Linkage Disequilibrium in Growing and Stable Populations

Montgomery Slatkin

Department of Integrative Biology, University of California, Berkeley, California 94720

Manuscript received October 19, 1993
Accepted for publication January 24, 1994

ABSTRACT

Nonrandom associations between alleles at different loci can be tested for using Fisher's exact test. Extensive simulations show that there is a substantial probability of obtaining significant nonrandom associations between closely or completely linked polymorphic neutral loci in a population of constant size at equilibrium under mutation and genetic drift. In a rapidly growing population, however, there will be little chance of finding significant nonrandom associations even between completely linked loci if the growth has been sufficiently rapid. This result is illustrated by the analysis of mitochondrial DNA sequence data from humans. In comparing all pairs of informative sites, fewer than 5% of the pairs show significant disequilibrium in Sardinians, which have apparently undergone rapid population growth, while 20% to 30% in !Kung and Pygmies, which apparently have not undergone rapid growth, show significance. The extent of linkage disequilibrium in a population is closely related to the gene genealogies of the loci examined, with "starlike" genealogies making significant linkage disequilibrium unlikely.

LINKAGE disequilibrium is the term most commonly used to describe the nonrandom association of alleles at different loci (LEWONTIN and KOJIMA 1960), although "gametic disequilibrium" or "gametic phase disequilibrium" are used for the same purpose. Linkage disequilibrium or the equivalent terms are, however, used in two ways that are related but still differ. The coefficient of linkage disequilibrium, usually denoted by D, is the difference between the observed frequency of a gametic type and the frequency expected on the basis of random associations of alleles in gametes (LEWONTIN and KOJIMA 1960). This quantity arises naturally in dynamic equations describing the evolution of gamete frequencies under various selection models (KIMURA 1956). Larger (absolute) values of D indicate that there is more linkage disequilibrium in a population. There are numerous other measures of linkage disequilibrium but they usually differ from D only by a normalizing factor. Although the term has been used for more than 30 years, there is still considerable controversy over the best or most useful measure of the extent of linkage disequilibrium (HEDRICK 1987; LEWONTIN 1988).

The second use of the term linkage disequilibrium is in the sense of statistical significance. In this usage, there is linkage disequilibrium between two loci if a statistical test shows that there is significant nonrandom association between alleles at two loci. Both the chi-square test and the Fisher's exact test can be used on the contingency table made up of gametic types (WEIR 1990, Ch. 3). Although the chi-square test has been the most commonly employed because of its computational simplicity, the recent development of rapid algorithms for performing Fisher's exact test and Monte Carlo methods for approximating the results from the exact test make the use of that test much easier (*e.g.*, MEHTA and PATEL 1983; GUO and THOMPSON 1992). The chi-square test is adequate when the expected numbers of each gametic type are not too small (LEWONTIN and FELSENSTEIN 1965) but the Fisher's exact test can be used with confidence for any table of gametic types.

Of course, these two uses of the term linkage disequilibrium are not unrelated. Larger values of D are generally associated with statistically significant disequilibrium. But it is easy to forget that very small values of D may also be consistent with significant linkage disequilibrium. For example, with two alleles at a locus and the numbers of the four gametic types of 66 (AB), 0 (Ab), 31 (aB) and 3 (ab) in a sample of 100, $D = 0.0198$, yet the probability obtained from the Fisher's exact test is only 0.037.

The distinction between the two uses of the term linkage disequilibrium becomes important in considering closely linked neutral loci. Well established population genetics theory, initiated by HILL and ROBERTSON (1968) and OHTA and KIMURA (1969), show that under genetic drift and mutation, values of D are expected to be small in a large population even between closely linked loci. For example, with two alleles at each of two loci, with mutations at a rate μ and in a population of constant size N, the expected value of D^2 is just 0.043 between completely linked loci and drops to 0.00056 for $R = 4Nc = 50$ (EWENS 1979, p. 203). GRIFFITHS (1981) provides general results of this type for a wide variety of mutation models. This theory is correct, but it is sometimes interpreted as meaning that statistically significant nonrandom associations are not likely to be found in such cases. The results presented here show that this conclusion is not justified. In fact,

there is a substantial probability of detecting significant linkage disequilibrium between closely linked polymorphic loci in a population of constant size. The reason that the analytic theory gives a different impression is that the expectation of D^2 predicted by the analytic theory includes cases in which one or both loci are fixed, while in practice only polymorphic loci would be tested for disequilibrium. The results presented below also show that there might be much less significant disequilibrium in a rapidly growing population.

SIMULATION MODEL

The results described below are based on data generated by a simulation program that uses a coalescent model, similar to that described by HUDSON and KAPLAN (1985). The simulation program generated the allelic states of two linked loci in a sample. In the simulations, the basic quantities were the allelic states of a certain number, i, of gametes that carried at least one gene ancestral to those in a sample. Some gametes had ancestors of alleles in the sample at both loci (denoted by AB), others of them had ancestors of alleles in the sample at only the A locus (denoted by $A-$), and others had ancestors at only the B locus (denoted by $-B$). Going backward in time, there were two kinds of events that could occur. There could be a coalescent event, in which case two gametes were descended from a common ancestral gamete. Or there could be a recombination event, in which case the ancestor of each gamete was assumed to have been derived by recombination from two other gametes.

What happened during a coalescent event depended on the kinds of gametes that coalesced. For example, if an AB gamete coalesced with another of the same type (say $A'B'$), then the ancestral gamete is of the same type ($A''B''$), there were coalescent events at both loci, and the number of ancestral gametes was reduced by 1. If instead an AB gamete coalesced with $A'-$, then only the A locus coalesced and the resulting ancestral gamete was $A''B$. If an $A-$ gamete coalesced with a $-B$ gamete, neither locus coalesced and the resulting gamete was AB. At any time in the past, each pair of gametes had the probability $1/(2N)$ of coalescing each generation. If there were i gametes carrying alleles that were ancestral to those in the sample, the total probability of a coalescent event in a generation was approximately $i(i-1)/(4N)$ if $i \ll N$.

In each generation going backward in time, each gamete in the population had a probability c of being involved in a recombination event. The value of c was always small enough that there was essentially no chance of more than one recombination event occurring in any generation, *i.e.*, $ic \ll 1$. The most likely type of recombination event is between one of the i gametes carrying ancestral alleles and one of the $2N - i$ gametes that are not. If an AB gamete is descended from a recombination event of this type, the two parental gametes were $A-$ and $-B$, thus increasing the value of i by 1. If an $A-$ or a $-B$ gamete was involved in a recombination event of this type, there was no effect. If an AB gamete and an $A'B'$ gamete are derived from a recombination event, the parental gametes were AB' and $A'B$. If an $A-$ gamete and a $-B$ gamete were derived in a recombination event, however, only the AB gamete remained because the other gamete contained no ancestors of alleles in the sample.

The mutation model used in this study was the K allele symmetric model, with K possible allelic states. A mutation occurred with probability μ per generation per locus, and when an allele mutated there was a probability of $1/(K-1)$ of becoming one of the other $K-1$ alleles. This model with $K = 4$ corresponds to the JUKES-CANTOR (1969) model of nucleotide evolution. Different values of K, K_1 and K_2, were allowed at the two loci. The results from each replicate were expressed as a contingency table of gametic types. That table could be as large as $K_1 \times K_2$ but in any replicate there might be fewer than the maximum possible number actually represented.

The Fisher's exact test was used to assign a probability value P to the contingency table of gametic types from each replicate. The value of P is the probability of finding a table with the same marginal totals which has a probability equal to or less than the probability of the observed table under the assumption that the entries are drawn from the appropriate hypergeometric distribution. If the table was relatively small, an exact P value was found using a program that implemented the algorithm of MEHTA and PATEL (1983). This program could analyze $2 \times n$ tables for $n \leq 8$, 3×3, and 4×4 tables. For larger tables, I wrote a program to perform a Monte Carlo estimation of the P value. The algorithm was based on a minor modification of the algorithm presented by GUO and THOMPSON (1992). In the Monte Carlo estimation of P values, the program did 1000 replicates to estimate each P value. As discussed by GUO and THOMPSON (1992), the results from a Monte Carlo simulation can be made arbitrarily close to those from the exact test by increasing the sample size.

The power of Fisher's exact test to detect significant non-random associations depends on the marginals of the table (*i.e.*, the row and column sums) (BENNETT and HSU 1960). For example, if there is only a single copy of one of two allele at a locus, it is very difficult to detect significant nonrandom association. To understand the importance of marginals in the contingency tables, I used different criteria to determine whether a locus was polymorphic. A locus was polymorphic if there were at least y copies of at least two alleles. The value of y was independent of K, the potential number of alleles at that locus, and there was no assurance that a locus that was polymorphic by this criterion would have all K allelic

types represented unless $K = 2$. At the end of each replicate, both loci were tested to determine whether they were polymorphic for the specified value of y. If either locus was not, then that replicate was discarded. The simulation program continued for each set of parameter values until 1000 contingency tables satisfying this criterion were produced.

The demographic history of the population strongly affected the results. Two demographic histories were used in this study. In a constant population, there were N individuals for all times in the past. In a growing population, there were N_0 individuals between the present and time τ in the past and N_1 individuals before time τ. The case of particular interest is the one in which $N_0 \gg N_1$ indicating that very rapid growth took place at time τ. ROGERS and HARPENDING (1992) have examined this model and shown that very rapid growth of this type can lead to a nearly starlike gene genealogy in which most coalescence events occur within a relatively narrow range of times. SLATKIN and HUDSON (1991) showed that sufficiently rapid exponential population growth over a long period of time also leads to a starlike gene genealogy. Thus the choice of the stepwise change in population size, although unrealistic, results in gene genealogies that are similar to those from more realistic demographic models. The value of τ used in the simulations, 10,000 generations, was chosen to be of the same order of magnitude as the apparent time of expansion of the population leading to modern Sardinians, assuming 15 years per generation (DI RIENZO and WILSON 1991).

RESULTS

Constant population size: For the case of no recombination, $c = 0$, a substantial number but by no means all pairs of loci show statistically significant linkage disequilibrium using the tail probability from Fisher's exact test ($P < 0.05$). The proportion depends somewhat on the mutation rate and decreases substantially for values of $4N\mu = \theta > 1$ (Figure 1). The maximum of these curves is at approximately $\theta = 1$, which would be expected on theoretical grounds discussed below. There is strong dependence on y for $\theta = 0.1$, less dependence for $\theta = 1$ and none for $\theta = 10$ (Figure 1). The decrease in number of significant P values with large values of θ is not surprising. High enough mutation rates would cause the states of each locus to be essentially random, but clearly values of θ much greater than 10 would be needed and those values would imply a much greater level of nucleotide diversity than is observed in current samples of both nuclear and mitochondrial DNA. There is relatively strong dependence on the numbers of alleles per locus, as shown in Table 1, with a much higher proportion of significant P values with larger values of K. Thus, more polymorphic loci would in general be more useful for detecting nonrandom associations.

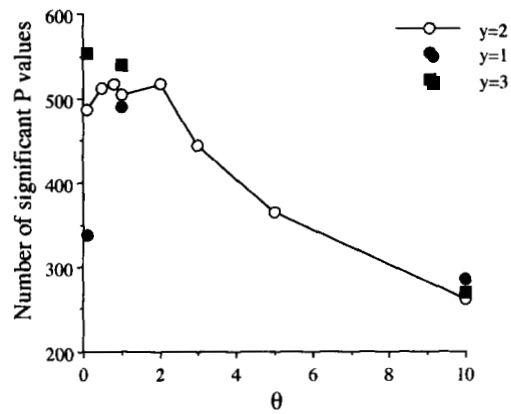


FIGURE 1.—Numbers of significant P values in simulations of a population of constant size under genetic drift and mutation. Each point represents the results of 1000 replicates for a set of parameter values. In all cases, there was no recombination ($c = 0$), 100 gametes were sampled, and $N = 10{,}000$. The results shown differ in the mutation rate ($\theta = 4N\mu$) which was the same at both loci, and y, the criterion for polymorphism, which was also the same for both loci.

Table 1 also shows that the results depend on the recombination rate, c. These results differ markedly from the predictions of the analytic theory, in which the expected value of D^2 decreases with $1/R$ for large R (EWENS 1979, Equation 6.82). Table 1 also shows that even with $R = 50$ there is a substantial chance of finding significant disequilibrium, especially when there are more alleles per locus. These results are in contrast to the expected values of D^2 for two of the same cases: with $\theta = 1$ and $K_1 = K_2 = 2$, $E(D^2) = 0.046$ for $R = 0$ and 0.00053 for $R = 50$. In making this comparison, it is worth recalling that the analytic theory is for the unconditional expectation, while the simulation results are for the case in which both loci are polymorphic. In these simulations, the same mutation rate μ was assumed for both loci, but comparable results were obtained when different mutation rates were assumed for the two loci.

There was some dependence of the results on sample size but not as much as might be expected. For example, with $\theta = 1$, $c = 0$, and $y = 2$, and $K_1 = K_2 = 2$, the number of significant P values for samples of 40, 60 and 80 gametes were 352, 426, 492 in 1000 replicates, compared with the value of 504 shown in Figure 1 for a sample size of 100.

Rapidly growing population: The results for a rapidly growing population are quite different. If $N_0 \gg N_1$ there is little detectable disequilibrium even when $R = 0$, as shown in Table 2. In Table 2, the proportion of significant P values decreases with μ. Thus for values of μ comparable to those used to obtain the results in Figure 1 for a constant population size, we would expect to find even fewer significant P values. With much smaller values of μ, the proportion of replicates in which both loci are polymorphic using the $y = 2$ criterion becomes very

TABLE 1

Dependence of number of significant *P* values on the number of alleles per locus

$K_1 \times K_2$	2 × 2	2 × 3	2 × 4	2 × 5	2 × 8	3 × 3	4 × 4	8 × 8
$R = 0$	504	607	680	707	725	714	766	840
$R = 50$	227	284	279	238	273	275	362	427

The values shown are the numbers of replicates out of 1000 in which the tail probabilities (the *P* values) from Fisher's exact test were less than 0.05, as described in the text. In all cases, a population of constant size $N = 10{,}000$ was simulated, 100 gametes were sampled, $\theta = 4N\mu = 1.0$, and $y = 2$ was used as the criterion for polymorphism. K_i is the maximum number of possible alleles at locus i.

TABLE 2

Significant nonrandom association in a rapidly growing population

Part A. $N_0 = 10^6$, $N_1 = 10^3$						
μ	10^{-5}	5×10^{-4}	10^{-4}	2×10^{-4}	5×10^{-4}	10^{-3}
	30	37	47	45	50	53
Part B. $N_1 = 10^3$ and $\mu = 2.5 \times 10^{-4}$						
N_0	10^6	10^5	10^5	10^4		
	47	54	95	404		

These results were obtained using the simulation model of a rapidly growing population, as described in the text. The numbers shown are the numbers of replicates (out of 1000) for which *P* values from Fisher's exact test were less than 0.05. In all cases 100 gametes were sampled, $K_1 = K_2 = 2$, $y = 2$ was the criterion for polymorphism at both loci, and the time τ at which the population size changed was 10,000 generations in the past.

small. The comparison between the results for constant and growing populations is more appropriately done for cases in which the product of the average coalescence time ($2N$ or τ) and the mutation rate are the same, because in those cases roughly the same levels of nucleotide diversity are expected. Viewed in this way, the results in Figure 1 and Table 2A are comparable.

Table 2B shows that as the ratio N_0/N_1 decreases, significant linkage disequilibrium becomes more frequent and the results become similar to those for a population of constant size. Comparable results were found for cases in which $K > 2$ at one or both loci. Not surprisingly, the proportion of significant *P* values was not affected by increases in the recombination rate when $N_0 \gg N_1$ (results not shown).

The results for a rapidly growing population clearly have to depend on the value of τ. When τ is extremely small, the model becomes equivalent to one in which the population is of constant size N_1 and when τ is extremely large the model becomes equivalent to one in which the population is of constant size N_0. As discussed by ROGERS and HARPENDING (1992), only for intermediate values of τ does a starlike genealogy result. The value of τ used was chosen to be of the same order of magnitude as the time of rapid growth in the Sardinian population discussed below.

APPLICATIONS

The preceding results suggest that for a given recombination rate between a pair of loci, the extent of significant disequilibrium depends strongly on the demographic history of a population. To test this prediction, I examined published mitochondrial DNA sequences in humans. There is apparently no recombination in mtDNA in humans and in some regions there are abundant sites that are polymorphic by any criterion. For this analysis, I used the sequences of the hypervariable region of the *D* loop of the 46 Sardinian individuals studied by DI RIENZO and WILSON (1991) and the sequences of the 37 !Kung and 34 Pygmies (both eastern and western) studied by VIGILANT *et al.* (1991). The Sardinians were chosen because the distribution of pairwise differences in DNA sequence is approximately a Poisson distribution, suggesting that there has been a period of rapid population growth in their history (DI RIENZO and WILSON 1991; SLATKIN and HUDSON 1991). In contrast, the distributions of pairwise differences in mtDNA sequence in the !Kung and the Pygmies suggest that those populations have experienced little or no population growth in their recent history (DI RIENZO and WILSON 1991). In analyzing these data, I took account of the fact that some of the haplotypes were found more than once. That is not necessary for inferring gene genealogies using parsimony, but it is for testing for linkage disequilibrium.

The results from analyzing these three data sets are shown in Table 3. As anticipated, there is little significant disequilibrium among pairs of polymorphic sites in the Sardinians, in fact no more than would be expected at random. In the !Kung and the Pygmies, however, there is significant disequilibrium found for a substantial number of pairs of sites. The dependence on y, particularly for the Pygmies, combined with the results in Figure 1 indicates that at numerous sites mutation rates are low enough that only one gamete carried a distinct nucleotide. In Figure 1, the strongest dependence on y was for $\theta = 0.1$. This conclusion is consistent with the results of WAKELEY (1993) who

TABLE 3

Nonrandom associations between polymorphic sites in mtDNA

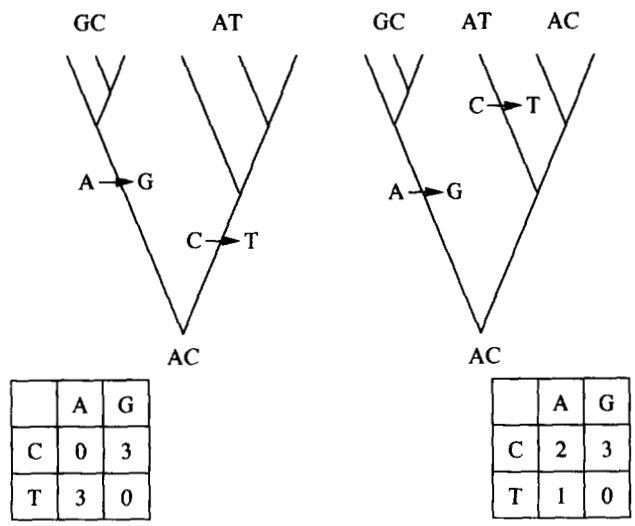
	$y = 1$	$y = 2$	$y = 3$
Sardinians	29/861 (3.3)[a]	9/190 (4.7)	4/91 (4.4)
Pygmies	58/1179 (4.9)	50/359 (13.9)	35/165 (21.2)
!Kung	38/409 (9.3)	35/158 (22.2)	15/46 (32.6)

The value of y is the minimum number of copies of at least two alleles at a locus in order for that locus to be counted as polymorphic. The numbers shown are the numbers of significant P values ($P < 0.05$) in Fisher's exact test found in all comparisons of polymorphic loci divided by the total number of pairwise comparisons of polymorphic loci. Because of missing data for some sites for some samples, not all pairs of sites that were polymorphic separately were polymorphic when considered together. Hence, there are not always $n(n-1)/2$ comparisons, for some value of n.

[a] Percentages are shown in parentheses.

found evidence of substantial variation in mutation rate at different sites in the hypervariable region of the D loop in human mtDNA.

It should be noted that these results are not completely comparable to those from the simulations. In each population, there is a single gene genealogy for all sites, which necessarily induces a correlation between them. The P values for different pairs of sites are not independent. In the simulations, on the other hand, the replicates were independent and hence all the P values were also.



	A	G
C	0	3
T	3	0

	A	G
C	2	3
T	1	0

FIGURE 2.—An illustration of a hypothetical gene genealogy of a sample of six gametes carrying two completely linked nucleotide sites. In both cases, the ancestral type was AC. In the left hand diagram, mutations occurred at both sites in the internal branch of the unrooted tree. As a result, the 2×2 table of gametic types is as shown, indicating perfect linkage disequilibrium between the two sites. In the right hand diagram, one mutation occurs on the internal branch and the other on a terminal branch, resulting in the other 2×2 table which shows much less evidence of linkage disequilbrium.

DISCUSSION

The simulation results presented here show that the extent of statistically significant disequilibrium depends both on the recombination rate between loci and on the demographic history of the population from which the samples were taken. A substantial fraction of polymorphic loci sampled from a population of constant size will exhibit significant disequilibrium but that will not necessarily be the case in a rapidly growing population. These predictions are confirmed by the analysis of mtDNA in humans. These results point to the importance of restricting the theoretical results to polymorphic loci, a point that has been made by HUDSON and KAPLAN (1985) also. The analytic theory predicts the expected value of D^2 under a variety of assumptions but that expectation includes cases in which one or both loci are fixed. In practice, calculations of D and tests for nonrandom associations between loci are done only between polymorphic loci. The present results also point to the importance of using different criteria for polymorphism at a locus.

These results can be easily understood by considering the source of linkage disequilibrium of a pair of completely linked genes. If there is complete linkage, alleles at both loci have the same gene genealogy. Suppose that there is single mutational event at each locus. As illustrated in Figure 2, significant disequilibrium between polymorphic loci is most likely to be detected if both mutations occur on the same internal branch and if there are several descendants from that branch. Locating the mutational events on the gene tree in other ways reduces the chance of detecting significant disequilibrium. The same principle holds if more than one mutational event occurs at each locus, as long as relatively few occur.

In a population of constant size, the longest internal branches of the gene tree are found near the root, because the probability of a coalescent event decreases almost with the square of the number of lineages (HUDSON, 1990). Thus, it should be of no surprise that significant disequilibrium is often but not always found between completely linked loci in a population of constant size. Furthermore, we can see why it is most likely that significant disequilibrium will be found when $\theta = 4N\mu = 1$ at both loci. In that case, there will be one mutation on average in the gene genealogies of each locus, because the total length of the gene genealogy is approximately $4N$ generations (HUDSON 1990). If more mutations occurred they would tend to randomize the allelic states and reduce the probability of significant disequilibrium, as found in the simulations. In a rapidly growing population, by contrast, there would probably be no long internal branches because the gene genealogy would be starlike. Thus significant disequilibrium is much less likely to be found.

These results should not be interpreted as meaning that there can be no linkage disequilibrium at closely

linked neutral loci in a population that has grown very rapidly. What they do mean is that, in such a population, linkage disequilibrium will not arise because of mutation and genetic drift alone. A starlike gene genealogy that results from rapid population growth means that genetic drift is effectively not occurring once the population starts to grow rapidly because there are few if any coalescent events after that time. In my simulations, all of the polymorphism was assumed to have arisen by mutation after the population was founded, so linkage disequilibrium did not often arise. If, however, a population is founded by a small group in which there is already linkage disequilibrium between a particular pair of loci, then the subsequent disequilibrium between those loci will be determined by the recombination rate between them and the time since the population started to grow rapidly. Very closely linked loci will remain in significant disequilibrium for a long time. Linkage disequilibrium in the founding population could have been present because the founders were an admixed group, because immigrants carried a novel gametic type to the population early in the growth phase, or because disequilibrium was maintained by selection. The difference between a rapidly growing population and one that remains of constant size is that substantial linkage disequilibrium between closely linked loci can be created by genetic drift alone in a population of constant size but not in one that has grown sufficiently rapidly.

This study grew out of discussions with N. B. FREIMER. I thank M. P. CUMMINGS for bringing a program to perform Fisher's exact test to my attention, J. P. WAKELEY for providing formatted copies of mtDNA sequence data for analysis, and M. P. CUMMINGS, N. B. FREIMER, S. P. OTTO, J. P. WAKELEY, B. S. WEIR and anonymous reviewers for helpful comments on earlier versions of this paper. This study has been supported in part by a grant from the National Institutes of Health.

LITERATURE CITED

BENNETT, B. M., and P. HSU, 1960 On the power function of the exact test for the 2 × 2 contingency table. Biometrika **47:** 393–398.

DI RIENZO A., and A. C. WILSON, 1991 The pattern of mitochondrial DNA variation is consistent with an early expansion of the human population. Proc. Natl. Acad. Sci USA **88:** 1597–1601.

EWENS, W. J., 1979 *Mathematical Population Genetics.* Springer-Verlag, Berlin.

GRIFFITHS, R. C., 1981 Neutral two-locus multiple allele models with recombination. Theor. Popul. Biol. **19:** 169–186.

GUO, S. W., and E. A. THOMPSON, 1992 Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics **48:** 361–372.

HEDRICK, P. W., 1987 Gametic disequilibrium measures: proceed with caution. Genetics **117:** 331–341.

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

HUDSON, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–44.

HUDSON, R. R., and N. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution in protein molecules, pp. 21–123 in *Mammalian Protein Metabolism,* edited by H. N. MUNRO, Academic Press, New York.

KIMURA, M., 1956 A model of a genetic system that leads to closer linkage by natural selection. Evolution **10:** 278–287.

LEWONTIN, R. C., 1988 On measures of gametic disequilibrium. Genetics **120:** 849–852.

LEWONTIN, R. C., and J. FELSENSTEIN, 1965 The robustness of homogeneity tests in 2 × *N* tables. Biometrics **21:** 19–33.

LEWONTIN, R. C., and K. KOJIMA, 1960 The evolutionary dynamics of complex polymorphisms. Evolution **14:** 458–472.

MEHTA, C. R., and N. R. PATEL, 1983 A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. J. Am. Stat. Assoc. Ser. B **78:** 427–434.

OHTA, T., and M. KIMURA, 1969 Linkage disequilibrium at a steady state determined by random genetic drift and recurrent mutation. Genetics **63:** 229–238.

ROGERS, A. R., and H. HARPENDING, 1992 Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9:** 552–569.

SLATKIN, M., and R. R. HUDSON, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129:** 555–562.

VIGILANT, L., M. STONEKING, K. HAWKES and A. C. WILSON, 1991 African populations and the evolution of human mitochondrial DNA. Science **253:** 1503–1507.

WAKELEY, J., 1993 Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. J. Mol. Evol. **37:** 613–623.

WEIR, B. S., 1990 *Genetic Data Analysis.* Sinauer Associates, Sunderland, Mass.

Communicating editor: B. S. WEIR