**Methods**

**Molecular signatures of balancing selection along the *fru* gene**

We investigated signatures of balancing selection along the *fru* gene in two wild population samples of *D. melanogaster* flies: a North American population sample of 205 genomes (RAL) and a Zambian population sample of 197 genomes (ZI) (Mackay et al. 2012; Lack et al. 2015). Elevated polymorphism and linkage disequilibrium can both indicate that a given region is under balancing selection (Fijarczyk & Babik 2015). We therefore estimated regional polymorphism (nucleotide diversity, Tajima’s D) and regional linkage disequilibrium (Kelly’s ZnS) over 1000bp windows (500bp step size) along the *D. melanogaster* (release 6) genome, in each population, using the PopGenome package (Pfeifer et al. 2014).

**Sanger sequencing of a candidate *fru* polymorphism**

We investigated in more detail the genetic structure of a ~400bp region of the *fru* gene, situated ~24Kb from its 5’ end. From population genetic analyses, this region was observed to exhibit unusually high levels of polymorphism and LD. A sample of 96 chromosomes was taken from LHM, a laboratory-adapted population of *D. melanogaster* (Rice et al. 2005). We did so using a 'hemiclonal' approach (Abbott & Morrow 2011), which relies on purpose-built 'clone generator' to manipulate haploid chromosome sets (X, II, III). Individual hemiclonal males were crossed with females from a deficiency strain (*Df(3R)BSC509*), which carries a deletion spanning the *fru* gene and a TM6 balancer complement marked with *Stubble* (*Sb*). DNA from the hemiclone/deletion heterozygote offspring of this cross was extracted using standard protocols. The *fru* region was then PCR-amplified and Sanger-sequenced using the following primers: 5’-CACCCAACGCCACCTAGTTA-3’ (forward) and 5’-CGCCACTTGATTGCCACATT-3’ (reverse). Sanger sequencing revealed that the two allelic variants segregating for this region of *fru* are differentiated by a 43bp indel. As this indel produces a difference in fragment length between PCR product of the two alleles, we designated the two alleles ‘Long’ (L) and ‘Short’ (S). To infer the frequency of the *fru* indel polymorphism in the RAL and ZI populations in the absence of indel polymorphism data in both populations, we constructed a haplotype network using a set of 7 SNPs located in very close proximity to—and in perfect LD with—the indel.

**Fly culture and husbandry**

Unless otherwise stated, flies were maintained in 25°C constant temperature rooms at 50% humidity on a 12:12hr light-dark cycle. Flies were maintained on corn-agar-molasses medium with a light powderding of live yeast in either vials (8ml of media) or bottles (50ml). When required, flies were collected as virgins, every 0-6 hours post-eclosion until sufficient numbers were obtained. Virgin flies were anaesthetised using a CO2 pad for short periods of time and manipulated using a fly aspirator.

**Creation of isogenic allelic lines**

In order to assess the sex-specific fitness effects of the L and S alleles, we created fly lines homozygous for either one or the other allele but isogenic across the rest of their genome (‘isogenic allelic lines’). This process involved three phases (see Figure S1 for a schematic representation). In Phase 1, hemiclonal lines from the LHM population were extracted and genotyped for the allele of interest (L or S). In Phase 2, an isogenic background was repeatedly introgressed into the hemiclonal line while ensuring the transmission the original hemiclonal allele (L or S) through the use of a deficiency stock. In Phase 3, introgression lines (which were heterozygous for the allele of interest and the deficiency during the introgression process) were made homozygous for the allele of interest. Each step is described in more detail below.

Phase 1 involved the extraction of 96 hemiclonal lines from LHM (these lines were the same as those used in ‘Sanger sequencing of a candidate *fru* polymorphism’; the genotyping process is described in more detail there). From the 96 lines, three lines carrying the L allele and three lines carrying the S allele were kept for further analysis and maintained as hemiclones using standard hemiclonal amplification.

In Phase 2, the LHM genetic background of the six allelic lines was replaced with an isogenic Canton-S background. To achieve this, each of the six hemiclonal lines was repeatedly crossed with a strain that carries a deficiency spanning the *fru* polymorphism (*Df(3R)fru4-40*) in an isogenic Canton-S background. The third chromosomes of this strain, which carries the deficiency, is balanced with TM6B?, a standard balancer chromosome marked with the dominant mutation *Tubby* (*Tb*). Since balancer and deletion homozygotes are lethal in homozygous state and balancer chromosomes are marked, the offspring of a cross between a hemiclonal female and a *Df(3R)fru4-40*/TM6B? male are always identifiable as hemiclone/deletion heterozygotes. By repeatedly backcrossing hemiclone/deletion heterozygotes females to *Df(3R)fru4-40*/TM6 males, the original hemiclonal genome is gradually eroded through recombination in females and replaced with the isogenic Canton-S background of the deficiency line. After 7 generations of introgression, the allelic lines should carry on average less than 1% of the original hemiclonal haplotype.

In Phase 3, lines were created which were homozygous for the *fru* polymorphism. To achieve this, a two-step crossing procedure was performed. In the initial cross, the virgin balancer offspring of an introgression line x *Df(3R)fru4-40*/TM6 cross were set up in pairs (dyads A, B, C, see Fig. S1). Depending on the genotype of the introgression line parent, this cross can produce non-*Tb* F1 offspring of one of two genotypes, either introgression line homozygotes (the genotype interest to be kept) or introgression line/deficiency heterozygotes (to be discarded). Since neither of these two offspring genotypes bears an identifiable phenotypic marker, an additional ‘test cross’ was performed where both F1 offspring were backcrossed to *Df(3R)fru4-40*/TM6 males. Based on the F2 phenotype, the genotype of the F1 could be inferred, as homozygous introgression line F1 parents produce a 1:1 ratio of wild-type to *Tb* offspring, whereas introgression line/deficiency heterozygotes produce 1:2 ratio of wild-type to *Tb* offspring. Introgression line homozygote F1s which produced a ratio of wild-type to *Tb* that was significantly different from 1:2 (assessed from a *χ*2 test; Sup. Tab. 2) were kept for fitness assays.

**Genetic crosses and fly rearing**

Due to possible effects of inbreeding depression in the isogenic allelic lines detected in pilot experiments (data not shown), we performed fitness assays on flies generated by crossing individuals from the lines to flies from the *Df(3R)fru4-40/TM6B* stock. The resulting individuals carried the fru allele of a line either complemented by the Df(3R)fru4-40 deficiency (D) or by the TM6B balancer chromosome (B). This ensures that line-specific recessive deleterious alleles do not affect fitness measures and allows us to assess the effect of the fru alleles in a design similar to a quantitative complementation test [REF]. EXTRA EXPLANATION?

For these crosses, replicate vials were set up containing 10 virgin females from an allelic line (S1-3 or L1-3) and 10 *Df(3R)fru4-40/TM6B* males and left overnight for the flies to mate. To limit larval densities, we then twice transferred flies to fresh vials for 4-hour egg lays (~10am-2pm and ~2-6pm). Once pupae began to appear, these were removed and sorted into separate vials based on their *Tb* phenotype to establish groups of flies that carried the fru allele with either the D or the B complement (twelve groups in total, lines S1-3 and L1-3 in D or B background, hereafter referred to as S1/D, S1/B and so forth).

**Fitness assays**

Once flies began to eclose from the sorted pupae they were transferred to a new vial of standard media with added yeast where they have the opportunity to mate. After 24 hours the flies are removed and the sexes separated. Experiments were performed to assay egg laying rate as a proxy of female fitness and competitive fertilisation success as a proxy for male fitness.

Female fitness

Females were placed as triplets into vials containing 1% agar and fed by a capillary tube through the stopper containing a 4:1 yeast to sugar solution (6.5g yeast extract and 1.625g sugar per 100ml) at 250C and 80% humidity (to minimise evaporation). The triplets were maintained like this until the flies were 4-5 days old, with new food supplied every day. On the last day of this maturation period, triplets were transferred to new agar vials (0.8% agar) at approximately 4pm and allowed to lay eggs for 18 hours, after which time they were removed.

Assay vials were then photographed using webcamSeriesCapture (github.com/groakat/webcamSeriesCapture) software and a Logitech HD Pro webcam C920. In these pictures, eggs show up black against the clear white background of the illuminated agar. We used the machine learning program QuantiFly (github.com/dwaithe/quantifly) (Whaite et al. 2015) to count the number of eggs in each picture. A ‘predicted count’ of the number of eggs in each vial is produced and used as the response variable in this assay. Vials where a female died during the 18-hour period were not analysed. Similarly, vials where bubbles, debris or other contaminants caused counting problems were removed from the final dataset and not used in the analysis. Overall, 35 of our 898 vials were discarded (30 for deaths and 5 for picture contamination).

The aim was to have up to 30 measures of fecundity per line/genotype combination per block. However, this varied widely depending on the number of flies available and the num,ber fo deaths which occurred before and during the assay. Four replicate blocks for this experiment were performed.

Male fitness

Males were matured in standard food vials of 30 flies until 4-5 days old. To assay male fertilisation success, focal males were paired with a competitor male from the *Df(3R)fru4-40/TM6B* stock (reared in the same way as the focal males). The pairs of males were placed overnight in standard food vials. The next morning (~10-11am) a virgin *Df(3R)fru4-40/TM6B f*emale was added to the vial without CO2 anaesthesia and the two males were allowed to compete for this mate for 90mins. This time maximises the likelihood that as least one of the males will mate with the female, while keeping the rate of female re-mating to a negligible level. After the competition period, the two males were removed and the female left to lay eggs until she died. Once the larvae had developed and began to pupate, paternity was scored using the pupal phenotype. If all pupae displayed the *Tb* phenotype then paternity was assigned to the competitor (*Df(3R)fru4-40/TM6B*) male. If the pupae were a mixture of wildtype and *Tb* paternity was assigned to the focal male. Only vials with >10 pupa were included to ensure that any normal pupa would be observed. Therefore, 331 vials out of 1480 attempted trials were removed from the analysis.

This method allowed up to 12 replicates to be performed reliably per line/chromosome combination per day with several rounds of the assay repeated on consecutive days. In total there where three experimental blocks with 3-5 rounds per block.

Larval survival

Development time

Ageing

**Statistical analyses**

All statistical analyses were performed in *RStudio* (R Core Team 2015).

The frequency of the S allele in cage populations was modelled using a binomial GLMM, with time, manipulation (High-S or Low-S) and their interaction treated as fixed effects and cage treated as a random effect.

Egg count data for female fitness was analysed using a generalised linear mixed effects model fitted with the lmer function from the package lme4. The model included the flies' fruitless allele (L or S), the chromosomal complement (B or D) and their interaction as fixed effects, as well as line (L1-3, or S1-3) and block (1-4) as random factors. P-values were calculated using parametric bootstrapping in the package pbkrtest (Halekoh and Højsgaard 2012). The significance of each term was tested by comparing nested models including and excluding that term and based on 500 simulations.

For male competitive ability, paternity (focal vs. competitor) was analysed as a binary variable by fitting a generalised linear mixed model using the function glmer from the lme4 package. The model contained allele (L or S), chromosomal compliment (B or D) and their interaction as fixed effects, along with line (L1-3, or S1-3), block (1-4), and round (repeats within a blocks 1-5) as random factors. P-values were again calculated using parametric bootstrap as before.

**Results**

**Signatures of balancing selection at the *fru* polymorphism**

We found that the 1000bp-window encompassing the *fru* polymorphism (red dashed line) exhibits unusually high levels of polymorphism and local LD relative to the genome-wide average (Fig. 1A,B). This was true both in the RAL population (upper 2nd percentile of nucleotide diversity; upper 12th percentile of Tajima’s D; upper 5th percentile of Kelly’s ZnS), and in the ZI population (upper 5th percentile of nucleotide diversity; upper 11th percentile of Tajima’s D; upper 9th percentile of Kelly’s ZnS). A haplotype network constructed using SNPs situated in close proximity (<80bp; see Fig. 1A) to the *fru* polymorphism using polymorphism data from RAL and ZI reflects these patterns. Thus, it can be seen that haplotypes clearly cluster into allelic classes rather than by population (Fig. 1C). In other words, alleles linked to the *fru* polymorphism form the two major haplotypes present across both populations, such that each alternative allele is found at intermediate frequencies in both populations. Given the large evolutionary distances between the RAL and ZI populations used in the construction of the haplotype network, this is suggestive evidence that the *fru* polymorphism is under balancing selection. We therefore performed further experiments to test this hypothesis.

**Effect of *fru* alleles on sex-specific fitness**

Female fitness

In total we performed 863 successful fecundity trials over 4 experimental blocks. The number of successful fecundity trials varied from 30-96 vials for each of the twelve line/complement combinations.

Our models found no effect of the *fruitless* allele on the number of eggs laid (S: 27.92±0.78 eggs, L: 23.57±0.55 eggs; =3.63, p=0.11). However, there was an effect on fecundity due to the chromosomal complement carried by the flies, with D females laying more eggs than B females (D: 26.4±0.7 eggs, B: 24.6±0.6 eggs; =4.17, p=0.044). Furthermore, there was also a significant allele-by-complement interaction, whereby S/D flies laid more eggs than all other genotypes (S/D: 29.7 ±1.1 eggs, S/B: 26.±1.1 eggs, L/D: 23.6±0.8 eggs, L/B: 23.5±0.8; =6.33, p=0.022; Figure 1). Overall, we therefore see support for a positive effect on female reproductive output incurred by the S allele, which is exposed when assaying the allele in hemizygous state (D complement).

Male fitness

We obtained valid data on mating success for 1149 males, ranging from 60 to 107 males per line/complement combination.

As for female fecundity, there was no effect of the *fruitless* allele on male mating success (S: 0.436±0.02, L: 0.388±0.02; =0.38, p=0.6). But once again there were differences in success rate between the two chromosomal complements. However, this effect in males was in the opposite direction to that observed for females with the success rate of B males 12% higher than that of D males (B: 0.47±0.02, D: 0.35±0.02; =17.7, p<0.002). Finally, there was a tendency for an allele-by-complement interaction, with L/D males particularly poor competitors compared to other allele/complement males (L/D: 0.3±0.03; L/B: 0.47±0.03, S/D: 0.4±0.03, S/B: 0.47±0.03; =3.6, p=0.064; Figure 2). As with females, there appears to be an advantage to male fitness for flies carrying the S allele at the *fruitless* locus.

Female/male fitness correlations

The flies used in blocks 1-3 of both assays are full siblings. Therefore, we can compare our measures of fitness for both sexes within each block and see how these compare. Overall there is a negative correlation between the number of eggs laid and the success of focal males across all treatments (-0.3221952) (Figure 3). This relationship varied between B and D flies with a more negative relationship between male mating success and egg production for B flies (-0.4387948) compared to D flies (-0.1590864) (Figure 4).

Larval survival

Development time

Ageing