Am. J. Hum. Genet. 56:799–810, 1995

An E-M Algorithm and Testing Strategy for Multiple-Locus Haplotypes

Jeffrey C. Long,[1] Robert C. Williams,[2] and Margrit Urbanek[1]

[1]Laboratory of Neurogenetics, NIAAA/NIH, Rockville, MD; and [2]Department of Anthropology, Arizona State University, Tempe

Summary

This paper gives an expectation maximization (EM) algorithm to obtain allele frequencies, haplotype frequencies, and gametic disequilibrium coefficients for multiple-locus systems. It permits high polymorphism and null alleles at all loci. This approach effectively deals with the primary estimation problems associated with such systems; that is, there is not a one-to-one correspondence between phenotypic and genotypic categories, and sample sizes tend to be much smaller than the number of phenotypic categories. The EM method provides maximum-likelihood estimates and therefore allows hypothesis tests using likelihood ratio statistics that have χ^2 distributions with large sample sizes. We also suggest a data resampling approach to estimate test statistic sampling distributions. The resampling approach is more computer intensive, but it is applicable to all sample sizes. A strategy to test hypotheses about aggregate groups of gametic disequilibrium coefficients is recommended. This strategy minimizes the number of necessary hypothesis tests while at the same time describing the structure of disequilibrium. These methods are applied to three unlinked dinucleotide repeat loci in Navajo Indians and to three linked HLA loci in Gila River (Pima) Indians. The likelihood functions of both data sets are shown to be maximized by the EM estimates, and the testing strategy provides a useful description of the structure of gametic disequilibrium. Following these applications, a number of simulation experiments are performed to test how well the likelihood-ratio statistic distributions are approximated by χ^2 distributions. In most circumstances the χ^2 grossly underestimated the probability of type I errors. However, at times they also overestimated the type 1 error probability. Accordingly, we recommend hypothesis tests that use the resampling method.

Introduction

Many highly polymorphic loci are now available for linkage analyses, forensics, and other population-genetic applications (Weber and May 1989; Weissenbach et al. 1992; Gyapay et al. 1994). This wealth of information has created a need for efficient statistical methods and computer algorithms to estimate such basic quantities as allele and haplotype frequencies. While counting alleles directly provides maximum-likelihood (ML) allele-frequency estimates for a locus with all codominant alleles (Gart and Nam 1984), the method requires a one-to-one correspondence between genotypes and phenotypes. Consequently, direct haplotype counting is impossible for multiple loci because multiple-locus heterozygosity masks genotypic categories (Hill 1974). The correspondence between genotypic and phenotypic categories is further obscured by recessive alleles. This is important because many highly polymorphic genetic systems possess a battery of mutually codominant alleles and one recessive allele; these systems are commonly referred to as "generalized ABO-like" (Yasuda and Kimura 1968). Genetic estimation is also challenged by the fact that with high polymorphism sample sizes tend to be smaller than the number of genotypic categories, so that any particular genotype is likely to be unique or absent in a sample (Guo and Thompson 1992; Weir 1992). To illustrate these complexities, consider a typical HLA analysis; there might be 15 HLA-A alleles, 25 HLA-B alleles, and 7 HLA-C alleles occurring in a single sample. This allows the possibility of 2,625 three-locus haplotypes, 3,446,625 genotypes, and 701,932 phenotypes (assuming one recessive allele at each locus). Not all of these will be realized in large populations, let alone in samples drawn from them.

The likelihood functions for single- and multiple-locus genetic models are easily written, even with recessive alleles, but there are a number of analytical problems. For example, the multiple-locus likelihood function can have hundreds or thousands of parameters, and numerical methods must be used to solve for a maximum. Many numerical methods are sensitive to rounding errors, and it usually cannot be proved that a particular solution is the global maximum. Moreover, a very large number of hypotheses can be formulated with so many parameters, and the significance level for a set of hypotheses must be adjusted to account for the number of tests performed (Sokal and Rohlf 1981; Weir 1990).

The purposes of this paper are fourfold. First, an expectation maximization (EM) algorithm (Cepellini et al. 1955; Smith 1957; Dempster et al. 1977; Ott 1977) is described

Received May 12, 1994; accepted for publication December 9, 1994.

Address for correspondence and reprints: Dr. Jeffrey C. Long, Laboratory of Neurogenetics, NIAAA/NIH, 12501 Washington Avenue, Rockville, MD 20852.

0002-9297/95/5603-0031$02.00

for ML estimation of multiple-locus haplotype frequencies. Second, a likelihood ratio strategy is given for testing hypotheses about gametic disequilibrium. Third, data resampling techniques are used to evaluate the sampling distributions of the likelihood-ratio statistics. Fourth, two data sets are analyzed in order to illustrate the proposed methods. An EM algorithm similar to the one presented here was used by Baur and Danilovs (1980) to estimate three-locus HLA haplotype frequencies, and they were able show its superiority over a competing method (Piazza 1975). Unfortunately, these authors did not explain the algorithm in detail, nor did they connect it to ML theory. Consequently, its application was not wide in subsequent years. The ML connection is crucial because it shows the statistical soundness of the technique and it enables formal hypothesis testing.

Genetic and Statistical Background

We will describe the algorithm for a genetic model with three loci, with each locus possessing a battery of mutually codominant alleles and one allele that is recessive to all others (i.e., generalized ABO-like systems). This genetic model serves to illustrate most problems encountered in the estimation and testing process, but the algorithm's basic features are readily applied to both more and less complicated situations.

Consider three polymorphic loci designated A, B, and C, with n_A, n_B, and n_C alleles, respectively. The first allele in each series is recessive to all of the other alleles that are both detectable and mutually codominant. Recessive alleles are due to limitations of the laboratory method (e.g., serology) or to the absence of a gene (e.g., Rh-D negative), but they are not due to typing errors or data missing due to sample degradation, etc. Let p_a denote the frequency of the ath allele at the first locus ($a = A_0, A_1, \ldots, A_{n_A}$), q_b denote the frequency of the bth allele at the second locus ($b = B_0, B_1, \ldots, B_{n_B}$), and r_c denote the frequency of the cth allele at the third locus ($c = C_0, C_1, \ldots, C_{n_C}$). H_{abc} denotes the haplotype carrying the ath, bth, and cth nonallelic genes, and its frequency is f_{abc}. Bennett (1954) showed that this frequency can be decomposed into the product of the single-locus gene frequencies and appropriately weighted second- and third-order disequilibrium coefficients:

$$f_{abc} = p_a q_b r_c + p_a \mathrm{D(BC)}_{bc} + q_b \mathrm{D(AC)}_{ac} + r_c \mathrm{D(AB)}_{ab} + \mathrm{D(ABC)}_{abc} . \quad (1)$$

This construction is useful because it removes lower-order disequilibrium effects from higher-order disequilibrium components (Weir 1990). Allele frequencies at the individual loci and haplotype frequencies at pairs of loci are obtained by summing over the appropriate three-locus haplotype frequencies. The disequilibrium coefficients are obtained by simple algebra (see table 1). The second- and third-order disequilibrium coefficients can be considered measures of linkage disequilibrium, but gametic disequilibrium is a more appropriate term because nonallelic genes can be associated on gametes for reasons other than linkage (such as population structure). Moreover, application of basic principles shows that only very tight linkage results in disequilibrium (Hartl and Clark 1989).

ML estimates of allele frequencies and disequilibria are provided by applying the same algebra to ML estimates of haplotype frequencies as would be applied to the population parameters. Thus, an ML set of haplotype frequency estimates is sufficient to describe the entire system. As usual, it is necessary to estimate one parameter fewer than the number of haplotypes, because the haplotype frequencies sum to 1.0. For our purposes, statistics will be distinguished from their corresponding parameters by using primes (e.g., f'_{abc} estimates f_{abc}).

Methods

Estimation

Consider a random sample of N individuals taken with replacement from a large and random-mating diploid population. The logarithmic likelihood function of the haplotype model is

$$\ln L = \sum_{i=1}^{N} \ln Pr(P_i) , \quad (2)$$

where $\ln Pr(P_i)$ is the logarithm of the probability of the ith person's phenotype. $Pr(P_i)$ is calculated by summing the probabilities of all constituent genotypes (i.e., all the genotypes that can express the phenotype), and the genotype frequencies are given by Hardy-Weinberg expansion of haplotype frequencies. While log-likelihood functions are usually written as a summation over all possible phenotypes, it is more efficient with highly polymorphic systems to sum over individuals because there are fewer individuals sampled than there are potential phenotypes. In accordance with this version of the likelihood function, the EM algorithm described below processes the data by person at each iteration, rather than by phenotype. The expectation step of the algorithm is concerned with the quantities $E[N_{abc}|P_i]$, which are the expected numbers of haplotypes, given a phenotype, while the maximization step involves counting these expected numbers over all individuals.

The following data structure is useful for implementing the algorithm. A record consisting of six fields (a pair for each locus) is constructed for each phenotype. For each locus the two fields are scored as follows: If no alleles are detected, zeros are placed in both fields. If one allele is detected, its specificity is recorded in the first field and the second field is assigned a zero. If two alleles are detected, then each field records one of the specificities. A person is

Table 1

Summary Statistics from Haplotype Frequencies

Component	Formula
Allele frequencies	$\begin{cases} p_a = \sum_b \sum_c f_{abc} \\ q_b = \sum_a \sum_c f_{abc} \\ r_c = \sum_a \sum_b f_{abc} \end{cases}$
Two-locus haplotype frequencies	$\begin{cases} f_{ab} = \sum_c f_{abc} \\ f_{ac} = \sum_b f_{abc} \\ f_{bc} = \sum_a f_{abc} \end{cases}$
Pairwise disequilibria	$\begin{cases} \mathrm{D(AB)}_{ab} = f_{ab} - p_a q_b \\ \mathrm{D(AC)}_{ac} = f_{ac} - p_a r_c \\ \mathrm{D(BC)}_{bc} = f_{bc} - q_b r_c \end{cases}$
Three-way disequilibria	$\mathrm{D(ABC)}_{abc} = f_{abc} - p_a q_b r_c - p_a \mathrm{D(BC)}_{bc} - q_b \mathrm{D(AC)}_{ac} - r_c \mathrm{D(AB)}_{ab}$

heterozygous for detectable alleles at a locus if both fields are assigned nonzero values; a nonzero value followed by a zero indicates that the person is either homozygous for the detectable allele or heterozygous for the detectable allele and the recessive; zeros in both fields indicate that the person is homozygous for the recessive allele.

The specific steps of our E-M algorithm are as follows: (*a*) The alleles at each locus are numbered with consecutive integers, beginning with zero, which is reserved for the blank allele. (*b*) A set of trial haplotype frequencies is chosen. (*c*) A variable T_{abc} is created for each H_{abc} to keep a running total of its expected numbers. (*d*) For each phenotype in the sample, (i) the constituent genotypes are identified by placing the person's phenotype into 1 of the 27 categories of the generalized three-locus system (fig. 1). The genotypes for a particular phenotype are generated as shown in figure 2. (ii) The expected number of copies for each haplotype contributing to the constituent genotypes is calculated according to

$$E[n_{abc} \mid P_i] = \frac{2 f_{abc} \sum_{H_{a^*b^*c^*} \in P_i} f_{a^*b^*c^*}}{Pr(P_i)}, \qquad (3)$$

where $E[n_{abc} \mid P_i]$ is the expected number of copies of H_{abc} within P_i, and $f_{a^*b^*c^*}$ is the frequency of another haplotype $H_{a^*b^*c^*}$ that can combine with H_{abc} to form P_i (for homozygotes, $H_{abc} \equiv H_{a^*b^*c^*}$). The summation is taken over the set of haplotypes $H_{a^*b^*c^*}$ that can combine with H_{abc} to form P_i. (iii) T_{abc} is updated for each H_{abc} for which $E[n_{abc}] > 0$. (*e*) The initial haplotype frequency estimates are improved, by replacing them with $T_{abc}/2N$. (*f*) The log likelihood of the sample is evaluated according to equation (2). (*g*) Steps *c*–*f* are repeated until the log likelihood stabilizes.

In programming the algorithm, the variables T_{abc} can be placed in a three-dimensional array. By using step *a*, the allele designations on the haplotype record its array address. Step *di* (figs. 1 and 2) requires a large amount of program code, and it is impractical to apply with more than three loci. An alternative is to generate from the typing data all haplotypes that could form a genotype that is compatible with the multiple-locus phenotype and then to proceed with step *dii*. This approach requires substantially less code, and it is extended to more than three loci easily, by increasing array sizes. We have programmed both versions of the algorithm. Unfortunately, the latter method takes substantially longer to run because many genotypes that could not have produced the phenotype must be evaluated. Like other EM algorithms, the likelihood increases on each iteration, until a peak on the likelihood surface is reached (Dempster et al. 1977; Ott 1977), but there is a danger that a local extreme has been reached.

Comments

There are several features of this algorithm that deserve attention. First, it is unnecessary to specify in advance which haplotypes occur in the sample. The method calculates frequencies for all possible haplotypes. In practice, many estimated frequencies are zero. Second, it constrains all haplotype frequency estimates to nonnegative values. Third, for systems without null alleles, ML allele frequencies are provided after the first iteration. Haplotype frequencies are properly constrained by these marginal totals on all subsequent iterations. Fourth, the computational speed of a program that uses step *di* is nearly independent

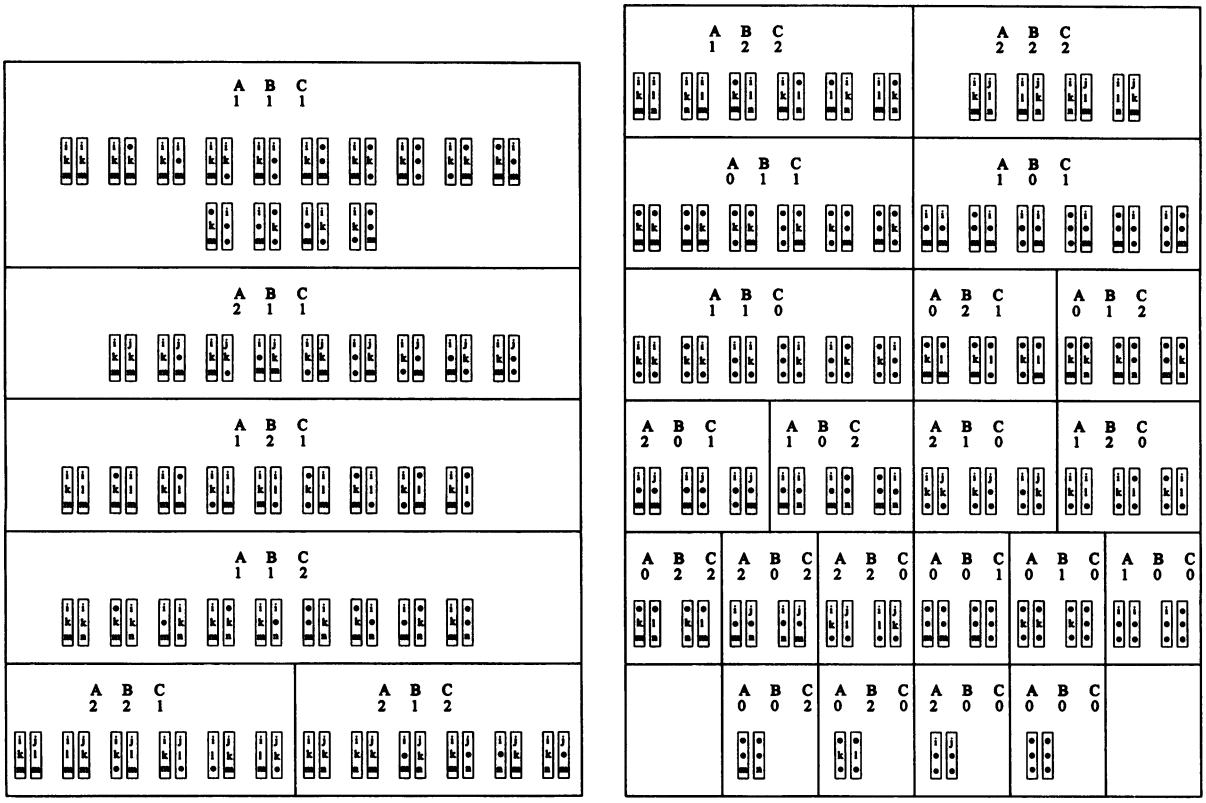

Figure 1 Twenty-seven phenotypic categories for the generalized three-locus model (see also Haseman and Elston 1972). The phenotype, according to the number of identifiable alleles at each locus, is given at the top of each large box. The small vertical boxes depict haplotypes. The letters within the boxes have the following meanings: i and j are detectable alleles at locus A; k and l are detectable alleles at locus B, and m and n are detectable alleles at locus C. Recessive alleles are represented by dots at all three loci. Actual typings are substituted for i, j, l, m, and n, as appropriate (see fig. 2).

of the number of alleles at the loci; it depends on the sample size.

Hypothesis Testing

We recommend a testing strategy that avoids focusing on the individual parameters but captures the essence of the system. Four disequilibrium-coefficient sets are defined. The first set contains all three-way coefficients (D_{abc}), and each of the next three sets include all coefficients between a particular pair of loci (e.g., set 2 contains all D_{ab}). Table 2 identifies 16 models incorporating some, or all, coefficient sets. Following a test for global equilibrium, a *forward-selection* testing strategy, whereby significant component sets are added to the most restricted model, is recommended. The test for global equilibrium is accomplished by contrasting M15, which is the full model defined by equation (1), with M0 (see table 2). If global equilibrium is rejected, then the locus pairs in disequilibrium are identified by contrasting M1–M3 with M0. Finally, three-way disequilibrium is established by testing a model with three-way disequilibrium (e.g., M9–M15) and all significant pairwise sets against an alternative with only the significant pairwise effects (e.g., M1–M7). This strategy provides a structured analysis to evaluate all levels of gametic disequilibrium while at the same time holding the number of hypothesis tests to a minimum. It is possible, although unlikely, that global equilibrium is rejected, but disequilibrium between specific pairs cannot be demonstrated. This can arise for two reasons. First, it is possible to have three-way disequilibrium while at the same time having equilibrium between all pairs. This can be demonstrated by contrasting M8 with M0. Second, the contrast of M15 with M0 provides the most powerful test. Failure to reject specific subhypotheses could result from reduced power. The contrast of M7 with M0 gives a more powerful test for pairwise disequilibrium, but it will not demonstrate which pairs of loci have nonrandomly associated alleles.

In all cases, the test statistic is twice the negative logarithmic likelihood ratio

$$G = -2(\ln L_{HR} - \ln L_{H\Omega}), \qquad (4)$$

where $lnL_{H\Omega}$ is the natural logarithm of the likelihood function computed under a general hypothesis and $\ln L_{HR}$ is the

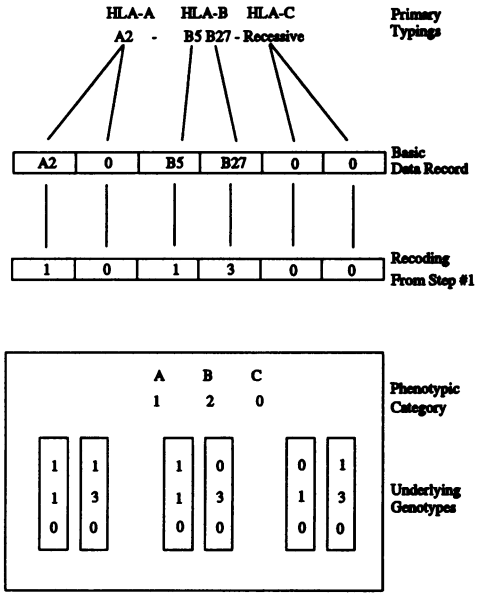

Figure 2 The method for identifying the constituent genotypes for a phenotype, illustrated for an individual who typed positive for HLA-A2, HLA-B5, HLA-B27, and recessive for all HLA-C alleles.

logarithm of the likelihood function computed under a restricted version of H_Ω. The null hypothesis tested by G is that the more general model does not fit the data significantly better than does the restricted model. With large samples, the distribution of G theoretically approximates a χ^2 distribution with df equal to the number of parameters eliminated from H_Ω in order to obtain H_R (Weir 1990).

Although the χ^2 approximations are appealing in principle, their large sample requirements may be unattainable in practice. An alternative mechanism for constructing statistical distributions is provided by resampling the observed data. Such empirical distributions avoid large sample assumptions at the expense of increased computer time. In addition, resampling is useful for determining what conditions are necessary for valid χ^2 approximations and for identifying when a χ^2 test is likely to be liberal (i.e., reject the null hypothesis too frequently) or conservative (i.e., maintain the null hypothesis too frequently) for a given level of type I error. In brief, the empirical distribution for G is built as follows: A replicated sample is constructed by drawing N pairs of haplotypes at random, with replication from the haplotype probability distribution specified by the null hypothesis, H_R. Each haplotype pair specifies a multiple-locus genotype for which the corresponding phenotype is recorded. G is computed for the replicated sample and saved. The preceding steps are repeated a large number of times, and the saved G values constitute the empirical distribution. The simulated G value above which the most extreme $100\alpha\%$ of the simulated statistics lie is the empirical $100\alpha\%$ significance level. This resampling procedure is an application of the more general statistical bootstrapping method (Efron and Tibshirani 1993).

The forward-parameter selection process advocated here is necessary for the resampling tests. This owes to the fact that two- and three-way disequilibrium coefficients are scale dependent. For example, the magnitude of pairwise disequilibrium depends on allele frequencies, and the magnitude of three-way disequilibrium depends on both allele frequencies and pairwise disequilibrium (Piazza 1975). Thus, all nonsignificant disequilibrium effects should be excluded when haplotype frequencies are computed for simulating a reduced model.

The expected haplotype frequencies for the reduced models (i.e., M0–M14) do not have direct EM estimates. They must be obtained by adjusting the haplotype frequency estimates from the full model. For M0–M3 this is accomplished by plugging only the specified components into equation (1). For models with disequilibrium between more than one pair of loci but without three-way disequilibrium (M4–M7), iterative proportional fitting (Deming and Stepan 1940) provides nonnegative three-locus haplotype frequencies with three-way equilibrium and the exact allele frequencies and pairwise disequilibria from the full model. Models with three-way disequilibrium but unsaturated for pairwise effects (M8–M14) require a procedure such as the Newton-Raphson iteration to meet these conditions (see Agresti 1990).

One advantage of this testing strategy is that it holds the

Table 2

Multiple-Locus Haplotype Models

Model	Set of Components				
	pqr	pD(BC)[a]	qD(AC)[b]	rD(AB)[c]	D(ABC)[d]
M0	1	0	0	0	0
M1	1	1	0	0	0
M2	1	0	1	0	0
M3	1	0	0	1	0
M4	1	0	1	1	0
M5	1	1	0	1	0
M6	1	1	1	0	0
M7	1	1	1	1	0
M8	1	0	0	0	1
M9	1	1	0	0	1
M10	1	0	1	0	1
M11	1	0	0	1	1
M12	1	0	1	1	1
M13	1	1	0	1	1
M14	1	1	1	0	1
M15	1	1	1	1	1

NOTE.—Each component set defined above (e.g., pD(BC)) includes values over all a, b, c (e.g., all p_aD(BC)$_{bc}$) as defined by equation (1).

[a] df = $(n_B - 1)(n_C - 1)$. n_A, n_B, and n_C refer to the no. of alleles at the first, second, and third loci, respectively.

[b] df = $(n_A - 1)(n_C - 1)$. See note a.

[c] df = $(n_A - 1)(n_B - 1)$. See note a.

[d] df = $(n_A - 1)(n_B - 1)(n_c - 1)$. See note a.

number of tests to a minimum. Nonetheless, testing several subhypotheses is still required, and the significance level for a specific test (α') requires the Bonferroni correction: $\alpha' = 1 - (1 - \alpha)^{1/k}$, where k is the number of tests performed (e.g., see Weir 1990). Another advantage of this testing strategy arises from the relation $-2[\ln L(M0) - \ln L(M15)] = -2[\ln L(M0) - \ln(M7)] - 2[\ln L(M7) - \ln L(M15)]$. The left-hand side of the equation is G for the global equilibrium hypothesis. The additive components on the right-hand side provide Gs for testing pairwise equilibrium and three-way equilibrium, respectively. This partition of the χ^2 for total gametic disequilibrium into additive components relating to two- and three-way interactions is a convenient description of the structure of disequilibrium. Documented Pascal programs for implementing the algorithm and testing strategy proposed here are available free of charge from the authors for DOS-operated PCs and Solaris-run Sun systems.

Applications

We have applied this algorithm- and hypothesis-testing strategy to two data sets in order to experience situations that will be encountered in real data analyses. We were most interested in determining (1) the algorithm's sensitivity to starting conditions, and (2) the correspondence between the simulated null distributions for G statistics and their theoretical χ^2s. Since the correspondence between the simulated and theoretical distributions was poor at times (see Results), we took the general characteristics of the data sets (e.g., sample sizes, numbers and frequencies of alleles, and presence of recessive alleles) as base lines for a number of simulation experiments. The simulation experiments were designed to reveal the conditions where the χ^2 distribution is most appropriate for G.

Data Sets

The first data set consists of typings at three loci encoding short tandem repeat (STR) polymorphisms (locus name/primers: D18S57/AFM147yg7 [Weissenbach et al. 1992], D20S115/AFM218yg3 [Weissenbach et al. 1992], and D22S274/AFM164th8 [Weissenbach et al. 1992]) in a sample of $N = 38$ Navajo Indians in New Mexico. Each of these three loci have dinucleotide repeat motifs. Aliquots of genomic DNA were PCR-amplified using *Taq* polymerase and fluorescent dye–labeled primers. Following amplification, PCR reaction products were identified using an Applied Biosystems (ABI) 373A DNA sequencer, and fragment size determinations were made using the ABI GENESCAN software. The laboratory procedures are fully described by Michelini et al. (in press).

The second data set consists of 619 three-locus phenotypes for the Class I HLA loci (HLA-A, HLA-B, and HLA-C) for members of the Gila River Indian Community in central Arizona. The histocompatibility alleles were detected serologically, and the methods of detection have been described elsewhere along with the other details of the sample (Williams and McCauley 1992). Both data sets were examined for recessive phenotypes, in order to determine whether the haplotype models should include recessive alleles. In the absence of recessive phenotypes, the Gart-Nam statistic was computed. Significance of the statistic was determined by comparison to the standard normal distribution. The Navajo (STR) and Gila River (HLA) data sets are summarized in table 3, which gives the alleles encountered and their frequencies.

The two data sets employed here illustrate the utility of the method. The Navajo (STR) analysis demonstrates the method with a relatively small sample size ($N = 38$) and with unlinked loci that are unlikely to be in gametic disequilibrium. By contrast, the Gila River (HLA) analysis demonstrates the technique with a large sample size ($N = 619$) and with closely linked loci that are likely to be in gametic disequilibrium. Moreover, the method's utility with systems possessing recessive alleles is demonstrated by the Gila River (HLA) data.

Simulations

The sampling distribution of G is potentially affected by numerous factors, such as sample size, numbers and frequencies of alleles, presence of recessives, and components of disequilibrium. In addition, since the simulation provides an estimate of the sampling distribution, the number of simulated replicate samples can affect the accuracy of the estimation. With these points in mind, it is clear that an exhaustive analysis of all factors and combinations of factors would be tedious, and such an analysis was not performed. However, we did perform some simulations to (1) determine whether sample size was a major factor, (2) determine the importance of the number of replicated samples, and (3) find conditions where the χ^2 approximation works well. Simulations were performed for the global equilibrium null hypothesis (M0) contrasted with the full model (M15), using the characteristics of the Gila River Indian sample. Either 1,000 or 5,000 replicate samples were evaluated using the procedure described in Methods.

Results

Recessive Alleles

Recessive phenotypes were observed among the Gila River HLA-C typings but were absent from the HLA-A and HLA-B typings. The Gart-Nam test revealed strong evidence for a recessive allele in the Gila River HLA-B data, but it failed to detect a recessive allele in the HLA-A typings. Accordingly, haplotype models for the Gila River HLA data included recessive alleles at HLA-B and HLA-C, but not at HLA-A. No recessive phenotypes were seen at the three Navajo STR loci. Moreover, the Gart-Nam test (Gart and Nam 1984) failed to provide any additional evidence

Table 3

Genetic Data Sets

Locus	Allele	Frequency	Tests for Recessive Alleles[a]			
			T	z	P	N
A: Navajo (STR)						
D22S274	A1	.026	.84	−.36	1.000	38
	A2	.026				
	A3	.013				
	A4	.053				
	A5	.289				
	A6	.158				
	A7	.250				
	A8	.184				
D18S57	B2	.079	.75	−.52	1.000	38
	B3	.026				
	B4	.132				
	B5	.329				
	B6	.013				
	B7	.013				
	B9	.132				
	B10	.276				
D20S115	C1	.013	1.42	1.30	.097	38
	C2	.447				
	C3	.158				
	C4	.368				
	C5	.013				
B: Gila River (HLA)						
HLA-A	A2	.561	1.00	−.06	1.000	619
	A24	.342				
	A31	.080				
	AR	.017				
	AX	.000				
HLA-B	B5	.075	1.81	6.36	.000	619
	BN21	.143				
	B27	.099				
	B35	.172				
	B39	.111				
	B40	.015				
	Bw48	.188				
	B51	.056				
	Bw60	.036				
	Bw61	.048				
	BR	.021				
	BX	.034				
HLA-C	Cw2	.098	No test	...	...	619
	Cw3	.221				
	Cw4	.152				
	Cw7	.115				
	Cw8	.170				
	CwR	.002				
	CX	.241				

[a] We use the Gart and Nam (1984) test statistic $T = \Sigma_i 2n_i/(G_i + n_i)$, where $G_i = n_i + \Sigma_{i<j} n_{ij}$, n_i is the number of phenotypes that show only the ith allele, and n_{ij} is the number of heterozygotes for the ith and jth alleles. Denoting the sample size by N and the number of codominant alleles by m, the null hypothesis is tested by the asymptotic normal score $z = (T - 1)N^{1/2}(m - 1)^{1/2}$. P = significance level; all hypothesis tests are one-sided.

Table 4

Navajo and Gila River Disequilibrium Analysis

Sample	H_R[a]	H_Ω[b]	Hypothesis	G[c]	df	P_{χ^2}[d]	P_s[e]
Navajo	M0	M15	Global equilibrium	102.55	301	1.000	.375
Gila River	M0	M15	Global equilibrium	2304.65	315	.000	.000
	M0	M1	B × C equilibrium	1929.37	66	.000	.000
	M0	M2	A × C equilibrium	113.72	18	.000	.000
	M0	M3	A × B equilibrium	177.88	33	.000	.000
	M7	M15	Three-way equilibrium	89.54	198	1.000	.000

[a] Null hypothesis.
[b] Alternative hypothesis (see eq. [4]).
[c] Test statistics.
[d] Statistical P-value estimated from χ^2 distribution.
[e] Statistical P-value estimated from the resampled distribution.

for them. Thus, recessive alleles were excluded from haplotypes at all three Navajo STR loci.

Optimization

Maxima on the Navajo (STR) log-likelihood function and the Gila River (HLA) log-likelihood function were obtained using starting allele frequencies equal to the reciprocal of the number of alleles at each locus, assuming multiple-locus equilibrium (that is, haplotype frequencies were the product of allele frequencies). In order to verify that these were indeed global maxima and to evaluate the sensitivity of the algorithm to starting values, we subjected each data set to 1,000 different sets of initial haplotype frequencies. Each of these sets were generated at random, under the constraint that the haplotype frequencies summed to 1. No constraints were placed on multiple-locus equilibrium, and, in fact, many of the random haplotype-frequency sets had profoundly irregular patterns of disequilibrium. Each set of initial haplotype frequencies for the Navajo (STR) data set generated the same optimization point. Thirty of the 1,000 initial haplotype frequency sets tried for the Gila River (HLA) sample found a slightly lower stable peak with very similar parameter estimates, while the remaining 970 sets of initial values found the slightly higher peak discovered on the first try. Thus, our initial impression that the algorithm is not overly sensitive to initial conditions was confirmed.

Allele Frequencies, Haplotype Frequencies, and Gametic Disequilibrium

We followed the steps of our recommended hypothesis-testing strategy. For the Navajo (STR) data (table 4), global equilibrium was not rejected, and the analysis was terminated. However, the statistical significance level obtained from the theoretical χ^2 ($P = 1.0$) does not agree with the resampling significance level ($P = .375$). For the Gila River (HLA) data (table 4), global equilibrium was rejected by both sampling distributions, thus warranting continuation of testing. Gametic disequilibrium between pairs of loci was tested for by contrasting M1–M3 with M0. The results of these tests were highly significant by both statistical distributions, leading us to test for three-way disequilibrium by comparing M15 with M7. For this last test, the statistical significance levels obtained from the χ^2 and resampling distributions were in extreme opposition. While the χ^2 P-value was 1.00, it was <.001 when resampling was used. In fact, the maximum G simulated under M7 (71.58), was considerably smaller than the observed value (89.54). Since we accept the resampling distribution as being more reliable than the χ^2 approximation, we also accept that the HLA three-way disequilibrium is significant. However, it accounts for a relatively small portion of the departure from equilibrium. The total disequilibrium $G = 2304.65$ was partitioned into 89.54 (4%) for three-way interactions and 2215.11 (96%) for two-way interactions (table 4).

The complete theoretical and empirical sampling distributions are shown for all tests in table 5. There are profound differences in every case. Furthermore, the apparent agreement between the resampling and χ^2 P-values that was seen for several tests is revealed to be an artifact of extreme disequilibrium among the HLA loci. Since it has been claimed that another χ^2 test works well with two multiallelic loci (Weir and Cockerham 1979), we collapsed our data into two-locus subsets and applied the likelihood-ratio test recommended. The χ^2 distribution worked much better with the two-locus data sets (table 5). It is nonetheless interesting that the χ^2 distributions provided more conservative tests than did the resampled distributions, when locus trios were analyzed together.

Simulations

In all cases we simulated the three-locus null distribution as specified by M0 (global equilibrium) and contrasted it with the full model as specified by M15. Our primary objective was to identify conditions where the theoretical χ^2s are good approximations of the empirical distributions.

Table 5

Disequilibrium Sampling Distributions

	Critical Values with Tail Probability of							
	.95	.90	.75	.50	.25	.10	.05	.01
Navajo (STR):								
M15 vs. M0	79.22	82.70	90.40	98.44	106.53	114.69	119.60	128.62
χ^2 (301)	261.81	270.02	284.11	300.33	317.17	332.84	342.46	361.00
Gila River (HLA):								
M15 vs. M0	222.48	229.30	240.87	253.98	268.05	279.62	286.61	302.48
χ^2 (315)	274.89	283.30	297.73	314.34	331.55	347.57	357.40	376.32
M3 vs. M0	13.67	15.89	20.73	26.53	32.42	38.71	41.39	46.18
(Two locus)	22.25	24.99	28.9	33.45	38.98	45.11	47.94	54.16
χ^2 (33)	20.87	23.12	27.22	32.34	38.06	43.75	47.4	54.78
M2 vs. M0	4.92	6.22	9.13	12.74	17.04	21.06	24.17	29.97
(Two locus)	9.37	10.96	13.67	17.29	21.49	26.05	29.14	35.37
χ^2 (18)	9.40	10.87	13.68	17.34	21.61	25.99	28.87	34.81
M1 vs. M0	36.09	40.06	45.54	52.75	60.5	70.61	73.97	83.31
(Two locus)	48.28	51.71	57.2	64.36	72.87	79.7	84.9	95.75
χ^2 (66)	48.31	51.78	57.94	65.34	73.35	81.09	85.97	95.63
M15 vs. M7	23.31	25.44	29.67	35.73	42.30	48.96	53.72	62.10
χ^2 (198)	166.44	172.96	184.24	197.33	211.03	223.89	231.83	247.21

NOTE.—Empirical and theoretical sampling distributions for hypothesis tests on differing levels of gametic disequilibrium are shown. Empirical distributions are grouped with the particular χ^2 that should approximate them. Each comparison of models within a block of rows tests a hypothesis about the same set of parameters. "(Two locus)" indicates that pairwise disequilibrium was tested using an appropriate two-locus model.

Following this, we attempted to identify the factors that caused the theoretical χ^2s to depart from the empirical distributions. A χ^2 test is classified *liberal* if it rejects the null hypothesis more often than does the resampling test. By contrast, a χ^2 test is classified *conservative* if it maintains the null hypothesis more often than does the resampling test. All simulation results are reported in table 6. In simulation experiment 1 we maintained the Gila River Indian sample size (N = 619) and replicated 5,000 three-locus data sets. Each locus had two codominant alleles with equal frequencies and no recessive alleles. The appropriate χ^2 distribution with 4 df provided an excellent approximation. In order to test whether 5,000 replicate samples are necessary to estimate the distribution adequately, the next experiment (2) maintained all of the conditions of experiment 1, except that only 1,000 three-locus data sets were generated.

Table 6

Critical Values of Simulated Sampling Distributions

						Critical Values with Tail Probability of							
Experiment	N[a]	R[b]	n_A-n_B-n_C[c]	Frequency[d]	Recessive?[e]	.950	.900	.750	.500	.250	.100	.050	.010
1	619	5,000	2-2-2	Even	No	.73	1.08	1.91	3.42	5.39	7.81	9.55	13.39
2	619	1,000	2-2-2	Even	No	.68	1.07	2.00	3.35	5.09	7.38	8.86	13.75
3	100	1,000	2-2-2	Even	No	.70	1.09	2.02	3.55	5.60	8.10	9.88	13.86
4	100	1,000	2-2-2	Skew	No	.80	1.14	1.96	3.19	4.62	6.61	7.85	10.96
5	1,000	1,000	2-2-2	Skew	No	.77	1.12	1.99	3.36	5.22	7.55	9.46	13.20
χ^2 (4 df)	...	...	...	...	...	.71	1.06	1.92	3.36	5.39	7.78	9.49	13.28
6	619	1,000	4-12-7	Even	No	380.87	388.12	401.25	416.11	432.75	448.85	456.58	474.43
7	619	1,000	4-12-7	HLA	No	239.10	245.44	257.49	271.61	285.73	297.89	305.21	316.97
8	619	1,000	4-12-7	HLA	Yes	222.48	229.30	240.87	253.98	268.05	279.62	286.61	302.48
χ^2 (315 df)	...	...	...	...	...	274.89	283.30	297.73	314.34	331.55	347.57	357.40	376.32

[a] Size of simulated samples.
[b] Number of replicated samples under the null hypothesis.
[c] Nos. of alleles at the first, second, and third loci, respectively.
[d] "Evenness" of allele frequencies (see text for details.)
[e] Indicates whether recessive alleles are included in haplotype models.

This estimated sampling distribution is still quite close to the theoretical χ^2. Accordingly, 1,000 replications were used in the subsequent simulation experiments.

The purpose of experiment 3 was to observe the effect of sample size (N) on the test statistic's distribution. The conditions of experiment 2 were maintained, except that N was reduced from 619 to 100. Higher values of G were observed more frequently than was predicted by χ^2 distribution. This result is important because the χ^2 approximation is liberal, while previously (table 5) it was very conservative. Simulation experiment 4 maintained the conditions of experiment 3 (i.e., $N = 100$) but substituted uneven allele frequencies (.9 and .1) at each locus. The resulting empirical distribution was shifted left, making the χ^2 approximation conservative. The results of experiments 3 and 4 suggest that the deviation of a true sampling distribution from its theoretical χ^2 will balance the conservative tendency due to uneven allele frequencies and the liberal tendency due to sample size. In order to verify this, simulation experiment 5 maintained the conditions of experiment 4 but increased the sample size to 1,000 individuals. As expected, the theoretical χ^2 closely approximated the empirical distribution.

The next simulation experiments dealt more directly with the conditions of the HLA data set. In simulation experiment 6, the Gila River Indian sample size ($N = 619$) and numbers of alleles at each locus ($n_A = 4$; $n_B = 12$; and $n_C = 7$) were maintained, but all alleles at a locus were assigned the same frequency (e.g., $1/n_A$), and all alleles were codominant. One thousand data sets were simulated under multiple-locus equilibrium. The distribution of the test statistic, under these circumstances was shifted far to the right of the χ^2 distribution (table 6). Thus, the theoretical distribution created a dangerously liberal test. The effect of uneven allele frequencies was assessed in simulation experiment 7. The conditions of experiment 6 were maintained, but the allele frequencies used were those observed in the Gila River (HLA) sample. As a result, the empirical distribution was dramatically shifted leftward. Thus, the χ^2 approximation was made extremely conservative, and the apparent bias in experiment 6 was reversed. Finally, to see the effect of recessive alleles, we simulated under the same conditions as experiment 6, but, following the observed HLA data structure, the last alleles in the HLA-B and HLA-C series were both recessive. The effect of the recessive alleles was to shift the distribution further leftward, thus making the test more conservative.

Discussion

The optima found by our algorithm for the Navajo (STR) and Gila River (HLA) likelihood equations appear to be global maxima. This is an important result because the algorithm is not guaranteed to yield a global maximum. In fact, Weir and Cockerham (1979) have identified data configurations for which another EM algorithm (see Hill 1974) fails to provide the global maximum for a genetic model with two codominant alleles at two loci. Their problems arose when the single-locus marginal totals exhibited substantial deviations from Hardy-Weinberg expectations. Weir and Cockerham (1979) provide direct solutions for their model's likelihood, but their approach is not feasible with multiple (>2) loci or with genetic systems containing recessive alleles. While our results are encouraging, we agree with Weir and Cockerham (1979) that blind applications of iterative procedures should be avoided and that multiple-locus analyses of data with significant deviations from Hardy-Weinberg expectations should be approached cautiously.

The ability to easily test for linkage disequilibrium in highly polymorphic systems with recessive alleles is an important advantage of the methods proposed here. Chakraborty et al. (1994) have demonstrated that failure to include recessives in statistical analyses, when they exist, can falsely reject the test of Brown et al. (1980) for linkage equilibrium. To the extent that this finding applies to other tests for linkage equilibrium, it is advisable to apply a test such as Gart and Nam's (1984) for hidden recessives and to include recessives in analyses whenever there is evidence for their existence. However, it should not be forgotten that genetic substructuring of populations causes a heterozygote deficiency that will spuriously appear as a recessive allele with the Gart-Nam test. Solid knowledge of the genetic systems being tested and the sample collection procedures can provide valuable indications about the likelihood of recessive alleles and population substructure.

The forward-selection testing strategy advocated here is unusual with multiple-factor categorical data analysis. While forward selection is necessary for constructing resampled distributions, *backward-selecting* strategies, in which nonsignificant components are eliminated from the most general model (e.g., Sokal and Rohlf 1981; Agresti 1990), are commonly used in other circumstances. The rationale behind backward selection is that higher-order disequilibrium obscures the meaning of lower-order disequilibrium. For example, three-way disequilibrium (D(ABC)) indicates that the disequilibrium between a pair of alleles at two loci is heterogeneous, depending on the allele present at a third locus (e.g., $D(AB)_{12}$ is positive on C1-bearing chromosomes but negative on C2-bearing chromosomes). However, this effect may be relatively weak. Consider the Gila River HLA example, the total G (2304.61) is partitioned into 2214.87 for disequilibrium between pairs of loci and 89.74 for three-way interaction (table 3). Although both components are statistically significant by resampling, high-order effects account for only about 4% of the total departure from multiple-locus equilibrium. Clearly, most of the deviation from the global null hypothesis is accounted for by pairwise disequilibrium, and analysis of the pairwise effects will serve at least for descrip-

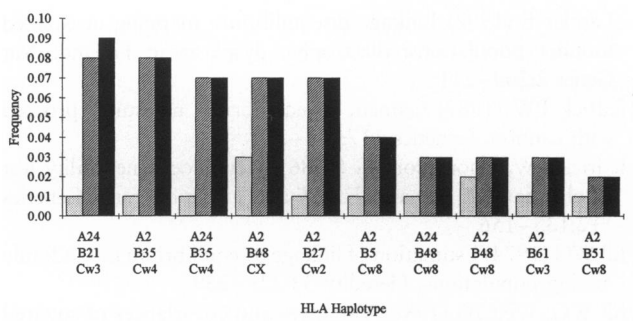

Figure 3 The 10 haplotypes estimated to be most frequent in the Gila River HLA data set. Expectations are given according to M0 (global equilibrium; dotted), M7 (pairwise disequilibrium only; striped), and M15 (full model, blackened).

tive purposes. Visual support for this interpretation is provided by figure 3, in which frequency estimates for the 10 most common haplotypes are plotted for three models. The estimates obtained for the full model (M15) are closely approximated by those obtained for the model allowing only pairwise disequilibrium (M7). In contrast, the global equilibrium model (M0) provides very different estimates.

The test statistic (G) distributions produced by computer resampling reveal conditions under which χ^2 approximations are useful and many conditions under which they will fail. The χ^2s worked well with the two-locus models considered here, but they were inadequate when a third locus was simultaneously considered. The success with two loci should be interpreted cautiously. In fact, Guo and Thompson (1992) have demonstrated that χ^2 distributions can be either too liberal or too conservative when applied to a two-way contingency table test for Hardy-Weinberg proportions. The most predominant factors affecting the χ^2's performance with three loci are sample size (N) and evenness of allele frequencies. Small sample size (N) gives the χ^2 test a liberal tendency (that is, the probability of a type I error is underestimated), but unevenness of allele frequencies renders the χ^2 test very conservative (that is, the probability of a type I error is over estimated). Increasing the number of alleles at the loci exacerbates both of these tendencies. Recessive alleles at some or all of the loci result in a tendency to underestimate α. Application of χ^2 to statistical decisions would have led to extremely conservative interpretations for both data sets analyzed here, but this may not be universal. Accordingly, we recommend that χ^2 be used only for preliminary screening of results and that formal acceptance or rejection of hypotheses be determined by the resampled distributions.

The haplotype-estimation algorithm- and hypothesis-testing strategy provided in this paper will enable detailed analyses of gametic disequilibrium. It is well known that pairwise disequilibrium coefficients (e.g., $D(AB)_{ab}$, $D(AC)_{ac}$, and $D(BC)_{bc}$) are dependent on allele frequencies, and different standardized measures have been recommended in order remove this dependence (Hedrick 1987; Weir 1990). Since the decision whether to standardize and the choice between different standardization methods depend on population-genetic and statistical assumptions appropriate to specific biological questions, we do not favor one method over another. Rather, we wish to note that all such standardizations utilize functions of the basic quantities that are estimated here. Thus, the analyses that we propose can serve as a springboard to any more extensive analysis of gametic disequilibrium. It should be noted also that, whether or not high-order disequilibrium coefficients are interesting in their own right, they are fundamental for computing the covariances between pairwise disequilibrium coefficients (Hill and Weir 1988; Weir 1990). Thus, analysis and interpretation of pairwise gametic disequilibrium when more than two loci are considered necessitates computing high-order disequilibrium coefficients.

The value of highly polymorphic loci for linkage analysis, human identification, and evolutionary genetics has been demonstrated in a number of recent publications (e.g., Ott 1992; Bowcock et al. 1994; Gill et al. 1994). These works underscore the necessity of good statistical methods and computer algorithms for such systems. In addition, we will presently review two areas of broad interest that will provide applications for our method. First, direct sequencing from genomic DNA provides ambiguities when an individual is heterozygous at two or more nucleotide positions. By treating each position as a polymorphic locus, our approach provides complete sequences of multiple alleles when population-based samples (as opposed to family-based samples) are used. Another, recently proposed method to accomplish this (Clark 1990) capitalizes on much the same information as does our method: haplotypes observed in homozygotes or single-locus heterozygotes are used to resolve multiple heterozygotes. This alternative method does not, however, provide frequency estimates for the haplotypes observed. Second, there has been recent interest in using linkage disequilibrium for gene mapping (Chakravarti et al. 1984; Jorde et al. 1993; Hill and Weir 1994). The basic principle underlying this approach is that chromosomes bearing disease-causing genes that are descended from a common mutation should show the ancestral haplotype in the vicinity of the disease gene. The extent to which the ancestral haplotype is preserved reflects recombination events occurring over the entire history of the population (Hästbacka et al. 1992). While the efficacy of the method has been questioned (Kaplan and Weir 1992; Hill and Weir 1994), future assessments of the method can only be improved by efficient methods for detecting haplotypes in multiple highly polymorphic genetic systems.

Acknowledgments

This work was supported in part by NSF grant BNS-9108422. We thank David Goldman and Raymond Peterson for their com-

ments on this paper. An earlier version of this paper was improved by suggestions offered by three anonymous reviewers.

References

Agresti A (1990) Categorical data analysis. John Wiley and Sons, New York

Baur MP, Danilovs JA (1980) Population analysis of HLA-A,B,C,DR and other genetic markers. In: Terasaki PI (ed) Histocompatibility testing 1980. UCLA, Los Angeles

Bennett JH (1954) On the theory of random mating. Ann Eugen 18:311–317

Bowcock AM, Ruiz-Linares A, Tomforhrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum.* Genetics 96:523–536

Ceppellini R, Siniscalco M, Smith CAB (1955) The estimation of gene frequencies in a random mating population. Ann Hum Genet 20:97–115

Chakraborty R, Zhong Y, Jin L, Budowle B (1994) Nondetectability of restriction fragments and independence of DNA fragment sizes within and between loci in RFLP typing of DNA. Am J Hum Genet 55:391–401

Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination within the human β-globin gene cluster. Am J Hum Genet 36:1239–1258

Clark AG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol 2:111–122

Deming WE, Stephan FF (1940) On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann Math Stat 11:427–444

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc B 39:1–38

Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman and Hall, New York

Gart JJ, Nam JA (1984) A score test for the possible presence of recessive alleles in generalized ABO-like genetic systems. Biometrics 40:887–894

Gill P, Ivanov P, Kimpton C, Piercy R, Benson N, Tully G, Evett I, et al (1994) Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6:130–135

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, et al (1994) The 1993–94 Généthon human genetic linkage map. Nat Genet 7:246–349

Hartl DL, Clark AG (1989) Principles of population genetics, 2d ed. Sinauer, Sunderland, MA

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 2:204–211

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331–341

Hedrick PW, Thompsom G (1986) A two-locus neutrality test Applications to humans, *E. coli* and *lodgepole* pine. Genetics 112:135–156

Hill WG (1974) Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229–239

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

——— (1994) Maximum-Likelihood estimation of gene location by linkage disequilibrium. Am J Hum Genet 54:705–714

Jorde LB, Watkins WS, Viskochil D, O'Connell P, Ward K (1993) Linkage disequilibrium in the neurofibromatosis I (NFI) region: implications for gene mapping. Am J Hum Genet 53:1038–1050

Kaplan N, Weir BS (1992) Expected behavior of conditional linkage disequilibrium. Am J Hum Genet 51:333–343

Michelini S, Urbanek M, Dean M, Goldman D. Polymorphism and genetic mapping of the human oxytocin receptor gene on chromosome 3. Neuropsychiatr Genet (in press)

Nam J, Gart JJ (1987) On two tests of fit for HLA data with no double blanks. Am J Hum Genet 41:70–76

Ott J (1977) Counting methods (EM algorithm) in human pedigree analysis: linkage and segregation analysis. Ann Hum Genet 40:443–454

——— (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. Am J Hum Genet 51:283–290

Piazza A (1975) Haplotypes and linkage disequilibria from three-locus phenotypes. In: Kissmeyer-Nielsen F (ed) Histocompatibility testing 1975. Munksgaard, Copenhagen, pp 923–927

Smith CAB (1957) Counting methods in genetical statistics. Ann Hum Genet 21:254–276

Sokal RR, Rohlf FJ (1981) Biometry. Freeman, San Francisco

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Weir BS, Cockerham CC (1979) Estimation of linkage disequilibrium in randomly mating populations. Heredity 42:105–111

Weir BS (1990) Genetic Data Analysis. Sinauer, Sunderland, MA

——— (1992) Independence of VNTR alleles defined as fixed bins. Genetics 130:873–887

Weir BS, Cockerham CC (1978) Testing hypotheses about linkage disequilibrium with multiple alleles. Genetics 88:633–642

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, et al (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Williams RC, McCauley J (1992) HLA class I variation controlled for genetic admixture in the Gila River Indian Community of Arizona: a model for the Paleo-Indians. Hum Immunol 33:39–46

Yasuda N, Kimura M (1968) A gene-counting method of maximum likelihood for estimating gene frequencies in ABO and ABO-like systems. Ann Hum Genet 13:409–420