

# Supporting Information

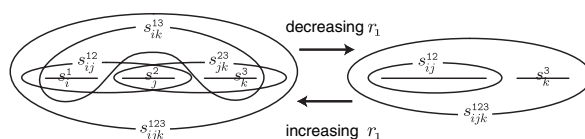
Livnat et al. 10.1073/pnas.0910734106

## SI Methods

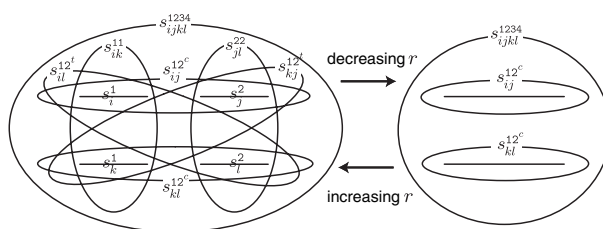
**Diploidy Measures.** Let  $w_{ijk...lmn...}$  be the fitness of a diploid genotype with alleles  $i, j, k, \dots$  at loci 1, 2, 3,  $\dots$  in the first haplotype and alleles  $l, m, n, \dots$  at loci 1, 2, 3,  $\dots$  in the second haplotype. For simplicity, we presently assume no position effects (the fitness does not change when, for any set of loci, the alleles that belong to the first haplotype are switched with their corresponding alleles on the second haplotype; e.g.,  $w_{ijk/lmn} = w_{ijn/lmk}$ ). On the first level, we now have in addition to  $s_i^1, s_j^2, s_k^3, \dots$  also  $s_l^1, s_m^2, s_n^3, \dots$ , and we take them to minimize the square error in  $w_{ijk...lmn...} = s^\emptyset + s_i^1 + s_j^2 + s_k^3 + \dots + s_l^1 + s_m^2 + s_n^3 + \dots$ , or equivalently  $w_{h',h''} = s^\emptyset + \sum_{1 \leq \alpha \leq m} (s_{h'_\alpha}^\alpha + s_{h''_\alpha}^\alpha)$ , where  $s^\emptyset$  is the average of all  $w$  values. On the second level, we have in addition to a term such as  $s_{ij}^{12}$  also the terms  $s_{lm}^{12}, s_{ij}^{12}, s_{lm}^{12}$ , as well as the within-gene interactions terms  $s_{il}^{11}$  and  $s_{jm}^{22}$ . Having fixed the first-level terms, we take all such second-level terms to minimize the error in

$$w_{h',h''} = s^\emptyset + \sum_{1 \leq \alpha \leq m} (s_{h'_\alpha}^\alpha + s_{h''_\alpha}^\alpha + s_{h'_\alpha h''_\alpha}^{\alpha\alpha}) + \sum_{\substack{1 \leq \alpha \leq m-1 \\ \alpha+1 \leq \beta < m}} (s_{h'_\alpha h'_\beta}^{\alpha\beta} + s_{h'_\alpha h''_\beta}^{\alpha\beta} + s_{h''_\alpha h'_\beta}^{\alpha\beta} + s_{h''_\alpha h''_\beta}^{\alpha\beta}). \quad [1]$$

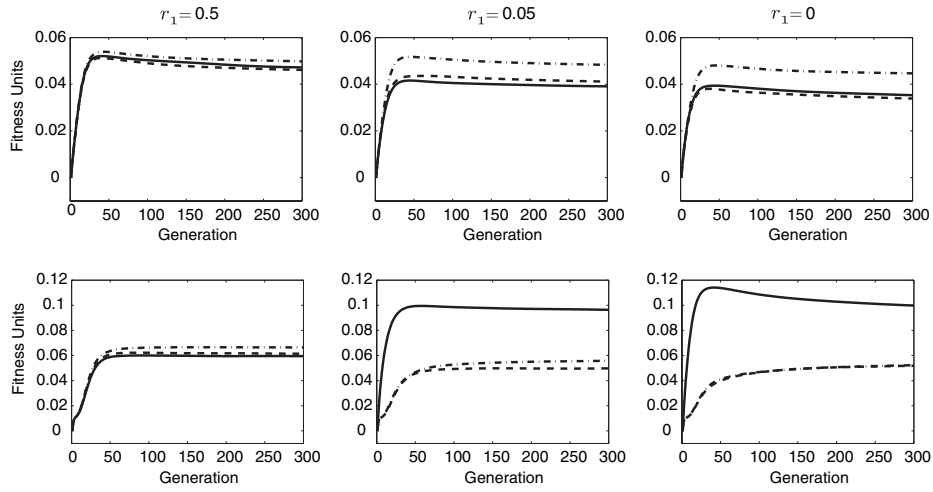
In general, for each level  $\ell$ , we have  $\sum_{0 \leq d \leq \ell/2} \binom{m}{d} \binom{m-d}{\ell-2d}$  different types of  $s$  terms, each defined by a string of superscripts which refers to: (i) a specific set of  $d$  loci of which both alleles are taken, one from each haplotype (these loci appear twice in the superscripts) and (ii) a specific set of  $(\ell-2d)$  loci of which one allele is taken, from either one or the other haplotype (these loci appear once in the superscripts); and each such  $s$  term type is repeated for each single fitness value prediction  $2^c$  times, where  $c$  is the number of loci appearing only once in the superscripts. For example, the third-level  $s$  terms predicting  $w_{ijk/lmn}$  include two instances of the  $s$ -term type 112:  $s_{ij}^{112}$  and  $s_{ilm}^{112}$ . Finally, if  $S_\ell$  is the sum of the terms of level  $\ell$ , we have:  $M_{1,t} = \sum_{ijk...lmn...} P_{ijk.../lmn...} S_{1,t}, M_{2,t} = \sum_{ijk...lmn...} P_{ijk.../lmn...} (S_1 + S_2)$ , etc.  $\Delta_1, \dots, \Delta_m$  are defined as before.



**Fig. S1.** Fusion and separation of genes 1 and 2 with decrease and increase of the recombination rate,  $r_1$ , between them in the three-locus haploid model. The horizontal lines represent genes, and the curved lines group them together per separate effect. Curves that run through a solid line are disallowed (e.g., an  $s_{ik}^{13}$  curve is disallowed on the right side), and curves that include just one of the genes 1, 2, and 3 are either disallowed or omitted for visual clarity. The fusion and separation of genes 1 and 2 are tied to the decrease and increase, respectively, of the relevance of  $s^1$  and  $s^2$  as compared to  $s^3$  and to the increase and decrease, respectively, of the relevance of  $s^{12}$  compared with  $s^{13}$  and  $s^{23}$ .



**Fig. S2.** Fusion and separation of genes with decrease and increase of the recombination rate,  $r$ , in the two-locus diploid model. The horizontal lines represent genes, and the curved lines group them together per separate effect. Curves that run through a solid line are disallowed, and curves that include just one of the four alleles are either disallowed or omitted for visual clarity. The fusion and separation of genes 1 and 2 are tied to the decrease and increase, respectively, of the relevance of  $s^1$  and  $s^2$  and to the increase and decrease, respectively, of the relevance of  $s^{12c}$  compared with  $s^{12e}$ ,  $s^{11}$ , and  $s^{22}$ .



**Fig. S3.** Varying  $r_1$  in the haploid three-locus model with four alleles per locus. (Upper) The across-trial averages  $\langle D^1 \rangle$  (solid),  $\langle D^2 \rangle$  (dashed) and  $\langle D^3 \rangle$  (dot-dash) for  $r_1 = 0.5$  (Left),  $r_1 = 0.05$  (Center) and  $r_1 = 0$  (Right). (Lower) The across-trial averages  $\langle D^{12} \rangle$  (solid),  $\langle D^{13} \rangle$  (dashed) and  $\langle D^{23} \rangle$  (dot-dash) for the same values of  $r$ . Results are based on 2,000 trials. With decreasing  $r_1$ ,  $D^1$  and  $D^2$  decrease relative to  $D^3$ , and  $D^{12}$  increases relative to  $D^{13}$  and  $D^{23}$ . Note also that the differences between  $D^1$  and  $D^2$  and between  $D^{13}$  and  $D^{23}$  reflect the random initial conditions and can be made arbitrarily small by increasing the number of trials, and that  $D^3$  decreases slightly on an absolute scale with decreasing  $r_1$  (this dependence may indicate that recombination between loci 1 and 2 helps to rebalance locus 3 more rapidly).