating cold, and fluctuating warm environments for female and male descendants of flies treated by antibiotics (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_∞) for large samples, assuming the validity of the asymptotic χ^2 distribution. The yellow background indicates statistical significance.

Tab 3:	(I) mixed model			(II) fixed effects ($\rho_g^n \equiv 0$)			(III) *lme4*		(IV) ANOVA	
group	W	P_∞	P	W	P_∞	P	W	P	F	P
FA	16.5667	0.0009	0.0147	15.8185	0.0012	0.0169	17.4096	0.0006	4.2570	0.0152
FN	1.7616	0.6233	0.6341	1.7296	0.6304	0.6406	1.7788	0.6195	0.4696	0.7059
MA	2.0993	0.5521	0.5715	2.1414	0.5436	0.5639	1.4547	0.6928	0.4156	0.7433
MN	10.0702	0.0180	0.0494	10.0724	0.0180	0.0521	8.9906	0.0264	2.6222	0.0702

456 **Methods** (3000 words)

457 **Experimental procedures**

458 Wild subpopulations of *D. melanogaster* were sampled in Australia. We sampled a hot
459 adapted subpopulation (H; Townsville: -19.26, 146.79) in the north-east, and a cool adapted
460 subpopulation (C; Melbourne: -37.99, 145.27) in the south of the continent. We collected
461 fertilised females and established 20 isofemale lineages from each wild population. After 3
462 generations of acclimatisation to laboratory conditions, we split each isofemale lineage into
463 two replicates, and treated one replicate of each lineage with 0.164 mg mL^{-1} tetracycline in
464 food for 3 generations to remove any intracellular and cytoplasmically-inherited bacteria,
465 such as *Wolbachia* (Clancy & Hoffman 1998). We then propagated these lineages for a
466 further 10 generations to mitigate any effects of the antibiotic treatment. Flies were reared at
467 25°C, on a 12:12 hour light:dark cycle, in 10 dram plastic vials, on a potato-dextrose-agar
468 medium, with *ad libitum* live yeast added to each vial. All isofemale lineages were then
469 transferred from our laboratories in Australia to those in Japan, and their food medium
470 changed to a corn flour-glucose-agar medium (Supplementary Table 2), with *ad libitum* live
471 yeast added to each vial. They acclimatized for a further 3 generations at 25°C, before
472 entering an admixture process described below, in order to set up a series of replicated
473 experimental populations.

474 We pooled 5 virgin females (♀) from each of 18 of the H isofemale lineages,
475 mentioned above, with 5 virgin males (♂) from each of 18 of the C lineages in one bottle
476 (HC = 18 x 5 ♀H + 18 x 5 ♂C), and 5 virgin females from each of the 18 C isofemale
477 lineages with 5 virgin males from each of the 18 H isofemale lineages in another bottle (CH =
478 18 x 5 ♀C + 18 x 5 ♂H), such that each bottle contained 90 males and 90 females. This step
479 was performed separately for flies sourced from the tetracycline-treated isofemale lineages
480 and flies sourced from untreated isofemale lineages, separately (Fig. 2; Supplementary Table

481 1). We then allowed the flies to lay eggs over 8 consecutive days, and transferred them to
482 fresh bottles as indicated in Supplementary Table 3. We reared all experimental populations
483 in 250 ml bottles on a corn flour-glucose-agar medium. In the next step, we mixed F1
484 offspring (25 virgin males and 25 virgin females) from the HC bottles with corresponding F1
485 offspring (25 virgin males and 25 virgin females) from the CH bottles (25 ♂CH + 25 ♀HC +
486 25 ♀CH + 25 ♂HC). We established 7 experimental populations from the tetracycline-treated
487 isofemale lineages and 8 experimental populations from the untreated lineages. We allowed
488 flies of these populations to mate and lay eggs at 25°C, and then we transferred bottles with
489 approximately 500 eggs into four thermal regimes, represented by cool versus warm
490 temperatures, on either a constant or fluctuating temperature cycle. Bottles maintained in the
491 cool and constant temperature were kept at a constant 19°C, and those in the warm and
492 constant temperature at 25°C temperature. We used Environmental Chambers (MIR-154,
493 Sanyo) to generate fluctuating thermal conditions that are common in areas of origin of our
494 experimental populations (The Australian Government, Bureau of Meteorology), Melbourne:
495 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);
496 16:00(16°C); 20:00(15°C); 22:00(14°C), and Townsville: 8:00(27°C); 10:30(28°C);
497 20:00(27°C); 22:30(26°C); 0:00(24°C); 8:00(26°C); 10:30(28°C); 20:00(27°C); 22:30(26°C);
498 0:00(25°C). The temperature in all conditions was continually monitored and in fluctuating
499 conditions recorded by Thermo-hydro SD Data Loggers (AD-5696; A&D Ltd). We
500 propagated all replicate populations for three months (3 or 7 successive generations
501 depending on the thermal condition; Supplementary Table 4). We regulated the size of each
502 population by trimming egg numbers per generation to approximately 500 eggs. At the end of
503 the experimental evolution period of three months, adult flies were collected and fixed in
504 95% ethanol.
505

506 **Data collection**

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

511 We mapped all reads on published mitogenomic sequence NC 001709 in
512 Geneious R6 (http://www.geneious.com, Kearse et al. 2012). We observed overall
513 mitogenomic variability and picked 14 mtDNA polymorphic sites (SNPs), which are not
514 unique to the H or C populations, which delineate all flies into one of two corresponding
515 mtDNA haplotypes, which we denote A and B, and which are described in the Results (Fig.
516 1, Supplementary Table 5), using multiplexPCR and MALDI/Tof (mass spectrometry;
517 Geneworks, Australia). We genotyped virtually all flies in the starting generation (formed by
518 50 males and 50 females per each of the 15 replicates), and then more than 24 males and 25
519 females per bottle in final generation of the experiment, upon completion of experimental
520 evolution. The minimal number of sequenced samples was estimated assuming the relative
521 thermal effect of 10% at power of $1 - \beta = 70\%$ and the F-distribution. In total, we
522 genotyped 4410 individuals (Supplementary Table 1).

523

524 **Data analysis**

525 A multilevel linear model showed a significant three-way interaction between
526 thermal regime, sex, and antibiotic treatment on the change in frequency of haplotype B
527 (Extended Table 1). Accordingly, we divided the total of 60 experimental populations into 4
528 groups: FA=females whose ancestors had been treated with antibiotics, FN=females
529 untreated, MA=males whose ancestors had been treated with antibiotics, and MN=males
530 untreated. Flies in each of these groups are propagated under 1 of 4 thermal treatments: 19°C,

25°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 $n = 1,2,\dots,N$ 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)

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

where θ_{gi} denotes the mean value of the frequency difference, obtained for thermal regime i in group g (averaged over N samples) and ϵ_{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

$$\langle \epsilon_{gi}^{n} \epsilon_{gj}^{n'} \rangle = \delta^{nn'} \sigma_{gi} \sigma_{gj} [\delta_{ij} (1 - \rho_g^n) + \rho_g^n \,] \tag{2}$$

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 ρ_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

$$\hat{\theta}_{gi} = N^{-1} \textstyle\sum_n y_{gi}^{n} \;\; , \;\; \hat{\sigma}_{gi}^2 = N^{-1} \textstyle\sum_n (y_{gi}^{n} - \hat{\theta}_{gi})^2 \tag{3}$$

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

$$12\rho_g^n = -\left(1+3\rho_g^n\right)\left(1-\rho_g^n\right)(\delta y_g^n)^{\intercal}\left[\Gamma\left(\rho_g^n\right)\right]\Omega\,[\Gamma\left(\rho_g^n\right)](\delta y_g^n) \qquad (4a)$$

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

$$\Gamma_{ii} = (1+2\rho_g^n)/[\hat{\sigma}_{gi}^2\left(3\rho_g^n+1\right)(1-\rho_g^n)],\ \Gamma_{i\neq j} = -\rho_g^n/[\hat{\sigma}_{gi}\hat{\sigma}_{gj}\left(3\rho_g^n+1\right)(1-\rho_g^n)] \qquad (4b)$$

and Ω is a traceless matrix of rank-1, such that $\Omega_{ij} = \hat{\sigma}_{gi}\hat{\sigma}_{gj}(1-\delta_{ij})$. 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:

$$(V_g)_{ij} \equiv \langle\delta\hat{\theta}_{gi}\delta\hat{\theta}_{gj}\rangle = \langle\left(\hat{\theta}_{gi}-\langle\hat{\theta}_{gi}\rangle\right)\left(\hat{\theta}_{gj}-\langle\hat{\theta}_{gj}\rangle\right)\rangle = -(H_g)_{ij}^{-1} = \left[\textstyle\sum_b \Gamma_{ij}(\hat{\rho}_g^n)\right]^{-1} \qquad (5)$$

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

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 $\hat{\theta}_{gi} = \mu_g + \alpha_{gi}$, where μ is the average over $T=4$ levels, $\mu_g \equiv \sum_i \hat{\theta}_{gi}/T$, and $\sum_{i=1}^T \alpha_{gi} = 0$. The F-statistic as defined by the standard 1-factor ANOVA, is then: $F_g =$ [(mean-square-error between levels)/(mean-square-error within levels)]. Namely, for each group separately,

$$F_g = [\textstyle\sum_{i=1}^T \alpha_{gi}^2/(T-1)]/\left\{\sum_{i=1}^T \hat{\sigma}_{gi}^2/[T(N-1)]\right\} \qquad (g=1,2,\ldots,4=G) \qquad (6)$$

with $\{\hat{\theta}_{gi}, \hat{\sigma}_{gi}\}$ 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 $= 78\%$). The *P*-value and power of F_g are given by: $P = 1 - \Phi_{\nu_1,\nu_2}(F_g, \lambda = 0)$ and power $\equiv 1 - \beta = 1 - \Phi_{\nu_1,\nu_2}(F_g, \lambda = TF_g)$, where $\Phi_{\nu_1,\nu_2}(\mathrm{f}, \lambda)$ is the cumulative non-central F-distribution with degrees-of-freedom $\nu_1 = T - 1$, $\nu_2 = T(N - 1)$, and non-centrality parameter λ.

In addition to F_g, we also studied the Wald-statistic of the thermal effects which is defined as $W_g \equiv \alpha_g^{\intercal}\left(\langle \delta\alpha_g \delta\alpha_g^{\intercal} \rangle\right)^{-1} \alpha_g$, where α_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 $\delta\alpha \equiv \alpha - \langle \alpha \rangle$. Introducing $\vec{\theta}_g \equiv (\theta_{g1}, \dots, \theta_{g4})$, $\vec{\tau}_g \equiv (\mu_g, \alpha_{g1}, \alpha_{g2}, \alpha_{g3})$ and changing variables, $\vec{\tau}_g = \mathcal{R}\vec{\theta}_g$, one finds that

$$W_g = (R\hat{\theta}_g)^{\intercal}\left(RV_gR^T\right)^{-1}(R\hat{\theta}_g) \qquad (g = 1,2,\dots,4 = G) \quad (7)$$

where R is a 3×4 transformation matrix $\left(R_{ii} = 3/4, R_{i\neq j} = -1/4\right)$, and V_g is the correlation matrix previously given in Eq. (5). We have applied Eq. (7) both to a mixed model [in which ρ_g^n are obtained using Eq. (4a)], and to a fixed-effects model (in which all ρ_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: $P = 1 - \Phi_{\nu_1,\nu_2}\left[\frac{W_g}{T-1}, \lambda = 0\right]$, power $= 1 - \beta = 1 - \Phi_{\nu_1,\nu_2}\left[\frac{W_g}{T-1}, \frac{\lambda_w}{T-1}\right]$, where $\nu_1 = T - 1$ and $\nu_2 = N(T - 2) - (T - 1)$. 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 χ^2 with (T-1) degrees of freedom. As a result, the corresponding *P*-values decrease significantly (see Table 1).

While the χ^2distribution 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 χ^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:

$$\hat{\theta} = [+0.0409, -0.0506, +0.0601, -0.1477], \hat{\sigma} = [\ 0.0936\ ,\ \ 0.1420,\ \ 0.0857,\ \ 0.1226]$$
$$\hat{\rho} = [+0.0408, +0.0427, +0.0917, +0.2002, -0.0771, +0.0758, -0.1162] \tag{8}$$

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 $\{\hat{\theta}_{gi}, \hat{\sigma}_{gi}, \hat{\rho}_g^n\}$ 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 χ^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 χ^2distribution. Consequenly, the *P*-value of the experiment decreases: $P = 0.0147 \rightarrow 9\times10^{-4}$ 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. 099.465 (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, $\hat{\theta}_{gi}^2 = \sum_i \hat{\sigma}_{gi}^2 \,/(NT)$. 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, $\left|\delta\hat{\theta}_{gi}(\mathrm{I}) - \delta\hat{\theta}_{gi}(\mathrm{II})\right| / \left|\hat{\theta}_{gi}\right| \leq 5\%$.

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 χ^2distribution and, as a result, the *P*-value decreases by an order of magnitude to give a $P = 6\times10^{-4} - 9\times10^{-4}$ and power $\cong$ 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.

MtDNA is transmitted mainly maternally, but paternal mtDNA 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

Bates, D., Mächler, M., Bolker, B., & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, i01(2015).

Camus, M.F., Wolff, J. N., Sgrò, C. M. & Dowling, D. K. Experimental evidence that thermal selection has shaped the latitudinal distribution of mitochondrial haplotypes in Australian fruit flies. Preprint at http://biorxiv.org/content/early/2017/04/26/103606 (2017).

Clancy, D. J. & Hoffmann, A. A. Environmental effects on cytoplasmic incompatibility and bacterial load in *Wolbachia*-infected *Drosophila simulans*. *Entomol. Exp. Appl.* **86**, 13-24 (1998).

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

Engle R. F. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. in *Handbook of Econometrics, Vol. II* (eds Grilliches Z. & Intriligator M. D.) 775-826 (North-Holland, 1984).

Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647-1649 (2012).

Kutner, M. H., Nachtsheim, C. J., Neder J. & Li, W. Applied linear statistical models, 5th Ed. (McGrew-Hill, 2013).

Nunes, M. S., Dolezal, M. & Schlötterer, C. Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*. *Mol. Ecol.* **22**, 2106–2117 (2013).

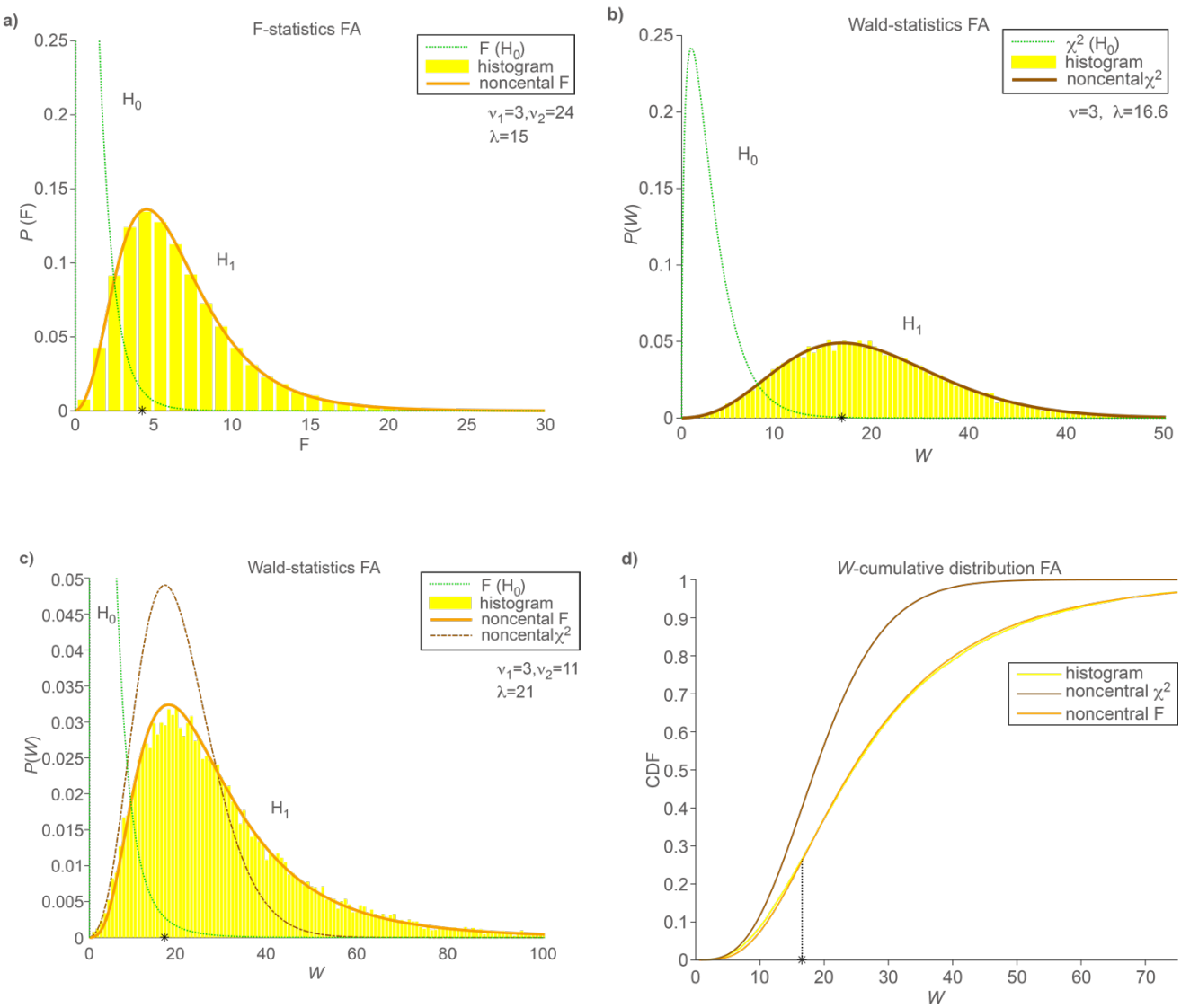
Parker, T. Finite-sample distributions of the Wald, likelihood ratio and Lagrange multiplier test statistics in the classical linear model. Commun. Stat. Theory Methods 46, 5195-5202 (2017).

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ (2015).

RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/ (2015).

The Australia Government, Bureau of Meteorology at http://www.bom.gov.au (2012)

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_g statistic (Eq. 6) compared to the non-central F-distribution (H_1) with parameters $\nu_1 = T - 1 = 3,\ \nu_2 = T(N-1) = 24, \lambda_F = \overline{m}_F \nu_1(\nu_2 - 2)/2 - \nu_1 = 15$. The probability distribution of the null hypothesis (H_0) is obtained by setting $\lambda_F = 0$.

(b) The non-central χ^2distribution (H_1) describing the Wald statistic W_g in the limit of large samples $(N \gg 1)$. Here, $\nu_1 = T - 1 = 3$, $\lambda_w = \overline{m}_w - \nu_1 = 16.6$. 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 $\nu_1 = T - 1 = 3, \nu_2 = N(T-2) - (T-1) = 11$. $\lambda_W = \overline{m}_W \nu_1(\nu_2 - 2)/[2(T-1)] - \nu_1 = 21$. For (H_0), $\lambda_W = 0$. The asymptotic χ^2distribution (b) is shown as a dashed line.

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

Note that the only free (fitting) parameters in (a-d) are the sampled means $\{\overline{m}_F, \overline{m}_W\}$ which are obtained by averaging over all realizations of $\{F_g, W_g\}$, respectively.

707 **Extended Table 1: Multilevel model examining the effect of sex, antibiotic treatment,**

708 **and thermal regime on B haplotype frequency change, as a response variable.**

709

	χ^2	d.f.	$P\ (>\chi^2)$
Intercept	0.1518	1	0.696824
sex	0.8269	1	0.363163
antibiotic treatment	0.0850	1	0.770666
thermal regime	1.9834	3	0.575866
antibiotic treatment: thermal regime	8.9186	3	0.030393
sex: antibiotic treatment	0.0326	1	0.856811
sex: thermal regime	15.4599	3	0.001463
sex: antibiotic treatment: thermal regime	10.7419	3	0.013207
	σ		
ID	0.119988		
Population	0.009593		

710

711

712 Sex, antibiotic treatment, and thermal regime were modelled as fixed effects. ID and

713 Population were modelled as random effects.

714 Population indicates the biological replicate i.e., group of 4 bottles descending from a single

715 starting bottle. ID pairs males and females which are sharing the same bottle. The yellow

716 background indicates statistical significance.

717

Extended Table 2: The fixed parameters, $\widehat{\theta}_{gi}$, and their errors, $\delta\widehat{\theta}_{gi}$, for 4 population groups (FA, FN, MA, MN), evaluated by three schemes of estimation: (I) linear mixed model (II) fix effects only (III) *lme4.*

model of error(†)	group(‡)	19°C		25°C		fluctuating cold		fluctuating warm	
		$\widehat{\theta}_1$	$\pm\delta\widehat{\theta}_1$	$\widehat{\theta}_2$	$\pm\delta\widehat{\theta}_2$	$\widehat{\theta}_3$	$\pm\delta\widehat{\theta}_3$	$\widehat{\theta}_4$	$\pm\delta\widehat{\theta}_4$
(I)	FA(7)	+0.0409	0.0348	-0.0506	0.0529	+0.0601	0.0319	-0.1477	0.0457
(II)			0.0354		0.0537		0.0324		0.0464
(III)			0.0428		0.0428		0.0428		0.0428
(I)	FN(8)	+0.0195	0.0584	-0.0150	0.0490	-0.0525	0.0443	+0.0400	0.0657
(II)			0.0589		0.0494		0.0446		0.0663
(III)			0.0554		0.0554		0.0554		0.0554
(I)	MA(7)	-0.0029	0.0438	-0.0563	0.0543	-0.0809	0.0335	-0.0266	0.0584
(II)			0.0442		0.0548		0.0338		0.0590
(III)			0.0489		0.0489		0.0489		0.0489
(I)	MN(8)	-0.0144	0.0409	+0.1329	0.0590	-0.0795	0.0405	+0.0866	0.0706
(II)			0.0418		0.0604		0.0414		0.0723
(III)			0.0555		0.0555		0.0555		0.0555

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

726

727 Supplementary Information Legends

728

729 **Supplementary Table 1: List of samples**

730 In experimental populations ATB1-ATB7, the ancestors had been exposed to antibiotic

731 treatment, while experimental populations UTR1-UTR8 correspond with untreated flies.

732 **Supplementary Table 2: Fly food composition**

733 **Supplementary Table 3: Starting dates of experimental populations.**

734 Foundation date marks the date at which virgin flies were combined in a bottle as outlined in

735 Admixture Step 2 (Fig. 2) to form the Starting generation. In Admixture Step 1, we allowed

736 their parents to lay eggs for about 1 day, and transferred them to a new bottle. This process

737 was repeated across nine days. We call the process by which we transfer the flies to a new

738 bottle a "tip". Virgin flies of each sex were sourced from several tips, in order to ensure we

739 had an adequate supply of flies to initiate the experimental populations. We show the dates of

740 maternal ovipositioning and virgin collection in a separate column (date). Number means the

741 number of virgin flies sourced from the tip.

742 **Supplementary Table 4: Propagation of experimental populations in dates.**

743 Foundation date corresponds with Supplementary Table 3. We imposed the four different

744 thermal regimes on multiple generations starting by eggs laid by the Starting generation.

745 Within each generation, the flies of each experimental bottle were transferred to new bottles,

746 and the column "tip number" reflects the "tip" that was used to propagate the next generation

747 per experimental population. Although the tip we used to initiate each experimental

748 population varied in Generation 1, in subsequent generations, we propagated experimental

749 populations mostly from the first tip.

750 **Supplementary Table 5: Selected SNPs characteristics and individual haplotypes.**

751 **Supplementary Table 6: The Wald statistics of the mixed model obtained with *lme4*,**

752 **when the frequencies f_{gi}^{n} are subtracted from the corresponding frequencies in the**

753 **starting generation (males and females combined).**