Supplement for: Conflict between heterozygote advantage and hybrid breakdown in haplodiploids (and sex chromosomes)

Ana-Hermina Ghenu*, Alexandre Blanckaert*, Roger Butlin, Jonna Kulmuni, and Claudia Bank

Supplementary Methods

Modeling of the life cycle in the absence of assortative mating

In the absence of assortative mating (i.e. $\alpha = 0$), equations (1)-(5) still hold. Nevertheless, tracking genotypes is cumbersome and not necessary in this context (as the production of gametes is independent of their probability of forming a zygote). Therefore, we present here the system of equations used to derive the random mating case. The main difference is where we place the observation point in the life cycle. In the main manuscript, we observe the genotype frequencies of adults and track them over time. Here, we track the frequencies of the gametes (the point in the life cycle when both sexes are haploid), reducing the previous system of 18 variables to 6.

In a given generation n, the frequencies of the male and female gametes are given by $\eta_k(n)$ and $\theta_i(n)$, respectively. Below, we present the modeled life cycle, in the absence of assortative mating, and how frequencies change from one generation to the next:

1. The first step is the formation of the zygote. In the absence of assortative mating, zygote formation depends only on the frequency of the different gametes in the population. In females, the frequency of a zygote carrying the ij genotype, with haplotype i inherited from the mother and j from the father is given by:

$$\theta^z_{ij}(n) = \theta_i(n) * \eta_j(n) \tag{S1}$$

   Males are formed from unfertilized female gametes. Therefore, the frequency of zygote males reflects perfectly the frequency of female gametes:

$$\eta^z_k(n) = \theta_k(n) \tag{S2}$$

2. Individuals of both sex are under viability (or survival) selection. The frequencies of female and male adults are given by $\theta^a_{ij}(n)$ and $\eta^a_i(n_a)$, respectively:

$$\theta^a_{ij}(n) = \frac{w^f_{ij}\theta^z_{ij}(n)}{\sum_{i=1}^4 \sum_{j=1}^4 w^f_{ij}\theta^z_{ij}(n)} \tag{S3}$$

with w^f_{ij} the fitness of the ij genotype. Note that there are no parental effects: $w^f_{ij} = w^f_{ji}$; we maintain the distinction only for modeling convenience and:

$$\eta^a_k(n) = \frac{w^m_k \eta^z_k(n)}{\sum_{k=1}^4 w^m_k \eta^z_k(n)} \tag{S4}$$

with w^m_k the fitness of haplotype k in males.

3. The last step corresponds to the production of gametes. As males are haploid, the gametes produced in the next generation reflect perfectly the distribution of adult males:

$$\eta_k(n+1) = \eta^a_k(n) \tag{S5}$$

Adult females produce eggs. Recombination happens during this stage. The frequency of the female gametes of the next generation, $f^f_i(n+1)$, is given by:

$$\begin{cases} \theta_i(n+1) = & \frac{1}{2}\sum_{j=1}^4 \left(\theta^a_{ij}(n) + \theta^a_{ji}(n)\right) - \frac{\rho}{2}\psi(n) \text{ if } i \in \{1,4\} \\ \theta_i(n+1) = & \frac{1}{2}\sum_{j=1}^4 \left(\theta^a_{ij}(n) + \theta^a_{ji}(n)\right) + \frac{\rho}{2}\psi(n) \text{ if } i \in \{2,3\} \end{cases} \tag{S6}$$

with $\psi(n) = \theta^a_{14}(n) + \theta^a_{41}(n) - \theta^a_{23}(n) - \theta^a_{32}(n)$.

By combining the different equations detailed above (i.e., substituting $\theta^a_{ij}(n)$ in equation (S6) and $\theta^z_{ij}(n)$ in (S3) for the female recursion and $\eta^a_k(n)$ in (S5) as well as $\eta^z_k(n)$ in (S4) for the male case), one obtains the following recursion when making the different fitness terms explicit:

$$
\begin{cases}
\theta_1(n+1) & = -\frac{1}{w^f}\left((\sigma+1)(-(\theta_2(n)+\theta_3(n)-1)\sigma\eta_1(n)+(\theta_2(n)+\theta_3(n)-1)\rho(\sigma+1)\eta_1(n)\right. \\
& \quad +\eta_1(n)+(\theta_3(n)\eta_2(n)+\theta_2(n)\eta_3(n))\rho(\sigma+1))+\theta_1(n)\left(-(2\eta_1(n)+\eta_2(n)+\eta_3(n)-1)\sigma^2\right. \\
& \quad \left.\left.-(4\eta_1(n)+\eta_2(n)+\eta_3(n)-2)\sigma+(2\eta_1(n)+\eta_2(n)+\eta_3(n)-1)\rho(\sigma+1)^2+1)\right) \\
\theta_2(n+1) & = \frac{1}{w^f}\left((\sigma+1)(((\theta_3(n)-1)\eta_1(n)+\theta_1(n)(2\eta_1(n)+\eta_3(n)-1))\rho(\sigma+1)\right. \\
& \quad +\eta_2(n)(\theta_1(n)\rho(\sigma+1)+\theta_3(n)(\sigma\rho+\rho-\sigma)-1))+\theta_2(n)\left(\eta_1(n)\rho(\sigma+1)^2\right. \\
& \quad \left.\left.+\eta_3(n)(\sigma\rho+\rho-\sigma)(\sigma+1)+2\eta_2(n)\sigma-\sigma+2\eta_2(n)\gamma_1-1)\right) \\
\theta_3(n+1) & = \frac{1}{w^f}\left((\sigma+1)(((\theta_2(n)-1)\eta_1(n)+\theta_1(n)(2\eta_1(n)+\eta_2(n)-1))\rho(\sigma+1)\right. \\
& \quad +\eta_3(n)(\theta_1(n)\rho(\sigma+1)+\theta_2(n)(\sigma\rho+\rho-\sigma)-1))+\theta_3(n)\left(\eta_1(n)\rho(\sigma+1)^2\right. \\
& \quad \left.\left.+\eta_2(n)(\sigma\rho+\rho-\sigma)(\sigma+1)+2\eta_3(n)\sigma-\sigma+2\eta_3(n)\gamma_2-1)\right) \\
\eta_1(n+1) & = -\frac{\theta_1(n)}{\theta_2(n)\gamma_1+\theta_3(n)\gamma_2-1} \\
\eta_2(n+1) & = \frac{\theta_2(n)(\gamma_1-1)}{\theta_2(n)\gamma_1+\theta_3(n)\gamma_2-1} \\
\eta_3(n+1) & = \frac{\theta_3(n)(\gamma_2-1)}{\theta_2(n)\gamma_1+\theta_3(n)\gamma_2-1}
\end{cases}
\tag{S7}
$$

with w^f the mean fitness of the female population:

$$
\begin{aligned}
w^f & = 2\left(\theta_3(n)\eta_1(n)\sigma^2-\eta_1(n)\sigma^2-\theta_3(n)\eta_2(n)\sigma^2-\theta_3(n)\sigma+2\theta_3(n)\eta_1(n)\sigma-2\eta_1(n)\sigma-\eta_2(n)\sigma\right. \\
& \quad +2\theta_3(n)\eta_3(n)\sigma-\eta_3(n)\sigma+\theta_1(n)(2\eta_1(n)+\eta_2(n)+\eta_3(n)-1)(\sigma+2)\sigma+\theta_2(n)\left((\eta_1(n)-\eta_3(n))\sigma^2\right. \\
& \quad \left.\left.+(2\eta_1(n)+2\eta_2(n)-1)\sigma+\eta_2(n)\gamma_1)+\theta_3(n)\eta_3(n)\gamma_2-1\right)
\end{aligned}
\tag{S8}
$$

Estimating genotype frequencies from a natural ant population

In order to compare our model with data from the natural, hybridizing Finnish ant population, we estimated the different genotype frequencies of parental *F. polyctena*-like and *F. aquilonia*-like individuals at pre-selection and post-selection life stages for males and females (Fig. S1(a)). We did not estimate the frequencies of introgressed or hybrid individuals. We used the genotype frequencies at different life-stages estimated in Kulmuni and Pamilo (2014) from nine microsatellite loci. For males, eggs were used to estimate pre-selection frequencies; the sum of adults and reproductive fathers was used to estimate post-selection frequencies. For females, eggs were used for pre-selection frequencies and the sum of young and old queens was used for post-selection frequencies. We used two different estimates for the number of parental females: individuals with exactly zero loci heterozygous for an introgressed allele, and individuals with one or more loci homozygous for the parental allele (i.e., the "diagnostic allele" in Kulmuni and Pamilo, 2014). In order to make

these data comparable to our model, we rescaled the genotype frequencies such that 10.3% of the population is from the *F. polyctena*-like sub-population and 89.7% from the *F. aquilonia*-like sub-population, as estimated from the observed abundances of *F. polyctena*-like and *F. aquilonia*-like individuals from nests in the hybrid population collected between 1996-2012 (Table S1). Assuming that the natural population is at equilibrium, we fit the data (Table S2) to the model by calculating the sum of squared differences between the observed data and predicted equilibrium frequencies from 40600 parameter combinations.

Supplementary Results

Fitting the model to natural population frequencies

We compared the pre- and post-selection haplodiploid model (Fig. S1(a)) predictions with the estimated genotype frequencies of the natural, hybridizing *Formica* wood ant population for eggs and reproductive life-stages of males and females (Table S2). The model predictions from the best-fit models are shown in Figures S7 and S8. The best-fit models had parameter values corresponding to single-locus polymorphism or asymmetric coexistence, regardless of how the female frequencies were estimated (Fig. S9). Since these outcomes can occur at a variety of parameter combinations, we were not able to infer any specific parameter estimates other than that large values appear to be preferred for γ_1 and recombination (Fig. S10-S13), consistent with the genomic architecture of the natural population, where multiple incompatibilities are likely to be spread across chromosomes (Kulmuni and Pamilo, 2014). Our model predicts less change in the genotype frequencies before vs. after selection as compared to the differential observed in the data for eggs vs. reproductive adults (Fig. S7(c) and S8(c)).

Supplementary Discussion

We fitted our model to the data from the natural ant population described in Kulmuni and Pamilo (2014) and Table S1 in a rather crude approach. In the fitting procedure, we ignored that the data contain information from marker loci rather than the selected alleles, and we summarized the data in categories to resemble our case of a two-locus interaction. Our model fitting results indicate that the unequal ratio of *F. polyctena*-like and *F. aquilona*-like types that is observed in the natural population could represent a stable equilibrium of asymmetric coexistence. In fact, the high recombination rates among diagnostic alleles and strong prezygotic mechanisms producing within-group zygotes exhibited in the natural population (Kulmuni et al., 2010; Kulmuni and Pamilo, 2014) correspond with an area in the parameter space where asymmetric coexistence can be stably maintained over a wide range of values for female hybrid advantage.

Our model fit does not perform well at predicting the number of introgressed and hybrid females in the population. We were not able to estimate the population frequencies

for introgressed and hybrid females with data from Kulmuni and Pamilo (2014), but we know from Kulmuni et al. (2010) that the vast majority of both *F. polyctena*-like and *F. aquilonia*-like females exhibit some introgression. Contrary to this observation in the natural population, our model fit predicts that introgressed *F. polyctena*-like females should be rare ($< 15\%$) and that pure *F. aquilona*-like females should be only slightly less common than the introgressed *F. polyctena*-like females (Fig. S8).

Supplementary Tables

Table S1: Table of adult frequencies for ant colonies in Långholmen in southern Finland. Individuals were genotyped and assigned to W- or R-type, where W indicates *F. aquilonia*-like parental individuals and R indicates *F. polyctena*-like parental individuals. Mean frequencies of *F. aquilonia*-like and *F. polyctena*-like types were calculated across castes and sampling years.

Sampling year	Nest code	Caste	No. W ind.	No. R ind.	Freq. W	Freq. R
1996	108_Year96	worker	7	0	1	0
1996	110_Year96	worker	14	0	1	0
1996	111_Year96	worker	6	0	1	0
1996	113_Year96	worker	0	4	0	1
1996	114_Year96	worker	6	0	1	0
1996	115_Year96	worker	6	0	1	0
1996	116_Year96	worker	7	0	1	0
1996	117_Year96	worker	1	0	1	0
1996	19_Year96	worker	7	0	1	0
1996	23_Year96	worker	7	0	1	0
1996	24_Year96	worker	7	0	1	0
1996	25_Year96	worker	1	2	0.333	0.667
1996	29_Year96	worker	5	3	0.625	0.375
1996	48_Year96	worker	6	0	1	0
2005	FA1	worker	4	0	1	0
2005	FA10	worker	4	0	1	0
2005	FA11	worker	4	0	1	0
2010	FA12	worker	2	36	0.298	0.702
2005	FA12	worker	15	4		
2005	FA13	worker	3	0	1	0
2005	FA14	worker	4	0	1	0
2005	FA15	worker	4	0	1	0
2005	FA16	worker	4	0	1	0
2008	FA17	queen (gyne)	10	0	1	0
2005	FA17	worker	4	0		
2008	FA18	queen (gyne)	19	0	1	0
2005	FA18	worker	4	0		
2008	FA19	queen (gyne)	2	0	1	0
2005	FA19	worker	4	0		
2005	FA2	worker	4	0	1	0

Continued on next page

Table S1 – continued from previous page

Sampling year	Nest code	Caste	No. W ind.	No. R ind.	Freq. W	Freq. R
2005	FA20	worker	4	0	1	0
2005	FA21	worker	4	0	1	0
2005	FA22	worker	4	0	1	0
2005	FA23	worker	4	0	1	0
2008	FA24	queen (gyne)	1	0	1	0
2005	FA24	worker	4	0		
2008	FA25	queen (gyne)	12	0	1	0
2012	FA26	worker	4	1	0.800	0.200
2012	FA27	worker	3	0	1	0
2012	FA29	worker	4	0	1	0
2005	FA3	worker	12	5	0.706	0.294
2012	FA31	worker	5	0	1	0
2012	FA32	worker	2	1	0.667	0.333
2012	FA33	worker	5	0	1	0
2012	FA35	worker	5	0	1	0
2012	FA39	worker	5	0	1	0
2010	FA4	worker	0	74	0.0645	0.935
2005	FA4	worker	6	13		
2005	FA5	worker	4	0	1	0
2005	FA6	worker	4	0	1	0
2005	FA7	worker	3	0	1	0
2008	FA8	queen (gyne)	1	0	1	0
2005	FA8	worker	4	0		
2005	FA9	worker	4	0	1	0
2014	FAuus2014A	worker	5	0	1	0
2004	Nest2	queen (gyne)	4	0	1	0
2004	Nest3	queen (gyne)	57	4	0.934	0.0667
2004	Nest4	queen (gyne)	6	0	1	0
2004	Nest5	queen (gyne)	6	21	0.222	0.778
Mean					**0.897**	**0.103**

Table S2: Table of estimated genotype frequencies used for data fitting. This data was calculated using the colony data from Table S1 above and the within-group genotype frequencies from Tables 1 & S7 of Kulmuni and Pamilo (2014). Following this same reference, "Introgressed" females refers to indviduals inferred to be purebred based on having no loci heterozygous for introgressed alleles. "Diagnostic" females refers to individuals inferred to be purebred based on having at least one locus homozygous for diagnostic alleles.

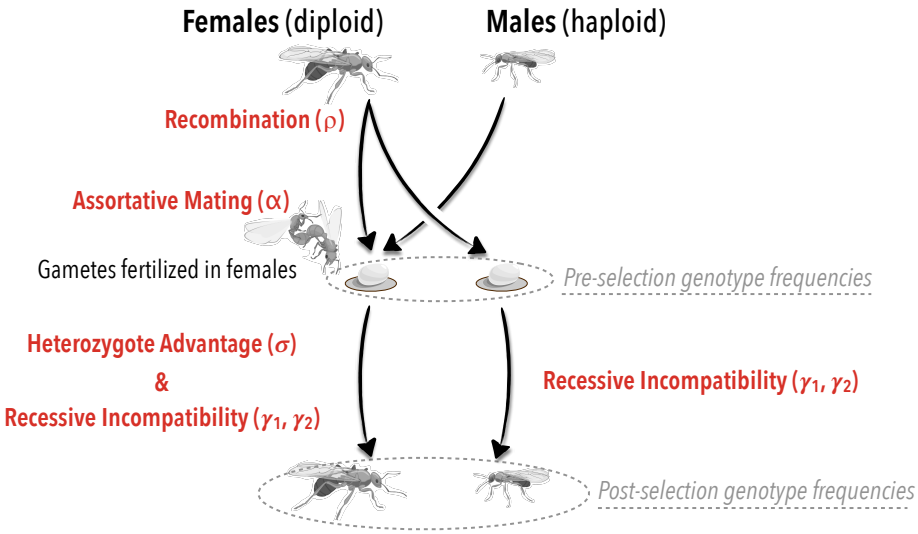
Males		
"Types"	**Before selection (Eggs)**	**After selection (Reproductive males)**
Parental *F. aquilonia*-like	0.637	0.828
Parental *F. polynecta*-like	0.0412	0.102
Summed hybrids	0.322	0.0701

"Introgressed" Females		
"Types"	**Before selection (Eggs)**	**After selection (Young and old queens)**
Parental *F. aquilonia*-like	0.511	0.424
Parental *F. polynecta*-like	0.0216	0
Summed all others	0.467	0.576
"Diagnostic" Females		
"Types"	**Before selection (Eggs)**	**After selection (Young and old queens)**
Parental *F. aquilonia*-like	0.314	0.189
Parental *F. polynecta*-like	0.0834	0.00214
Summed all others	0.603	0.809

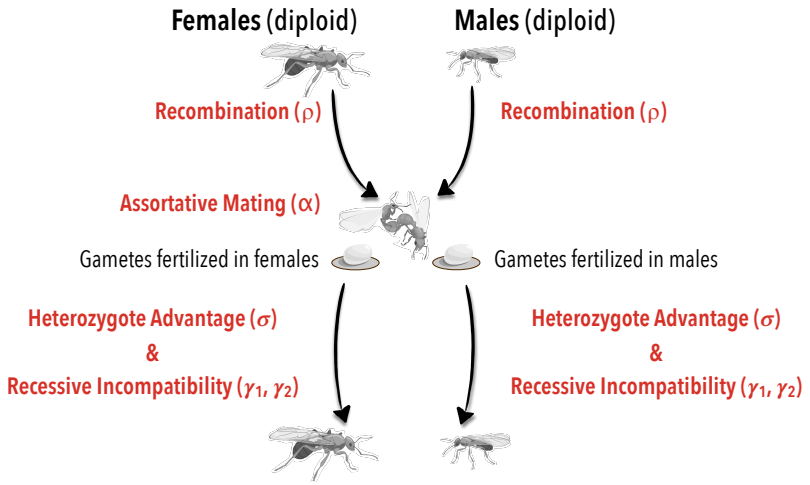
Table S3: Table that demonstrates why single-locus polymorphism can be invaded in diploids. This serves to explain why single-locus polymorphism is never stable in diploids.

Genotype	**Mutant**	**Fitness of the mutant / fitness of the mutated genotype**
$A_+A_+B_+B_+$	$A_-A_+B_+B_+$	$(1+\sigma) > 1$
$A_+A_+B_+B_-$	$A_-A_+B_+B_-$	$(1+\sigma) > 1$
$A_+A_+B_-B_-$	$A_-A_+B_-B_-$	$(1+\sigma)/(1-\gamma_1) > 1$
A_+B_+	A_-B_+	$(1-\gamma_2) < 1$
A_+B_-	A_-B_-	$1/(1-\gamma_1) > 1$

Supplementary Figures

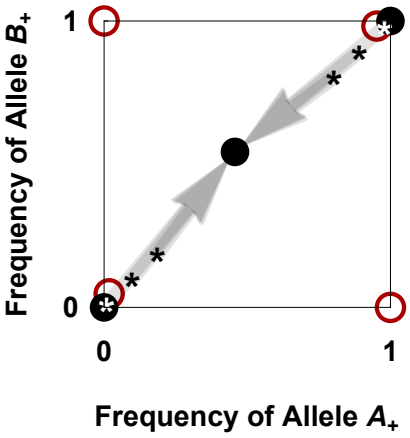


(a) Haplodiploid system

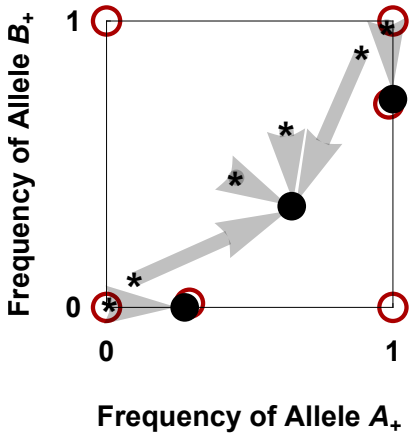


(b) Diploid system

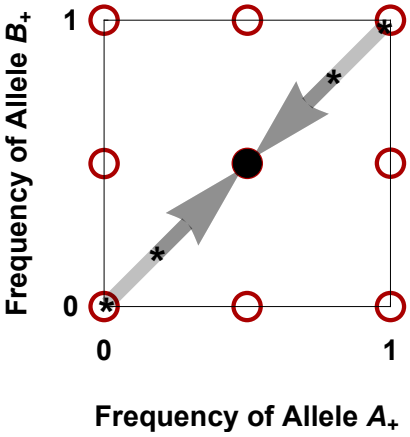
Figure S1: Panel (a) shows the lifecycle of haplodiploids, panel (b) shows the lifecycle of diploids. Events that occur in either females and/or males are illustrated separately on the left and right sides, respectively, in order to highlight the differences between the lifecycles. The pre- and post-selection genotype frequencies for haplodiploids are indicated because they were used for model fitting.



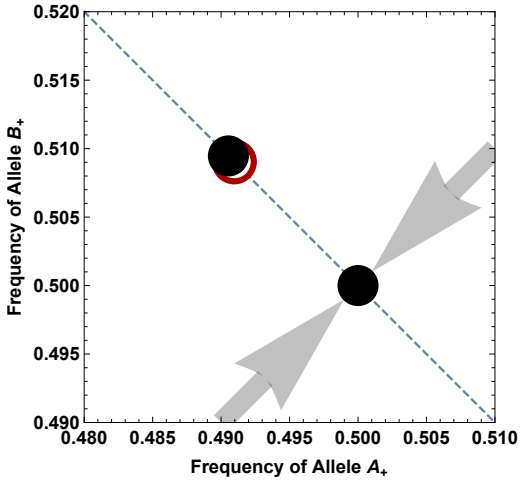
(a) Tristability of Exclusion and Symmetric Coexistence.



(b) Tristability of Single-Locus Polymorphism and Symmetric Coexistence



(c) Symmetric Coexistence with Bistability



(d) Inset of Symmetric Coexistence with Bistability

Figure S2: Phase-plane diagrams of evolutionary scenarios in the haplodiploid model that are not discussed in the main text. Figures are as described in figure 3 except that unstable equilibia are red. Panel (a): $\sigma = 0.05$, $\gamma_1 = 0.2$, $\gamma_2 = 0.11$, $\rho = 0.099$, $\alpha = 0$. Panel (b): $\sigma = 0.143$, $\gamma_1 = 0.46$, $\gamma_2 = 0.083$, $\rho = 0.455$, $\alpha = 0$. Panel (c): $\sigma = 0.2$, $\gamma_1 = 0.019$, $\gamma_2 = 0.011$, $\rho = 0.000001$, $\alpha = 0$. Panel (d): inset of panel (c) with a dashed line at $y = 1 - x$.

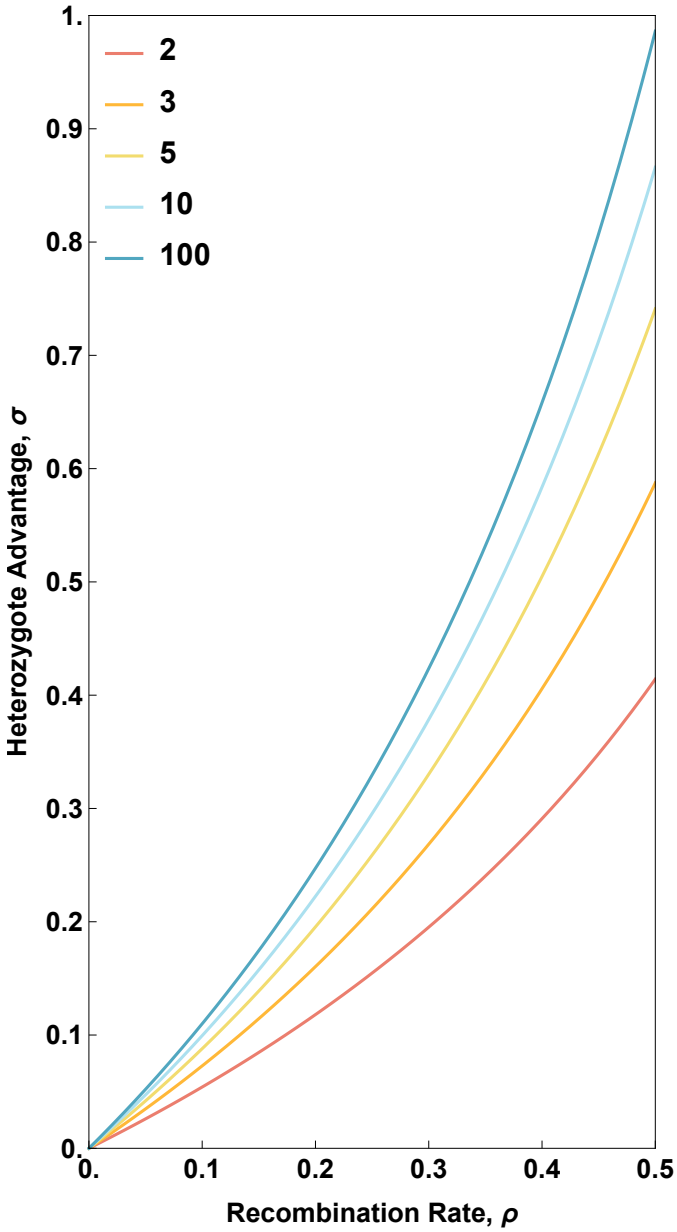


Figure S3: Conditions for stability of exclusion in the network case for haplodiploids with an increasing number of loci, as indicated in the legend. Exclusion represents a locally stable equilibrium for any parameter combination below the curve.

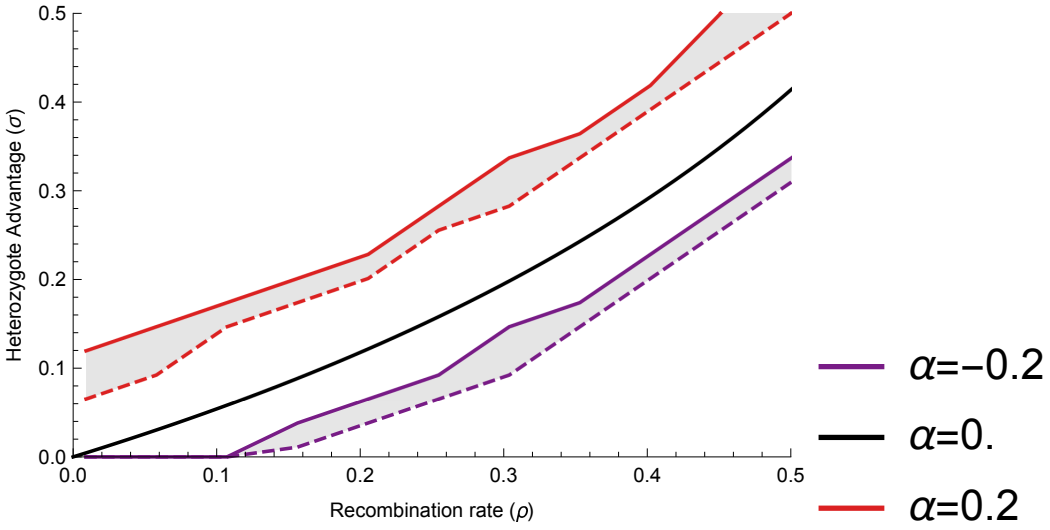
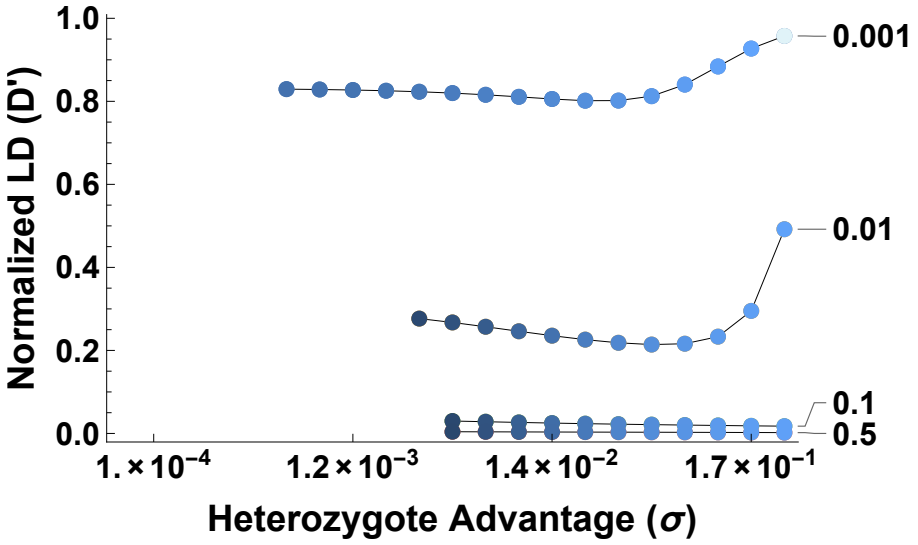
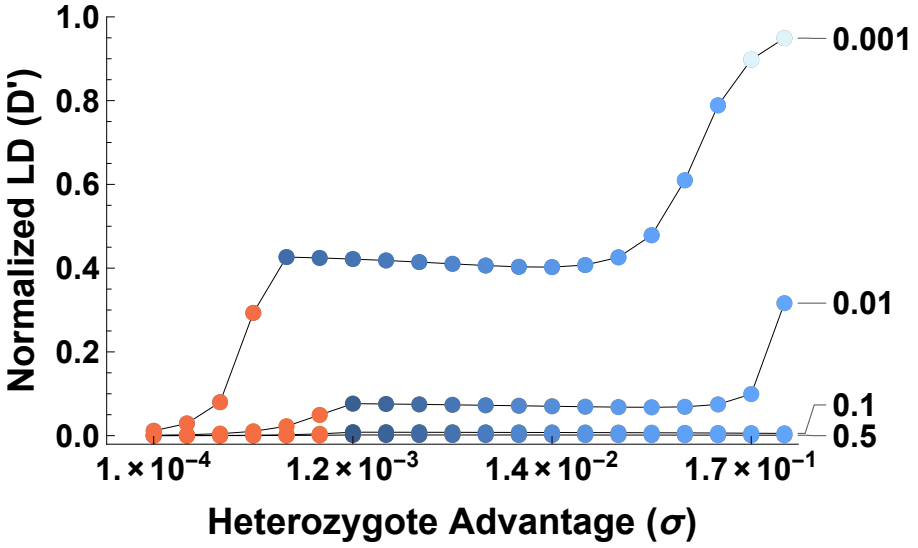


Figure S4: Plot of recombination (ρ) vs heterozygote advantage (σ) for lethal epistasis ($\gamma_1 = \gamma_2 = 1$) and varying strength of assortment, α. Solid lines show values at and above which symmetric coexistence occurs; dashed lines show values at and below which exclusion occurs; the shaded region between the solid and dashed lines shows values where asymmetric coexistence occurs. The result random mating ($\alpha = 0$) is given in equation 7 and so the black line is exactly the same as figure 4(a). This figure was generated with a precision of only ≈ 0.05 units of σ.



(a) Haplodiploids



(b) Diploids

Figure S5: The effect of heterozygote advantage on the normalized linkage disequilibrium ($D' = \frac{LD}{D_{\max}}$) for (a) haplodiploids and (b) diploids under varying recombination rates (values for ρ indicated at right of plot). Each point is shaded according to its evolutionary outcome following the color-scheme in figure 6. Normalization of LD is required since allele frequencies at the two loci can vary considerably (see shading of blue points in the plots above and in figure 6A). No values are shown for the single-locus polymorphism or exclusion outcomes because LD cannot exist when polymorphism is lost at one or both loci. ($\gamma_1 = 0.01, \gamma_2 = 0.002, \alpha = 0$).

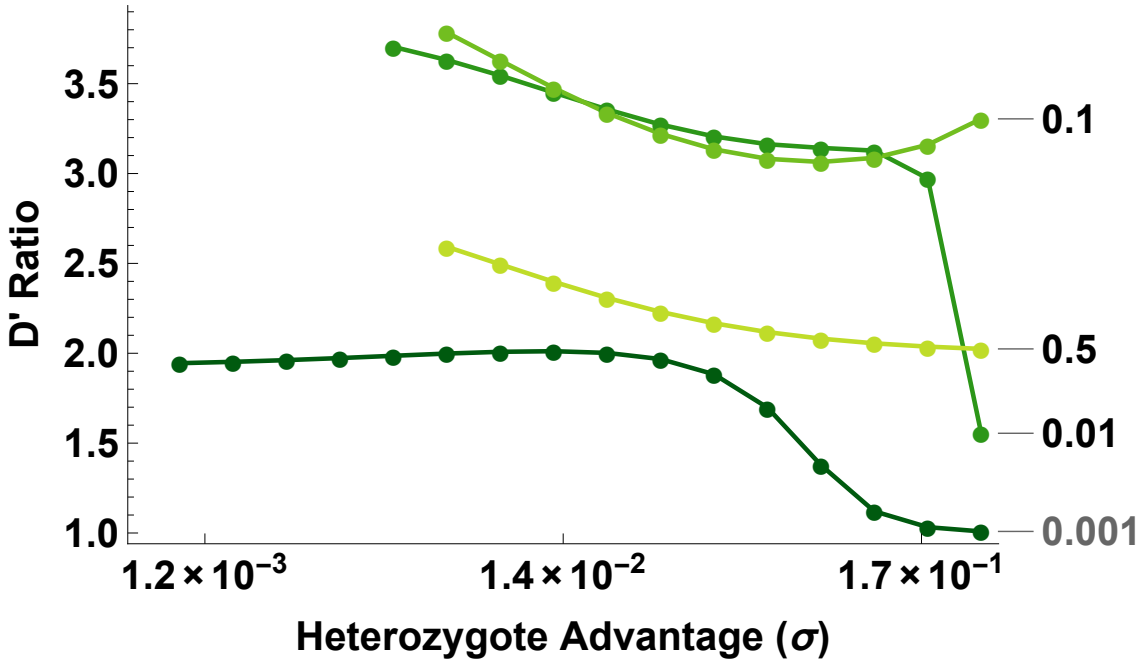
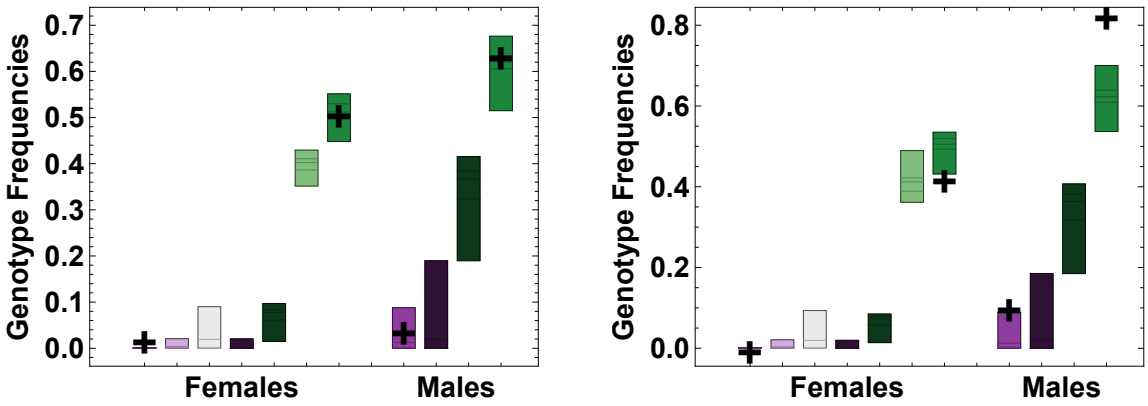
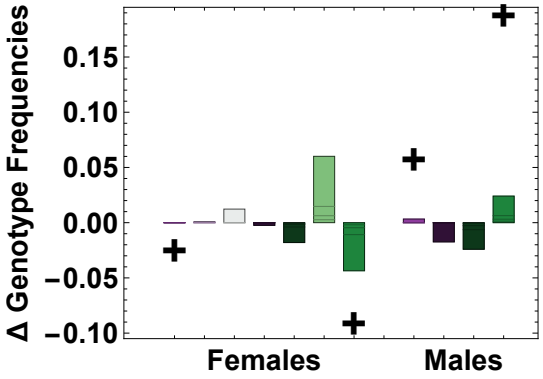


Figure S6: The effect of heterozygote advantage on the ratio of normalized linkage disequilibrium for haplodiploids over diploids ($\frac{D'_{\text{haplodiploids}}}{D'_{\text{diploids}}}$) under varying recombination rates. A ratio greater than 1 indicates more LD in haplodiploids; when polymorphism is maintained at both alleles, haplodiploids always have more or equal LD to diploids. Intermediate recombination values maximize the ratio. ($\gamma_1 = 0.01, \gamma_2 = 0.002, \alpha = 0$)



(a) Before viability selection

(b) After viability selection



(c) Genotype frequency difference

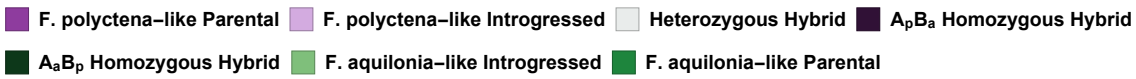
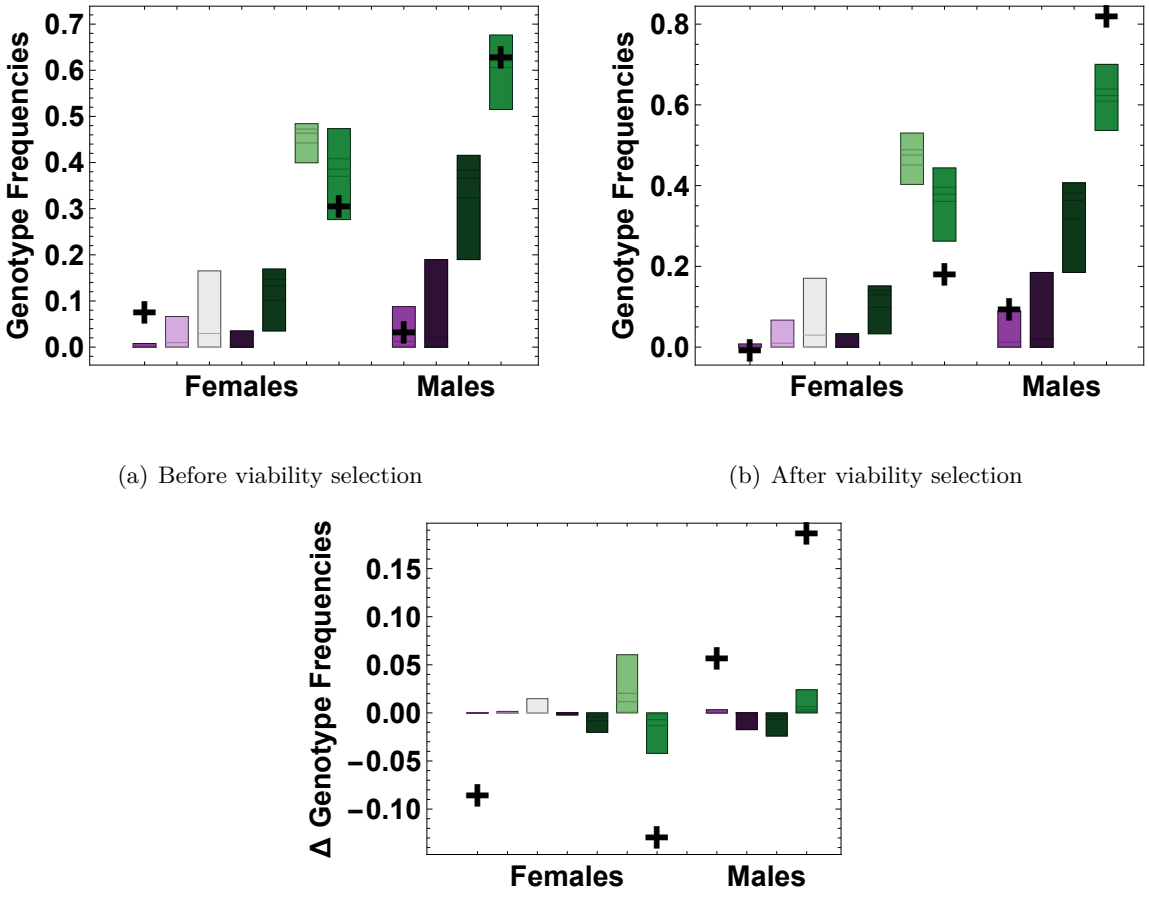


Figure S7: Comparison of model predictions (boxplots) to the data used for fitting the model (+) shows that the model is able to capture the high frequency of *F. aquilona*-like alleles (green shades) in the population. Boxplots show the genotype frequencies for females and males that are predicted from the distribution of the best fitting models. Model predictions before selection, after selection, and delta selection (after selection − before selection). For "introgressed markers" data with random mating ($\alpha = 0$).

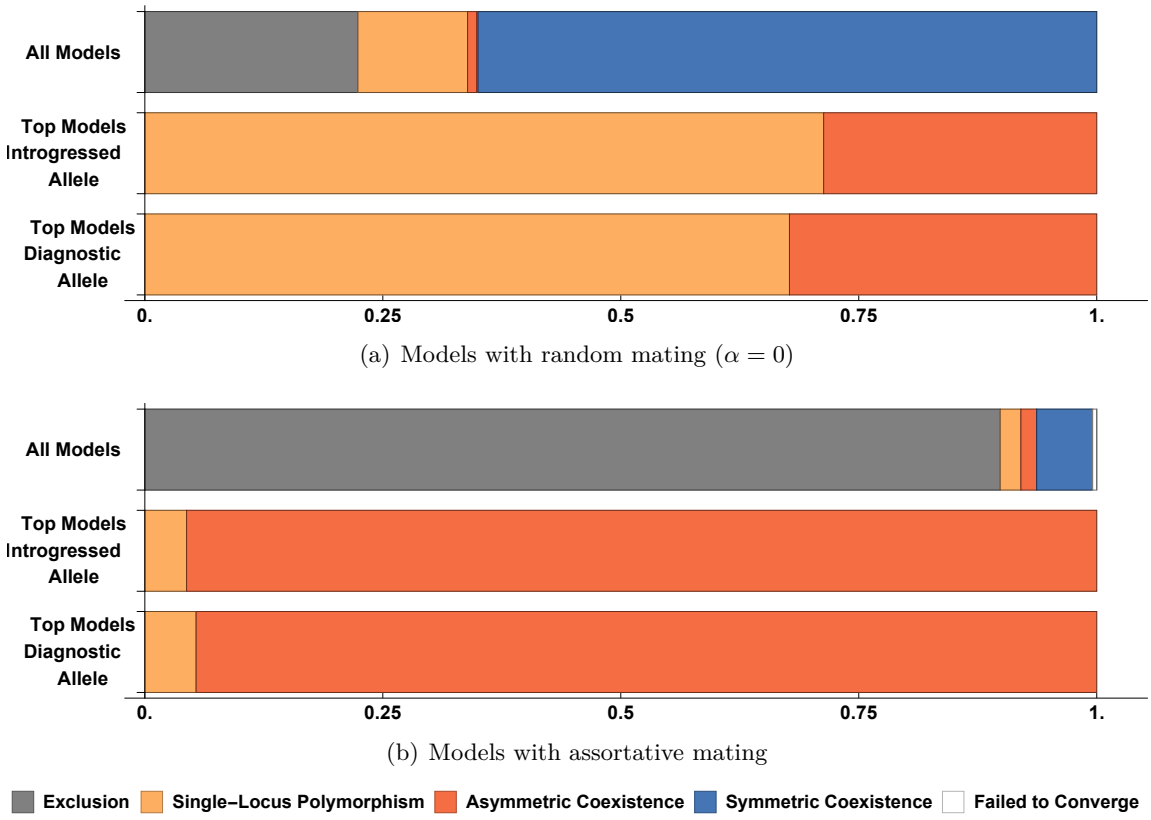


(a) Before viability selection

(b) After viability selection

(c) Genotype frequency difference

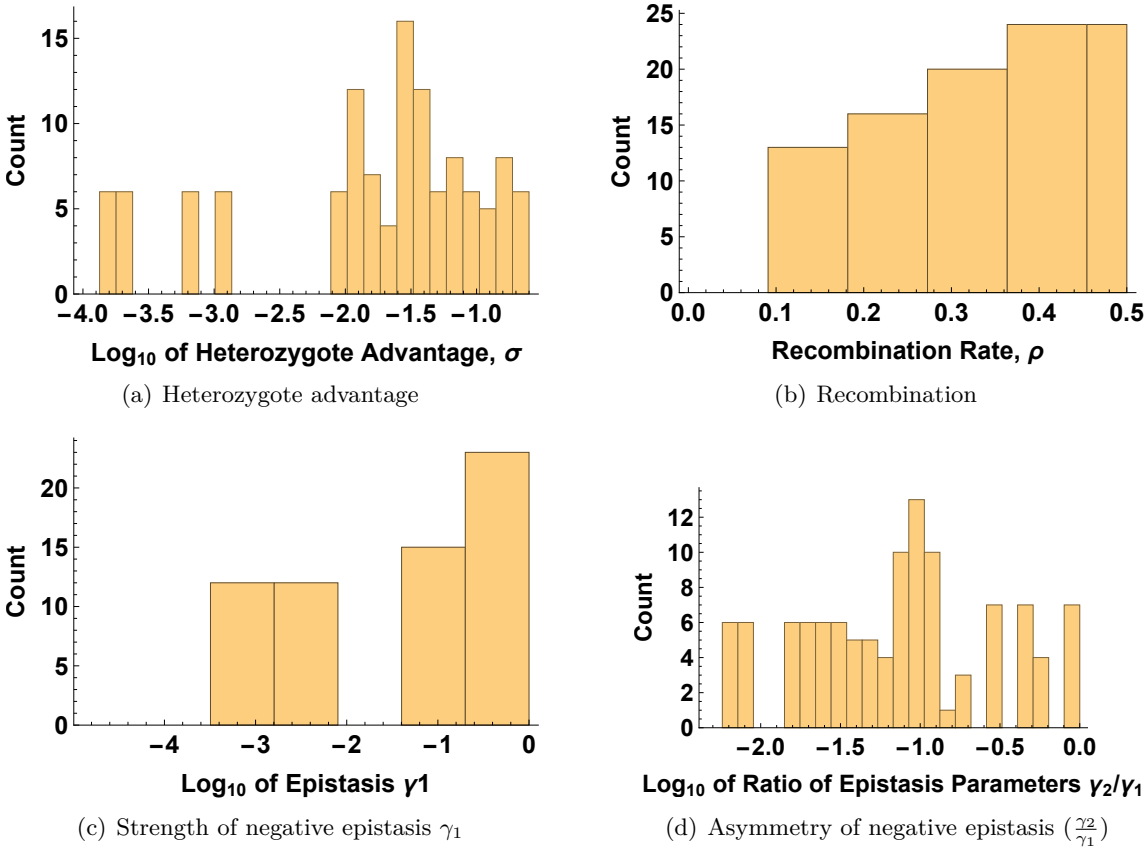
Figure S8: Comparison of model predictions (boxplots) to the data used for fitting the model (+) shows that the model is able to capture the high frequency of *F. aquilona*-like alleles (green shades) in the population. Boxplots show the genotype frequencies for females and males that are predicted from the distribution of the best fitting models. In this case parental genotype frequencies (shown on plot as +) are estimated using individuals with one or more loci homozygous for the parental allele ("diagnostic markers" data). Model predictions before selection, after selection, and delta selection (after selection − before selection). For "diagnostic markers" data with random mating ($\alpha = 0$).



(a) Models with random mating ($\alpha = 0$)

(b) Models with assortative mating

Figure S9: Bar plots show the proportion of parameter combinations that correspond to different evolutionary outcomes. "All models" is for all parameter combinations investigated, "Top models introgressed allele" is for model fitting to the "introgressed markers" data, and "Top models diagnostic allele" is for model fitting to the "diagnostic markers" data.



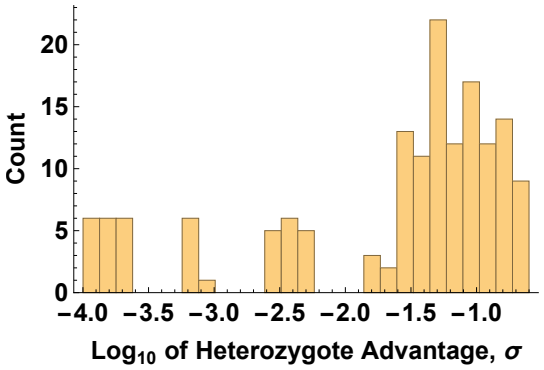
(a) Heterozygote advantage

(b) Recombination

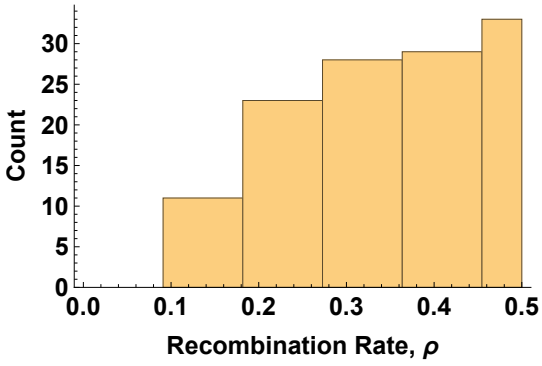
(c) Strength of negative epistasis γ_1

(d) Asymmetry of negative epistasis $(\frac{\gamma_2}{\gamma_1})$

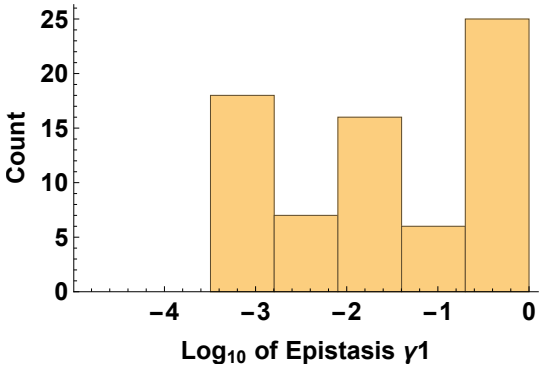
Figure S10: Distribution of the estimated parameter values for the model fitting "introgressed markers" data with random mating ($\alpha = 0$).



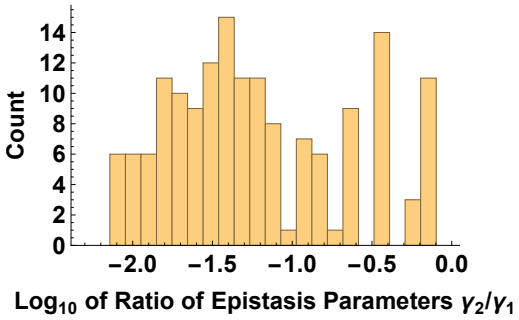
(a) Heterozygote advantage



(b) Recombination

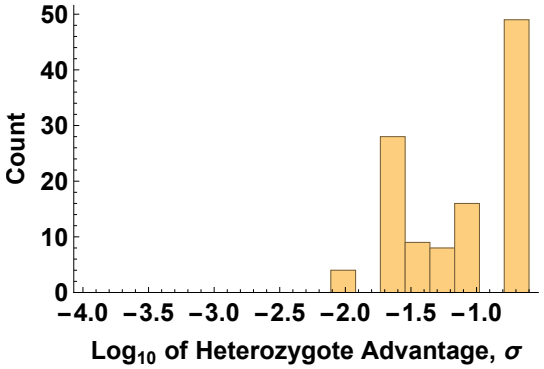


(c) Strength of negative epistasis γ_1

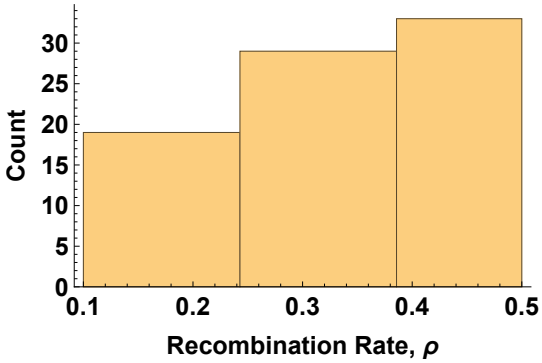


(d) Asymmetry of negative epistasis ($\frac{\gamma_2}{\gamma_1}$)

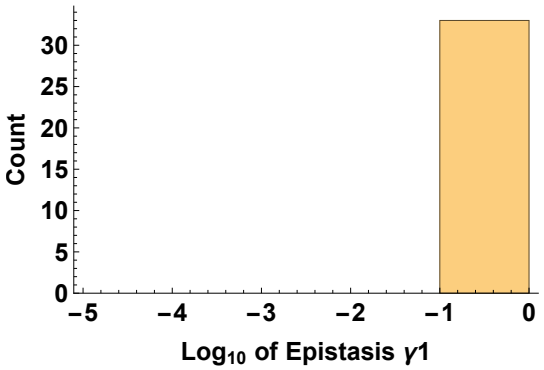
Figure S11: Distribution of the estimated parameter values for the model fitting "diagnostic markers" data with random mating ($\alpha = 0$).



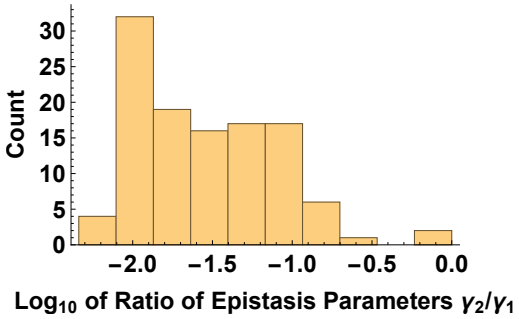
(a) Heterozygote advantage



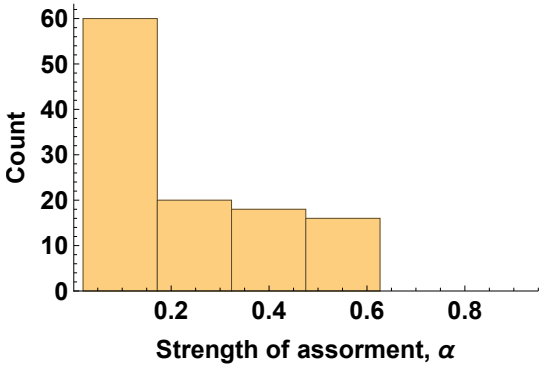
(b) Recombination



(c) Strength of negative epistasis γ_1

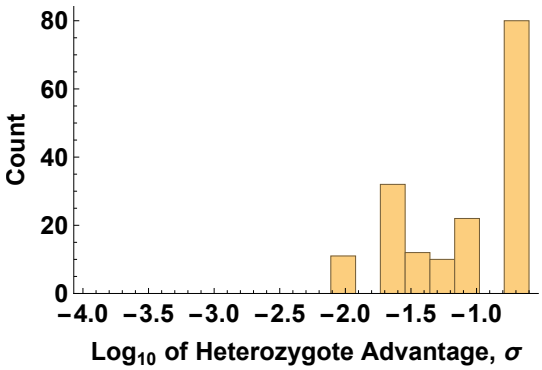


(d) Asymmetry of negative epistasis ($\frac{\gamma_2}{\gamma_1}$)

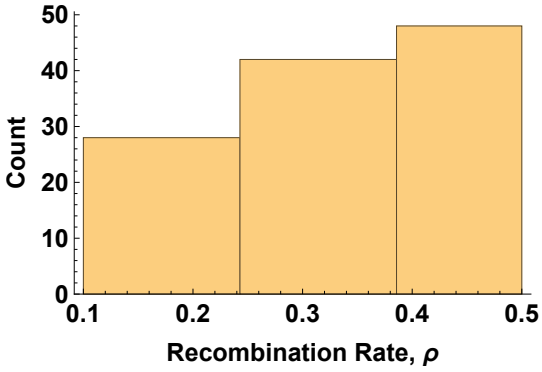


(e) Strength of assortment

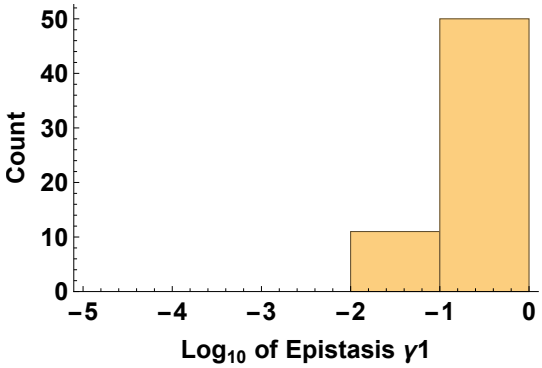
Figure S12: Distribution of the estimated parameter values for the model fitting "introgressed markers" data with assortative mating.



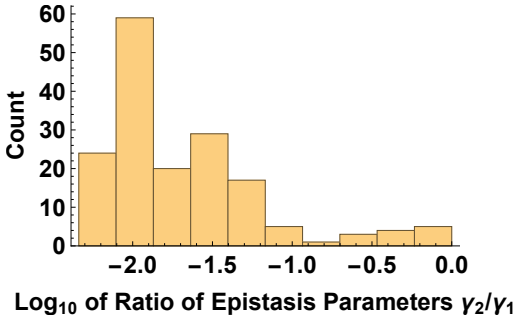
(a) Heterozygote advantage



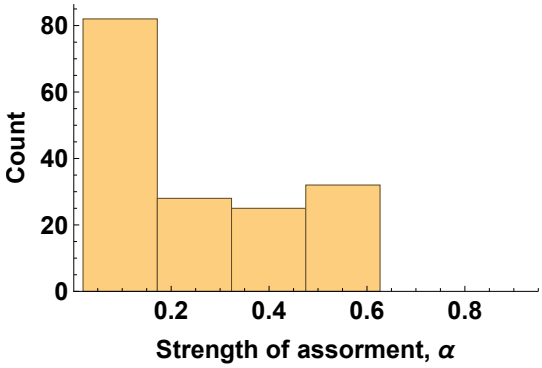
(b) Recombination



(c) Strength of negative epistasis γ_1



(d) Asymmetry of negative epistasis $(\frac{\gamma_2}{\gamma_1})$



(e) Strength of assortment

Figure S13: Distribution of the estimated parameter values for the model fitting "diagnostic markers" data with assortative mating.

References

J Kulmuni and P Pamilo. Introgression in hybrid ants is favored in females but selected against in males. *Proceedings of the National Academy of Sciences*, 111(35):12805–10, 2014. doi: 10.1073/pnas.1323045111.

J Kulmuni, B Seifert, and P Pamilo. Segregation distortion causes large-scale differences between male and female genomes in hybrid ants. *Proceedings of the National Academy of Sciences*, 107(16):7371–6, 2010. doi: 10.1073/pnas.0912409107.