Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm

MONTGOMERY SLATKIN* & LAURENT EXCOFFIER†

Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, U.S.A. and †Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, 12, Rue G. Revilliod, 1227 Carouge, Switzerland

We generalize an approach suggested by Hill (*Heredity*, **33**, 229–239, 1974) for testing for significant association among alleles at two loci when only genotype and not haplotype frequencies are available. The principle is to use the Expectation-Maximization (EM) algorithm to resolve double heterozygotes into haplotypes and then apply a likelihood ratio test in order to determine whether the resolutions of haplotypes are significantly nonrandom, which is equivalent to testing whether there is statistically significant linkage disequilibrium between loci. The EM algorithm in this case relies on the assumption that genotype frequencies at each locus are in Hardy–Weinberg proportions. This method can accommodate X-linked loci and samples from haplodiploid species. We use three methods for testing the significance of the likelihood ratio: the empirical distribution in a large number of randomized data sets, the χ^2 approximation for the distribution of likelihood ratios, and the Z^2 test. The performance of each method is evaluated by applying it to simulated data sets and comparing the tail probability with the tail probability from Fisher's exact test applied to the actual haplotype data. For realistic sample sizes (50–150 individuals) all three methods perform well with two or three alleles per locus, but only the empirical distribution is adequate when there are five to eight alleles per locus, as is typical of hypervariable loci such as microsatellites. The method is applied to a data set of 32 microsatellite loci in a Finnish population and the results confirm the theoretical predictions. We conclude that with highly polymorphic loci, the EM algorithm does lead to a useful test for linkage disequilibrium, but that it is necessary to find the empirical distribution of likelihood ratios in order to perform a test of significance correctly.

Keywords: EM algorithm, likelihood ratio test, linkage disequilibrium, microsatellite loci.

Introduction

There has been a revival of interest in linkage disequilibrium because of the vast numbers of closely linked loci that can now be examined and because of the possibility of using disequilibrium to map genetic diseases (Jorde, 1995). This revival of interest raises several theoretical issues. We will be concerned here with the question of what can be inferred from data in which genotypes but not haplotypes are observed, with particular emphasis on the case where there are numerous alleles per locus, as is typical with microsatellite and other hypervariable loci.

Most of the theory and most of our understanding of linkage disequilibrium is in the context of two alleles at each of two loci. The coefficient of linkage disequilibrium, D, is the difference between the frequency of a haplotype and its expected frequency under the assumption of random association of alleles (Lewontin & Kojima, 1960). The value of D indicates the extent of linkage disequilibrium. If haplotypes are known, a variety of statistical methods, including the χ^2-test, Fisher's exact test and several resampling tests, can be used to test the statistical significance of the linkage disequilibrium (Weir, 1990). If statistical significance is detected, then it can be concluded that D is significantly different from zero.

If only genotypic data are available, Hill (1974) suggested that the value of D could be estimated by a method that is equivalent to what was later called

the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). Because this method maximizes the likelihood of the data (under the assumption of random mating), Hill suggested that a test of the significance of D could be carried out using a likelihood ratio test. That is, twice the logarithm of the ratio of the likelihood found by the EM algorithm to the likelihood under the assumption that $D = 0$ would have a χ^2 distribution with one degree of freedom. Hill (1974) showed that, asymptotically, the power of this test would be comparable to a χ^2-test on the contingency table of haplotypes if the sample size of the genotypic sample were $\sqrt{2}$ larger. Hill concluded that, in general, it would not be worth the extra effort to collect haplotype data by isolating isofemale lines or using information from large pedigrees, because simply increasing the sample size by a modest amount would have the same effect.

With more than two alleles per locus, the problem becomes much more difficult. With k alleles at one locus and l alleles at the other, there are kl D-values, one for each pair of alleles, although only $(k-1)(l-1)$ of them are independent (Weir, 1979; Long et al., 1995). With two highly polymorphic loci, it is easy to reach a situation in which there are more possible haplotypes than there are gametes represented in the sample. One solution is to choose one allele of interest at each locus and use the two-allele theory to infer that D-value and test for its significance. That approach is appropriate if there is some other reason to be interested in that particular pair of alleles. But if the goal is to determine whether there is overall evidence of nonrandom association of alleles at the two loci, then grouping the alleles will reduce the ability to detect significance (Watkins et al., 1994). If haplotypes can be resolved, Fisher's exact test can be performed on tables of arbitrary size using a Monte Carlo algorithm (Guo & Thompson, 1992), and a simulation study by Slatkin (1994) showed that increasing the numbers of allele per locus substantially increases the probability of detecting a significant association between closely linked neutral loci.

Long et al. (1995) have recently considered the problem of inferring haplotypes and testing for nonrandom associations among alleles using an approach nearly the same as that developed here, and many of their conclusions are the same as ours. Following Hill (1974), they used the EM algorithm (described in the next section) and a test for the significance of the resulting likelihood ratios. Long et al. (1995) found, as we have, that the χ^2 approximation is inadequate when there are several alleles

per locus. Our paper differs from that of Long et al. (1995) in three ways. First, for reasons discussed later, we think that the bootstrap method they propose for testing significance of likelihood ratios may not be adequate and we use a randomization test for significance instead. Secondly, we evaluate the performance of our method on simulated data for which the true significance of the association between loci, as determined by Fisher's exact test, is known. This allows us to explore the performance of this method as a function of sample size and allele number. Thirdly, we do not consider the problem of higher order disequilibria, which Long et al. (1995) examined using an approach similar to that of Hill (1975).

The EM algorithm

We will consider the case of only two loci with k alleles at the first (the A locus) and l alleles at the second (the B locus). The sample consists of counts of the numbers of individuals ($n_{ii'jj'}$) with the genotype $A_iA_{i'}B_jB_{j'}$ in a sample of diploid individuals. From these, we can compute the allele frequencies at the two loci, p_i and q_j. The next step is to use the EM algorithm to find the maximum likelihood estimates of the frequencies of A_iB_j, f_{ij}. The general procedure which is applicable to data with an arbitrary number of loci is described by Hill (1975), Long et al. (1995) and Excoffier and Slatkin (1995).

An initial guess of the haplotype frequencies, $f_{ij}^{(0)}$, is chosen. Then these haplotype frequencies are used to determine the resolution of each double heterozygote in the sample by assuming that the probability of each of the two possible resolutions is in proportion to the expectation given the current estimates of the haplotype frequencies. For example the probability that the $A_1A_2B_1B_2$ genotype is resolved into A_1B_1 and A_2B_2 is $f_{11}^{(0)}f_{22}^{(0)}/(f_{11}^{(0)}f_{22}^{(0)}+f_{12}^{(0)}f_{21}^{(0)})$. Given the resolution of every double heterozygote, we can then compute a new estimate of the fs, $f_{ij}^{(1)}$. Dempster et al. (1977) show that this method increases the likelihood with each iteration, so when the iteration converges the resulting frequencies represent at least local maximum likelihood estimates of the haplotype frequencies. The likelihood is, under the assumption of multinomial sampling, proportional to

$$L \propto \prod_{ii'jj'} (c_{ii'jj'}f_{ij}f_{i'j'})^{n_{ii'jj'}}$$

where the product is taken over all genotypes and $c_{ii'jj'} = 1$ if $i = i'$ and $j = j'$ (double homozygotes), 4 if $i \neq i'$ and $j \neq j'$ (double heterozygotes) and 2

otherwise. There is no guarantee that there is a single maximum and it is best to try numerous initial guesses for the haplotype frequencies (Long *et al.*, 1995).

In this analysis, like that of Long *et al.* (1995) but unlike that of Hill (1974), we do not estimate the disequilibrium coefficients directly but instead work with the haplotype frequencies. This is simply for notational and programming convenience. Our concern here is not with estimating the Ds but with testing for nonrandom association of alleles at the two loci. To do this we take the ratio of the final likelihood L^* (computed using the haplotype frequencies found by the EM algorithm) to L_0, the likelihood under the assumption of no disequilibrium ($f_{ij} = p_i q_j$). From this procedure, we obtain a test statistic, $S = 2\log_e(L^*/L_0)$. This test statistic is chosen because, for large sample sizes, it will have a χ^2 distribution.

We performed three tests to test for significant associations between loci.. The first is the standard χ^2-test with $(k-1)(l-1)$ d.f., which would be valid if sample sizes were sufficiently large. The result of this test is a P-value which we will denote by P_C (C for χ^2). Secondly, we performed a randomization test. For each data set analysed, we randomly permuted alleles at each locus (although randomizing one locus would have been sufficient) and then ran the EM algorithm to obtain a value of S for that replicate. In a set of 1000 randomizations, we obtained an empirical distribution of S under the null hypothesis of no association and hence could find the proportion of replicates which produced values of S equal to or greater than that from the data. We called that proportion P_R (R for randomized).

Our third test, which is an alternative to the log-likelihood ratio test, is one described by Elandt-Johnson (1971, p. 356) for comparing simultaneously the maximum likelihood (ML) haplotype frequencies to their expectations under the hypothesis of Hardy–Weinberg (HW) equilibrium. Let $\mathbf{f}_{ML}$ be the vector of size $(k \times l) - 1$ functionally independent maximum likelihood haplotype frequencies, and $\mathbf{f}_{HW}$ be the vector of expected frequencies under HW equilibrium obtained as $f_{ij} = p_i p_j$. In the absence of linkage disequilibrium, the statistic

$$Z^2 = (\mathbf{f}_{ML} - \mathbf{f}_{HW})' \mathbf{I}_{HW} (\mathbf{f}_{ML} - \mathbf{f}_{HW})$$

follows approximately a χ^2 distribution with $(k-1) \times (l-1)$ d.f., where $\mathbf{I}_{HW}$ is the inverse of the haplotype frequencies variance–covariance matrix. The variance–covariance matrix need not be computed and $\mathbf{I}_{HW}$ can be obtained directly as the expected information matrix of negative second partial derivatives of the log-likelihood function evaluated for HW haplotype frequencies. Note that the elements of the vector $(\mathbf{f}_{ML} - \mathbf{f}_{HW})$ are of the form $f_{ij} - p_i p_j$ and are thus the expected D-values. Therefore, Z^2 can be seen to be a weighted sum of squared linkage disequilibrium values, very similar to the X_T^2 statistic used to test for global disequilibrium between two loci having multiple alleles when the gametic phase is known (Weir, 1990, p. 94). The tail probability obtained by comparing the observed value of Z^2 will be denoted by P_Z.

To describe the performance of each of these three methods for detecting significance, we compared the three P-values (P_C, P_R, and P_Z) with the tail probability from Fisher's exact test, P_E (E for exact). With two and three alleles per locus, we used a program that enumerated all possible tables and provided the value of P_E. With more alleles per locus, we used a Monte Carlo program written by one of us (M.S.). This program had already been tested extensively against published results and the results from the other program. When the Monte Carlo program was used, 17 000 replicates were used for each case to ensure that the resulting estimate of P_E was accurate to within 1 per cent (Guo & Thompson, 1992).

Slatkin (1994) and Peterson *et al.* (1995) have argued that when haplotype frequencies are known, P_E provides a useful description of the extent of nonrandom association between loci, particularly when the goal is to compare the extent of association between different pairs of loci in the same sample of individuals. Other authors (e.g. Watkins *et al.*, 1994) have used a measure of association based on the χ^2 statistic for the contingency table of haplotype numbers (which is not to be confused with the χ^2-test for the likelihood ratio statistic). With large sample sizes, using P_E is equivalent to using the χ^2 statistic from the contingency table because for a given number of alleles the χ^2 statistic can be converted to a tail probability that is asymptotically equal to P_E. The advantage of using the tail probability from the exact test is that there are no problems created by low frequency alleles or haplotypes, and hence there is no need arbitrarily to group low frequency alleles in order to eliminate cells with low expected counts.

Simulation study

To test the performance of the procedure described above, we carried out an extensive simulation study of data sets in which we knew the haploytpes and

then grouped them arbitrarily into genotypes. With each data set, we first performed Fisher's exact test to obtain the tail probability, P_E. We then compared the value of P_E with those of P_C, P_R and P_Z. To find P_C, we used 100 random starting conditions for the EM algorithm and used the smallest resulting value of P_C (which resulted from the largest of the likelihood values found). Even if the global maximum was not found, we can still be sure that the value of P_C is no larger than that which would have been obtained if the global maximum likelihood solution had been found. To ensure that these results were correct we each wrote a program to implement the EM algorithm and carry out the χ^2-test. The results from these two programs agreed in detail.

To compute P_R in the large number of data sets we examined, it was not feasible to use 100 random starting conditions in each randomized replicate. The results from the computations of P_C also indicated the numbers of local maxima found when 100 random initial conditions we used. Based on those results, we found it was sufficient to use a single initial condition for cases with two, three and five alleles per locus and 10 initial conditions for those with eight alleles per locus, for 100 and 200 gamete samples, and five in the cases when 300 gametes were sampled.

We assume that the value of P_E provides a useful measure of the degree of nonrandom association between the two loci and a useful test of significance of that association, although many other measures can be and have been used. Maiste (1993) reviews and tests several possibilities. We used two measures of performance of each method. One is the fraction of replicates (ψ_{SS}) in which P_C, P_R or P_Z indicated that there was significant disequilibrium given that P_E indicated the same level of significance (i.e. given $P_E < \alpha$, ψ_{SS} is the fraction of replicates for which P_C, P_R or P_Z was also less than α). A larger value of ψ_{SS} would indicate a larger chance of successfully detecting significant nonrandom association when it is actually present (according to the exact test). The second is the fraction of replicates (ψ_{SN}) in which P_E indicates that there is no significant association, whereas P_C, P_R or P_Z indicates that there is (i.e. given $P_E > \alpha$, ψ_{SN} is the fraction in which P_C, P_R or P_Z is less than α). A larger value of ψ_{SN} indicates a greater chance of a false positive identification of significant disequilibrium. We will present results for $\alpha = 0.05$ but the same patterns were found for $\alpha = 0.01$.

We analysed data obtained using the simulation program described by Slatkin (1994). This program used a coalescent method to generate data at two linked loci under a variety of assumptions about recombination rates, mutation rates and demographic histories of a randomly mating population. In all cases, we assumed the same number of alleles per locus ($k = l$) and all combinations of $k = 2$, 3, 5 and 8, with sample sizes of $n = 100$, 200 and 300 gametes. For each combination of n and k, the parameters of the simulation program were adjusted so that roughly equal numbers of data sets with significant ($P_E < 0.05$) and nonsignificant ($P_E > 0.05$) degrees of association were obtained. A total of 1000 data sets were generated for each combination. Then the χ^2, randomization, and Z^2 tests were performed by other programs to generate values of P_C, P_R or P_Z as described above, and finally a third program took the combined tables of P-values to compute ψ_{SS} and ψ_{SN}. These results are shown in Table 1.

Results

Table 1 shows the performance of these different methods as compared to Fisher's exact test. We can see that for $k = 2$ and $k = 3$, all the methods perform about equally well and show that there is a substantial probability of detecting significance if it is present (ψ_{SS}) and with an acceptably low probability of false positives (ψ_{SN}). The values of ψ_{SS} are somewhat difficult to interpret because it was not possible to control for the distribution of values of P_E in the simulated data sets. With more alleles per locus ($k = 5$ and $k = 8$) some important differences among the methods are apparent. The probability based on the resampling method (P_R) shows that there is still a substantial chance of detecting significance if it is present but that the probability of false positives is somewhat larger than 5 per cent, which is the significance level used, although it is never greater than roughly 10 per cent. The performance of the other two methods becomes sufficiently poor as to make them unusable. As was also found by Long et al. (1995), P_C can either show too much or too little significance depending on the values of n and k. The value of P_Z tends always to show too much significance.

Discussion

These simulations confirm Hill's (1974) conclusion that with two alleles per locus, the EM algorithm applied to genotypic data combined with the likelihood ratio test provides a sufficiently accurate way to test for linkage disequilibrium between loci and that it is probably not worth the extra trouble to

Table 1 Comparison of results obtained using the empirical (P_R), asymptotic (P_C) and Z^2 (P_Z) tail probabilities with results obtained used Fisher's exact test (P_E). In all cases, n is the number of gametes in the sample and k is the number of alleles at both loci. (a) Values of ψ_{SS} (the fraction of times significance was detected with one of the other tests, given that it was detected with the exact test) in the 1000 artificial data sets are generated as described in the text. The sample size is the number of cases out of 1000 in which $P_E < 0.05$. (b) Values of ψ_{SN} (the fraction of times significance was detected with the likelihood ratio test, given that it was not detected with the exact test) in the sample of 1000 artificial data sets generated as described in the text. The sample size is the number of cases out of 1000 in which $P_E > 0.05$

k	Empirical	Asymptotic	Z^2	Sample size
(a)				
$n = 100$				
2	0.682	0.700	0.660	421
3	0.386	0.454	0.408	152
5	0.603	0.561	0.803	544
8	0.772	0.405	0.987	659
$n = 200$				
2	0.783	0.788	0.769	632
3	0.403	0.445	0.445	238
5	0.490	0.651	0.669	625
8	0.380	0.638	0.831	597
$n = 300$				
2	0.788	0.780	0.660	613
3	0.435	0.481	0.408	370
5	0.386	0.518	0.803	479
8	0.449	0.753	0.745	774
(b)				
$n = 100$				
2	0.0597	0.0760	0.0570	579
3	0.0483	0.0707	0.0767	848
5	0.0601	0.0394	0.3202	456
8	0.0811	0.0000	0.8475	341
$n = 200$				
2	0.0598	0.0652	0.0462	368
3	0.0656	0.0787	0.0944	762
5	0.0640	0.1600	0.1833	375
8	0.1156	0.4937	0.4888	403
$n = 300$				
2	0.0672	0.0698	0.0620	387
3	0.0714	0.0762	0.0810	630
5	0.0595	0.1286	0.1958	521
8	0.1087	0.3628	0.7467	226

resolve the genotypes into haplotypes if the only goal is to test for linkage disequilibrium. They also show that the same finding holds if there are more alleles per locus, provided that a resampling method is used to test the significance of the resulting likelihood ratios. The χ^2 approximation to the distribution of likelihood ratios is clearly inadequate when there are several alleles per locus.

The resampling method we used is different from the bootstrap method described by Long *et al.* (1995). We think that the problem with the bootstrap method is that it assumes that the gamete frequencies inferred by the EM algorithm are correct, and then it resamples from inferred haplotype composition to determine the distribution of the likelihood ratios under sampling. In addition, the number of alleles per locus may differ from one bootstrap sample to another, so the number of d.f. will also change, and hence the distribution of χ^2-values in the bootstrap sample would not necessarily be the null distribution. In contrast, our resampling method gives an approximation to the null distribution of likelihood ratios because it allows for the possibility that the inferred haplotype frequencies may not be correct. We think that it is important to allow for this possibility because, in examining the detailed output of several data sets, we found that the inferred frequencies of rare haplotypes were often incorrect even when the global maximum likelihood was found. Nevertheless, our conclusions about the usefulness of the EM algorithm and the inadequacy of the χ^2 approximation for the distribution of likelihood ratios are the same as those of Long *et al.* (1995).

X-linked loci and haplodiploid species

In some situations, phase information about some gametes in a sample is available. That is true for X-linked loci on gametes sampled from the heterogametic sex (males in humans and *Drosophila*), and from males in haplodiploid species such as thrips and the Hymenoptera. It can also be true if haplotypes have been resolved using pedigrees or chromosome extractions for some but not all individuals in a sample of diploids. All of these cases can be easily accommodated in the EM algorithm. It is necessary only to treat some haplotypes as known. They then contribute to the haplotype frequencies which are then used to resolve haplotypes in doubly heterozygous individuals for whom the gametic phases are unknown. In doing this for X-linked loci and haplodiploids, it is necessary to assume the same haplotype frequencies in males and females.

Table 2 An application of the EM algorithm and the likelihood ratio test to the data of Peterson *et al.* (1995). For both P_C and P_R, 100 random initial conditions were used. The results are shown only for adjacent loci for which P_E (from Fisher's exact test) <0.05. k and l are the numbers of alleles at the two loci. The numbers of equilibria of the EM algorithm in 100 randomly chosen initial conditions are also presented

	k		l	P_E	P_C	P_R	No. equil.
D4S1601	6	D4S467	5	0.0196	0.1718	0.079	1
D4S551	12	D4S544	4	0.0265	0.2119	0.052	2
D4S419	6	D4S1562	5	0.0060	0.0764	0.029	1
D4S425	7	D4S616	10	0.0088	0.2800	0.039	3
D4S391	13	D4S230	12	0.0055	0.5197	0.004	6
D4S230	12	D4S418	9	0.0210	0.5649	0.053	17
D4S418	9	D4S595	5	0.0008	0.3039	0.102	3
D4S885	8	D4S1632	6	0.0357	0.2660	0.044	1
D4S405	9	D4S190	8	0.0020	0.2537	0.003	4
D4S414	12	D4S423	8	0.0000	3.585×10^{-10}	0.000	1
D4S423	8	D4S410	4	0.0330	0.1947	0.114	1
D4S410	4	D4S1089	6	0.0429	0.3837	0.267	1

Application

To illustrate the application of this method to experimental data, we re-examined the data described by Peterson *et al.* (1995). The data were from 32 microsatellite loci on chromosome 4 in a sample of 50 unrelated Finnish individuals. There was sufficient information from a pedigree so that the genotypes could be almost completely resolved into haplotypes. There were between 4 and 16 alleles per locus. Peterson *et al.* (1995) performed Fisher's exact test on all pairs of loci, after the few remaining ambiguities in haplotype identification were randomly resolved, and found low P_E values between several pairs of loci. They used these results to show that a physical map of this region of chromosome 4 was more likely to be accurate than a genetic map.

The question here is whether it was necessary to use related individuals to resolve the haplotypes in this sample. To answer this question, we reanalysed the data using the EM algorithm described above. As with the simulated data sets, we used 100 randomly chosen starting conditions and kept track of all equilibria found. Results are shown in Table 2 for the 12 pairs of adjacent loci that showed significant disequilibrium using the exact test. For six out of 12, P_R indicated significance at the 5 per cent level when P_E did, and in two other cases the values of P_R were close to 5 per cent. The values of P_C indicate that the χ^2 approximation for large sample sizes is not valid and if used alone would be misleading. For none of the 20 pairs of adjacent loci for which $P_E > 0.05$ did P_R or P_C indicate significance. These results are consistent with the simulation results presented above. They show that some significant disequilibrium could be detected from the genotypic data but that additional information was gained from the haplotypes.

Conclusions

Our results show that the generalization of Hill's (1974) method, namely the EM algorithm, does lead to a useful test of nonrandom associations among pairs of loci when only genotypic data are available. We have shown that this method can be used for hypervariable loci such as microsatellite loci, even with modest sample sizes, but that a resampling method is necessary to carry out a statistical test.

Acknowledgements

References

DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B.*, **39**, 1–38.

ELANDT-JOHNSON, R. C. 1971. *Probability Models and Statistical Methods in Genetics.* Wiley & Sons, New York.

EXCOFFIER, L. AND SLATKIN, M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, **12**, 921–927.

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

HILL, W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. *Heredity*, **33**, 229–239.

HILL, W. G. 1975. Tests for association of gene frequencies at several loci in random mating diploid populations. *Biometrics*, **31**, 881–888.

JORDE, L. B. 1995. Linkage disequilibrium as a gene mapping tool. *Am. J. Hum. Genet.*, **56**, 11–14.

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

LONG, J. C., WILLIAMS, R. C. AND URBANEK, M. 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. *Am. J. Hum. Genet.*, **56**, 799–810.

MAISTE, P. J. 1993 *Comparisons of Statistical Tests for Independence at Genetic Loci with Many Alleles.* Ph.D. Thesis, North Carolina State University.

PETERSON, A. C., LEHESJOKI, A. E., DE LA CHAPELLE, A., DI RIENZO, A., SLATKIN, M. AND FREIMER, N. B. 1995. The distribution of linkage disequilibrium over anonymous genome regions. *Hum. Mol. Genet.*, **4**, 887–894.

SLATKIN, M. 1994. Linkage disequilibrium in growing and stable populations. *Genetics*, **137**, 331–336.

WATKINS, W. S., ZENGER, R., O'BRIEN, E. ET AL. 1994. Linkage disequilibrium patterns vary with chromosomal location: a case study from the von Willebrand factor region. *Am. J. Hum. Genet.*, **55**, 348–355.

WEIR, B. S. 1979. Inferences about linkage disequilibrium. *Biometrics*, **25**, 235–254.

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