

# Letter to the Editor

## The Effect of Gene Conversion on Intralocus Associations

Peter Andolfatto\* and Magnus Nordborg\*,†

\*Committee on Genetics, Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637 and

†Department of Genetics, Lund University, Sölvegatan 29, 223 62 Lund, Sweden

Manuscript received July 12, 1997

Accepted for publication November 24, 1997

SEVERAL predictions of the genomic pattern of nucleotide polymorphism include the local recombination rate,  $r$ , as a parameter. In typical population genetic models, the parameter  $r$ , for two specified loci (or sites), is properly defined as the probability that a randomly selected gamete produced by a double heterozygote is a recombinant. In a given region,  $r$  can be measured directly through genetic crosses. This is typically done by observing markers spaced several centimorgans apart, in which case recombinants are produced almost exclusively by classical crossing-over. Over small intervals it may be assumed that crossing-over events are equally likely to occur at any point between two markers, and the probability of observing more than one can be neglected. Thus  $r$ , to a reasonable approximation, increases linearly over short physical distances. The problem we wish to draw attention to is that this model is often extrapolated to distances far too small to ignore the added effect of gene conversion events on the overall probability of producing a recombinant.

In fungi and *Drosophila*, a common feature of models of homologous recombination is that Holliday junctions are resolved either as gene conversion alone or as gene conversion with the accompanying exchange of flanking markers (Carpenter 1984). Thus some fraction of genetic exchanges will involve the transfer of short tracts of information from one gamete to another (*i.e.*, gene conversion) without concurrent crossing-over. For clarity, we will refer to gene conversion without crossing-over as “gene conversion,” and gene conversion accompanied by crossing-over as “crossing-over.” Here we show that, at intragenic distances, gene conversion, rather than crossing-over, is likely to be the dominant force that breaks up associations among sites. We discuss implications for population genetic predictions for the behaviour of neutral sites closely linked to a site under balancing or directional selection.

**A crude model:** It is intuitive that gene conversion should increase the rate of exchange for closely linked sites, but should have negligible effects for more distant sites. Assume that over very short intervals, the probability of crossing-over increases linearly with  $d$ , the physical distance in base pairs between two sites. Thus the probability of producing a recombinant by crossing-over equals  $\rho d$ , where  $\rho$  is the probability of a crossing-over event between two given sites per base pair per generation. We suggest that the following crude model describes the added effect of gene conversion events on the rate of exchange between two sites. Define  $\gamma$  as the probability per generation that a given site is included in a gene conversion tract of length  $L$  base pairs, where  $L$  is taken to be fixed. The probability per generation that a gene conversion tract will produce a recombinant for two markers  $d$  base pairs apart will then be roughly  $2\gamma d/L$ , when  $d < L$ . Thus, incorporating both crossing-over and gene conversion,  $r$  depends on  $d$  as follows:

$$\begin{aligned} r &= \rho d + 2\gamma \frac{d}{L} & \text{if } d < L, \\ r &= \rho d + 2\gamma & \text{if } d \geq L. \end{aligned} \quad (1)$$

Reasonable values for  $\rho$  and  $\gamma$  in *Drosophila* are  $\sim 10^{-8}$  and  $\sim 10^{-5}$  per generation, respectively (Ashburner 1989; Hilliker and Chovnick 1981), and  $L$  is estimated to be about 350 base pairs on average (Hilliker *et al.* 1994). It is clear from Equation 1 that gene conversion contributes significantly to  $r$  only when  $d$  is of the same order as  $L$  or smaller.

**Some implications:** Several attempts have recently been made to interpret estimated levels of linkage disequilibrium between polymorphic sites at the intragenic level in the context of competing population genetic models (*e.g.*, Schaeffer and Miller 1993; Miyashita *et al.* 1993; Begun and Aquadro 1995). Ohta and Kimura (1971) showed that in a neutral finite population at equilibrium,

$$E(\sigma^2) \approx \frac{1}{1 + 4Nr}, \quad (2)$$

Corresponding author: Peter Andolfatto, Department of Ecology and Evolution, 1101 E. 57th Street, University of Chicago, Chicago, IL 60637. E-mail: pandolfa@midway.uchicago.edu

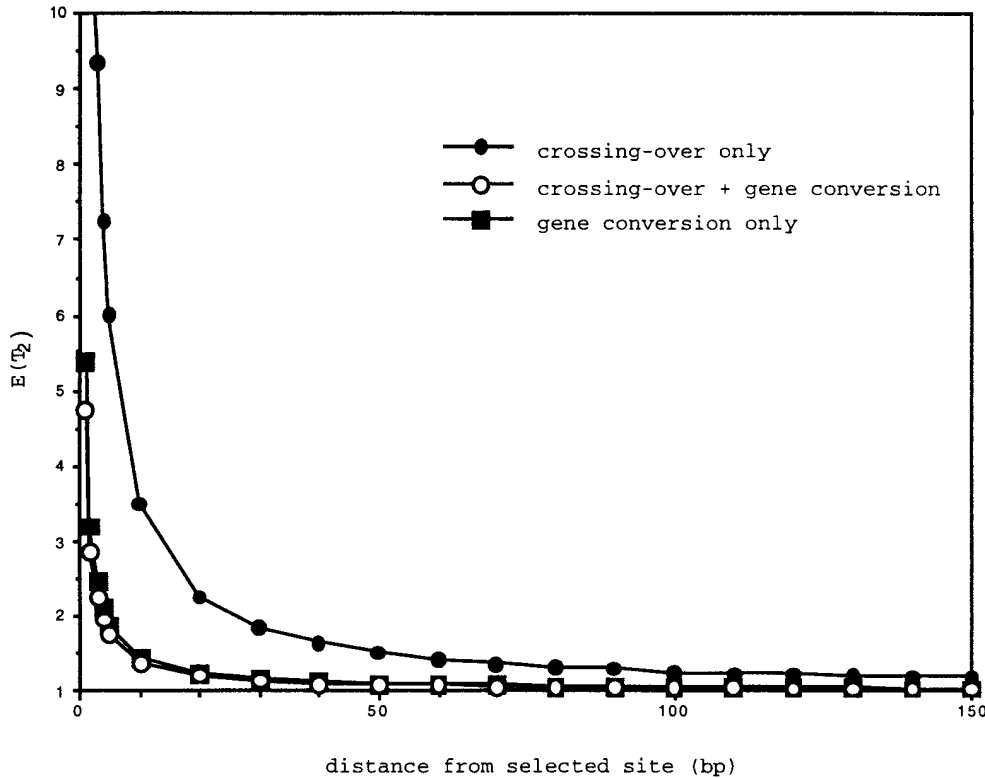


Figure 1.—The expected pairwise coalescence time,  $E(T_2)$ , as a function of physical distance for neutral sites linked to a balanced polymorphism (Equation 3) under various models of recombination:  $r = \rho d$  for crossing-over only,  $r = 2\gamma d/L$  for gene conversion only and see Equation 1 for crossing-over + gene conversion. The parameters used are  $N = 10^6$ ,  $\rho = 10^{-8}$ ,  $\gamma = 10^{-5}$  and  $L = 350$ .

where  $\sigma^2$  is the squared correlation between linked sites and  $N$  is the effective population size. We see that Equation 2 predicts stronger associations, on average, in regions of the genome with “low recombination,” such as the tip of the *X* chromosome in *Drosophila*. However, “low recombination” refers only to empirically measured rates of crossing-over in the region. When rates of crossing-over are low, rates of gene conversion need not be similarly affected; the two mechanisms are to some extent separable (Carpenter 1984; Engebrecht *et al.* 1990). Under Equation 1, we expect less linkage disequilibrium among closely linked sites than expected under a model that only considers rates of crossing-over. It is easy to see that if  $N\gamma$  is sufficiently large, a low local rate of crossing-over will have little effect on the expected level of intragenic disequilibrium. In support of this view, Begun and Aquadro (1995) observed extensive haplotype shuffling in the *y-ac-su(f)* region, despite its position in a region of low rates of crossing-over. This observation is also consistent with other available data for patterns of linkage disequilibria at the tip of the *X* chromosome in *Drosophila melanogaster* (C. H. Langley, personal communication). It should also be noted that because gene conversion will tend to decrease the expected level of linkage disequilibrium, tests for selective sweeps based on haplotypic structure that ignore gene conversion (*e.g.*, Hudson *et al.* 1994) may be conservative.

Another implication concerns our ability to detect balancing selection. If some form of balancing selection acts on a given polymorphism for a sufficiently long

time, a build-up of differences between allelic classes, centered on the selected site, is expected. Consider the expected pairwise coalescence time for a neutral site linked to a two-allele polymorphism with both alleles maintained at equal frequencies, given by

$$E(T_2) = 1 + \frac{1}{4Nr}, \quad (3)$$

where  $r$  is the rate of recombination between the neutral site and the site under selection (Strobeck 1983; Hudson and Kaplan 1988; Hey 1991; Nordborg 1997). Since the expected number of pairwise differences ( $\pi$ ) between alleles is proportional to  $E(T_2)$ , Equation 3 implies that a “peak of polymorphism” surrounding the selected site is expected. Figure 1 illustrates this effect and also demonstrates that gene conversion, even in the absence of crossing-over, will narrow the peak considerably.

Strong signals of balancing selection have been observed at MHC loci in humans (Hughes and Nei 1988) and at plant self-incompatibility loci (see Charlesworth and Guttman 1997 for a review). In *D. melanogaster*, a few loci with polymorphisms thought to be maintained by selection have been investigated, but so far, only the *Adh* locus shows significant evidence for balancing selection (Begun and Aquadro 1994; Eanes *et al.* 1996; Hudson *et al.* 1987; Hudson and Kaplan 1988; Kreitman and Hudson 1991). This has led to speculation that either balancing selection is not a prevalent force maintaining variation or that it only rarely acts long enough for the equilibrium pattern predicted

by Equation 3 to develop (Kreitman and Akashi 1995; Hudson 1996). In light of Figure 1, we suggest an alternative explanation: Perhaps the rate of gene conversion in *Drosophila* is simply too high, given the population size, for this phenomenon to be readily observed. The striking signal seen at *Adh* remains an enigma as it is larger than expected even under a model that does not include gene conversion (Hudson and Kaplan 1988). Possible explanations include past geographic subdivision or the association of *Adh* alleles with the common polymorphic rearrangement, *In(2L)t*.

In conclusion, methods for detecting balancing selection and positive directional selection that rely on  $Nr$  being sufficiently small may not have much power in organisms such as *Drosophila*. An exception may be genomic regions where gene conversion is likely to be suppressed, such as near an inversion breakpoint (see, for instance, Hasson and Eanes 1996). As Hudson and Kaplan (1988) point out, picking an organism with smaller  $N$  may not help, because such organisms are expected to harbor less polymorphism, which reduces our power to detect selection. A better solution may be to study organisms with lower  $r$ , such as partially selfing organisms, where the effective rate of exchange is reduced because most individuals are homozygous (Nordborg *et al.* 1996).

We thank Joy Bergelson, Marty Kreitman, Molly Przeworski, and Jeff Wall for helpful discussions. Comments from Brian Charlesworth, Richard Hudson, Steven Orzack, Baharé Rashidi, and an anonymous reviewer significantly improved this manuscript. P.A. holds a Postgraduate Scholarship from the National Science and Engineering Research Council of Canada.

#### LITERATURE CITED

- Ashburner, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, New York.
- Begun, D. J., and C. F. Aquadro, 1994 Evolutionary inferences from DNA variation at the *6-phosphogluconate dehydrogenase* locus in natural populations of *Drosophila*: Selection and geographic differentiation. *Genetics* **136**: 155–171.
- Begun, D. J., and C. F. Aquadro, 1995 Evolution at the tip and base of the X chromosome in an African population of *Drosophila melanogaster*. *Mol. Biol. Evol.* **12**: 382–390.
- Carpenter, A. T. C., 1984 Meiotic roles of crossing-over and of gene conversion. *Cold Spring Harbor Symp. Quant. Biol.* **49**: 23–26.
- Charlesworth, D., and D. S. Guttman, 1997 Seeing selection in S allele sequences. *Curr. Biol.* **7**: R34–R37.
- Eanes, W. F., M. Kirchner, J. Yoon, C. H. Biermann and I. N. Wang *et al.*, 1996 Historical selection, amino acid polymorphism and lineage-specific divergence at the *G6pd* locus in *Drosophila melanogaster* and *D. simulans*. *Genetics* **144**: 1027–1041.
- Engelbrecht, J., J. Hirsch and G. S. Roeder, 1990 Meiotic gene conversion and crossing-over: their relationship to each other and to chromosome synapsis and segregation. *Cell* **62**: 927–937.
- Hasson, E., and W. F. Eanes, 1996 Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. *Genetics* **144**: 1559–1564.
- Hey, J., 1991 A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. *Theor. Pop. Biol.* **39**: 30–48.
- Hilliker, A. J., and A. Chovnick, 1981 Further observations of intragenic recombination in *Drosophila melanogaster*. *Genet. Res.* **38**: 281–296.
- Hilliker A. J., G. Harauz, A. G. Reaume, M. Gray, S. H. Clark *et al.*, 1994 Meiotic gene conversion tract length distribution within the rosy locus of *Drosophila melanogaster*. *Genetics* **137**: 1019–1026.
- Hudson, R. R., M. Kreitman and M. Aguadé, 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–160.
- Hudson, R. R., and N. L. Kaplan, 1988 The coalescent process in models with selection and recombination. *Genetics* **120**: 831–840.
- Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski and F. J. Ayala, 1994 Evidence for positive selection in the *Superoxide Dismutase (Sod)* region of *Drosophila melanogaster*. *Genetics* **136**: 1329–1340.
- Hudson, R. R., 1996 Molecular population genetics of adaptation, pp. 291–309 in *Adaptation*, edited by M. R. Rose and G. V. Lauder. Academic Press, San Diego.
- Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167–170.
- Kreitman, M., and R. R. Hudson, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**: 565–582.
- Kreitman, M., and H. Akashi, 1995 Molecular evidence for natural selection. *Ann. Rev. Ecol. Syst.* **26**: 403–422.
- Miyashita, N.T., M. Aguadé, and C. H. Langley, 1993 Linkage disequilibrium in the white locus region of *Drosophila melanogaster*. *Genet. Res.* **62**: 101–109.
- Nordborg, M., B. Charlesworth and D. Charlesworth, 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proc. R. Soc. Lon. Ser. B.* **263**: 1033–1039.
- Nordborg, M., 1997 Structured coalescent processes on different time scales. *Genetics* **146**: 1501–1514.
- Ohta, T., and M. Kimura, 1971 Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics* **68**: 571–580.
- Schaeffer, S. W., and E. L. Miller, 1993 Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Genetics* **135**: 541–552.
- Strobeck, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal rearrangement. *Genetics* **103**: 545–555.

Communicating editor: R. R. Hudson