DOI: 10.1534/genetics.108.092270

Note

The Influence of Gene Conversion on Linkage Disequilibrium Around a Selective Sweep

Danielle A. Jones[1] and John Wakeley

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

Manuscript received June 4, 2008
Accepted for publication July 17, 2008

ABSTRACT

In a 2007 article, McVean studied the effect of recombination on linkage disequilibrium (LD) between two neutral loci located near a third locus that has undergone a selective sweep. The results demonstrated that two loci on the same side of a selected locus might show substantial LD, whereas the expected LD for two loci on opposite sides of a selected locus is zero. In this article, we extend McVean's model to include gene conversion. We show that one of the conclusions is strongly affected by gene conversion: when gene conversion is present, there may be substantial LD between two loci on opposite sides of a selective sweep.

MCVEAN (2002) showed that predictions for r^2, a commonly used measure of LD introduced by HILL and ROBERTSON (1968), depend on the correlations in coalescence times for a pair of loci, which in turn depend on the recombination rate between the loci. In applying this result to LD near a locus that has undergone a selective sweep, MCVEAN (2007) developed a new model that features two neutral loci partially linked to the selected locus. He assumed that recombination could occur between each pair of loci and that the sweep had a particularly simple structure: a star tree. Figure 1 depicts the model, in which a sample at the two neutral loci is taken at the present time 0, just after the sweep has finished. The sweep is assumed to have occurred quickly and to have begun at time t_M in the past, measured in units of $2N_e$ generations, where N_e is the (diploid) coalescent effective population size (SJÖDIN *et al.* 2005). On the basis of the work of MAYNARD SMITH and HAIGH (1974) and others (KAPLAN *et al.* 1989; STEPHAN *et al.* 1992; DURRETT and SCHWEINSBERG 2004), MCVEAN (2007) used $t_M = 0.1$, and we adopt this value below. MCVEAN (2007) tested the validity of this approximate model by comparing its predictions to the results of fully stochastic simulations of a sweep and found them to be largely accurate.

MCVEAN (2007) allowed for recombination (reciprocal exchange of genetic material as in a single crossover event) but not for gene conversion (nonreciprocal exchange of short tracks of genetic material). However, there is a growing body of evidence for the importance of gene conversion in shaping genetic variation in humans (FRISSE *et al.* 2001; JEFFREYS and MAY 2004; PADHUKASAHASRAM *et al.* 2004; CHEN *et al.* 2007; GAY *et al.* 2007) and models that do not feature gene conversion, therefore, do not completely capture the biological causes, and expectations, of genetic variability. Our aim here is to incorporate gene conversion into the model and to ask whether this changes the results. We focus on the case in which variation at the two neutral loci is due to mutations that occurred during the neutral phase shown in Figure 1B. In this case, the two neutral loci can be polymorphic only if they do not coalesce along with the selected allele at time t_M in Figure 1B. Without recombination or gene conversion, present-day samples at the two neutral loci will always remain linked to the selected allele and will certainly coalesce with the selected allele. Recombination and gene conversion allow the loci to "escape" the sweep with some probability and to coalesce during the neutral phase, where they might also experience mutations.

To make a prediction for r^2—specifically σ_d^2 of OHTA and KIMURA (1971)—it is necessary to compute the expectation of the product of the coalescence times at two loci for each of the three sample configurations (A, B, and C) in Figure 1A (MCVEAN 2002). Briefly, in the three-locus model of MCVEAN (2007), for each of these three sampling configurations, we must compute the probability that the two neutral loci are in configuration A, B, or C at the start of the neutral phase looking back (*i.e.*, time t_M), at which point all chromosomes in the

[1] *Corresponding author:* Biolabs 4092, 16 Divinity Ave., Cambridge, MA 02138. E-mail: djones@fas.harvard.edu

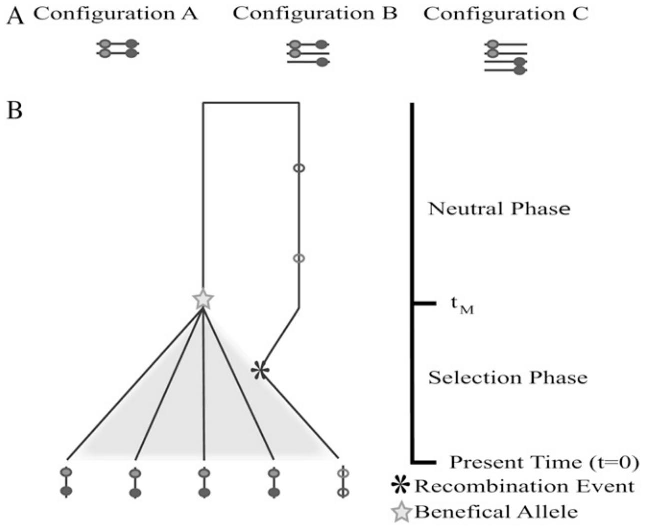

Figure 1.—This is an adaptation of Figure 1 of McVean (2007). (A) The three configurations are A, two neutral loci are sampled from two chromosomes; B, two neutral loci are sampled from three chromosomes; and C, two neutral loci are sampled from four chromosomes. (B) The selection event occurs as a rapid selective sweep during which only crossing-over events can occur. During the neutral phase, coalescent events can also occur.

population have the ancestral type, or wild type, at the selected locus. There are nine such probabilities in total, one for each pair of configurations. These nine probabilities are denoted using ϕ, with subscripts to represent configurations. The expected product of coalescence times at the two neutral loci, sampled at present when all chromosomes possess the selected allele (denoted by the subscript S), are the averages over the three ancestral configurations. The expected coalescence time for two chromosomes, i and j, sampled at locus X is written as t^x_{ij} and for chromosomes k and l, at locus Y, it is written as t^y_{kl}. For configuration A, we have

$$E_S[t^x_{ij}t^y_{ij}] = \phi_{AA}E_W[t^x_{ij}t^y_{ij}] + \phi_{AB}E_W[t^x_{ij}t^y_{ik}] + \phi_{AC}E_W[t^x_{ij}t^y_{kl}],$$

which is Equation 9 in McVean (2007). The probabilities ϕ_{AA}, ϕ_{AB}, and so on, depend on t_M and on the rates of recombination and gene conversion. The expected values on the right-hand side above are those expected during the neutral phase (W stands for wild-type allele) and are given in Equation 10 of McVean (2007).

The predictions about LD depend strongly on the relative position of the selected locus compared to the neutral loci. McVean considered two cases: (1) the selected locus is located halfway between the two neutral loci (NSN) and (2) the selected locus is located on one side of the two neutral loci (SNN). All of the derivations described above are done separately for these two cases. Considering only recombination, McVean predicted substantial LD for SNN, because in this case both neutral loci can escape the sweep yet remain linked to each other at the beginning of the neutral phase. For the NSN case, the model without gene conversion predicts no LD between the two neutral loci. As McVean

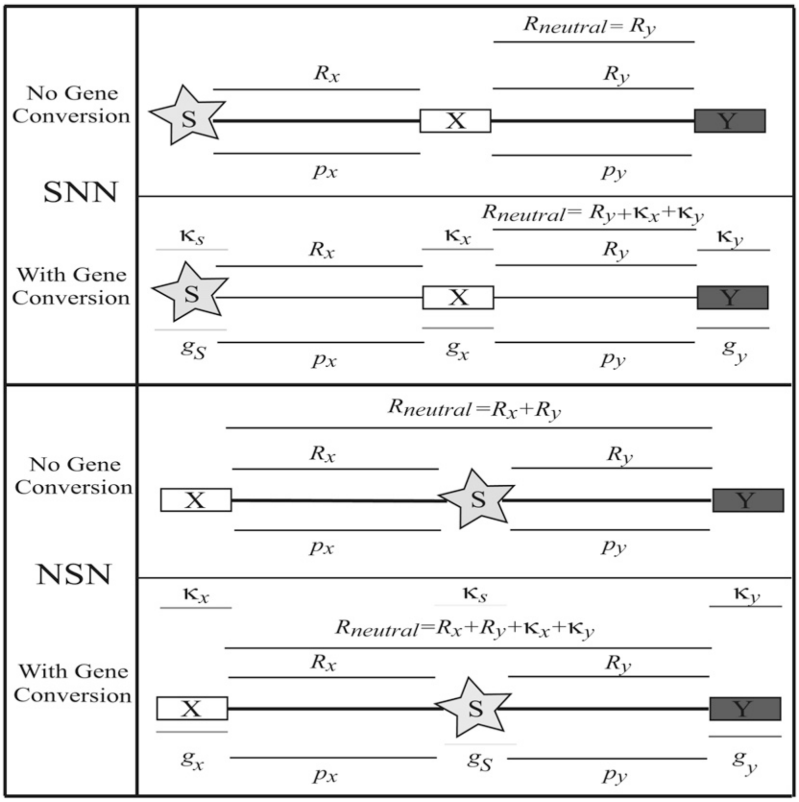

Figure 2.—There are four cases: SNN without gene conversion and with gene conversion and NSN without gene conversion and with gene conversion. In all cases, during the selective sweep phase, there are two parameters that describe the probability of escape via recombination: $p_x = 1 - e^{-(R_x/2)t_M}$ and $p_y = 1 - e^{-(R_y/2)t_M}$. During the neutral phase, the probability of recombination is $R_x = 4N_e r$ and $R_y = 4N_e r$, where r is the per generation probability of recombination. The recombination distance between the two neutral loci is kept constant regardless of whether they are in the SNN or the NSN case. Thus for NSN, $R_x = R_y = \frac{1}{2}4N_e r$. When gene conversion is added to the model for both SNN and NSN, the gene conversion is restricted to the individual loci. The overall rate of gene conversion is $\kappa_X = \kappa_S = \kappa_Y = 4N_e\gamma$ for each locus, where γ is the per generation probability of gene conversion.

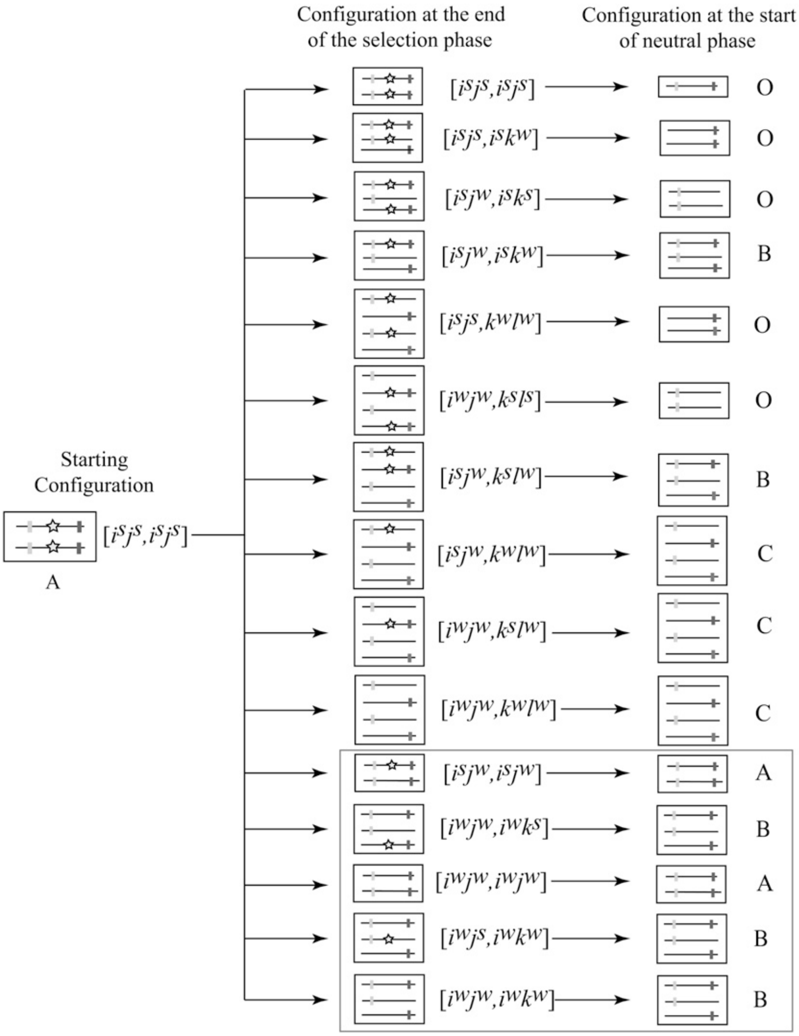

FIGURE 3.—This is an adaptation of Figure 2 of McVEAN (2007) but, importantly, it includes five novel transitions from configuration A to configurations A and B at the start of the neutral phase that are not possible when only recombination is present. The transitions represented correspond to the transition probabilities present in Table A1 and are in the same order—excluding the five cases where the transition probability is zero—as in Table A1. Also shown in brackets is the notation used for each configuration (McVEAN 2007). The notation for configuration A is $[i^sj^s, i^sj^s]$ which indicates that two chromosomes, *i* and *j*, were sampled at locus X and locus Y, and both chromosomes possess the selected allele. The sampling configurations at locus X and locus Y are separated by a comma.

notes, this reflects the symmetric nature of the recombination process for the NSN case: the probabilities of each of the three configurations at the beginning of the neutral phase (A, B, or C) are the same for each of the three sample configurations, so that $\sigma_{\mathrm{d}}^2 = 0$ (see Equation 14 of McVEAN 2007). This symmetry breaks down when gene conversion is included in the model because gene conversion at the selected locus allows both neutral loci to escape the sweep yet remain linked at the beginning of the neutral phase.

Figure 2 gives a graphical representation of the model for SNN and NSN, with and without gene conversion. We assume that when gene conversion occurs, it copies a tract length of *m* nucleotides, which is greater than the size of each locus and less than the distance between the loci. Thus, gene conversion acts on single loci independently. During the neutral phase, on the coalescent timescale, recombination events occur with rates R_x and R_y, and gene conversion events occur at rates κ_{s}, κ_x, and κ_y. Following McVEAN (2007), during the selection phase with duration t_{M} there are two probabilities of escape via recombination: $p_x = 1 - e^{-t_{\mathrm{M}}R_x/2}$ and $p_y = 1 - e^{-t_{\mathrm{M}}R_y/2}$. To these we add three probabilities of escape by gene conversion: $g_{\mathrm{s}} = 1 - e^{-t_{\mathrm{M}}\kappa_{\mathrm{s}}/2}$, $g_x = 1 - e^{-t_{\mathrm{M}}\kappa_x/2}$, $g_y = 1 - e^{-t_{\mathrm{M}}\kappa_y/2}$.

Tables A1–A6 in the APPENDIX give all the terms needed to compute the probabilities ϕ_{AA}, ϕ_{AB}, and so on, for both SNN and NSN. We follow McVEAN (2007) in computing transitions between the present and the start of the neutral phase, using the intermediate configuration at the end of the selection phase; our Tables A1–A3 correspond to each of the three columns of Appendix A in McVEAN (2007) and our Tables A4–A6 correspond to each of the three columns of Appendix B in McVEAN (2007). Again, the key difference between the models is that in the NSN case gene conversion allows present-day configuration A to remain in configuration A at the start of the neutral phase, whereas this is not possible by recombination alone. Figure 3 shows the configurations that can be reached from sampling configuration A at the present; it is analogous to Figure 2 in McVEAN (2007) and shows five of the six novel con-

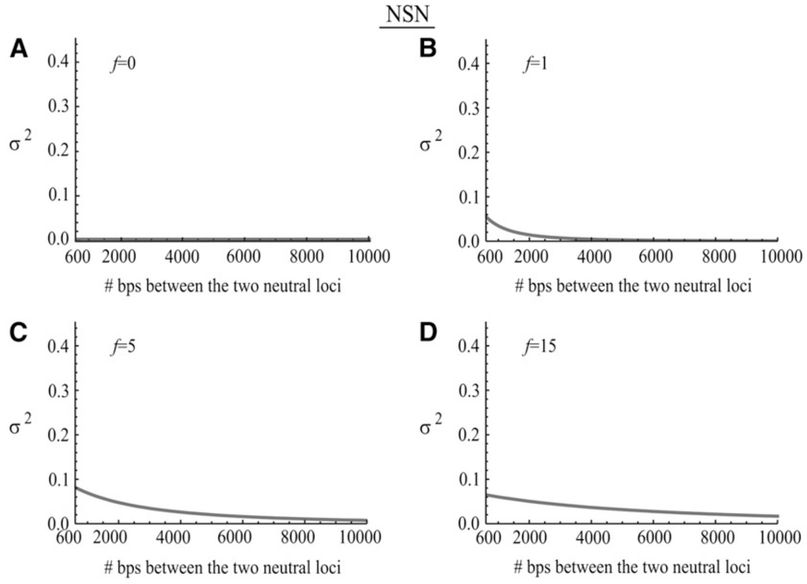

Figure 4.—Expected LD for the NSN case. The y-axis is the amount of LD as predicted by σ_D^2. The x-axis is the distance, in base pairs, between the two neutral loci starting from a distance of 2 × tract length, m, between them and increasing to 10,000 bp. In this example, $m = 300$ bp. The distance between either neutral locus and the selected site must be at least a tract length, so that any given gene conversion event converts only one locus; the two neutral loci are separated by a minimum of two tract lengths, 600 bp. For this distance range, $R_{\mathrm{Neutral}} = n\rho + 2m\kappa$. (A) $f = 0$, there is no gene conversion. (B) $f = 1$, there is the same amount of gene conversion as there is recombination. (C) $f = 5$, there is 5 times as much gene conversion as recombination. This is a reasonable ratio for human data. (D) $f = 15$, there is 15 times as much gene conversion as recombination.

figurations (the bottom five transitions encompassed in a box) that can be reached when gene conversion is included in the model.

Following McVean (2007), to generate predictions applicable to molecular data, we assume that the rates of recombination, R_x and R_y, depend linearly on the distance between the loci. For example, if locus X is n nucleotides away from locus S, then $R_x = n\rho$, where ρ is the rate of recombination between adjacent nucleotides on the coalescent timescale. In addition, we assume that each locus is a single-nucleotide site, so that in the neutral phase all three loci have the same rate of conversion: $\kappa_s = \kappa_x = \kappa_y = m\kappa$, where κ is the rate of initiation of a gene conversion event between two adjacent nucleotides on a coalescent timescale. Finally, we include a parameter, $f = \kappa/\rho$, which is the ratio of gene conversion to recombination. Note that, with this parameterization, the per generation probabilities of recombination and gene conversion in Figure 2 are given by $r = n\rho/4N_e$ and $\gamma = m\kappa/4N_e$.

To predict the likely effect of gene conversion on LD in human populations, we substituted plausible genetic parameters from humans: $\rho \sim 0.0005$/bp (Frisse *et al.* 2001); the ratio of gene conversion to recombination, f, has been estimated to be between ~1.5 and 14 (Frisse *et al.* 2001; Jeffreys and May 2004; Padhukasahasram *et al.* 2004; Chen *et al.* 2007; Gay *et al.* 2007); and typical gene conversion tract lengths range from 50 to 500 bp

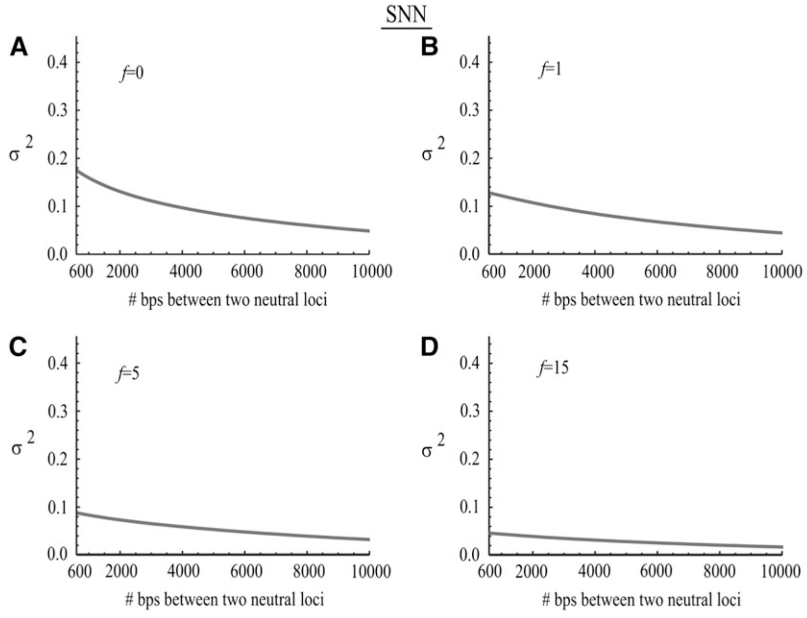

Figure 5.—Expected LD for the SNN case. The y-axis is the amount of LD as predicted by σ_D^2. The x-axis is the distance, in base pairs, between the two loci starting from a distance of at least two tract lengths between them, to allow for comparison to the NSN case, and increasing to 10,000 bp. In this example, $m = 300$ bp; therefore, the starting distance between the two neutral loci is 600 bp. For this distance range, $R_{\mathrm{Neutral}} = n\rho + 2m\kappa$. (A) $f = 0$, there is no gene conversion. (B) $f = 1$, there is the same amount of gene conversion as there is recombination. (C) $f = 5$, there is 5 times as much gene conversion as recombination. (D) $f = 15$, there is 15 times as much gene conversion as recombination.

(Jeffreys and May 2004). To simplify our analysis, we assume a fixed tract length of 300 bp (Jeffreys and Neumann 2002). Repeating our analysis with a 50-bp tract length led to slightly higher levels of LD (results not shown).

As Figures 4 and 5 show, adding gene conversion affects LD in both the NSN and the SNN case. In the SNN case, the effect of gene conversion is similar to the effect of recombination, so that increasing f decreases LD (Figure 5, B–D). Adding gene conversion in the SNN case creates more opportunities for the two neutral loci to escape the sweep independently, so LD between them is reduced. In the NSN case, gene conversion *increases* LD (Figure 4, B–D). In this case, gene conversion has a qualitatively different effect. Gene conversion events at the middle (S) locus allow the two neutral loci to escape the sweep together. Present-day samples of this type may then be samples of an ancestral haplotype that would otherwise have been lost during the sweep. Looking forward in time, in the NSN case gene conversion can preserve the preexisting correlated genealogical structure between the outer loci.

By incorporating gene conversion into the three-locus model of McVean (2007), we have shown that LD is expected between two loci on opposite sides of a selected locus that has undergone a sweep. Although we have focused only on a single pair of neutral loci, our results have implications for genomic scans for selective sweeps using extended haplotype homozygosity (Sabeti *et al.* 2002), integrated extended haplotype homozygosity (Voight *et al.* 2006), and long-range haplotypes (Sabeti *et al.* 2007). In particular, gene conversion *at the selected site* will cause some fraction of present-day chromosomes to show the selected allele but while sitting on an ancestral haplotype. Using $t_M = 0.1$ as in McVean (2007), and assuming a tract length of $m = 300$ and a ratio of gene conversion to recombination of $f = 5$, the probability of sampling such a chromosome is $1 - e^{-0.0375} \approx 0.037$. Then, among 100 chromosomes that all possess the selected allele, we would expect to see about four of these aberrant haplotypes, and the chance that all 100 chromosomes would show the classic, recombination-only sweep pattern would be $0.9625^{100} \approx 0.024$. Thus, it is possible that many selected loci have been missed in the recent genomic scans for selection.

LITERATURE CITED

Chen, J.-M., D. N. Copper, N. Chuzhanova, C. Ferec and G. P. Patrinos, 2007 Gene conversion: mechanisms, evolution and human disease. Nature **8:** 762–775.

Durrett, R., and J. Schweinsberg, 2004 Approximating selective sweeps. Theor. Popul. Biol. **66:** 129–138.

Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack *et al.*, 2001 Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium. Am. J. Hum. Genet. **69:** 831–843.

Gay, J., S. Myers and G. McVean, 2007 Estimating meiotic gene conversion rates from population genetic data. Genetics **177:** 881–894.

Hill, W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

Jeffreys, A. J., and C. May, 2004 Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat. Genet. **36:** 151–156.

Jeffreys, A. J., and R. Neumann, 2002 Reciprocal crossover asymmetry and meiotic drive in human recombination hot spot. Nat. Genet. **31:** 267–271.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The hitchhiking effect revisited. Genetics **123:** 887–899.

Maynard Smith, J., and J. Haigh, 1974 The hitchhiking effect of a favourable gene. Genet. Res. **23:** 23–35.

McVean, G., 2002 A genealogical interpretation of linkage disequilibrium. Genetics **16:** 987–991.

McVean, G., 2007 The structure of linkage disequilibrium around a selective sweep. Genetics **175:** 1395–1406.

Ohta, T., and M. Kimura, 1971 Linkage disequilibrium between two segregating nucleotide sites under steady flux of mutations in a finite population. Genetics **68:** 571–580.

Padhukasahasram, B., P. Marjoram and M. Nordborg, 2004 Estimating the rate of gene conversion on human chromosome 21. Am. J. Hum. Genet. **75:** 386–397.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419:** 832–837.

Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter *et al.*, 2007 Genome-wide detection and characterization of positive selection in human populations. Nature **449:** 913–918.

Sjödin, P., I. Kaj, S. Krone, M. Lascoux and M. Nordborg, 2005 On the meaning and existence of an effective population size. Genetics **169:** 1061–1070.

Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphisms: analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254.

Voight, B. F., S. Kudaravalli, X. Wen and J. K. Pritchard, 2006 A map of recent positive selection in the human genome. PLoS Biol. **4**(3): e72.

Communicating editor: N. Takahata

APPENDIX

Notation

The transition equations used in Tables A1–A6 are complicated by the addition of gene conversion events. In an attempt to simplify the equations used to build the tables, a new notation, which incorporates the notation of McVean (2007), is used. The new notation is outlined and compared to McVean's below. All of the symbols used refer to escape probabilities during the selection phase that result from either recombination or gene conversion.

For NSN (Tables A1–A3): *McVean (2007):* q_x is used to indicate that no recombination event has occurred between locus X and locus S. p_x is used to indicate that a recombination event has occurred between locus X and locus S. q_y is used to indicate that no recombination event has occurred between locus S and locus Y. p_y is used to indicate that a recombination event has occurred between locus S and locus Y.

TABLE A1

Transition probabilities for NSN for the starting configuration A to the four states, A, B, C, or O, present at the beginning of the neutral phase

Configuration at the end of selection phase	Probability given starting configuration $[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}j^{\mathrm{S}}]$	Configuration at the start of the neutral phase
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}j^{\mathrm{S}}]$	$(q_x^*(1-g_{\mathrm{s}})q_y^*)^2$	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}k^{\mathrm{W}}]$	$2(q_x^*(1-g_{\mathrm{s}})q_y^*)(q_x^*(1-g_{\mathrm{s}})p_y^*)$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}k^{\mathrm{S}}]$	$2(q_x^*(1-g_{\mathrm{s}})q_y^*)(p_x^*(1-g_{\mathrm{s}})q_y^*)$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}k^{\mathrm{W}}]$	$2(q_x^*(1-g_{\mathrm{s}})q_y^*)(p_x^*p_y^* + p_x^*q_y^*g_{\mathrm{s}} + q_x^*p_y^*g_{\mathrm{s}})$	B
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$((q_x^*(1-g_{\mathrm{s}})p_y^*))^2$	O
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	$(p_x^*(1-g_{\mathrm{s}})q_y^*)^2$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	$2(q_x^*(1-g_{\mathrm{s}})p_y^*)(p_x^*(1-g_{\mathrm{s}})q_y^*)$	B
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$2(q_x^*(1-g_{\mathrm{s}})p_y^*)(p_x^*p_y^* + p_x^*q_y^*g_{\mathrm{s}} + q_x^*p_y^*g_{\mathrm{s}})$	C
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	$2(p_x^*p_y^* + p_x^*q_y^*g_{\mathrm{s}} + q_x^*p_y^*g_{\mathrm{s}})(p_x^*(1-g_{\mathrm{s}})q_y^*)$	C
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$(p_x^*p_y^* + p_x^*q_y^*g_{\mathrm{s}} + q_x^*p_y^*g_{\mathrm{s}})^2$	C
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}k^{\mathrm{S}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}j^{\mathrm{W}}]$	$2(q_x^*(1-g_{\mathrm{s}})q_y^*)(q_x^*g_{\mathrm{s}}q_y^*)$	A
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}k^{\mathrm{S}}]$	$2(q_x^*g_{\mathrm{s}}q_y^*)(p_x^*(1-g_{\mathrm{s}})q_y^*)$	B
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}j^{\mathrm{W}}]$	$(q_x^*g_{\mathrm{s}}q_y^*)^2$	A
$[i^{\mathrm{W}}j^{\mathrm{S}}, i^{\mathrm{W}}k^{\mathrm{W}}]$	$2(q_x^*g_{\mathrm{s}}q_y^*)(q_x^*(1-g_{\mathrm{s}})p_y^*)$	B
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}k^{\mathrm{W}}]$	$2(q_x^*g_{\mathrm{s}}q_y^*)(p_x^*p_y^* + p_x^*q_y^*g_{\mathrm{s}} + q_x^*p_y^*g_{\mathrm{s}})$	B
$[i^{\mathrm{W}}j^{\mathrm{S}}, i^{\mathrm{W}}k^{\mathrm{S}}]$	0	A

O corresponds to a state where at least one of the two neutral loci has coalesced. There are six new states created by the addition of gene conversion that are not present when recombination is the only crossing-over event option. These are the six last states. This table corresponds to the transition probabilities in the second column of Appendix A of McVean (2007) and the same notation, explained in detail in Figure 3, is used.

Notation with gene conversion: The terminology is the same as that in McVean (2007) but with the additional consideration that escape from the selection sweep can also come from gene conversion.

$q_x^* = q_x(1 - g_x)$: there is no recombination between locus X and locus S and no gene conversion at locus X.

$p_x^* = 1 - q_x^* = [p_x(1 - g_x) + (1 - p_x)g_x + p_xg_x]$: there is either at least one recombination event between locus X and locus S or at least one gene conversion at locus X.

$q_y^* = q_y(1 - g_y)$: there is no recombination between locus S and locus Y and no gene conversion at locus Y.

$p_y^* = 1 - q_y^* = [p_y(1 - g_y) + (1 - p_y)g_y + p_yg_y]$: there is either at least one recombination event between locus S and locus Y or at least one gene conversion at locus Y.

For SNN (Tables A4–A6): *McVean (2007):* q_x and p_x have the same meaning as their NSN counterparts. q_y is used to indicate that no recombination event has occurred between locus X and locus Y. p_y is used to indicate that at least one recombination event has occurred between locus X and locus Y.

Notation with gene conversion:

q_x^* and p_x^* have the same meaning as their NSN counterparts.

$q_y^* = q_y(1 - g_y)$: there is no recombination between locus X and locus Y and no gene conversion at locus Y.

$p_y^* = 1 - q_y^* = [p_y(1 - g_y) + (1 - p_y)g_y + p_yg_y] = [p_y + g_y - p_yg_y]$: there is either at least one recombination event between locus X and locus Y or at least one gene conversion at locus Y.

There are two additional probabilities present in the case of SNN:

$q_{\mathrm{s}} = (1 - p_x)(1 - g_{\mathrm{s}})$ is the probability that no recombination event has occurred between locus S and locus X and no gene conversion event has occurred at locus S.

$p_{\mathrm{s}} = [p_x(1 - g_{\mathrm{s}}) + (1 - p_x)g_{\mathrm{s}} + p_xg_{\mathrm{s}}] = [p_x + g_{\mathrm{s}} - p_xg_{\mathrm{s}}]$ is the probability that a recombination event between locus S and locus X or a gene conversion event at locus S or both has occurred.

TABLE A2

Transition probabilities for NSN for the starting configuration B

Configuration at the end of selection phase	Probability given starting configuration $[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}k^{\mathrm{S}}]$	Configuration at start of neutral phase
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}j^{\mathrm{S}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}k^{\mathrm{W}}]$	$q_x^*(1-g_{\mathrm{s}})q_y^*q_x^*(1-g_{\mathrm{s}})(p_y^* + p_y^*q_y^*g_{\mathrm{s}})$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}k^{\mathrm{S}}]$	$q_x^*(1-g_{\mathrm{s}})q_y^*(p_x^* + g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}})$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}k^{\mathrm{W}}]$	$q_x^*(1-g_{\mathrm{s}})q_y^*((p_x^* + q_x^*g_{\mathrm{s}}))(p_y^* + q_y^*g_{\mathrm{s}})$	B
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$q_x^*(1-g_{\mathrm{s}})p_y^*q_x^*(1-g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})$	O
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	$p_x^*(1-g_{\mathrm{s}})q_y^*(p_x^* + g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}})$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	$(q_x^*(1-g_{\mathrm{s}})p_y^*(p_x^* + g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}}) + p_x^*(1-g_{\mathrm{s}})$ $q_y^*q_x^*(1-g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}}) + (p_x^*p_y^* + p_x^*q_y^*g_{\mathrm{s}} +$ $q_x^*p_y^*g_{\mathrm{s}})q_x^*(1-g_{\mathrm{s}})^2q_y^*)$	B
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$(q_x^*(1-g_{\mathrm{s}})p_y^*(p_x^* + q_x^*g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}}) + (p_x^*p_y^* +$ $p_x^*q_y^*g_{\mathrm{s}} + q_x^*p_y^*g_{\mathrm{s}})q_x^*(1-g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}}))$	C
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	$(p_x^*(1-g_{\mathrm{s}})q_y^*(p_x^* + q_x^*g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}}) + (p_x^*p_y^* +$ $p_x^*q_y^*g_{\mathrm{s}} + q_x^*p_y^*g_{\mathrm{s}})(p_x^*+ g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}})$	C
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$(p_x^*p_y^* + p_x^*q_y^*g_{\mathrm{s}} + q_x^*p_y^*g_{\mathrm{s}})(p_x^* + q_x^*g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})$	C
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}k^{\mathrm{S}}]$	$q_x^*(1-g_{\mathrm{s}})q_y^*q_x^*(1-g_{\mathrm{s}})^2q_y^*$	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	$q_x^*(1-g_{\mathrm{s}})p_y^*q_x^*(1-g_{\mathrm{s}})^2q_y^*$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	$p_x^*(1-g_{\mathrm{s}})q_y^*q_x^*(1-g_{\mathrm{s}})^2q_y^*$	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}j^{\mathrm{W}}]$	0	A
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}k^{\mathrm{S}}]$	$q_x^*g_{\mathrm{s}}q_y^*(p_x^*+ g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}})$	B
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}j^{\mathrm{W}}]$	0	A
$[i^{\mathrm{W}}j^{\mathrm{S}}, i^{\mathrm{W}}k^{\mathrm{W}}]$	$q_x^*g_{\mathrm{s}}q_y^*q_x^*(1-g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})$	B
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}k^{\mathrm{W}}]$	$q_x^*g_{\mathrm{s}}q_y^*(p_x^* + q_x^*g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})$	B
$[i^{\mathrm{W}}j^{\mathrm{S}}, i^{\mathrm{W}}k^{\mathrm{S}}]$	$q_x^*g_{\mathrm{s}}q_y^*q_x^*(1-g_{\mathrm{s}})^2q_y^*$	A

The six new states created by the addition of gene conversion are the last six rows of the table. This corresponds to the transition probabilities in the third column of Appendix A of McVean (2007).

TABLE A3

Transition probabilities for NSN for the starting configuration C

Configuration at the end of selection phase	Probability given starting configuration $[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	Configuration at the start of neutral phase
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}j^{\mathrm{S}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}k^{\mathrm{W}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}k^{\mathrm{S}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}k^{\mathrm{W}}]$	0	B
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$(q_x^*(1-g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}}))^2$	O
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	$((p_x^* + g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}}))^2$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	$2(q_x^*(1-g_{\mathrm{s}})^2q_y^*(p_x^* + q_x^*g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}}) + q_x^*(1-g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})$ $(p_x^* + g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}}))$	B
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$2q_x^*(1-g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})(p_x^* + q_x^*g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})$	C
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	$2(p_x^* + g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}})(p_x^* + q_x^*g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})$	C
$[i^{\mathrm{W}}j^{\mathrm{W}}, k^{\mathrm{W}}l^{\mathrm{W}}]$	$((p_x^* + q_x^*g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}}))^2$	C
$[i^{\mathrm{S}}j^{\mathrm{S}}, i^{\mathrm{S}}k^{\mathrm{S}}]$	0	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{S}}l^{\mathrm{W}}]$	$2q_x^*(1-g_{\mathrm{s}})^2q_y^*q_x^*(1-g_{\mathrm{s}})(p_y^* + q_y^*g_{\mathrm{s}})$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	$2q_x^*(1-g_{\mathrm{s}})^2q_y^*(p_x^* + g_{\mathrm{s}}q_x^*)q_y^*(1-g_{\mathrm{s}})$	O
$[i^{\mathrm{S}}j^{\mathrm{S}}, k^{\mathrm{S}}l^{\mathrm{S}}]$	$(q_x^*(1-g_{\mathrm{s}})^2q_y^*)^2$	O
$[i^{\mathrm{S}}j^{\mathrm{W}}, i^{\mathrm{S}}j^{\mathrm{W}}]$	0	A
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}k^{\mathrm{S}}]$	0	B
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}j^{\mathrm{W}}]$	0	A
$[i^{\mathrm{W}}j^{\mathrm{S}}, i^{\mathrm{W}}k^{\mathrm{W}}]$	0	B
$[i^{\mathrm{W}}j^{\mathrm{W}}, i^{\mathrm{W}}k^{\mathrm{W}}]$	0	B
$[i^{\mathrm{W}}j^{\mathrm{S}}, i^{\mathrm{W}}k^{\mathrm{S}}]$	0	A

There are six new states created by the addition of gene conversion; these are described in the last six rows of the table. This corresponds to the transition probabilities in the fourth column of Appendix A of McVean (2007).

TABLE A4

Transition probabilities for SNN when the starting configuration is A

Configuration at the end of selection phase	Probability given starting configuration $[i^{S}j^{S}, i^{S}j^{S}]$	Configuration at the start of neutral phase
$[i^{S}j^{S}, i^{S}j^{S}]$	$(q_x^* q_y^* (1-g_s))^2$	O
$[i^{S}j^{W}, i^{S}j^{W}]$	$2q_x^* q_y^* (1-g_s)(p_x + g_s - p_x g_s)(1-g_x) q_y^*$	A
$[i^{S}j^{S}, i^{S}k^{W}]$	$2(q_x^* q_y^* (1-g_s))(q_x^* (1-g_s) p_y^*)$	O
$[i^{S}j^{W}, i^{S}k^{W}]$	$2q_x^* q_y^* (1-g_s)((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*)$	B
$[i^{W}j^{W}, i^{W}j^{W}]$	$((p_x + g_s - p_x g_s)(1-g_x) q_y^*)^2$	A
$[i^{W}j^{S}, i^{W}k^{W}]$	$2(p_x + g_s - p_x g_s)(1-g_x) q_y^* q_x^* (1-g_s) p_y^*$	B
$[i^{S}j^{S}, k^{W}l^{W}]$	$(q_x^* (1-g_s) p_y^*)^2$	O
$[i^{S}j^{W}, k^{W}l^{W}]$	$2q_x^* (1-g_s) p_y^* ((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*)$	C
$[i^{W}j^{W}, i^{W}k^{W}]$	$2(p_x + g_s - p_x g_s)(1-g_x) q_y^* ((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*)$	B
$[i^{W}j^{W}, k^{W}l^{W}]$	$(((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*))^2$	C
$[i^{S}j^{S}, i^{S}k^{S}]$	0	O
$[i^{S}j^{W}, i^{S}k^{S}]$	$2(q_x^* q_y^* (1-g_s))(q_s^* g_x q_y^*)$	O
$[i^{S}j^{S}, k^{S}l^{W}]$	0	O
$[i^{W}j^{S}, i^{W}k^{S}]$	0	A
$[i^{S}j^{W}, k^{S}l^{W}]$	$2q_s^* g_x q_y^* q_x^* (1-g_s) p_y^*$	B
$[i^{W}j^{W}, i^{W}k^{S}]$	$2(p_x + g_s - p_x g_s)(1-g_x) q_y^* q_s^* g_x q_y^*$	B
$[i^{W}j^{W}, k^{S}l^{W}]$	$2q_s^* g_x q_y^* ((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*)$	C
$[i^{S}j^{W}, k^{S}l^{S}]$	0	O
$[i^{W}j^{W}, k^{S}l^{S}]$	$(q_s^* g_x q_y^*)^2$	O
$[i^{S}j^{S}, k^{S}l^{S}]$	0	O

There are no new states created with the addition of gene conversion but there are new transition probabilities. This table corresponds to the transition probabilities in the second column of Appendix B of McVean (2007).

TABLE A5

Transition probabilities for SNN when the starting configuration is B

Configuration at the end of selection phase	Probability given starting configuration $[i^{S}j^{S}, i^{S}k^{S}]$	Configuration at the start of neutral phase
$[i^{S}j^{S}, i^{S}j^{S}]$	0	O
$[i^{S}j^{W}, i^{S}j^{W}]$	0	A
$[i^{S}j^{S}, i^{S}k^{W}]$	$(q_x^* q_y^* (1-g_s))(q_x^* (1-g_s))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y)$	O
$[i^{S}j^{W}, i^{S}k^{W}]$	$q_x^* q_y^* (1-g_s)(p_x^* + (1-g_x) g_s (1-p_x))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y)$	B
$[i^{W}j^{W}, i^{W}j^{W}]$	0	A
$[i^{W}j^{S}, i^{W}k^{W}]$	$(p_x + g_s - p_x g_s)(1-g_x) q_y^* (q_x^* (1-g_s))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y)$	B
$[i^{S}j^{S}, k^{W}l^{W}]$	$q_x^* (1-g_s) p_y^* (q_x^* (1-g_s))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y)$	O
$[i^{S}j^{W}, k^{W}l^{W}]$	$q_x^* (1-g_s) p_y^* (p_x^* + (1-g_x) g_s (1-p_x))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y) + ((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*)(q_x^* (1-g_s))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y)$	C
$[i^{W}j^{W}, i^{W}k^{W}]$	$(p_x + g_s - p_x g_s)(1-g_x) q_y^* (p_x^* + (1-g_x) g_s (1-p_x)) (q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y)$	B
$[i^{W}j^{W}, k^{W}l^{W}]$	$((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*)(p_x^* + (1-g_x) g_s (1-p_x))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y)$	C
$[i^{S}j^{S}, i^{S}k^{S}]$	$q_x^* q_y^* (1-g_s) q_x^* q_x (1-g_s)^2 q_y^*$	O
$[i^{S}j^{W}, i^{S}k^{S}]$	$(q_x^* q_y^* (1-g_s))(p_x^* + g_s q_x^*)(q_x(1-g_s) q_y^*)$	O
$[i^{S}j^{S}, k^{S}l^{W}]$	$q_x^* (1-g_s) p_y^* q_x^* q_x (1-g_s)^2 q_y^*$	O
$[i^{W}j^{S}, i^{W}k^{S}]$	$(p_x + g_s - p_x g_s)(1-g_x) q_y^* q_x^* q_x (1-g_s)^2 q_y^*$	A
$[i^{S}j^{W}, k^{S}l^{W}]$	$q_s^* g_x q_y^* (q_x^* (1-g_s))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y) + q_x^* (1-g_s) p_y^* (p_x^* + g_s q_x^*)(q_x(1-g_s) q_y^*) + ((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*) q_x^* q_x (1-g_s)^2 q_y^*$	B
$[i^{W}j^{W}, i^{W}k^{S}]$	$(p_x + g_s - p_x g_s)(1-g_x) q_y^* (p_x^* + g_s q_x^*)(q_x(1-g_s) q_y^*)$	B
$[i^{W}j^{W}, k^{S}l^{W}]$	$q_s^* g_x q_y^* (p_x^* + (1-g_x) g_s (1-p_x))(q_y^* q_x g_s + q_x q_y g_y + p_x q_y + q_x p_y + p_x p_y) + ((g_s + p_x - p_x g_s)(p_y^* + q_y^* g_x) + q_x(1-g_s) g_x p_y^*)(p_x^* + g_s q_x^*)(q_x(1-g_s) q_y^*)$	C
$[i^{S}j^{W}, k^{S}l^{S}]$	$q_s^* g_x q_y^* q_x^* q_x (1-g_s)^2 q_y^*$	O
$[i^{W}j^{W}, k^{S}l^{S}]$	$q_s^* g_x q_y^* (p_x^* + g_s q_x^*)(q_x(1-g_s) q_y^*)$	O
$[i^{S}j^{S}, k^{S}l^{S}]$	0	O

There are no new states created with the addition of gene conversion but there are no new transition probabilities. This corresponds to the transition probabilities in the third column of Appendix B of McVean (2007).

TABLE A6

Transition probabilities for SNN when the starting configuration is C

Configuration at the end of selection phase	Probability given starting configuration $[i^{S}j^{S},\ k^{S}l^{S}]$	Configuration at the start of neutral phase
$[i^{S}j^{S},\ i^{S}j^{S}]$	0	O
$[i^{S}j^{W},\ i^{S}j^{W}]$	0	A
$[i^{S}j^{S},\ i^{S}k^{W}]$	0	O
$[i^{S}j^{W},\ i^{S}k^{W}]$	0	B
$[i^{W}j^{W},\ i^{W}j^{W}]$	0	A
$[i^{W}j^{S},\ i^{W}k^{W}]$	0	B
$[i^{S}j^{S},\ k^{W}l^{W}]$	$((q_x^*(1-g_s))(q_y^*q_xg_s + q_xq_yg_y + p_xq_y + q_xp_y + p_xp_y))^2$	O
$[i^{S}j^{W},\ k^{W}l^{W}]$	$2(q_x^*(1-g_s))(q_y^*q_xg_s + q_xq_yg_y + p_xq_y + q_xp_y + p_xp_y)(p_x^* + (1-g_x)$ $g_s(1-p_x))*(q_y^*q_xg_s + q_xq_yg_y + p_xq_y + q_xp_y + p_xp_y)$	C
$[i^{W}j^{W},\ i^{W}k^{W}]$	0	B
$[i^{W}j^{W},\ k^{W}l^{W}]$	$(p_x^* + (1-g_x)g_s(1-p_x))(q_y^*q_xg_s + q_xq_yg_y + p_xq_y + q_xp_y + p_xp_y))^2$	C
$[i^{S}j^{S},\ i^{S}k^{S}]$	0	O
$[i^{S}j^{W},\ i^{S}k^{S}]$	0	O
$[i^{S}j^{S},\ k^{S}l^{W}]$	$2q_x^*q_x(1-g_s)^2q_y^*(q_x^*(1-g_s))(q_y^*q_xg_s + q_xq_yg_y + p_xq_y + q_xp_y + p_xp_y)$	O
$[i^{W}j^{S},\ i^{W}k^{S}]$	0	A
$[i^{S}j^{W},\ k^{S}l^{W}]$	$2(q_x^*(1-g_s))(q_y^*q_xg_s + q_xq_yg_y + p_xq_y + q_xp_y + p_xp_y)(p_x^* + g_sq_x^*)$ $(q_x(1-g_s)q_y^*) + 2q_x^*q_x(1-g_s)^2q_y^*(p_x^* + (1-g_x)g_s(1-p_x))$ $(q_y^*q_xg_s + q_xq_yg_y + p_xq_y + q_xp_y + p_xp_y)$	B
$[i^{W}j^{W},\ i^{W}k^{S}]$	0	B
$[i^{W}j^{W},\ k^{S}l^{W}]$	$2(p_x^* + g_sq_x^*)(q_x(1-g_s)q_y^*)(p_x^* + (1-g_x)g_s(1-p_x))(q_y^*q_xg_s + q_xq_yg_y +$ $p_xq_y + q_xp_y + p_xp_y)$	C
$[i^{S}j^{W},\ k^{S}l^{S}]$	$2q_x^*q_x(1-g_s)^2q_y^*(p_x^* + g_sq_x^*)(q_x(1-g_s)q_y^*q_y^*)$	O
$[i^{W}j^{W},\ k^{S}l^{S}]$	$((p_x^* + g_sq_x^*)(q_x(1-g_s)q_y^*))^2$	O
$[i^{S}j^{S},\ k^{S}l^{S}]$	$(q_x^*q_x(1-g_s)^2q_y^*)^2$	O

There are no new states created with the addition of gene conversion but there are new transition probabilities. This corresponds to the transition probabilities in the fourth column of Appendix B of McVean (2007).