FILE S1 Supplementary Information for B. Charlesworth 2020 How good are predictions of the effects of selective sweeps on levels on neutral diversity?

S1. Derivation of Frisse et al. formula for the effect of gene conversion

Consider two sites (1 and 2, with 1 to the left of 2) separated by z basepairs; let the rate of initiation of a gene conversion tract be r_g and the mean tract length be d_g. Assume an exponential distribution of tract lengths, with rate parameter $\lambda = 1/d_g$. If a tract is initiated to the left of 1, it has a probability of ½ of moving towards 1. If it is initiated at distance y from 1, the probability that it includes 1 but falls short of 2, resulting in the conversion of 1 but not 2, is given by:

$$\lambda \int_{y}^{y+z} \exp(-\lambda x)\ \mathrm{d}x = \exp(-\lambda y)[1 - \exp(-\lambda z)] \qquad \text{(S1)}$$

The net probability of conversion of 1 from this class of event is then given by:

$$\tfrac{1}{2} r_g \int_0^{\infty} \exp(-\lambda y)\,[1 - \exp(-\lambda z)]\,\mathrm{d}y = \tfrac{1}{2} r_g d_g \left[1 - \exp\left(-z/d_g\right)\right] \qquad \text{(S2)}$$

We also need to consider the class of events where a tract is initiated between sites 1 and moves the right including 2. If it starts at a distance u from 1, the probability that it includes 2 is given by:

$$\lambda \int_{z-u}^{\infty} \exp(-\lambda x)\ \mathrm{d}x = \exp[-\lambda(z-u)] \qquad \text{(S3)}$$

The net probability of a tract of this class converting site 2 but not 1 is:

$$\tfrac{1}{2} r_g \int_0^{z} \exp[-\lambda(z-u)]\,\mathrm{d}u = \tfrac{1}{2} r_g d_g [1 - \exp(-z/\,d_g)] \qquad \text{(S4)}$$

The same argument applies to tracts moving from right to left, so the net probability of recombination between sites 1 and 2, due to one site but not the other being included in a conversion tract, is:

$$2 r_g d_g [1 - \exp(-z/\,d_g)] \qquad \text{(S5)}$$

S2. Probability of recombination during a sweep, for large values of *r/s*

A semi-dominant ($h = 0.5$) autosomal locus with random mating is assumed. Using the notation described in the main text, together with Equations A1a, A2a, and A3a, Equation 10 can be written as:

$$P_r = Re^{\alpha}\gamma^{-\alpha}\int_{q_1}^{q_2} q^{R+\alpha-1}p^{-\alpha}\exp\ -(\alpha\, q^{-1})\ \mathrm{d}q \qquad \text{(S6)}$$

Assuming that $\gamma >> 1$ (so that $\alpha << 1$), and $R > 1$, the integrand can be approximated by:

$$f(q) = \ q^{R-1}p^{-\alpha}[1 - \alpha q^{-1}]$$

Expanding $p^{-\alpha}$ as a binomial series in powers of q, the integrand can be written as:

$$q^{R-1} - \alpha\, q^{R-2} \ + \ \alpha q^{R}[1 - \alpha q^{-1}] + \sum_{i=2}^{\infty}\frac{\alpha(\alpha+1)\ldots(\alpha+i-1)}{i!}[q^{R+i-1} - \alpha q^{R+i-2}] \qquad \text{(S7a)}$$

Neglecting higher powers in α, this can be approximated by:

$$q^{R-1} - \alpha\, q^{R-2} + \ \alpha q^{R} + \alpha\sum_{i=2}^{\infty} i^{-1}\, q^{R+i-1} \qquad \text{(S7b)}$$

This expression can be integrated term by term to yield the following indefinite integral:

$$R^{-1}q^{R} - \alpha(R-1)^{-1}\, q^{R-1} + \ \alpha(R+1)^{-1}q^{R+1} + \alpha\sum_{i=2}^{\infty}[i(R+i)]^{-1}\, q^{R+i} \qquad \text{(S8a)}$$

We can then use $q_1 = p_2 = 1/\gamma$ to obtain the definite integral in Equation S6, again neglecting higher order terms in α:

$$R^{-1} - \ \alpha(R-1)^{-1} + \ \alpha(R+1)^{-1} + \alpha\sum_{i=2}^{\infty}[i(R+i)]^{-1} \qquad \text{(S8b)}$$

Simplifying the first two terms in α in this expression, and multiplying by the terms outside the integral in Equation A4, we obtain the following expression for the probability of recombination during a sweep:

$$P_r \approx e^{\alpha}\gamma^{-\alpha}\{1 - 2\alpha R[(R-1)(R+1)]^{-1} + \alpha R\sum_{i=2}^{\infty}[i(R+i)]^{-1}\} \qquad \text{(S9)}$$

S3. Increase in expected diversity of recombinant alleles at the start of a sweep

At the start of a sweep, the expected diversity (relative to the neutral expectation) is π_1. If there is sufficient recombination that recombinant alleles have an effective population size of B_1 relative to neutrality, their expected relative pairwise diversity relative to neutrality at a given time T since the start of the sweep is given by:

$$\pi(T)B_1^{-1} = 1 - (1 - \pi_1 B_1^{-1})\exp(-B_1^{-1}T) \qquad \text{(S10a)}$$

(Compare with Equation 11 of CC).

For sweep of type i, with duration T_{di} and mean time to the first recombination event T_{ri}, the expected diversity at the time of the recombination event is thus:

$$\begin{aligned}\pi(T_{ri}) &= B_1\{1 - (1 - \pi_1 B_1^{-1})\exp[-B_1^{-1}(T_{di} - T_{ri})]\} \\ &= B_1[1 - (1 - \pi_1 B_1^{-1})E_i]\end{aligned} \qquad \text{(S10b)}$$

If we assume that this is approximately the same as the mean coalescent time associated with multiply-recombinant haplotypes, the diversity at the end of a sweep of this type can be written as:

$$\begin{aligned}\pi_{0i} &\approx (P_{ri} - P_{rsi})B_1[1 - (1 - \pi_1 B_1^{-1})E_i] + P_{rsi}\pi_1 + T_{si} + P_{rsi}T_{di} \\ &= [P_{rsi} + (P_{ri} - P_{rsi})E_i\,]\,\pi_1 + (P_{ri} - P_{rsi})B_1(1 - E_i) + T_{si} + P_{rsi}T_{di}\end{aligned} \qquad \text{(S11)}$$

If the rate of occurrence of sweeps of type i is ω_i, their frequency among all sweeps is equal to $f_i = \omega_i/\omega$. The expected diversity at the end of a sweep is given by:

$$\pi_0 \approx [\bar{P}_{rs} + G]\;\pi_1 + (\bar{P}_r - \bar{P}_{rs})B_1 - GB_1 + \bar{T}_s + \overline{P_{rs}T_d} \qquad \text{(S12a)}$$

where

$$G = \sum_i f_i(P_{ri} - P_{rsi})E_i \qquad \text{(S12b)}$$

From Equation 10 of CC, we also have:

$$\pi_1 = B_1[1 - (1 - \pi_0 B_1^{-1})A) \quad \text{(S13a)}$$

where

$$A = \omega/(1 + B_1^{-1}) \quad \text{(S13b)}$$

Furthermore, by Equation 10 of CC and Equation 20 of the main text, the expected diversity, π, is given by:

$$\pi = [1 - (1 - \pi_0 B_1^{-1})]B_1^2 \omega I(\omega, B_1) \quad \text{(S14)}$$

where I is the integral defined in Equation 10 and section 5 of File S1of CC.

As in CC, these expressions allow π_0, π_1 and π to be determined explicitly. Let $C_1 = \bar{P}_{rs} + G$ and $C_2 = (\bar{P}_r - \bar{P}_{rs})B_1 - GB_1 + \bar{T}_s + \bar{P}_{rs}\bar{T}_d$. Using Equations S10 and S13, we have:

$$\pi_1 = [AC_2 + (1 - A)B_1]/(1 - AC_1) \quad \text{(S15)}$$

Substituting this expression into Equation S12a yields an expression for π_0, which in turn allows π to be determined from Equation S14.

S4. Rates of substitution of new favorable mutations for arbitrary genetic systems

Following Charlesworth *et al*. (2018), let N_H be the total number of haploid copies among breeding adults for a given genetic system, where $N_H = 2N$ for A, $N_H = 3N/2$ for X, and N is the number of breeding adults. Let the effective population size for the genetic system in question be kN_{e0}, where N_{e0} is the effective population size for A. Under the selection model described in the text (Equation 6), the probability of fixation of a new mutation is $4as(kN_{e0}/N_H)$ (Charlesworth 2020), and the rate of entry of new favorable mutations into the population each generation is $N_H up_a$, where p_a is the proportion of mutations that are favourable. The rate of substitution of favorable mutations per generation is thus $4N_H up_a (as\ kN_{e0}/N_H) = 4kN_{e0} up_a (as) = 2up_a(ak\gamma_0)$, where γ_0 is the scaled selection coefficient for autosomal mutations. The expected number of substitutions over one unit of coalescent time for the given genetic system is thus $4kN_{e0} \times up_a(ak\gamma_0) = \pi p_a(ak\gamma_0)$, where $\pi = 4kN_{e0}u$ is the expected neutral nucleotide site site diversity with mutation rate u. The relevant formulae for

a described in the text can be used to obtain the substitution rates for the genetic system of interest.

S5. Effect of the β parameter on R_{XA}

It is sufficient to consider the partial derivative of R_{XA} with respect to $1/\beta$, as given by Equation 19 of the main text. We have:

$$\frac{\partial R_{XA}}{\partial \beta^{-1}} \propto \left(1 - \alpha\, e^{-\beta^{-1}T}\right) k^{-1} e^{-k^{-1}\beta^{-1}T} \; - \left(1 - \alpha\, e^{-k^{-1}\beta^{-1}T}\right) e^{-\beta^{-1}T}$$

$$= k^{-1} e^{-k^{-1}\beta^{-1}T} - e^{-\beta^{-1}T} \; + \alpha(1 - k^{-1}) e^{-\left(1+k^{-1}\right)\beta^{-1}T} \qquad \text{(S16)}$$

As T tends to 0, this expression approaches $(1 - \alpha)(k^{-1} - 1)$. Provided that $k < 1$, this is positive if $\alpha < 1$ (a population expansion), and negative if $\alpha > 1$ (a population contraction). The product of $\exp(\beta T)$ and the derivative is equal to:

$$k^{-1} e^{-(k^{-1}-1)\beta^{-1}T} - 1 \; + \alpha(1 - k^{-1}) e^{-k^{-1}\beta^{-1}T}$$

For $T >> k\beta$ and $k < 1$, this quantity approaches -1, which implies that the derivative must be negative; a negative value will be reached at smaller values of T, the larger α. In the case of a population expansion, this corresponds to a smaller increase in population size. The effect of β on R_{XA} for a given value of k is therefore dependent on the time since the start of the expansion.

S6. Supplementary Tables

Table S1 Results for a single sweep of an autosomal locus in a randomly mating population with $N = 10^6$

$\gamma = 125$

$h = 0.1$, $T_d = 0.169$, approx. $T_s = 0.107$, $P_{nc} = 0.284$

r/s	P_c	P_{nr}	P_r	P_{rs}	T_s	T_c/P_c	T_r
0	0.716	1	0	0	0.114	0.130	–
0.04	0.427	0.354	0.472	0.324	0.0607	0.102	0.0902
0.08	0.280	0.125	0.685	0.304	0.0311	0.0898	0.0821
0.16	0.154	0.0157	0.842	0.137	0.0112	0.0682	0.0698
0.32	0.0849	0.0002	0.915	0.0154	0.0038	0.0447	0.0566
0.64	0.0534	0.0000	0.945	0.0001	0.0017	0.0302	0.0456
1.28	0.0408	0.0000	0.960	0.0000	0.0009	0.0022	0.0369

$h = 0.5$, $T_d = 0.154$, approx. $T_s = 0.0978$, $P_{nc} = 0.118$

r/s	P_c	P_{nr}	P_r	P_{rs}	T_s	T_c/P_c	T_r
0	0.882	1	0	0	0.125	0. 121	–
0.04	0.558	0.463	0.388	0.292	0.0607	0.113	0.0956
0.08	0.373	0.214	0.602	0.331	0.0428	0.104	0.0906
0.16	0.198	0.0457	0.797	0.216	0.0177	0.0852	0.0820
0.32	0.101	0.0021	0.899	0.0482	0.0568	0.0559	0.0698
0.64	0.0655	0.0000	0.935	0.0014	0.0024	0.0368	0.0570
1.28	0.0277	0.0000	0.953	0.0000	0.0014	0.0277	0.0454

$h = 0.9$, $T_d = 0.169$, approx. $T_s = 0.0617$, $P_{nc} = 0.109$

r/s	P_c	P_{nr}	P_r	P_{rs}	T_s	T_c/P_c	T_r
0	0.891	1	0	0	0.143	0.140	–
0.04	0.586	0.523	0.358	0.274	0.0874	0.133	0.112
0.08	0.401	0.274	0.569	0.326	0.0547	0.124	0.107
0.16	0.108	0.0748	0.775	0.234	0.0237	0.103	0.0962
0.32	0.101	0.0056	0.892	0.0629	0.0568	0.0073	0.0795
0.64	0.0664	0.0000	0.935	0.0025	0.0028	0.0423	0.0593
1.28	0.0439	0.0000	0.959	0.0000	0.0013	0.0299	0.0406

$\gamma = 500$

$h = 0.1$, $T_d = 0.0695$, approx. $T_s = 0.0441$, $P_{nc} = 0.192$

r/s	P_{c1}	P_{nr}	P_r	P_{rs}	T_s	T_c / P_{c1}	T_r
0	0.808	1	1	0	0.0533	0.0495	–
0.04	0.274	0.136	0.701	0.338	0.0132	0.0417	0.0314
0.08	0.123	0.0188	0.874	0.175	0.0044	0.0334	0.0267
0.16	0.0496	0.0004	0.951	0.0262	0.0011	0.0216	0.0216
0.32	0.0257	0.0000	0.975	0.0004	0.0003	0.0130	0.0175
0.64	0.0175	0.0000	0.983	0.0000	0.0002	0.0090	0.0145
1.28	0.0136	0.0000	0.987	0.0000	0.0001	0.0069	0.0123

$h = 0.5$, $T_d = 0.0497$, approx. $T_s = 0.0316$, $P_{nc} = 0.129$

r/s	P_{c1}	P_{nr}	P_r	P_{rs}	T_s	T_c / P_{c1}	T_r
0	0.871	1	0	0	0.0434	0.0425	–
0.04	0.433	0.370	0.519	0.368	0.0200	0.0407	0.0319
0.08	0.230	0.137	0.753	0.361	0.0096	0.0380	0.0300
0.16	0.0848	0.0188	0.913	0.177	0.0027	0.0306	0.0270
0.32	0.0335	0.0004	0.969	0.0234	0.0006	0.0178	0.0232
0.64	0.0223	0.0000	0.978	0.0002	0.0003	0.0117	0.0197
1.28	0.0178	0.0000	0.983	0.0000	0.0002	0.0093	0.0166

$h = 0.9$, $T_d = 0.0695$, approx. $T_s = 0.0441$, $P_{nc} = 0.124$

r/s	P_{c1}	P_{nr}	P_r	P_{rs}	T_s	T_c / P_{c1}	T_r
0	0.876	1	0	0	0.0642	0.0634	–
0.04	0.484	0.453	0.460	0.336	0.0335	0.0612	0.0525
0.08	0.281	0.205	0.694	0.361	0.0180	0.0579	0.0504
0.16	0.116	0.0421	0.879	0.212	0.0059	0.0479	0.0463
0.32	0.0469	0.0018	0.951	0.0389	0.0015	0.0289	0.0399
0.64	0.0340	0.0000	0.968	0.0008	0.0007	0.0198	0.0321
1.28	0.0252	0.0000	0.979	0.0000	0.0004	0.0154	0.0242

T_d and T_s are the expected durations of the deterministic phase of the sweep and pairwise coalescent time associated with the sweep, respectively; P_{nc} is the probability of no coalescence during the sweep, in the absence of recombination; P_{nr} is the probability of no recombination during the sweep, in the absence of coalescence; P_{c1} is the probability of coalescence during the sweep; P_r is the probability of a least one recombination event during the sweep; T_c / P_{c1} is the mean time to coalescence during the sweep, conditioned on coalescence; T_r is the mean time to the first recombination event, conditioned on the occurrence of a recombination event.

S10. Supplementary Figures

Figure S1 The reduction in diversity (relative to the neutral value) at the end of a sweep for an autosomal locus (Y-axis, linear), as a function of the ratio of the frequency of recombination (*r*) to the selection coefficient for homozygotes (*s*) (X-axis, linear scale). A randomly mating population of size 5000 is assumed, with three different values of the scaled selection coefficient ($\gamma = 2N_e s$): 125 (top panel), 250 (middle panel) and 500 (bottom panel), and three different values of the dominance coefficient (*h*), increasing from left to right. The filled red circles are the mean values from computer simulations, using Tajima's algorithm; the open blue circles and black circles are the *C1* and *C2* predictions, respectively; the open blue squares are the *NC* predictions.

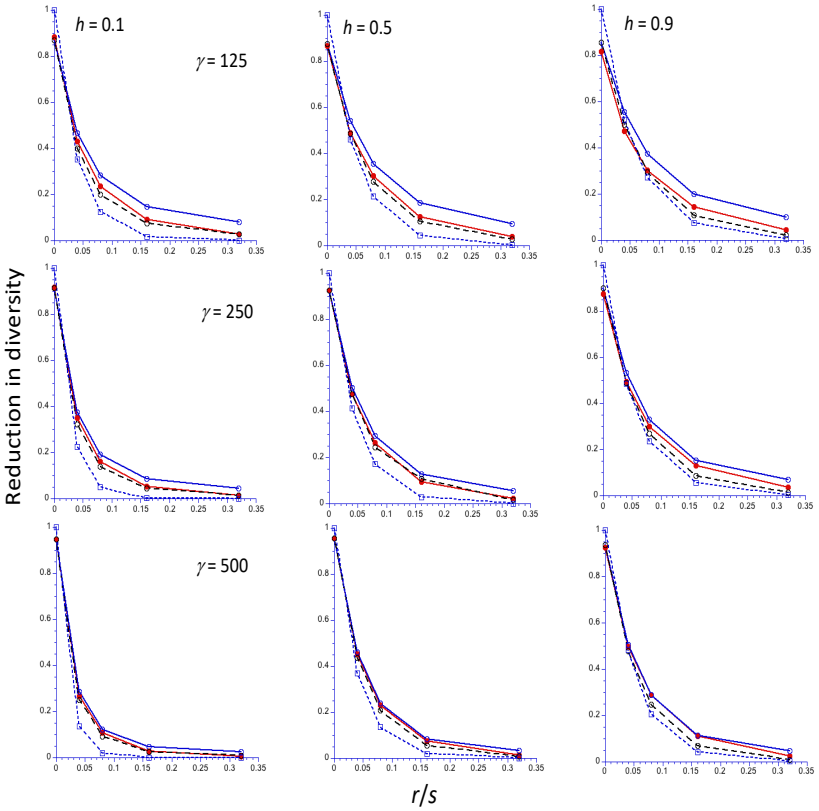

Figure S2 The reduction in diversity (relative to the neutral value) at the end of a sweep for an autosomal locus, as a function of the scaled rate of recombination ($\rho = 2N_e r$). The results for three different values of the dominance coefficient (h) are displayed. A randomly mating population of size 5000 is assumed, with a scaled selection coefficient ($\gamma = 2N_e s$) of 500. The filled red circles and black lozenges are the mean values from computer simulations, using Tajima's algorithm and the results of Hartfield and Bataillon (2020), respectively; the open blue circles and black circles are the *C1* and *C2* predictions, respectively; the open blue squares are the *NC* predictions.

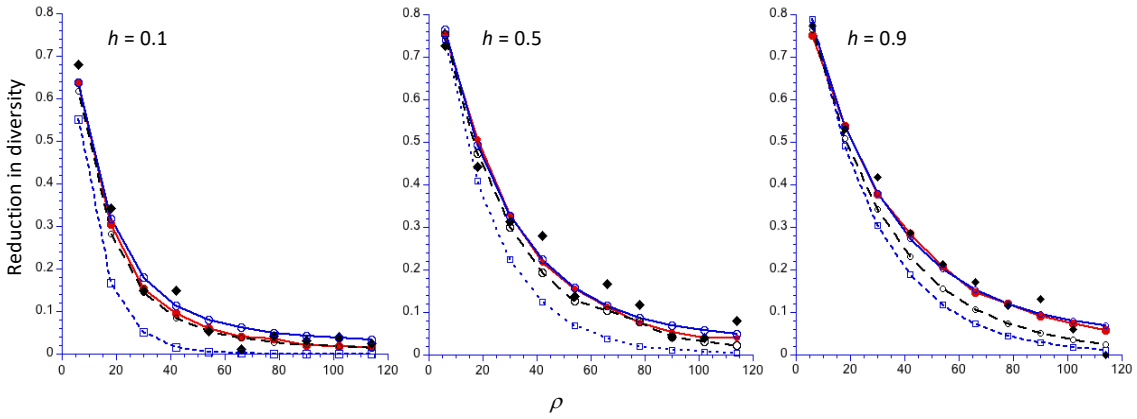

Figure S3 The reduction in diversity (relative to the neutral value) at the end of a sweep for an X-linked locus (Y-axis, $\log_{10}$ scale), as a function of the ratio of the frequency of recombination (r) to the selection coefficient for homozygotes (s) (X-axis, $\log_2$ scale). The mammalian recombination model is assumed. The results for mutations with no sex limitation are shown in the left-hand panels; those for male-limited and female-limited mutations are shown in the middle and right-hand panels, respectively. A population size of 5000 is assumed, with a scaled selection coefficient for an autosomal mutation in a randomly mating population ($\gamma = 2N_e s$) of 250 for the cases with no sex-limitation. For the sex-limited cases, $\gamma = 500$ to ensure comparability to sex-limited autosomal mutations. Results for three different values of the dominance coefficient (h) are shown, with h increasing from top to bottom. The filled red circles are the mean values from computer simulations, using Tajima's algorithm; the open filled blue circles and black circles are the *C1* and *C2* predictions, respectively; the open blue squares are the *NC* predictions. Values of the reduction in diversity less than 0.001 have been reset to 0.001.

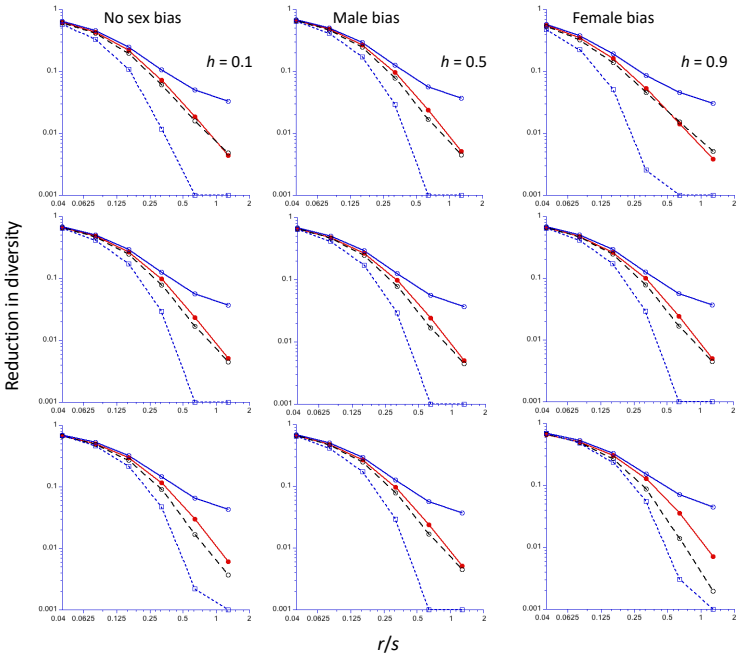

Figure S4 Reductions in diversity (relative to the neutral value) at the end of a sweep for an X-linked locus in a randomly mating population, as a function of the ratio of the autosomal value of the rate of crossing over to the homozygous selection coefficient (r/s), for three different dominance coefficients (h). N_e for the X chromosome is three-quarters of that for the autosomes. Linear scales are used for both X and Y axes. The upper panel is for a *Drosophila* model, with no crossing over in males; the lower panel is for a mammalian model, with equal rates of crossing over for autosomes in both sexes. The red and blue colors denote the *C1* and *C2* predictions, respectively. The full bars denote results for mutations with equal effects in both sexes, with γ = 250. The stippled bars denote male-limited mutations and the hatched bars denote female limited mutations. For the sex-limited cases, γ = 500 to ensure comparability with sex-limited autosomal mutations.

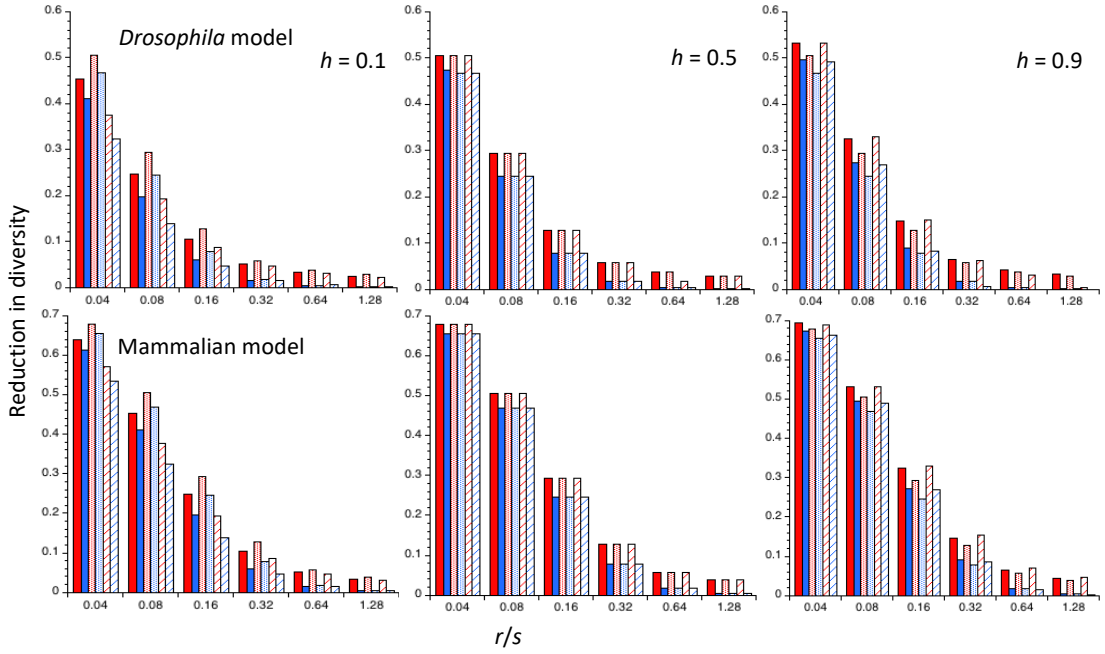

Figure S5 Reductions in diversity (relative to the neutral value) at the end of a sweep for an X-linked locus in a randomly mating population for the mammalian recombination model, as a function of the ratio of the autosomal value of the rate of crossing over to the homozygous selection coefficient (r/s), for three different dominance coefficients (h). Gene conversion is absent. The upper panel is for the case when N_e for the X chromosome is half of that for the autosomes; in the lower panel, N_e is the same for both X and A. Red and blue denote the *C1* and *C2* predictions, respectively. The full bars denote results for mutations with equal effects in both sexes, with γ = 250. The stippled bars denote male-limited mutations, and the hatched bars denote female-limited mutations. For the sex-limited cases, γ = 500 to ensure comparability with sex-limited autosomal mutations.

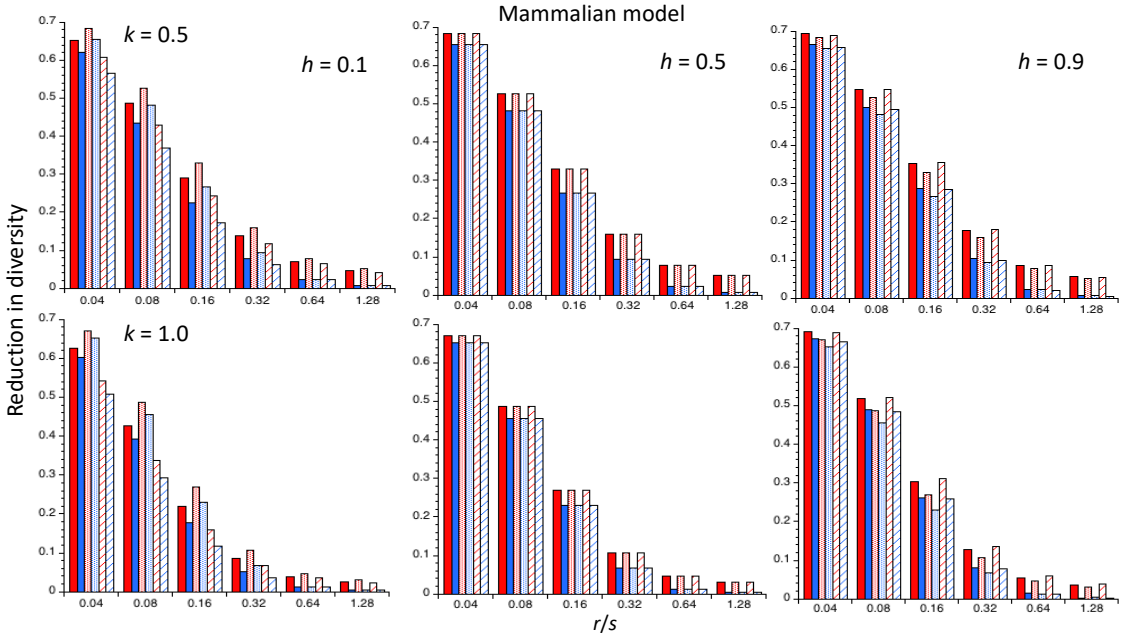

Figure S6 Reductions in diversity (relative to neutrality) under recurrent sweeps at autosomal and X-linked loci for the *Drosophila* recombination model, using the *C2* theoretical predictions with no gene conversion and five different rates of crossing over relative to the autosomal standard value (the X-linked rates of crossing over were chosen to give the same sex-averaged effective recombination rates as for autosomes). N_e for the X chromosome is three-quarters of that for the autosomes. The upper panel is for cases without BGS; the lower panel is for cases with BGS (using the parameters described in the main text). The filled red bars are for autosomal mutations, the hatched blue bars are for X-linked mutations with no sex-limitation, the stippled gray bars are for male-limited mutations X-linked mutations, and the hatched green bars are for female-limited X-linked mutations. For the sex-limited cases, γ = 500 to ensure comparability with sex-limited autosomal mutations.

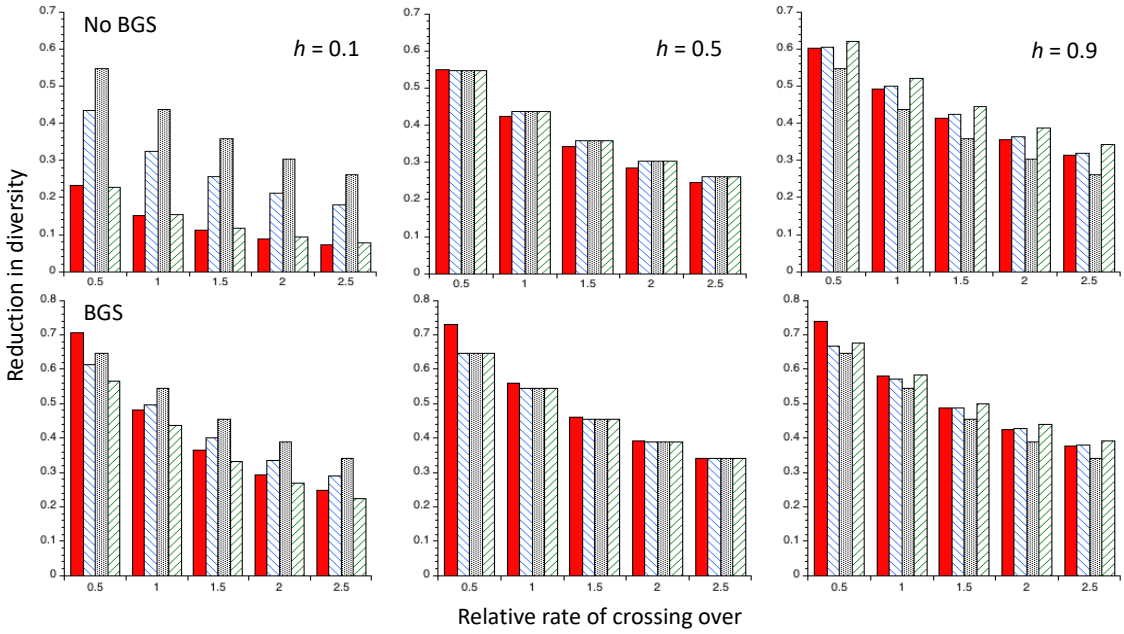

Figure S7 The ratios of X chromosome to autosome nucleotide site diversities under recurrent sweeps for the *Drosophila* model, using the *C2* theoretical predictions with no gene conversion and five different rates of crossing over relative to the autosomal standard value (the X-linked rates of crossing over were chosen to give the same sex-averaged effective rates as for the autosomes). The upper panel is for cases without BGS; the lower panel is for cases with BGS. The filled red bars are for h = 0.1 (the dominance coefficient of favorable mutations), the hatched blue bars are for h = 0.5 and the stippled gray bars are for h = 0.9. The other details are as for Figure S6.

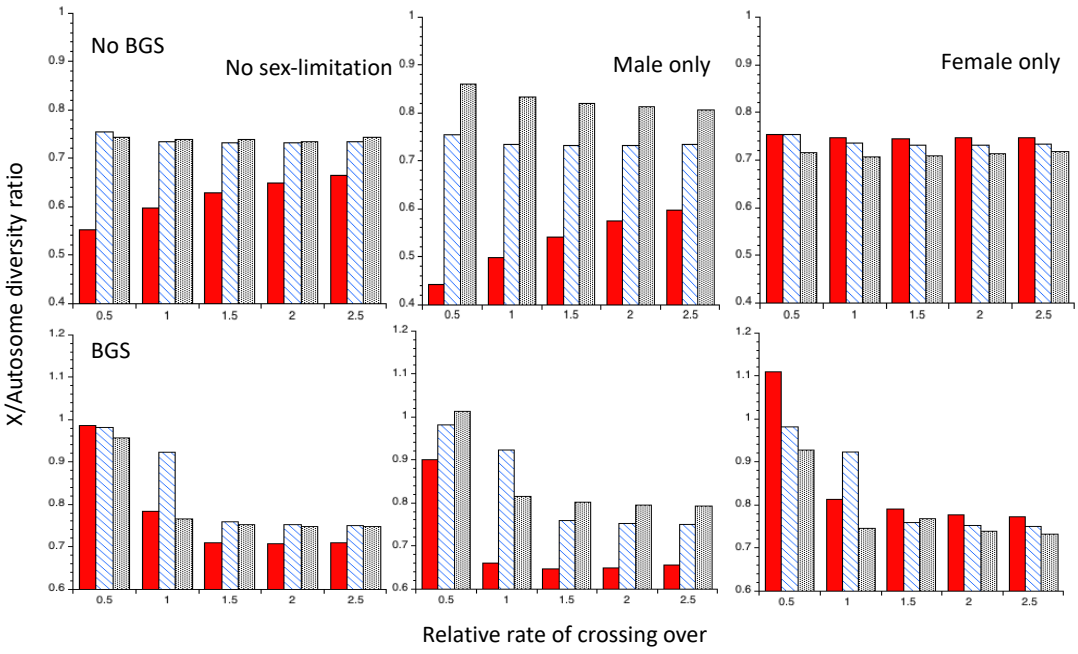

Figure S8 Reductions in diversity (relative to neutrality) under recurrent sweeps at autosomal and X-linked loci for the *Drosophila* model, using the *C2* theoretical predictions with gene conversion and five different rates of crossing over relative to the autosomal standard value (the X-linked rates of crossing over and gene conversion were chosen to give the same sex-averaged effective rates as for the autosomes). BGS is assumed to be present, with the parameters described in the main text. The upper panel is for the case when N_e for the X chromosome is half of that for the autosomes; in the lower panel, N_e is the same for both X and A. The filled red bars are for autosomal mutations, the hatched blue bars are for X-linked mutations with no sex-limitation, the stippled gray bars are for male-limited mutations, and the hatched green bars are for female-limited mutations. For the sex-limited cases, γ = 500 to ensure comparability with the autosomal and non-sex-limited X-linked mutations.

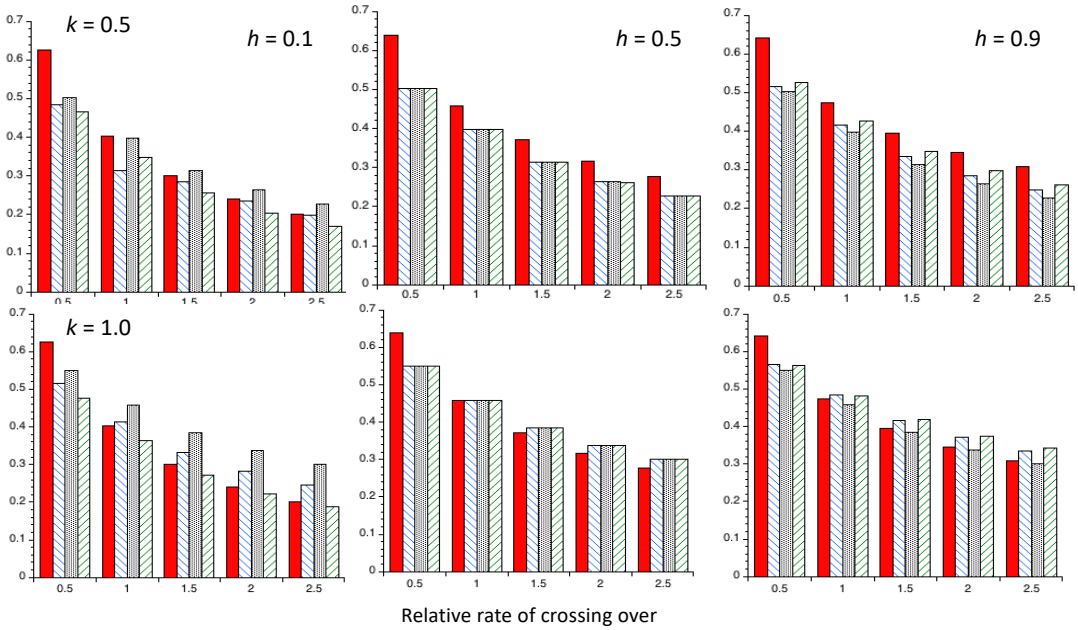

Figure S9 The ratios of X chromosome to autosome nucleotide site diversities for the *Drosophila* model with recurrent sweeps, using the *C2* theoretical predictions with gene conversion and five different rates of crossing over relative to the autosomal standard value (the X-linked rates of crossing over and gene conversion were chosen to give the same sex-averaged effective rates as the autosomal rates). BGS is assumed to be present. The upper panel is for the case when N_e for the X chromosome is half that for the autosomes; in the lower panel, N_e is the same for both X and A. The filled red bars are for *h* = 0.1 (the dominance coefficient of favorable mutations), the hatched blue bars are for *h* = 0.5 and the stippled gray bars are for *h* = 0.9. The other details are as for Figure S8.