LETTER

doi:10.1038/nature12678

# Genetic incompatibilities are widespread within species

Russell B. Corbett-Detig[1], Jun Zhou[1], Andrew G. Clark[2,3], Daniel L. Hartl[1] & Julien F. Ayroles[1,2,4]

**The importance of epistasis—non-additive interactions between alleles—in shaping population fitness has long been a controversial topic, hampered in part by lack of empirical evidence[1–4]. Traditionally, epistasis is inferred on the basis of non-independence of genotypic values between loci for a given trait. However, epistasis for fitness should also have a genomic footprint[5–7]. To capture this signal, we have developed a simple approach that relies on detecting genotype ratio distortion as a sign of epistasis, and we apply this method to a large panel of *Drosophila melanogaster* recombinant inbred lines[8,9]. Here we confirm experimentally that instances of genotype ratio distortion represent loci with epistatic fitness effects; we conservatively estimate that any two haploid genomes in this study are expected to harbour 1.15 pairs of epistatically interacting alleles. This observation has important implications for speciation genetics, as it indicates that the raw material to drive reproductive isolation is segregating contemporaneously within species and does not necessarily require, as proposed by the Dobzhansky–Muller model, the emergence of incompatible mutations independently derived and fixed in allopatry. The relevance of our result extends beyond speciation, as it demonstrates that epistasis is widespread but that it may often go undetected owing to lack of statistical power or lack of genome-wide scope of the experiments.**

The role of epistasis in shaping genetic variation and contributing to observable differences within and between populations has been the focus of much debate[1–3]. In complex trait genetics, the additive paradigm used in genome-wide association studies[10] has recently been challenged by mounting evidence highlighting the importance of non-additive interactions between alleles[4]. Although the debate has been centred on the relative contribution of epistasis to the genetic variance, we still have a poor grasp of the extent to which epistasis affects the mean genotypic values of traits, an important step towards understanding the genetic basis of complex traits and the organization of molecular pathways[5]. Although epistasis is widely accepted to underlie the genetic basis of speciation, many details of this phenomenon remain poorly understood[2,3,5]. In particular, the evolutionary origins of the alleles that cause reproductive isolation are largely unidentified. Therefore, the importance of epistasis in shaping fitness within and between populations remains an important question in evolutionary biology.

Our understanding of the contribution of epistasis and the molecular details underlying non-additive genetic interactions is limited largely by the scarcity of available data. Although the idea that populations may harbour alleles with epistatic fitness effects has existed in the literature for some time, very few examples have been dissected at the genetic level (except for individual cases[6,11]). Furthermore, as yet, no systematic surveys have been conducted in diploid out-crossing species that have adequate statistic power to detect small fitness effects or to finely map interacting loci.

The traditional approach used to detect epistasis by statistical means relies on the observation of non-additivity of genotypic values between loci for a given phenotype. However, epistasis for fitness should have a genomic signature, regardless of our ability to measure a given phenotype[5–7]. In particular, it is expected that unfavourable allelic combinations will be under-represented, and this should precipitate a deviation from Mendelian proportions among unlinked incompatible alleles (detected by performing a screen for statistical association between alleles at loci that are not physically linked; Methods). Hereafter we refer to such deviations as genotype ratio distortion (GRD). In natural populations an exhaustive search for GRD is computationally intractable, statistically underpowered, or both[6]. By contrast, model organisms allow us to create experimental populations in which the amount of genetic variation and recombination can be controlled, thereby amplifying the signature of epistasis in a background of reduced dimensionality.

Here we apply tests of epistasis to the *Drosophila* Synthetic Population Resource (DSPR)[8,9] (Extended Data Fig. 1). To create the DSPR, two sets of eight highly inbred strains of diverse geographic origins were independently crossed in a round-robin design. Each set was duplicated and maintained for 50 generations in large freely-mating population cages. Subsequently, approximately 400 recombinant inbred lines (RILs) in each of four independent panels were created through 20 generations of sib-mating (generating four 'panels'; A-1, A-2 and B-1, B-2). After inbreeding, each RIL was genotyped at densely spaced markers, allowing a description of the genome of each RIL as a genetic mosaic of the eight founding lines originally crossed (Extended Data Fig. 1). The 50 generations of recombination and the large number of RILs within a panel provides replication over random allelic permutations. This replication is essential to attain statistical power for the detection of small effect epistasis.

We first excluded the possibility that residual population structure within the DSPR created association among alleles in the absence of epistasis by performing principal component analysis (Extended Data Fig. 2, Methods). Subsequently, we identified 22 pairs of epistatically interacting alleles in the DSPR (Fig. 1, Extended Data Table 1, Extended Data Fig. 3). Importantly, of the 44 incompatible alleles, 27 appear to be shared between two or more strains (Extended Data Table 1). This indicates that incompatible alleles are segregating at polymorphic frequencies in natural populations, and are not a result of inbreeding or long-term maintenance at small population size. On the basis of the frequencies in the founder strains, we estimate that any pairwise combination of founders has, on average, 1.15 pairs of epistatically interacting alleles. This is probably an underestimate, both because our statistical approach is conservative and because selectively disfavoured allelic combinations may be purged by selection during the free-recombination phase of the DSPR.

We next sought to confirm the predicted effect on reproductive fitness and to identify the underlying phenotype of two pairs of incompatible haplotypes (Fig. 2a, c). Using the original founder strains that contributed the putatively interacting alleles, we performed experimental crosses, and in both cases, we discovered that the negative interaction is caused by the minor alleles at each locus (genotype *aabb* in Fig. 2b, d; Extended Data Fig. 1). Specifically, in the case of one incompatibility

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853, USA. [3]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA. [4]Harvard Society of Fellows, Harvard University, Cambridge, Massachusetts 02138, USA.

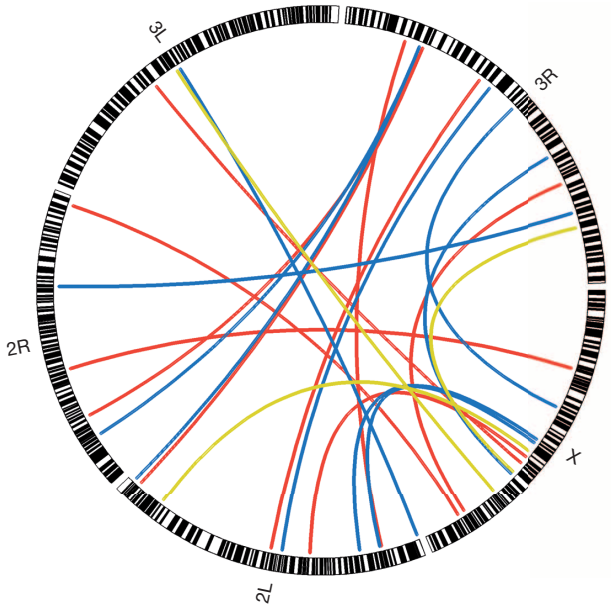



**Figure 1 | Locus pairs showing significant GRD across the DSPR lines of *Drosophila*.** The outer circle represent each chromosome arm. Each link represents a locus pair showing significant two-locus GRD. Yellow, blue and red links correspond respectively to RIL panel A-2, B-1 and B-2 (5% FDR corrected $P < 0.05$).

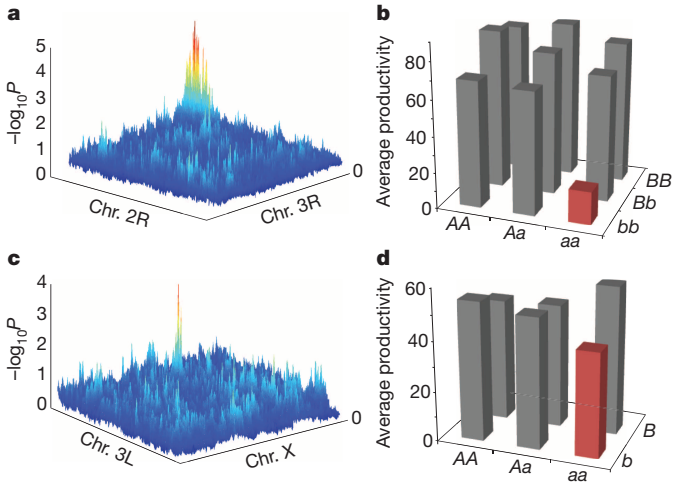



**Figure 2 | From missing genotypes to epistasis.** **a**, GRD signature between all genotyped loci on chromosomes (Chr.) 2R and 3R in RIL panel B-2. *P*, *P* value. **b**, Average productivity of each genotypic class recovered from 318 $F_2$ single-pair matings (progeny counts are $F_3$). As predicted from the GRD signal (in **a**), haplotypes tagged by single nucleotide polymorphism (SNPs) at positions 2R:4806926 and 3R:5870973 show strong negative epistasis for the *aa;bb* genotypes, $P = 5.51121 \times 10^{-9}$ LRT[29] (indicated by the red bar). **c**, GRD between loci on chromosomes 3L and X in RIL panel A-2. **d**, Average productivity of each genotypic class recovered from 401 $F_2$ single-pair matings. Haplotypes tagged by SNPs at positions 3L:11510853 and X:16483812 show strong negative epistasis for the minor alleles on each haplotype *aa;bb*, $P = 8.25 \times 10^{-5}$ LRT[29] (indicated by the red bar).

between chromosomes 2 and 3, males that are homozygous for both incompatible alleles produce on average 74% fewer offspring compared to all other allelic combinations ($P = 5.51121 \times 10^{-9}$ Likelihood Ratio Test, LRT, Fig. 2b, Extended Data Fig. 4). No significant effect was detected in females for any combination of genotypes. Using the same approach we validated a second instance of GRD, selected in the low range of effect size, between a haplotype on chromosomes X and 3 (Extended Data Fig. 5). We again observe a significant decrease (22%) in $F_2$ male fertility ($P = 8.25 \times 10^{-5}$ LRT, Fig. 2b, Extended Data Fig. 4), suggesting that GRD is a reliable signature of epistasis. The 'faster-males' theory[2,12] and subsequent experimental confirmations (reviewed in ref. 13) predict that male infertility will evolve more rapidly than other forms of post-zygotic reproductive isolation. Although we only have phenotypic data for our confirmed examples, the fact that both implicate male fertility as the underlying phenotype suggests that this effect may extend to within-species fitness epistasis.

The DSPR was intercrossed for sufficiently many generations (in excess of 50) that little linkage disequilibrium remains; hence this approach allows us to narrow down likely candidate genes associated with epistatic interaction for male fecundity. In total, there are three genes within the haplotype on chromosome arm 2R (~40 kb). The gene *notopleural* (*np*) is at the peak of this region; it is expressed in mature sperm[14] with alleles that are known to affect viability and sterility[15]. Notably, the human orthologue of *np* is associated with sperm-dysfunction in humans[16]. The interacting haplotype on chromosome arm 3R contains only two genes. In the centre of this region is *Cyp12e1*, a P450-cytochrome associated with electron transport in the mitochondria[17]. Interestingly, *Cyp12e1* harbours a non-synonymous mutation in a highly conserved protein domain. Mitochondrial dysfunction is commonly associated with male sterility in humans, plants and *D. melanogaster*[18], and therefore seems a plausible candidate phenotype.

To confirm that these observations were not specific to the *Drosophila* DSPR, we used the same method to screen for GRD in two additional RIL panels: the MAGIC panel in *Arabidopsis*[19] and the NAM panel in maize[20]. We found 7 instances of GRDs in *Arabidopsis* and 5 in maize (Extended Data Table 2). Although we have not validated these results, they suggest that GRD is present in other species as well.

Although the contribution of epistasis to variation in fitness is controversial in some fields[21], the Dobzhansky–Muller incompatibility (DMI) model[2,22,23] is a widely accepted guiding principle for biologists studying of the genetic basis of intrinsic, post-zygotic reproductive isolation. Largely motivated by this model, which predicts that alleles causing hybrid incompatibility are derived and fixed after population divergence, much empirical work in speciation genetics has been dedicated to mapping DMIs between species that diverged relatively long ago on an evolutionarily timescale[1,2] (Extended Data Fig. 5). However, it is unclear if these known examples of so-called 'speciation genes'[1,2,23,24] are an accurate representation of the earliest events in speciation, which have the greatest biological significance[2]. Even species that have diverged for only ~250,000 years have evolved complete male sterility an estimated 15 times over[25]. A reasonable interpretation of this evidence may concede that known 'speciation genes' are unlikely to be the same as those that initially contributed to reproductive isolation, but that these examples are instructive as to the properties of those genes[2]—an argument that closely mirrors our own.

Our central finding, that fitness epistasis is widespread within natural populations, indicates that the raw material to drive reproductive isolation is segregating contemporaneously within species and does not necessarily require, as proposed by the DMI model[22], the emergence of genetically incompatible mutations independently derived and fixed in allopatric lineages[23]. It is therefore necessary to explore the possibility that reproductive isolation could be achieved through divergence in frequencies of numerous pre-existing, polymorphic, small-effect incompatibilities[26–28] (Fig. 3). The implications of the present results go beyond understanding the role of intra-specific incompatibility in the context of speciation. Our work shows that epistasis for fitness-related traits has a detectable genomic footprint, and supports the idea that latent incompatibilities often exist between segregating variation within populations, only to be released when divergent lineages hybridize. This discovery highlights the importance of understanding the contribution of epistasis to observable phenotypic differences within and between populations.

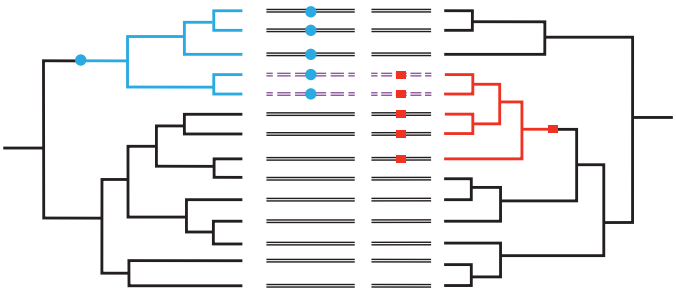


**Figure 3 | Model for unlinked loci with segregating pairs of incompatible alleles.** The dendrograms on the left and the right represent the genealogies of two haplotypes segregating within a species. The blue dot and the red rectangle indicate the origins of incompatible mutations on each respective genealogy. On the left, derived blue alleles are incompatible with derived red alleles on the right. These genealogies yield the individuals shown in the centre, wherein each line segment corresponds to a chromosome and each coloured square indicates the derived incompatible allele. Importantly, these incompatible allele pairs are polymorphic in this sample of individuals, thus individuals who inherit both incompatible alleles have lower fitness than those with either none or only a single incompatibility.

## METHODS SUMMARY

We genotyped the RILs of the DSPR by requiring that each putative variant be supported by a minimum of five reads. All sites wherein two or more alleles are supported by five reads were discarded. We confirmed that the RIL panels were free of cryptic population structure by performing principal component analysis (Extended Data Fig. 2). We next excluded sites wherein fewer than 150 individuals have a supported genotype, where the minor allele was present in fewer than 10 individuals, or where more than 15% of individuals with data had heterozygous genotypes. Following this, we assessed statistical significance for non-independence between pairwise combinations of alleles using a $\chi^2$ test, and applied a 5% false discovery rate (FDR) to correct for multiple testing. To reduce type 1 error, we restricted our search to inter-chromosomal comparisons and required that each putative instance of GRD be consistent with signal from adjacent variants (see Methods).

To confirm the predictions of the GRD scan, we first crossed the two DSPR founder strains that contributed the predicted interacting alleles. We then intercrossed the $F_1$ progeny to produce $F_2$ offspring. Virgin $F_2$ females were then individually and randomly mated to a single $F_2$ male. After mating for 4 days, the $F_2$ pairs were individually genotyped at known variable sites near the interacting alleles. We recorded the number of progeny of each pair to assay productivity. We used TaqMan kits to perform qPCR on the $F_2$ parents, and performed numerous statistical analyses[5] to quantify epistatic effects as a product of genotypes at the two sites (see Supplementary Methods).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 1 April; accepted 17 September 2013.**
**Published online 6 November 2013.**

1. Presgraves, D. C. The molecular evolutionary basis of species formation. *Nature Rev. Genet.* **11,** 175–180 (2010).
2. Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer Associates, 2004).
3. Cutter, A. D. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends Ecol. Evol.* **27,** 209–218 (2012).
4. Carlborg, O. & Haley, C. S. Epistasis: too often neglected in complex trait studies? *Nature Rev. Genet.* **5,** 618–625 (2004).
5. Phillips, P. C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9,** 855–867 (2008).
6. Bomblies, K. *et al.* Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol.* **5,** e236 (2007).
7. Payseur, B. A. & Hoekstra, H. E. Signatures of reproductive isolation in patterns of single nucleotide diversity across inbred strains of house mice. *Genetics* **171,** 1905–1916 (2005).
8. King, E. G., Macdonald, S. J. & Long, A. D. Properties and power of the *Drosophila* Synthetic Population Resource for the routine dissection of complex traits. *Genetics* **191,** 935–949 (2012).
9. King, E. G. *et al.* Genetic dissection of a model complex trait using the *Drosophila* Synthetic Population Resource. *Genome Res.* **22,** 1558–1566 (2012).
10. Zuk, O. *et al.* The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109,** 1193–1198 (2012).
11. Bikard, D. *et al.* Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323,** 623–626 (2009).
12. Palopoli, M. F. & Wu, C. I. Genetics of hybrid male sterility between *Drosophila* sibling species: a complex web of epistasis is revealed in interspecific studies. *Genetics* **138,** 329–341 (1994).
13. Wu, C. I., Johnson, N. A. & Palopoli, M. F. Haldane's rule and its legacy: why are there so many sterile males? *Trends Ecol. Evol.* **11,** 281–284 (1996).
14. Wasbrough, E. R. *et al.* The *Drosophila melanogaster* sperm proteome-II (DmSP-II). *J. Proteomics* **73,** 2171–2185 (2010).
15. Dockendorff, T. C., Robertson, S. E., Faulkner, D. L. & Jongens, T. A. Genetic characterization of the 44D–45B region of the *Drosophila melanogaster* genome based on an F2 lethal screen. *Mol. Gen. Genet.* **263,** 137–143 (2000).
16. Netzel-Arnett, S. *et al.* The glycosylphosphatidylinositol-anchored serine protease PRSS21 (testisin) imparts murine epididymal sperm cell maturation and fertilizing ability. *Biol. Reprod.* **81,** 921–932 (2009).
17. Kasai, S. & Tomita, T. Male specific expression of a cytochrome P450 (Cyp312a1) in *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **300,** 894–900 (2003).
18. Meiklejohn, C. D., Montooth, K. L. & Rand, D. M. Positive and negative selection on the mitochondrial genome. *Trends Genet.* **23,** 259–263 (2007).
19. Kover, P. X. *et al.* Multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **5,** e1000551 (2009).
20. McMullen, M. D. *et al.* Genetic properties of the maize nested association mapping population. *Science* **325,** 737–740 (2009).
21. Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4,** e1000008 (2008).
22. Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, 1937).
23. Orr, H. A. & Turelli, M. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* **55,** 1085–1094 (2001).
24. Presgraves, D. C. & Stephan, W. Pervasive adaptive evolution among interactors of the *Drosophila* hybrid inviability gene, *Nup96*. *Mol. Biol. Evol.* **24,** 306–314 (2007).
25. Tao, Y. *et al.* Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics* **164,** 1383–1397 (2003).
26. Fitzpatrick, B. M. Hybrid dysfunction: population genetic and quantitative genetic perspectives. *Am. Nat.* **171,** 491–498 (2008).
27. Demuth, J. P. & Wade, M. J. On the theoretical and empirical framework for studying genetic interactions within and among species. *Am. Nat.* **165,** 524–536 (2005).
28. Reed, L. K. & Markow, T. A. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. *Proc. Natl Acad. Sci. USA* **101,** 9009–9012 (2004).
29. Cheverud, J. M. & Routman, E. J. Epistasis and its contribution to genetic variance components. *Genetics* **139,** 1455–1461 (1995).

**Acknowledgements** We are grateful to T. Long, S. Macdonald and E. King for creating the DSPR and sharing the RILs and founder strains with us. We thank C. Jones, B. de Bivort, T. Sackton, S. D. Kocher, J. Grenier and N.E. Soltis for comments and discussions. We thank X. Shi for technical assistance. This work was supported by grants: NIH GM065169 and GM084236 to D.L.H., HD059060 to A.G.C., and Harvard Society of Fellows Fellowship and Harvard Milton Funds to J.F.A. R.B.C.-D. was supported by a Harvard Prize Fellowship.

**Author Contributions** J.F.A. conceived the idea of the project, R.B.C.-D. and J.F.A. conceived and designed experiments and analyses. R.B.C.-D. and J.F.A. conducted bioinformatics and statistical analyses; R.B.C.-D., J.F.A. and J.Z. performed experiments; J.Z. carried out molecular work; A.G.C. and D.L.H. gave analytical and conceptual advice throughout the project.

**Author Information** All code used and generated for this study is available upon request. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F.A. (ayroles@fas.harvard.edu).

## METHODS

**The DSPR resource.** The *Drosophila* Synthetic Population Resource (DSPR) is described in detail elsewhere[30,31]. This panel of recombinant inbred lines (RILs) was generated by first crossing eight highly-inbred strains in a round-robin design and subsequently maintaining in a freely-mating large population cage. Initially two sets of eight lines were independently crossed (panels A and B). Following the round-robin cross, each panel was subdivided into two replicate sets (we refer to these panels as A-1, A-2 for set A and B-1 and B-2 for set B). The geographic origin of each line is described elsewhere[30]. After 50 generations of recombination, approximately 400 lines were inbred by 20 generations of full sib mating. The genome of the resulting RILs is a mosaic of the original eight founder strains within each set. The DSPR resource is composed of a total of four sets of approximately 400 RILs.

We downloaded the genome sequences of the founding strains as well as the SNP pileups from the DSPR website (http://wfitch.bio.uci.edu/~dspr/index.html). Briefly, King *et al.*[30] sequenced each RIL at a subset of markers selected via a restriction enzyme associated digest (RAD; hence RAD-seq)[32]. This is followed by sequencing of sites adjacent to restriction enzyme cut sites (*Sgr*AI) using 100-bp single-end reads. In addition, King *et al.*[30] sequenced each founder genome to approximately 50× depth, allowing for accurate genotyping of the founding strains. Thus for all sites assayed via RAD-seq in the RIL panels, it is possible to identify which founders may have contributed that particular variant.

**Genotypic data.** We began by applying various quality filters to RAD-seq based genotypes of each RIL. Because the error properties and ascertainment bias of RAD-seq methodologies are generally poorly understood, we adopted a conservative genotyping approach to analyse the SNP data. For each site, we required that a minimum of five reads support a genotype call in order to consider that site in a given RIL. If at a given site an RIL has two genotypes supported by five or more reads, the site was considered heterozygous and excluded from any further analysis. More than 95% of genotype calls were homozygous, which is consistent with the results of King *et al.*[30], who reported very high rates of homozygosity, and confirming a largely successful inbreeding. We excluded all sites for which the minor allele was present in fewer than 10 lines and for which the number of heterozygous genotypes across RILs within a panel exceeded 15% of the total.

**Detecting GRD.** We used the resulting SNP genotype calls to search between pairs of variable sites for non-independent allelic segregation[33], which we call genotype ratio distortion (GRD).

GRD was detected by computing a $\chi^2$ test between each pair of alleles on different chromosomes. Significant GRD reflects a deviation from expected Mendelian genotypic ratio under independent segregation between loci. In order to remain conservative, and to ensure that significant allelic pairs were not physically linked, we restricted this search to inter-chromosomal pairs. We also excluded all pairwise comparisons with fewer than 150 total genotypes or for which any allele's frequency was less than 0.05. Statistical significance was assessed for all inter-chromosomal comparisons via a $\chi^2$ test and we applied a 5% false discovery rate (FDR) correction for multiple-testing[34]. Subsequently, we only report pairs of alleles for which at least three adjacent SNPs are in local linkage and also show significant GRD.

**Assessing frequency of incompatible alleles in founder lines.** For each instance of GRD, we attempted to estimate the frequency of the alleles in the founder population. Although many of the SNPs we identified in linkage disequilibrium with incompatible haplotypes are present in more than one of the founder strains, this does not necessarily indicate that both founders contained the interacting allele. In some cases one or more founder haplotypes may not be present in the RILs. To account for this problem, we confirmed that each parental haplotype was present in the RILs by confirming that SNPs near to instances of GRD and unique to each potentially interacting founder strain are present in the RILs and that SNPs unique to each founder also show strong associations with their predicted inter-chromosomal interactions when the other founders haplotypes are excluded.

We further confirmed our inferences of GRD by searching for associations between haplotypes as identified by the hidden Markov model implemented in ref. 30 for the analysis of the DSPR. Because haplotype probabilities are 'soft' (that is, the maximum genotype probability for any one haplotype is 0.995), we only considered individuals at sites that have a greater than 95% probability of a single haplotype (probably homozygous). We searched for GRD between all possible sets of parental haplotypes at each locus (that is, all 8 individuals, 8 choose 2 pairs, 8 choose 3 trios, and so on), and we excluded all sets that contained founder haplotypes that were not represented in the RIL panel. Here again we required that minor allele frequencies of any pairwise comparisons be a minimum of 0.05. In 66.67% of cases, estimates of founder allele frequency from the maximally significant SNP and maximally significant set of haplotypes matched perfectly. For the remaining 33.33% of cases, a likely explanation is that most significant SNPs were not in perfect linkage disequilibrium with the incompatible allele in the founders' genomes. In all cases we detected a significant interaction between a similar set of lines as we predicted based on the SNP data. Results from these analyses are summarized in Extended Data Table 2. Importantly, our results indicate that many incompatible alleles are present in more than one founder, suggesting that they were segregating in natural populations before being 'captured' within isofemale lines.

**Experimental validation.** In order to experimentally validate two specific instance of GRD, we sought to identify the causative phenotype underlying two of the interactions that we discovered in the SNP data (in panel B-2 between 2R:4806926 and 3R:5870973 and panel A-1 between 3L:11510853 and X:16272168). The first instance was chosen for the strength of the interaction and the interesting biology associated with the gene harboured by each haplotype. The second one was chosen at random but aimed to represent the average magnitude of disequilibrium we found across all the interacting pairs uncovered. S. MacDonald provided the founder strains used in the construction of the DSPR.

For both incompatibilities, we initially crossed the two strains that contributed the interacting haplotypes. We then inter-crossed the $F_1$ progeny to produce $F_2$s. Both parental and $F_1$ crosses were performed using five males and five virgin females per vial. We collected 318 (2R–3R interaction) and 401 (3L–X interaction) virgin $F_2$ females and maintained them in female-only vials for between four and seven days. Males were kept for the same time in male-only vials. If after this time we did not observe any larvae, each female was then mated individually to a single $F_2$ male. After four days, we removed the parents and extracted DNA from each $F_2$ individually. Experimenters were blind to the fly genotypes until the end of the experiment. At six and twelve days after removing the parents, we cleared each vial and recorded the number of offspring that had been produced. All crosses were performed on standard medium supplemented with yeast. We maintained all vials for crosses and for ageing virgin flies on a 12 h light:dark cycle at a constant 25 °C.

Following mating, we ground single flies in 50 µl of appropriate buffer (10 mM Tris, pH 8.0; 1 mM EDTA; 25 mM NaCl; 0.5% SDS). 2 µl protease K (20 mg ml$^{-1}$) were added and the samples were incubated for 30 min at 37 °C, and 5 min at 95 °C. Genotyping was performed using Taqman genotyping assays at sites that differentiated the founder strains near to the instances of GRD (catalogue no. 4371353, Applied Biosystems). This assay uses two probes that differ at the SNP site of interest, with each probe complementary to one allele. We used a 5 µl reaction volume containing 2.25 µl DNA, 2.5 µl TaqMan master mix, and 0.25 µl Custom TaqMan probes. We placed a 384-well plate in the Applied Biosystems 7900HT fast real-time PCR system for qPCR reactions. The program began with a step at 95 °C for 10 min, following 40 cycles of 15 s at 92 °C and 1 min at 60 °C. We assigned genotypes to individuals using TaqMan Genotyper Software V1.3. We required that each genotype have a minimum 0.95 posterior probability of being correctly called before productivity analyses.

Various definitions of epistasis have led to many approaches to statistical detection of epistatic effects (reviewed in ref. 35). Given that epistasis will be present when the combined effect of a particular pair of alleles is different from what would be expected under additivity (that is, whether alleles at loci A and B were considered together or independently, the phenotypic effects would be equivalent), many tests rely on detecting departures of the means of the genotypic classes.

We tested for the presence of statistical epistasis by implementing the method of Cheverud and Routman[36] who proposed an intuitive approach which consists of fitting a linear model containing additive, dominance and interaction effects—tested against the null in which there are only additive effects. The model for autosomal loci is of the following form:

$$Y_{ijkl} = a_{ij} + a_{kl} + d_{ij} + d_{kl} + a_{ij}a_{kl} + a_{ij}d_{kl} + a_{kl}d_{ij} + d_{ij}d_{kl} + e$$

Here, $Y_{ijkl}$ corresponds to the productivity value of flies with genotype $ij$ at loci 1 and $kl$ at loci 2; $a_{ij}$ and $a_{kl}$ for the additive effects of loci 1 and 2, respectively; $d_{ij}$ and $d_{kl}$ for the dominance effects of loci 1 and 2; $a_{ij}a_{kl}$ for the additive by additive epistatic effects between loci 1 and 2; $a_{ij}d_{kl}$ and $a_{kl}d_{ij}$ for the additive by dominant effects and finally $d_{ij}d_{kl}$ for the dominance by dominance effects ($e$, the residual error). We can then perform a likelihood ratio test to ask whether a model including interaction effect provides a better fit to the data. The model fitted to the incompatibility involving a hemizygous chromosome (between chromosomes X and 3L) was adjusted accordingly to reflect the absence of dominance and dominance interaction on the X.

For the incompatibility between chromosome 2R and 3R, we obtained significant evidence for epistasis ($P = 5.51121 \times ^{10-9}$; LRT) definitions above. Similarly, the incompatibility between chromosomes X and 3L was also significant ($P = 8.25 \times 10^{-5}$; LRT).

As we have shown, recombinant inbred lines are a powerful tool for detecting epistasis statistically. Nonetheless, the additional test-crosses we used to validate two instances of GRD were necessary to determine the role of dominance in each case of epistasis, something that cannot be assessed solely from inbred lines.

**Assembly methods.** To identify candidate polymorphisms in regions of GRD between 2R and 3R, we obtained short-read data for the founder strains from

the DSPR website (http://wfitch.bio.uci.edu/~dspr/). We aligned all data to the *D. melanogaster* reference genome v5.42[37] using stampy[38], which first maps reads using BWA[39], and subsequently attempts to map those reads which BWA fails to confidently map under more permissive conditions. All alignments were performed using the default parameters of each program. We first realigned indels in each line using the 'RealignerTargetCreator' and 'IndelRealigner' tools contained within the Genome Analysis Toolkit (GATK[40]). We called the consensus for each line using the GATK genotyping function, 'UnifiedGenotyper'. Genotyping was performed using the default parameters of the program except that we called each line as a haploid genome (—sample_ploidy 1), which is justified because all strains are highly inbred in the regions we focused on[30]. Finally, we predicted the effect of all discovered variants in regions of strong GRD using SnpEff[40].

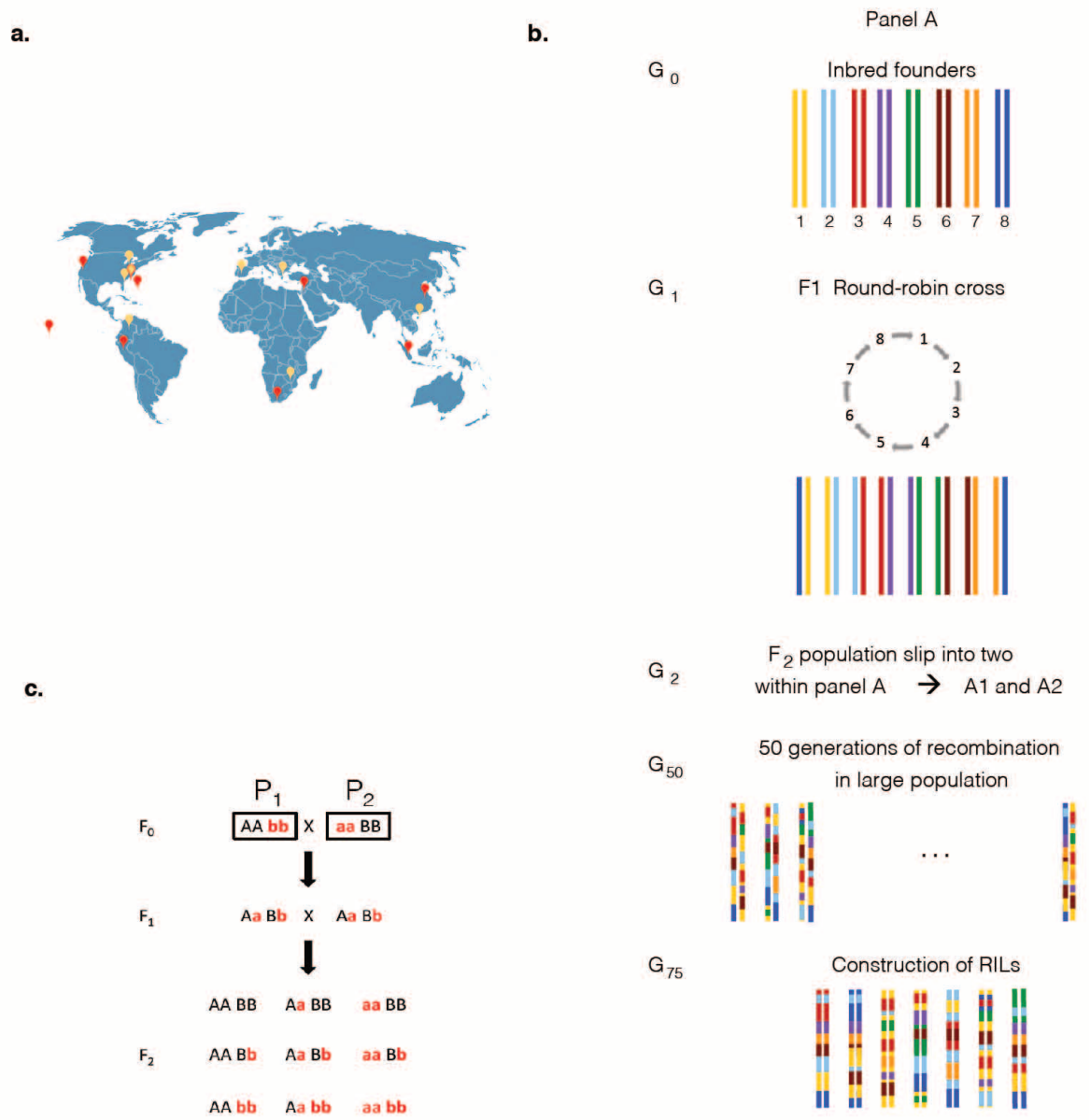
**Homology search.** We used blastp[35] to query the sequence of Cyp12e1 against the non-redundant protein database maintained by the NCBI. We used COBALT[41] to align these protein sequences and to identify highly conserved protein domains. To determine if the particular mutation that we observed is unusual at this site, we counted the number of occurrences of this residue among the top 250 most closely related proteins.

**Maize incompatibilities.** We searched for GRD in *Zea mays* using RIL genotype data generated for the Nested-Association Mapping panel[42]. This panel is composed of 25 sets of RILs made up of 200 RILs each. Each set was produced by crossing a single strain against one of 25 other highly-inbred parents, followed by several generations of self-fertilization. Known single nucleotide variants and short repeats were then assayed using by PCR to establish the haplotype structure of each RIL.

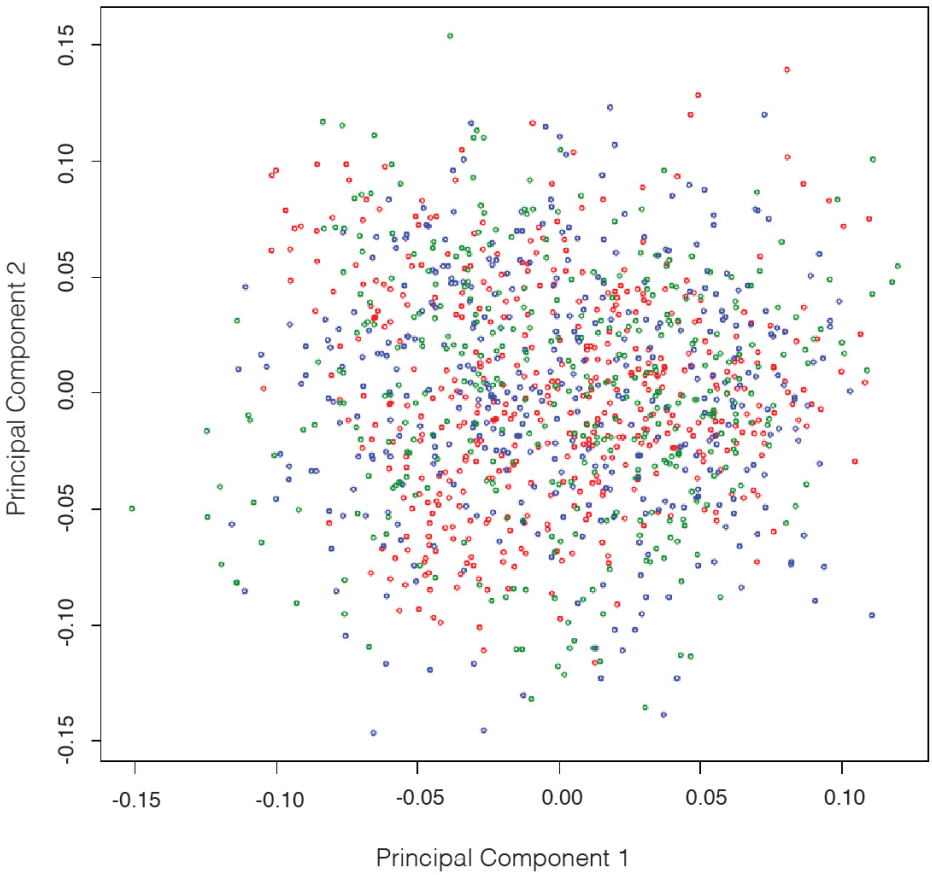
For these analyses, we used the 'imputed' genotype calls (described elsewhere[42]), as the 'raw' genotype data may be unreliable (E. Buckler, personal communication). We acquired the imputed data for the RILs from www.panzea.org. We then filtered data following a similar procedure to that we implemented for the *Drosophila* data (described above). We required that each site in any one individual be homozygous for consideration in downstream analyses. We further required each allele in any pairwise comparison have a minimum frequency of 0.05. Statistical significance was assessed using a $\chi^2$ test and corrected for multiple testing by applying a 10% FDR correction within each set of RILs[34]. Here we used a slightly higher FDR than for the *Drosophila* data because the NAM RIL panels are smaller and therefore have reduced power to detect interacting alleles.

***Arabidopsis* incompatibilities.** The multi-parental advanced generation intercross panel of *Arabidopsis thaliana* RILs[43] was produced in a similar way to the DSPR panel. Initially 18 founder strains were intercrossed followed by several generations of self-fertilization to produce approximately 500 mostly independent RILs. Each RIL was then genotyped at known variable sites using PCR[43]. Once again, in analysing these data, we required that each site be homozygous in order to be considered in downstream analyses and that each allele within any pairwise comparison be present at a frequency greater than 0.05. Afterwards, we identified significant GRD between haplotypes using $\chi^2$ tests and applied a 10% FDR correction[33].

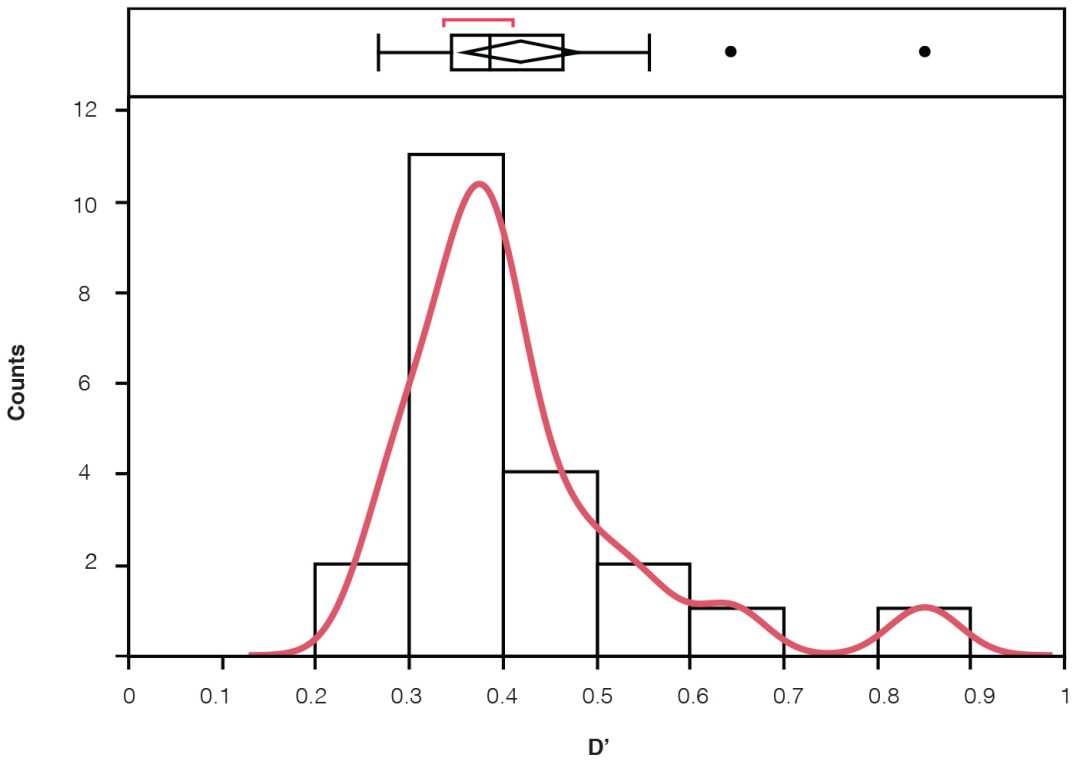

30. King, E. G. *et al.* Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. *Genome Res.* **22,** 1558–1566 (2012).
31. King, E. G., Macdonald, S. J. & Long, A. D. Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits. *Genetics* **191,** 935–949 (2012).
32. Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3,** e3376 (2008).
33. Ackermann, M. & Beyer, A. Systematic detection of epistatic interactions based on allele pair frequencies. *PLoS Genet.* **8,** e1002463 (2012).
34. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
35. Phillips, P. C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9,** 855–867 (2008).
36. Cheverud, J. M. & Routman, E. J. Epistasis and its contribution to genetic variance components. *Genetics* **139,** 1455–1461 (1995).
37. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287,** 2185–2195 (2000).
38. Lunter, G. & Goodson, M. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21,** 936–939 (2011).
39. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
40. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).
41. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain $w^{1118}$; iso-2; iso-3. *Fly (Austin)* **6,** 80–82 (2012).
42. McMullen, M. D. *et al.* Genetic properties of the maize nested association mapping population. *Science* **325,** 737–740 (2009).
43. Kover, P. X. *et al.* A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **5,** e1000551 (2009).



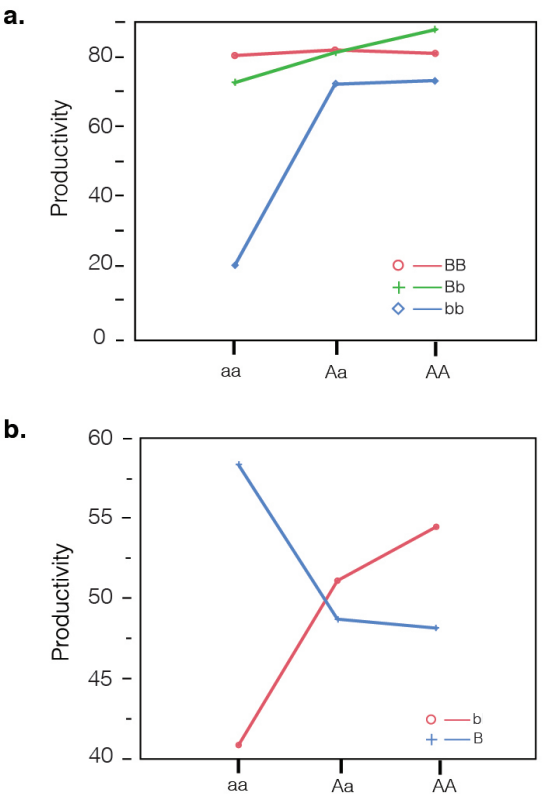

**Extended Data Figure 1 | Description of the DSPR and validation scheme.** **a**, Geographic distribution of the DSPR founding strains (orange, panel A; red, panel B). **b**, Construction of the recombinant inbred lines. For each panel all founder strains were crossed in a round-robin design (line 1 ♀ × line 2 ♂, line 2 ♀ × line 3 ♂,…, line 8 ♀ × line 1 ♂) to produce $F_1$s, and the $F_1$s were then allowed to mate free to produce an $F_2$ population. In each panel A and B, these $F_2$ populations were split into two independent population to create panels A1, A2 and B1, B2. Each was allowed to recombine freely for 50 generations, in very large population. After 50 generations, for each replicate panel, about 400 isofemale lines were inbred for 25 generations to create the 4 panels of RILs used in this study. **c**, Crossing scheme used to validate epistatic effects. A pair of founder segregating incompatible alleles was selected and crossed to produce $F_1$s; we then intercrossed the $F_1$ progeny to produce a large $F_2$ population, segregating all possible allelic combinations between alleles at loci 1 and 2. We then counted the progeny each pair produced by intercrossing a large number of $F_2$s which were later genotyped at sites near to the predicted interacting loci.



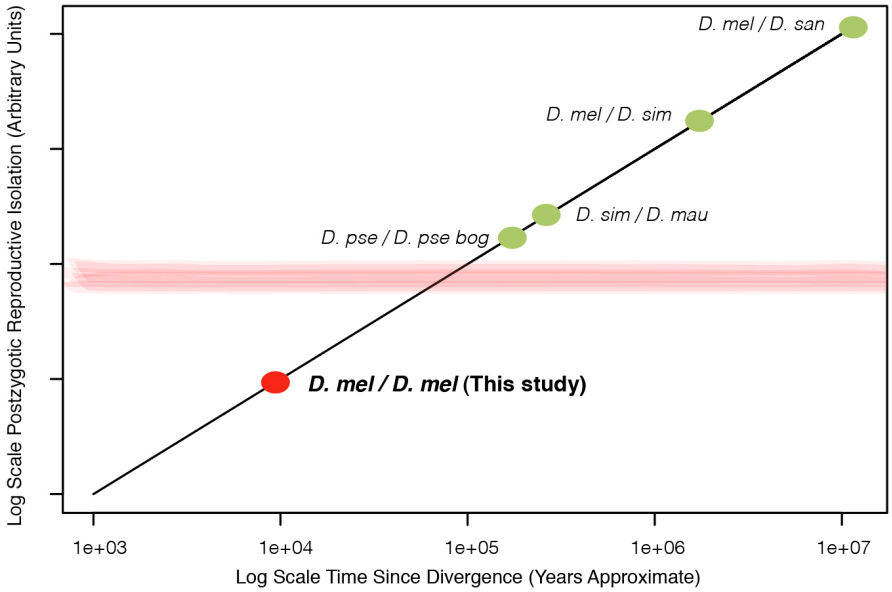

**Extended Data Figure 2 | Principal component analysis of all three DSPR RIL panels.** Green, panel A-2; blue, panel B-1; and red, panel B-2. No evidence of population structure is shown.




**Extended Data Figure 3 | D′ distribution for significant GRD.** Data are plotted across DSPR panels. On the *x* axis, D′ is a measure of the disequilibrium between interacting alleles. The red curve corresponds to a smooth curve fit using non-parametric density estimation. An outlier box-plot is presented above the histogram (the lozenge represent the mean and 95% CI, the edge of the rectangle represent the 25% and 75% percentile, the vertical bar within the median, the dots are possible outlier and the red bracket represents the shortest length that contain 50% of the data).




**Extended Data Figure 4 | Epistasis plot for each validated instance of GRD.** On the *y* axes are the productivity measurements that correspond to each genotypic class across both chromosomes. The *x* axes correspond to the genotypes on one of the chromosomes, the other genotype is represented by the colour indicated inside the plot (for example, genotype *AA*,*bb* in panel **a** is found in the lower left corner, where *AA* is read from the *x* axis and *bb* from the blue colour). **a**, GRD between chromosomes 2R and 3R (tagged by SNPs 2R:4806926, on the X axis and 3R:5870973, coloured lines) shows strong negative epistasis due to the low fitness of the *aa;bb* genotype. The additive-by-additive genetic effect is equal to −13.75 (in the sense of refs 5 and 29). **b**, GRD between chromosomes 3L and X (tagged by SNPs 3L: 11510853, on the X axis and X: 16483812, coloured lines) also shows negative epistasis. Here the additive-by-additive genetic effect equals −5.94.




**Extended Data Figure 5 | The accumulation of post-zygotic reproductive isolation through time (note log scale on axes).** Approximate divergence times of commonly studied *Drosophila* species are indicated by green circles, and the red circle indicates a reasonable expectation for divergence times of stocks used to found the DSPR (~10,000 years). The horizontal red area indicates a very approximate 'speciation threshold', and indicates that many species pairs that are commonly studied substantially exceed this threshold.

## Extended Data Table 1 | List of all significant inter-chromosomal GRD identified in the DSPR

**SNP based analysis**

| Panel | Chromosome 1 | Position 1 | Chromosome 2 | Position 2 | Number of RILs counted | 1st Major Allele | 1st Minor Allele | 2dn Major Allele | 2dn Minor Allele | 1st Major allele freq | 1st Minor allele freq | 2dn Major allele freq | 2dn Major allele freq | Major-Major frequency | Major-Minor frequency | Minor-Major frequency | Minor-Minor frequency | d | d' | r | Chi-Square | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B-2 | 2L | 2767815 | 3R | 4492436 | 391 | C | A | A | T | 0.91 | 0.09 | 0.92 | 0.08 | 0.86 | 0.05 | 0.06 | 0.03 | 0.02 | 0.27 | 0.25 | 23.70 | 5.85E-07 |
| B-2 | 2L | 8027605 | X | 13053319 | 418 | C | T | C | T | 0.94 | 0.06 | 0.88 | 0.12 | 0.85 | 0.10 | 0.03 | 0.03 | 0.02 | 0.38 | 0.25 | 26.89 | 1.12E-07 |
| B-2 | 2L | 10869984 | 3R | 10633352 | 177 | C | G | T | indel | 0.94 | 0.06 | 0.95 | 0.05 | 0.92 | 0.03 | 0.03 | 0.02 | 0.02 | 0.41 | 0.39 | 26.77 | 1.18E-07 |
| B-2 | 2L | 21657908 | 3R | 5870973 | 444 | A | C | G | A | 0.78 | 0.22 | 0.85 | 0.15 | 0.70 | 0.08 | 0.15 | 0.07 | 0.04 | 0.34 | 0.26 | 30.69 | 1.56E-08 |
| B-2 | 2R | 4806926 | 3R | 5870973 | 443 | G | A | G | A | 0.50 | 0.50 | 0.85 | 0.15 | 0.49 | 0.01 | 0.36 | 0.14 | 0.06 | 0.85 | 0.36 | 58.05 | 1.30E-14 |
| B-2 | 2R | 8464341 | X | 5753834 | 457 | A | G | T | A | 0.89 | 0.11 | 0.77 | 0.23 | 0.72 | 0.17 | 0.05 | 0.06 | 0.03 | 0.39 | 0.25 | 29.29 | 3.22E-08 |
| B-2 | 2R | 20512785 | X | 19647595 | 436 | T | G | A | C | 0.64 | 0.36 | 0.59 | 0.41 | 0.33 | 0.32 | 0.27 | 0.09 | -0.06 | -0.40 | -0.24 | 25.79 | 1.98E-07 |
| B-2 | 3L | 9627942 | X | 13622563 | 263 | G | T | A | G | 0.94 | 0.06 | 0.94 | 0.06 | 0.89 | 0.04 | 0.04 | 0.02 | 0.02 | 0.31 | 0.31 | 24.99 | 3.00E-07 |
| B-2 | 3R | 20437352 | X | 19127400 | 385 | C | T | A | T | 0.90 | 0.10 | 0.95 | 0.05 | 0.87 | 0.03 | 0.08 | 0.02 | 0.02 | 0.36 | 0.26 | 26.13 | 1.65E-07 |
| | | | | | | | | | | | | | | | | | | | | | | |
| B-1 | 2L | 66907 | 3L | 11787066 | 256 | A | G | C | T | 0.92 | 0.08 | 0.89 | 0.11 | 0.85 | 0.07 | 0.04 | 0.04 | 0.03 | 0.39 | 0.33 | 27.30 | 9.03E-08 |
| B-1 | 2L | 22131178 | 3R | 5598460 | 375 | A | G | A | T | 0.75 | 0.25 | 0.90 | 0.10 | 0.71 | 0.04 | 0.19 | 0.06 | 0.04 | 0.48 | 0.28 | 28.94 | 3.85E-08 |
| B-1 | 2L | 2896002 | X | 11823233 | 274 | A | G | T | G | 0.95 | 0.05 | 0.93 | 0.07 | 0.90 | 0.05 | 0.03 | 0.02 | 0.02 | 0.36 | 0.31 | 26.88 | 1.12E-07 |
| B-1 | 2L | 10065007 | 3R | 11588046 | 283 | C | A | T | G | 0.92 | 0.08 | 0.87 | 0.13 | 0.83 | 0.08 | 0.04 | 0.04 | 0.03 | 0.45 | 0.35 | 35.10 | 1.61E-09 |
| B-1 | 2L | 4140219 | X | 11514794 | 361 | G | A | T | indel | 0.94 | 0.06 | 0.95 | 0.05 | 0.91 | 0.03 | 0.04 | 0.02 | 0.02 | 0.33 | 0.30 | 33.13 | 4.43E-09 |
| B-1 | 2R | 3232234 | 3R | 5598051 | 356 | G | A | G | A | 0.75 | 0.25 | 0.92 | 0.08 | 0.72 | 0.03 | 0.19 | 0.06 | 0.04 | 0.56 | 0.29 | 30.98 | 1.34E-08 |
| B-1 | 2R | 14543771 | 3R | 22691609 | 355 | A | G | T | C | 0.85 | 0.15 | 0.93 | 0.07 | 0.81 | 0.04 | 0.12 | 0.04 | 0.03 | 0.41 | 0.27 | 26.33 | 1.49E-07 |
| B-1 | 3R | 13807981 | X | 8763898 | 151 | T | C | C | T | 0.91 | 0.09 | 0.88 | 0.12 | 0.84 | 0.07 | 0.04 | 0.05 | 0.04 | 0.51 | 0.45 | 30.06 | 2.17E-08 |
| B-1 | 3R | 18284739 | X | 14686047 | 378 | G | T | G | A | 0.94 | 0.06 | 0.92 | 0.08 | 0.88 | 0.06 | 0.04 | 0.02 | 0.02 | 0.30 | 0.25 | 23.46 | 6.62E-07 |
| | | | | | | | | | | | | | | | | | | | | | | |
| A-2 | 2L | 19531958 | X | 12584624 | 326 | G | T | C | G | 0.94 | 0.06 | 0.94 | 0.06 | 0.91 | 0.04 | 0.03 | 0.02 | 0.02 | 0.35 | 0.34 | 37.94 | 3.74E-10 |
| A-2 | 3L | 11510853 | X | 16483812 | 354 | A | T | G | A | 0.92 | 0.08 | 0.91 | 0.09 | 0.86 | 0.06 | 0.05 | 0.03 | 0.02 | 0.28 | 0.27 | 26.16 | 1.62E-07 |
| A-2 | 3R | 23793328 | X | 14472525 | 64 | A | T | C | T | 0.84 | 0.16 | 0.84 | 0.16 | 0.80 | 0.05 | 0.05 | 0.11 | 0.08 | 0.64 | 0.64 | 26.58 | 1.31E-07 |
| A-2 | 2L | 16549805 | 3L | 10566820 | 236 | C | T | A | G | 0.92 | 0.08 | 0.92 | 0.08 | 0.87 | 0.05 | 0.05 | 0.03 | 0.03 | 0.37 | 0.37 | 32.37 | 6.55E-09 |

**HMM based analysis**

| Panel | Chromosome 1 | Position 1 | Chromosome 2 | Position 2 | Number of RILs counted | 1st Major Allele | 1st Minor Allele | 2dn Major Allele | 2dn Minor Allele | Incompat. Panel-1 A | Incompat. Panel-1 B | Incompat. Panel-2 A | Incompat. Panel-2 B | Maximun \ Chi-square | Maximun P-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B-2 | 2L | 2767815 | 3R | 4492436 | 391 | C | A | A | T | 2,3 | 6,7,8 | 1,5 | 2,3,4,6,7 | 22.285 | 1.22E-06 |
| B-2 | 2L | 8027605 | X | 13053319 | 418 | C | T | C | T | 2,7 | 1,3,6,8 | 5,6 | 1,2,3,7 | 28.552 | 4.71E-08 |
| B-2 | 2L | 10869984 | 3R | 10633352 | 177 | C | G | T | indel | 1,3 | 6,7 | 2 | 3,4,5,6,7 | 14.007 | 9.69E-05 |
| B-2 | 2L | 21657908 | 3R | 5870973 | 444 | A | C | G | A | 1,2,7 | 3,6,8 | 2 | ,3,4,5,6, | 31.207 | 1.19E-08 |
| B-2 | 2R | 4806926 | 3R | 5870973 | 443 | G | A | G | A | 3 | 1,2,6,7,8 | 2 | ,3,4,5,6, | 62.734 | 1.20E-15 |
| B-2 | 2R | 8464341 | X | 5753834 | 457 | A | G | T | A | 1,3,8 | 2,5,6 | 1,4,5 | 2,3,7 | 37.070 | 5.84E-10 |
| B-2 | 2R | 20512785 | X | 19647595 | 436 | T | G | A | C | 4,8 | 2,3,5,6,7 | 3,5 | 1,2,4,6 | 28.937 | 3.86E-08 |
| B-2 | 3L | 9627942 | X | 13622563 | 263 | G | T | A | G | 2,3,5 | 1,6,7,8 | 5,7 | 1,3,6 | 17.625 | 1.41E-05 |
| B-2 | 3R | 20437352 | X | 19127400 | 385 | C | T | A | T | 3,5 | 1,4,6 | 8 | ,2,4,5,6, | 10.122 | 7.95E-04 |
| | | | | | | | | | | | | | | | |
| B-1 | 2L | 66907 | 3L | 11787066 | 256 | A | G | C | T | 7 | 1,2,3,6,8 | 5 | 2,3,4,6,7 | 16.936 | 2.04E-05 |
| B-1 | 2L | 22131178 | 3R | 5598460 | 375 | A | G | A | T | 1,2,3,6 | 4,7,8 | 1,2,3,6 | 4,7,8 | 28.154 | 5.79E-08 |
| B-1 | 2L | 2896002 | X | 11823233 | 274 | A | G | T | G | 6,7,8 | 2,3,4,5 | 2,8 | 1,3,5,6,7 | 27.374 | 8.67E-08 |
| B-1 | 2L | 10065007 | 3R | 11588046 | 283 | C | A | T | G | 1,7 | 2,3,4,6 | 2,8 | 1,2,3,6,7 | 29.341 | 3.13E-08 |
| B-1 | 2L | 4140219 | X | 11514794 | 361 | G | A | T | indel | 3 | 2,4,6,7,8 | 3,5,6 | 1,2,7,8 | 44.297 | 1.44E-11 |
| B-1 | 2R | 3232234 | 3R | 5598051 | 356 | G | A | G | A | 4,7,8 | 1,2,3,6 | 4,7,8 | 1,2,3,6 | 30.049 | 2.17E-08 |
| B-1 | 2R | 14543771 | 3R | 22691609 | 355 | A | G | T | C | 2,6 | 1,3,5,8 | 4 | 2,3,5,6,7 | 26.540 | 1.34E-07 |
| B-1 | 3R | 13807981 | X | 8763898 | 151 | T | C | C | T | 3,4 | 1,2,6,7,8 | 5,7,8 | 1,3,6 | 18.243 | 1.02E-05 |
| B-1 | 3R | 18284739 | X | 14686047 | 378 | G | T | G | A | 2,7 | 1,4,6,8 | 5 | 1,3,6,8 | 29.620 | 2.71E-08 |
| | | | | | | | | | | | | | | | |
| A-2 | 2L | 19531958 | X | 12584624 | 326 | G | T | C | G | 1 | 2,4,5,7 | 5 | ,2,3,4,6, | 42.031 | 4.59E-11 |
| A-2 | 3L | 11510853 | X | 16483812 | 354 | A | T | G | A | 4 | 2,3,5,6 | 5 | 1,3,4,6,7 | 34.180 | 2.58E-09 |
| A-2 | 3R | 23793328 | X | 14472525 | 64 | A | T | C | T | 5 | 3,4,6,7 | 1,5 | 2,3,4,6,7 | 12.913 | 1.74E-04 |
| A-2 | 2L | 16549805 | 3L | 10566820 | 236 | C | T | A | G | 1,8 | 2,4,5,7 | 4 | 3,5 | 20.660 | 2.86E-06 |

**Extended Data Table 2 | List of significant inter-chromosomal GRD in the *Arabidopsis* MAGIC panel and maize NAM panel**

**MAGIC panel**

| Chromosome 1 | Position 1 | Chromosome 2 | Position 2 | Number of RIL counted | Chi Square | P-value |
|---|---|---|---|---|---|---|
| 4 | 15897262 | 5 | 25151311 | 522 | 53.75 | 2.27E-13 |
| 1 | 4947328 | 4 | 11984761 | 513 | 47.07 | 6.85E-12 |
| 1 | 5447591 | 4 | 11984761 | 513 | 45.49 | 1.54E-11 |
| 2 | 5666979 | 3 | 15317766 | 442 | 41.02 | 1.51E-10 |
| 3 | 8190257 | 5 | 23705451 | 524 | 32.09 | 1.47E-08 |
| 1 | 30044327 | 3 | 21116640 | 517 | 24.91 | 6.00E-07 |
| 1 | 2211127 | 4 | 3781192 | 522 | 19.58 | 9.64E-06 |

**NAM panel**

| PANEL | Marker 1 ID | Chromosome 1 | Position 1 | Marker 2 ID | Chromosome 2 | Position 2 | Number of RIL counted | Chi Square | P-value |
|---|---|---|---|---|---|---|---|---|---|
| 4 | PZB01647.1 | 1 | 137.6 | PZA01951.1 | 8 | 32.3 | 175 | 27.50 | 1.57E-07 |
| 25 | PZA03032.19 | 3 | 82.1 | PHM7898.10 | 7 | 111.8 | 164 | 25.00 | 5.73E-07 |
| 12 | PZA00494.2 | 3 | 97.8 | PZA01964.29 | 8 | 106.9 | 136 | 24.88 | 6.09E-07 |
| 7 | PZA01753.1 | 2 | 40.6 | PZB00752.1 | 7 | 71.2 | 175 | 24.70 | 6.72E-07 |
| 4 | PZA00261.6 | 5 | 66.8 | PZA00090.1 | 8 | 70.6 | 160 | 23.58 | 1.20E-06 |