Genet. Res., Camb. (1988), **51**, *pp.* 55–58 *With* 1 *text-figure Printed in Great Britain*

Unusual behaviour of linkage disequilibrium in two-locus gene conversion models

J. BRUCE WALSH
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

(*Received 30 June 1987 and in revised form 2 September 1987*)

Summary

The amount of linkage of disequilibrium maintained in a two-locus infinite population model by gene conversion and recombination is examined. Intrachromosomal conversion (conversion between different loci on the same chromosome) generates positive linkage disequilibrium. Specifically, $\hat{D} = p(1-p)[1-r/(\gamma+r-\gamma r)]$, where p is the frequency of allele A at both loci, r is the recombination rate between loci and γ is the per-gamete conversion rate. Somewhat unexpectedly, interchromosomal conversion (conversion between loci on different chromosomes) also generates positive disequilibrium, albeit very small. More interestingly, the behaviour of this disequilibrium as a function of recombination is unusual. If β is the interchromosomal conversion rate between a pair of loci, then $\hat{D} = p(1-p)[r\beta/(\beta+r-\beta r)]$. $\hat{D}$ *increases* with increasing recombination, being zero for the case of complete linkage ($r = 0$), and maximized at $r = 1/2$. This unusual behaviour can be accounted for by the generation of excess coupling gametes when an interchromosomal conversion event is followed by recombination.

1. Introduction

The amount of disequilibrium between members of a multigene family generated by gene conversion has received increased attention recently. For example, Slatkin (1986) has pointed out the connection between the amount of linkage disequilibrium and cohesiveness assumptions required for 'molecular drive' (Dover, 1982). The average amount of pairwise linkage disequilibrium ($\bar{D}$) can be extracted from gene family models based on identity coefficients (Slatkin, 1986). If C_1 is the identity coefficient for different loci on the same chromosome and C_2 is the identity coefficient for nonhomologous loci on different chromosomes, $\bar{D} = (C_1 - C_2)/2$. C_1 and C_2 correspond to Ohta's (1983) notation, with $C_1 = g$, $C_2 = l$ in Nagylaki's (1984*a*, *b*) notation. These models compound the effects of finite population size and mutation on the amount of disequilibrium generated solely by gene conversion balanced by recombination. Here we examine a simple deterministic two-locus selectively neutral model to obtain the amount of disequilibrium maintained by the balance between conversion and recombination. Interestingly, interchromosomal conversion generates a very small amount of positive linkage disequilibrium. Surprisingly, the amount of disequilibrium increases as recombination increases.

2. Analysis of Models

(i) *General structure*

Consider two loci which exchange alleles through gene conversion. Conversion can be either intrachromosomal (between loci on the same chromosome) or interchromosomal (conversions between maternally and paternally derived homologous chromosomes); see Fig. 1. We examine the consequences of these different forms of conversion separately.

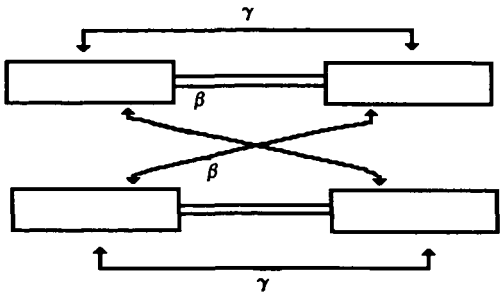

Fig. 1. Intra- and interchromosomal conversion. Intrachromosomal conversion occurs independently on each chromosome with rate γ per chromosome per generation. Interchromosomal conversion occurs independently between each set of nonhomologous loci with rate β per pair per generation.

The basic structure of our models is as follows. An infinite population with discrete generations is assumed. Both loci share the same two alleles, A and a. Using standard two-locus notation, we follow the gamete frequencies x_i and the amount of linkage disequilibrium, D. Specifically, $x_1 = \text{freq}(AA)$, $x_2 = \text{freq}(Aa)$, $x_3 = \text{freq}(aA)$, $x_4 = \text{freq}(aa)$, $D = x_1x_4 - x_2x_3$. The life cycle is:

$$\underset{x_i,\,D}{\text{Adults}} \xrightarrow{\text{Conversion}} \underset{x_i',\,D'}{\text{Adults}} \xrightarrow{\text{Recombination}} \underset{x_i'',\,D''}{\text{Gametes}}$$

$$\xrightarrow{\text{Random union of gametes}} \underset{x_i'',\,D''}{\text{Zygotes.}}$$

Conversion and recombination are assumed to be independent (see Fink & Petes, 1984), and there is no conversion bias. If bias occurs, the equilibrium is trivial, the population being fixed for the allele favoured by bias. Conversion is assumed to occur via a double-strand gap repair model (Szostak *et al.* 1983), so that at most only a single allele is altered per conversion event. This ensures that two alleles do not switch places, for example an Aa gamete cannot be transformed by a single conversion event into an aA gamete.

(ii) *Intrachromosomal conversion*

Intrachromosomal conversion occurs at rate γ per chromosome, and conversion occurs independently on each chromosome (see Fig. 1). Under these assumptions, conversion does not alter an AA chromosome, while 1/2 of Aa and aA chromosomes undergoing conversion become AA chromosomes. Proceeding in a similar fashion for the other gamete classes we obtain

$$x_1' = x_1 + (\gamma/2)(x_2 + x_3), \tag{1a}$$

$$x_2' = (1-\gamma)x_2, \tag{1b}$$

$$x_3' = (1-\gamma)x_3, \tag{1c}$$

$$x_4' = x_4 + (\gamma/2)(x_2 + x_3). \tag{1d}$$

The frequency of the genotype x_i/x_j after conversion is simply $2x_i'x_j'$ for $i \neq j$, and $(x_i')^2$ for $i = j$. This can be seen either by noting that the conversion process acts independently on each gamete, or by writing out all genotypes after conversion and verifying this directly. It follows that gamete frequencies after recombination are

$$x_i'' = x_i' + \eta_i r(x_1' \times x_4' - x_2' \times 3') \quad \text{for} \quad i = 1,2,3,4, \tag{2}$$

$$\eta_i = -1 \quad \text{for} \quad i = 1,4; \quad \eta_i = 1 \quad \text{for} \quad i = 2,3.$$

Allele frequencies at a single locus are not generally conserved. For example, freq (A at locus 1) $= x_1 + x_2$ changes over time. However, the total frequency of an allele in the system (i.e. summed over both loci) is conserved. Define

$$p = x_1 + (1/2)(x_2 + x_3). \tag{3a}$$

From (1) and (2),

$$p'' = p. \tag{3b}$$

From (1b), (1c) and (2),

$$x_2(t+1) - x_3(t+1) = (1-\gamma)(x_2(t) - x_3(t)), \tag{4a}$$

yielding

$$x_2(t) - x_3(t) = (1-\gamma)^t(x_2(0) - x_3(0)). \tag{4b}$$

Since p remains constant, it follows from (3) that

$$x_1(t) = x_4(t) + 2p - 1. \tag{5}$$

Denote the equilibrium value of the gamete frequencies by $\hat{x}_i$ and likewise denote the equilibrium for D by $\hat{D}$. From (4b) and (5),

$$\hat{x}_2 = \hat{x}_3, \tag{6a}$$

$$\hat{x}_4 = \hat{x}_1 - 2p + 1. \tag{6b}$$

From the definition of D, p and (6),

$$\hat{D} = p(1-p) - \hat{x}_2. \tag{7}$$

At equilibrium, from (1b), (2) and (6) we have

$$\hat{x}_2 = (1-\gamma)\hat{x}_2 + r[\hat{D} + \gamma\hat{x}_2(\hat{x}_1 + \hat{x}_4) + 2\gamma(\hat{x}_2)^2]. \tag{8a}$$

Using $\hat{x}_1 + \hat{x}_4 = 1 - 2\hat{x}_2$, (8$a$) rearranges to give

$$\hat{x}_2 = r\hat{D}/\gamma(1-r) \quad \text{or} \quad \hat{D} = \hat{x}_2\gamma(1-r)/r. \tag{8b}$$

Combining (7) and (8b) and recalling (6) gives

$$\hat{x}_1 = p[1 - (1-p)r/(r + \gamma - \gamma r)], \tag{9a}$$

$$\hat{x}_2 = \hat{x}_3 = p[1-p][r/(r + \gamma - \gamma r)], \tag{9b}$$

$$\hat{x}_4 = [1-p][1 - pr/(r + \gamma - \gamma r)], \tag{9c}$$

$$\hat{D} = p[1-p][1 - r/(r + \gamma - \gamma r)]. \tag{9d}$$

Observe from (9d) that the standardized linkage disequilibrium, $\hat{D}/(p[1-p])$, is independent of the initial allelic frequencies. Finally, we generally expect that conversion is rarer than recombination, giving

$$\hat{D} \cong p[1-p]\gamma/r \quad \text{for} \quad r \gg \gamma. \tag{9e}$$

(iii) *Interchromosomal conversion*

Interchromosomal conversion occurs independently between each alternate pair of loci at rate β per generation (see Fig. 1). We ignore multiple conversion events (i.e. terms of order β^2). Consider the genotype AA/Aa. After conversion, $(1-\beta)$ of the genotypes are AA/Aa, $\beta/2$ of the genotypes are AA/AA, and $\beta/2$ are aA/Aa. Recombination acting on aA/Aa produces all four classes of gametes. Putting these together gives the resulting gamete frequencies as $(1/2)[1 + r\beta/2]AA$, $(1/2)[1 - \beta(1+r)/2]\ Aa$, $(\beta/4)(1-r)aA$ and $\beta r/4\ aa$.

https://doi.org/10.1017/S0016672300023946 Published online by Cambridge University Press

Table 1. *Expected gamete frequencies under interchromosomal conversion and recombination*

Genotype frequency	Gamete types produced			
	AA	*Aa*	*aA*	*aa*
x_1^2†	1	0	0	0
$2x_1x_2$	$(1/2)[1+r\beta/2]$	$(1/2)[1-\beta(1+r)/2]$	$(\beta/4)(1-r)$	$\beta r/4$
$2x_1x_3$	$(1/2)[1+r\beta/2]$	$(\beta/4)(1-r)$	$(1/2)[1-\beta(1+r)/2]$	$\beta r/4$
$2x_1x_4$	$(1/2)[1-\beta-r+2\beta r]$	$(1/2)[\beta+r-2\beta r]$	$(1/2)[\beta+r-2\beta r]$	$(1/2)[1-\beta-r+2\beta r]$
x_2^2	$\beta/2$	$(1-\beta)$	0	$\beta/2$
$2x_2x_3$	$r/2$	$(1/2)(1-r)$	$(1/2)(1-r)$	$r/2$
$2x_2x_4$	$\beta r/4$	$(1/2)[1-\beta(1+r)/2]$	$(\beta/4)(1-r)$	$(1/2)[1+\beta r/2]$
x_3^2	$\beta/2$	0	$1-\beta$	$\beta/2$
$2x_3x_4$	$\beta r/4$	$(\beta/4)(1-r)$	$(1/2)[1-\beta(1+r)/2]$	$(1/2)[1+\beta r/2]$
x_4^2	0	0	0	1

† x_1^2 is the frequency of AA/AA, $2x_1x_2$ is the frequency of AA/Aa, etc.

Table 1 gives the gametic outcomes for the other nine genotypes, which after a few simplifications yield

$$x_1'' = x_1-(\beta+r-2\beta r)D+(\beta/2)(1-r)(x_2-x_3)^2 + (r\beta/2)(x_2+x_3) \quad (10a)$$

$$x_2'' = x_2+(\beta+r-2\beta r)D-(\beta/2)(1-r)(x_2-x_3)^2 - (r\beta/2)(x_2+x_3)+(\beta/2)(x_3-x_2), \quad (10b)$$

$$x_3'' = x_3+(\beta+r-2\beta r)D-(\beta/2)(1-r)(x_2-x_3)^2 - (r\beta/2)(x_2+x_3)-(\beta/2)(x_3-x_2), \quad (10c)$$

$$x_4'' = x_4-(\beta+r-2\beta r)D+(\beta/2)(1-r)(x_2-x_3)^2 + (r\beta/2)(x_2+x_3). \quad (10d)$$

Observe that $p'' = x_1''+(1/2)(x_2''+x_3'') = x_1+(1/2)(x_2+x_3) = p$, thus like intrachromosomal conversion, allele frequencies at single loci are not conserved, but the total frequency of an allele in the system is. It immediately follows that

$$x_1(t) = x_4(t)+2p-1 \quad (11)$$

From (10*b*) and (10*c*),

$$x_2(t+1)-x_3(t+1) = (1-\beta)(x_2(t)-x_3(t)) \quad (12a)$$

or

$$x_2(t)-x_3(t) = (1-\beta)^t(x_2(0)-x_3(0)). \quad (12b)$$

We therefore have $\hat{x}_2 = \hat{x}_3$ and $\hat{x}_4 = \hat{x}_1-2p+1$, and relationships (6) and (7) above hold. From these, (10*a*) reduces at equilibrium to

$$r\beta\hat{x}_2 = (\beta+r-2\beta r)\hat{D}. \quad (13)$$

Using (13) along with (6) and (7) yields

$$\hat{x}_1 = p[p+(1-p)r\beta/(r+\beta-\beta r)], \quad (14a)$$

$$\hat{x}_2 = \hat{x}_3 = p[1-p][1-r\beta/(r+\beta-\beta r)], \quad (14b)$$

$$\hat{x}_4 = [1-p][(1-p)+pr\beta/(r+\beta-\beta r)], \quad (14c)$$

$$\hat{D} = p[1-p][r\beta/(\beta+r-\beta r)]. \quad (14d)$$

As for intrachromosomal conversion, the standardized linkage disequilibrium is independent of allele frequencies. Generally, the amount of positive disequilibrium maintained under interchromosomal conversion/recombination balance is very small. If conversions are much rarer than recombination,

$$\hat{D} \cong p(1-p)\beta \quad \text{for} \quad r \gg \beta. \quad (14e)$$

Typically β values are very small, approaching 10^{-2} as an upper limit. Thus the amount of disequilibrium generated would be very difficult to detect.

At equilibrium D has unusual behaviour under interchromosomal conversion. If there is no recombination, $\hat{D} = 0$, while $\hat{D}$ is maximized at $r = 1/2$. Likewise, if $\beta = 0$, $\hat{D} = 0$, as expected. To see the reason for this unusual behaviour, consider the genotype Ax/yz, where x, y and z are unspecified alleles. An interchromosomal conversion event changes the genotype to Ax/yA, and if recombination follows, an AA coupling gamete is produced. Notice that two events are required to generate the coupling gamete: interchromosomal conversion followed by recombination. If no recombination occurs, no such gametes are generated, and for fixed β their formation is maximized by maximizing recombination.

3. Discussion and Conclusions

For both intra- and interchromosomal conversion, starting allele frequencies at the individual loci are not conserved, but p, the system frequency of an allele, is. Both intra- and interchromosomal conversion generate positive D at equilibrium. Denoting by $\hat{D}^*$ and $\hat{D}^{**}$ the equilibrium D under intra- and interchromosomal conversion respectively, from (9*e*) and (14*e*), given the same p value and provided that $r \gg \beta$, γ, we have: $\hat{D}^{**}/\hat{D}^* = r(\beta/\gamma)$. If inter- and intrachromosomal conversion occur at the same rate, then $\hat{D}^{**}/D^* = r$. For loose linkage, the amount of disequilibrium generated by both types of conversion is similar and near zero, while for tight linkage intrachromosomal conversion generates much higher linkage disequilibrium.

The finding of positive D at equilibrium under interchromosomal conversion is interesting given other reports that $\hat{D} = 0$ under interchromosomal conversion. Slatkin (1986) showed for a deterministic

https://doi.org/10.1017/S0016672300023946 Published online by Cambridge University Press

model that $\hat{D} = 0$ when $r = 0$, and made the apparently logical assumption that $\hat{D} = 0$ for arbitrary r. Given the very small value of $\hat{D}$, using $\hat{D} = 0$ is not likely to produce a large error. Finite population models of interchromosomal conversion using identity coefficients give $l = g$ and thus $\bar{D} = (g-l)/2 = 0$ (Nagylaki, 1984*b*). However, Nagylaki explicitly states that he assumes that rates of both recombination and conversion are small, and hence ignores terms of order $r\beta$. A more exact analysis for this case shows a non-zero $\bar{D}$ (Nagylaki, personal communication).

Finally, it should be pointed out that other cases of unusual behaviour of linkage disequilibrium are likely to exist in gene family models. Nagylaki (1984*a*, eq. 22) showed that intrachromosomal conversion, under certain conditions, gives $l > g$, and hence $\hat{D} < 0$. Thus, negative linkage disequilibrium can apparently be generated by intrachromosomal conversion. A necessary condition is the formation of some symmetric heteroduplexes (i.e. the possibility that two alleles switch positions), which our models did not allow for.

Thanks to three anonymous reviewers for excellent criticism, and to Tom Nagylaki for useful discussions.

References

Dover, G. A. (1982). Molecular drive: a cohesive model of species evolution. *Nature* **299**, 111–117.

Fink, G. R. & Petes, T. D. (1984). Gene conversion in the absence of reciprocal recombination. *Nature* **310**, 728–729.

Nagylaki, T. (1984*a*). The evolution of multigate families under intrachromosomal gene conversion. *Genetics* **106**, 529–548.

Nagylaki, T. (1984*b*). The evolution of multigene families under interchromosomal gene conversion. *Proceedings of the National Academy of Sciences, USA* **81**, 3796–3800.

Ohta, T. (1983). On the evolution of multigene families. *Theoretical Population Biology* **23**, 216–240.

Slatkin, M. (1986). Interchromosomal biased gene conversion, mutation and selection in a multigene family. *Genetics* **112**, 681–698.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell* **33**, 25–35.