

Are there Physical Linkages between Genes that have Synergistic Fitness Effects?

Juliet Byrnes, John Murray, Mark M. Tanaka, Ben Goldys, Antony Bellanto, Luis Cayetano, William Sherwin

Abstract

Many of the effects on fitness in population genetics are due not to single locations in the genome, but to the interaction of genetic variants at multiple locations in the genome. Of particular interest are ‘completely epistatic’ interactions, where a combination of genetic variants is required to produce an effect, and the effect cannot occur with any other combination. In diploids, epistasis is strongly connected to meiotic recombination, a process which can both assemble and destroy beneficial combinations of genetic variants. Additionally, epistatic interactions can be hard to detect in empirical studies, and mathematical models of epistasis and recombination are challenging to analyse, so despite their ubiquity epistatic interactions are regularly not considered. As a result, there is little consensus on when high levels of recombination might be expected, or how strongly recombination affects beneficial or deleterious fitness effects controlled by epistatic interactions. We address this question by conducting a meta-analysis and simulations. The meta-analysis used data drawn and curated from *Drosophila melanogaster* studies in Flybase. We extracted studies relating genetic combinations and phenotypically detectable effects on fitness, then analysed the relationship between the rate of recombination and effect on fitness with a statistical model. We also ran simulations under a two-locus Wright-Fisher model with recombination and epistatic selection. The results of both approaches indicated a tendency for genetic combinations with an epistatic effect on fitness to occur in an environment of reduced meiotic recombination. Two possible explanations for this are that the variants controlling such interactions are selected for in regions where there is little recombination, or that such interactions lead to selection for lower rates of recombination in the regions where those variants appear.

Introduction

Theories of the interaction of meiotic recombination and selection have had a long and tumultuous history, with almost a century of research dedicated to understanding how a system as

reproductively costly as recombination could have evolved and why it is pervasive in so many species (Otto and Lenormand 2002). Recombination creates but also destroys ‘haplotypes’, DNA molecules carrying combinations of alleles at different loci. Selection is also complex, especially when based on allelic variation at two or more loci in haplotypes. Of particular importance is epistasis, a multilocus selection effect in which alleles at different loci interact to produce an effect on fitness, either synergistically or antagonistically, that neither could produce in the absence of the other. Synergistic epistasis means that if each locus has one allele with the same type of fitness effect (i.e. both are either beneficial or deleterious), the combination of those alleles will have a stronger effect on fitness than effects of both the alleles if they are assumed to act independently. In contrast, under antagonistic epistasis, that combination of alleles will have a weaker effect. When neither allele has a fitness effect by itself, with the effect only appearing when both alleles are present, this is known as complete epistasis. If we arbitrarily label the alleles at each locus ‘0’ and ‘1’, then we call the genetic types (or ‘haplotypes’) with a ‘0’ at both loci or a ‘1’ at both loci the ‘coupling haplotypes’, and the remaining two haplotypes the ‘repulsion haplotypes’ [TABLE 1]. Henceforth we consider the coupling haplotype with ‘0’s at both loci to be under selection that is completely epistatic (no single locus selection effects occur) and label it as the ‘beneficial haplotype’ or the ‘deleterious haplotype’ depending on the direction of the selection effect. We refer to the haplotype with ‘1’s at both loci as the ‘non-beneficial coupling haplotype’ or the ‘non-deleterious coupling haplotype’ also depending on the direction of the same selection effect, meaning that it has the same fitness value as the repulsion haplotypes. Epistasis in its various forms can speed up or slow down selection, can induce selection plateaus (Moore and Williams 2015), and is regularly considered as a possible driver of the evolution of recombination rates.

		Locus 1	
		Allele 0	Allele 1
Locus 2	Allele 0	(0,0) Coupling, beneficial or deleterious	(0,1) Repulsion
	Allele 1	(1,0) Repulsion	(1,1) Coupling, non-beneficial or non-deleterious

[TABLE 1] Caption: Table describing the names of the multilocus haplotypes we use in this paper.

These are consistent with their usual names in the population genetics literature.

Possible explanations for the prevalence of recombination are broadly divided into three categories, with epistasis often playing a prominent role (Otto 2009, Hartfield and Keightley 2012).

- 1) It has been proposed that recombination 1) prevents beneficial mutations at separate loci from competing with one another and 2) allows the creation of chromosomes with no deleterious mutations from two chromosomes that each have a deleterious mutation at different loci. These phenomena are termed ‘clonal interference’ (Hill and Robertson 1966) and ‘Muller’s Ratchet’ respectively (Muller 1932, Felsenstein 1974). A recurring theme in the recombination literature deals with the speed of production of double mutants - i.e. a haplotype containing novel mutations at two different loci - particularly if those mutations are beneficial. An extensive body of theoretical literature has shown that higher rates of recombination cause a double mutant to occur much earlier in an evolutionary history than otherwise (Barton 1995, Suzuki 1997, Christiansen *et al.* 1998, Weissman *et al.* 2010).
- 2) Another argument proposes that recombination makes a species more successful under fluctuating selection, such as by allowing rapid production of novel, possibly adaptive haplotypes when a species is co-evolving with its parasite (Kondrashov & Yampolsky 1996, Peters and Lively 1999). Therefore some authors have considered models with fluctuating levels of selection (Sasaki and Iwasa 1986), epistasis (Suzuki 1997, Weissmann *et al.* 2010) or fluctuating selection *and* epistasis (Peters and Lively 1997, Otto and Michalakis 1998, Gandon and Otto 2007).
- 3) A third proposal is that recombination increases the effectiveness of selection in purging deleterious mutations when they are common and act synergistically (Kondrashov 1988, Gandon and Otto 2007). Empirical explorations of this by Fry *et al.* (2004) and Rosa *et al.* (2005) found evidence of synergistic epistasis in *Drosophila melanogaster*, though Rosa *et al.* contend that this is too occasional to explain the origin or maintenance of meiotic recombination.

The situation is complicated by the fact that early models of recombination, which did not include genetic drift, showed no benefit to fitness of recombination (Felsenstein 1974; Otto and Barton 2001), despite the fact that recombination is ubiquitous and highly variable at every scale (Stapley *et al.* 2017). Later stochastic models of recombination, which do include genetic drift, pose strong mathematical difficulties, though there has been some progress in the mathematical analysis of these models (Mano 2013, Jenkins and Song 2015, Esser *et al.* 2016). An important simplification is to restrict consideration to populations that are at ‘linkage equilibrium’, i.e. the proportion of a

haplotype (i, j) in the population is equal to the proportion of haplotypes possessing the i allele at the first locus times the proportion of haplotypes possessing the j allele at the second locus ($p_{ij} = p_i p_j$). When this occurs and there is no selection, evolution at each locus can be modelled independently (Durrett 2008, Mano 2013). As a consequence, the quantity $p_{ij} - p_i p_j$ is useful in modelling recombination, and is known as a measure of ‘linkage disequilibrium’.

However, recombination can have positive, negative or equivocal effects. For example, recombination can clearly be a disadvantage due to breaking up selectively advantageous haplotypes. Therefore, the faster production of double mutants in the presence of recombination would be offset by a relative increase in the time taken for the double mutant to reach fixation in the population when compared to the non-recombining case (Barton 1995, Suzuki 1997, Christiansen *et al.* 1998). Additionally, Weissman *et al.* (2010) contend that there is an optimal rate of recombination that causes the appearance of the first double mutant to occur much earlier if the alleles experience epistatic selection – but when some ‘critical’ recombination rate is exceeded, the first double mutant will take an exponentially longer time to first occur. However, these models have also been criticised for being too restrictive for an underlying biological mechanism to have evolved in the first place (Baumgardner *et al.* 2013) or for requiring a parameter space that is too restrictive to allow persistence (Otto 2009).

It can be seen from the above that plausibly there are both advantages and disadvantages for recombination between epistatically interacting loci. The fitness effects epistasis could have on ecological and genetic processes mean that there are important theoretical and empirical implications from understanding the relationship between epistasis and recombination. We therefore investigated the relationship between recombination, epistasis, selection and drift in two ways: 1) via a meta-analysis using Flybase and 2) via simulations of small randomly mating populations under epistatic selection at two recombining loci. Specifically, we address the following three questions about pairs of loci that interact synergistically to have an effect on fitness:

1. If the interaction is deleterious, do the loci recombine more often than the average pair of loci, as proposed by Kondrashov? Or will they have a lower rate of recombination, as suggested by the body of evidence in Otto and Lenormand’s review (2002)?
2. If the interaction is beneficial, do the loci recombine more often than the average pair of loci, as suggested by clonal interference and Muller’s Ratchet? Or will they have a lower rate of recombination?

3. If the interaction is beneficial, do the loci have an optimal rate of recombination, as suggested by Weissman *et al.* (2010)?

Methods

Flybase Meta-analysis: Data on Interactions and Recombination

We were interested in whether the fitness of a phenotype determined by pairs of genetic loci had an effect on the amount of recombination between those loci. We reviewed a number of different databases with genetic data on various species (The Arabidopsis Information Resource (Berardini *et al.* 2004), MaizeGDB (Andorf *et al.* 2015), ZFIN (Howe *et al.* 2013), Flybase (Gramates *et al.* 2016)) to determine suitability of use in conducting this investigation. Our criteria for selection were: 1) the database links alleles to observed phenotypes in the study species; 2) the expression of several observed phenotypes depends on the presence of particular alleles at 2 or more loci; and 3) the alleles at these loci were naturally occurring and the genetic loci involved had been mapped genetically, so that we could determine recombination rates. These criteria were satisfied only by Flybase, a database containing an extensive set of genetic and phenotype data on *D. melanogaster*.

First, we identified the pairs of genetic loci that determined a phenotype and whether there was epistasis between them; second, we estimated the recombination rates between these genetic loci based on genetic map data; thirdly and finally, we fitted a regression model to analyse the data with recombination rates as the response and ‘epistasis class’ as the predictor variables, where the ‘epistasis class’ of a locus pair is defined as the effect of that locus pair on the genotype of a fly. Epistasis class was determined by extracting data from Flybase on phenotypes, the genetic loci that determine that phenotype and the particular phenotype information associated with certain combinations of alleles at those loci. We considered phenotypes determined by pairs of loci only, and sorted them into three epistasis classes; those where the genetic loci had 1) a phenotypically discernible interaction, with an impact on fitness (abbreviation PDFI); 2) a phenotypically discernible interaction, which may or may not have an impact on fitness (abbreviation PDNF); or 3) with no phenotypically discernible interaction (abbreviation NDI), i.e. the genetic loci acted on the phenotype independently but not epistatically. We processed the data by extracting the results of studies on the effect of various genetic loci on phenotypes, excluding pairs involving alleles which did not occur naturally, and stratified them into types of interaction.

We searched for haplotypes consisting of alleles at two or more loci, and whether these haplotypes had any epistatic effects on fitness. We used a genetics interaction file previously computed by Flybase on allele interactions and the phenotypes associated with them. We refer to this as [GENE_INT]. We also used the Flybase SQL database for the Flybase meta-analysis. The files used are the February 2017 release. The table in [GENE_INT] contains 4 columns: 1) the symbol (a human readable name) of an allele in the interaction; 2) the unique Flybase reference of the allele; 3) the description of the type of interaction and all genetic loci and alleles involved; and 4) the unique Flybase reference of the study which determined the genetic loci involved in the phenotype. An example of the tabulation in Flybase is given in [TABLE 2].

##allele_symbol	allele_FBal#	Interaction	FBrf#
Abi[P1]	FBal0239154	Abi[+]/Abi[P1] is a suppressor partially of lethal phenotype of Abl[4]/Abl[1]	FBrf0208583
Abl[1]	FBal0028708	Abl[1], N[55e11], Nrt[M54] has eye phenotype	FBrf0102318
Abl[unspecified]	FBal0089849	Abl[unspecified], Nrt[unspecified] has lethal phenotype	FBrf0052941

[TABLE 2] Caption: An example of what a phenotype-gene association looks like when obtained from Flybase. The first column gives a locus name followed by the allele name in square brackets. The second column gives the unique Flybase reference for this allele. The third column gives the names of all alleles and the loci they exist on involved in the interaction as well as the type of interaction. In the first row, third column, there is a lethal interaction between the loci ‘Abi’ and ‘Abl’. In the second row, third column, there is a non-fitness interaction between three loci; ‘Abl’, ‘N’ and ‘Nrt’. In the third row, third column, there is a lethal interaction between the loci ‘Abl’ and ‘Nrt’.

Each data entry contains a phenotype and the alleles associated with it, so there may be multiple data entries which involve the same sets of loci. Our search method identified the names of the genetic loci involved in each data entry. Details of this search method can be found in the supplement [Code Supplement, Part I: Input and Search Patterns]. There were 98087 entries for *D. melanogaster* locus pairs. We identified 81265 phenotypes determined by sets of loci containing two loci, and 9585 containing three or more. We removed phenotypes involving more than three loci for simplicity. In practice, 6754 out of the 9585 (70.5%) cases where there was a three-way

interaction also contained a two-way interaction. We then standardised the names of the genetic loci by using a SQL query to locate their unique Flybase references and name by symbol. This was necessary because a single locus can have multiple different symbolic names in the Flybase database, but will always have a unique Flybase reference ID. Because we were only interested in alleles which occur in natural populations or spontaneously in cultivated populations, we used the SQL database to extract the list of mutagens associated with each allele. We then removed any entries where a genotype had an allele with a mutagen listed other than ‘natural population’. Of the original 98087 locus pairs, only 5346 locus pairs were associated with natural mutations.

We then classified these 5346 locus pair entries into three ‘epistasis classes’ based on their effects on the phenotype. If an entry contained the terms ‘non-suppressor’, ‘non-enhancer’, ‘non-enhanceable’ or ‘non-suppressible’, this meant that the allele at one of the loci in the data entry had no discernible effect on the expression of a phenotype associated with the allele at the other locus, i.e.: they acted on the phenotype independently but not epistatically. We were able to detect these in entries by reducing them to the stems ‘non-enh’ or ‘non-suppr’. Furthermore, the fitness effects recorded for the data entries consisted of four types. They were: 1) lethality: where flies possessing the genotype had a higher death rate than those without the genotype; 2) sterility: where flies possessing the genotype had fewer offspring than those without the genotype; 3) fertility: where flies possessing the genotype had *more* offspring than those without the genotype; and 4) viability: where flies possessing the genotype were more likely to survive than those without the genotype. We classified these effects into epistasis classes according to the following protocol:

1. PDFI: phenotypically detectable fitness interactions: the alleles at the genetic loci interacted and this had a discernible impact on fitness in the form of survival or fertility. These did not contain the stems ‘non-enh’, ‘non-suppr’, but did contain the terms ‘lethal’, ‘fertile’, ‘sterile’, or ‘viable’.
2. PDNF: phenotypically detectable interactions with no known impact on fitness: the alleles interacted, but their interactions had no visible impact on fitness. These did not contain the word stems ‘non-enh’, ‘non-suppr’, and did not contain the terms ‘lethal’, ‘fertile’, ‘sterile’ or ‘viable’.
3. NDIs: No phenotypically discernible interaction: the alleles at the genetic loci did not interact. In the Flybase data, these all contained the abbreviated stems ‘non-enh’ or ‘non-suppr’, indicating that the alleles at each loci acted only independently on the phenotype and

not epistatically (the allele at one locus neither enhanced nor suppressed the phenotypic effect of the allele at the other locus).

Where a single pair of loci had different allele combinations appearing in the data that would allow entry to more than one of the three categories above, we used the following decision method [FIGURE 3]:

1. If any observed combinations of alleles at the locus pair were classified as PDFI, the locus pair was classified as PDFI.
2. If any observed combinations of alleles at the locus pair were classified as PDNF, and none were classified as PDFI, the locus pair was classified as PDNF.
3. If no observed combinations of alleles at the locus pair were classified as PDFI or PDNF, the locus pair was classified as NDI.

As a result of this process, many (5076) duplicate locus pair entries were removed, leaving a total of 270 unique locus pairs. We classified 39 locus pairs as PDFI, 184 locus pairs as PDNF, and 47 locus pairs as NDI.

We used the genetic map data from the precomputed Flybase file to filter for locus pairs that were on the same chromosome, because meiotic homologous recombination only occurs between loci lying on the same chromosome. The genetic map data also gave the average rate of recombination between each locus and a designated end of the chromosome (telomere). These data may have been determined by experiment or computed using information from nearby loci, depending on the locus of interest. The unit for recombination is centimorgans (cM), where 1 cM indicates that the average number of recombinations expected in a generation is 0.01. We then use this to approximate the probability of recombination between two loci – if two loci are 1 cM apart, the probability of a recombination occurring in a single generation between the two loci is 0.01. To reflect the biological reality of the maximum possible observed recombination probability in diploid organisms, we truncated the probabilities of recombination at 0.5. We discarded locus pairs where recombination data were not available for one or more of the loci involved. This resulted in the removal of 21 data points – 2 (out of 47) NDIs and 19 (out of 184) PDNFs. Finally, we removed interactions where one locus was nested inside another (as loci can span several megabases). This removed 3 (out of 39) PDFIs. In the final extracted dataset there were 36 PDFIs, 165 PDNFs and 45 NDIs [TABLE 7]. 53 locus pairs were on the X chromosome, 100 pairs were on chromosome 2, and 93 pairs were on chromosome 3. There were no locus pairs located on chromosome 4. The final dataset contained the recombination rate between each pair of loci (c), the epistasis class of the two

Flybase Meta-analysis: Regression Model to Test for Effects of Epistasis Class and Chromosome Number on Recombination

For our study, we wished to know whether the epistasis class of a locus pair drove the probability of recombination between the two loci. We also needed to control for chromosomal identity due to the large variance in recombination rates between chromosomes. We needed a model that could deal with data having a large number of points on 0 and 0.5 plus a spread of values in between.

Therefore we analysed the extracted dataset using a partially stratified zero-and-one-inflated beta regression model (Ospina and Ferrari 2012). The zero-and-one-inflated beta regression model contains sub-units which we will call ‘components’, and each component contains one response variable, a set of explanatory variables, a link function, and parameters associated with each explanatory variable in the generalised linear model framework. Therefore when modelling the recombination rate as a response of the zero-and-one-inflated beta regression model the response variables of the four sub-model components are:

- 1) the mean of the intermediate recombination rates (i.e. recombination rates that were not 0 or 0.5);
- 2) the precision of the intermediate recombination rates, defined by $\text{precision} = [\text{mean}(1-\text{mean}) / \text{variance}] - 1$;
- 3) the relative probability that $c = 0$ rather than $0 < c < 0.5$; and
- 4) the relative probability that $c = 0.5$ rather than $0 < c < 0.5$.

The first and second components model the recombination rates that are not at the extremes of 0 and 0.5 (the Intermediate Recombination Rates) according to a beta distribution. The precision variable describes the relationship between the variance and the mean and is allowed to vary with different explanatory variables. The third component models the probability of the recombination being 0 divided by the probability of the recombination rate lying in the intermediate range of $0 < c < 0.5$ (the Odds of Minimum Recombination). The fourth component models the probability of the recombination rate being 1 divided by the probability of the recombination rate lying in the intermediate range of $0 < c < 0.5$ (the Odds of Maximum Recombination).

The link functions for each component response variable are given in [TABLE 4] Each component sub-model shares the same set of explanatory variables. These were epistasis class, chromosomal

identity and the interaction of epistasis class and chromosomal identity. However, each sub-model can - and in practice, will - have different parameter fits for each explanatory variable.

Component Name	Link Function
Intermediate Recombination Mean	Logit
Intermediate Recombination Precision	Logit
Odds of Minimum Recombination	Log
Odds of Maximum Recombination	Log

[TABLE 4] Caption: The link functions for each component of the zero-and-one-inflated beta regression model. Each component submodel is of the form: $link(response) = \sum_i a_i x_i$ where the x_i 's are the explanatory variables, the a_i 's are the parameters of the sub-model component, and i is the index of the covariates included in the component.

Epistasis class was coded as two binary variables. The first epistasis variable was 1 if the epistasis class was PDFI, and 0 otherwise; we labelled this variable as 'Fitness'. The second variable took the value 1 if the epistasis class was PDNF *or* PDFI, and 0 otherwise; we labelled this variable 'Interaction'. If both variables were zero, the epistasis class was NDI. There were also two binary variables coding chromosomal identity. The first variable took the value 1 if the locus pair was on chromosome 2 and 0 otherwise; we labelled this variable 'Chrom2'. The second variable took the value 1 if the locus pair was on chromosome 3 and 0 otherwise; we labelled this variable 'Chrom3'. If both of these variables were 0, the locus pair was on chromosome 1 (the X chromosome). The parameter coding is shown in [TABLE 5].

We could not include all explanatory variables in every sub-model because our sample size was very small (246) compared to the number of parameters described by having two categorical variables with three levels (epistasis class and chromosomal identity) and the interaction of those variables (8 parameters), and having four model components (32 parameters in total). We therefore required a model selection procedure. We used a generalized Akaike information criterion (GAIC) provided by the R package GAMLSS (Rigby and Stasinopoulos 2005) to simultaneously perform model selection and fit the parameters of the model. The GAIC model selection procedure excludes explanatory variables from each sub-model if they do not contribute substantially to the model fit.

	PDFI	PDNF	NDI
Chromosome 1	Intercept + Fitness +	Intercept + Interaction	Intercept

	Interaction		
Chromosome 2	Intercept + Fitness + Interaction + Chrom2 + Fitness*Chrom2 + Interaction*Chrom2	Intercept + Interaction Chrom2 + Interaction*Chrom2	Intercept + Chrom2
Chromosome 3	Intercept + Fitness + Interaction + Chrom3 + Fitness*Chrom3 + Interaction*Chrom3	Intercept + Interaction Chrom3 + Interaction*Chrom3	Intercept + Chrom3

[TABLE 5] Caption: Coding of the variable parameters for the nine different categories for a component sub-model containing all the explanatory variables stated above (epistasis class, chromosomal identity and their interactions). The ‘Intercept’ parameter is *not* a true baseline, because the variables are categorical. Four components result in a total of 32 parameters to fit if all explanatory variables are included in all the component sub-models. The parameter for a locus pair on chromosome 1 in the NDI class is coded as the intercept. Variables with a ‘*’ between them indicate interactions of those two variables.

We provide a brief explanation of the interpretation of the parameters that describe the relative probabilities of the minimum and maximum recombination rates. We assign each locus pair to a category, which is the particular combination of chromosome and epistasis class to which it belongs [TABLE 7]. For instance, a locus pair on Chromosome 3 and in epistasis class NDI would be in category (NDI, 3). We consider the component (4) above, which describes the relative probability that $c = 0.5$ rather than $0 < c < 0.5$. The response variable of this component sub-model is the log of the probability that pairs of loci in a particular category freely recombine (at the maximum recombination value, $c = 0.5$) divided by the probability that pairs of loci in the same category recombine at an intermediate rate ($0 < c < 0.5$), i.e.:

$$\log\left(\frac{\text{probability of a locus pair in category A recombining at the maximum rate}}{\text{probability of a locus pair in category A recombining at an intermediate rate}}\right)$$

$$= \log\left(\frac{p_{\text{maximum}}(A)}{p_{\text{intermediate}}(A)}\right)$$

=

$$\text{Intercept} + a_{\text{Interaction}} * \text{Interaction}(A) + a_{\text{Fitness}} * \text{Fitness}(A) + a_{\text{Chrom2}} * \text{Chrom2}(A) + a_{\text{Chrom3}} * \text{Chrom3}(A) + a_{\text{Interaction*Chrom2}} * (\text{Interaction} * \text{Chrom2})(A) + a_{\text{Fitness*Chrom2}} *$$

$$(Fitness * Chrom2)(A) + a_{Interaction*Chrom3} * (Interaction * Chrom3)(A) + a_{Fitness*Chrom3} * (Fitness * Chrom3)(A)$$

Note that if we consider the category to be the (NDI, 1) category, we get:

$$\log\left(\frac{p_{maximum}(NDI, 1)}{p_{intermediate}(NDI, 1)}\right) = Intercept$$

Similarly, if we consider category to be the (PDNF, 3) category, we get:

$$\log\left(\frac{p_{maximum}(PDNF, 3)}{p_{intermediate}(PDNF, 3)}\right) = Intercept + a_{Interaction} + a_{Chrom3} + a_{Interaction*Chrom3}$$

Each parameter or some sum of parameters (excluding the intercept) therefore has an interpretation as the log of an odds ratio. For instance:

$$\begin{aligned} a_{Interaction} &= \log\left(\frac{p_{maximum}(PDNF, 1)}{p_{intermediate}(PDNF, 1)}\right) - \log\left(\frac{p_{maximum}(NDI, 1)}{p_{intermediate}(NDI, 1)}\right) \\ &= \log\left(\frac{\frac{p_{maximum}(PDNF, 1)}{p_{intermediate}(PDNF, 1)}}{\frac{p_{maximum}(NDI, 1)}{p_{intermediate}(NDI, 1)}}\right) \end{aligned}$$

So $e^{a_{Interaction}}$ is a ratio of odds. The numerator of this ratio is the odds of a locus pair in the (PDNF,1) category recombining at the maximum rate over recombining at an intermediate rate. The denominator of this ratio is the odds of a locus pair in the (NDI,1) category recombining at the maximum rate over recombining at an intermediate rate. If the ratio is greater than 1 (>1), this means that the odds of a locus pair in the (PDNF,1) category recombining at the maximum rate over recombining at an intermediate rate is greater than the corresponding odds for a locus pair in the (NDI,1) category.

The conclusion is the opposite if the ratio is less than 1 (<1) and the odds are equal if the ratio is equal to 1. This interpretation is why the component is referred to as the ‘Odds of Maximum Recombination Component’. To illustrate this, suppose that the probability of a locus pair in the (PDNF,1) recombining at a maximum rate is 0.2 and the probability of an intermediate rate is 0.1.

Then the odds for this is 2. Further suppose that the probability of a locus pair in the (NDI,1) class recombining at a maximum rate is 0.3 and the probability of an intermediate rate is 0.3. Then the odds is 1. Now the odds ratio is 2, and we can follow the interpretation above, but the *probability* of a maximum rate for a locus pair in the (PDNF,1) class is in fact *smaller* than the probability of a maximum rate for a locus pair in the (NDI,1) class. Some care must therefore be taken with the interpretation of results. All the arguments above are the same with 0s instead of 0.5s for the Odds of Minimum Recombination Component.

Simulations of Epistatic Selection and Recombination

We simulated the evolution of a small population at two loci according to a single-sex Wright-Fisher haploid model with recombination and synergistic epistatic selection. The Wright-Fisher haploid model assumes a fixed population size and random mating. The next generation is determined by randomly drawing individuals from the previous generation under a multinomial distribution with probabilities determined by the proportion of the haplotype in that generation and the selection strength. With recombination, the model is modified to include a binomial element – for each new haploid individual, there is a probability ‘ c ’ of recombination while reproducing. When a recombination event occurs, another individual is drawn and the haplotype is modified by recombination between the two individuals drawn. We further modified this model by using a beneficial complete epistasis fitness scheme. This scheme did not include sex and is technically a haploid model with recombination. In general, however, haploid models provide very good approximations to diploid models of half the size in most cases and it is a convention to use haploid models where possible (Durrett 2008).

Population size was varied in multiples of 50, with the smallest containing 50 individuals and the largest containing 500 individuals. We modelled haplotypes of two biallelic loci, with haplotypes labelled as in [TABLE 1]. We set the relative fitness of the beneficial haplotype (0,0) to be $1 + s$, where s was allowed to vary from 0 to 0.05, and the other three haplotypes had fitness 1. We varied the recombination probability c from 0 to 0.025. We then investigated higher recombination probabilities with simulations in which recombination probabilities ranged from 0.01 to 0.5, but only with a single population size (200). We also considered twelve different cases for the initial genotype frequencies with different levels of linkage disequilibrium and relative proportions of the beneficial genotype in the population. The details can be found in [TABLE 6].

Each simulation ran for 50 generations, with 10,000 replicates per simulation. This period of time was sufficient for several simulations to reach approximate linkage equilibrium; these patterns are reflected in our graphs [FIGURE 10, FIGURE S3 Cases]We assessed the behaviour of the model by examining the probability that beneficial haplotype becomes fixed in the population (100%) and the mean time for this fixation, if it occurred. The code for the simulations is written entirely in R and can be found in the appendix.

Case #	Proportion of Haplotype (0,0) (beneficial, coupling)	Haplotype (0,1) (repulsion)	Haplotype (1,0) (repulsion)	Haplotype (1,1) (coupling)	LD
1	0	0.5	0.5	0	-0.25
2	0.24	0.26	0.26	0.24	-0.01
3	0.5	0	0	0.5	0.25
4	0.24	0.5	0	0.26	0.0624
5	0.1	0.2	0.2	0.5	0.01
6	0.5	0.5	0	0	0
7	0.1	0	0	0.9	0.09
8	0.1	0.2	0	0.7	0.07
9	0	0.24	0.26	0.5	-0.0624
10	0.02	0.48	0.5	0	-0.24
11	0	0.1	0.1	0.8	-0.01
12	0.02	0.1	0.1	0.78	0.0056

[TABLE 6]

Caption: The initial haplotype proportions for each of the twelve cases, and their linkage disequilibrium, calculated by $p_{00} - (p_{01} + p_{00})(p_{10} + p_{00})$

Results

Flybase Meta-analysis

The zero-and-one-inflated beta regression model appeared to be an appropriate model for the recombination data. A model where either the distribution of link function was incorrectly specified would likely display a pattern in the residuals. The residual plots for the fitted model showed no detectable trend, supporting the choice of the zero-and-one-inflated beta regression model [FIGURE S5].

There were 36 locus pairs with phenotypically detectable fitness interactions [PDFIs], 165 locus pairs with phenotypically detectable interactions without a fitness effect [PDNFs] and 45 locus pairs with no phenotypically detectable interaction [NDIs] in the final data set. (These abbreviations are defined in Methods). Out of a total of 246 locus pairs, there were 19 locus pairs (7.7%) with the minimum recombination rate of $c = 0$ and 51 locus pairs (20.7%) with the maximum recombination rate of $c = 0.5$.

All observed fitness interactions in the final dataset were lethal, i.e. the epistasis was deleterious. There was substantial variability in the locations of the locus pairs across the genome. Most pairs of loci in the extracted dataset were on chromosome 2 (see [TABLE 7]). On each chromosome the sampled loci could be quite close to a telomere or the centromere. Distance in base pairs between the paired loci ranged from zero to almost three-quarters of the chromosome [FIGURE S1]. Due to the large parameter space relative to sample size, we used a generalized Akaike information criterion (GAIC) to select the best model (see in Methods: Flybase Meta-analysis: Regression Model to Test for Effects of Epistasis Class and Chromosome Number on Recombination). According to the GAIC the best model was the intercept-only model for the intermediate recombination mean, intermediate recombination precisions and odds of minimum recombination components, and for the odds of maximum recombination component included the epistasis class and chromosomal identity variables without interactions ([TABLE 8], [FIGURE S2] & [TABLE 9]).

	PDFI	PDNF	NDI
Chromosome 1	6	40	7
Chromosome 2	12	73	15
Chromosome 3	18	52	23

[TABLE 7] Caption: Numbers of locus pairs in each of the nine possible categories after filtering.

Model	Intermediate recombination mean	Intermediate recombination precision	Odds of minimum recombination	Odds of maximum recombination	GAIC
Null (1)	Intercept	Intercept	Intercept	Intercept	353.67
Full (2)	Chromosomal Identity, Epistasis Class, Interactions	Chromosomal Identity, Epistasis Class, Interactions	Chromosomal Identity, Epistasis Class, Interactions	Chromosomal Identity, Epistasis Class, Interactions	360.23
Selected (3)	Intercept	Intercept	Intercept	Epistasis	327.79

				Class, Chromosomal Identity	
4	Chromosomal Identity, Epistasis Class, Interactions	Intercept	Intercept	Intercept	360.50
5	Epistasis Class	Epistasis Class	Epistasis Class	Epistasis Class	348.11
6	Epistasis Class	Intercept	Intercept	Intercept	357.26
7	Chromosomal Identity	Chromosomal Identity	Chromosomal Identity	Chromosomal Identity	339.57
8	Chromosomal Identity	Intercept	Intercept	Intercept	356.99

Caption: [TABLE 8]: Comparison of models for the effect of chromosomal identity and epistasis class on the recombination rate between pairs of loci. Column headings refer to components defined in Methods (Flybase Meta-Analysis: Regression Model to Test for Effects of Epistasis Class and Chromosome Number on Recombination): 1) the mean of the intermediate recombination rates (i.e. recombination rates that were not 0 or 0.5); 2) the precision of the intermediate recombination rates 3) the relative probability that $c = 0$ rather than $0 < c < 0.5$; and 4) the relative probability that $c = 0.5$ rather than $0 < c < 0.5$. The names for them are: 1) Intermediate Recombination Mean, 2) Intermediate Recombination Precision, 3) Odds of Minimum Recombination, 4) Odds of Maximum Recombination. The final column shows the GAIC with the best fit (i.e. lowest value) in bold. The rows describe the variables included in each component of the model. An ‘Intercept’-only component means that none of the independent variables had an effect on that component of the recombination rate.

Component	Covariate	Estimate	Confidence Interval
<i>Intermediate Recombination Mean</i>	Intercept	0.407	(0.378,0.434)
<i>Intermediate Recombination Precision</i>	Intercept	0.614	(0.561,0.667)
<i>Odds of Minimum Recombination</i>	Intercept	0.108	(0.057, 0.159)
<i>Odds of Maximum Recombination</i>	Intercept	0.177	(0,0.374)
	Interaction	0.867	(0.126,1.609)
	Fitness	0.179	(0,0.473)
	Chrom2	4.774	(0.241,9.307)
	Chrom3	0.778	(0,1.660)

CAPTION: [TABLE 9] : Best model for the effect of chromosomal identity and epistasis class on the recombination rate between pairs of loci. This is model 4 from [TABLE 8] & [FIGURE S2]. Column headings are variables included in the model, its parameter estimate (as defined in

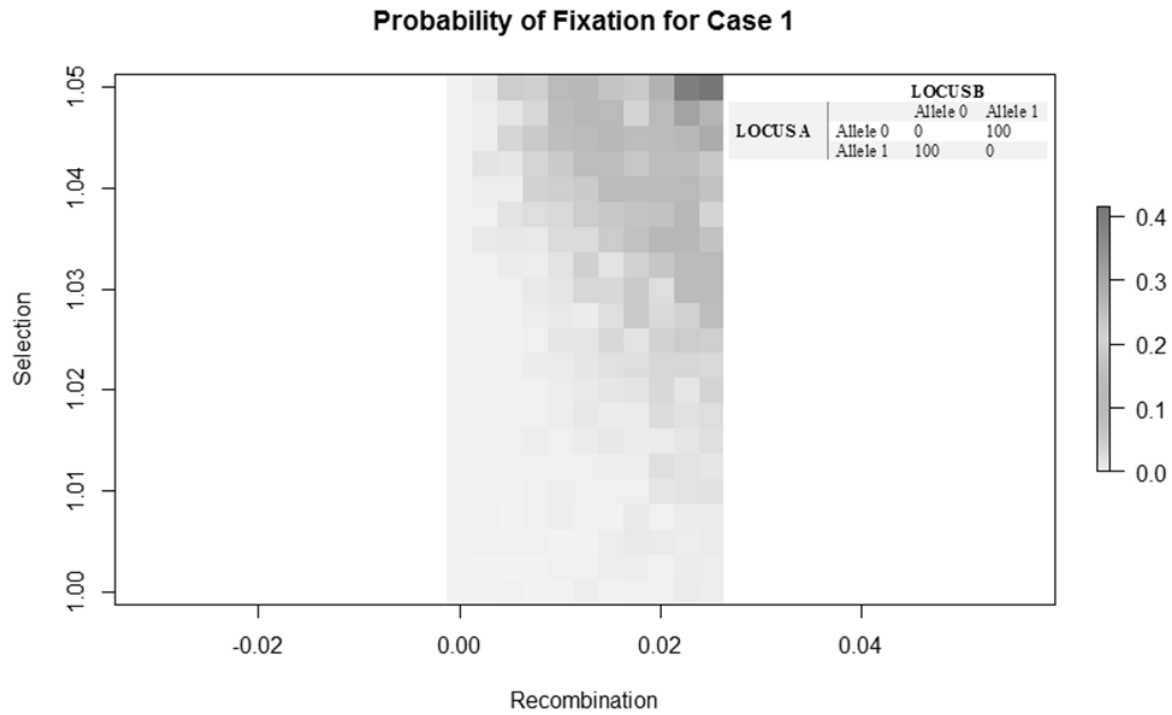
appendix II) and its standard error and 95% confidence intervals. Note that ‘component’ refers to any of the four parts of the response variable (recombination rate); these components are shown in the row headings and are defined in the Methods (Flybase Meta-Analysis: Regression Model to Test for Effects of Epistasis Class and Chromosome Number on Recombination): Intermediate Recombination Mean, Intermediate Recombination Precision, Odds of Minimum Recombination, Odds of Maximum Recombination. In contrast to ‘components’, ‘variable’ refers to any of the explanatory variables, that is the Intercept, Interaction, Fitness, Chrom2 and Chrom3 variables. The estimates given are transformed back to their original scale using the inverse of the link function and confidence intervals are calculated using the Delta Method.

The best fitting model (number 4 in [TABLE 8]) appeared to be a good fit (GAIC = 327.79); each component sub-model except the Odds of Maximum Recombination component only included the Intercept, and the Odds of Maximum Recombination component included the Fitness, Interaction, Chrom2 and Chrom3 variables but no interaction variables. In other words, a constant model was sufficient to explain the distribution of recombination rates of locus pairs except for locus pairs which recombined at the maximum rate. Within the Odds of Maximum Recombination component, only the Fitness variable was significant (transformed effect: 0.179, 95% CI: (0,0.473)). We considered an effect significant if it appeared in both the transformed scale and the original scale, i.e. it was robust to being transformed as discussed in Methods.

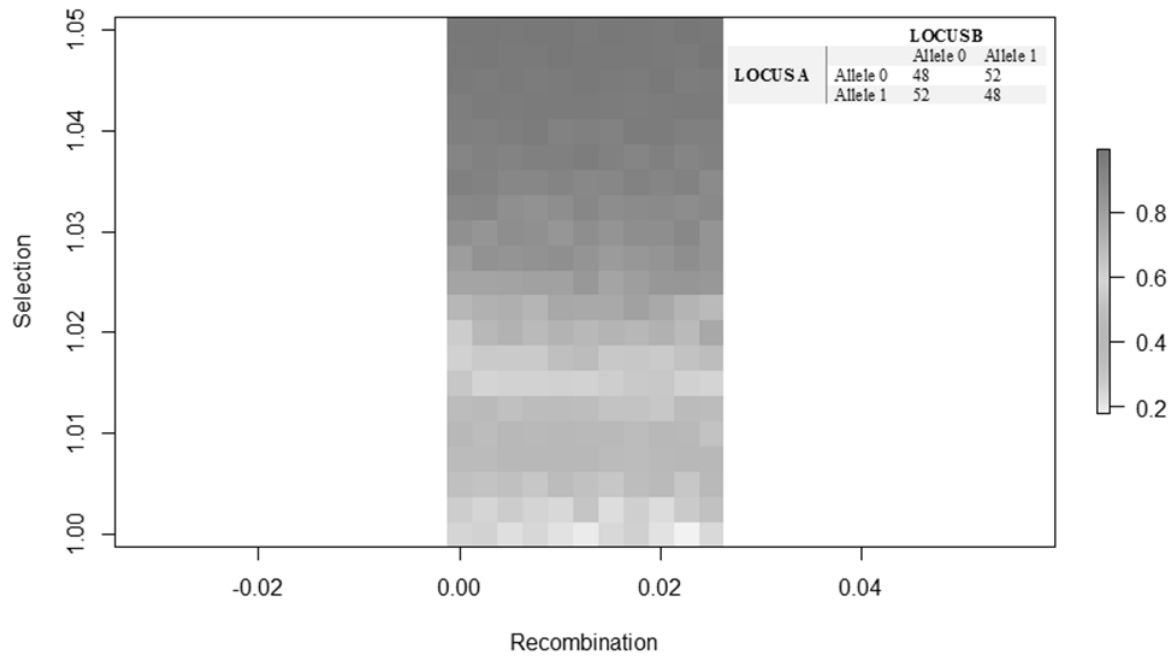
Simulations

In the simulations performed, the fixation probability of the beneficial genotype could display very different behaviours depending on the recombination rate, how close the initial population was to linkage equilibrium, and how many individuals bore the beneficial haplotype. [FIGURE 10] gives a graphical description of the behaviour with different initial population distributions, all with a population size of 200. In all cases the beneficial haplotype was (0,0). In each raster the x-axis is the recombination rate used for each simulation, the y-axis is the $1 + s$ in each simulation, and the colour is the probability of fixation of (0,0). The table in the right-hand corner gives the initial number of each haplotype in every simulation depicted. Higher rates of selection always increased the probability of the fixation of the beneficial haplotype, which is an intuitive and well-known result. When the population started close to linkage equilibrium [FIGURE 10, Case 2], recombination had no effect on the probability of fixation. When the population started far from linkage equilibrium, and had a small number of individuals of the beneficial genotype *and* a relatively small number of individuals bearing the non-beneficial coupling haplotype (1,1), higher

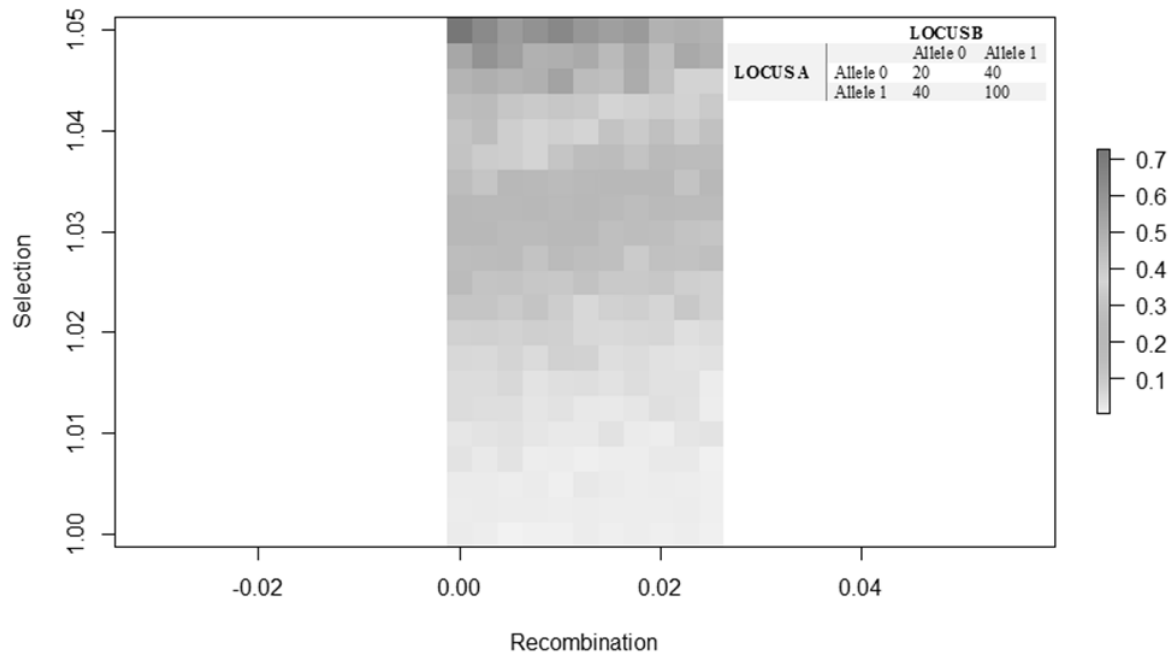
rates of recombination gave a higher probability of fixation [FIGURE 10, Cases 1 & 10]. When the initial population started far from linkage equilibrium, and had at least one individual bearing the beneficial haplotype and a large number of individuals bearing the non-beneficial coupling haplotype, higher recombination rates reduced the probability of fixation [FIGURE 10, Case 5 and 12]. These patterns were also reflected in the simulations with large recombination rates [FIGURE S3].



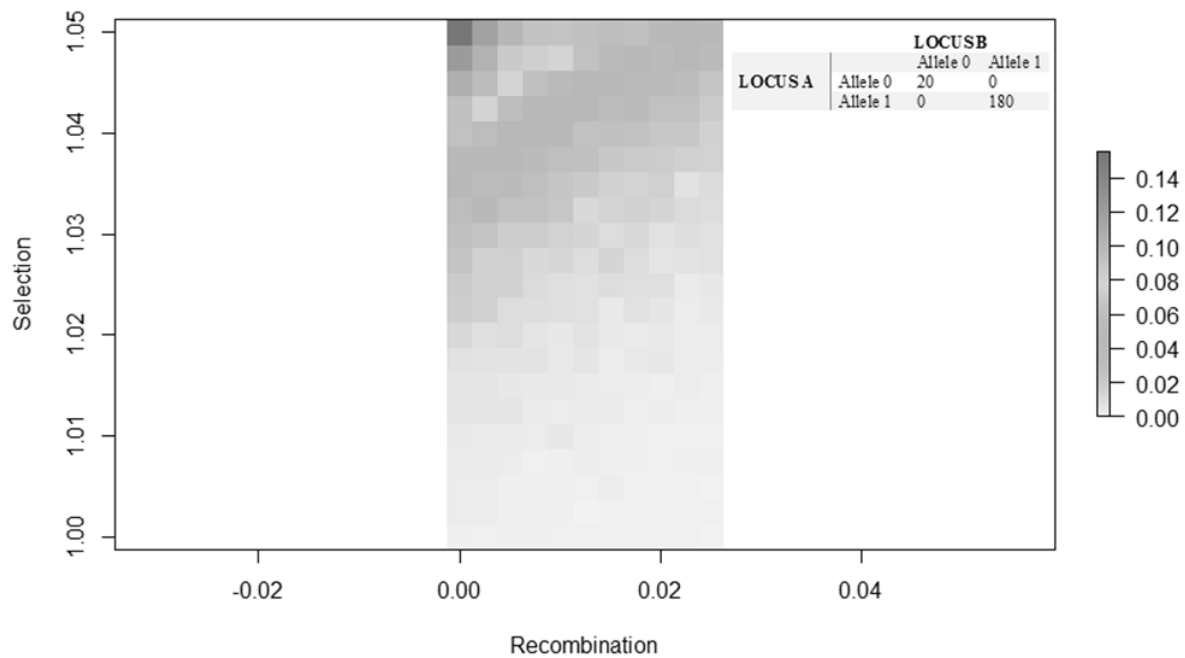
Probability of Fixation for Case 2

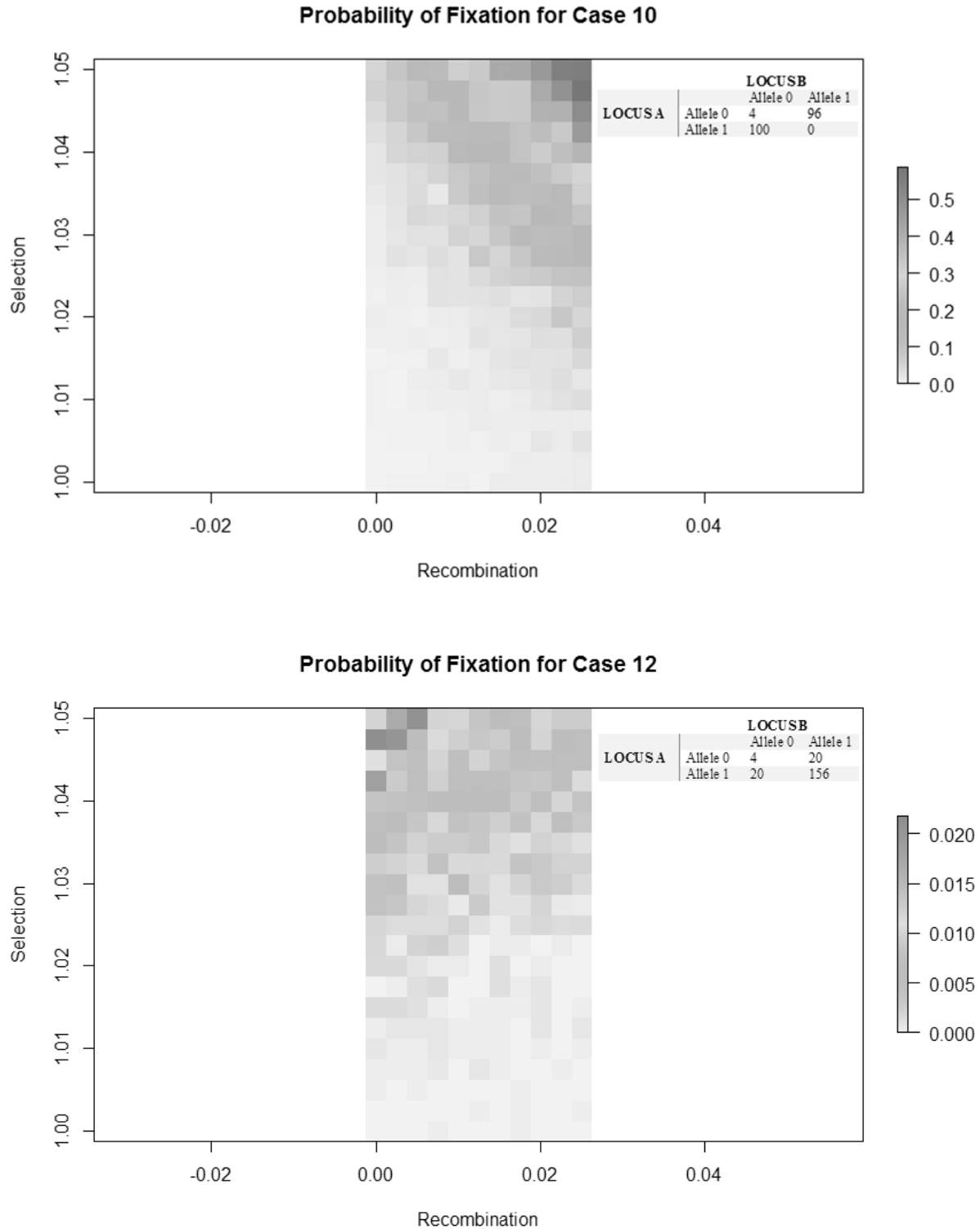


Probability of Fixation for Case 5



Probability of Fixation for Case 7





[FIGURE 10] Caption: Probability of fixation on two axes – recombination and selection. Each column of shading is centred on a recombination value on the horizontal axis. In all cases the beneficial haplotype was (0, 0). A darker colour indicates a higher probability of fixation (and thus

advantage of that recombination and selection combination). The table in the top right corner gives the initial population counts. All graphs are of populations with 200 individuals. The Supplement gives more plots for the results of the cases not shown here [FIGURE S3].

Discussion

The results of the meta-analysis indicate that the odds of a locus pair belonging to the PDFI epistasis class with a deleterious fitness effect recombining at the maximum rate over recombining at an intermediate rate are about one-fifth the corresponding odds of a pair belonging to the PDNF class (odds ratio: 0.179, 95% CI: 0 – 0.473, see [TABLE 9]). The confidence interval of this effect is very large, which is most likely due to a relatively small number of PDFI loci (36), but nonetheless seems to be strong evidence of an effect.

Two possible explanations for the reduced recombination rate are as follows. First, the non-lethal interactions are more likely to be maintained in regions of low recombination, possibly because if they occur in a region of high recombination, they would be quickly broken up. Second, regions with this kind of deleterious interaction experience selection against recombination, resulting in reduced recombination in that region. The two explanations are not mutually exclusive.

We did not observe a difference between epistasis classes when comparing locus pairs that did not recombine at all ($c = 0$) to pairs that recombined in the range (0,0.5) [TABLE 9]. A possible explanation for this is that there is no qualitative difference between zero recombination and an intermediate recombination rate when it comes to lethal interactions – all that matters is whether a locus pair recombines at the maximum rate or not. Another possibility is that we simply did not detect a difference where it existed, due to the relatively small number of locus pairs with zero recombination rate vs those recombining freely (19 vs 51).

There were a number of non-significant variables in the model which we included out of interest: the Chrom2 variable increased the odds of a locus pair recombining at the maximum rate over recombining at an intermediate rate (transformed value: 4.774); on the other hand, the Chrom3 and Interaction variables all decreased these odds (with transformed values of 0.778, 0.867 respectively).

It is important to note in the studies we curated and analysed, the selection pressures have to be quite strong to be detected. There is some evidence from mathematical modelling that other weaker

epistatic fitness effects could show different patterns but such effects are much more difficult to detect than lethality.

Our simulations indicate that there is a broad range of cases in which a high recombination rate is advantageous and another broad range of cases in which it is not. This is consistent with the lack of consensus in the previous literature. Whether recombination is advantageous or not is mainly controlled by whether there is a substantial number of the beneficial genotype in the initial population. If initially there is a large number of beneficial genotypes, recombination is less advantageous [Figure 10, “Case 3” & “Case 6”]. The latter is also determined by the relative proportion of repulsion genotypes. If repulsion genotypes are common recombination becomes advantageous [Figure 10, “Case 9”, “Case 11” & “Case 12”]. This means that the threshold between advantageous and disadvantageous occurs at linkage equilibrium, which agrees with prior mathematical literature (Mano 2013).

In our simulations the effect of beneficial epistatic interactions could favour increased rates of recombination in cases where the population was far from linkage equilibrium and there were not too many (relative to population size) of either the beneficial haplotype or the non-beneficial coupling haplotype. The effect of linkage disequilibrium is likely explained by the fact that, in the neutral case, the marginal distributions of the alleles evolve independently – that is, a population at linkage equilibrium is likely to stay in linkage equilibrium even as it moves to fixation in one of the types (Durrett and Mano 2013). Our simulations suggest that this remains (roughly) true even with epistatic selection. This is because for those of our simulations where the population was initially in linkage equilibrium, the fixation probabilities at each locus decoupled and became independent of one another at 50 generations, as would be expected of a population that remains in linkage equilibrium. This can clearly be seen in the fixation probability graphs of these cases, where there are horizontal rather than diagonal lines visible, indicating that fixation probability no longer depends on recombination rate. [Figure 10, “Case 2”, “Case 3”, “Case 4” & “Case 5”]

The reduced effectiveness of recombination in facilitating selection when there are several of the non-beneficial coupling haplotype is most likely because the beneficial haplotype and the non-beneficial coupling haplotype recombine to make the repulsion haplotypes, meaning that on average there will be fewer individuals bearing the beneficial haplotype than if there were no recombination.

Results from our simulations and meta-analysis were complementary, as genotypes from the meta-analysis were lethal but beneficial genotypes were studied in the simulations. There was broad

agreement between the two - *D. melanogaster* genotypes with lethal phenotypes displayed tight linkage, which is very similar what our simulations indicate would occur in a population with a large number of a beneficial genotype ((0,0)) or the genotype which opposes it ((1,1)) [FIGURE 10, “Case 7” & “Case 8]. In both cases non-beneficial or lethal genotypes very quickly reduce in frequency and are eliminated from the population with little to no recombination.

It is not clear how well the meta-analysis results will generalise to other species. Less closely related organisms are likely to display different recombination and selection patterns to *Drosophila melanogaster* (Dumont and Payseur 2008). In particular, mammals have so-called ‘recombination hotspots’ throughout the genome, with regions of size 1-2 kilobases displaying very high rates of recombination, but with very large regions with very little recombination. In humans the recombination rate within such a hotspot can be as high as 50 times the average of the rate over the entire chromosome (Kauppi *et al.* 2004). *Drosophila melanogaster* does not display such punctuated recombination rates (Comeron *et al.* 2012).

Population size reductions are a common occurrence, happening regularly as part of a cyclical process, or in response to some kind of threat such as habitat loss. A population would be more likely to persist if beneficial combinations of alleles lost to such a population size reduction were regenerated via recombination within a few generations. This would require relatively high rates of recombination between pairs of loci that interact to have an effect on fitness. Unfortunately, our meta-analysis of *D. melanogaster* data suggests that the opposite is in fact the case – that in fact the rates of recombination between pairs of loci with a fitness interaction are particularly low [RESULTS TABLE]. The simulations show conditions under which recombination would be an advantage, which depend upon linkage disequilibrium and allele proportions for loci with epistatic effects on fitness. As more of these loci are discovered in natural populations of non-model species, it will be useful to assess whether these species are likely to experience effects that favour or select against recombination between epistatically interacting loci. This will have important implications for evolution in general, and for conservation management, because, as stated above, population persistence would be more likely if beneficial combinations of alleles lost during a population size reduction were regenerated rapidly via recombination.

This study has taken some of the first steps in empirically analysing the relationship between epistatic selection and recombination, but much more needs to be done, because epistatic selection and recombination are both ubiquitous and their interaction is both important and poorly understood.

Experimental work on this type of problem may be difficult, given how difficult it is to identify epistatic genetic interactions (Chanda *et al.* 2007), but it may be possible to do a similar analysis on other model species as more data becomes available. Studies of speciation may also yield some important results as experimental evidence has shown that hybrid male sterility is often controlled by epistatic interactions (Palopoli *et al.* 2004, Chang *et al.* 2010).

Author Contributions

William Sherwin, John Murray, Mark Tanaka and Benjamin Goldys initiated this project as part of a UNSW mathematical biology initiative. Antony Bellanto carried out preliminary investigation of Flybase. Luis Cayetano performed preliminary simulations.

Acknowledgements

Some early programming work was performed by Julian Marchal and Camilo Cassel.

References

- Agrawal, N., Joshi, S., Kango, M., Saha, D., Mishra, A., & Sinha, P. (1995). Epithelial hyperplasia of imaginal discs induced by mutations in *Drosophila* tumor suppressor genes: Growth and pattern formation in genetic mosaics. *Developmental Biology*.
<http://doi.org/10.1006/dbio.1995.1155>
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <http://doi.org/10.1109/TAC.1974.1100705>
- Akong, K., McCartney, B. M., & Peifer, M. (2002). *Drosophila* APC2 and APC1 have overlapping roles in the larval brain despite their distinct intracellular localizations. *Developmental Biology*, 250(1), 71–90. <http://doi.org/10.1006/dbio.2002.0777>
- Akten, B., Suh, J., Genova, G. K., Roberts, M. a, & Jackson, F. R. (2002). RNA interference from a CAX trinucleotide repeat. *Genesis (New York, N.Y. □: 2000)*, 34(1–2), 156–9.
<http://doi.org/10.1002/gene.10126>
- Ashton-Beaucage, D., Udell, C. M., Gendron, P., Sahmi, M., Lefrançois, M., Baril, C., ... Therrien, M. (2014). A Functional Screen Reveals an Extensive Layer of Transcriptional and Splicing Control Underlying RAS/MAPK Signaling in *Drosophila*. *PLoS Biology*, 12(3).
<http://doi.org/10.1371/journal.pbio.1001809>
- Ashton-Beaucage, D., Udell, C. M., Lavoie, H., Baril, C., Lefrançois, M., Chagnon, P., ... Therrien, M. (2010). The Exon Junction Complex Controls the Splicing of *mapk* and Other Long Intron-

- Containing Transcripts in *Drosophila*. *Cell*, *143*(2), 251–262.
<http://doi.org/10.1016/j.cell.2010.09.014>
- Bajusz, I., Sipos, L., Györgypál, Z., Carrington, E. A., Jones, R. S., Gausz, J., & Gyurkovics, H. (2001). The Trithorax-mimic allele of Enhancer of zeste renders active domains of target genes accessible to Polycomb-group-dependent silencing in *Drosophila melanogaster*. *Genetics*, *159*(3), 1135–1150.
- Baonza, A., & Freeman, M. (2005). Control of cell proliferation in the *Drosophila* eye by notch signaling. *Developmental Cell*, *8*(4), 529–539. <http://doi.org/10.1016/j.devcel.2005.01.019>
- Barrios, N., González-Pérez, E., Hernández, R., & Campuzano, S. (2015). The Homeodomain Iroquois Proteins Control Cell Cycle Progression and Regulate the Size of Developmental Fields. *PLoS Genetics*, *11*(8), 1–20. <http://doi.org/10.1371/journal.pgen.1005463>
- Barton, N. H. (1995). A general model for the evolution of recombination. *Genetical Research*, *65*(2), 123–145. <http://doi.org/10.1017/S0016672300033140>
- Bartoszewski, S., Luschnig, S., Desjeux, I., Grosshans, J., & Nüsslein-Volhard, C. (2004). *Drosophila* p24 homologues eclair and baiser are necessary for the activity of the maternally expressed Tkv receptor during early embryogenesis. *Mechanisms of Development*, *121*(10), 1259–1273. <http://doi.org/10.1016/j.mod.2004.05.006>
- Baum, B., & Perrimon, N. (2001). Spatial control of the actin cytoskeleton in *Drosophila* epithelial cells. *Nature Cell Biology*, *3*(10), 883–890. <http://doi.org/10.1038/ncb1001-883>
- Baumgardner, J. R., Brewer, W. H., & Sanford, J. C. (2013). Can Synergistic Epistasis Halt Mutation Accumulation? Results from Numerical Simulation. *Biological Information: New Perspectives*, 312–337. http://doi.org/10.1142/9789814508728_0013
- Bejsovec, A., & Chao, A. T. (2012). crinkled reveals a new role for Wingless signaling in *Drosophila* denticle formation. *Development*, *139*(4), 690–698.
<http://doi.org/10.1242/dev.074013>
- Berardini, T., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L., ... Rhee, S. (2004). Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiology*, *135*(2), 1–11.
- Bergmann, A., Tugentman, M., Shilo, B. Z., & Steller, H. (2002). Regulation of cell number by MAPK-dependent control of apoptosis: A mechanism for trophic survival signaling. *Developmental Cell*, *2*(2), 159–170. [http://doi.org/10.1016/S1534-5807\(02\)00116-8](http://doi.org/10.1016/S1534-5807(02)00116-8)
- Bhat, K. M., van Beers, E. H., & Bhat, P. (2000). Sloppy paired acts as the downstream target of wingless in the *Drosophila* CNS and interaction between sloppy paired and gooseberry inhibits sloppy paired during neurogenesis. *Development (Cambridge, England)*, *127*(3), 655–65. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10631185>
- Bianco, A., Poukkula, M., Cliffe, A., Mathieu, J., Luque, C. M., Fulga, T. A., & Rørth, P. (2007). Two distinct modes of guidance signalling during collective migration of border cells. *Nature*, *448*(7151), 362–365. <http://doi.org/10.1038/nature05965>

- Biehs, B., Sturtevant, M. a, & Bier, E. (1998). Boundaries in the *Drosophila* wing imaginal disc organize vein-specific genetic programs. *Development (Cambridge, England)*, *125*(21), 4245–4257.
- Biryukova, I., & Heitzler, P. (2005). The *Drosophila* LIM-homeodomain protein Islet antagonizes proneural cell specification in the peripheral nervous system. *Developmental Biology*, *288*(2), 559–570. <http://doi.org/10.1016/j.ydbio.2005.09.033>
- Bossing, T., & Brand, A. H. (2002). Dephrin, a transmembrane ephrin with a unique structure, prevents interneuronal axons from exiting the *Drosophila* embryonic CNS. *Development (Cambridge, England)*, *129*(18), 4205–4218.
- Braver, G. (1953). Phenotypic detection of heterozygosity for wa in v/v. *Drosophila Information Service*, (27), 86.
- Bridges, C. B., & Brehme, K. S. (1944). *The Mutants of Drosophila melanogaster*. Carnegie Institution of Washington.
- Büchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G., & Dorn, R. (2000). Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in *Drosophila*. *Genetics*, *155*(1), 141–157.
- Buratovich, M. A., & Bryant, P. J. (1997). Enhancement of overgrowth by gene interactions in lethal(2)giant discs imaginal discs from *Drosophila melanogaster*. *Genetics*, *147*(2), 657–670.
- Busturia, a, Casanova, J., Sánchez-Herrero, E., González, R., & Morata, G. (1989). Genetic structure of the abd-A gene of *Drosophila*. *Development (Cambridge, England)*, *107*(3), 575–583. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2575516>
- Carrillo, R. A., ??zkan, E., Menon, K. P., Nagarkar-Jaiswal, S., Lee, P. T., Jeon, M., ... Zinn, K. (2015). Control of Synaptic Connectivity by a Network of *Drosophila* IgSF Cell Surface Proteins. *Cell*, *163*(7), 1770–1782. <http://doi.org/10.1016/j.cell.2015.11.022>
- Casal, J., Lawrence, P. A., & Struhl, G. (2006). Two separate molecular systems, Dachous/Fat and Starry night/Frizzled, act independently to confer planar cell polarity. *Development*, *133*(22), 4561–4572. <http://doi.org/10.1242/dev.02641>
- Cassidy, J. J., Jha, A. R., Posadas, D. M., Giri, R., Venken, K. J. T., Ji, J., ... Carthew, R. W. (2013). MiR-9a minimizes the phenotypic impact of genomic diversity by buffering a transcription factor. *Cell*, *155*(7), 1556–1567. <http://doi.org/10.1016/j.cell.2013.10.057>
- Chanda, P., Zhang, A., Brazeau, D., Sucheston, L., Freudenheim, J. L., Ambrosone, C., & Ramanathan, M. (2007). Information-Theoretic Metrics for Visualizing Gene-Environment Interactions. *The American Journal of Human Genetics*, *81*(5), 939–963. <http://doi.org/10.1086/521878>
- Chang, A. S., Bennett, S. M., & Noor, M. A. F. (2010). Epistasis among *Drosophila* persimilis factors conferring hybrid male sterility with *D. pseudoobscura bogotana*. *PLoS ONE*, *5*(10), 1–5. <http://doi.org/10.1371/journal.pone.0015377>

- Chanut, F., Woo, K., Pereira, S., Donohoe, T. J., Chang, S. Y., Lavery, T. R., ... Heberlein, U. (2002). Rough eye is a gain-of-function allele of *amos* that disrupts regulation of the proneural gene *atonal* during *Drosophila* retinal differentiation. *Genetics*, *160*(2), 623–635.
- Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F., & Plaza, S. (2006). Shavenbaby couples patterning to epidermal cell shape control. *PLoS Biology*, *4*(9), 1549–1561. <http://doi.org/10.1371/journal.pbio.0040290>
- Chapman, C. H., & Bingham, P. M. (1985). Evidence that the locus of a novel type of suppressor mutation regulates transcription of the white locus. *Drosophila Information Service*, (61), 48–50.
- Chase, B. A., & Kankel, D. R. (1987). A genetic analysis of glutamatergic function in *Drosophila*. *Journal of Neurobiology*, *18*(1), 15–41. <http://doi.org/10.1002/neu.480180104>
- CHATTERJEE, R. N., CHATTERJEE, P., KUTHE, S., ACHARYYA-ARI, M., & CHATTERJEE, R. (2015). Intersex (*ix*) mutations of *Drosophila melanogaster* cause nonrandom cell death in genital disc and can induce tumours in genitals in response to decapentaplegic (*dpp*^{disk}) mutations. *Journal of Genetics*, *94*(2), 207–220. <http://doi.org/10.1007/s12041-015-0503-3>
- Chen, F., Williams, A., & Rebay, I. (1999). Characterization of a new Ras pathway component EY2-3, an enhancer of Yan/ACT. In *Program and Abstracts. 40th Annual Drosophila Research Conference*.
- Chen, G. C., Gajowniczek, P., & Settleman, J. (2004). Rho-LIM kinase signaling regulates ecdysone-induced gene expression and morphogenesis during *Drosophila* metamorphosis. *Current Biology*, *14*(4), 309–313. [http://doi.org/10.1016/S0960-9822\(04\)00065-X](http://doi.org/10.1016/S0960-9822(04)00065-X)
- Christiansen, F. B., Otto, S. P., Bergman, A., & Feldman, M. W. (1998). Waiting with and without recombination: the time to production of a double mutant. *Theoretical Population Biology*, *53*(3), 199–215. <http://doi.org/10.1006/tpbi.1997.1358>
- CM, A., EK, C., JL, P., JM, G., LC, H., ML, S., ... CJ, L.-D. (2016). MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic acids research. Nucleic Acids Research*, *44*(D1), D1195-1201.
- Coelho, C. M. A. (2005). Growth and cell survival are unevenly impaired in *pixie* mutant wing discs. *Development*, *132*(24), 5411–5424. <http://doi.org/10.1242/dev.02148>
- Coleman, H. A., Labrador, J.-P., Chance, R. K., & Bashaw, G. J. (2010). The Adam family metalloprotease Kuzbanian regulates the cleavage of the roundabout receptor to control axon repulsion at the midline. *Development*, *137*(14), 2417–2426. <http://doi.org/10.1242/dev.047993>
- Curry, V. S. (1939). [new mutants report]. *Drosophila Information Service*, (12), 45–47.
- DG, H., YM, B., T, C., AE, E., D, F., K, F., ... M, W. (2013). ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic Acids Research*, *41*(D1), D854–D860.

- Dimitriadi, M., Sleight, J. N., Walker, A., Chang, H. C., Sen, A., Kalloo, G., ... Hart, A. C. (2010). Conserved genes act as modifiers of invertebrate SMN loss of function defects. *PLoS Genetics*, 6(10), 1–16. <http://doi.org/10.1371/journal.pgen.1001172>
- Doyle, H. J., & Bishop, J. M. (1993). Torso, a receptor tyrosine kinase required for embryonic pattern formation, shares substrates with the sevenless and EGF-R pathways in *Drosophila*. *Genes and Development*, 7(4), 633–646. <http://doi.org/10.1101/gad.7.4.633>
- Du, C., McGuffin, M. E., Dauwalder, B., Rabinow, L., & Mattox, W. (1998). Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*. *Molecular Cell*, 2(6), 741–50. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9885562>
- Dudek, S. M., Chiang, E. T., Camp, S. M., Guo, Y., Zhao, J., Brown, M. E., ... Garcia, J. G. N. (2010). Abl tyrosine kinase phosphorylates nonmuscle Myosin light chain kinase to regulate endothelial barrier function. *Molecular Biology of the Cell*, 21(22), 4042–4056. <http://doi.org/10.1091/mbc.E09>
- Duncan, J. E., Lytle, N. K., Zuniga, A., & Goldstein, L. S. B. (2013). The Microtubule Regulatory Protein Stathmin Is Required to Maintain the Integrity of Axonal Microtubules in *Drosophila*. *PLoS ONE*, 8(6), 1–20. <http://doi.org/10.1371/journal.pone.0068324>
- Emoto, K., Parrish, J. Z., Jan, L. Y., & Jan, Y. N. (2006). The tumour suppressor Hippo acts with the NDR kinases in dendritic tiling and maintenance. *Nature*, 443(7108), 210–213. <http://doi.org/10.1038/nature05090>
- Esser, M., Probst, S., & Baake, E. (2016). Partitioning, duality, and linkage disequilibria in the Moran model with recombination. *Journal of Mathematical Biology*, 73(1), 161–197. <http://doi.org/10.1007/s00285-015-0936-6>
- Estes, P. S., Jackson, T. C., Stimson, D. T., Sanyal, S., Kelly, L. E., & Ramaswami, M. (2003). Functional dissection of a eukaryotic dicistronic gene: Transgenic stonedB, but not stonedA, restores normal synaptic properties to *Drosophila* stoned mutants. *Genetics*, 165(1), 185–196.
- Fei, X., He, B., & Adler, P. N. (2002). The growth of *Drosophila* bristles and laterals is not restricted to the tip or base. *Journal of Cell Science*, 115, 3797–3806. <http://doi.org/10.1242/jcs.00065>
- Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, 78, 737–756.
- Fetting, J. L., Spencer, S. A., & Wolff, T. (2009). The cell adhesion molecules Echinoid and Friend of Echinoid coordinate cell adhesion and cell signaling to regulate the fidelity of ommatidial rotation in the *Drosophila* eye. *Development*, 136(19), 3323–3333. <http://doi.org/10.1242/dev.038422>
- Fox, A. N., & Zinn, K. (2005). The heparan sulfate proteoglycan Syndecan is an in vivo ligand for the *Drosophila* LAR receptor tyrosine phosphatase. *Current Biology*, 15(19), 1701–1711. <http://doi.org/10.1016/j.cub.2005.08.035>
- Freeman, A., Franciscovich, A., Bowers, M., Sandstrom, D. J., & Sanyal, S. (2011). NFAT regulates pre-synaptic development and activity-dependent plasticity in *Drosophila*. *Molecular and Cellular Neuroscience*, 46(2), 535–547. <http://doi.org/10.1016/j.mcn.2010.12.010>

- Fry, J. D. (2004). On the Rate and Linearity of Viability Declines in *Drosophila* Mutation-Accumulation Experiments: Genomic Mutation Rates and Synergistic Epistasis Revisited. *Genetics*, *166*(2), 797–806. <http://doi.org/10.1534/genetics.166.2.797>
- Gaengel, K. (2003). Egfr signaling regulates ommatidial rotation and cell motility in the *Drosophila* eye via MAPK/Pnt signaling and the Ras effector *Canoe/AF6*. *Development*, *130*(22), 5413–5423. <http://doi.org/10.1242/dev.00759>
- Gao, X., & Pan, D. (2001). Antagonize Insulin Signaling in Cell Growth. *Genes and Development*, *13*83–1392. <http://doi.org/10.1101/gad.901101.BP>
- Gause, M., Georgieva, S., & Georgiev, P. (1996). Phenotypic reversion of the gypsy-induced mutation *scD1* of *Drosophila melanogaster* by replicative transposition of a *sc* enhancer to the yellow gene and by mutations in the enhancer of yellow and zeste loci. *Molecular & General Genetics*: *MGG*, *253*(3), 370–6. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9003324>
- Genova, J. L., & Fehon, R. G. (2003). Neuroglian, Gliotactin, and the Na⁺/K⁺ ATPase are essential for septate junction function in *Drosophila*. *Journal of Cell Biology*, *161*(5), 979–989. <http://doi.org/10.1083/jcb.200212054>
- Gertler, F. B., Doctor, J. S., & Hoffmann, F. M. (1990). Genetic suppression of mutations in the *Drosophila* *abl* proto-oncogene homolog. *Science*, *248*(4957), 857–60. <http://doi.org/10.1126/science.2188361>
- Ghiglione, C. (2003). Mechanism of inhibition of the *Drosophila* and mammalian EGF receptors by the transmembrane protein Kekk1. *Development*, *130*(18), 4483–4493. <http://doi.org/10.1242/dev.00617>
- Gibert, J. M., Marcellini, S., David, J. R., Schlötterer, C., & Simpson, P. (2005). A major bristle QTL from a selected population of *Drosophila* uncovers the zinc-finger transcription factor *Poils-au-dos*, a repressor of *achaete-scute*. *Developmental Biology*, *288*(1), 194–205. <http://doi.org/10.1016/j.ydbio.2005.09.032>
- Gramates, L. S., Marygold, S. J., Santos, G. dos, Urbano, J.-M., Antonazzo, G., Matthews, B. B., ... R.P., Z. (2016). Flybase at 25: looking to the future. *Nucleic Acids Research*, *44*(October 2016). <http://doi.org/10.1093/nar/gkw1016>
- Green, M. M. (1952). Mutant isoalleles at the vermilion locus in *Drosophila melanogaster*. *Proc. N. A. S.*, *38*(X), 300–305.
- Gritzan, U., Hatini, V., & DiNardo, S. (1999). Mutual antagonism between signals secreted by adjacent wingless and engrailed cells leads to specification of complementary regions of the *Drosophila* parasegment. *Development (Cambridge, England)*, *126*(18), 4107–4115.
- Halsell, S. R., & Kiehart, D. P. (1998). Second-site noncomplementation identifies genomic regions required for *Drosophila* nonmuscle myosin function during morphogenesis. *Genetics*, *148*(4), 1845–4863.
- Hannah-alava, A. (1964). Interaction of non-allelic loci in expression of the extra-sexcomb phenotype in *Drosophila melanogaster*. *Z. Vererbungslehre*, *95*(1), 1–9.

- Hart, K., Klein, T., & Wilcox, M. (1993). A Minute encoding a ribosomal protein enhances wing morphogenesis mutants. *Mechanisms of Development*, 43(2–3), 101–110. [http://doi.org/10.1016/0925-4773\(93\)90028-V](http://doi.org/10.1016/0925-4773(93)90028-V)
- Hartfield, M., & Keightley, P. D. (2012). Current hypotheses for the evolution of sex and recombination. *Integrative Zoology*, 7, 192–209. <http://doi.org/10.1111/j.1749-4877.2012.00284.x>
- He, B., & Adler, P. N. (2002). The genetic control of arista lateral morphogenesis in drosophila. *Development Genes and Evolution*, 212(5), 218–229. <http://doi.org/10.1007/s00427-002-0229-0>
- Held, L. I. (1990). Arrangement of bristles as a function of bristle number on a leg segment in *Drosophila melanogaster*. *Roux Archives of Developmental Biology*, 199(1), 48–62. <http://doi.org/10.1007/BF01681532>
- Henikoff, S. (1979). Position effects and variegation enhancers in an autosomal region of *Drosophila melanogaster*. *Genetics*, 93(1), 105–115.
- Herbst, R., Carroll, P. M., Allard, J. D., Schilling, J., Raabe, T., & Simon, M. A. (1996). Daughter of sevenless is a substrate of the phosphotyrosine phosphatase corkscrew and functions during sevenless signaling. *Cell*, 85(6), 899–909. [http://doi.org/10.1016/S0092-8674\(00\)81273-8](http://doi.org/10.1016/S0092-8674(00)81273-8)
- Hilfiker, a, Amrein, H., Dübendorfer, A., Schneiter, R., & Nöthiger, R. (1995). The gene *virilizer* is required for female-specific splicing controlled by Sxl, the master gene for sexual development in *Drosophila*. *Development (Cambridge, England)*, 121(12), 4017–26. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8575302>
- Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, 8(2), 269–294. <https://doi.org/10.1017/S001667230800949X>
- Hiraizumi, Y., Martin, D. W., & Eckstrand, I. A. (1980). A modified model of segregation structure in *Drosophila melanogaster*. *Genetics*, 95(3), 693–706.
- House, V. L. (1961). Mutant effects in multiple heterozygotes of recessive venation mutants in *Drosophila melanogaster*. *Records of the Genetics Society of America*, 30.
- Hsu, J. C., Wu, C. W., Perrimon, N., & Bai, J. M. (1998). *echinoid*, a negative regulator of the *Drosophila* EGF receptor signaling pathway. In *Program and Abstracts. 39th Annual Drosophila Research Conference*.
- Hummel, T., Schimmelpfeng, K., & Klämbt, C. (1999). Commissure formation in the embryonic CNS of *Drosophila*. I. Identification of the required gene functions. *Developmental Biology*, 209(2), 381–398. <http://doi.org/10.1006/dbio.1999.9235>
- Ilius, M., Wolf, R., & Heisenberg, M. (2007). The central complex of *Drosophila melanogaster* is involved in flight control: Studies on mutants and mosaics of the gene *Ellipsoid body open*. *Journal of Neurogenetics*, 21(4), 321–338. <http://doi.org/10.1080/01677060701693503>
- Jaffe, A. B., & Jongens, T. A. (2001). Structure-specific abnormalities associated with mutations in a DNA replication accessory factor in *Drosophila*. *Developmental Biology*, 230(2), 161–176. <http://doi.org/10.1006/dbio.2000.0117>

- Jauffred, B., Llense, F., Sommer, B., Wang, Z., Martin, C., & Bellaiche, Y. (2013). Regulation of centrosome movements by Numb and the Collapsin Response Mediator Protein during *Drosophila* sensory progenitor asymmetric division. *Development*, *140*(13), 2657–2668. <http://doi.org/10.1242/dev.087338>
- Jaźwińska, a, Rushlow, C., & Roth, S. (1999). The role of brinker in mediating the graded response to Dpp in early *Drosophila* embryos. *Development (Cambridge, England)*, *126*(15), 3323–3334.
- Jenkins, P. A., Fearnhead, P., & Song, Y. S. (2015). Tractable diffusion and coalescent processes for weakly correlated loci. *Electronic Journal of Probability*, *20*, 1–26. <http://doi.org/10.1214/EJP.v20-3564>
- Johansen, K. A., Green, R. B., Iwaki, D. D., Hernandez, J. B., & Lengyel, J. A. (2003). The Drm-Bowl-Lin relief-of-repression hierarchy controls fore- and hindgut patterning and morphogenesis. *Mechanisms of Development*, *120*(10), 1139–1151. <http://doi.org/10.1016/j.mod.2003.08.001>
- Johnston, D. S., Road, T. C., & Kingdom, U. (2002). *Drosophila* Nicastrin Is Essential for the Intramembranous Cleavage of Notch. *Developmental Cell*, *2*, 79–89.
- Joly, W., Chartier, A., Rojas-Rios, P., Busseau, I., & Simonelig, M. (2013). The CCR4 deadenylase acts with Nanos and Pumilio in the fine-tuning of Mei-P26 expression to promote germline stem cell self-renewal. *Stem Cell Reports*, *1*(5), 411–424. <http://doi.org/10.1016/j.stemcr.2013.09.007>
- Jouliia, L., Deutsch, J., Bourbon, H. M., & Cribbs, D. L. (2006). The specification of a highly derived arthropod appendage, the *Drosophila* labial palps, requires the joint action of selectors and signaling pathways. *Development Genes and Evolution*, *216*(7–8), 431–442. <http://doi.org/10.1007/s00427-006-0086-3>
- Joyce, E. F., & McKim, K. S. (2009). *Drosophila* PCH2 is required for a pachytene checkpoint that monitors double-strand-break-independent events leading to meiotic crossover formation. *Genetics*, *181*(1), 39–51. <http://doi.org/10.1534/genetics.108.093112>
- Judd, B. H. (1995). Mutations of zeste that mediate transvection are recessive enhancers of position-effect variegation in *Drosophila melanogaster*. *Genetics*, *141*(1), 245–253.
- Kaplan, N. A., Colosimo, P. F., Liu, X., & Tolwinski, N. S. (2011). Complex interactions between GSK3 and aPKC in *drosophila* embryonic epithelial morphogenesis. *PLoS ONE*, *6*(4). <http://doi.org/10.1371/journal.pone.0018616>
- Kauppi, L., Jeffreys, A. J., & Keeney, S. (2004). Where the crossovers are: recombination distributions in mammals. *Genetics*, *5*(June), 413–424. <http://doi.org/10.1038/nrg1346>
- Kelemen-Valkony, I., Kiss, M., Csiha, J., Kiss, A., Bircher, U., Szidonya, J., ... Mink, M. (2012). *Drosophila* basement membrane collagen col4a1 mutations cause severe myopathy. *Matrix Biology*, *31*(1), 29–37. <http://doi.org/10.1016/j.matbio.2011.09.004>
- Khalsa, O., Yoon, J. W., Torres-Schumann, S., & Wharton, K. a. (1998). TGF-beta/BMP superfamily members, Gbb-60A and Dpp, cooperate to provide pattern information and

establish cell identity in the *Drosophila* wing. *Development (Cambridge, England)*, *125*(14), 2723–2734.

- Klarsfeld, A. (2004). Novel Features of Cryptochrome-Mediated Photoreception in the Brain Circadian Clock of *Drosophila*. *Journal of Neuroscience*, *24*(6), 1468–1477. <http://doi.org/10.1523/JNEUROSCI.3661-03.2004>
- Kolodkin, a L., Pickup, a T., Lin, D. M., Goodman, C. S., & Banerjee, U. (1994). Characterization of Star and its interactions with sevenless and EGF receptor during photoreceptor cell development in *Drosophila*. *Development (Cambridge, England)*, *120*(7), 1731–1745.
- Kondrashov, A. S., & Yampolsky, L. Y. (1996). Evolution of amphimixis and recombination under fluctuating selection in one and many traits. *Genet. Res.*, *68*(2), 165–173. <http://doi.org/10.1017/S0016672300034054>
- Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, *336*(6198), 435–440. <http://doi.org/10.1038/336435a0>
- Kuzin, B. A., Nikitina, E. A., Cherezov, R. O., Vorontsova, J. E., Slezinger, M. S., Zatsepina, O. G., ... Savvateeva-Popova, E. V. (2014). Combination of hypomorphic mutations of the drosophila homologues of aryl hydrocarbon receptor and nucleosome assembly protein family genes disrupts morphogenesis, memory and detoxification. *PLoS ONE*, *9*(4). <http://doi.org/10.1371/journal.pone.0094975>
- Kuzin, B., Doszhanov, K., & Mazo, A. (1997). Interaction between spineless-aristopedia gene and genes from Antennapedia and bithorax complexes of *Drosophila melanogaster*. *International Journal of Developmental Biology*, *41*(6), 867–875.
- Labrador, J. P., O'Keefe, D., Yoshikawa, S., McKinnon, R. D., Thomas, J. B., & Bashaw, G. J. (2005). The homeobox transcription factor even-skipped regulates netrin-receptor expression to control dorsal motor-axon projections in *Drosophila*. *Current Biology*, *15*(15), 1413–1419. <http://doi.org/10.1016/j.cub.2005.06.058>
- Lake, C. M., Nielsen, R. J., & Hawley, R. S. (2011). The drosophila zinc finger protein trade embargo is required for double strand break formation in meiosis. *PLoS Genetics*, *7*(2). <http://doi.org/10.1371/journal.pgen.1002005>
- Lamb, R. S., Ward, R. E., Schweizer, L., & Fehon, R. G. (1998). *Drosophila* coracle, a member of the protein 4.1 superfamily, has essential structural functions in the septate junctions and developmental functions in embryonic and adult epithelial cells. *Molecular Biology of the Cell*, *9*(12), 3505–3519. <http://doi.org/10.1091/MBC.9.12.3505>
- Lee, H., & Adler, P. N. (2002). The function of the frizzled pathway in the *Drosophila* wing is dependent on inturned and fuzzy. *Genetics*, *160*(4), 1535–1547.
- Lee, H., & Adler, P. N. (2004). The grainy head transcription factor is essential for the function of the frizzled pathway in the *Drosophila* wing. *Mechanisms of Development*, *121*(1), 37–49. <http://doi.org/10.1016/j.mod.2003.11.002>
- Lee, J. D., Amanai, K., Shearn, A., & Treisman, J. E. (2002). The ubiquitin ligase Hyperplastic discs negatively regulates hedgehog and decapentaplegic expression by independent mechanisms. *Development*, *129*(24), 5697–5706. <http://doi.org/10.1242/dev.00159>

- Lewis, E. B. (1945). The Relation of Repeats to Position Effect in *Drosophila Melanogaster*. *Genetics*, 30(2), 137–166.
- Li, J., & Li, W. X. (2003). *Drosophila* gain-of-function mutant RTK Torso triggers ectopic Dpp and STAT signaling. *Genetics*, 164(1), 247–258.
- Li, W., Ohlmeyer, J. T., Lane, M. E., & Kalderon, D. (1995). Function of protein kinase A in hedgehog signal transduction and *Drosophila* imaginal disc development. *Cell*, 80(4), 553–562. [http://doi.org/10.1016/0092-8674\(95\)90509-X](http://doi.org/10.1016/0092-8674(95)90509-X)
- Liang, L., Diehl-Jones, W., & Lasko, P. (1994). Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. *Development (Cambridge, England)*, 120(5), 1201–1211.
- Lin, H. V, Rogulja, A., & Cadigan, K. M. (2004). Wingless eliminates ommatidia from the edge of the developing eye through activation of apoptosis. *Development (Cambridge, England)*, 131(10), 2409–18. <http://doi.org/10.1242/dev.01104>
- Lin, T.-Y., Huang, C.-H., Kao, H.-H., Liou, G.-G., Yeh, S.-R., Cheng, C.-M., ... Juang, J.-L. (2009). Abi plays an opposing role to Abl in *Drosophila* axonogenesis and synaptogenesis. *Development (Cambridge, England)*, 136(18), 3099–107. <http://doi.org/10.1242/dev.033324>
- Lindsley, D. L., & Grell, E. H. (1968). *Genetic Variations of Drosophila melanogaster*. Carnegie Institution of Washington.
- Lindsley, D. L., & Zimm, G. (1992). *The genome of Drosophila melanogaster* (1st ed.). San Diego: Academic Press.
- Lloyd, V. K., Sinclair, D. A., Alperyn, M., & Grigliatti, T. A. (2002). *Enhancer of garnet /δAP-3* is a cryptic allele of the *white* gene and identifies the intracellular transport system for the white protein. *Genome*, 45(2), 296–312. <http://doi.org/10.1139/g01-139>
- Lu, B. Y., Emtage, P. C., Duyf, B. J., Hilliker, a J., & Eissenberg, J. C. (2000). Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in *Drosophila*. *Genetics*, 155(2), 699–708.
- Ma, C., Liu, H., Zhou, Y., & Moses, K. (1996). Identification and characterization of autosomal genes that interact with glass in the developing *Drosophila* eye. *Genetics*, 142(4), 1199–1213.
- Mace, K., & Tugores, A. (2004). The product of the split ends gene is required for the maintenance of positional information during *Drosophila* development. *BMC Developmental Biology*, 4, 1–19. <http://doi.org/10.1186/1471-213X-4-15>
- Mach, J. M., & Lehmann, R. (1997). An egalitarian-BicaudalD complex is essential for oocyte specification and axis determination in *Drosophila*. *Genes and Development*, 11(4), 423–435. <http://doi.org/10.1101/gad.11.4.423>
- Magie, C. R., Meyer, M. R., Gorsuch, M. S., & Parkhurst, S. M. (1999). Mutations in the Rho1 small GTPase disrupt morphogenesis and segmentation during early *Drosophila* development. *Development (Cambridge, England)*, 126(23), 5353–64. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10556060%5Cnhttp://journals.bmn.com/medline/search/results?uid=MDLN.20025934>

- Mano, S. (2013). Duality between the two-locus wright-fisher diffusion model and the ancestral process with recombination. *Journal of Applied Probability*, 50(1), 256–271. <http://doi.org/10.1239/jap/1363784437>
- Maor, G., Rencus-Lazar, S., Filocamo, M., Steller, H., Segal, D., & Horowitz, M. (2013). Unfolded protein response in Gaucher disease: From human to *Drosophila*. *Orphanet Journal of Rare Diseases*, 8(1), 1. <http://doi.org/10.1186/1750-1172-8-140>
- Melnikova, L., Kulikov, A., & Georgiev, P. (1996). Interactions between cut wing mutations and mutations in *zeste*, and the enhancer of yellow and Polycomb group genes of *Drosophila melanogaster*. *Molecular and General Genetics*, 252(3), 230–236. <http://doi.org/10.1007/s004380050224>
- Merianda, T. T., Botta, V., & Bhat, K. M. (2005). Patched regulation of axon guidance is by specifying neural identity in the *Drosophila* nerve cord. *Development Genes and Evolution*, 215(6), 285–296. <http://doi.org/10.1007/s00427-005-0475-z>
- Meyer, H. U. (1952). [new mutants report]. *Drosophila Information Service*, (26), 111.
- Moore, J. H., & Williams, S. M. (2015). *Epistasis: Methods and Protocols*. Springer. <http://doi.org/10.1007/978-1-4939-2155-3>
- Morris, J. Z. (2005). *twin*, a CCR4 homolog, regulates cyclin poly(A) tail length to permit *Drosophila* oogenesis. *Development*, 132(6), 1165–1174. <http://doi.org/10.1242/dev.01672>
- Mount, S. M., Green, M. M., & Rubin, G. M. (1988). Partial revertants of the transposable element-associated suppressible allele *white-apricot* in *Drosophila melanogaster*: structures and responsiveness to genetic modifiers. *Genetics*, 118(2), 221–234.
- Muller, H. J. (1932). Some Genetic Aspects of Sex. *The American Naturalist*, 66(703), 118–138. <http://doi.org/10.1086/285850>
- Nagel, A. C., Szawinski, J., Zimmermann, M., & Preiss, A. (2016). *Drosophila* cyclin G is a regulator of the Notch signalling pathway during wing development. *PLoS ONE*, 11(3), 1–17. <http://doi.org/10.1371/journal.pone.0151477>
- Nöthiger, R., Leuthold, M., Andersen, N., Gerschwiler, P., Grüter, A., Keller, W., ... Schmid, H. (1987). Genetic and developmental analysis of the sex-determining gene ‘double sex’ (*dsx*) of *Drosophila melanogaster*. *Genetical Research*, 50(2), 113–123. <http://doi.org/10.1017/S001667230002351X>
- O’Hare, K. (1991). On the identity of the *white-eosin* and *white-cherry*. *Drosophila Information Service*, (70), 160–161.
- Ohlmeyer, J. T., & Schüpbach, T. (2003). *Encore* facilitates SCF-Ubiquitin-proteasome-dependent proteolysis during *Drosophila* oogenesis. *Development (Cambridge, England)*, 130(25), 6339–6349. <http://doi.org/10.1242/dev.00855>
- Oikemus, S. R., Queiroz-Machado, J., Lai, K., McGinnis, N., Sunkel, C., & Brodsky, M. H. (2006). Epigenetic telomere protection by *Drosophila* DNA damage response pathways. *PLoS Genetics*, 2(5), 693–706. <http://doi.org/10.1371/journal.pgen.0020071>

- Oldham, S., Stocker, H., Laffargue, M., Wittwer, F., Wymann, M., & Hafen, E. (2002). The *Drosophila* insulin/IGF receptor controls growth and size by modulating PtdInsP(3) levels. *Development (Cambridge, England)*, *129*, 4103–4109.
- Oliver, B., & Pauli, D. (1998). Suppression of distinct ova phenotypes in the *Drosophila* female germline by maleless- and sex-lethal. *Developmental Genetics*, *23*(4), 335–346. [http://doi.org/10.1002/\(SICI\)1520-6408\(1998\)23:4<335::AID-DVG8>3.0.CO;2-M](http://doi.org/10.1002/(SICI)1520-6408(1998)23:4<335::AID-DVG8>3.0.CO;2-M)
- Ospina, R., & Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis*, *56*(6), 1609–1623. <http://doi.org/10.1016/j.csda.2011.10.005>
- Otto, S. P., & Barton, N. H. (2001). Selection for recombination in small populations. *Evolution*, *55*(10), 1921–1931. <http://doi.org/10.2307/2680441>
- Otto, S. P., & Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nature Reviews Genetics*, *3*(4), 252–261. <http://doi.org/10.1038/nrg761>
- Otto, S. P., & Michalakis, Y. (1998). The evolution of recombination in changing environments. *Trends in Ecology and Evolution*, *13*(4), 145–151. [http://doi.org/10.1016/S0169-5347\(97\)01260-3](http://doi.org/10.1016/S0169-5347(97)01260-3)
- Otto, S. P. (2009). The Evolutionary Enigma of Sex. *The American Naturalist*, *174*(S1), S1–S14. <http://doi.org/10.1086/599084>
- Palacios, I. M., Gatfield, D., Johnston, D. S., & Izaurralde, E. (2004). An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature*, *427*(February), 753–757. <http://doi.org/10.1038/nature02305.1>
- Palopoli, M. F., & Wu, C. I. (1994). Genetics of hybrid male sterility between *Drosophila* sibling species: A complex web of epistasis is revealed in interspecific studies. *Genetics*, *138*(2), 329–341. [http://doi.org/10.1016/0168-9525\(95\)90591-X](http://doi.org/10.1016/0168-9525(95)90591-X)
- Parker, D. S., Jemison, J., & Cadigan, K. M. (2002). Pygopus, a nuclear PHD-finger protein required for Wingless signaling in *Drosophila*. *Development (Cambridge, England)*, *129*(11), 2565–76. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12015286>
- Penton, A., Chen, Y., Staehling-Hampton, K., Wrana, J. L., Attisano, L., Szidonya, J., Hoffmann, F. M. (1994). Identification of two bone morphogenetic protein type I receptors in *Drosophila* and evidence that Brk25D is a decapentaplegic receptor. *Cell*, *78*(2), 239–250. [http://doi.org/10.1016/0092-8674\(94\)90294-1](http://doi.org/10.1016/0092-8674(94)90294-1)
- Peschel, N., Veleri, S., & Stanewsky, R. (2006). Veela defines a molecular link between Cryptochrome and Timeless in the light-input pathway to *Drosophila*'s circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(46), 17313–17318. <http://doi.org/10.1073/pnas.0606675103>
- Peters, A. D., & Lively, C. M. (1999). The Red Queen and Fluctuating Epistasis: A Population Genetic Analysis of Antagonistic Coevolution. *The American Naturalist*, *154*(4), 393–405. <http://doi.org/10.1086/303247>

- Petrovich, T. Z., Merakovsky, J., & Kelly, L. E. (1993). A genetic analysis of the stoned locus and its interaction with dunce, shibire and suppressor of stoned variants of *Drosophila melanogaster*. *Genetics*, *133*(4), 955–965.
- Quinn, L. M., Herr, A., McGarry, T. J., & Richardson, H. (2001). The *Drosophila* Geminin homolog: Roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. *Genes and Development*, *15*(20), 2741–2754. <http://doi.org/10.1101/gad.916201>
- Rangone, H., Wegel, E., Gatt, M. K., Yeung, E., Flowers, A., Debski, J., ... Glover, D. M. (2011). Suppression of scant identifies endos as a substrate of greatwall kinase and a negative regulator of protein phosphatase 2a in mitosis. *PLoS Genetics*, *7*(8). <http://doi.org/10.1371/journal.pgen.1002225>
- Rawls, A. S., Schultz, S. A., Mitra, R. D., & Wolff, T. (2007). Bedraggled, a putative transporter, influences the tissue polarity complex during the R3/R4 fate decision in the *Drosophila* eye. *Genetics*, *177*(1), 313–328. <http://doi.org/10.1534/genetics.107.075945>
- Ren, N., He, B., Stone, D., Kirakodu, S., & Adler, P. N. (2006). The shavenoid gene of *Drosophila* encodes a novel actin cytoskeleton interacting protein that promotes wing hair morphogenesis. *Genetics*, *172*(3), 1643–1653. <http://doi.org/10.1534/genetics.105.051433>
- Riede, I. (2001). Gene combinations inducing neoplasms in *Drosophila*. *Drosophila Information Service*, (84), 103–114.
- Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, *54*(3), 507–554.
- Rogge, R., Green, P. J., Urano, J., Horn-Saban, S., Mlodzik, M., Shilo, B. Z., ... Banerjee, U. (1995). The role of yan in mediating the choice between cell division and differentiation. *Development (Cambridge, England)*, *121*(12), 3947–3958.
- Rosa, J. M., Camacho, S., & García-Dorado, A. (2005). A measure of the within-chromosome synergistic epistasis for *Drosophila* viability. *Journal of Evolutionary Biology*, *18*(4), 1130–1137. <http://doi.org/10.1111/j.1420-9101.2005.00892.x>
- Rutledge, B. J., Mortin, M. A., Schwarz, E., Thierry-Mieg, D., & Meselson, M. (1988). Genetic interactions of modifier genes and modifiable alleles in *Drosophila melanogaster*. *Genetics*, *119*(2), 391–397.
- Rutschmann, S., Kilinc, A., & Ferrandon, D. (2002). Cutting Edge: The Toll Pathway Is Required for Resistance to Gram-Positive Bacterial Infections in *Drosophila*. *The Journal of Immunology*, *168*(4), 1542–1546. <http://doi.org/10.4049/jimmunol.168.4.1542>
- Sambandan, D., Yamamoto, A., Fanara, J. J., Mackay, T. F. C., & Anholt, R. R. H. (2006). Dynamic genetic interactions determine odor-guided behavior in *Drosophila melanogaster*. *Genetics*, *174*(3), 1349–1363. <http://doi.org/10.1534/genetics.106.060574>
- Sasaki, A., & Iwasa, Y. (1987). Optimal recombination rate in fluctuating environments. *Genetics*, *115*(2), 377–388.

- Schimmelpfeng, K., Gögel, S., & Klämbt, C. (2001). The function of leak and kuzbanian during growth cone and cell migration. *Mechanisms of Development*, *106*(1–2), 25–36. [http://doi.org/10.1016/S0925-4773\(01\)00402-6](http://doi.org/10.1016/S0925-4773(01)00402-6)
- Schultz, J. (1929). The Minute reaction in the development of *Drosophila melanogaster*. *Genetics*, *14*, 366–419.
- Schulze, S., Sinclair, D. A. R., Silva, E., Fitzpatrick, K. A., Singh, M., Lloyd, V. K., ... Honda, B. M. (2001). Essential genes in proximal 3L heterochromatin of *Drosophila melanogaster*. *Molecular and General Genetics*, *264*(6), 782–789. <http://doi.org/10.1007/s004380000367>
- Siddall, N. a, Behan, K. J., Crew, J. R., Cheung, T. L., Fair, J. a, Batterham, P., & Pollock, J. A. (2003). Mutations in lozenge and D-Pax2 invoke ectopic patterned cell death in the developing *Drosophila* eye using distinct mechanisms. *Development Genes and Evolution*, *213*(3), 107–19. <http://doi.org/10.1007/s00427-003-0295-y>
- Simpson, J. H., Kidd, T., Bland, K. S., & Goodman, C. S. (2000). Short-range and long-range guidance by Slit and its robo receptors: Robo and Robo2 play distinct roles in midline guidance. *Neuron*, *28*(3), 753–766. [http://doi.org/10.1016/S0896-6273\(00\)00151-3](http://doi.org/10.1016/S0896-6273(00)00151-3)
- Sinclair, D. A. R., Schulze, S., Silva, E., Fitzpatrick, K. A., & Honda, B. M. (2000). Essential genes in autosomal heterochromatin of *Drosophila melanogaster*. *Genetica*, *109*(9), 9–18. <http://doi.org/10.1007/s004380000367>
- Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160455. <http://doi.org/10.1098/rstb.2016.0455>
- Stern, M., Kreber, R., & Ganetzky, B. (1990). Dosage effects of a *Drosophila* sodium channel gene on behavior and axonal excitability. *Genetics*, *124*(1), 133–143.
- Stevens, A., & Jacobs, J. R. (2002). Integrins regulate responsiveness to slit repellent signals. *Journal of Neuroscience*, *22*(11), 4448–4455. <http://doi.org/20026413>
- Swanson, L. E., Yu, M., Nelson, K. S., Laprise, P., Tepass, U., & Beitel, G. J. (2009). *Drosophila* convoluted/dALS is an essential gene required for tracheal tube morphogenesis and apical matrix organization. *Genetics*, *181*(4), 1281–1290. <http://doi.org/10.1534/genetics.108.099531>
- Taylor, J., Abramova, N., Charlton, J., & Adler, P. N. (1998). Van Gogh: A new *Drosophila* tissue polarity gene. *Genetics*, *150*(1), 199–210.
- Terriente-Félix, A., & de Celis, J. F. (2009). Osa, a subunit of the BAP chromatin-remodelling complex, participates in the regulation of gene expression in response to EGFR signalling in the *Drosophila* wing. *Developmental Biology*, *329*(2), 350–361. <http://doi.org/10.1016/j.ydbio.2009.03.010>
- Thompson, P. E. (1963). [new mutants report]. *Drosophila Information Service*, (38), 28.
- Tseng, A. S. K., Tapon, N., Kanda, H., Cigizoglu, S., Edelmann, L., Pellock, B., ... Hariharan, I. K. (2007). Capicua Regulates Cell Proliferation Downstream of the Receptor Tyrosine

Kinase/Ras Signaling Pathway. *Current Biology*, 17(8), 728–733.
<http://doi.org/10.1016/j.cub.2007.03.023>

- Tsubota, S. I., Vogel, A. C., Phillips, A. C., Ibach, S. M., Weber, N. K., Kostrzebski, M. A., & Spencer, S. A. (2011). Interactions between enhancer of rudimentary and notch and deltex reveal a regulatory function of enhancer of rudimentary in the notch signaling pathway in *Drosophila melanogaster*. *Fly*, 5(4), 275–284. <http://doi.org/10.4161/fly.5.4.17807>
- van Impel, A., Schumacher, S., Draga, M., Herz, H.-M., Grosshans, J., & Muller, H. A. J. (2009). Regulation of the Rac GTPase pathway by the multifunctional Rho GEF Pebble is essential for mesoderm migration in the *Drosophila* gastrula. *Development*, 136(5), 813–822.
<http://doi.org/10.1242/dev.026203>
- Waddington, C. H. (1943). THE INTERACTIONS OF SOME MORPHOGENETIC GENES IN *DROSOPHILA MELANOGASTER*. *Journal of Genetics*, 51(2), 243–258.
- Wang, Z., Pan, Y., Li, W., Jiang, H., Chatzimanolis, L., Chang, J., ... Liu, L. (2008). Visual pattern memory requires foraging function in the central complex of *Drosophila*. *Learning & Memory*, 15(3), 133–142. <http://doi.org/10.1101/lm.873008>
- Wehn, A., & Campbell, G. (2006). Genetic interactions among scribbler, Atrophin and groucho in *Drosophila* uncover links in transcriptional repression. *Genetics*, 173(2), 849–861.
<http://doi.org/10.1534/genetics.105.055012>
- Weissman, D. B., Feldman, M. W., & Fisher, D. S. (2010). The rate of fitness-valley crossing in sexual populations. *Genetics*, 186(4), 1389–1410. <http://doi.org/10.1534/genetics.110.123240>
- Wiellette, E. L., Harding, K. W., Mace, K. a, Ronshaugen, M. R., Wang, F. Y., & McGinnis, W. (1999). spen encodes an RNP motif protein that interacts with Hox pathways to repress the development of head-like sclerites in the *Drosophila* trunk. *Development (Cambridge, England)*, 126(23), 5373–5385.
- Willecke, M., Hamaratoglu, F., Kango-Singh, M., Udan, R., Chen, C. lin, Tao, C., ... Halder, G. (2006). The Fat Cadherin Acts through the Hippo Tumor-Suppressor Pathway to Regulate Tissue Size. *Current Biology*, 16(21), 2090–2100. <http://doi.org/10.1016/j.cub.2006.09.005>
- Wills, Z., Emerson, M., Rusch, J., Bikoff, J., Baum, B., Perrimon, N., & Van Vactor, D. (2002). A *Drosophila* homolog of cyclase-associated proteins collaborates with the Abl tyrosine kinase to control midline axon pathfinding. *Neuron*, 36(4), 611–622. [http://doi.org/10.1016/S0896-6273\(02\)01022-X](http://doi.org/10.1016/S0896-6273(02)01022-X)
- Wilson, T. G., Yerushalmi, Y., Donnell, D. M., & Restifo, L. L. (2006). Interaction between hormonal signaling pathways in *Drosophila melanogaster* as revealed by genetic interaction between Methoprene-tolerant and Broad-Complex. *Genetics*, 172(1), 253–264.
<http://doi.org/10.1534/genetics.105.046631>
- Winberg, M. L., Tamagnone, L., Bai, J., Comoglio, P. M., Montell, D., & Goodman, C. S. (2001). The transmembrane protein off-track associates with plexins and functions downstream of semaphorin signaling during axon guidance. *Neuron*, 32(1), 53–62.
[http://doi.org/10.1016/S0896-6273\(01\)00446-9](http://doi.org/10.1016/S0896-6273(01)00446-9)

- Wittwer, F., Jaquenoud, M., Brogiolo, W., Zarske, M., Wüstemann, P., Fernandez, R., ... Hafen, E. (2005). Susi, a negative regulator of Drosophila PI3-Kinase. *Developmental Cell*, 8(6), 817–827. <http://doi.org/10.1016/j.devcel.2005.04.002>
- Wu, Y. C., Chen, C. H., Mercer, A., & Sokol, N. S. (2012). Let-7-Complex MicroRNAs Regulate the Temporal Identity of Drosophila Mushroom Body Neurons via chinmo. *Developmental Cell*, 23(1), 202–209. <http://doi.org/10.1016/j.devcel.2012.05.013>
- Yun, B., Lee, K., Farkaš, R., Hitte, C., & Rabinow, L. (2000). The LAMMER protein kinase encoded by the Doa locus of Drosophila is required in both somatic and germline cells and is expressed as both nuclear and cytoplasmic isoforms throughout development. *Genetics*, 156(2), 749–761.
- Zeng, W., Wharton, K. a, Mack, J. a, Wang, K., Gadabaw, M., Suyama, K., ... Scott, M. P. (2000). naked cuticle encodes an inducible antagonist of Wnt signalling. *Nature*, 403(February 2000), 789–795. <http://doi.org/10.1038/35001615>
- Zeng, X., Singh, S. R., Hou, D., & Hou, S. X. (2010). Tumor suppressors Sav/Scrib and oncogene ras regulate stem-cell transformation in adult Drosophila malpighian tubules. *Journal of Cellular Physiology*, 224(3), 766–774. <http://doi.org/10.1002/jcp.22179>
- Zenvirt, S., Nevo-Caspi, Y., Rencus-Lazar, S., & Segal, D. (2008). Drosophila LIM-only is a positive regulator of transcription during thoracic bristle development. *Genetics*, 179(4), 1989–1999. <http://doi.org/10.1534/genetics.108.090076>
- Zhang, S., Xu, L., Lee, J., & Xu, T. (2002). Drosophila Atrophin homolog functions as a transcriptional corepressor in multiple developmental processes. *Cell*, 108(1), 45–56. [http://doi.org/10.1016/S0092-8674\(01\)00630-4](http://doi.org/10.1016/S0092-8674(01)00630-4)
- Zhao, H., Zheng, X., Yuan, X., Wang, L., Wang, X., Zhong, Y., ... Tully, T. (2009). ben Functions with Scamp during Synaptic Transmission and Long-Term Memory Formation in Drosophila. *Journal of Neuroscience*, 29(2), 414–424. <http://doi.org/10.1523/JNEUROSCI.5036-07.2009>

Code Supplement to ‘Are there Physical Linkages between Genes that have Synergistic Fitness Effects?’

I. Input and Search Patterns

```
# Cleaned up and complete file draft

# Takes as input - the FLYBASE SQL database, and the precomputed file
allele_genetic_interactions.tsv
# writes three sets of files - the phenotypes considered, those with interactions, and those with
fitness interactions
# neglects interactions with more than two loci, and those involving alleles with a known
artificial mutagen.
# Additionally, it neglects those not located on the same chromosome.

# The table in each file is of the form:
# Column 1: Locus 1 name
# Column 2: Locus 2 name
# Column 3: Locus 1 FLYBASE id
# Column 4: Locus 2 FLYBASE id
# Column 5: Locus 1 chromosome location (X, 2, 3, 4)
# Column 6: Locus 2 chromosome location (X, 2, 3, 4)
# Column 7: Locus 1 recombination map data
# Column 8: Locus 2 recombination map data
# Column 9: Recombination probability between locus 1 and locus 2
# Column 10: Interaction information
# Column 11: Flybase id of study determining interaction

# If there are multiple types of interactions between a pair of loci depending on the alleles
involved, we choose a fitness interaction
# prioritising lethal|viable|sterile|fertile, and label all others either 'non-fitness' or 'no
interaction'.

# Set working directory to wherever you have chosen to place your data.
setwd("folder-where-data-is")

##### Load the file.
library(data.table)
interdat <- fread("allele_genetic_interactions_fb_2017_02.tsv", header=T, sep="\t",
stringsAsFactors = F,
                data.table = F, encoding="UTF-8")

##### Convert Greek characters so that we do not have issues with character
encodings later.

greek_index <- c(0:9, "A", "B", "C", "D", "E", "F")
greek_names <- c("alpha", "beta", "gamma")
greek_iter <- c(9, "A", "B", "C")
```

```
greek_holder <- c(NA, rep(length(greek_iter)*length(greek_index)))
greek_letters <- c(NA, rep(length(greek_iter)*length(greek_index)))

j <- 0
for(outer in greek_iter){
  i <- 1
  for(inner in greek_index){
    startstr <- paste("\u03", outer, inner, sep="")
    nextstr <- parse(text= paste0(" ", startstr, " "))
    greek_holder[i + j*length(greek_index)] <- sum(grepl(nextstr, interdat[,3]))
    greek_letters[i + j*length(greek_index)] <- nextstr[[1]]
    i <- i + 1
  }
  j <- j+1
}
```

This allows us to work out what greek letters are in the file

```
interdat[,3] <- gsub("\u03B1", "alpha", interdat[,3])
interdat[,3] <- gsub("\u03B2", "beta", interdat[,3])
interdat[,3] <- gsub("\u03B3", "gamma", interdat[,3])
interdat[,3] <- gsub("\u03B4", "delta", interdat[,3])
interdat[,3] <- gsub("\u03B5", "epsilon", interdat[,3])
interdat[,3] <- gsub("\u03B6", "zeta", interdat[,3])
interdat[,3] <- gsub("\u03BC", "mu", interdat[,3])
interdat[,3] <- gsub("\u03C9", "omega", interdat[,3])
interdat[,3] <- gsub("\u0394", "Delta", interdat[,3])
```

Additionally, PostgreSQL requires an additional escape character for single quotes
interdat[,3] <- gsub("'", "''", interdat[,3])

Remove data from other types of flies.

```
# Gene symbols are appended with Nnnn\ if they are foreign, with the exception of
# Dmel\, which is Drosophila melanogaster (not normally used, see FlyBase Nomenclature)
melano <- nrow(interdat)
interdat <- interdat[!grepl("Nnnn\\", interdat[,3]), ]
non_melano <- melano - nrow(interdat)
melano <- nrow(interdat)
```

Determine matched pairs, remove rows with descriptions containing more than two loci.

```
# Involves a complicated regular expression, which basically amounts to:
# "Is preceded by a space, the start of the string, or a forward slash,
# contains the numbers 0-9 or a opening square bracket,
# does not contain a space, a slash, or a comma,
# then is succeeded by a forward slash, a space, a comma, or the end of the string".
```

```
xm <- 1:nrow(interdat)
```

```
spec_pattern <- "(?<=\\ |^\\|\\|)[^ /,]*(?:[0-9]\\|\\|)[^ /,]*(?:=\\|\\| \\|,|$)"
spec_matches <- gregexpr(spec_pattern, interdat[,3], perl=T)
spec_set <- regmatches(interdat[xm,3],
  gregexpr("(?<=\\ |^\\|\\|)[^ /,]*(?:[0-9]\\|\\|)[^ /,]*(?:=\\|\\| \\|,|$)",
    interdat[xm,3], perl=T))

##### Get the loci names and mutagen data, etc via SQL queries.

# various queries before loop
sql_template <- "
SELECT DISTINCT (f.uniquename)
FROM feature f, feature_synonym fs, synonym s, cvterm cvt, cvterm cvt2
WHERE f.type_id = cvt.cvterm_id AND cvt.name IN
('gene','chromosome_structure_variation') AND
f.feature_id = fs.feature_id AND fs.synonym_id = s.synonym_id AND
fs.is_current = 't' AND fs.is_internal = 'f' AND
s.type_id = cvt2.cvterm_id AND cvt2.name = 'symbol' AND
f.is_obsolete = 'f' AND s.name IN
"

ordering <- "ORDER BY f.uniquename;"

sql_gene <- "
SELECT g.uniquename, g.name, f.uniquename
FROM feature f, feature g, feature_relationship fr, cvterm cvt
WHERE g.feature_id = object_id AND subject_id = f.feature_id AND
fr.type_id = cvt.cvterm_id AND cvt.name = 'alleleof' AND
f.uniquename IN
"

# mutation query, provided you have an ID on the allele.
sql_mute <- "
SELECT DISTINCT f.uniquename, f.name, cvt.name
FROM cvterm cvt, cvtermprop cvtp, feature f, feature_cvterm fcv, pub p
WHERE f.feature_id = fcv.feature_id AND fcv.cvterm_id = cvt.cvterm_id AND
cvt.cvterm_id = cvtp.cvterm_id AND fcv.pub_id = p.pub_id AND
cvtp.VALUE = 'origin_of_mutation' AND f.uniquename IN
"

loci_list <- list()
names_list <- list()
mutagen_list <- list()
no_allele_data <- NULL

# basic function to test for mutagens.
# just checks a string, basically (going to have to manually look through mutagens... sigh)
is_artificial <- function(mute_input){
  return(F)
}
```

```
#install.packages("RPostgreSQL")
library(RPostgreSQL)

# log on to PostgreSQL database
pw <- {"12345"}

# load driver
drv <- dbDriver("PostgreSQL")
# create connection
con <- dbConnect(drv, dbname="fb2017_02", host="localhost", port=5432,
                 user="postgres", password=pw)
# removes password
rm(pw)

# check for existence of feature table
dbExistsTable(con, "feature")
# if true, we're good to go.
allele_list <- list()

for(i in 1:length(spec_set)){

  loci_index <- paste("", spec_set[[i]], "", sep="", collapse=",")
  loci_query <- paste(sql_template, "(", loci_index, ")", ordering, sep="")

  # gets the allele ids
  allele_data <- dbGetQuery(con, loci_query)
  allele_list[[i]] <- allele_data

  if(nrow(allele_data) > 0){
    # gene and mutation data resulting from that
    gene_query <- paste(sql_gene, "(", paste("", allele_data[,1], "", sep="", collapse=","),
                       ")\n", ordering, sep="")
    mute_query <- paste(sql_mute, "(", paste("", allele_data[,1], "", sep="", collapse=","),
                       ")\n", ordering, sep="")

    gene_data <- dbGetQuery(con, gene_query)
    mute_data <- dbGetQuery(con, mute_query)

    # check if it contains a mutagen
    # otherwise:
    if(nrow(gene_data) > 0){
      index <- !duplicated(gene_data[,1]) & !duplicated(gene_data[,2])
      gene_data <- gene_data[index,]
      loci_list[[i]] <- sort(gene_data[,1])
      names_list[[i]] <- gene_data[order(gene_data[,1]), 2]

      if(nrow(mute_data) > 0)
        mutagen_list[[i]] <- mute_data[,3]
      else
        mutagen_list[[i]] <- NA
    }
  }
}
```

```
}
else{
  loci_list[[i]] <- NA
  names_list[[i]] <- NA
  mutagen_list[[i]] <- NA
}

# should probably keep the name data somewhere so I don't need to look it up again.
}
else
  no_allele_data <- c(no_allele_data, i)
}

# when finished, disconnect
dbDisconnect(con)
dbUnloadDriver(drv)

##### Restrict to two loci [note that NON-ENHANCERS of phenotypes show up
later]
# We would like to determine the number of three locus cases, and look at subsets here.

two_indexer <- sapply(1:length(loci_list), function(x) length(loci_list[[x]]) == 2)

# Check how many interactions have 3 or more loci involved.
more_indexer <- sapply(1:length(loci_list), function(x) length(loci_list[[x]]) > 2)
n_three_or_more <- sum(more_indexer)

extract_pairs <- function(x){
  list_subset <- loci_list[more_indexer][[x]]
  hold <- combn(1:length(list_subset), 2)
  return(list(split(list_subset[hold], col(hold))))
}

# Get the subsets of pairs of loci.
more_subsets <- sapply(1:length(loci_list[more_indexer]), extract_pairs)
# Now wish to check whether all of these subsets occur in pairs of interactions or not.
test_subsets <- sapply(1:length(more_subsets), function(x) unlist(more_subsets[x],
recursive=FALSE))
# For each three or > interaction, we wish to test if that ALL or ANY pairs occur in the
regular list.
# WARNING: VERY SLOW.
length_holder <- sapply(1:length(test_subsets), function(x) length(test_subsets[[x]]))
start_holder <- cumsum(length_holder) - length_holder[1] + 1
end_holder <- cumsum(length_holder)

# separate both into matrices and sort row-wise. Remove duplicates
mat_test_subsets <- matrix(unlist(test_subsets), ncol=2, byrow=T)
mat_test_subsets <- matrix(sapply(1:nrow(mat_test_subsets), function(x)
sort(mat_test_subsets[x,])), ncol=2, byrow=T)
```

```
#mat_test_subsets <- mat_test_subsets[!duplicated(mat_test_subsets),]

mat_two_subsets <- matrix(unlist(loci_list[two_indexer]), ncol=2, byrow=T)
mat_two_subsets <- matrix(sapply(1:nrow(mat_two_subsets), function(x)
sort(mat_two_subsets[x,])), ncol=2,byrow=T)

mat_two_subsets <- mat_two_subsets[!duplicated(mat_two_subsets),]

# test which of these second order effects are in the interactions
checker <- sapply(1:nrow(mat_test_subsets), function(x) mat_test_subsets[x,] %in%
mat_two_subsets)
check_ref <- sapply(1:length(length_holder), function(x)
sum(!checker[start_holder[x]:end_holder[x]]))

sum(check_ref > 0)
# conclude 2831/9585 contain pairwise interactions not included in the pairs. (i.e. ~30%),
which is acceptable.
sum(!checker)
# this accounts for 7983/63916 of pairwise interactions, if we consider it that way

##### Remove entries containing alleles with a mutagen.

# issues with NAs, so replace with "none"

mutagen_list <- sapply(1:length(mutagen_list),
function(x) replace(mutagen_list[[x]], is.na(mutagen_list[[x]]), "none"))

# mutation occur either 1) spontaneously; 2) in natural populations; or 3) no cause is given

mutagen_less <- sapply(1:length(mutagen_list),
function(x) all(grepl("spontaneous|natural population|none",
mutagen_list[[x]][!is.na(mutagen_list[[x]])])))

n_mute <- sum(!mutagen_less)
tote_mute <- length(mutagen_list)

##### Classify interactions as: non-interactions, fitness interactions, non-fitness
interactions
# Note that we have not yet removed duplicates - this will come later.
check_inter <- sapply(1:nrow(interdat), function(x) grepl("non\\-enh|non\\-suppr",
interdat[x,3]))
check_fit <- sapply(1:nrow(interdat), function(x) grepl("lethal|viab|fertil|steril",
interdat[x,3]))

# We then have to classify the types of interaction of loci
# 1) Type 0: no interaction
# 2) Type 1: interaction, but not a fitness one
# 3) Type 2: fitness interaction
```

```
# Note that an interaction will be listed multiple times: we go fitness > normal interaction >
no interaction
# Start putting the table together:

new_interdat <- interdat[two_indexer&mutagen_less,]
new_interdat <- cbind(new_interdat, sapply(which(two_indexer&mutagen_less), function(x)
loci_list[[x]][1]))
new_interdat <- cbind(new_interdat, sapply(which(two_indexer&mutagen_less), function(x)
names_list[[x]][1]))
new_interdat <- cbind(new_interdat, sapply(which(two_indexer&mutagen_less), function(x)
loci_list[[x]][2]))
new_interdat <- cbind(new_interdat, sapply(which(two_indexer&mutagen_less), function(x)
names_list[[x]][2]))

# Put in a column describing the type of interaction. ["no"]=0, "non fit=1", "fit=2"]
new_interdat <- cbind(new_interdat, ((1 - check_inter)*(1 +
check_fit))[which(two_indexer&mutagen_less)])
colnames(new_interdat)[5:9] <- c("Locus 1 id", "Locus 1 name", "Locus 2 id", "Locus 2
name", "Fitness type")
# Now start merging interactions according to a protocol

# Go with first listed, I think? Or maybe lowest number on FBref. Doesn't matter too much.
library(plyr)
# match_df to find the things that match it (in particular the fitness types)

# set to max of fitness, then take first column
# so that we have access to the previous table, make a new one.
second_interdat <- new_interdat

for(k in 1:nrow(second_interdat)){
  second_interdat[k,9] <- max(match_df(new_interdat, new_interdat[k,c(5,7)])[,9])
}

second_interdat <- second_interdat[!duplicated(second_interdat[,c(5,7)]),]

##### Load map data - restrict to interactions on same chromosome, and get
recombination rates.
##### Additionally, get sequence location data and check for gene overlap.
map_table <- fread("gene_map_table_fb_2017_02.tsv", header=T, sep="\t", stringsAsFactors
= F,
                 data.table = F, encoding="UTF-8")

# add new columns to store recombination data and cytogenetic data
second_interdat[,10:21] <- rep(NA, nrow(second_interdat))
colnames(second_interdat)[10:21] <- c("Locus 1 chromosome", "Locus 1 recomb", "Locus 2
chromosome", "Locus 2 recomb",
    "Loc1cyt1", "Loc1cyt2", "Loc1seq1", "Loc1seq2",
    "Loc2cyt1", "Loc2cyt2", "Loc2seq1", "Loc2seq2")
```



```
# NOTE THAT WE NOW NEED TO SORT THE ROWS TO GET CONSISTENT RESULTS
```

```
# Some cytogenetic data is unhelpful and needs to be done by hand  
hand_cyto <- c()
```

```
# extract recombination data and sequence location data from table  
for(k in 1:nrow(second_interdat)){  
  # First locus information  
  recomb_holder <- map_table[match(second_interdat[k, 6], map_table[,1]), 3]  
  second_interdat[k,10] <- substr(recomb_holder, 1, 1)  
  second_interdat[k,11] <- gsub("\\[\\]", "", substr(recomb_holder, 3, 10))
```

```
  # Cytogenetic location  
  trimmed_loc <- map_table[match(second_interdat[k, 6], map_table[,1]), 4]  
  if(nchar(trimmed_loc) == 0)  
    second_interdat[k,14:15] <- c(NA, NA)  
  else{  
    second_interdat[k,14:15] <- strsplit(trimmed_loc, "\\-")[[1]][c(1,2)]  
    if(length(strsplit(trimmed_loc, "\\-")[[1]])>2)  
      hand_cyto <- c(hand_cyto,k)  
  }  
}
```

```
  # Sequence location  
  trimmed_loc <- map_table[match(second_interdat[k, 6], map_table[,1]), 5]  
  if(nchar(trimmed_loc) == 0)  
    second_interdat[k,16:17] <- c(NA, NA)  
  else{  
    trimmed_loc <- gsub(".*\\:|\\(..*\\)", "", trimmed_loc, perl=TRUE)  
    second_interdat[k,16:17] <- strsplit(trimmed_loc, "\\.|\\.")[[1]]  
  }  
}
```

```
  # Second locus information  
  recomb_holder <- map_table[match(second_interdat[k, 8], map_table[,1]), 3]  
  second_interdat[k,12] <- substr(recomb_holder, 1, 1)  
  second_interdat[k,13] <- gsub("\\[\\]", "", substr(recomb_holder, 3, 10))
```

```
  # Cytogenetic location  
  trimmed_loc <- map_table[match(second_interdat[k, 8], map_table[,1]), 4]  
  if(nchar(trimmed_loc) == 0)  
    second_interdat[k,18:19] <- c(NA, NA)  
  else{  
    second_interdat[k,18:19] <- strsplit(trimmed_loc, "\\-")[[1]][c(1,2)]  
    if(length(strsplit(trimmed_loc, "\\-")[[1]])>2)  
      hand_cyto <- c(hand_cyto,k)  
  }  
}
```

```
  # Sequence location  
  trimmed_loc <- map_table[match(second_interdat[k, 8], map_table[,1]), 5]
```

```
if(nchar(trimmed_loc) == 0)
  second_interdat[k,20:21] <- c(NA, NA)
else{
  trimmed_loc <- gsub(".*\\:|\\(..*\\)", "", trimmed_loc, perl=TRUE)
  second_interdat[k,20:21] <- strsplit(trimmed_loc, "\\.|\\.")[[1]]
}
}
```

This cytogenetic data contains notes and can be hard to parse
Put in the code manually, but SHOULD BE CHECKED AND CHANGED
Indices can be found in the variable 'hand_cyto'

```
second_interdat[11,18:19] <- c("21C1", "25A1") # Abnormal-abdomen
second_interdat[101,14:15] <- c("70D2", "89B9") # ale
second_interdat[103, 14:15] <- c("70DC", "89B9") # ale
second_interdat[187, 18:19] <- c("5A8", "7B3") # Su(tor)1-1
second_interdat[188, 18:19] <- c("52D3", "59E1") #Su(tor)
second_interdat[288, 14:15] <- c("70DC", "89B9") #Mosaic
second_interdat[410, 14:15] <- c("7D1", "8A2") #Su(stn)
second_interdat[545, 18:19] <- c("1B1", "3C2") #dcm
second_interdat[567, 14:15] <- c("8D8", NA) #dvr
second_interdat[674, 14:15] <- c("100A2", "100C3") #ld
second_interdat[787, 18:19] <- c("13F1", "15F4") #vb
```

some unavailable, don't worry about for now, have chromosome data for all of them

```
second_sum <- nrow(second_interdat)
removed_match_chroms <- sum(second_interdat[,10]!=second_interdat[,12])
second_interdat <- second_interdat[second_interdat[,10]==second_interdat[,12],]
# turns out this gives us all the recombination data as well for these pairs.
```

Some recombination data missing, unfortunately, due to lack of gene annotation: drop

```
removed_recomb_data <- sum(nchar(second_interdat[,11])==0 |
nchar(second_interdat[,13])==0)
final_interdat <- second_interdat[!(second_interdat[,11]=="" | second_interdat[,13]==""),]
# Lose 2 non-interactions, 17 non-fitness interaction, 0 fitness interactions
```

Finally, remove points with overlapping locations.
Use sequence location data first

```
does_overlap <- ((pmin(final_interdat[,16], final_interdat[,17]) < pmax(final_interdat[,20],
final_interdat[,21])) &
  (pmax(final_interdat[,16], final_interdat[,17]) > pmax(final_interdat[,20],
final_interdat[,21]))) |
  ((pmax(final_interdat[,16], final_interdat[,17]) > pmin(final_interdat[,20],
final_interdat[,21])) &
  (pmax(final_interdat[,16], final_interdat[,17]) < pmax(final_interdat[,20],
final_interdat[,21])))
overlap_seq_num <- sum(does_overlap[!is.na(does_overlap)])
```

Check to see which sequence data is missing

```
# It turns out that sequence data is much more likely to be missing than cytological data;
assume that they don't
# overlap
na_indexer <- rowSums(is.na(final_interdat[,14:21]))>0
na_seq_indexer <- rowSums(is.na(final_interdat[,c(16,17,20,21)]))>0
seq_loc_missing <- sum(na_seq_indexer)

# Deal with the NAs and remove overlaps
does_overlap[is.na(does_overlap)] <- 0
final_interdat <- final_interdat[!does_overlap,]

# Scale to lie within 1; also include truncated value.
final_interdat[,22:23] <- rep(0, nrow(final_interdat))

final_interdat[,22] <- pmin(abs(as.numeric(final_interdat[,11]) -
as.numeric(final_interdat[,13]))/100, 0.5)
final_interdat[,23] <- pmin(abs(as.numeric(final_interdat[,11]) -
as.numeric(final_interdat[,13]))/100, 1)
```

II. Data Analysis

```
library(ggplot2)
library(scales)

#unscaled histogram
#unscaled_hist <- ggplot() + geom_histogram(aes(x=final_interdat[final_interdat[,9]==0,
14]), fill="red", alpha=0.2) +
# geom_histogram(aes(x=final_interdat[final_interdat[,9]==1, 14]), fill="blue", alpha=0.2) +
#geom_histogram(aes(x=final_interdat[final_interdat[,9]==2, 14]), fill="green", alpha=0.2)
+
#labs(title="Histogram of non-interactions, non-fitness interactions, and fitness
interactions",
# x="Recombination rate", y="Interaction counts")
#scaled histogram
#ggplot() + geom_histogram(aes(x=final_interdat[final_interdat[,9]==0, 14],
y=..count../sum(..count..)), fill="red", alpha=0.2) +
# geom_histogram(aes(x=final_interdat[final_interdat[,9]==1, 14],
y=..count../sum(..count..)), fill="blue", alpha=0.2) +
#geom_histogram(aes(x=final_interdat[final_interdat[,9]==2, 14],
y=..count../sum(..count..)), fill="green", alpha=0.2) +
#labs(title = "Scaled histogram of non-interactions, non-fitness interactions, and fitness
interactions", x = "Recombination probability", y="Proportion")

ana_data <- final_interdat[,c(23,22,9,10,6,8,4)]
colnames(ana_data) <- c("recomb", "recomb_trunc2", "interaction", "chromosome", "Locus 1
name", "Locus 2 name", "study_ref")
ana_data$recomb_trunc2 <- ana_data$recomb_trunc2*2

# we want to model: response is recombination distance
# effect of interaction; fitness (note fitness is nested in interaction, so mixed effect)
# control for chromosome (though histograms don't indicate anything in particular);
categorical variable

# Issues: presence of zeroes (19), all in one category
# Presence of ones (2), all in one category
# Additionally if using truncation, 51 recomb distances are > 0.5, resulting in aggregation
there.
# Finally the design is random (i.e. the number of each was not known in advance)

library(gamlss)

# There are multiple formulas: one of the mean of the beta distribution, its variance,
# The zeros, as well as the ones.
```

```
# In particular, the model for the zeroes is quite important, as they are ALL in one category
(interactions).
# We also consider a very basic censored model, where 0 values are assumed to be randomly
and uniformly distributed
# in 0.001-0.003 (where 0.003 is the minimum value is observed) & the 1's are truncated to
0.999, allowing a beta model.
# Finally, we check against a model where all values are truncated to 0.5.
# Check against truncated as well? Probably won't qualitatively change results...
# We consider models with interaction effects on each chromosome, and then as a simple
effect.

# Truncating because centimorgan distances mean nothing about 50cM (cannot observe more
than half getting into the zygote)

set.seed(42)

# Truncated models
model.null <- gamlss(recomb_trunc2~1, data=ana_data, family=BEINF())

# Full linear model
model.full <- gamlss(recomb_trunc2 ~ as.factor(interaction)*as.factor(chromosome),
  nu.formula = ~ as.factor(interaction)*as.factor(chromosome),
  tau.formula = ~ as.factor(interaction)*as.factor(chromosome),
  sigma.formula = ~ as.factor(interaction)*as.factor(chromosome),
  data=ana_data, family=BEINF())
model.best.1 <- stepGAICAll.A(model.full)
model.best.3 <- stepGAICAll.B(model.full)
model.best.2 <- stepGAICAll.A(model.full, k=log(nrow(ana_data)))

# extract starts and ends for each chromosome
# order by start
# Red (area spanned by leftmost gene), Blue (area spanned by right-most gene) and Green
(area between)

library(IRanges)
library(ggplot2)

# extract sequences and ignore NAs
# Maybe actually get the loci names and indices as well...
interv_setup <- final_interdat[,c(10,16,17,20,21)]
interv_setup <- interv_setup[which(rowSums(is.na(interv_setup)) == 0),]
interv_setup <- sapply(interv_setup, as.numeric)

# switch around sequences not in the right order (i.e.)
correct_index <- interv_setup[,2] <= interv_setup[,4]
interv_setup[!correct_index,c(1:5)] <- interv_setup[!correct_index, c(1,4,5,2,3)]
interv_setup <- interv_setup[order(interv_setup[,2]),]

chrom1_interv <- interv_setup[interv_setup[,1]==1,]
```

```
chrom2_interv <- interv_setup[interv_setup[,1]==2,]
chrom3_interv <- interv_setup[interv_setup[,1]==3,]

# another problem... ordering... sigh. May need to swap around.

#bins1 <- disjointBins(IRanges(chrom1_interv[,2], chrom1_interv[,5]))
bins1 <- 1:nrow(chrom1_interv)
#bins2 <- disjointBins(IRanges(chrom2_interv[,2], chrom2_interv[,5]))
bins2 <- 1:nrow(chrom2_interv)
bins3 <- disjointBins(IRanges(chrom3_interv[,2], chrom3_interv[,5]))
bins3 <- 1:nrow(chrom3_interv)

ggplot() + geom_rect(aes(xmin=chrom1_interv[,3], xmax=chrom1_interv[,4], ymin=bins1,
ymax=bins1+0.9), colour="grey", fill="grey") +
  geom_rect(aes(xmin=chrom1_interv[,2], xmax=chrom1_interv[,3], ymin=bins1,
ymax=bins1+0.9), colour="red", fill="red") +
  geom_rect(aes(xmin=chrom1_interv[,4], xmax=chrom1_interv[,5], ymin=bins1,
ymax=bins1+0.9), colour="blue", fill="blue") +
  labs(title="X Chromosome locus pairs", x="Sequence location", y=NULL)

ggplot() + geom_rect(aes(xmin=chrom2_interv[,3], xmax=chrom2_interv[,4], ymin=bins2,
ymax=bins2+0.9), colour="grey", fill="grey") +
  geom_rect(aes(xmin=chrom2_interv[,2], xmax=chrom2_interv[,3], ymin=bins2,
ymax=bins2+0.9), colour="red", fill="red") +
  geom_rect(aes(xmin=chrom2_interv[,4], xmax=chrom2_interv[,5], ymin=bins2,
ymax=bins2+0.9), colour="blue", fill="blue") +
  labs(title="Chromosome 2 locus pairs", x="Sequence location", y=NULL)

ggplot() + geom_rect(aes(xmin=chrom3_interv[,3], xmax=chrom3_interv[,4], ymin=bins3,
ymax=bins3+0.9), colour="grey", fill="grey") +
  geom_rect(aes(xmin=chrom3_interv[,2], xmax=chrom3_interv[,3], ymin=bins3,
ymax=bins3+0.9), colour="red", fill="red") +
  geom_rect(aes(xmin=chrom3_interv[,4], xmax=chrom3_interv[,5], ymin=bins3,
ymax=bins3+0.9), colour="blue", fill="blue") +
  labs(title="Chromosome 3 locus pairs", x="Sequence location", y=NULL)
```

```
# Check residual plots, GAIC
```

B. Fitted model

```
Mu link function: logit
Mu Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.37635    0.08882  -4.237 3.24e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
-----
Sigma link function: logit
Sigma Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.46515    0.07183   6.475 5.36e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
-----
Nu link function:  log
Nu Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.2260     0.2415  -9.218  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
-----
Tau link function:  log
Tau Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.7314     0.5690  -3.043  0.00260 **
as.factor(interaction)1 -0.1424     0.4361  -0.326  0.74436
as.factor(interaction)2 -1.7207     0.8381  -2.053  0.04114 *
as.factor(chromosome)2  1.5632     0.4845   3.227  0.00143 **
as.factor(chromosome)3 -0.2505     0.5775  -0.434  0.66490
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
-----
No. of observations in the fit:  246
Degrees of Freedom for the fit:   8
      Residual Deg. of Freedom: 238
              at cycle: 4
```

```
Global Deviance:  311.7911
      AIC:         327.7911
      SBC:         355.8337
```

III. Simulation code and analysis

A. Simulation code

```
# Simulation function, called for multiple population sizes and rates
mono_haploid_sim <- function(pop_size, recomb_rate, sel_strength, init_pop, num_reps,
num_gens){
```

```
  # set the working directory, etc.
  setwd("C:/Working-directory-here")
  # create a new folder with date and time
```

```
  start_time <- Sys.time()
  n <- unlist(strsplit(toString(start_time), " "))
  substr(n[2], 3, 3) <- "-"
  substr(n[2], 6, 6) <- "-"
```

```
  dir_name <- paste(n[1], n[2], sep="_")
  dir.create(paste(dir_name))
```

```
setwd(dir_name)

# set the random number seed
random_seed <- 42

# write to text file, also time and date (later)
# to get time and date: Sys.Date(), Sys.time()

# typical landscape with just one beneficial genotype
# randomly initialising for the time being.

# now, what do I want to start with?
if(any(is.na(init_pop)))
  start_pop <- rmultinom(n=1, size=pop_size, prob=c(1,1,1,1))
else
  start_pop <- init_pop
sel_numer <- 1 + sel_strength

# wright-fisher model
pop_record <- matrix(0, nrow=num_gens, ncol=4*num_reps)
pop_record[1,] <- rep(start_pop, num_reps)

# recombination binomials drawn in advance
recomb_counter <- matrix(rbinom(n=(num_gens-1)*num_reps, size=pop_size,
prob=recomb_rate), ncol=num_reps, nrow=num_gens-1)
non_recomb_counter <- pop_size - recomb_counter
ref_index <- seq(1, 4*num_reps - 3, by = 4)
prob_draw <- c(rep(NA, length(ncol(pop_record))))

recomb_probs <- c(rep(NA, length(ncol(pop_record))))

# note: selective one assumed to be first entry, i.e. [00, 01, 10, 11]

# evolve and record
for(i in 1:(num_gens-1)){
  # calculate relevant probabilities:
  denoms <- pop_size + sel_strength * pop_record[i,]
  prob_draw <- pop_record[i,]/denoms
  prob_draw[ref_index] <- sel_numer*prob_draw[ref_index]

  # divide into - recombining half, etc
  pop_record[i+1,] <- sapply(1:num_reps,
    function(x) rmultinom(1, size=non_recomb_counter[i,x],
prob=prob_draw[(4*x-3):(4*x)]))
  # get the number of draws for each
  recomb_probs[ref_index] <- prob_draw[ref_index]*(1 - prob_draw[ref_index + 3]) +
  prob_draw[ref_index + 1] * prob_draw[ref_index + 2]
```



```
    recomb_probs[ref_index + 1] <- prob_draw[ref_index + 1]*(1 - prob_draw[ref_index + 2])
+ prob_draw[ref_index] * prob_draw[ref_index + 3]
    recomb_probs[ref_index + 2] <- prob_draw[ref_index + 2]*(1 - prob_draw[ref_index + 1])
+ prob_draw[ref_index] * prob_draw[ref_index + 3]
    recomb_probs[ref_index + 3] <- prob_draw[ref_index + 3]*(1 - prob_draw[ref_index]) +
prob_draw[ref_index + 1] * prob_draw[ref_index + 2]

    pop_record[i+1, ] <- pop_record[i+1, ] + sapply(1:num_reps,
          function(x) rmultinom(1, size=recomb_counter[i,x],
            prob=recomb_probs[(4*x-3):(4*x)]))
  }

# now write data frame to file
write.csv(pop_record, "sim_data.csv")

fin_time <- Sys.time()

# save data to two files: explanation and actual data
sim_exp <- file("sim_exp.txt")
writeLines(c("Recombination simulation in R by Julian Byrnes", paste("Population size: ",
pop_size, sep=""),
  paste("Selection strength: ", sel_numer, sep=""),
  paste("Recombination rate: ", recomb_rate, sep=""),
  paste("Random number seed: ", random_seed, sep=""),
  paste("Starting population: ", toString(start_pop), sep=""),
  paste("Start time: ", toString(start_time), sep=""),
  paste("Finish time: ", toString(fin_time), sep=""),
  paste("Time elapsed: ", as.numeric(fin_time - start_time), " seconds", sep="")),
sim_exp)
close(sim_exp)

# machine readable version
write.csv(as.matrix(c(pop_size, sel_numer, recomb_rate, random_seed, start_pop)),
"sim_exp2.csv")

# clear workspace
#rm(list=ls())

}

# collection of population size, initial populations, sel strengths, recomb rates
# num reps is always 10000, num gens is always 100
# actually, num_gens should scale with population size (comments on diffusion)

pop_sizes <- seq(from=50, to=500, by = 50)
recomb_rates <- (0:10)*0.0025
sel_strengths <- (0:20)*0.0025
```

```
# note that code prior to 03/05/2017 has wrong levels of selection
# let init_pop deal be random for the time being? No, want all sims to be comparable:
generate for smallest, mult.
```

```
# letting init_pop be random means they all cluster around LE. No.
```

```
# Case 1: complete LD
# 0 25 25 0
# Case 2: roughly evenly divided
# 12 13 13 12
# Case 3: in opposition
# 25 0 0 25
# Case 4: one mid dominates
# 12 25 0 13
# Case 5: very few
# 5 10 10 25
# Case 6: two alleles
# 25 25 0 0
# Case 7: on the verge
# 5 0 0 45
# Case 8 tilted verge
# 5 10 0 35
# Case 9
# 0 12 13 25
# Case 10
# 1 24 25 0
# Case 11
# 0 5 5 40
# Case 12
# 1 5 5 39
```

```
set.seed(42)
```

```
start_pops <- matrix(c(0, 25, 25, 0, 12, 13, 13, 12, 25, 0, 0, 25, 12, 25, 0, 13,
                      5, 10, 10, 25, 25, 25, 0, 0, 5, 0, 0, 45, 5, 10, 0, 35, 0, 12, 13, 25,
                      1, 24, 25, 0, 0, 5, 5, 40, 1, 5, 5, 39), nrow=4, byrow=F)
```

```
for(pop_count in 1:length(pop_sizes)){
  for(recomb_count in 1:length(recomb_rates)){
    for(sel_count in 1:length(sel_strengths)){
      for(init_count in 1:ncol(start_pops)){

        mono_haploid_sim(pop_size=pop_sizes[pop_count],
                        recomb_rate=recomb_rates[recomb_count],
                        sel_strength=sel_strengths[sel_count],
                        init_pop=start_pops[init_count]*pop_count, num_reps=10000,
num_gens=min(100, pop_sizes[pop_count]))
```

```
}  
}  
}  
}
```

B. Analysis code

```
library(raster)  
  
# load simulation results file as sim_res  
sim_res <- read.csv("C:/Users/z5048271/Desktop/flybase-  
manuscript/recombination_simulation_results.csv", header=T)  
  
# Compile into correct format for retrieval.  
sim_index <- sim_res[,1]  
sim_res <- sim_res[,-1]  
colnames(sim_res) <- rep(colnames(sim_res)[1:11], 10)  
  
new_sim_res <- sim_res[,1:11]  
  
for(i in 2:10){  
  print(((i-1)*11 + 1):((i-1)*11 + 11))  
  new_sim_res <- rbind(new_sim_res, sim_res[(((i-1)*11 + 1):((i-1)*11 + 11)])  
}  
  
sim_res <- new_sim_res  
new_sim_res <- NULL  
sim_index <- rep(sim_index, 10)  
  
sel_index <- sort(unique(sim_res[,2]))  
rec_index <- sort(unique(sim_res[,3]))  
  
# Case 1: complete LD  
case_1 <- c(0,25,25,0)  
# Case 2: roughly evenly divided  
case_2 <- c(12,13,13,12)  
# Case 3: in opposition  
case_3 <- c(25,0,0,25)  
# Case 4: one mid dominates  
case_4 <- c(12,25,0,13)  
# Case 5: very few  
case_5 <- c(5,10,10,25)  
# Case 6: two alleles  
case_6 <- c(25,25,0,0)  
# Case 7: on the verge  
case_7 <- c(5,0,0,45)  
# Case 8 tilted verge  
case_8 <- c(5,10,0,35)  
# Case 9  
case_9 <- c(0,12,13,25)
```

```
# Case 10
case_10 <- c(1,24,25,0)
# Case 11
case_11 <- c(0,5,5,40)
# Case 12
case_12 <- c(1,5,5,39)

case_all <- rbind(case_1, case_2, case_3, case_4, case_5, case_6, case_7, case_8, case_9,
                 case_10,
                 case_11, case_12)

# This is a table with several entries.
# The columns contain: 1) Population size 2) Selection strength 3) Recombination rate
# 4-7) Starting population 8) Probability of fixation in favoured 9) Probability of fixation
# 10) Mean time of fixation in the favoured genotype 11) Mean time of fixation (if it occurs)

# We create a number of rasters to examine the probabilities of fixation in various scenarios
and population sizes.
ld_vals <- (sim_res[,4]*sim_res[,7] - sim_res[,5]*sim_res[,6])/sim_res[,1]^2

# Multiple scenarios, should be on the same colour scale.
# Time plots for each case (overall fixation time)

cond_prob <- sim_res[,8]/sim_res[,9]
prob_col <- colorRampPalette(c("white", "green"))(max(cond_prob[!is.na(cond_prob)]))
# Fixation probability for each case (conditional given fixation occurs)
# Population size 100
for(i in 1:12){
  case_indexer <- i + (1:(nrow(sim_res)/12) - 1)*12
  pop_indexer <- which(sim_res[,1] == 100)
  testdf <- sim_res[intersect(case_indexer,pop_indexer),c(8,2,3)]
  testdf[,1] <- testdf[,1] / sim_res[intersect(case_indexer, pop_indexer),9]
  colnames(testdf) <- c("p", "s", "r")
  testdf <- testdf[!is.na(testdf$p),]

  coordinates(testdf) <- ~ r + s
  gridded(testdf) <- TRUE
  rasterDF <- raster(testdf)
  plot(rasterDF, xlab = "Recombination", ylab = "Selection")

  # insert title here
  title(main = paste("Probability of Fixation", "for Case", i, sep=" "))
}

# Overall fixation time
# Population size 100
time_col <- colorRampPalette(c("black", "white"))(max(sim_res[!is.na(sim_res[,11]),11]))

for(i in 1:12){
```

```
case_indexer <- i + (1:(nrow(sim_res)/12) - 1)*12
pop_indexer <- which(sim_res[,1] == 100)
testdf <- sim_res[intersect(case_indexer,pop_indexer),c(10,2,3)]
#testdf[is.na(testdf[,1]),1] <- 200
colnames(testdf) <- c("t", "s", "r")

coordinates(testdf) <- ~ r + s
gridded(testdf) <- TRUE
rasterDF <- raster(testdf)
plot(rasterDF, xlab = "Recombination", ylab = "Selection")

# insert title here
title(main = paste("Expected Time of Fixation", "for Case", i, sep=" "))
}
```

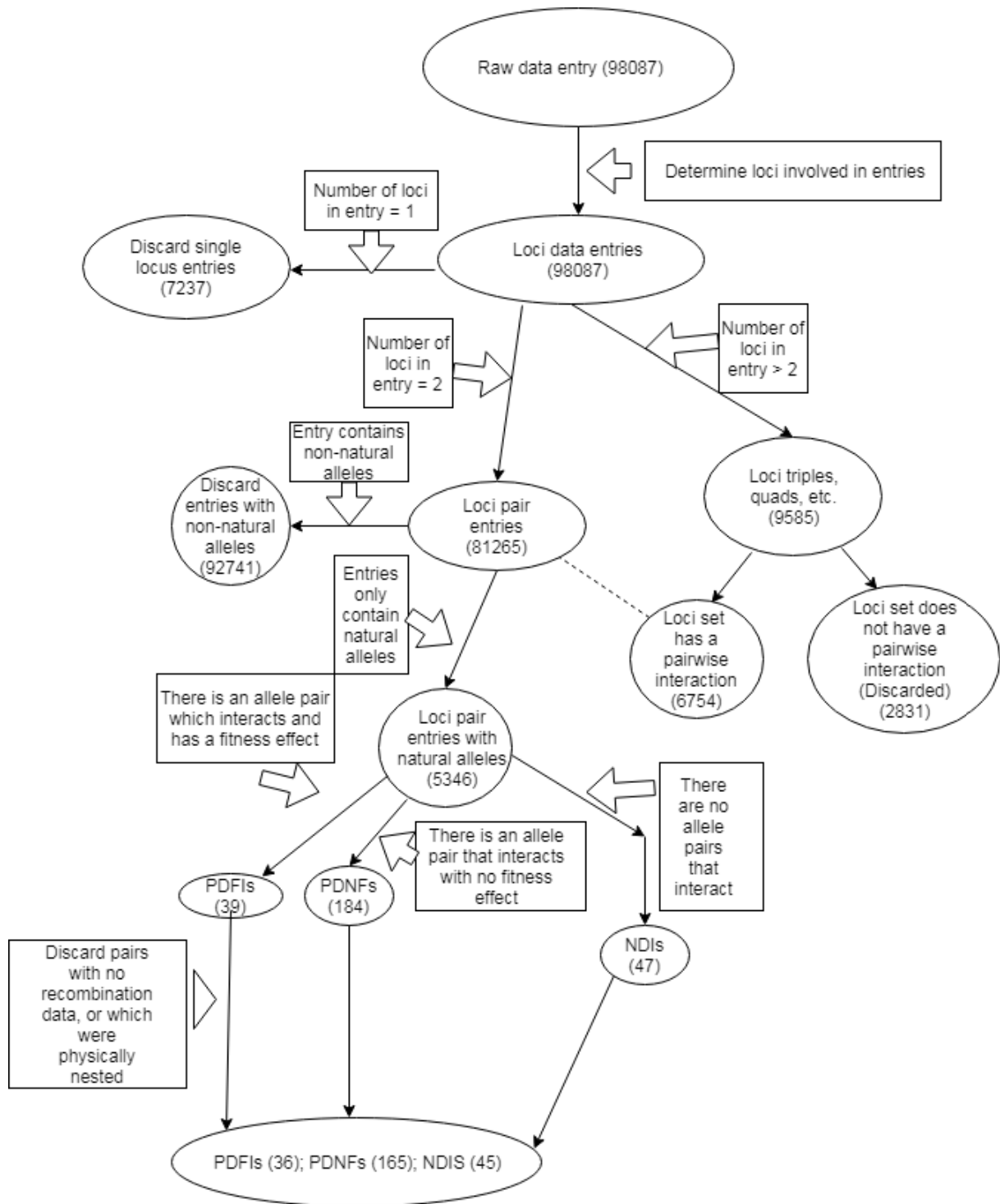
Figures Supplement to ‘Are there Physical Linkages between Genes that have Synergistic Fitness Effects?’

		Locus 1	
		Allele 0	Allele 1
Locus 2	Allele 0	(0,0) Coupling, beneficial or deleterious	(0,1) Repulsion
	Allele 1	(1,0) Repulsion	(1,1) Coupling, non-beneficial or non-deleterious

[TABLE 1] Caption: Table describing the names of the multilocus haplotypes we use in this paper. These are consistent with their usual names in the population genetics literature.

##allele_symbol	allele_FBa1#	Interaction	FBrf#
Abi[P1]	FBa10239154	Abi[+]/Abi[P1] is a suppressor partially of lethal phenotype of Abl[4]/Abl[1]	FBrf0208583
Abl[1]	FBa10028708	Abl[1], N[55e11], Nrt[M54] has eye phenotype	FBrf0102318
Abl[unspecified]	FBa10089849	Abl[unspecified], Nrt[unspecified] has lethal phenotype	FBrf0052941

[TABLE 2] Caption: An example of what a phenotype-gene association looks like when obtained from Flybase. The first column gives a locus name followed by the allele name in square brackets. The second column gives the unique Flybase reference for this allele. The third column gives the names of all alleles and the loci they exist on involved in the interaction as well as the type of interaction. In the first row, third column, there is a lethal interaction between the loci ‘Abi’ and ‘Abl’. In the second row, third column, there is a non-fitness interaction between three loci; ‘Abl’, ‘N’ and ‘Nrt’. In the third row, third column, there is a lethal interaction between the loci ‘Abl’ and ‘Nrt’.



[FIGURE 3] Caption: A flowchart describing the data extraction process and the number of data points at each step. NDIs are cases where both loci affect the phenotype, but there is no epistatic interaction. The bottom discard step applies to all classes.

Component Name	Link Function
----------------	---------------

Intermediate Recombination Mean	Logit
Intermediate Recombination Precision	Logit
Odds of Minimum Recombination	Log
Odds of Maximum Recombination	Log

[TABLE 4] Caption: The link functions for each component of the zero-and-one-inflated beta regression model. Each component submodel is of the form: $link(response) = \sum_i a_i x_i$ where the x_i 's are the explanatory variables and the a_i 's are the parameters of the sub-model.

	PDFI	PDFN	NDI
Chromosome 1	Intercept + Fitness + Interaction	Intercept + Interaction	Intercept
Chromosome 2	Intercept + Fitness + Interaction + Chrom2 + Fitness*Chrom2 + Interaction*Chrom2	Intercept + Interaction + Chrom2 + Interaction*Chrom2	Intercept + Chrom2
Chromosome 3	Intercept + Fitness + Interaction + Chrom3 + Fitness*Chrom3 + Interaction*Chrom3	Intercept + Interaction + Chrom3 + Interaction*Chrom3	Intercept + Chrom3

[TABLE 5] Caption: Coding of the variable parameters for the nine different categories for a component sub-model containing all the explanatory variables stated above (epistasis class, chromosomal identity and their interactions). The 'Intercept' parameter is NOT a true baseline, because the variables are categorical. Four components means that there are a total of 32 parameters to fit if all explanatory variables are included in all the component sub-models. The parameter for a locus pair on chromosome 1 in the NDI class is coded as the intercept. Variables with a '*' between them indicate interactions of those two variables.

Case #	Haplotype (0,0) (beneficial, coupling)	Haplotype (0,1) (repulsion)	Haplotype (1,0) (repulsion)	Haplotype (1,1) (coupling)	LD
1	0	0.5	0.5	0	-0.25
2	0.24	0.26	0.26	0.24	-0.01
3	0.5	0	0	0.5	0.25
4	0.24	0.5	0	0.26	0.0624
5	0.1	0.2	0.2	0.5	0.01

6	0.5	0.5	0	0	0
7	0.1	0	0	0.9	0.09
8	0.1	0.2	0	0.7	0.07
9	0	0.24	0.26	0.5	-0.0624
10	0.02	0.48	0.5	0	-0.24
11	0	0.1	0.1	0.8	-0.01
12	0.02	0.1	0.1	0.78	0.0056

[TABLE 6]

Caption: The initial haplotype proportions for each of the twelve cases, and their linkage disequilibrium, calculated by $p_{00} - (p_{01} + p_{00})(p_{10} + p_{00})$

	PDFI	PDNF	NDI
Chromosome 1	6	40	7
Chromosome 2	12	73	15
Chromosome 3	18	52	23

[TABLE 7] Caption: Numbers of locus pairs in each of the nine possible categories after filtering.

Model	Intermediate recombination mean	Intermediate recombination precision	Odds of minimum recombination	Odds of maximum recombination	GAIC
Null (1)	Intercept	Intercept	Intercept	Intercept	353.67
Full (2)	Chromosomal Identity, Epistasis Class, Interactions	Chromosomal Identity, Epistasis Class, Interactions	Chromosomal Identity, Epistasis Class, Interactions	Chromosomal Identity, Epistasis Class, Interactions	360.23
Selected (3)	Intercept	Intercept	Intercept	Epistasis Class, Chromosomal Identity	327.79
4	Chromosomal Identity, Epistasis Class, Interactions	Intercept	Intercept	Intercept	360.50
5	Epistasis Class	Epistasis Class	Epistasis Class	Epistasis Class	348.11
6	Epistasis Class	Intercept	Intercept	Intercept	357.26
7	Chromosomal Identity	Chromosomal Identity	Chromosomal Identity	Chromosomal Identity	339.57
8	Chromosomal	Intercept	Intercept	Intercept	356.99

	Identity				
--	----------	--	--	--	--

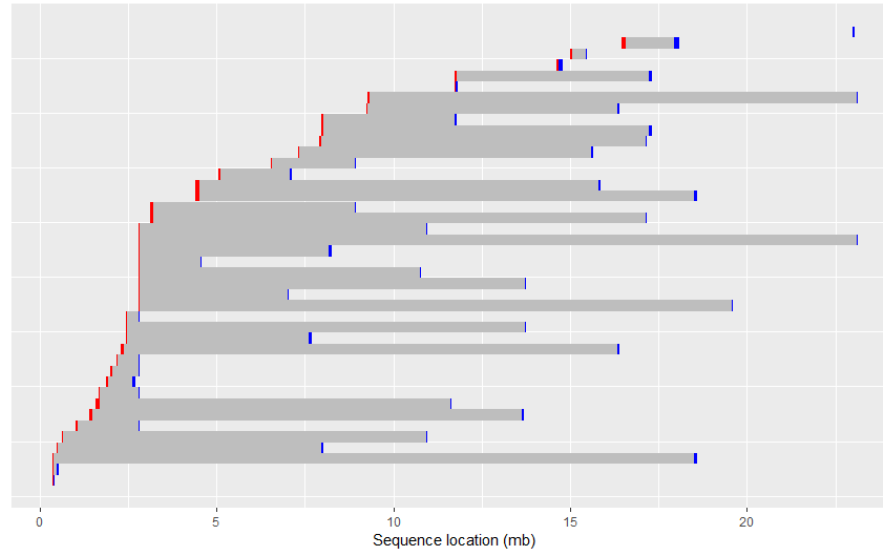
Caption: [TABLE 8]: Comparison of models for the effect of chromosomal identity and epistasis class on the recombination rate between pairs of loci. Column headings refer to components defined in methods (A TEST FOR THE EFFECTS OF EPISTASIS AND CHROMOSOME ON RECOMBINATION RATE): 1) the mean of the intermediate recombination rates (i.e. recombination rates that were not 0 or 0.5); 2) the precision of the intermediate recombination rates 3) the relative probability that $c = 0$ rather than $0 < c < 0.5$; and 4) the relative probability that $c = 0.5$ rather than $0 < c < 0.5$. The names for them are: 1) Intermediate Recombination Mean, 2) Intermediate Recombination Precision, 3) Odds of Minimum Recombination, 4) Odds of Maximum Recombination. The final column shows the GAIC with the best fit (i.e. lowest value) in bold. The rows describe the variables included in each component of the model. An ‘Intercept’-only component means that none of the independent variables had an effect on that component of the recombination rate.

Component	Covariate	Estimate	Confidence Interval
<i>Intermediate Recombination Mean</i>	Intercept	0.407	(0.378,0.434)
<i>Intermediate Recombination Precision</i>	Intercept	0.614	(0.561,0.667)
<i>Odds of Minimum Recombination</i>	Intercept	0.108	(0.057, 0.159)
<i>Odds of Maximum Recombination</i>	Intercept	0.177	(0,0.374)
	Interaction	0.867	(0.126,1.609)
	Fitness	0.179	(0,0.473)
	Chrom2	4.774	(0.241,9.307)
	Chrom3	0.778	(0,1.660)

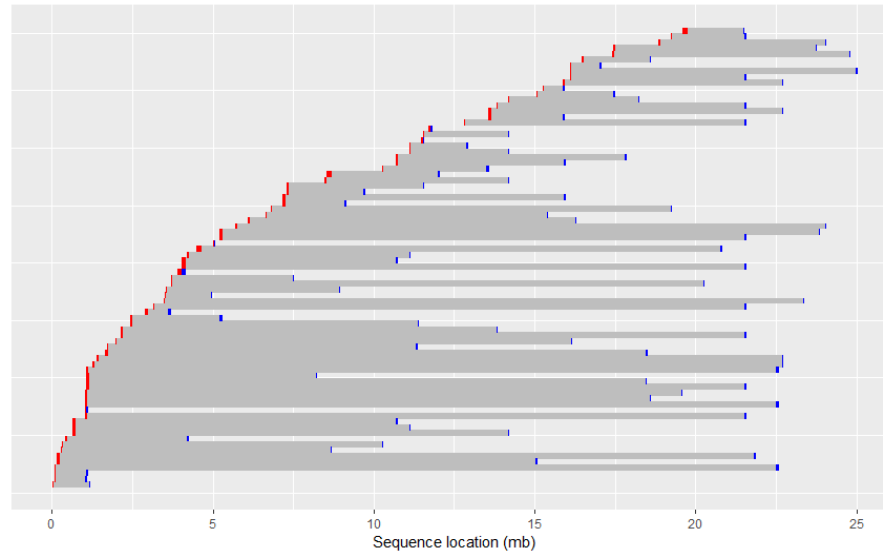
CAPTION: [TABLE 9] : Best model for the effect of chromosomal identity and epistasis class on the recombination rate between pairs of loci. This is model 4 from [TABLE 8] & FIGURE GAIC. Column headings are variables included in the model, its parameter estimate (as defined in appendix II) and its standard error and 95% confidence intervals. Note that ‘component’ refers to any of the four parts of the response variable (recombination rate); these components are shown in the row headings and are defined in the Methods (A TEST FOR EPISTASIS AND CHROMOSOME ON RECOMBINATION RATE): Intermediate Recombination Mean, Intermediate Recombination Precision, Odds of Minimum Recombination, Odds of Maximum Recombination. In contrast to ‘components’, ‘variable’ refers to any of the explanatory variables, that is the Intercept, Interaction, Fitness, Chrom2 and Chrom3 variables (further definition in Methods A TEST FOR EPISTASIS...). The

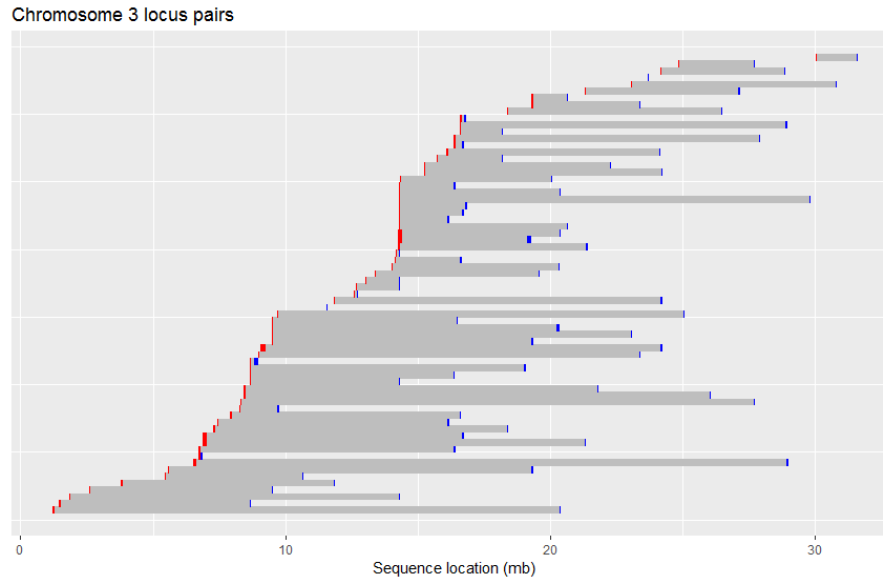
estimates given are transformed back to their original scale using the inverse of the link function and confidence intervals are calculated using the Delta Method.

Chromosome 1 locus pairs

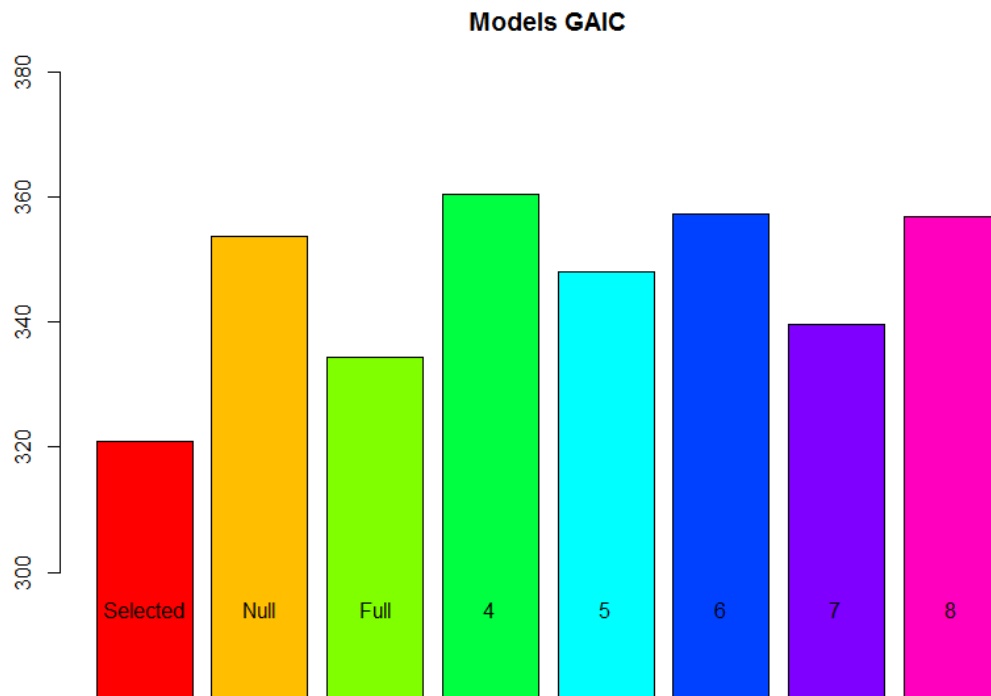


Chromosome 2 locus pairs



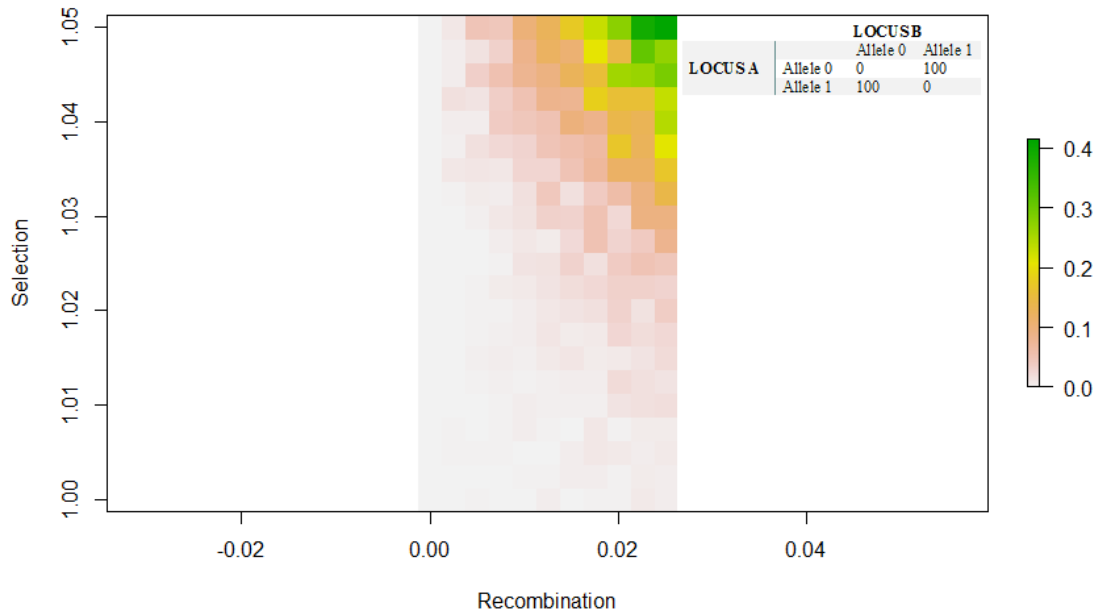


[FIGURE S1] Caption: Locations of the locus pairs on each chromosome. Red indicates the 'leftmost' locus and blue indicates the 'rightmost' locus involved in the particular phenotype. The scale is given in megabases (mb).

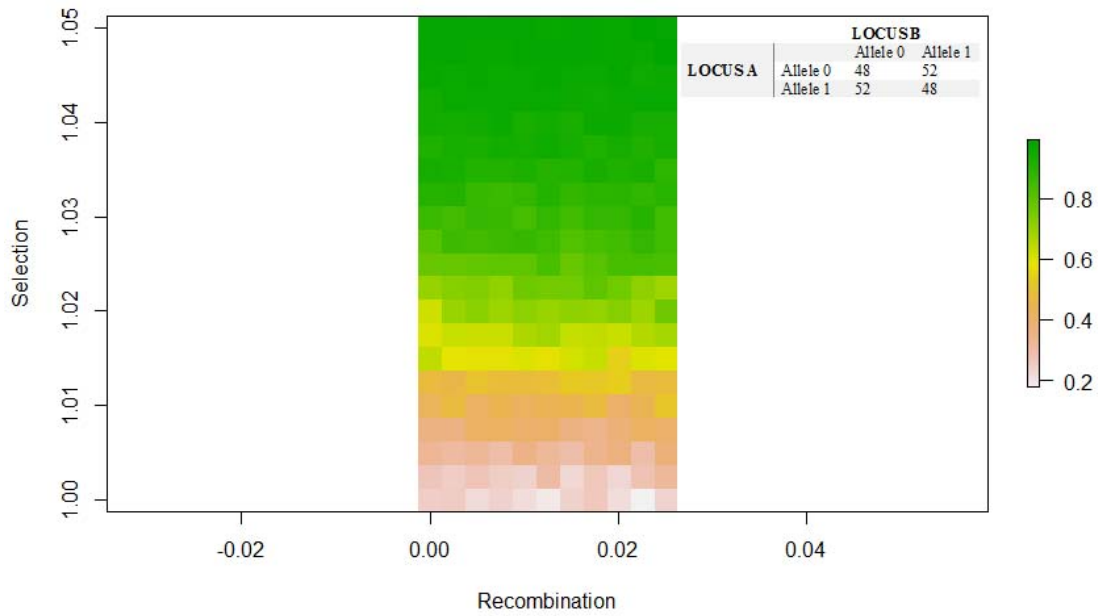


[FIGURE S2] A graphical comparison of the GAICs of the models in [TABLE 6] of the main text.

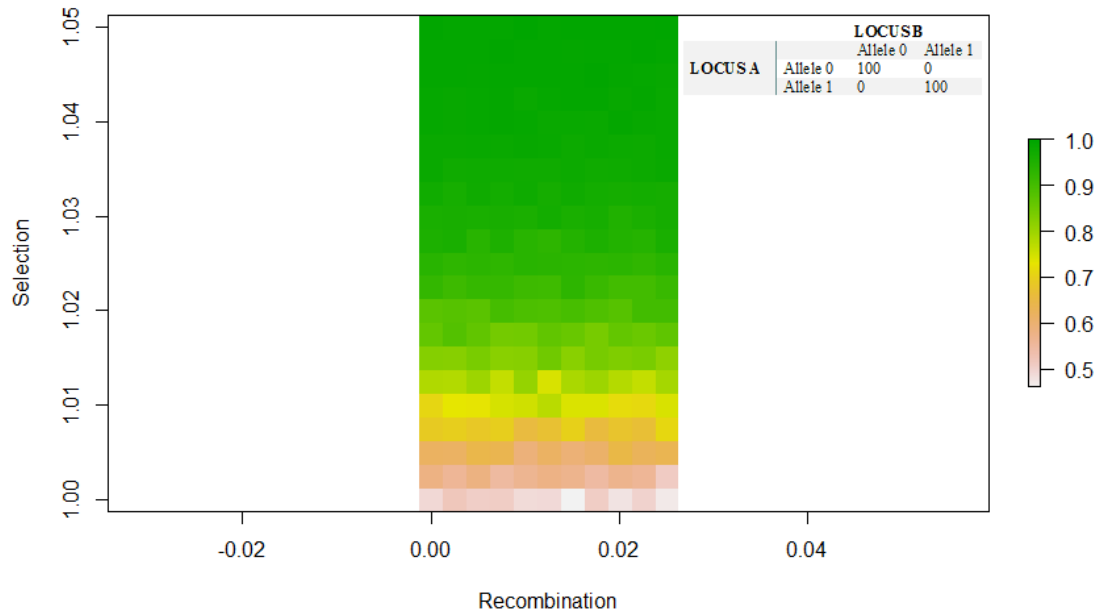
Probability of Fixation for Case 1



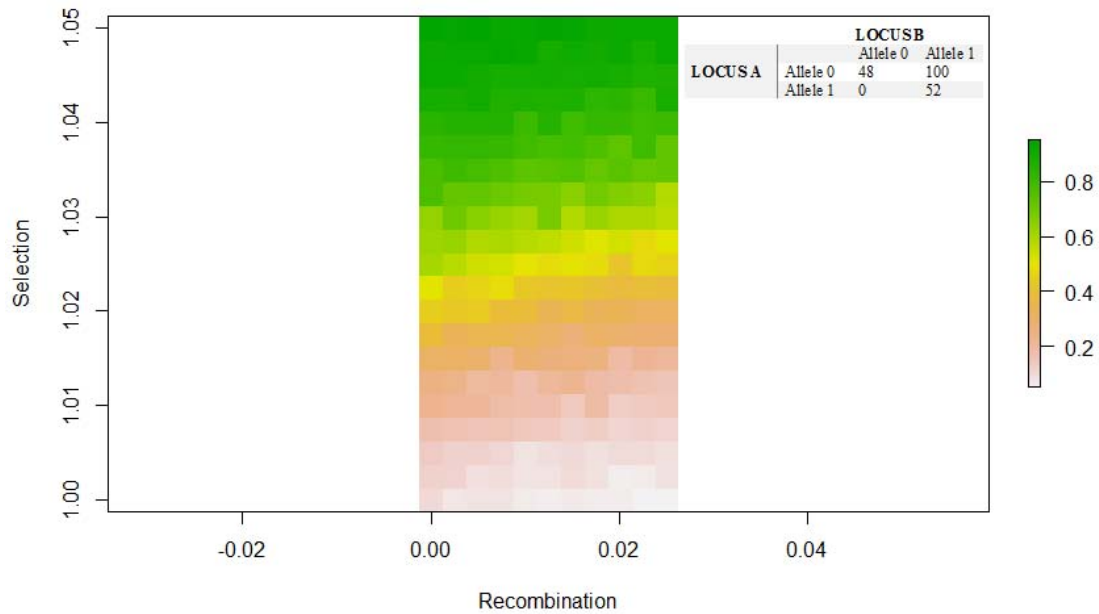
Probability of Fixation for Case 2



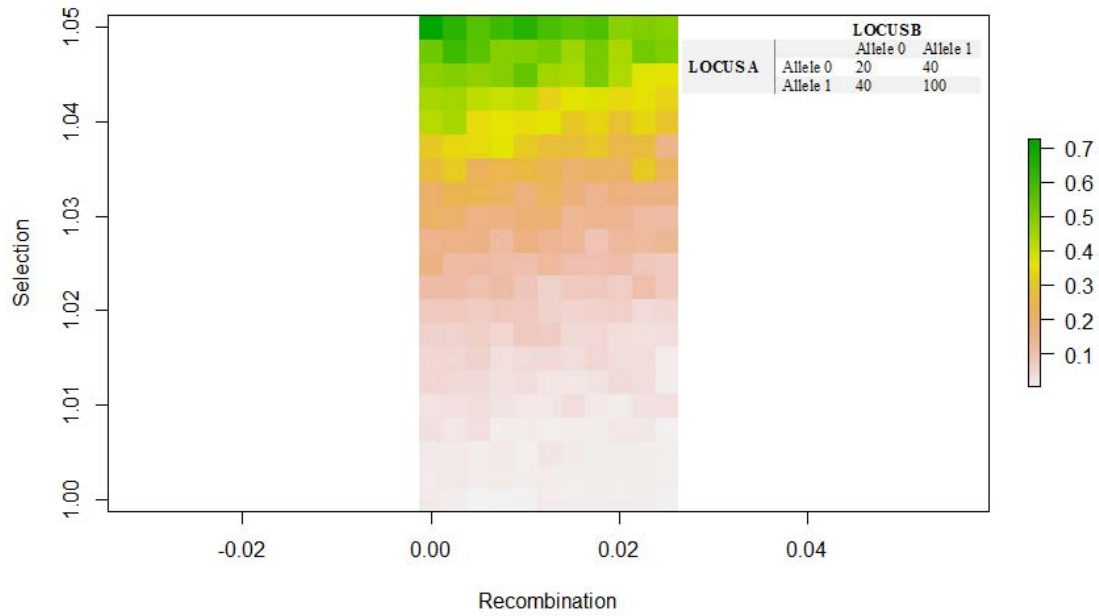
Probability of Fixation for Case 3



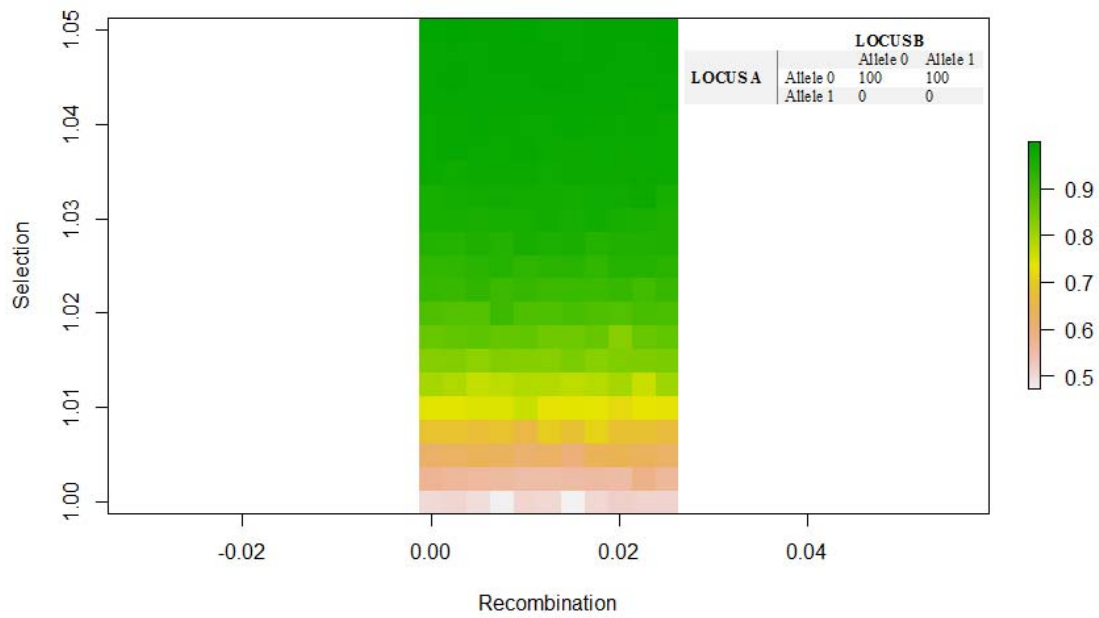
Probability of Fixation for Case 4



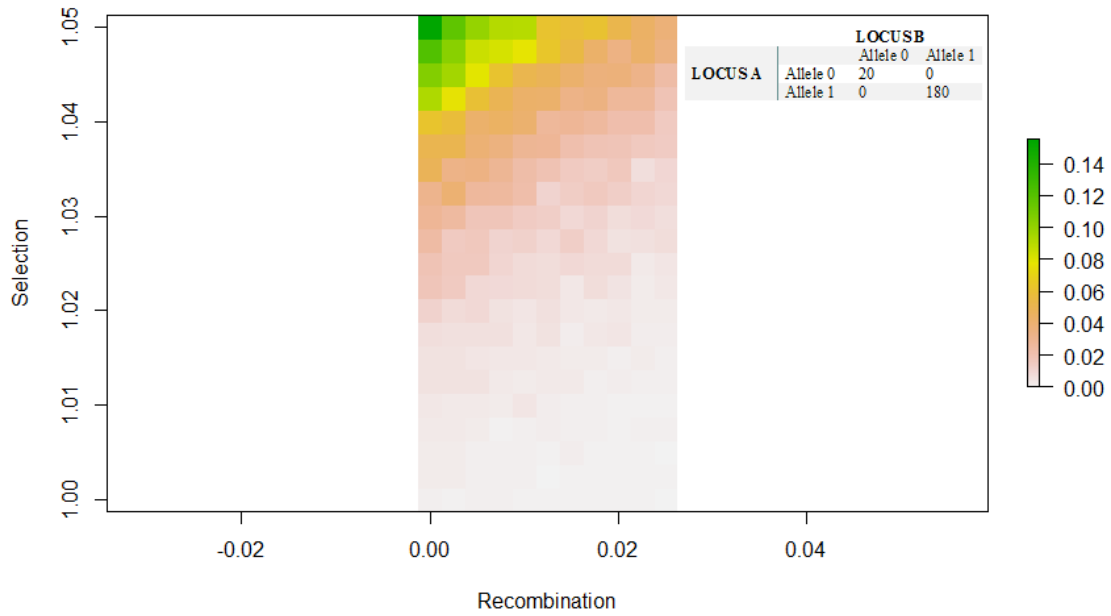
Probability of Fixation for Case 5



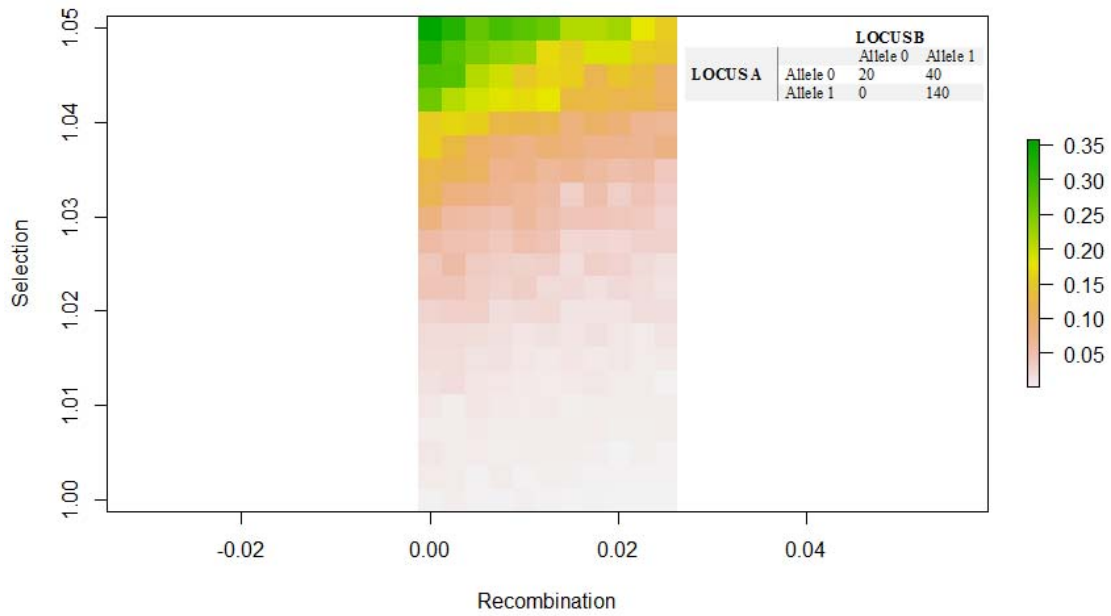
Probability of Fixation for Case 6



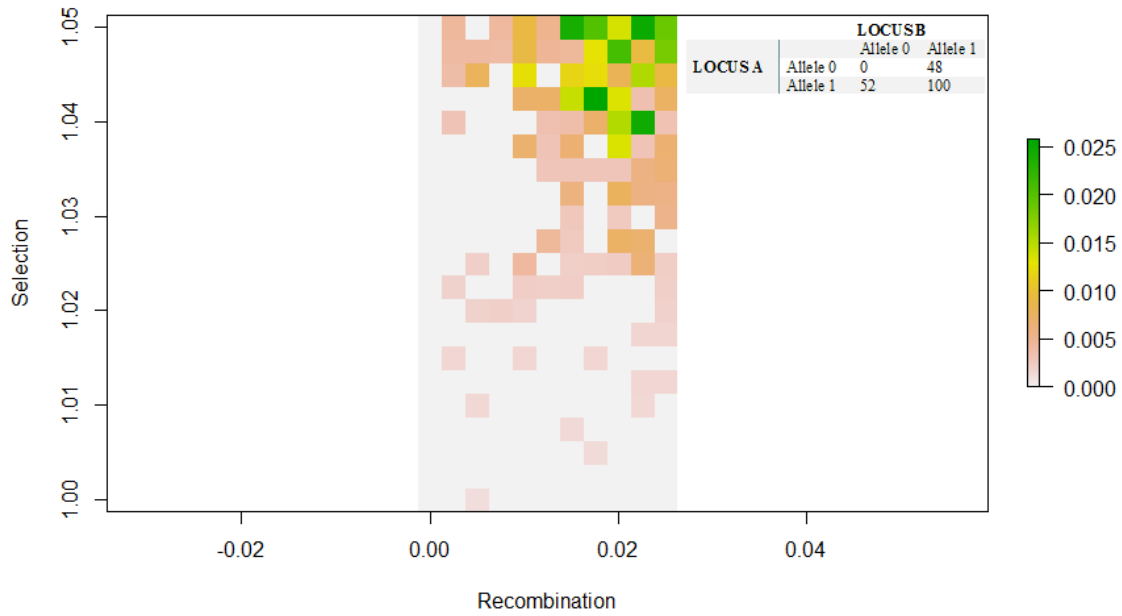
Probability of Fixation for Case 7



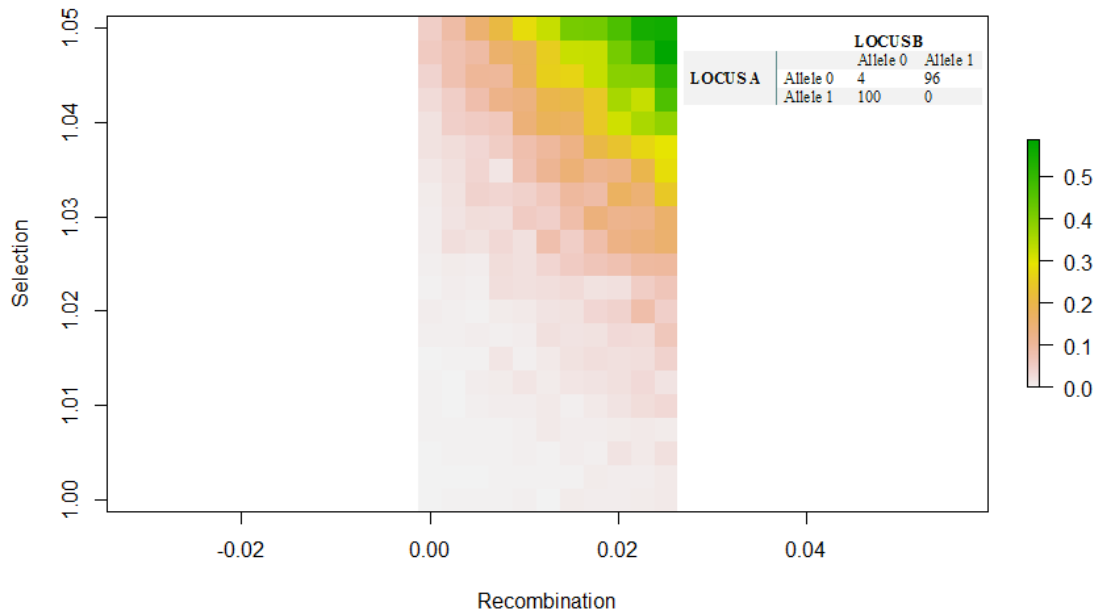
Probability of Fixation for Case 8

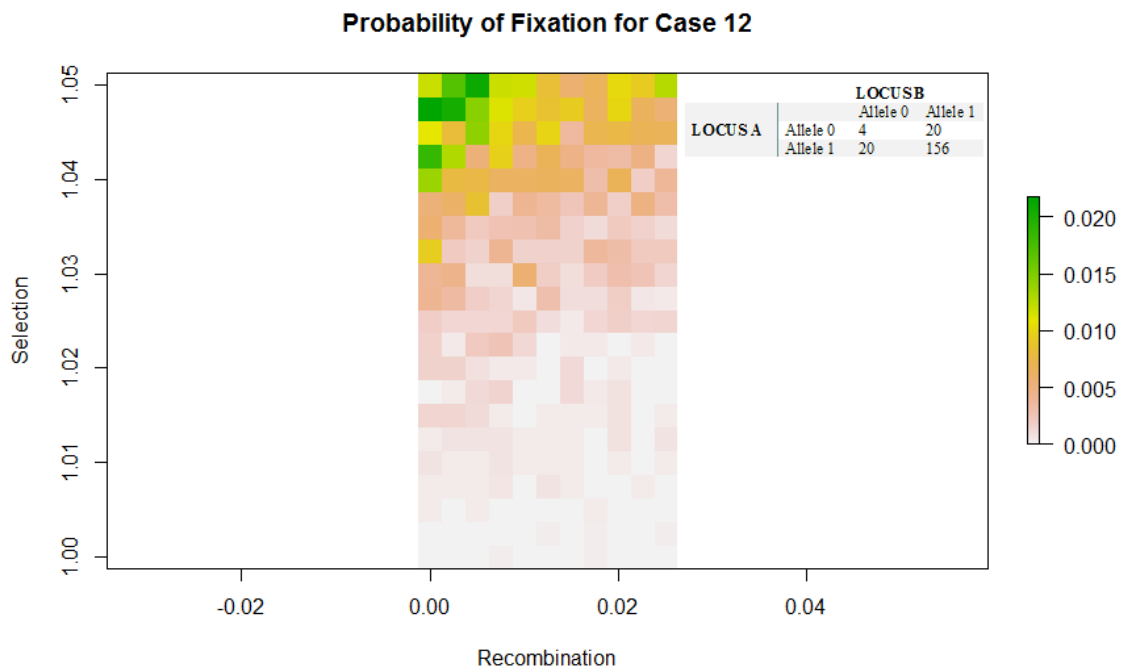
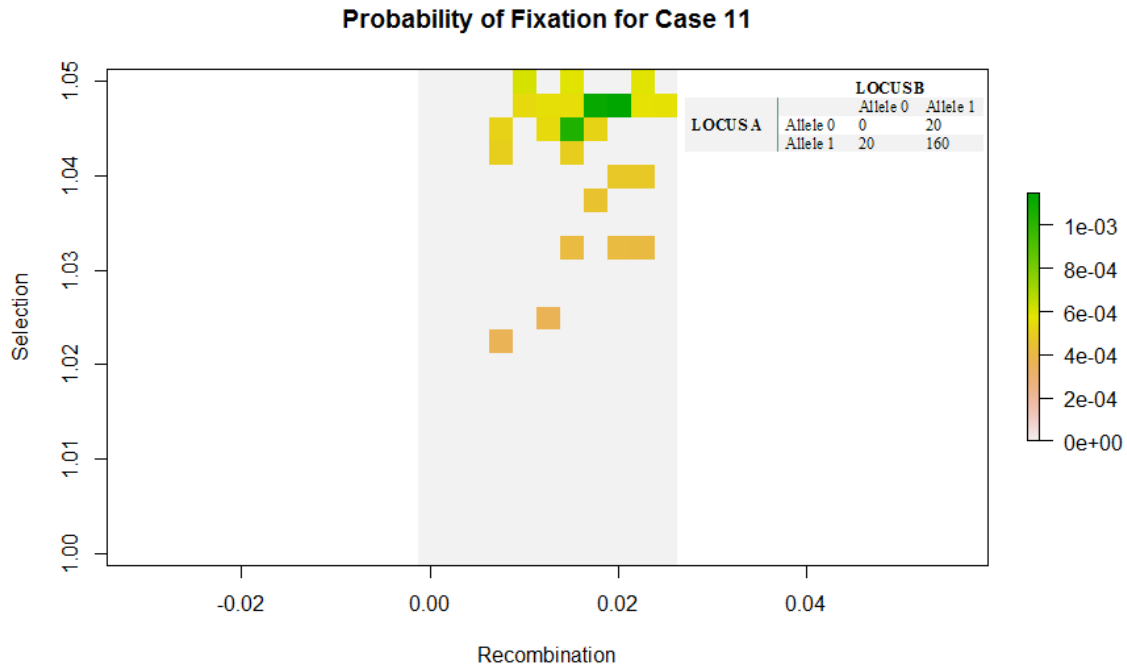


Probability of Fixation for Case 9

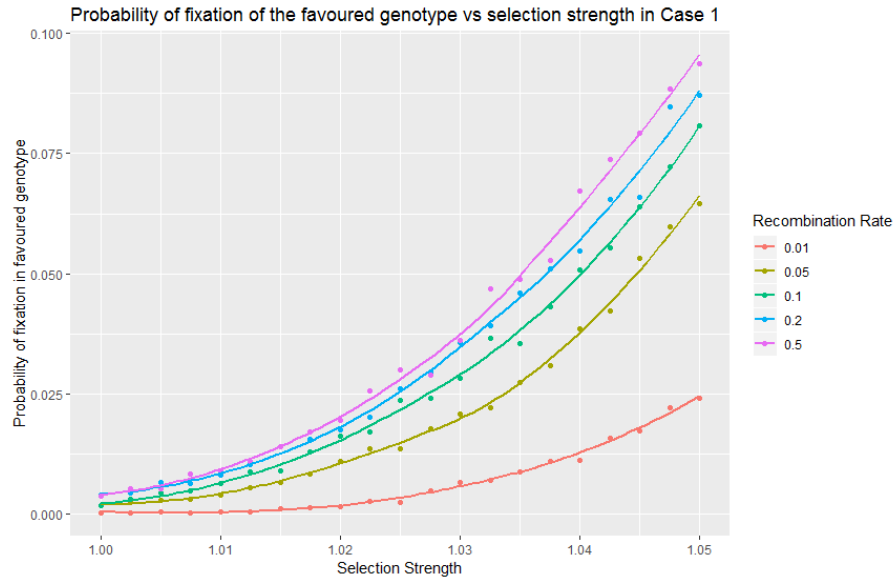


Probability of Fixation for Case 10

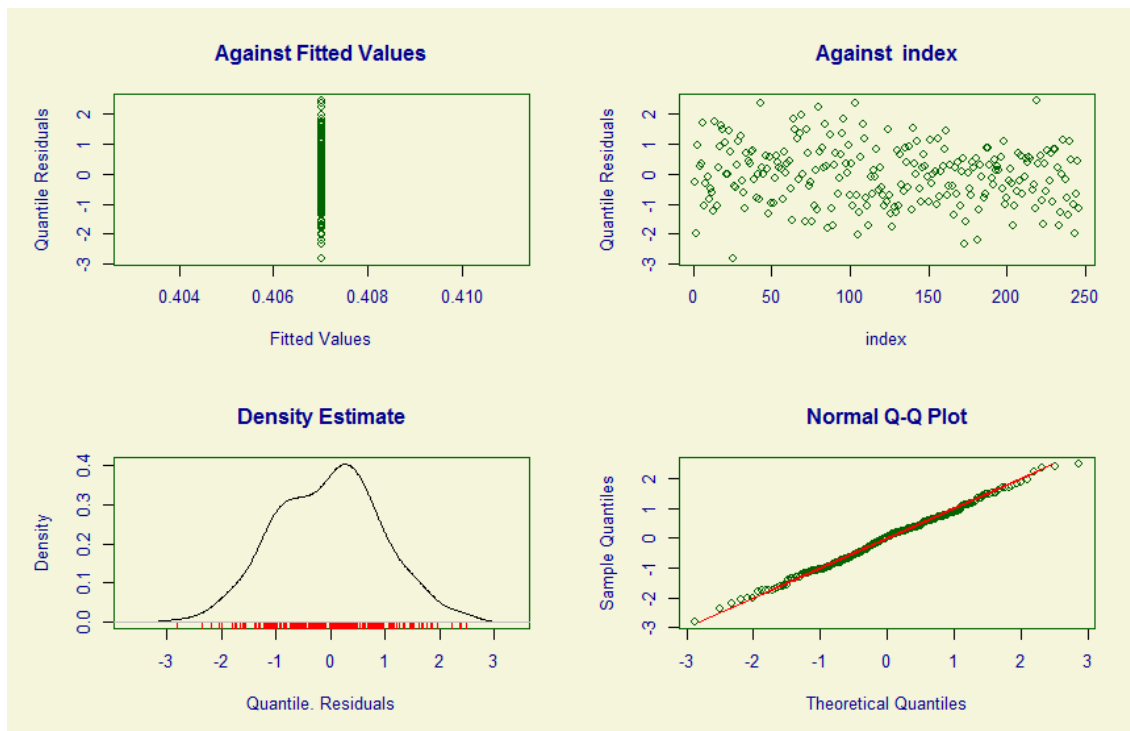




[FIGURE S3] Caption: Rasters of probability of fixation on two axes – recombination and selection. A darker colour indicates a higher probability of fixation (and thus advantage of that recombination and selection combination). The table in the top right corner gives the initial population counts. All graphs are of populations of size 200.



[FIGURE S4] Caption: Graph of probability of fixation with higher levels of recombination in [FIGURE S3, Case 1]. These reflect similar patterns to the lower recombination rates.



[FIGURE S5] Caption: Residual plots for the fitted model. We see that the residuals appear to show no pattern and fit the Q-Q plot very well, indicating a good fit.