

A Combination of *cis* and *trans* Control Can Solve the Hotspot Conversion Paradox

A. D. Peters¹

Department of Zoology and Center of Rapid Evolution, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received November 1, 2007

Accepted for publication January 2, 2008

ABSTRACT

There is growing evidence that in a variety of organisms the majority of meiotic recombination events occur at a relatively small fraction of loci, known as recombination hotspots. If hotspot activity results from the DNA sequence at or near the hotspot itself (in *cis*), these hotspots are expected to be rapidly lost due to biased gene conversion, unless there is strong selection in favor of the hotspot itself. This phenomenon makes it very difficult to maintain existing hotspots and even more difficult for new hotspots to evolve; it has therefore come to be known as the “hotspot conversion paradox.” I develop an analytical framework for exploring the evolution of recombination hotspots under the forces of selection, mutation, and conversion. I derive the general conditions under which *cis*- and *trans*-controlled hotspots can be maintained, as well as those under which new hotspots controlled by both a *cis* and a *trans* locus can invade a population. I show that the conditions for maintenance of and invasion by *trans*- or *cis*-plus-*trans*-controlled hotspots are broader than for those controlled entirely in *cis*. Finally, I show that a combination of *cis* and *trans* control may allow for long-lived polymorphisms in hotspot activity, the patterns of which may explain some recently observed features of recombination hotspots.

THERE is growing evidence from several model systems across the eukaryotic phylogeny that meiotic recombination events, rather than being distributed uniformly across the genome, are largely concentrated into relatively small regions known as “recombination hotspots.” Hotspots in yeast (MALONE *et al.* 1994; WU and LICHTEN 1995; PETES 2001; CROMIE *et al.* 2005), mice (GUILLON and DE MASSY 2002; KELMENSEN *et al.* 2005; SHIFMAN *et al.* 2006; BAUDAT and DE MASSY 2007), and humans (JEFFREYS *et al.* 1998, 2000, 2001, 2005; CRAWFORD *et al.* 2004; McVEAN *et al.* 2004; MYERS *et al.* 2005; CONRAD *et al.* 2006; INTERNATIONAL HAPMAP CONSORTIUM 2007) have now been well characterized, and evidence suggests that hotspots also exist in chimpanzees (PTAK *et al.* 2005) and several plants (DOONER and MARTINEZ-FEREZ 1997; OKAGAKI and WEIL 1997; YAO *et al.* 2002; DROUAUD *et al.* 2006). While some other well-studied eukaryotes (*e.g.*, *Drosophila melanogaster* and *Caenorhabditis elegans*) show no evidence of hotspots (HEY 2004), the phenomenon is widespread enough across the tree of life to merit substantial study.

These hotspots pose a variety of interesting questions for population geneticists, not the least of which is that of their continued existence. One striking characteristic of hotspots is that they are subject to a form of meiotic drive: when a DSB occurs at a hotspot heterozygous for

an active (“hot”) and an inactive (“cold”) hotspot allele, the cold allele tends to appear in a higher proportion of the offspring than does the hot allele (often in an ~3:1 ratio rather than the expected 2:2) (CATCHESIDE 1975; NICOLAS *et al.* 1989; GRIMM *et al.* 1991; MALONE *et al.* 1994; GUILLON and DE MASSY 2002; JEFFREYS and NEUMANN 2002; JEFFREYS and MAY 2004; CROMIE *et al.* 2005). This is likely the result of the mechanism by which recombination is thought to be initiated: a double-strand break (DSB) forms on one chromatid; this break extends a variable distance in the 5' direction on each strand; and the sequence of the nonsister chromatid is used as a template to repair the gap. The physical connections (Holliday junctions) between the two chromatids that form as a result of this repair often (but not always) result in a crossover event (SZOSTAK *et al.* 1983); however, a more direct consequence is that a stretch of DNA sequence on the chromatid that experiences the initial DSB is replaced (“converted”) by homologous sequence from the nonsister chromatid: while the converted sequence was originally present on two of four chromatids (a 2:2 ratio), after DSB repair it is present on only one of four (a 3:1 ratio). Figure 1 shows a diagram of this process. Since DSBs occur at hot alleles more frequently than at cold alleles, they are converted in this manner more frequently as well and are therefore expected to rapidly decrease in frequency when there is a cold allele present in the population. The continued existence of recombination hotspots in the face of this biased gene conversion has been termed the “hotspot conversion paradox” (BOULTON *et al.* 1997).

¹Address for correspondence: Department of Zoology, Birge Hall, 430 Lincoln Dr., University of Wisconsin, Madison, WI 53706.
E-mail: adpeters1@wisc.edu

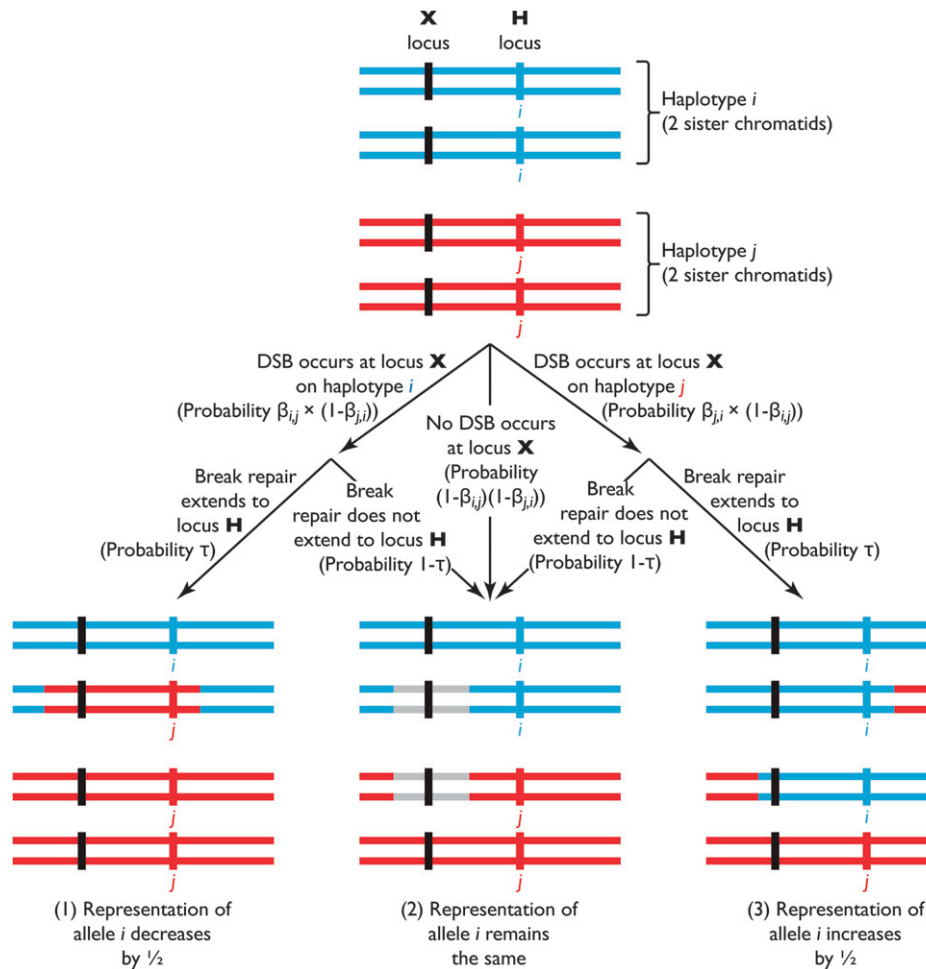


FIGURE 1.—Schematic of double-strand breaks (DSBs) and gene conversion. Shown are homologous regions of two pairs of sister chromatids (each pair is referred to here as a haplotype), coded by color/shading, lined up as for metaphase I of meiosis. This region carries two loci of interest: the **H** locus controls the rate of DSBs, which occur at the **X** locus. We are interested in the rate at which allele *i* (carried by the blue/lightly shaded haplotype) at the **H** locus changes as a result of gene conversion. On this measure, there are three possible outcomes: (1) one copy of allele *i* can be converted to allele *j*, which decreases the representation of allele *i* by a factor of $\frac{1}{2}$; (2) no conversion, which causes no change to the representation of allele *i*; and (3) one copy of allele *j* on one chromatid can be converted to allele *i*, which increases the representation of allele *i* by $\frac{1}{2}$. The schematic details the paths by which these outcomes can occur and the associated probabilities. The diagram also shows that some, but not all, conversion events can be accompanied by crossover (recombination) events; in this case, a crossover occurs in the rightmost set of chromosomes (3) but not in the left two (1 and 2). The gray regions in the middle set

of chromosomes (2) may come from either original chromatid, depending on the path followed to yield this outcome. Note that the probability that DSBs occur in both haplotypes is ignored in this figure but not in the models presented in the text.

This paradox has been addressed in several recent theoretical studies (BOULTON *et al.* 1997; PINEDA-KRCH and REDFIELD 2005; CALABRESE 2007; COOP and MYERS 2007), but a general solution to the problem remains elusive. Simulation studies show that selection favoring DSBs can counteract the process of biased gene conversion and slow or stop the loss of a hotspot, but only if there is a strong fitness benefit directly associated with activity at the hotspot itself (BOULTON *et al.* 1997; PINEDA-KRCH and REDFIELD 2005). This appears to largely preclude the classical evolutionary explanation for recombination, which appeals to the ability of recombination to break down associations between alleles at multiple loci (linkage disequilibrium) and therefore potentially increase the ability of a population to respond to natural selection (BARTON 1995). Since this classical explanation works through indirect fitness effects, in which alleles that increase recombination become associated with high-fitness genotypes at other loci, it is unlikely to provide strong enough selection on the recombination alleles (*i.e.*, hotspots) themselves to overcome the force of biased gene conversion (BOULTON *et al.* 1997).

There is some reason to think that DSBs (and therefore hotspots) may provide direct fitness benefits. In species with recombination hotspots (but not those without), DSBs serve as a trigger for the formation of the synaptonemal complex and therefore guarantee the correct segregation of chromosomes in meiosis (HEY 2004). This potentially provides a source of selection directly on hotspots; unfortunately, previous models have suggested that this selection is unlikely to provide strong enough selection to overcome biased gene conversion (BOULTON *et al.* 1997; PINEDA-KRCH and REDFIELD 2005; COOP and MYERS 2007).

Because selection must be unrealistically strong to provide a way around the conversion paradox by itself, recent theoretical work has concentrated on other potential escapes. One appealing approach is to allow hotspot activity to be controlled not by the sequence at the hotspot itself (*in cis*), but by the sequence at a site some distance away (*in trans*). Under such circumstances, DSB repair at the hotspot may only infrequently (or never) extend to the controlling locus; biased gene conversion at the hotspot itself would then not necessarily translate to biased gene conversion of hotspot activity.

But is *trans* control of hotspots likely? The mode of control of hotspots remains poorly understood. It is clear that *cis* control is important in many cases: several specific sequences (often, 6- to 10-bp motifs) have been shown to be associated with hotspot activity (PETES 2001; JEFFREYS and NEUMANN 2002; MYERS *et al.* 2005), although importantly these motifs appear to be neither necessary nor sufficient to form hotspots (PETES 2001; MYERS *et al.* 2005). More direct evidence for *cis* control comes from several examples in which polymorphisms in hotspot activity have been shown to be associated with sequence polymorphisms at the hotspots themselves (JEFFREYS and NEUMANN 2002, 2005; BAUDAT and DE MASSY 2007). However, there is also substantial scope for *trans* control: in yeast, many hotspots require the action of sequence-specific transcription factors (PETES 2001); changes in the specificity or activity of these or any other DSB-repair-specific proteins would be expected to have remote-acting effects on hotspot activity. Similarly, the details of chromatin structure and modification in a region appear to affect hotspot activity in some cases; in particular, regions with a relatively open chromatin structure provide better access for the recombination machinery (PETES 2001). Mutations at remote loci that change the patterns of histone modification, for example, might open up entirely new regions as potential hotspots. More specifically, recent evidence suggests that some polymorphisms in hotspot activity map only partially or not at all to local sequence variation (NEUMANN and JEFFREYS 2006; BAUDAT and DE MASSY 2007), suggesting that variation in *trans* or epigenetic variation is responsible for some variation in hotspot activity. A combination of *cis* and *trans* control is therefore likely to be a fruitful direction for theoretical exploration.

Recent comparisons of hotspot distributions in humans and chimpanzees have highlighted further mysteries about recombination hotspots. Many hotspots are identified by the “chunks” of linkage disequilibrium—“haplotype blocks”—they cause in the genome. The fact that hotspots cause these signatures implies that they are maintained for thousands of generations. However, a large fraction of hotspots that are present in humans are absent in chimpanzees (PTAK *et al.* 2004, 2005; WINCKLER *et al.* 2005), suggesting that hotspots have been gained (and/or lost) over the scale of hundreds of thousands of generations. Combined with evidence, based on the comparison of historical rates of recombination (the strength of signal from haplotype blocks) with current rates (from sperm-typing studies), that the “heat” of at least some hotspots is decreasing in humans (JEFFREYS *et al.* 2005), this suggests a picture in which hotspots arise and are then slowly lost over thousands of generations, causing turnover of hotspots at larger timescales.

But what can explain this pattern? Biased gene conversion can easily explain the loss of hotspots, but what can explain the appearance of new ones? Given that the

conversion paradox is likely to be at its strongest when a hot allele is rare (so that the hot allele is always found in heterozygotes and is therefore constantly being lost to conversion), explaining the spread of a new hotspot is more difficult than explaining the existence of an old one. Recent theoretical studies exploring the action of genetic drift on modifiers undergoing biased gene conversion suggest that only when the force of conversion (that is, the product of the DSB rate and the probability that the control allele is converted when a DSB occurs) is weak can a hotspot spread or be maintained at high frequency (CALABRESE 2007; COOP and MYERS 2007). That is, very-hot *cis*-controlled hotspots act essentially as strongly deleterious mutations, which almost never fix due to drift. One possible mechanism by which very hot hotspots may arise is if biased gene conversion does not act against them on their initial spread through the population (COOP and MYERS 2007): again, this might be achieved if the hotspot is, at least initially, controlled from a locus located in *trans*.

Theoretical approaches to the hotspot conversion paradox to date have consisted of simulations of finite populations (BOULTON *et al.* 1997; PINEDA-KRCH and REDFIELD 2005) and diffusion approximations focusing on the stochastic fate of hotspots (CALABRESE 2007; COOP and MYERS 2007). Here, I present an analytical model of the deterministic processes underlying the evolution of a single recombination hotspot, to clearly define the conditions under which selection can overcome the effects of biased gene conversion and allow the maintenance of an existing hotspot and/or the spread of a new one. I focus on the evolution of hotspots under *cis* control, *trans* control, or a combination of the two, with emphasis on those conditions under which a hotspot under complete or partial *trans* control can spread but one under pure *cis* control cannot.

ONE-LOCUS MODEL

I begin by examining the fate of two alleles— H (the hot allele) and h (the cold allele)—at locus \mathbf{H} (for the hotspot-control or simply the “control” locus), under the forces of mutation, selection, and biased gene conversion. The genotype at the \mathbf{H} locus determines the DSB rate at the potential hotspot, which I refer to as locus \mathbf{X} (as in, \mathbf{X} marks the spot); the DSB rate in turn determines fitness and the rate of gene conversion. There is no allelic variation at the \mathbf{X} locus *per se*; \mathbf{X} is simply the physical location at which DSBs occur, although the \mathbf{H} locus and the \mathbf{X} locus may be one and the same, in which case the control of the hotspot is in *cis*. Alternatively, the \mathbf{H} locus may be located some arbitrary distance from the \mathbf{X} locus (it may, in fact, be located on a different chromosome, although to minimize confusion in the text I refer to the \mathbf{X} and \mathbf{H} loci as being on the same chromosome), in which case the control locus is in *trans* to the hotspot. In the following I describe in

TABLE 1
Notation used in the text and the model

Notation	Definition
X	Hotspot locus
H , H , h	One-locus model: control locus, cold allele, hot allele Modifier model: <i>cis</i> locus, cold allele, hot allele
M , M , m	Modifier locus, active allele, inactive allele
h_i	Dominance of i allele over alternative allele (at H locus in one-locus model and at M locus in modifier model)
β_H , β_h	“Hot” double-strand break rate, “cold” double-strand break rate
β_{ij}	Per-haplotype double-strand break rate of chromosome i in diploid genotype ij
p_i , p_{ij}	Frequency of allele i , frequency of chromosome ij
\hat{p}_i , \hat{p}_{ij}	Equilibrium frequencies
σ	Selection gradient: slope of relationship between fitness and DSB rate
μ , ν	Mutation rates at H locus ($H \rightarrow h$, $h \rightarrow H$)
τ	One-locus model: probability that locus H is converted when a DSB occurs at locus X Modifier model: probability that locus M is converted when locus H is converted
r	Rate of recombination between H and M loci (modifier model only)

more detail the workings of the model. A summary of all important notation in this article can be found in Table 1, and a diagram of key features of gene conversion can be seen in Figure 1. All analyses described below were performed using Mathematica 6.0 (WOLFRAM RESEARCH 2007).

Double-strand break rate: The per-chromatid DSB rate at the hotspot locus **X** of a given homolog in a diploid genotype is determined by the allele carried at the control locus, **H**, on the same homolog, possibly with some dominance effect arising from the allele carried at locus **H** on the other homolog. This requires two DSB rates for each diploid genotype ij : $\beta_{i,j}$ which describes the DSB rate on the homolog carrying allele i when the other homolog carries allele j ; and $\beta_{j,i}$ which describes the DSB rate on the homolog carrying allele j when the other homolog carries allele i . Individuals homozygous for allele H have a per-chromatid DSB rate of $\beta_H \equiv \beta_{H,H}$, while those homozygous for allele h have a per-chromatid DSB rate of $\beta_h \equiv \beta_{h,h}$ ($\beta_H > \beta_h$). DSB rates in heterozygotes are determined by a pair of “dominance coefficients”: h_H , which denotes the extent to which the presence of allele H on the homologous chromosome increases the DSB rate associated with the h allele, and h_h , which denotes the extent to which the presence of allele h on the homologous chromosome decreases the DSB rate associated with the H allele (in this article, one or both of the dominance coefficients are always set to zero, and $0 \leq h_i \leq 1$). All in all, in heterozygotes the DSB rate associated with allele H is $\beta_{H,h} = \beta_H - h_h(\beta_H - \beta_h)$, and that associated with allele h is $\beta_{h,H} = \beta_h + h_H(\beta_H - \beta_h)$. Table 2 summarizes the values of $\beta_{i,j}$ for each diploid one-locus genotype.

Gene conversion at the **H locus:** The probability that the control locus, **H**, undergoes conversion when a DSB occurs at locus **X** is given by τ . This parameter can be thought of as the *trans* parameter: the greater the physical distance is between loci **X** and **H**, the lower the

probability that a DSB at locus **X** causes a conversion at **H**. $\tau = 1$ therefore implies that control of the hotspot is completely in *cis*, while $\tau = 0$ implies that control is completely in *trans*. Note that this yields a very broad definition of “*trans*” (and a correspondingly narrow definition of “*cis*”): the **H** locus may be only a few kilobases away from the **X** locus and still have $\tau = 0$. Given these assumptions, the differential equation describing the rate of change of the frequency of allele H (p_H) due to conversion is

$$\frac{d_c p_H}{dt} = (1 - p_H)p_H \frac{\tau}{2} (\beta_{h,H} - \beta_{H,h}), \quad (1)$$

where d_c denotes change due to conversion; similar subscripts are used for selection and mutation below.

Selection: In reality, selection on overall DSB rate within a chromosomal region likely results from some combination of the direct benefits (*e.g.*, proper chromosome segregation) and the indirect benefits [which arise in turn from linkage disequilibria between loci flanking the DSB site and require models explicitly including selection on two or more viability loci to evaluate completely (BARTON 1995)] of crossovers. These overall rates seem likely to be under stabilizing selection; however, as with any quantitative trait under stabilizing selection,

TABLE 2
Double-strand break rates in one-locus models

	$i = h$ ($\beta_{h,j}$)	$i = H$ ($\beta_{H,j}$)
$j = h$ ($\beta_{i,h}$)	β_h	$\beta_H - h_h(\beta_H - \beta_h)$
$j = H$ ($\beta_{i,H}$)	$\beta_h + h_H(\beta_H - \beta_h)$	β_H

$\beta_{i,j}$ gives the per-haplotype DSB rate of haplotypes carrying the i allele when the other haplotype in a diploid genotype carries the j allele. In the H -dominant model, h_h is set to zero; in the h -dominant model, h_H is set to zero; in the no-dominance model, both h_H and h_h are set to zero.

TABLE 3
Stability conditions

	No dominance ($h_H = h_h = 0$)	H partially dominant ($h_h = 0$; $0 < h_H < 1$)	h partially dominant ($h_H = 0$; $0 < h_h < 1$)
A. One-locus model			
$\hat{p}_H \approx 1$ stable (hotspot maintained)	$\sigma > \frac{\tau}{2(1 - \beta_H)}$	$\sigma > \frac{\tau}{2(1 - \beta_H)}$	$\sigma > \frac{\tau(1 - h_h)}{2(1 + h_h - \beta_H)}$
$\hat{p}_H \approx 0$ unstable (hotspot increases in frequency)	$\sigma > \frac{\tau}{2}$	$\sigma > \frac{\tau(1 - h_H)}{2(1 + h_H(1 - \beta_H))}$	$\sigma > \frac{\tau}{2}$
$\hat{p}_H \approx 0$ unstable (assuming β_H and $\beta_h - \beta_n$ are small)	$\sigma > \frac{\tau}{2}$	$\sigma > \frac{\tau(1 - h_H)}{2(1 + h_H)}$	$\sigma > \frac{\tau}{2}$
B. Modifier model			
$p_M = 0$ unstable (hotspot-activating modifier increases in frequency)	$\sigma > \frac{\tau}{2}$	$\sigma > \frac{\tau(1 - h_H)}{2(1 + h_H)}$	$\sigma > \frac{\tau}{2}$

any single allele causing a small difference in DSB rates relative to the total DSB rate in the region may be under directional selection (FALCONER and MACKAY 1996). Because I am interested in evaluating the strength of selection required for hotspots to be maintained or increase in frequency, I assume that there is directional selection on DSB rate at the \mathbf{X} locus. In particular, I assume that fitness increases as a linear function of the probability that a DSB occurs at the \mathbf{X} locus; that is, the fitness of diploid genotype ij is $1 + s_{ij}$, where the selection coefficient is $s_{ij} = (1 - (1 - \beta_{i,j})(1 - \beta_{j,i}))\sigma$ and σ is the selection gradient for DSB rate at locus \mathbf{X} (*i.e.*, the slope of the relationship between DSB rate and fitness). Under these conditions, the differential equation for the change in p_H due to selection is

$$\frac{d_s p_H}{dt} = (1 - p_H)p_H(s_{HH} - s_{hh} + p_H(s_{HH} - 2s_{Hh} + s_{hh})). \quad (2)$$

Mutation: Finally, I assume that mutation from $H \rightarrow h$ occurs at rate μ , and mutation from $h \rightarrow H$ occurs at rate ν , yielding the differential equation

$$\frac{d_m p_H}{dt} = \nu - (\mu + \nu)p_H. \quad (3)$$

Mutation–selection–conversion equilibria: The overall equation describing the rate of change of the frequency of allele H is $dp_H/dt = d_s p_H/dt + d_m p_H/dt + d_c p_H/dt$. I used Taylor series to approximate this and all further equations under the assumption that the DSB rate induced by the cold allele and mutation are both weak forces [*i.e.*, that β_h , μ , and ν are $O(\zeta)$, where $\zeta \ll 1$]. Numerical analysis suggested that there is an equilibrium near $p_H = 1$ and one near $p_H = 0$ (*i.e.*, near-fixation and near-loss of the hot allele). To express these equilibria analytically, I approximated the differential equation under further assumptions: to find the equilibrium

when $p_H \approx 1$ (near-fixation, *i.e.*, maintenance of the hotspot), I approximated around the point $p_H \approx 1$ [*i.e.*, assuming that $1 - p_H$ is $O(\zeta)$]; to find the equilibrium when $p_H \approx 0$ (loss of the hotspot), I approximated around the point $p_H \approx 0$ [*i.e.*, assuming that p_H is $O(\zeta)$]. Under these assumptions, the first-order approximation of the equilibrium near $p_H = 0$ is

$$\hat{p}_H \approx \frac{\nu}{\beta_H(\tau((1 - (h_h + h_H))/2) - \sigma(1 - h_h + h_H - \beta_H(h_H - h_h h_H)))} \quad (4)$$

and that of the equilibrium near $p_H = 1$ is

$$\hat{p}_H \approx 1 - \frac{\mu}{\beta_H(\sigma(1 + h_h - h_H - \beta_H(1 - h_H + h_h h_H)) - \tau((1 - (h_h + h_H))/2))}. \quad (5)$$

Recall that for this analysis $h_H = 0$ (h partially dominant), $h_h = 0$ (H partially dominant), or $h_H = h_h = 0$ (no dominance); these substitutions simplify these expressions considerably.

These equilibria are similar in form to the classic diploid mutation–selection balance, in which the disfavored allele is held at a frequency of μ/h_s , where μ is the mutation rate to the disfavored allele and h_s is the strength of selection against the heterozygote. Here, which allele is “disfavored” is determined by the relative strengths of conversion (the τ -terms in Equations 4 and 5) and selection (the σ -terms).

Of more interest than the expressions describing the precise frequencies at equilibrium are the conditions under which equilibrium (5) is stable (*i.e.*, the hotspot can be maintained at high frequency) and those under which equilibrium (4) is unstable (*i.e.*, the hotspot can increase in frequency when rare). These conditions are summarized in Table 3A. Also included in Table 3A is a further approximation of the conditions under which Equation 4 is unstable; this approximation assumes that

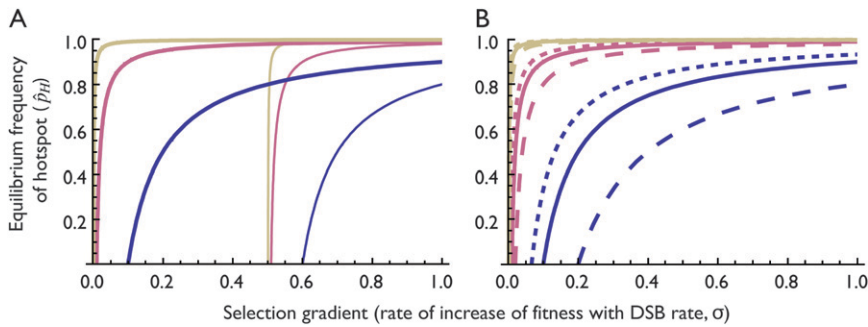


FIGURE 2.—Stable “near-fixation” equilibria in the one-locus model. Lines show the values of the $\hat{p}_H \approx 1$ equilibrium over the range for which it is valid and stable for a variety of parameter values. (A) *cis* vs. *trans* control with no dominance ($h_H = h_h = 0$). Thin lines represent *cis* control ($\tau = 1$); thick lines represent *trans* control ($\tau = 0$). (B) The effects of dominance under *trans* control ($\tau = 0$). Solid lines represent no dominance ($h_H = h_h = 0$; the same as seen in A); short dashes represent *H* partially dominant ($h_h = 0, h_H = \frac{1}{2}$); long dashes represent *h* partially dominant ($h_H = 0, h_h = \frac{1}{2}$).

represent *h* partially dominant ($h_H = 0, h_h = \frac{1}{2}$). (A and B) Color/shading represents the DSB rate of the hot allele (β_H): blue/dark lines, $\beta_H = 10^{-6}$; magenta/medium lines, $\beta_H = 10^{-5}$; gold/light lines, $\beta_H = 10^{-4}$. Higher DSB rates ($\beta_H > 10^{-4}$) closely resemble the pattern for $\beta_H = 10^{-4}$. Other parameter values: $\mu = \nu = 10^{-7}$. The drop to zero at low values of σ in the *trans* case is a side effect of the approximation; see supplemental material for exact results.

the difference in DSB rate between the *H* and *h* alleles is small [*i.e.*, $\beta_H - \beta_h$ is $O(\xi)$] and therefore β_H itself is small (but still greater than β_h). This approximation is included to facilitate comparison with the modifier analysis below, which makes the same assumption.

A detailed examination of the stability conditions in Table 3A leads to several conclusions. First, as shown by simulation studies (BOULTON *et al.* 1997; PINEDA-KRCH and REDFIELD 2005), selection must be strong to maintain a hotspot whose activity is controlled by *cis* sequence ($\tau = 1$). If we assume that *cis*-acting sequences experience no dominance effects ($h_H = h_h = 0$), to maintain a *cis* hotspot selection must be strong enough that $\sigma > \frac{1}{2}(1 - \beta_H)$ (that is, $\sigma > \sim \frac{1}{2}$). Recall that σ represents the rate of increase of fitness with DSB rate; similarly, under *cis* control, $\frac{1}{2}$ represents the rate of increase of conversions with DSB rate (Figure 1). Thus, near-fixation is stable (and hotspots are maintained) whenever fitness increases more rapidly than conversion with increases in DSB rate. Put more concretely, this can be visualized as requiring that individuals that experience a DSB at the **X** locus have 50% higher fitness than those that do not (although note that formally the increase in fitness could be due to effects of the *H* allele other than DSBs *per se*). In other words, the fitness advantage must exceed the 3:1 transmission disadvantage of the hot allele in heterozygotes undergoing a DSB.

Another important conclusion from Table 3A is that the conditions for maintenance ($p_H \approx 1$ stable) or invasion ($p_H \approx 0$ unstable) of the hotspot are more stringent for *cis* control than for *trans* control. As the probability that the control (**H**) locus is converted alongside the hotspot (τ) decreases, the strength of selection required for maintenance or invasion decreases proportionally: for fully *trans* loci ($\tau = 0$), any $\sigma > 0$ suffices. In general, dominance ($h_H > 0$ or $h_h > 0$) also decreases the strength of selection required for invasion or maintenance of a hotspot, largely because dominance decreases the difference in DSB (and therefore conversion) rate of the hot and cold alleles in heterozygotes, which is where biased gene conversion occurs.

Figure 2 shows the “near-fixation” equilibrium frequency of allele *H* as a function of σ for the range over which it is stable for a variety of parameter values. As seen in Table 3A, *trans* control allows the maintenance of hotspots over a broader range of selection strengths than does *cis* control; Figure 2A demonstrates in addition that *trans*-controlled hotspots are held at higher frequencies at equilibrium than are *cis*-controlled hotspots. In addition, Figure 2B demonstrates that partial dominance at the **H** locus affects the equilibrium frequency of the *H* allele under *trans* control. In particular, if the *H* allele is partially dominant ($h_h = 0; 0 < h_H < 1$), the frequency of *H* at equilibrium is higher than under no dominance; if the *h* allele is partially dominant ($h_H = 0; 0 < h_h < 1$), the frequency of *H* is lower than under no dominance. This is due to the fact that when allele *H* is at high frequency, its fitness advantage is slightly elevated when it is dominant and slightly decreased when it is recessive.

Finally, it is worth noting that, although these equilibria are determined under the assumption that the equilibrium frequency of the hotspot is very close to one ($p_H \approx 1$), the actual equilibrium values can be substantially less than one, particularly for *trans*-controlled hotspots when the fitness advantage of increasing DSB rate (σ) and/or the DSB rate of the hot allele (β_H) are low (Figure 2). This is because under these conditions the standard mutation–selection balance assumption that selection is a much stronger force than mutation is violated: the total strength of selection in this model is on the order of $\beta_H \times (\sigma - \tau/2)$, which may be quite small under reasonable parameter values. One consequence of this is that it potentially allows for hotspot polymorphisms to be maintained deterministically; however, because this tends to occur when selection is weak, drift may be more important than selection under these parameter values (CALABRESE 2007; COOP and MYERS 2007). A comparison of these approximate results to the exact solutions, along with confirmation through numerical results, can be found in the supplemental material.

TWO-LOCUS (CIS HOTSPOT AND TRANS MODIFIER) MODEL

The one-locus model described above confirms that strong selection is required if a *cis*-controlled hotspot is to be maintained at an appreciable frequency, while *trans*-controlled hotspots can be maintained over a much broader range of selection strengths. Thus, in this purely deterministic model, there are two solutions to the hotspot conversion paradox: (1) allow strong selection on modifiers or (2) allow hotspot activity to be controlled in *trans*. Unfortunately, neither of these solutions offers a complete escape from the paradox: selection on individual hotspots is not expected to be particularly strong (COOP and MYERS 2007) (but see DISCUSSION below), and there is convincing evidence that many hotspots are at least partially under *cis* control (PETES 2001; JEFFREYS and NEUMANN 2002, 2005; MYERS *et al.* 2005; BAUDAT and DE MASSY 2007). Thus, the hotspot paradox remains for *cis*-controlled modifiers where selection is weak ($\sigma < \tau/2(1 - \beta_H)$).

One possible explanation for the existence of hotspots under such circumstances may be that hotspots are under some combination of *cis* and *trans* control; that is, there must be a specific sequence at the hotspot locus itself to induce an increased frequency of DSBs, but that frequency increase occurs only when a particular allele is carried at another (*trans*) locus. This might be the case if, for example, a *trans*-acting regulating protein induces DSBs at a specific *cis* sequence or if the *trans*-acting protein affects histone modifications in a chromosomal region, which in turn control the accessibility of *cis* trigger sequences in that region to the DSB machinery.

To explore hotspot dynamics under such circumstances, I extend the one-locus model described above. First, I assume that the **H** locus is completely *cis* to the **X** locus (that is, that DSBs occur at the **H** locus—**H** is now both the control locus and the hotspot itself, so we ignore locus **X**) and that the effects of the two alleles at this locus (*h* and *H*) are determined by the genotype at “modifier” locus **M**, which has two alleles—an “active” allele (*M*) and an “inactive” allele (*m*). I am particularly interested in the extent to which the *trans* modifier can lead to the formation of new hotspots under circumstances where individual hotspots under purely *cis* control are expected to be lost.

Extensions of the one-locus model: To build the modifier model, I generalize the one-locus model in several ways. First, I define separate DSB rates for every possible chromosome (the “inactive-cold” chromosome, *hm*; the “inactive-hot” chromosome, *Hm*; the “active-cold” chromosome, *Hm*; and the “active-hot” chromosome, *HM*) for every possible diploid genotype in which it might be found (note that, as for the **X** and **H** loci in the one-locus model, while I refer to the **H** and **M** loci as being on the same chromosome for clarity, the mathematical ma-

TABLE 4

DSB rates in *cis*-plus-*trans* modifier models

	<i>i = hm</i> ($\beta_{hm,i}$)	<i>i = hM</i> ($\beta_{hM,i}$)	<i>i = Hm</i> ($\beta_{Hm,i}$)	<i>i = HM</i> ($\beta_{HM,i}$)
<i>j = hm</i> ($\beta_{i,hm}$)	β_h	β_h	β_h	$\beta_H - h_m(\beta_H - \beta_h)$
<i>j = hm</i> ($\beta_{i,hM}$)	β_h	β_h	$\beta_h + h_M(\beta_H - \beta_h)$	β_H
<i>j = HM</i> ($\beta_{i,Hm}$)	β_h	β_h	β_h	$\beta_H - h_m(\beta_H - \beta_h)$
<i>j = HM</i> ($\beta_{i,HM}$)	β_h	β_h	$\beta_h + h_M(\beta_H - \beta_h)$	β_H

$\beta_{i,j}$ gives the per-haplotype DSB rate at the **H** locus of chromosome *i* when in a diploid genotype with chromosome *j*. For example, in a diploid genotype made up of chromosomes *HM* and *hm*, to determine the DSB rate of the *HM* chromosome, *i = HM* and *j = hm*, yielding $\beta_{HM,hm} = \beta_H - h_m(\beta_H - \beta_h)$; to determine the DSB rate of the *hm* chromosome, *i = hm* and *j = HM*, yielding $\beta_{hm,HM} = \beta_h$. In the *M*-dominant model, *h_m* is set to zero; in the *m*-dominant model, *h_M* is set to zero; in the no-dominance model, both *h_M* and *h_m* are set to zero.

chinery does not require them to be on the same chromosome; it applies as well to any pair of alleles at the **H** and **M** loci that are inherited from the same parent). In this article, all of these DSB rates are defined as functions of β_H and β_h , the hot and cold DSB rates as seen in the one-locus model: individuals homozygous for the inactive modifier allele (*m*) experience the cold DSB rate no matter what alleles they carry at the **H** locus, while individuals homozygous for the active modifier allele (*M*) undergo DSB rates depending on their genotypes at the **H** locus as in the one-locus, no-dominance model above. DSB rates in *Mm* heterozygotes depend on the genotype at the **H** locus, as well as on dominance at the **M** locus, as given by dominance coefficients *h_M* and *h_m* in a manner analogous to *h_H* and *h_h* in the one-locus model. There is assumed to be no dominance at the **H** locus (*i.e.*, *h_H* = *h_h* = 0). Table 4 shows the specific values of these DSB rates for all chromosomes in all genotypes. While the **H** locus is assumed to be in *cis*, the **M** locus may be in *cis* or *trans*; similar to the one-locus model, the probability that the **M** locus undergoes conversion when a DSB occurs at the **H** locus is described by the parameter τ . The differential equation describing the rate of change of the frequency of haplotype *xy* due to conversion is

$$\frac{d_c p_{xy}}{dt} = p_{xy} \times \left\{ p_{xy} + p_{xy'} \left(1 - \frac{\tau \beta_{xy,xy'}}{2} + \frac{\tau \beta_{xy',xy}}{2} \right) + p_{x'y} \left(1 - \frac{\beta_{xy,x'y}}{2} + \frac{\beta_{x'y,xy}}{2} \right) + p_{x'y'} \left(1 - \frac{\beta_{xy,x'y'}}{2} + \frac{\tau \beta_{x'y',xy}}{2} \right) \right\} + p_{xy'} p_{x'y} \frac{(1 - \tau) \beta_{x'y,xy'}}{2} - p_{x,y} \tag{6}$$

where x' and y' represent the alternative alleles at loci **H** and **M**, respectively.

The fitnesses of diploid genotypes are then defined as a linear function of the probability that a DSB occurs in that genotype, as defined in the one-locus model: $w_{ij} = 1 + s_{ij}$, where $s_{ij} = (1 - (1 - \beta_{i,j})(1 - \beta_{j,i}))\sigma$; i and j now refer to haplotypes (e.g., hM), and $\beta_{i,j}$ and $\beta_{j,i}$ refer to the DSB rates at the **H** locus of haplotypes i and j , respectively, when found in genotype ij (Table 4). The differential equation describing the change in frequency of haplotype xy due to selection is therefore

$$\frac{d_s p_{xy}}{dt} = p_{xy} \left(\sum_{i \in \{hM, Hm, HM\}} p_i (s_{i,xy} - s_{hm,xy}) + s_{hm,xy} - \bar{s} \right), \quad (7)$$

where $\bar{s} = \sum_{j \in \mathbb{G}} p_j \sum_{k \in \mathbb{G}} p_k s_{j,k}$ is the mean selection coefficient, and \mathbb{G} is the set of all chromosomes $\{hm, hM, Hm, HM\}$.

Mutation is assumed not to occur at the **M** locus, but occurs at the **H** locus at rates μ and ν as defined in the one-locus model; that is, mutation will convert only between haplotypes $hm \leftrightarrow Hm$ or $hM \leftrightarrow HM$. The differential equation describing the change of the frequency of haplotypes HM and Hm due to mutation is

$$\frac{d_m p_{Hy}}{dt} = \nu p_{hy} - \mu p_{Hy} \quad (8)$$

and that describing the change of frequency of haplotypes hM and hm is

$$\frac{d_m p_{hy}}{dt} = \mu p_{Hy} - \nu p_{hy}, \quad (9)$$

where y indicates the allele carried at the **M** locus.

Finally, I assume that the rate of recombination between the **H** and **M** loci is r . Note that, while DSB activity at the **H** locus is expected to affect the rate of recombination between loci on opposite sides of the **H** locus, it is not necessarily expected to affect the rate of recombination between **H** itself and other loci; hence, for this model, I treat r as independent of $\beta_{i,j}$. In addition, because I concentrate primarily on fully *trans* modifiers, which I imagine to be on the order of kilobases or more away from the hotspot and therefore well outside the range of repair of DSBs originating at that hotspot ($\tau = 0$) but still potentially linked ($r < \frac{1}{2}$), I treat r as independent of τ . The differential equation describing the rate of change of the frequency of haplotype xy due to recombination is

$$\frac{d_r p_{xy}}{dt} = r(p_{x'y} p_{xy'} - p_{xy} p_{x'y'}), \quad (10)$$

where x' and y' represent the alternative alleles at loci **H** and **M**, respectively.

The modifier analysis proceeds by determining under what conditions a rare mutant at the **M** locus can increase in frequency (invade) a population at equilibrium at the **H** locus (the “resident population”). This requires a set of three equations that describe the changes in frequency of four chromosome types (as opposed to the one-locus model, which required one equation to describe the changes in frequency of two alleles). Complete details of this analysis can be found in the supplemental material.

Since I am interested here in the origin of new hotspots, I assume that the resident population is fixed for the inactive (m) allele at the **M** locus, which causes both the H and the h alleles at the **H** locus (and therefore all diploid genotypes) to have the same DSB rate (β_H). Under these circumstances there is no biased gene conversion and no selection, so the equilibrium frequency of allele H is determined by mutation alone: $\hat{p}_H = \nu/\mu + \nu$. To examine the ability of a modifier to spread when an allele at the **H** locus is at an arbitrary frequency (e.g., fixed), I also consider populations with no mutation (in which any allele frequency is an equilibrium in the absence of gene conversion/selection).

This process involves showing that the resident population is in equilibrium under the full system of three equations when the frequency of the M allele (p_M) is set to zero and then determining under what conditions the M allele will increase in frequency, and at what rate (OTTO and DAY 2007). To yield rates that are amenable to interpretation, in addition to the approximations described for the one-locus model, I approximate under the assumption that the modifier is weak; that is, $\beta_H - \beta_h$ is small ($O(\xi)$), and therefore β_H itself is small (but still greater than β_h). In a manner similar to the one-locus model, I separately consider the cases in which the M allele is partially dominant to the m allele ($h_m = 0$ and $0 < h_M < 1$ in Table 4), the m allele is partially dominant to the M allele ($h_M = 0$ and $0 < h_m < 1$ in Table 4), or neither modifier allele is dominant ($h_M = h_m = 0$ in Table 4).

Under these assumptions, over the vast majority of the parameter space the first-order approximation of the rate of change of the frequency of allele M (the active modifier allele) when the resident population is fixed for allele m is

$$\lambda \approx p_{Hm,0} \left(\sigma(1 - h_m + h_M) - \frac{\tau}{2}(1 - h_m - h_M) \right) (\beta_H - \beta_h), \quad (11)$$

where $p_{Hm,0}$ is the initial frequency of allele H (and therefore of chromosome Hm). This equation describes the rate of change of the M allele both when the resident population is at its mutational equilibrium (in which case $p_{Hm,0} = \nu/\mu + \nu$) and when we ignore mutation (i.e., $\mu = \nu = 0$, in which case $p_{Hm,0}$ can be any frequency). The active modifier M will increase in frequency when

$\lambda > 0$; the conditions under which this requirement is met (assuming that $\beta_H > \beta_h$; *i.e.*, the hot allele increases DSB rate when found with the active modifier) are given in Table 3B.

A comparison of Table 3B with Table 3A shows that the conditions allowing for the spread of a *trans* modifier that affects the DSB rate of *cis* allele *H* are identical to those allowing for the spread of the hot allele in the one-locus model, when both are subject to the same approximation (*i.e.*, comparing the third row of Table 3A with Table 3B). As with the one-locus model, a modifier located completely in *trans* of the hotspot ($\tau = 0$) can spread under any selective conditions as long as there is some fitness benefit to DSB formation ($\sigma > 0$). This allows a new, *cis*-controlled hotspot to initially increase in frequency even when an increase in DSB rate causes only a slight increase in fitness: that is, it may offer a way around the hotspot conversion paradox even for *cis*-controlled hotspots under weak selection. Whether this initial frequency increase translates into measurable frequencies of such hotspots in real populations requires exploration of longer-term processes, as discussed below.

Overall, Equation 11 and the associated invasion conditions show that, as with Equations 4 and 5 and their stability conditions, the fate of the system is largely determined by the difference between a selection (σ) term and a conversion (τ) term. One potentially surprising aspect to this result is the fact that the rate of recombination between the **H** and **M** loci (r) does not affect the rate of spread in the first-order approximation (but note that, when r is very low, a more permissive invasion condition takes effect under the no-mutation model; see supplemental material. This condition is not discussed in detail here, because the parameter range under which it is relevant is quite restrictive). In “traditional” modifier models of the evolution of recombination, the rate of recombination between the modifier locus and viability loci is often an important parameter in determining whether or not the modifier can spread (BARTON 1995). The fact that r is not important in this case is almost certainly due to the fact that the modifier has a direct effect on fitness in this model (*i.e.*, the fitness of individuals with the *HM* chromosome is by definition higher than those without it), whereas in standard recombination modifier models any fitness benefit to the modifier must arise through associations with high-fitness genotypes (which are broken down more rapidly when r is high). Because my primary goal in the current study is to determine the strength of selection on the hotspot itself that is required for it to spread, and because there is reason to think that there may be direct selection on hotspot activity due to its importance in proper chromosome segregation, these effects are beyond the scope of the current model.

Long-term fate of hotspots: It is important to note that this modifier analysis predicts only the short-term

behavior of the system: the ability of a modifier to invade does not necessarily imply that it will go to fixation. In fact, even if the active modifier does fix (or nearly so), the system then reduces to the one-locus case: a *cis*-controlled hotspot established by the spread of a *trans* modifier will be lost to biased gene conversion except under strong selection, as shown in Table 3A. Thus, the rise of a new hotspot through the spread of a *trans* modifier is at best a temporary escape from the hotspot conversion paradox. But can an active hotspot—that is, active-hot chromosome *HM*—remain at high frequency for long enough to be observed in natural populations?

To address this question, I examined the numerical solutions to the exact differential equations approximated above for a range of parameter values. In particular, I considered the case in which the population is fixed or nearly fixed for the hot allele at the **H** locus ($p_H = 1$ or $p_H = 0.99$) and the active modifier allele (*M*) is introduced at a low frequency ($p_{HM} = 0.001 \times p_H$). This is relevant to situations in which a “potentially hot” sequence has fixed, or nearly so (*e.g.*, due to drift), in a chromosomal region in which it does not act as a hotspot; the modifier then “activates” this hotspot.

Figure 3 shows a representative set of frequency changes for the range of selection strengths under which hotspots are predicted to be unstable in the *cis* case, but a *trans* modifier is expected to spread ($0 < \sigma \leq \frac{1}{2}$). In all cases, active modifier *M* increases in frequency initially, as predicted by the analytical results: because initially all (or nearly all) copies of the active modifier (*M*) allele are found with hot *cis* (*H*) alleles, the active-hot chromosome *HM* increases in frequency, displacing the inactive resident chromosomes. Eventually, however, as mutation produces cold allele *h*, active-cold chromosome *hM* begins to displace active-hot chromosome *HM* due to biased gene conversion (that is, the hotspot conversion paradox takes effect); once hot allele *H* is lost due to gene conversion, the spread of the active modifier *M* stops, often at an intermediate frequency.

The final frequency of active modifier *M*, and more importantly the length of time during which active-hot chromosome *HM* exists at high frequency in the population, is determined largely by the relative strengths of mutation, selection, and biased gene conversion: if selection is strong relative to the rate of mutation, the *M* allele may spread to near-fixation before the *h* allele appears at high enough frequency to cause substantial amounts of biased gene conversion. This allows chromosome *HM* to exist at relatively high frequency for hundreds or even thousands of generations if selection is strong enough ($\sigma > \frac{1}{10}$ and $\beta_H > 10^{-4}$). If selection is weak, however ($\sigma \leq \frac{1}{50}$ and $\beta < 10^{-4}$), chromosome *HM* never reaches high frequencies. In addition, even small deviations from fixation of the *H* allele at time zero (*e.g.*, $p_H = 0.99$; solid lines in Figure 3) can magnify the effects

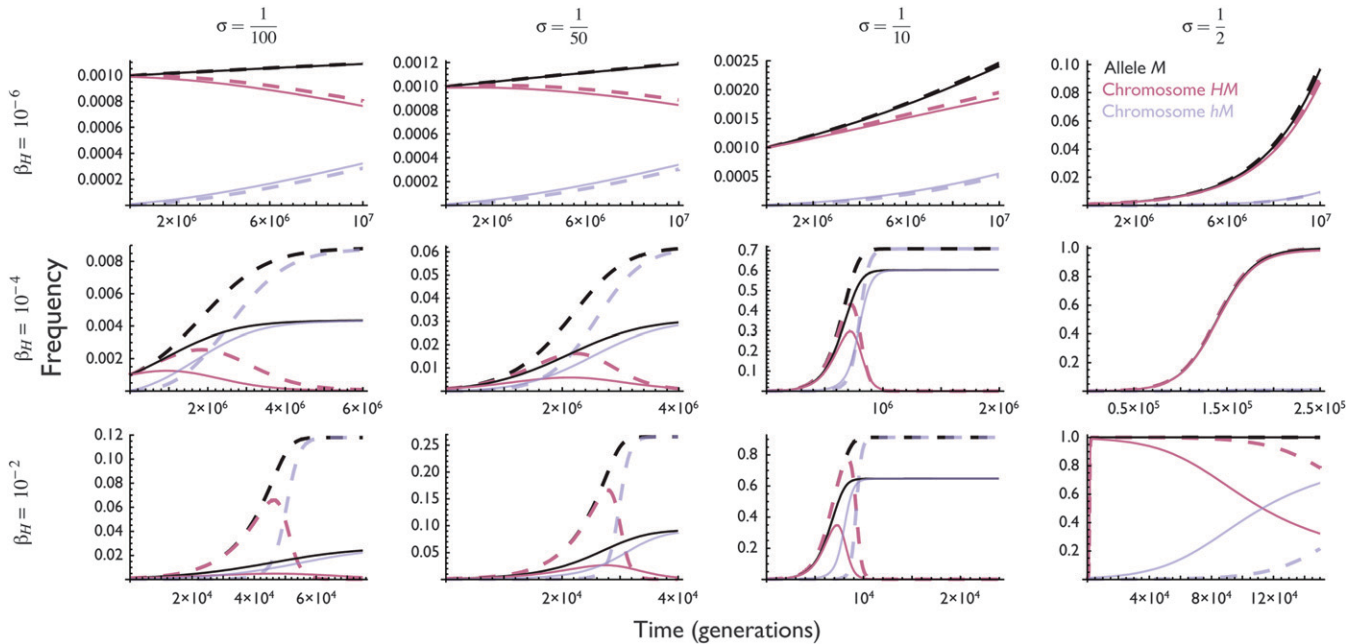


FIGURE 3.—Dynamics of an active *trans* ($\tau = 0$) modifier allele M (black/dark lines) and the two chromosomes carrying it: the “active-hot” chromosome (HM , magenta/medium lines) and the “active-cold” chromosome (hM , blue/light lines) under a variety of selection regimes. Note that the scales of both the x - and the y -axes change from section to section. Exact (nonapproximated) differential equations were solved numerically, starting from fixation (thick dashed lines) or near-fixation ($p_{HM} = 10^{-3}$; $p_{Hm} = 1 - 10^{-3}$) of the H allele and near-fixation of the m allele ($p_{HM} = 10^{-3}$; $p_{Hm} = 1 - 10^{-3}$). Other parameter values are $\mu = 10^{-6}$, $\nu = 10^{-7}$, $\beta_h = 10^{-8}$, $h_M = h_m = 0$, and $r = \frac{1}{2}$; varying these parameters varies the rates of these processes but not the range of outcomes.

of weaker selection, substantially limiting the spread of chromosome HM . As such, this mechanism is likely to be effective only on hotspot loci at which the (inactive) hot allele is truly at fixation.

These patterns of allele and chromosome frequencies can lead to complicated patterns of hotspot activity polymorphism, particularly when selection is strong (Figure 4). Polymorphism in hotspot activity can result from polymorphism at the M locus (*trans* polymorphism) when the frequency of hot allele H is high or from polymorphism at the H locus (*cis* polymorphism) when the frequency of the active modifier, M , is high. Because the active modifier, M , spreads when the frequency of the hot allele, H , is high, after which the cold allele, h , spreads through the population, this can lead to two periods during which there is polymorphism for hotspot activity: a period of *trans* polymorphism followed by a period of *cis* polymorphism. Under most parameter values considered here, these two periods largely overlap, although if selection is strong and the hotspot is very hot (e.g., $\sigma = \frac{1}{2}$, $\beta_H = 10^{-2}$), there can be two very distinct peaks of polymorphism separated by thousands of generations.

As with the analytical results, these patterns depend largely on the relationship between selection (σ) and conversion (τ) and the magnitude of the modifier effect (β_H). The other parameters of the model (cold DSB rate β_h , mutation rates μ and ν , dominance parameters h_M and h_m , recombination rate r) affect the rates of change

and therefore the duration of polymorphism and the maximum frequency of the active-hot chromosome (HM). In addition, the initial frequencies at the H locus have a strong effect on these results: for example, if the resident population is at its mutational equilibrium ($\hat{p}_H = \nu/(\nu + \mu)$), the frequency of cold allele h likely begins at a high enough frequency that biased gene conversion takes effect immediately, preventing active-hot chromosome HM from reaching high frequencies. Thus, this process is most likely to be important when populations are finite, such that there is an appreciable rate of turnover of neutral alleles (in particular, leading to some likelihood of fixation of the H allele when on the inactive-hot chromosome), but large enough that selection on the order of $\beta_H \times (\sigma - \tau/2)$ can act.

DISCUSSION

These results demonstrate the ability of *trans* control, on its own or in combination with *cis* control, to expand the conditions under which recombination hotspots can persist or increase under the influence of purely deterministic forces. When fitness increases rapidly enough with DSB rate ($\sigma \approx 0.1$) and hotspot activity is under a combination of *cis* and *trans* control, these processes can explain both the appearance and the loss of hotspots, even (in fact, particularly) “very-hot” hotspots ($\beta_H \geq 10^{-2}$; see Figure 3). This suggests a process of constant turnover of hotspots, in which individual

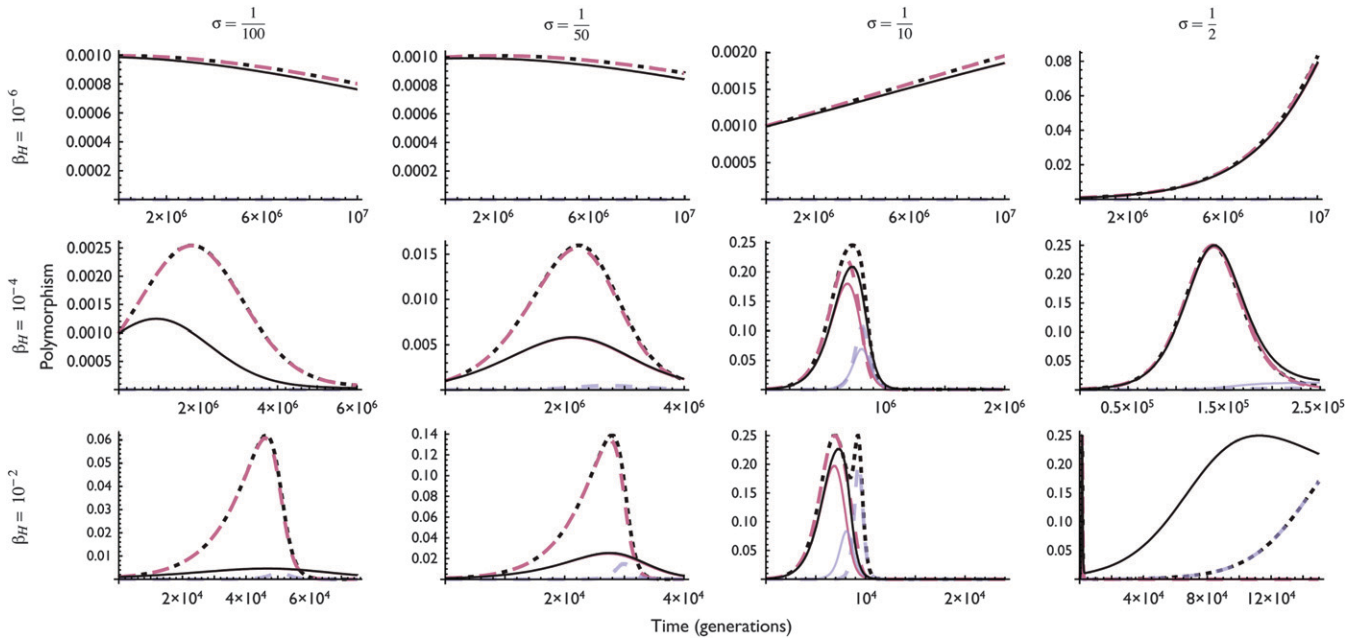


FIGURE 4.—Dynamics of polymorphism over the lifespan of a hotspot under the *cis-plus-trans* modifier model. Panels and parameter values correspond to those in Figure 3. Black/dark line shows total polymorphism for the hotspot phenotype, measured as $p_{HM} \times (1 - p_{HM})$; magenta/medium line shows polymorphism due to variation at the *trans* (modifier) locus, measured as $p_{HM} \times (p_{hm} + p_{hm})$; blue/light line shows polymorphism due to variation at the *cis* (hotspot) locus, measured as $p_{HM} \times p_{hm}$. Note that the scales of both the x - and the y -axes change from section to section.

hotspots are constantly disappearing due to biased gene conversion, and *trans* modifiers spread infrequently, allowing new hotspots to arise. Weaker selection ($0 < \sigma < \sim 0.1$), while it is still able to explain the existence of hotspots controlled purely in *trans* (Table 3A; Figure 2), is unlikely to cause the appearance of new hotspots under both *cis* and *trans* control: although alleles in *trans* that activate such hotspots can spread initially (Table 3B), cold *cis* alleles arise and spread too quickly for active hotspots to reach any appreciable frequency (Figure 3).

The obvious question arising from these results is, “How strong is selection likely to be?” Since there are no direct estimates of the fitness effects of single hotspot alleles, there is no clear answer to this question. However, we may be able to get some indication of whether the results presented here fall within a range that is ever likely to be relevant. To begin with, it may be helpful to make the required σ -values above as concrete as possible. Recall that σ represents the slope of the relationship between fitness increase and DSB rate—that is, it is a selection gradient. Put simply, if fitness differences arise strictly from differences in DSB rate, σ represents the increase in fitness accruing to an individual that does undergo a DSB at locus **X**, relative to one that does not. $\sigma = 0.1$, therefore, implies that the fitness of individuals experiencing a DSB at locus **X** is 1.1 relative to those that do not, $\sigma = 0.5$ implies that this relative fitness is 1.5, etc.

Now we must consider the various mechanisms by which selection might favor a hotspot. In the broadest

and most general sense, there are two: indirect benefits deriving from the breakdown of linkage disequilibrium and direct benefits, largely deriving from the association of DSBs and crossovers with the correct segregation of chromosomes. The degree of effective selection from the former depends strongly on assumptions about how linkage disequilibrium forms among loci flanking the hotspot. As an example, however, one recent model providing a relatively strong advantage to recombination finds that the effective selection on a recombination modifier under reasonable parameter values is $s_e = 0.008$ times the amount by which the modifier increases recombination (BARTON and OTTO 2005). Assuming that a DSB has a 50% chance of resolving as a crossover, this translates to $\sigma = 0.004$. This number decreases as the recombination rate between the modifier and the selected loci (or, in the present model, between the **H** and the **M** loci) increases.

It seems clear that indirect fitness effects of recombination alone are unlikely to cause selection strong enough to allow the evolution of new, *cis-plus-trans*-controlled modifiers as seen above. This leaves us with direct effects. Here, although there are no formal models to guide us, some rough calculations may be enough to determine whether such selection is ever likely to be strong enough to yield these patterns. I begin with the assumption that at least one crossover per chromosome arm is required to yield correct chromosome segregation (HASSOLD and HUNT 2001) and that individuals that segregate correctly have a fitness of one, while those

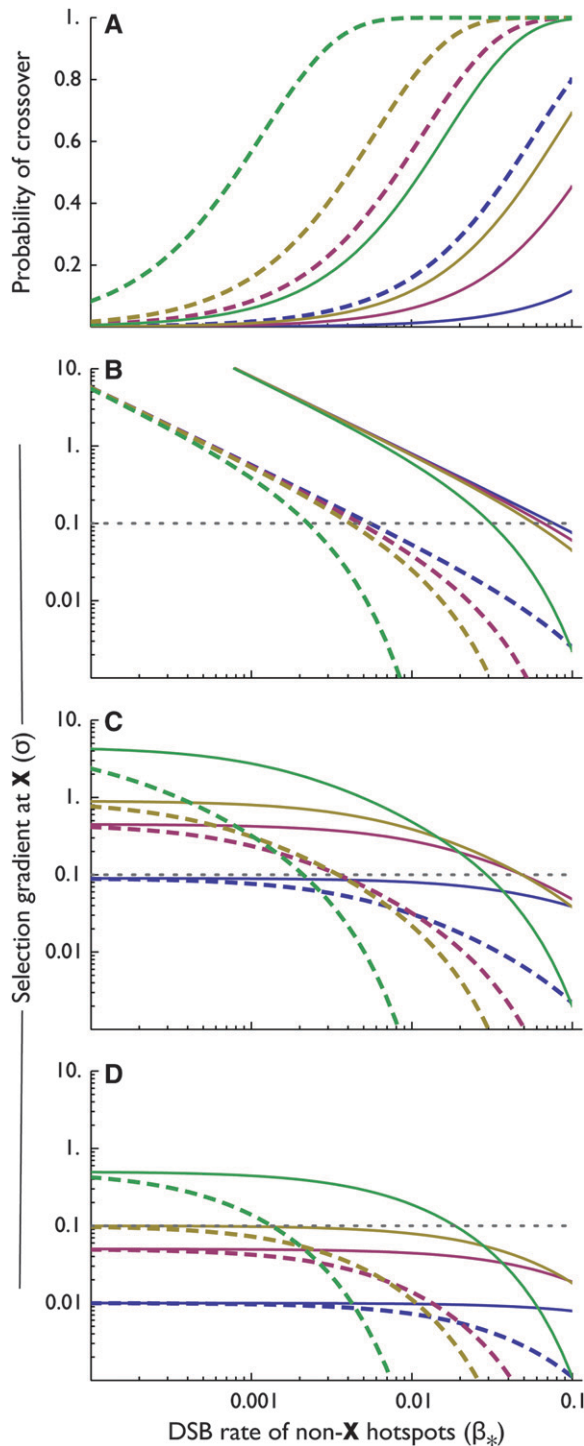


FIGURE 5.—Effects of DSBs other than at locus **X** on the probability of crossover occurrence and on the selection gradient (σ) at locus **X**. (A) Probability that a crossover occurs on the chromosome arm carrying locus **X**, assuming no DSB occurs at locus **X**, as a function of DSB rate at other hotspots on the same chromosome arm. (B–D) Log–log plots of selection gradient (σ) of DSB rate at locus **X** under direct selection for crossovers (*e.g.*, due to a requirement for proper segregation) as a function of DSB rate at other hotspots on the same chromosome arm. B–D correspond to different fitnesses when there is no crossover (w_0): (B) $w_0 = 0$; (C) $w_0 = 0.1$; (D) $w_0 = 0.5$. The gray dotted line at $\sigma = 0.1$ indicates the approximate cutoff for the *cis-plus-trans* model to allow the

that do not have fitness w_0 . The probability that a crossover occurs in the chromosome arm carrying locus **X** is dictated not only by the probability of a DSB at **X**, but also by the number of other hotspots on that chromosome arm, their DSB rates, the probability that a DSB resolves as a crossover, and the background (nonhotspot) crossover rate. The probability that a crossover occurs at a hotspot other than locus **X** is $r_* = 1 - (1 - \beta_*\rho)^{2L}$, where β_* is the per-haplotype-per-locus probability of a DSB at these hotspots (*i.e.*, the probability that a DSB occurs at a given copy of a given hotspot), ρ is the probability that a DSB at a hotspot resolves as a crossover, and L is the number of hotspot loci (not including **X**). Finally, I assume that the probability of a crossover event somewhere other than a hotspot is a fraction b of the probability that a crossover occurs in a hotspot. The selection gradient can now be calculated as $\sigma = w_{\mathbf{X}}/w_* - 1$, where

$$w_{\mathbf{X}} = (1 - (1 - \rho)(1 - r_*)(1 - br_*)) + w_0(1 - \rho)(1 - r_*)(1 - br_*)$$

is the fitness of individuals experiencing a DSB at locus **X**, and

$$w_* = (1 - (1 - r_*)(1 - br_*)) + w_0(1 - r_*)(1 - br_*)$$

is the fitness of individuals with no DSB at locus **X**.

To place values on these numbers, we can turn to recent data from humans. In particular, recent studies suggest that there are 25,000–33,000 hotspots in the human genome (MYERS *et al.* 2005; INTERNATIONAL HAPMAP CONSORTIUM 2007), translating to several hundred per chromosome arm ($L \leq \sim 700$), and that 60–80% of recombination occurs in hotspots (and therefore $\sim \frac{1}{4} \leq b \leq \sim \frac{2}{3}$). Although the DSB rates (β_*) of all of these hotspots are not clear, well-characterized hotspots in humans range from $\beta_* \approx 5 \times 10^{-6}$ to $\beta_* \approx 5 \times 10^{-3}$ (COOP and MYERS 2007). Finally, the probability that a DSB resolves as a crossover may be as low as $\rho \approx 0.1$ (JEFFREYS and MAY 2004); the maximum possible probability is $\rho = \frac{1}{2}$. Figure 5 shows the value of σ over parameter values in these ranges. When the fitness of individuals with no crossover (w_0) is zero, as the number of hotspots, or their heat (measured either as DSB rate, β_* , or the probability that a DSB resolves as a crossover, ρ) decreases, the probability of correct segregation without a DSB at **X** decreases as well; this causes the strength of selection on DSB rate at locus **X** (σ) to

spread of new hotspots to appreciable frequencies starting from fixation of the H allele. In all sections, line dashing corresponds to number of other hotspots (L): dashed, 50; solid, 700. Color/shading corresponds to the probability that a DSB resolves as a crossover (ρ): blue/dark lines, 0.01; magenta/semidark lines, 0.05; gold/semilight lines, 0.1; green/light lines, 0.5. In all cases, the ratio of the frequency of crossovers outside of hotspots to that inside them (b) is $2/3$.

increase. The primary effect of increasing the fitness of individuals with no crossover is to flatten out the relationship between non-X DSB rate (β_*) and the selection gradient σ at low values of β_* ; this effect is stronger when the probability that a DSB resolves as a crossover (ρ) is low (hence the crossing of the lines when $w_0 > 0$). Perhaps surprisingly, σ falls within the range under which rapid turnover of *cis-plus-trans*-controlled hotspots is expected under deterministic processes alone ($\sigma \geq 0.1$, assuming allele *H* is fixed prior to the introduction of the modifier) over a variety of reasonable parameter values ($\beta_* \leq \sim 10^{-2}$ with 500 hotspots/chromosome arm; $\beta_* \leq 10^{-1}$ with 50 hotspots/chromosome arm).

These rough calculations point out an important pattern: as the heat (β_*) and/or number (L) of hotspots on a chromosomal arm decreases due to biased gene conversion, conditions become increasingly favorable for the spread of a new hotspot under any control scenario (*cis*, *trans*, or *cis-plus-trans*). In fact, these parameters (number and heat of hotspots) may not have to decrease particularly far, if at all, from their current levels for these conditions to be achieved.

In addition, the *cis-plus-trans* model provides another avenue by which the effective strength of selection might be increased: a *trans* modifier is likely to affect more than one hotspot locus. For example, such a modifier might affect an entire chromosomal region or globally change the specificity of a protein involved in the DSB or histone modification machinery. Selection on a *trans* modifier might therefore be stronger than that on any individual hotspot, allowing the modifier to spread to higher frequencies before individual hotspots begin to be lost to gene conversion and decreasing the per-hotspot selection gradient (σ) required for new hotspots to arise. This may be particularly important in cases where the *H* allele at one or more hotspots is not initially fixed. A full analysis of this process requires an explicit model of multiple hotspots, which is beyond the scope of this article.

Overall, it appears that a new, active hotspot under a combination of *cis* and *trans* control can spread to high frequencies under the deterministic forces of drift and gene conversion under reasonable parameter values, particularly as old hotspots become colder and are ultimately lost to biased gene conversion. In circumstances under which selection is weaker, of course, there is the potential for drift to contribute to the appearance of new hotspots; however, rather than requiring the fixation of effectively deleterious mutations (COOP and MYERS 2007), the *cis-plus-trans* model allows the hotspot initially to be a mildly beneficial mutation, which may occasionally spread faster than expected due to drift. That is, a *trans* modifier that might spread only to low frequencies under purely deterministic processes may occasionally spread to higher frequencies under drift, yielding a high frequency of the active-hot modifier be-

fore biased gene conversion causes the hot allele to be lost. Again, an explicitly stochastic model of the *cis-plus-trans* case will be required to analyze this possibility fully.

Interestingly, the predictions of the *cis-plus-trans* model converge with a pair of recent observations regarding hotspot polymorphisms in mammals. One clear requirement of this model is that hotspots be under some combination of *cis* and *trans* control. Recent data from mice suggest just that: activity at recombination hotspot *Psm9* appears to be under control of a *trans*-acting enhancer that increases crossover rate ~ 2000 -fold, specifically at the *Psm9* locus. In addition, variation at the hotspot locus itself (in *cis*) can change the crossover rate by a factor of 10 (BAUDAT and DE MASSY 2007).

A more subtle prediction of the *cis-plus-trans* model involves the relationship between polymorphism and hotspot age. If an active hotspot originally arises through the spread of a *trans* modifier and then is lost to biased gene conversion in *cis*, two (possibly discrete) periods of polymorphism arise: initially, as the active *trans* modifier spreads, there is a period of polymorphism due to the fact that the modifier has not fixed; later, as the hot *cis* allele is lost to biased gene conversion, a second period of polymorphism occurs due to the *cis* allele being at intermediate frequencies (Figure 4). In other words, relatively young hotspots are more likely to show polymorphisms due to variation in *trans*, while older hotspots are more likely to show polymorphisms due to variation in *cis* (and possibly also in *trans*, depending on the maximum frequency reached by the modifier).

In a tantalizing recent analysis of two closely spaced hotspots in humans (NEUMANN and JEFFREYS 2006)—a putatively young hotspot (MSTM1a) and an older hotspot (MSTM1b)—both show evidence of *trans* control (more precisely, non-*cis* control, which could also include purely epigenetic control), in that there is significant variation in crossover frequency among individuals with identical DNA sequences in a large interval around the hotspot. In addition, for neither hotspot is there a significant relationship between *cis* sequence and hotspot activity. However, this evidence appears stronger for the young hotspot, which shows a strictly “on-off” polymorphism in which individuals with active hotspots have no unique sequence difference that distinguishes them from those with inactive hotspots. The older hotspot shows much more continuous variation in hotspot activity, and only individuals classified as having “high” and “low” activity were examined for diagnostic sequence differences; this would appear to leave open the possibility that some variation in hotspot activity is due to DNA sequence variation in *cis*. It is possible, therefore, that the younger hotspot shows a polymorphism in *trans* while the older hotspot shows polymorphisms both in *cis* and in *trans*. It is also possible that these two hotspots are under joint control by the same *trans* element [the authors discount this possibility, but

this is based on a marginally nonsignificant ($P = 0.08$) rank correlation on 26 individuals; it is therefore possible that this conclusion was based on a statistical test with marginal power]. This admittedly optimistic interpretation is precisely the prediction of a *cis-plus-trans* model in which the *trans*- and *cis*-polymorphism phases overlap. Further experiments explicitly comparing the nature of polymorphisms in young *vs.* old hotspots are likely to be a fruitful (if challenging) direction for exploring the realism of these models.

Another consequence of the prediction that older hotspots are more likely to show *cis* polymorphisms is that the majority of identified hotspots may provide a nonrandom sample for testing the mechanisms of hotspot control. Most hotspots are identified due to the blocks of linkage disequilibrium they cause over thousands of generations; however, new hotspots have not had the time to cause such blocks to form and are therefore unlikely to be detected by even the most sensitive linkage disequilibrium mapping techniques; only relatively intensive approaches such as sperm typing are likely to find these newer hotspots (JEFFREYS *et al.* 2005). Any failure to find *trans* polymorphisms for hotspots identified by linkage-disequilibrium mapping cannot necessarily be taken as evidence that *trans* control has not been important to the evolution of these hotspots: the *cis-plus-trans* model specifically predicts that *cis* polymorphisms will become more prominent in older hotspots.

These models, particularly the *cis-plus-trans* control model, demonstrate that there is scope for the rise of new recombination hotspots under purely deterministic forces, given reasonable parameter values. Unlike recent stochastic models (CALABRESE 2007; COOP and MYERS 2007), it allows for the fixation of very-hot hotspots, assuming selection is strong enough (in fact, very-hot modifiers are more likely to fix under a given selection strength than less-hot modifiers; see Figure 3). Of course, this model requires several simplifying assumptions. As mentioned above, it models the benefits to hotspots as direct fitness effects rather than the indirect effects traditional in recombination modifier models; however, given that the effective selection coefficients in such models are so weak relative to the force of conversion (BARTON 1995; BARTON and OTTO 2005), ignoring these effects is probably appropriate to a first approximation. Also as discussed above, the explicit consideration of multiple hotspot loci under the control of a single modifier seems likely to expand the conditions under which the *cis-plus-trans* model can bring a new hotspot to high frequencies; this is therefore likely to be a useful direction for future modeling efforts. Another dimension that may be of importance is the possibility that there are more than two levels of activity possible at a hotspot. As one example, if there are more than two alleles in *cis*, each with its own level of heat, then the secondary hotspot decline seen as

biased gene conversion removes the hot allele in *cis* may not lead, at least immediately, to a complete reversion of a hotspot to its historical (premodifier) levels of recombination.

Finally, of course, the importance of stochastic processes cannot be denied: full understanding of the expected distribution of outcomes of any evolutionary process requires a formal inclusion of genetic drift. However, it is important to have a full understanding of the underlying deterministic processes to build a complete stochastic model. In this article, I have provided the first analytical approach to the deterministic processes behind the evolution of recombination hotspots and have shown that the inclusion of *trans* control can substantially broaden the circumstances under which hotspots are expected to be present in populations. Further, I have shown that *cis-plus-trans* control can explain the turnover of hotspots observed at the species level, even without the inclusion of stochastic processes. Further expansions of this model seem likely to expand these conditions.

I thank B. Payseur for invaluable discussions leading up to the preparation of this model and helpful comments throughout; B. Dumont, A. R. Ives, and R. T. Gilman for discussion; and S. P. Otto, Aneil Agrawal, J. Crow, R. Redfield, and an anonymous reviewer for comments on earlier versions of this manuscript. This work was supported by funds from the Department of Zoology and the Graduate School at the University of Wisconsin, Madison.

LITERATURE CITED

- BARTON, N. H., 1995 A general model for the evolution of recombination. *Genet. Res.* **65**: 123–144.
- BARTON, N. H., and S. P. OTTO, 2005 Evolution of recombination due to random drift. *Genetics* **169**: 2353–2370.
- BAUDAT, F., and B. DE MASSY, 2007 *Cis*- and *trans*-acting elements regulate the mouse *Psmb9* meiotic recombination hotspot. *PLoS Genet.* **3**: 1–11.
- BOULTON, A., R. S. MYERS and R. J. REDFIELD, 1997 The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl. Acad. Sci. USA* **94**: 8058–8063.
- CALABRESE, P., 2007 A population genetics model with recombination hotspots that are heterogeneous across the population. *Proc. Natl. Acad. Sci. USA* **104**: 4748–4752.
- CATCHESIDE, D. G., 1975 Occurrence in wild strains of *Neurospora crassa* of genes controlling genetic recombination. *Aust. J. Biol. Sci.* **28**: 213–225.
- CONRAD, D. F., M. JAKOBSSON, G. COOP, X. WEN, J. D. WALL *et al.*, 2006 A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat. Genet.* **38**: 1251–1260.
- COOP, G., and S. MYERS, 2007 Live hot, die young: transmission distortion in recombination hotspots. *PLoS Genet.* **3**: 377–386.
- CRAWFORD, D. C., T. BHANGALE, N. LI, G. HELLENTHAL, M. J. RIEDER *et al.*, 2004 Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* **36**: 700–706.
- CROMIE, G. A., C. A. RUBIO, R. W. HYPPA and G. R. SMITH, 2005 A natural meiotic DNA break site in *Schizosaccharomyces pombe* is a hotspot of gene conversion, highly associated with crossing over. *Genetics* **169**: 595–605.
- DOONIE, H. K., and I. M. MARTINEZ-FEREZ, 1997 Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. *Plant Cell* **9**(9): 1633–1646.
- DROUAUD, J., C. CAMILLERI, P.-Y. BOURGUIGNON, A. CANAGUIER, A. BERARD *et al.*, 2006 Variation in crossing-over rates across chro-

- mosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination "hot spots". *Genome Res.* **16**(1): 106–114.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Pearson, Harlow, UK.
- GRIMM, C., P. SCHAER, P. MUNZ and J. KOHLI, 1991 The strong *ADH1* promoter stimulates mitotic and meiotic recombination at the *ADE6* gene of *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **11**: 289–298.
- GUILLOIN, H., and B. DE MASSY, 2002 An initiation site for meiotic crossing-over and gene conversion in the mouse. *Nat. Genet.* **32**: 296–299.
- HASSOLD, T., and P. HUNT, 2001 To ert (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.* **2**: 280–291.
- HEY, J., 2004 What's so hot about recombination hotspots? *PLoS Biol.* **2**: 730–733.
- INTERNATIONAL HAPMAP CONSORTIUM, 2007 A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**: 851–862.
- JEFFREYS, A. J., and C. A. MAY, 2004 Intense and highly localized gene conversion activity in human meiotic crossover hotspots. *Nat. Genet.* **36**: 151–156.
- JEFFREYS, A. J., and R. NEUMANN, 2002 Reciprocal crossover asymmetry and meiotic drive in a human recombination hotspot. *Nat. Genet.* **31**: 267–271.
- JEFFREYS, A. J., and R. NEUMANN, 2005 Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.* **14**: 2277–2287.
- JEFFREYS, A. J., J. MURRAY and R. NEUMANN, 1998 High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell* **2**: 267–273.
- JEFFREYS, A. J., A. RITCHIE and R. NEUMANN, 2000 High resolution analysis of haplotype diversity and meiotic crossover in the human *TAP2* recombination hotspot. *Hum. Mol. Genet.* **5**: 725–733.
- JEFFREYS, A. J., L. KAUPPI and R. NEUMANN, 2001 Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* **29**: 217–222.
- JEFFREYS, A. J., R. NEUMANN, M. PANAYI, S. MYERS and P. DONNELLY, 2005 Human recombination hot spots hidden in regions of strong marker association. *Nat. Genet.* **37**: 601–606.
- KELMENSEN, P. M., P. PETKOV, X. WANG, D. C. HIGGINS, B. J. PAIGEN *et al.*, 2005 A torrid zone on mouse chromosome 1 containing a cluster of recombinational hotspots. *Genetics* **169**: 833–841.
- MALONE, R. E., S. KIM, S. A. BULLARD, S. LUNDQUIST, L. HUTCHINGS-CROW *et al.*, 1994 Analysis of a recombination hotspot for gene conversion occurring at the *HIS2* gene of *Saccharomyces cerevisiae*. *Genetics* **137**: 5–18.
- MCVEAN, G. A. T., S. R. MYERS, S. HUNT, P. DELOUKAS, D. R. BENTLY *et al.*, 2004 The fine-scale structure of recombination rate variation in the human genome. *Science* **304**: 581–584.
- MYERS, S., L. BOTTOLO, C. FREEMAN, G. MCVEAN and P. DONNELLY, 2005 A fine-scale map of recombination hotspots across the human genome. *Science* **310**: 321–324.
- NEUMANN, R., and A. J. JEFFREYS, 2006 Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. *Hum. Mol. Genet.* **15**: 1401–1411.
- NICOLAS, A., D. TRECO, N. P. SCHULTES and J. W. SZOSTAK, 1989 An initiation site for meiotic gene conversion in the yeast *Saccharomyces cerevisiae*. *Nature* **338**: 35–39.
- OKAGAKI, R. J., and C. F. WEIL, 1997 Analysis of recombination sites within the maize waxy locus. *Genetics* **147**(2): 815–821.
- OTTO, S. P., and T. DAY, 2007 *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press, Princeton, NJ.
- PETES, T. D., 2001 Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* **2**: 360–369.
- PINEDA-KRCH, M., and R. J. REDFIELD, 2005 Persistence and loss of meiotic recombination hotspots. *Genetics* **169**: 2319–2333.
- PTAK, S. E., A. D. ROEDER, M. STEPHENS, Y. GILAD, S. PÄÄBO *et al.*, 2004 Absence of the *TAP2* human recombination hotspot in chimpanzees. *PLoS Biol.* **2**: 849–855.
- PTAK, S. E., D. A. HINDS, K. KOEHLER, B. NICKEL, N. PATIL *et al.*, 2005 Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.* **37**: 429–434.
- SHIFMAN, S., J. T. BELL, R. R. COPLEY, M. S. TAYLOR, R. W. WILLIAMS *et al.*, 2006 A high-resolution single nucleotide polymorphism genetic map of the mouse chromosome. *PLoS Biol.* **4**: 2227–2237.
- SZOSTAK, J. W., T. L. ORR-WEAVER and R. J. ROTHSTEIN, 1983 The double-strand-break repair model for recombination. *Cell* **33**: 25–35.
- WINCKLER, W., S. R. MYERS, D. J. RICHTER, R. C. ONOFRIO, G. J. McDONALD *et al.*, 2005 Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* **308**: 107–111.
- WOLFRAM RESEARCH, 2007 *Mathematica, Version 6.0*. Wolfram Research, Champaign, IL.
- WU, T.-C., and M. LICHTEN, 1995 Factors that affect the location and frequency of meiosis-induced double-strand breaks in *Saccharomyces cerevisiae*. *Genetics* **140**: 55–66.
- YAO, H., Q. ZHOU, J. LI, H. SMITH, M. YANDEAU *et al.*, 2002 Molecular characterization of meiotic recombination across the 140-kb multigenic *a1-sh2* interval of maize. *Proc. Natl. Acad. Sci. USA* **99**(9): 6157–6162.

Communicating editor: M. S. McPEEK