Electronic supplementary material for:

A non-coding indel polymorphism in the *fruitless* gene of *Drosophila melanogaster* exhibits antagonistically pleiotropic fitness effects

Michael D. Jardine[1,2], Filip Ruzicka[3], Charlotte Diffley[1], Kevin Fowler[1,2], Max Reuter[1,2]

[1]Department of Genetics, Evolution and Environment, University College London, London, United Kingdom

[2]Centre for Life's Origins and Evolution, University College London, London, United Kingdom

[3]School of Biological Sciences and Centre for Geometric Biology, Monash University, Clayton, Australia

DOI: 10.1098/rspb.2020.2958

Methods S1 – Identification of a polymorphic indel in the *fruitless* gene

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) [1,2]. Elevated polymorphism and linkage disequilibrium (LD) can both indicate that a given region is under balancing selection [3]. We therefore estimated regional polymorphism (nucleotide diversity, Tajima's D) and regional LD (Kelly's ZnS) over 1000bp windows (500bp step) along the *D. melanogaster* (release 6) genome, in each population, using PopGenome [4].

Sanger sequencing of a candidate *fru* region

A ~1000bp region of the *fru* gene was identified as exhibiting elevated levels of polymorphism and LD in both North American and Zambian population samples (Figure 1, main text; Results S1). To investigate this region in more detail, 96 chromosomes were sampled from LH_M, a laboratory-adapted North American population of *D. melanogaster* [5].

Sampling was performed using a 'hemiclonal' approach, in which purpose-built 'clone generator' flies are used to manipulate haploid chromosome sets (X, II, III) [6]. Individual hemiclonal males were crossed with females from a deficiency strain (*Df(3R)BSC509*), which carries a deletion spanning the *fru* gene and a TM6C balancer complement marked with *Stubble* (*Sb*). DNA from the hemiclone/*Df(3R)BSC509* heterozygote offspring of this cross was extracted using standard protocols (see "Phase 1" in Figure S1). A ~400bp region of the *fru* gene was then PCR-amplified and Sanger-sequenced using the following primers: 5'-CACCCAACGCCACCTAGTTA-3' (forward) and 5'-CGCCACTTGATTGCCACATT-3' (reverse).

Balancer stock genotyping

To ascertain the *fru* allele carried by the TM6B balancer, DNA was extracted from several *Df(3R)fru*$^{4\text{-}40}$*/TM6B* flies and the indel region was then PCR-amplified as above. The size of the PCR product was checked on an agarose gel (using control reaction with L- and S-bearing DNA templates as controls) and Sanger-sequenced to confirm allelic identity.

Results S1 - Identification of a polymorphic indel in the *fruitless* gene

We found that a 1000bp-window of the *fru* gene exhibited unusually high levels of polymorphism and local LD relative to the genome-wide average (red dashed line in Figure 1, main text). 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). Sanger sequencing further revealed that this polymorphic region of *fru* segregates for a 43bp indel, producing fragment length differences between the PCR products of the two alternative haplotypes in this region. We therefore designated these haplotypes 'Long' (L) and 'Short' (S), respectively.

To infer the frequency of the *fru* indel polymorphism in the RAL and ZI populations in the absence of direct indel polymorphism data, we examined the frequency of SNPs located in very close proximity to (<80bp) and in tight LD (in LH_M) with the indel (Figure 1, main text). A haplotype network constructed from these SNPs showed that haplotypes do not cluster by population but fall into divergent allelic classes that occur at intermediate frequencies in

both populations (Figure 1, main text). 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* indel (and/or alleles linked to it) are under some form of antagonistic and/or balancing selection. We therefore performed further experiments to test this hypothesis.

Methods S2 – Creation of isogenic lines

To assess the sex-specific fitness effects of the L and S alleles, we created fly lines homozygous for each allele but otherwise isogenic for a Canton-S background across the rest of their genome ('isogenic allelic lines'; see Figure S1 for the full crossing scheme).

First, we randomly selected three lines carrying the S allele and three lines carrying the L allele among the 96 sequenced hemiclonal lines (see Methods S1, "Sanger sequencing of a candidate *fru* region") and introgressed these alleles into an isogenic background, as described below. Introgression of the *fru* allele was performed with the help of a $Df(3R)fru^{4-40}$/TM6B deficiency stock, carrying a deletion spanning the *fru* locus (see Figure S1) in a Canton-S background, complemented with the third-chromosome balancer TM6B marked with the dominant mutation *Tubby* (*Tb*). Introgression of the *fru* allele onto the deficiency chromosome and into the Canton-S background was achieved by repeatedly backcrossing: (i) females heterozygous for a third chromosome carrying a focal *fru* allele ($fru^{S/L}$) and the $Df(3R)fru^{4-40}$ deficiency (themselves obtained by mating the hemiclonal line and females from the $Df(3R)fru^{4-40}$ deficiency stock), with (ii) males from $Df(3R)fru^{4-40}$ deficiency stock (see Figure S1). Since balancer and deficiency chromosomes are lethal in homozygous state and balancers carry the dominant *Tb* marker, the wild-type offspring of a hemiclone/$Df(3R)fru^{4-40}$ x $Df(3R)fru^{4-40}$/TM6B cross are always identifiable as $fru^{S/L}/Df(3R)fru^{4-40}$ heterozygotes. By repeatedly backcrossing $fru^{S/L}/Df(3R)fru^{4-40}$ heterozygote females to $Df(3R)fru^{4-40}$/TM6B males, the original hemiclonal genome carrying the focal *fru* allele is gradually eroded through recombination in females and replaced with the isogenic Canton-S background of the $Df(3R)fru^{4-40}$ deficiency line. After 7 generations of backcrossing, the allelic lines should carry on average less than 1% of the original hemiclonal haplotype (i.e. 1% of the original X-II-III complement).

Having introgressed the *fru* allele into the Canton-S background of *Df(3R)fru$^{4-40}$*, we created lines homozygous for the *fru* allele (as opposed to *fru$^{S/L}$*/*Df(3R)fru$^{4-40}$* heterozygotes). Because *fru$^{S/L}$*/*Df(3R)fru$^{4-40}$* heterozygotes and *fru$^{S/L}$*/ *fru$^{S/L}$* homozygotes are phenotypically indistinguishable, this was achieved through a two-step crossing procedure. An initial cross served to identify pairs of parents in which both individuals carried a focal *fru* allele. Virgin *Tb*-carrying offspring of a *fru$^{S/L}$*/*Df(3R)fru$^{4-40}$* x *Df(3R)fru$^{4-40}$*/TM6B cross (either *fru$^{S/L}$*/TM6B or *Df(3R)fru$^{4-40}$*/TM6B) were set up in pairs (dyads A, B, C, see "Phase 3" in Figure S1). Depending on the genotypes of the F1 pair, this cross can either produce: (i) 100% *Tb* F2s, if both F1 parents were *Df(3R)fru$^{4-40}$*/TM6B—these were discarded, or (ii) some fraction of non-*Tb* F2s, if the F1 pair were *fru$^{S/L}$*/TM6B+*Df(3R)fru$^{4-40}$*/TM6B or *fru$^{S/L}$*/TM6B+*fru$^{S/L}$*/TM6B. To distinguish the two latter cases and identify pairs of *fru$^{S/L}$*/TM6B individuals that are capable of producing the *fru$^{S/L}$*/ *fru$^{S/L}$* individuals we required, an additional 'test cross' was performed where F2s were backcrossed to *Df(3R)fru$^{4-40}$*/TM6 males. Based on the F3 phenotype, the genotype of the F2 could be inferred, as *fru$^{S/L}$*/ *fru$^{SL/}$* F2s produce a 1:1 ratio of wild-type to *Tb* F3s, whereas *fru$^{S/L}$*/*Df(3R)fru$^{4-40}$* heterozygotes produce 1:2 ratio of wild-type to *Tb* F3s. F2s producing a ratio of wild-type to *Tb* F3s that was significantly less than 1:2 (as assessed from a χ^2 test) were used to establish isogenic allelic lines.

References

1. Mackay TFC *et al.* 2012 The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173–178. (doi:10.1038/nature10811)
2. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015 The *Drosophila* Genome Nexus: A population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* **199**, 1229–1241. (doi:10.1534/genetics.115.174664)
3. Fijarczyk A, Babik W. 2015 Detecting balancing selection in genomes: Limits and prospects. *Mol. Ecol.* **24**, 3529–3545. (doi:10.1111/mec.13226)
4. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014 PopGenome: An efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936. (doi:10.1093/molbev/msu136)
5. Rice WR, Linder JE, Friberg U, Lew TA, Morrow EH, Stewart AD. 2005 Inter-locus antagonistic coevolution as an engine of speciation: assessment with hemiclonal analysis. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6527–6534. (doi:10.1073/pnas.0501889102)
6. Abbott JK, Morrow EH. 2011 Obtaining snapshots of genetic variation using hemiclonal analysis. *Trends Ecol. Evol.* **26**, 359–368. (doi:10.1016/j.tree.2011.03.011)

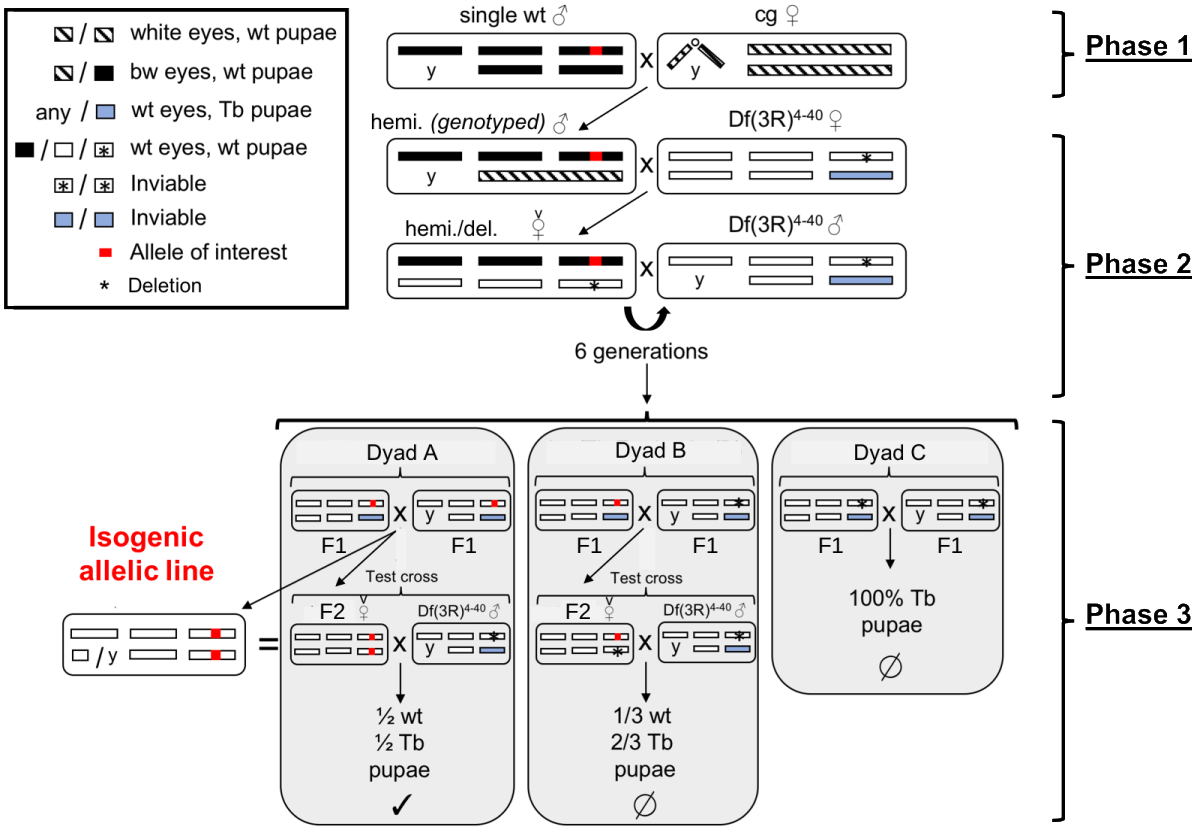

Figure S1. Crossing scheme used to create isogenic lines. See Methods S2 for details.

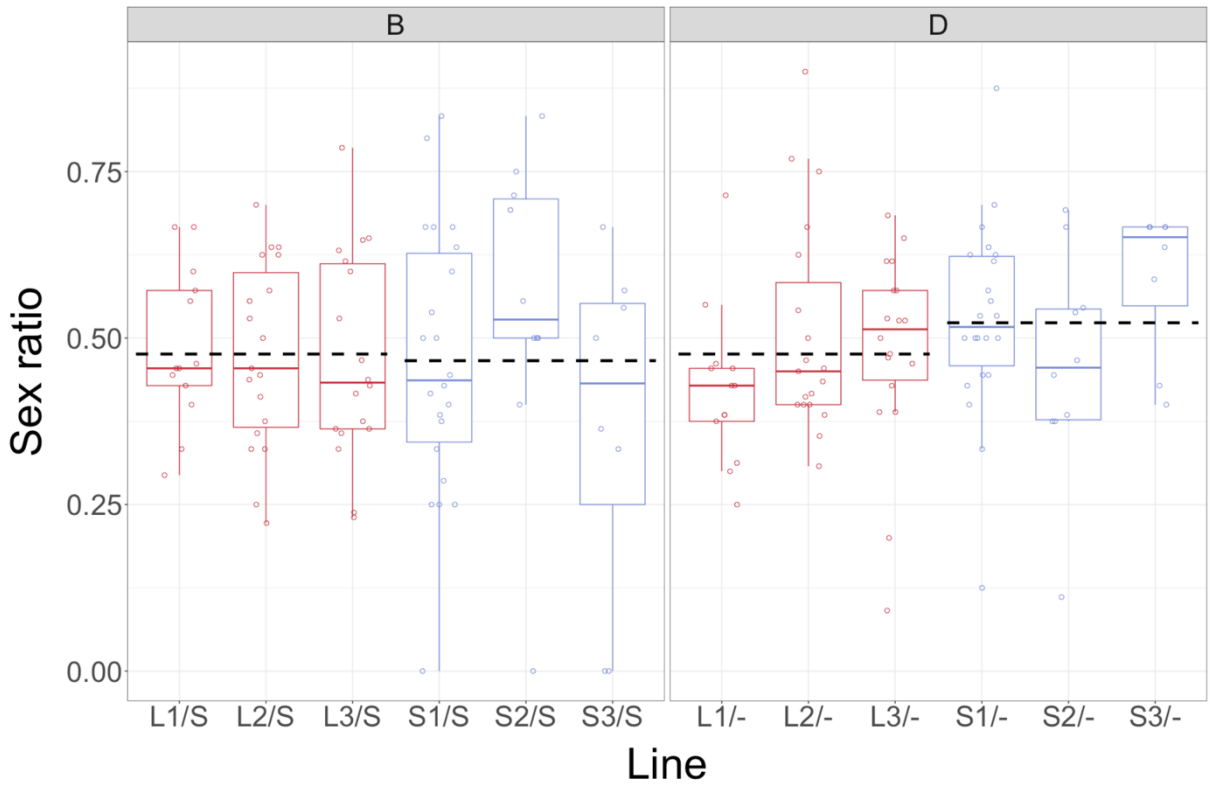

Figure S2. Sex ratio among surviving offspring presented for each line (L1-3 and S1-3) and chromosomal complement (B and D). Allelic means represented by dashed lines (L/B: 0.476±0.019; S/B: 0.466±0.035; L/D: 0.477±0.021; S/D: 0.0523±0.024). Sex ratio is defined as the proportion of males among offspring at eclosion.

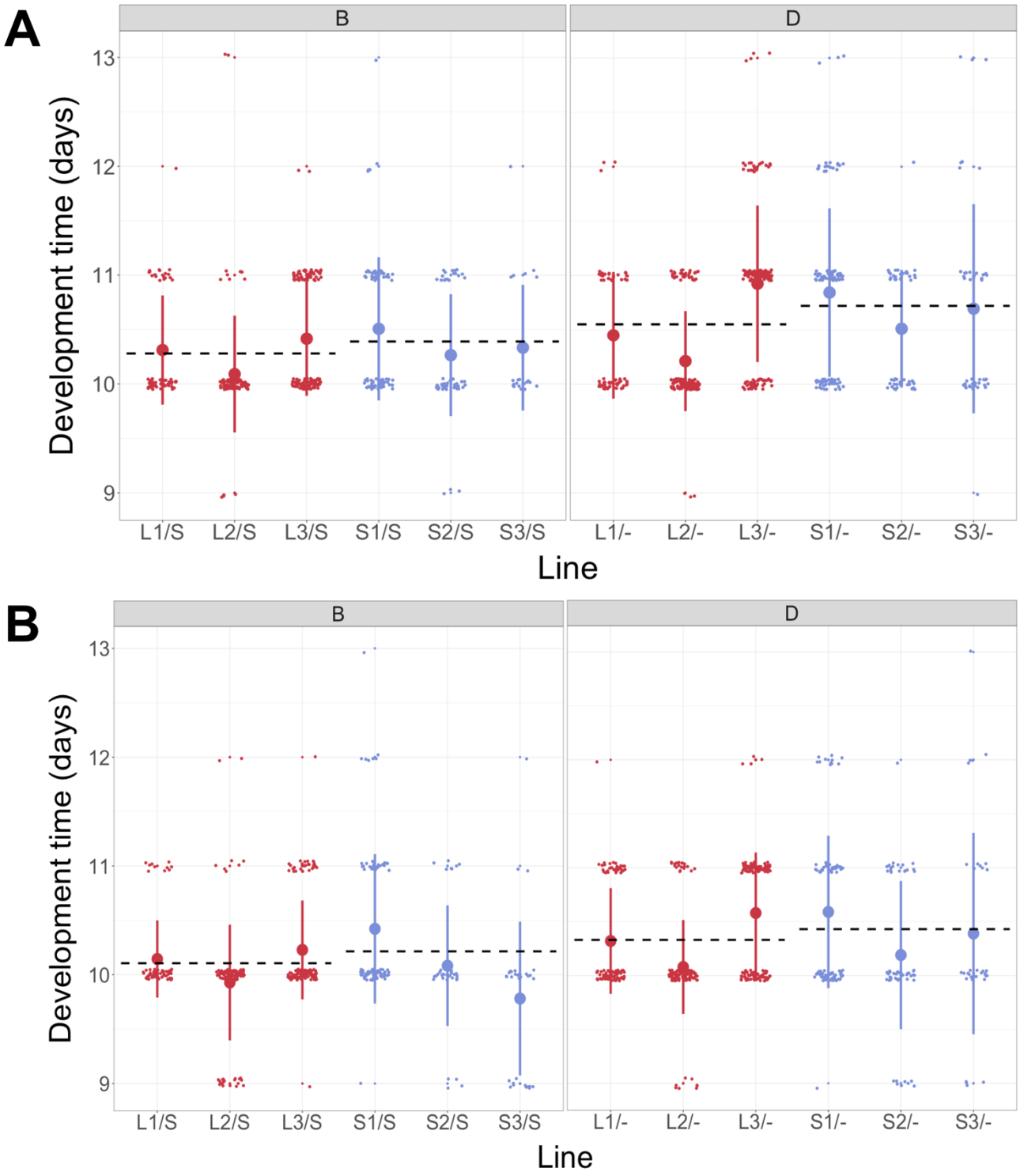

Figure S3. Development time (days ±standard error) of *fru* allelic lines (L1-3 and S1-3), for each chromosomal complement (B and D). Allelic means represented by dashed lines. Since sex was the most important factor in determining development time, this data is presented with the sexes separated: **A)** male flies (L/B: 10.28±0.03; S/B: 10.4±0.05; L/D: 10.55±0.036;

S/D: 10.72±0.054), and **B)** female flies (L/B: 10.1±0.027; S/B: 10.22±0.056; L/D: 10.33±0.028; S/D: 10.42±0.056).

~~Supplementary~~ **Table S1.** Results from Cox Proportional Hazard (CPH) models applied to lifespan data. Five models were used. One was for all flies and then the data was split to have separate models for each chromosome complement (B and D) and sex (female or male). The first column indicates the set of data the model is applied to, while the second column indicates the term being tested in that model. CPH models use one level of a term as the reference level with a value of one. Other levels are then compared to this. The comparison made is shown in brackets as: (compared level:reference). Each term in a model has a hazard-ratio (H-R), a 95% confidence interval and a H-R p-value, which indicates if the compared level differs from the reference level. Also presented are χ_1^2 and its p-value, indicating the contribution of each term to the overall risk of mortality.

Model	Term (comparison)	HR	95%-CI	HR p-value	χ_1^2	p-value
All flies	*fru* allele (S:L)	1.318	1.126-1.544	<0.001	0.139	0.71
	Complement (D:B)	0.519	0.44-0.612	<0.001	43.79	<0.001
	Sex (M:F)	0.531	0.449-0.627	<0.001	31.886	<0.001
	Allele x complement (S/D:L/F)	0.693	0.57-0.841	<0.001	10.411	0.0013
	Allele x sex (S/D:L/B)	0.821	0.676-0.997	0.046	4.856	0.0276
	Complement x sex (D/M:B/F)	2.624	2.154-3.198	<0.001	90.752	<0.0001
	Allele x complement x sex (S/D/M:L/B/F)	1.258	0.852-1.856	0.249	1.331	0.249
B only	*fru* allele (S:L)	1.386	1.16-1.655	<0.001	3.848	0.049
	Sex (M:F)	0.572	0.472-0.692	<0.001	105.65	<0.001
	Allele x sex (S/D:L/B)	0.731	0.561-0.953	0.02	5.368	0.021
D only	*fru* allele (S:L)	0.87	0.715-1.059	0.164	5.317	0.021
	Sex (M:F)	1.32	1.081-1.614	0.0066	10.705	0.001
	Allele x sex (S/D:L/B)	0.927	0.696-1.234	0.604	0.269	0.604
Females only	*fru* allele (S:L)	1.381	1.157-1.65	<0.001	2.334	0.127
	Complement (D:B)	0.542	0.449-0.655	<0.001	14.879	<0.001
	Allele x complement (S/D:L/F)	0.611	0.469-0.798	<0.001	13.127	<0.0001
Males only	*fru* allele (S:L)	1.039	0.854-1.263	0.705	1.276	0.259
	Complement (D:B)	1.301	1.061-1.595	0.011	3.119	0.077
	Allele x complement (S/D:L/F)	0.772	0.58-1.029	0.077	3.117	0.077

164