ECOLOGY LETTERS: Supporting Information

The maintenance of mitochondrial genetic variation by negative frequency-dependent selection

Erem Kazancıoğlu and Göran Arnqvist

DETAILED MATERIALS AND METHODS

Constructing introgression lines

All seed beetles were reared on black-eyed beans (*Vigna unguiculata*) in climate chambers (Sanyo MLR-352H) at 29°C, 55% RH and a 12:12 diurnal light cycle. Polymorphism in mtDNA is pronounced in seed beetles, both within (Tuda *et al*. 2004) and between (Arnqvist & Tuda 2010; Arnqvist *et al*. 2010) natural populations. For example, Tuda *et al*. (2004) found that natural populations hosted up to nine distinct mtDNA haplotypes. To date, 68 distinct extant mtDNA haplotypes have been sequenced from *C. maculatus*. To construct the cytonuclear introgression lines used in our experiment, we followed a modified version of the protocol detailed in Dowling *et al*. (2007). In short, we used three different outbred stock populations of *C. maculatus* to generate 9 fully crossed combinations of three distinct mitochondrial and nuclear lineages (stocks B, Y and C [Arnqvist *et al*. 2010]), with 2 replicates per combination, resulting in a

total of 18 distinct introgression lines. The three stock populations were selected from a larger set of potential populations as previous research has shown that their mtDNA genotypes differ (see below), that mitonuclear genetic variation across these populations affects life history phenotypes and that mtDNA sequence divergence across haplotypes is associated with life history phenotype divergence (Dowling *et al*. 2007; Arnqvist *et al*. 2010; Dowling *et al*. 2010). We first mated two randomly selected virgin females from each stock population each to a single male from their own stock population. These females were essentially mitochondrial "Eves" for subsequent introgression lines carrying their specific mitochondrial haplotype. We then set up distinct lines by mating groups of full-sib virgin daughters of mitochondrial "Eves" to males from one of the three stock populations, in all 9 possible combinations (2 replicates of each combination, starting with different mitochondrial "Eves"). In each subsequent generation, we mated 3-5 virgin females from each of the 18 lines to the same number of randomly selected males from the same stock population as their fathers. In each generation of introgression, we placed mated beetles together into closed petri dishes with 36 grams of black-eyed beans, *Vigna unguiculata*. We used this repeated introgressive backcrossing scheme for 15 subsequent generations to disassociate each of the sampled mitochondrial genomes from the nuclear genome that it was originally co-expressed with. After 15 generations of backcrossing, more than 99.99 % of the nuclear genome originally associated with each mitochondrial haplotype will have been replaced with nuclear genes derived from one of the three stock populations, assuming no strong selection for specific mitonuclear combinations during the introgression. We observed no elevated rates of hatching failure of eggs or juvenile mortality during the

introgressive backcrossing and adult viability and fecundity were not depressed. Nevertheless, each of the specific 9 introgression crosses was replicated twice (see above) to allow a statistical evaluation of the potential effects of capricious line specific adaptation (i.e., independent of mitochondrial and nuclear genotype) on genotype frequency changes during the selection experiments.

Many insects are infected by maternally inherited, cytoplasmic bacteria such as *Wolbachia*, which can confound the results of experiments that aim to investigate mitochondrial effects on fitness. *Wolbachia* infections have been screened for in many *C. maculatus* populations, including those used here, but have never been detected (Kondo *et al*. 1999; Kageyama *et al*. 2010; Tuda *et al*. 2006). To preclude the possibility that our introgression lines nevertheless harboured cytoplasmic bacterial infections, we treated all introgression lines with an antibiotic treatment between generations 13 and 14. To ensure that the antibiotic was ingested by beetle larva, we followed the protocol proved to be efficient by Kondo et al. (2002). Briefly, black-eyed beans were ground and tetracycline was added to the resulting bean powder along with water. This mix was used to mould pellets of a size and shape that resemble black-eyed beans. We then freeze-dried these pellets and coated them with a cellulose solution to reproduce the peel of black-eyed beans. These manufactured and "artificial" beans were then used to generate the adults of generation 14 of our introgression lines. Antibiotic treatments can affect overall mitochondrial function in insects (Ballard & Melvin 2007). However, we are unaware of any evidence suggesting that such an effect can be differential across different mtDNA haplotypes and, more importantly, such an effect cannot by itself generate negative frequency

dependent selection across haplotypes. We also note that (i) we allowed for four generations in between the antibiotic treatment and the start of our experiments and (ii) metabolic performance in antibiotic-treated *C. maculatus* lines is significantly related to mtDNA sequence divergence across haplotypes (Arnqvist *et al*. 2010), supporting a causal link between mtDNA genotype and phenotype.

After 15 generations of backcrossing, we placed mated beetles into glass jars with 85 grams of black-eyed beans to increase population size such that enough beetles were produced for the selection experiment. At generation 17, we once again crossed introgression line females with randomly selected males from respective stock populations to further prevent the possibility of line specific adaptation. We started the selection experiment at generation 18.

mtDNA sequencing

Previous mtDNA sequencing efforts of two mitochondrial genes (1005 bp of cytochrome oxidase subunit I [COI] and 473 bp of cytochrome b [Cyt-b]) have documented substantial genetic variation across the three mtDNA haplotypes used here (Arnqvist & Tuda 2010; Arnqvist *et al*. 2010). Approximately 2-3% of all mtDNA sites are polymorphic and the haplotypes are known to differ also in encoded amino acids. For example, one of the 335 amino acids encoded by the sequenced segment of COI is polymorphic (valine in C and isoleucine in the other two haplotypes) and one of 157 amino acids encoded by the sequenced segment of Cyt-b is polymorphic (serine in Y and asparagine in the other two haplotypes) (Arnqvist *et al*. 2010).

Here, we determined the mitochondrial haplotype of individual beetles using targeted sequencing of a region of COI previously shown to be polymorphic among these haplotypes (Tuda *et al*. 2006; Arnqvist & Tuda 2010; Arnqvist *et al*. 2010). To validate the mitochondrial DNA distinctiveness of each stock population (B, Y and C), we first collected all females from each line four days after mating at the start of generation 16, when most mating and egg laying was complete but beetles were still alive. We extracted DNA from fresh beetles using QIAGEN DNeasy Blood and Tissue kit and used the primers modC1J-1751 and revC1J-2441 described by Tuda *et al*. (2006) to amplify a region of COI. During PCR, we subjected 25ul reactions (2.5ul 10X Taq buffer, 1.5ul 25mM $MgCl_2$, 2ul 10mMdNTP mix, 10ng of each primer, 0.125ul of Taq polymerase [50units/ul], dH_20) to 35 cycles of 94^oC for 30s, 45^oC for 1m, and 60^oC for 3m in a thermal cycler (GeneAmp PCR System, Applied Biosystems). We cleaned the PCR products using an Exo-SAP protocol and obtained ~650bp long sequences using an automated DNA sequencer (ABI3730XL DNA Analyzer, Applied Biosystems). Analysis of these DNA sequences showed 16 sites that were polymorphic among the three stock populations, recovering the sequence variation previously described from the dominating mtDNA haplotype within each of these populations (Arnqvist & Tuda 2010; Arnqvist *et al*. 2010). We thus used these sites to unambiguously determine the population origin of the mitochondrial haplotype of focal individuals. The sequencing effort at generation 16 (prior to the selection experiment) fully validated the within line homogeneity, and between line distinctiveness, of mtDNA haplotypes of the lines used.

Experimental evolution

We set up a very large number of experimental populations, each with a population size of 500 beetles, in the laboratory. We refer to each such experimental population as a replicate. Briefly, the exact mtDNA haplotype frequency at the start of the experiment was known and we then estimated the net change in haplotype frequency that occurred over 10 generations of evolution in each replicate. Each replicate was founded by a mix of beetle individuals from two different introgression lines, at a precisely known frequency (see below), such that all individuals in a given replicate population shared the same nuclear genetic background while carrying one of two possible alternative mitochondrial haplotypes. We ensured that, at the onset of the experiment, the frequency of one of the two mitochondrial haplotypes was precisely four times as high as the alternative one by founding each replicate with 4 × 1 ml (i.e., 4 ml) of beetles from an introgression line that we term "common" and 1 × 1 ml (i.e., 1 ml) of beetles from another introgression line that we term "rare" (1ml of beetles corresponds to ≈ 100 beetles). Thus, the population level mtDNA haplotype frequency at the start of the experiment was exactly 20% of the rare haplotype and 80% of the common haplotype. We made every conceivable effort to treat beetles assigned to be "common" and "rare" identically during the founding of our replicate populations. We employed a fully crossed three-way design in which the factors were mitochondrial haplotypes in competition (three levels: B vs Y; B vs C; Y vs C), nuclear genetic background (three levels; B, Y, C) and starting frequency of the focal haplotype (two levels; common or rare), resulting in a total of 3 × 3 × 2 = 18 treatment level combinations. We replicated each cell in the design 10 times (N = 180 replicate

independently evolving populations). For each cell in our design, half of the replicates were performed with one of the two replicate introgression lines (see above) and half with the other, to allow assessment of any line specific adaptation (see below).

Replicate populations were reared in 1 litre plastic jars (95 × 95 × 110 mm) kept in climate chambers (Sanyo MLR-352H) at 29°C, 55% RH and a 12:12 diurnal light cycle. As laboratory environments are typically relatively homogenous, we elevated environmental heterogeneity within each evolving replicate population in two ways. First, populations were maintained on a heterogeneous food resource, consisting of 200 grams (per replicate and generation) of an equal-weights mixture of black-eyed beans (*V. unguiculata*), adzuki beans (*V. angularis*) and mung beans (*V. radiata*). The model species used here, *C. maculatus*, utilizes all of these seeds as hosts in nature. However, egg production, growth rate and development time of *C. maculatus* varies across these hosts (Giga & Smith 1987). Second, we created thermal heterogeneity by means of an air circulation system, in which a system of plastic tubes carrying cold air at a constant flow rate was piped through all rearing jars (along the bottom of one side). Temperature measurements showed that this created a thermal gradient within each jar of 1.5 - 2^{o}C. A temperature difference of this magnitude affects the average growth rate and development time of the beetles, corresponding to a range of the emergence time of beetles of approximately 2-3 days (Stillwell *et al*. 2007). We note that temperature and host type have a suite of main and interactive effects on life history traits of *C. maculatus* (e.g., Giga & Smith 1987; Stillwell *et al*. 2007).

To seed subsequent generations, we collected all beetles that emerged in a replicate once every day and placed them into collection jars (one per replicate) during day 20 – 27 after the start of each generation. Beetles in collection jars were provided with deionized water and 20% sucrose solution (to avoid starvation and dehydration) but contained no beans. As *C. maculatus* females do not oviposit unless seed resources are present, this prevented reproduction in collection jars. At day 28, we seeded the next generation by collecting a random sample of 5 ml of beetles (500 individuals) from each collection jar and introducing these into a clean experimental jar with 200 grams of a fresh bean mixture. Because 200 grams of the bean mixture corresponds to approximately 2500 beans and 250 *C. maculatus* females lay approximately 25000 eggs, juvenile resource competition was pronounced in our evolving populations. In each generation of the experiment, we randomly assigned the position of all replicates within the climate chambers, to prevent systematic positioning differences between replicate lines from affecting our results. We repeated this procedure for 10 subsequent generations.

Sampling and sequencing

At the end of the experimental evolution experiment, at generation 10, we froze beetles by placing the collection jars in a freezer (-20^{o}C) at day 28 in the generation cycle. From each jar, we then randomly sampled 10 individuals and placed these individually to 0.2 ml Eppendorf tubes filled with 96% ethanol. We stored beetles at -20^{o}C until DNA extraction. We extracted DNA from whole beetles using Invitrogen PureLink Pro 96 Genomic DNA Purification Kit. As an mtDNA haplotype marker, we amplified and sequenced a region of the

mitochondrial COI gene using the molecular protocol described above. We aligned the DNA sequences with the reference sequences that we obtained earlier and determined the mtDNA haplotype (B, Y or C) carried by each individual. In total, thus, we sequenced the target COI gene segment of 1800 beetles. Two out of our 180 replicates contained one or more individuals that yielded poor or ambiguous sequence data. We excluded these two replicates from further analyses.

Our experimental and sampling design reflects an overall constraint on the total number of individuals we could sequence (i.e., $N \approx 1800$). Pre-experimental power analyses showed that the balance struck in our experiment between the number of replicate populations ($N = 180$) and the number of individuals sequenced in each population ($N = 10$) yielded the highest statistical power.

Statistical analyses

Our analyses focused on changes in the frequency of mitochondrial haplotypes over 10 generations of evolution. The key issues were to determine whether the relative fitness of an mtDNA haplotype depended upon (i) if it was common or rare at the start of the experiment, (ii) which mtDNA haplotype it was competing with and (iii) in which nuclear genetic background it was expressed. To enable statistical modelling of these questions, we first had to organize the data from our experiment into three subsets that were analysed separately, each composed of only those replicates that contained a given focal mtDNA haplotype (i.e. B, Y or C). Each such subset, thus, contained data from 120 entirely independent replicates, 40 of which were unique to the subset in question, 40 of which were

shared which the second subset and 40 of which were shared with the third subset. The three main inferential models were thus not entirely independent, and we base our conclusions on the general pattern that emerged from these models. As a response variable, we used

$\Delta f = f_{10} - f_0$,

where f_{10} is the observed relative frequency of the focal haplotype at generation 10 in a given sample of 10 individuals and f_0 is the starting frequency of the focal haplotype at generation 0 in the replicate in question (i.e., 0.2 if rare and 0.8 if common); Δf thus estimates the change in relative frequency of a given mitochondrial haplotype in a given replicate population over 10 generations.

We tested two specific key predictions of the hypothesis of frequency-dependent selection on mtDNA haplotypes: (i) that the change in the frequency of the focal haplotype should be affected by its starting frequency and (ii) that the mean Δf was larger than zero when rare and smaller than zero when common. We first modelled changes in haplotype frequency, in separate models for each of the three subsets of data, using two alternative modelling approaches. We used REML estimation to fit linear mixed models (JMP 9.0.0), which included starting frequency, nuclear genetic background, competing mtDNA genotype and their interactions as fixed effects and line identities of the focal and competing lines as two random effects variables. The latter terms were included to account for the fact that our design included two replicates of each particular mitonuclear introgression. These models form our main inferential models (Table 1). We note here that the line specific effects were neither biologically (on average 4.7% of the total variance in Δf) nor statistically (95% CI's of variance components

enclosed zero in all cases) significant, which is unsurprising given the 15 generations of backcrossing and the fact that we used an additional backcross to randomly selected outbred stock population males in the generation immediately preceding our experiment to further avoid the possibility of line specific adaptation. Because of the lack of line specific effects, and to validate our linear mixed effect modelling, we also modelled changes in haplotype frequency in conventional linear fixed effects model analogues of Δf (model I ANOVA, using type III SS), including starting frequency, nuclear genetic background, competing mtDNA genotype and their interactions (but excluding line specific effects). These models yielded results that were quantitatively very similar, and qualitatively identical (in terms of our ability or inability reject null hypotheses), to the linear mixed models with one single exception: the fixed effects model with the B mtDNA haplotype as the focal haplotype showed a significant effect of nuclear genetic background ($F_{2,106} = 3.89$, $P = 0.023$) on Δf which was not significant in the linear mixed model (Table 1). Visual inspections of residual plots of the fixed effects models suggested that data was well behaved and fulfilled the assumption of homoscedasticity (e.g., residual distributions were symmetrical and bell-shaped), but the residuals differed slightly but significantly from normality in one of the three models (Shapiro Wilk's test; B: $W = 0.949$, $p < 0.05$; C: $W = 0.987$, $p = 0.31$; Y: $W = 0.991$, $p = 0.60$). To ensure that our results were not biased as a result of this, we also evaluated the fixed effects models with randomization tests using 10000 permutations of the original dataset. This mode of testing for effects in our fixed effects models yielded quantitatively very similar and qualitatively identical results to those of the conventional F - tests.

To confirm that the expected mean of Δf in our experiment is indeed zero under the null hypothesis that gene frequencies are affected by genetic drift alone, we performed Monte Carlo simulations that mimicked our experiment. In these, a starting population of 250 adult females carrying either of two mtDNA haplotypes at relative frequencies of 0.2 and 0.8 were used to generate 250 female offspring drawn at random, by asexual reproduction, which then formed the adult females of the next generation. Each simulation involved 10 subsequent generations (at a fixed population size of N = 250 females) and we ran 10000 independent simulations and extracted Δf at generation 10. The mean Δf across these simulation was 0.00 (s.e.m. = 0.001). In no case was one of the haplotypes driven to fixation over 10 generations (i.e., the probability of fixation due to drift, $p < 0.0001$).

Given that the change in the relative frequency of a given focal haplotype is affected by its starting frequency (Table 1 and Table S1) and given that $\Delta f = 0$ under the null hypothesis, we asked whether the frequency of a focal haplotype increased when it was rare in the population (i.e., $\Delta f > 0$) and decreased when it was common (i.e., $\Delta f < 0$). To do this, we first partitioned data into six subsets, classified by focal haplotype and whether it was rare or common. We then performed one sample t-tests of H_0; $\Delta f = 0$. We also calculated the bootstrap 95% CI of the mean Δf using 10000 bootstrap replicates of each subset of the data. We note here that, because each replicate is represented only once, these t-tests are independent from one another within a given category of initial frequency of the focal haplotype (common or rare). Therefore, we performed two separate

combined probability tests (Whitlock 2005) to test the common hypotheses that the frequency of mitochondrial haplotypes increased when rare and decreased when common, respectively.

SUPPORTING RESULTS

In addition to the inferential statistical modelling based on mean Δf (see above), we tested whether the observed frequency distribution of f_{10} across replicates differed from that expected under either (i) a pure binomial probability distribution or (ii) a Monte Carlo simulation mimicking our experiment, by calculating exact goodness-of-fit probabilities (StatXact 9.0.0). The exact probabilities represent the probability that the observed frequency distribution (ranging between 0 and 1 with 0.1 increments) equals the expected and are thus non-directional. For expected frequencies under a binomial probability distribution, we simply used a sample of 10 from a standard binomial distribution with a success rate of 0.8 when focal haplotypes are common and 0.2 when focal haplotypes are rare. This simple method yields the expected frequency distribution under sampling variance, but it does not take several generations of sequential sampling into account. The Monte Carlo simulations were identical to those detailed above, with two exceptions: (i) the population size was not constant at $N = 250$ but was drawn randomly every generation from a binomial distribution with $p = 0.5$ and $N = 500$ trials to allow for random variation in sex ratio across generations; (ii) we drew a random sample of 10 individuals in generation 10 and calculated the haplotype frequency of this

sample (i.e., f_{10}) to generate the expected distribution of samples (rather than the expected value of Δf) across simulations. We ran 15000 independent simulations and the expected frequency distribution was based on probabilities derived from the frequency distribution of these 15000 observations. This yields the expected frequency distribution under both sampling variance, random variation in female population size and genetic drift during the experimental evolution. The mean f_{10} in samples across all simulations were 0.20 and 0.80, respectively (s.e.m. = 0.001), again confirming that the expected $\Delta f = 0$ (see above).

The observed frequency distribution of f_{10} across replicates differed significantly from that expected both under a binomial probability distribution and the Monte Carlo simulations. The exact probabilities were invariably very low (binomial: B rare: $p < 0.001$; B common: $p = 0.005$; C rare: $p = 0.005$; C common: $p = 0.002$; Y rare: $p < 0.001$; Y common: $p < 0.001$) (Monte Carlo: B rare: $p = 0.006$; B common: $p = 0.006$; C rare: $p = 0.005$; C common: $p = 0.006$; Y rare: $p = 0.008$; Y common: $p = 0.003$), due to an overrepresentation of higher-than-expected f_{10} observations when a given haplotype was rare and an overrepresentation of lower-than-expected f_{10} observations when a given haplotype was common (see Fig. S1). These analyses thus fully validated the conclusions based on our statistical modelling of Δf.

REFERENCES

Arnqvist, G. & Tuda, M. (2010). Sexual conflict and the gender load: correlated evolution between population fitness and sexual dimorphism in seed beetles. *Proc. R. Soc. Lond. B*, 277, 1345-1352.

Arnqvist, G., Dowling, D.K., Eady, P., Gay, L., Tregenza, T., Tuda, M. & Hosken, D.J. (2010). Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect. *Evolution,* 64, 3354-3363.

Ballard, J.W.O. & Melvin, R.G. (2007) Tetracycline treatment influences mitochondrial metabolism and mtDNA density two generations after treatment in Drosophila. *Insect Mol. Biol.*, 16, 799-802.

Dowling, D.K., Abiega, K.C. & Arnqvist, G. (2007). Temperature-specific outcomes of cytoplasmic-nuclear interactions on egg-to-adult development time in seed beetles. *Evolution*, 61, 194-201.

Dowling, D.K., Meerupati, T. & Arnqvist, G. (2010). Cytonuclear interactions and the economics of mating in seed beetles. *Am. Nat.*, 176, 131-140.

Kageyama, D., Narita, S., Imamura, T. & Miyanoshita, A. (2010). Detection and identification of *Wolbachia* endosymbionts from laboratory stocks of stored-product insect pests and their parasitoids. *J. Stored Prod. Res*., 46, 13-19.

Kondo, N., Shimada, M. & Fukatsu, T. (1999). High prevalence of *Wolbachia* in the azuki bean beetle *Callosobruchus chinensis* (Coleoptera, Bruchidae). *Zool. Sci.*, 16, 955-962.

Tuda, M., Wasano, N. Kondo, N., Horng, S.B., Chou, L.Y. & Tateishi, Y. (2004). Habitat-related mtDNA polymorphism in the stored-bean pest *Callosobruchus chinensis* (Coleoptera: Bruchidae). *Bull. Entomol. Res*., 94, 75-80.

Tuda, M., Rönn, J., Buranapanichpan, S., Wasano, N. & Arnqvist, G. (2006). Evolutionary diversification of the bean beetle genus *Callosobruchus* (Coleoptera: Bruchidae): traits associated with stored-product pest status. *Mol. Ecol.*, 15, 3541-3551.

Kondo, N., Nikoh, N., Ijichi, N., Shimada, M. & Fukatsu, T. (2002). Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc. Natl. Acad. Sci. USA*, 99, 14280-14285.

Giga, D.P. & Smith, R.H. (1987). Egg production and development of *Callosobruchus rhodesianus* (Pic) and *Callosobruchus maculatus* (F.) (Coleoptera: Bruchidae) on several commodities at two different temperatures. *J. stored Prod. Res.*, 23, 9-15.

Stillwell, R.C., Wallin, W.G., Hitchcock, L.J. & Fox, C.W. (2007). Phenotypic plasticity in a complex world: interactive effects of food and temperature on fitness components of a seed beetle. *Oecologia*, 153, 309-321.

Whitlock, M.C. (2005). Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.*, 18, 1368-1373.

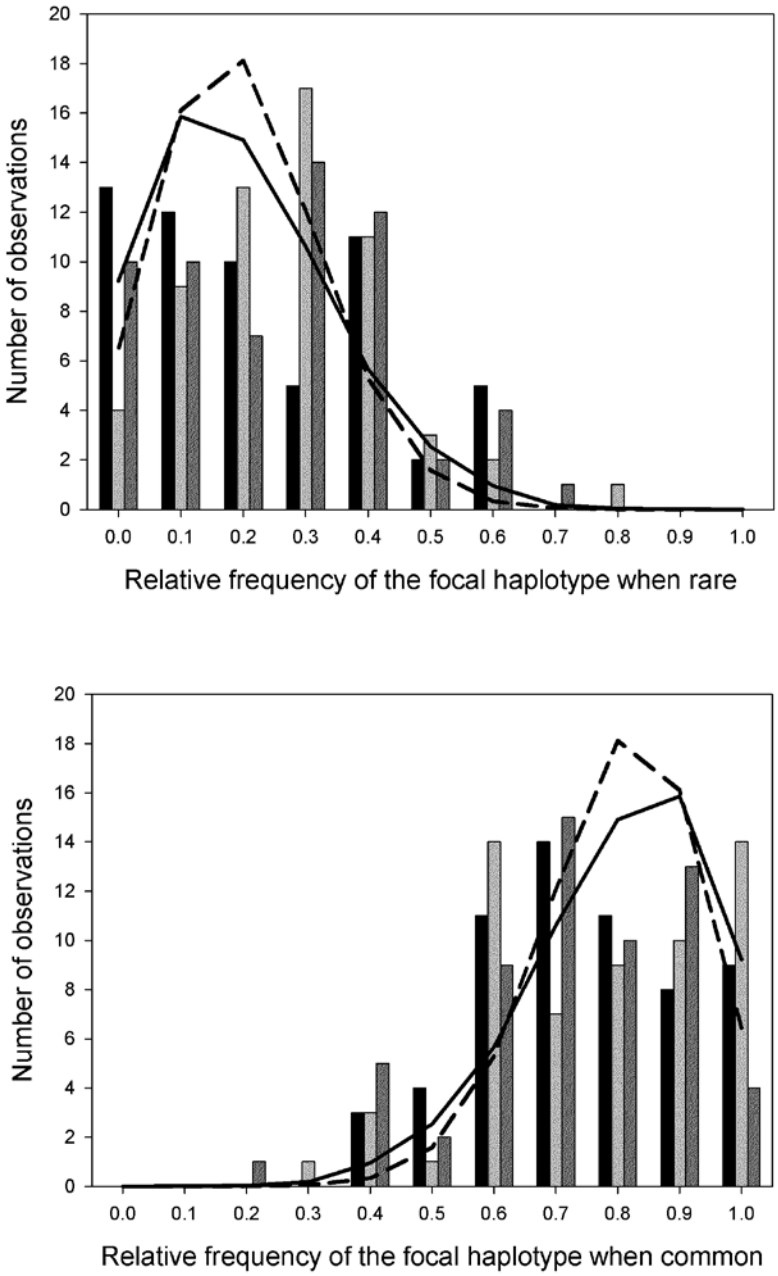

Figure S1 The expected (binomial probabilities: dashed line; Monte Carlo simulation: solid line) and observed frequency distributions of f_{10} across replicate populations where focal haplotypes (black bars = B; light grey bars = C; dark grey bars = Y) were either rare (top panel) or common (bottom panel) at the start of the experiment.

Table S1. Average Δf for all cells in the design, where "Common" and "Rare" refers to the starting frequency of the focal haplotype.

Focal mtDNA haplotype:	C		B		Y	
Competitor mtDNA haplotype:	B	Y	C	Y	C	B
Common C nDNA background	0.03	-0.10	-0.06	-0.03	-0.09	-0.02
Rare	0.06	0.09	-0.03	0.02	0.10	0.03
Common B nDNA background	-0.02	-0.02	0.01	-0.07	-0.08	-0.22
Rare	-0.01	0.08	0.02	0.22	0.02	0.07
Common Y nDNA background	-0.01	-0.05	-0.11	-0.08	-0.12	0.07
Rare	0.11	0.12	0.01	-0.07	0.05	0.08