Am J Hum Genet 38:918–939, 1986

The Use of Restriction Fragment Length Polymorphisms in Paternity Analysis

PETER E. SMOUSE[1] AND RANAJIT CHAKRABORTY[2]

SUMMARY

This paper examines the utility of restriction fragment length polymorphisms (RFLPs) for paternity analysis. While, on the average, 99% of falsely accused males can be excluded with the standard battery of blood group antigens, red cell enzymes, serum proteins, and HLA antigens, there are still mother-child pairs for whom the exclusion probability is not high. It has been suggested that additional resolution would be available with RFLPs. We have examined the strategic aspects of using RFLPs for paternity analysis, comparing the efficacy and cost of a multimarker haplotypic set with those of a comparable set of unlinked RFLPs, using published frequencies for the β-globin complex, the serum albumin region, and the growth hormone region. There are four major findings. (1) Greater resolution is obtained with a carefully chosen set of tightly linked RFLPs producing chromosomal haplotypes than with a comparable set (same allele frequencies) of unlinked markers, but only if it is possible to establish linkage phase unambiguously. (2) Assay of linked sets is cheaper than is the assay of unlinked markers, but the cost advantage is optimized with sets of no more than two or three linked markers. (3) Also, with more than two or three tightly linked markers, the haplotypic frequencies are too poorly estimated to provide a reliable measure of the probability of paternity for unexcluded males, given the sample sizes likely to be available in the near future. (4) Optimal resolution, minimal cost, and acceptable accuracy are obtained with several independent sets of no more than two or three tightly linked RFLP markers each. With current technology, RFLP analysis is more expensive for the same level

Received August 7, 1985; revised November 19, 1985.

This study was supported by grant 86-ER60089 from the Department of Energy (P. E. S.) and grant R01-GM-20293 from the National Institutes of Health (R. C.).

[1] Department of Human Genetics and Division of Biological Sciences, University of Michigan, Ann Arbor, MI 48109.

[2] Center for Demographic and Population Genetics, University of Texas Health Science Center, Houston, TX 77225.

of genetic resolution than is the standard battery, but gradual replacement of the latter can be anticipated as economies of scale reduce the cost of the DNA technology.

INTRODUCTION

It is now possible, by using information on blood group antigens, red cell enzymes, serum proteins, and HLA antigens, to exclude over 99% of falsely accused men in paternity cases in the United States [1–3]. There nevertheless remain a considerable number of mother-child pairs with genotypic combinations yielding low exclusion probabilities for falsely accused men and low probabilities of paternity for those not excluded [4]. The only way to improve genetic resolution for these refractory cases is to assay additional polymorphic loci, and the availability of additional protein polymorphisms has become problematic.

It has recently been suggested [5, 6] that an additional battery of markers is available for paternity analysis, namely, a set of DNA restriction fragment length polymorphisms (RFLPs). There seems to be no dearth of these polymorphisms in man [7], and given enough of them, one should (in principle) be able to achieve any desired level of resolution for paternity analysis. There are some real opportunities with RFLPs, but there are also some limitations, and our purpose here is to examine both. Our objective is to determine the best way to use RFLPs in paternity analysis, and we shall consider the following features of the analysis.

(1) It is well known that the additional resolution decreases with each successive polymorphic marker added to the battery; the rate of information drop-off depends critically on the allele frequencies of the genetic markers in question. How fast is this information drop-off with unlinked RFLPs of the sort currently being reported?

(2) The cost of assay for RFLPs is high, compared with that of the standard battery, but there are strategies available that reduce the cost. One of these is to look at tightly linked markers along a single sector of DNA, utilizing several restriction enzymes with a single probe. What are the relative costs of adding a tightly linked marker, relative to that of adding an unlinked marker?

(3) By using sets of tightly linked RFLP markers, we can deploy a powerful haplotypic analysis, but resolution is reduced if linkage phase is ambiguous. Is the extra cost and effort needed to recover that linkage phase information justified, or would we be better off with a carefully chosen set of unlinked markers?

(4) All of our probability calculations are critically dependent on reliable population estimates of allele and haplotype frequencies, but RFLP sample sizes will remain small for some time to come. Can we correct for or avoid the resulting biases in our estimates?

THE BACKGROUND THEORY

We begin with a brief recapitulation of the standard formulation, in an effort to set the notation and partition the problem into more manageable components. We begin with a trio of individuals: a child (C), a mother (M), and a putative father (F). It is usually assumed that maternity is established beyond question.

The argument is usually posed in terms of the paternity index, PI, defined as the posterior odds in favor of paternity. One might impose a prior as well, but this is usually avoided. If we denote the observed RFLP phenotypes of the putative father, mother, and child by (F = f), (M = m), and (C = c), respectively, then the odds in favor of paternity (over nonpaternity) are given by the index (λ):

$$\lambda = \frac{\Pr(F = f, M = m, C = c \mid \text{Paternity})}{\Pr(F = f, M = m, C = c \mid \text{Nonpaternity})} = \frac{X}{Y} . \tag{1a}$$

All of the genetic evidence is contained in X and Y. The term $X = \Pr(F = f, M = m, C = c \mid \text{Paternity})$, given Hardy-Weinberg population frequencies and random mating with respect to the genetic markers under consideration, is $\Pr(F = f) \cdot \Pr(M = m) \cdot \Pr(C = c \mid M = m, F = f)$, where $\Pr(F = f)$ and $\Pr(M = m)$ depend only on population frequencies and $\Pr(C = c \mid M = m, F = f)$ depends only on segregation probabilities. The term $Y = \Pr(F = f, M = m, C = c \mid \text{Nonpaternity}) = \Pr(F = f) \cdot [\Pr(M = m) \cdot \Pr(C = c \mid M = m)]$ with the same population level assumptions. The first term of the right-hand side of Y is the probability that the man (falsely) indicated as the father is of phenotype (f). The term in brackets is the probability that the woman will be of phenotype (m) and that she will have a child of phenotype (c). Whenever phenotypes are genetically *unambiguous*, as will generally be the case with RFLP data, where there are no recessive (blank) alleles, $\Pr(F = f) \cdot \Pr(M = m)$ cancels from equation (1a), yielding:

$$\lambda = \frac{\Pr(C = c \mid M = m, F = f)}{\Pr(C = c \mid M = m)} . \tag{1b}$$

If loci are segregating independently, then both X and Y can be partitioned locus by locus to yield:

$$X = \prod_{l=1}^{L} X_l \quad \text{and} \quad Y = \prod_{l=1}^{L} Y_l , \tag{2}$$

which reduces equation (1a) to the form:

$$\lambda = \prod_{l=1}^{L} \frac{X_l}{Y_l} . \tag{1c}$$

Given a two-allele codominant locus, with the frequency of the (+) allele denoted as P_l and the frequency of the (−) allele denoted as $Q_l = 1 - P_l$, the exclusion probability (E_l for an average trio is known to be [8, 9]:

$$E_l = P_l Q_l [1 - P_l Q_l] \ , \tag{3}$$

and that for the full set of L unlinked and independent loci:

$$E(\text{L loci}) = 1 - \prod_{l=1}^{L} (1 - E_l) \ . \tag{4}$$

The relationship between the probability of exclusion (E) and the (λ) ratio for each locus or for a combined set of loci is also worth noting. Although this dependence has also been alluded to by other authors [10, 11], we will follow Chakraborty and Ferrell [12], since their table 1 explicitly gives the relevant distributions from which the relationship may be computed.

It is easy to show that for nonexcluded non-fathers of random genotype, the relationship is given by:

$$E\left[\frac{X_l}{Y_l}\right]_{NF} = \frac{1}{1 - E_l} \ , \tag{5}$$

irrespective of the number of or dominance relations among alleles at the locus.

From equations (2) and (4), it is possible to construct a relationship between E(L loci) and E[X(L loci)/Y(L loci)], since the distribution of (X_l/Y_l) is independent across unlinked loci:

$$E\left[\frac{X(\text{L loci})}{Y(\text{L loci})}\right]_{NF} = \prod_{l=1}^{L} E\left[\frac{X_l}{Y_l}\right]_{NF} = \prod_{l=1}^{L} (1 - E_l)^{-1} = \frac{1}{1 - E(\text{L loci})} \ , \tag{6a}$$

more conveniently presented in log-linear form as:

$$\log\left\{E\left[\frac{X(\text{L loci})}{Y(\text{L loci})}\right]\right\}_{NF} = -\log[1 - E(\text{L loci})] = \epsilon(\text{L loci}) \ , \tag{6b}$$

from which we assert that the average probability of paternity for nonexcluded non-fathers increases with the average exclusion probability; it follows that $\log[\bar{\lambda}_{TF}] \geq \log[\bar{\lambda}_{NF}]$. It is convenient to use the logarithmic form, ϵ(L loci), for most of what follows.

UNLINKED LOCI

Two-Allele Loci

We present in table 1 a sample of RFLPs that have already been described whose frequencies are reasonably well estimated, along with their E_I-values. We have chosen sets of tightly linked markers from three particular chromosomal regions to facilitate some of the comparisons that follow, but will treat these markers as though they were unlinked and independent for present expository purposes. The β-globin data are those presented for the β^A-bearing chromosomes in Greeks and Italians [13], pooled to raise the sample size to number (N) = 65. The serum albumin frequencies are those reported for Caucasians [14], N = 110. The growth hormone data are those for northern Europeans and Mediterraneans [15], again pooled to increase the sample size to N = 46. These sample sizes are all rather small (although large by usual RFLP haplotype criteria), and very much larger sample sizes are needed before RFLP data are adequate to support the paternity testing enterprise. We will have

TABLE 1

A LISTING OF RFLPS AVAILABLE FROM THREE SETS OF LINKED MARKERS: (a) THE β-GLOBIN COMPLEX, (b) THE SERUM ALBUMIN REGION, AND (c) THE GROWTH HORMONE REGION, WITH THE FREQUENCIES OF THE (+) STATE AND INDIVIDUAL EXCLUSION PROBABILITIES (E)

RFLP marker	Allele frequency (%)	Exclusion probability
(a) β-Globin region:		
*Hin*dIII-$^{G}\gamma$	44.6	.186
*Hin*dIII-$^{A}\gamma$	20.0	.134
*Hinc*III-ψβ1	27.7	.160
*Hinc*III-3′ψβ1	35.4	.176
*Hin*fI-5′β	95.4	.042
*Hgi*AI-β	81.5	.128
*Ava*II-β	81.5	.128
*Bam*HI-3′β	78.5	.140
*Rsa*I-3′β	24.6	.151
(b) Serum albumin region:		
*Pst*Ia	46.4	.187
*Msp*I	3.6	.034
*Pst*Ib	96.4	.034
*Hae*IIIa	53.6	.187
*Hae*IIIb	50.0	.188
*Sac*I	50.0	.188
*Hae*IIIc	53.6	.187
*Eco*RV	13.6	.104
(c) Growth hormone region:		
*Hinc*II	32.6	.171
*Msp*Ia	52.2	.187
*Msp*Ib	47.8	.187
*Bgl*IIa	69.6	.167

NOTE: (a) from [13], (b) from [14], (c) from [15]. Markers are presented in map order 5′ → 3′.

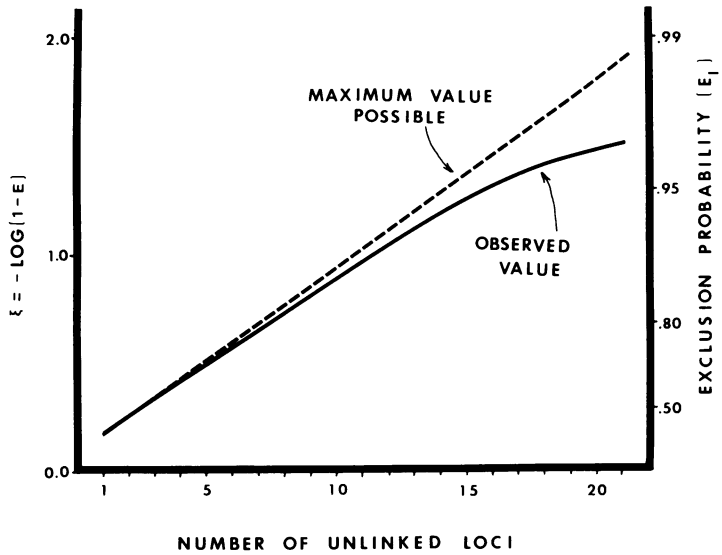

FIG. 1.—The impact of increasing nos. unlinked, two-allele loci on paternity resolution, where the loci are those of table 1, added in decreasing order of marginal resolution: a plot of $\epsilon = -\log[1 - E]$ against the number of loci.

more to say about sample sizes under STATISTICAL CONSIDERATIONS, but for now will treat these frequency estimates as though they were parametric (known without error). There is no reason why we could not add other markers, but the point of the exercise is illustration rather than exhaustive enumeration. We assume that the markers presented are a reasonable sample of the spectrum of unlinked RFLPs that might be available in the near future.

Adding unlinked RFLPs to the battery will improve both the overall exclusion rate and the average probability of paternity for nonexcluded males. Assuming that RFLPs will be successively added in decreasing order of their E_I-values, we should expect the relationship portrayed in figure 1, where ϵ(L loci) is plotted against the number of RFLPs employed. We also present the maximum value possible for ϵ(L loci), realizable only if all loci have ($P_I = \frac{1}{2} = Q_I$), as a convenient frame of reference. As the number of loci increases, the realized value of ϵ(L loci) drops progressively below this maximum figure. Careful choice of loci will minimize these losses of information. It is important to remember, however, that even the maximum attainable values of ϵ (linear in the number of loci) represent diminishing returns on exclusion probability. To improve our resolution from E = .1875 to E = .3398 requires the addition of a single two-allele locus; to improve from E = .3398 to E = .5642 requires the addition of two additional loci, and so on; to improve from E = .9896 to E = .9991 requires 12 additional loci.

Multiple-Allele Loci

Resolution can be improved if there are more than two alleles per locus. Given a set of K alleles whose frequencies are denoted $P_k: k = 1, \ldots, K$, the

exclusion probability is given by [6]:

$$E(K \text{ alleles}) = a_1 - 2a_2 + a_3 + 3(a_2a_3 - a_5) - 2(a_2^2 - a_4) \quad (7a)$$

where:

$$a_i = \sum_{k=1}^{K} P_k^i \ . \quad (8)$$

The maximum exclusion probability is achieved where the alleles are all equally frequent, that is, with $P_k = K^{-1}$ for all k. In this case, it is known [16] that the exclusion probability E(K alleles) is given by:

$$E_H(K \text{ alleles}) = \frac{(K-1)(K^3 - K^2 - 2K + 3)}{K^4} \ . \quad (7b)$$

There are, as yet, only a few multiple-allelic systems of RFLP markers, some involving small rearrangements of length, such as the *Eco*RI alleles of the DS14S1 marker [17] and the *Bgl*I alleles of the insulin region [18], others involving variable numbers of small tandem repeat sequences, such as the *Sac*I alleles of the ζ-globin gene complex [19] and the minisatellite alleles described for several unnamed loci by Jeffries et al. [20]. None of these systems has yet been characterized well enough for unambiguous allelic resolution, and frequency data are too limited to justify formal analysis at this time, but these sorts of multiple-allelic systems hold future promise for paternity analysis.

TIGHTLY LINKED LOCI

The use of unlinked RFLPs requires that one employ multiple DNA probes and multiple restriction enzymes. The realities of Southern gel analysis are such that it is attractive, given a probe-enzyme combination that identifies an RFLP, to add additional restriction enzymes to the assay, rather than deploying a new probe for an unlinked marker. This strategy of identifying a set of tightly linked RFLPs has some distinct advantages, but it also has disadvantages, to both of which we now turn.

Linkage Phase Known

If we can unambiguously determine linkage phase for a set of closely linked genetic markers, we can treat a set of linked markers as a haplotype, which reduces the problem to that of multiple codominant alleles. Considering just a single probe and a set of closely linked markers, each assayed with a different restriction enzyme, denote the available haplotypes as h_k and their population frequencies as P_k: $k = 1, \ldots, K = 2^L$. Then, the results of the multiple-allele section can be used without alteration.

If the haplotypes are all equi-frequent, then E_H(L loci) $\geq E_I$(L loci), where E_I(L loci) denotes the exclusion probability of a set of independent (unlinked)

TABLE 2

HAPLOTYPES (5′ → 3′), THEIR FREQUENCIES AND AVERAGE EXCLUSION PROBABILITIES (E_H) FOR THREE LINKED SYSTEMS OF RFLP MARKERS: (a) THE β-GLOBIN REGION, (b) THE SERUM ALBUMIN REGION, AND (c) THE GROWTH HORMONE REGION; THE AVERAGE EXCLUSION PROBABILITIES FOR THE SAME SET OF MARKERS, TREATED AS INDEPENDENT (E_I), ARE ALSO PRESENTED, USING THE ALLELE FREQUENCIES DRAWN FROM TABLE 1

REGION AND HAPLOTYPE	HAPLOTYPE FREQUENCY (%)	REGION AND HAPLOTYPE	HAPLOTYPE FREQUENCY (%)
(a) β-Globin region: (N = 65)		β-Globin region (continued)	
− − − − + + + + −	29.2	+ − − − + + + + −	1.5
− − − − + − − + −	12.3	+ − + + + + + − −	1.5
+ − + + + + + + −	12.3	− − − − + + + + +	1.5
− − − − + + + − +	7.7	+ + − − + + + + −	1.5
+ + + − + + + + −	6.2	+ − + + + + + + +	1.5
+ + − + + + + + +	3.1	+ − + + − + + − +	1.5
+ + − + + + + − +	3.1	+ − − + + + + + −	1.5
− − − − + + + − −	3.1	+ + − − + − − + −	1.5
+ − + + + − − + −	3.1	+ + − + − + + + +	1.5
+ + − + + − − + −	1.5	+ + − + − + + − +	1.5
− − − + + + + − +	1.5	+ − + + + + + − +	1.5
	[E_H = .751 > E_I = .743]		
(b) Serum albumin region: (N = 110)		(c) Growth hormone region: (N = 46)	
− − + + − − + −	49.1	− − − +	43.5
+ − + − + + − −	31.8	+ + + −	28.3
+ − + − + + − +	13.6	− + + +	13.0
− + − + + + + −	3.6	− + − +	6.5
+ − + + − − + −	0.9	+ + + +	2.2
− − + − + + − −	0.9	+ − − +	2.2
[E_H = .374 < E_I = .703]		− + + −	2.2
		− − + +	2.2
		[E_H = .479 < E_I = .544]	

NOTE: (a) from [13], (b) from [14], (c) from [15].

loci with the same constellation of *allele* frequencies ($P_l = \frac{1}{2} = Q_l$) exhibited by the individual RFLP markers. In general, however, the direction of the inequality depends intimately on the haplotype frequencies themselves, and it is *not* always the case that E_H(L loci) ≥ E_I(L loci).

To illustrate the power of the haplotype approach, we use three examples drawn from the literature: (1) the β-globin complex [13], (2) the serum albumin region [14], and (3) the growth hormone region [15]. The haplotype frequencies for these three systems are presented in table 2, along with the value of E_H(L loci) for each system. Recall that we used these same haplotypes to generate the allele frequencies in table 1, and we have already presented the exclusion probabilities for individual markers. To facilitate the comparison of unlinked and haplotypic *sets* of markers, we also present the value of E_I(L loci) computed from equation (4) for the same RFLP markers (table 1), treated as though they were unlinked and independent.

It is evident that while the β-globin haplotype system is slightly more informative than a corresponding set of unlinked markers, this is not the case for either the serum albumin or growth hormone systems. It is possible, however, that by judicious choice of some subset of these serum albumin and growth hormone markers, we could reverse the inequalities. The idea is to choose the markers in sequential fashion in such a fashion as to maximize the evenness of the haplotype frequency distribution at each step. (Maximum resolution is achieved if all haplotypes are equally frequent.) This is done by starting with the best single marker for any linked system, for example, the *Hae*IIIb or *Sac*I markers of the serum albumin region. One then tries various choices of second marker, computing E_H(2-loci) for each in combination with the first marker. Having chosen that second marker which maximizes E_H(2-loci), one then tries various choices of third marker, computing E_H(3-loci) for each in combination with the first two, and so on. This stepwise algorithm can be shown to be the optimal search procedure. We present in figure 2 the increasing values of ϵ_H(L loci) available from adding successive markers to the haplotype and compare the pattern with that available from treating the same markers as though they were independent. The results presented in figure 2 show that the use of haplotypes is efficacious for the β-globin complex but not for the serum albumin or growth hormone regions. Some care is necessary in the choice of which haplotype systems to deploy. Even with a useful haplotype system, such as that for the β-globin complex, the frequency array of available haplotypes is such that it is not worthwhile to add more than just a few RFLP markers. Given more than two or three linked markers, it is usually better to add an unlinked locus.

Linkage Phase Unknown

If linkage phase is unknown, then genotypes become ambiguous, and E(L loci) is reduced. Chakravarti and Li [6] and Chakraborty and Hedrick [21] examined the loss of information due to linkage phase ambiguity for a pair of linked markers. A pair of inequalities conveys the essential feature of the situation. Denote the average exclusion probability for a pair of loci as E_A(2-loci) if they are absolutely linked and *ambiguous* as to linkage phase, as E_H(2-loci) if they are absolutely linked and *unambiguous* as to linkage phase, and as E_I(2-loci) if they are unlinked. With the same allele frequencies and any array of haplotype frequencies derived from L = 2 loci $\{P_k : k = 1, \ldots, 4 = 2^2\}$, we have:

$$E_H(\text{2-loci}) \geq E_A(\text{2-loci}) \leq E_I(\text{2-loci}) \ . \tag{9a}$$

The extension to L loci would be extremely difficult to solve in closed form, but the directions of the inequalities would be maintained:

$$E_H(\text{L loci}) \geq E_A(\text{L loci}) \leq E_I(\text{L loci}) \ . \tag{9b}$$

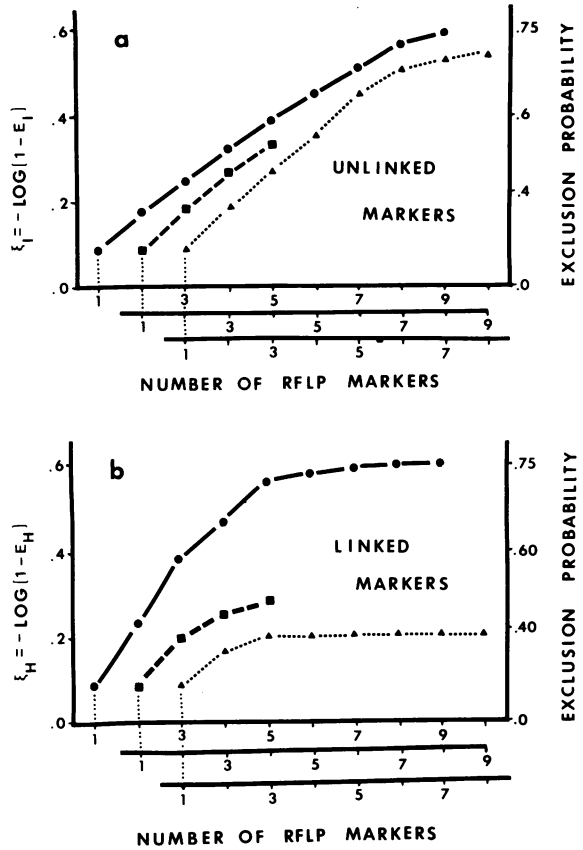

FIG. 2.—A plot of ϵ_H(L loci) with three sets of linked RFLP markers (linkage phase known), compared with ϵ_I(L loci) from a corresponding set of unlinked markers: (●) the β-globin complex, (■) the serum albumin region, and (▲) the growth hormone region. The markers are added in decreasing order of additional information.

Solely on the basis of genetic resolution, there is no reason to choose closely linked (but phase-ambiguous) loci over a comparable set of unlinked loci. The loss of resolution due to phase ambiguity increases as we add linked loci to the battery. The procedures whereby linkage phase is determined are described below. Suffice it here that it is better to use unlinked loci than tightly linked but phase-ambiguous loci.

The Determination of Linkage Phase

There are at least two basic ways in which linkage phase information can be extracted from a set of closely linked RFLPs: (1) the characterization of relatives, and (2) the use of double-digest techniques. Given the grandparents of the child of interest or a set of siblings or both, one can often determine the linkage phase of a set of linked markers [22–24]. This is a common analytic strategy in the analysis of HLA haplotypes [25, 26] and is also useful for Rh,

Gm, and MNS haplotypes, but there are three limitations. (1) We are forced to the assumption that all of the auxiliary individuals are related to the child of interest in the manner specified, that is, there is no other nonpaternity in the family. (2) Not all haplotypic ambiguity can be removed with the use of relatives, and the fraction of refractory cases increases with the number of linked markers. This is evident in many reports of RFLP haplotypes and is an unavoidable limitation. (3) The number of individuals who must be assayed for genotype rises with this approach, thus increasing the cost.

The other method that is useful for the elucidation of RFLP haplotypes is the double-digest procedure so often used in mapping them. One may digest DNA spanning two tightly linked RFLP sites with a pair of restriction enzymes and can distinguish between coupling $(++/--)$ and repulsion $(+-/-+)$ double heterozygotes, thus establishing linkage phase. There are some limitations of the procedure that restrict its applicability. The assay requires extra lanes on a given gel, one extra lane for each pair of heterozygous RFLPs requiring phase resolution. All double-digest fragments must be distinguishable with a single probe in a single lane, and so on. While there are limitations, however, double-digest procedures should prove quite helpful for haplotype delineation.

There is one promising "wrinkle" on the double-digest idea that merits mention. Note from table 1 that there are two *Msp*I sites in the growth hormone region. With a single enzyme identifying more than one site, a certain amount of haplotype information can be obtained from the basic assay, with little or no extra effort. This strategy has been used to good effect in the elucidation of multipoint haplotypes for the μ-switch gene [27] and the $^{C}\gamma$-gene [28]. Note that while there are also two *Hin*dIII sites and two *Hin*cIII sites in the β-globin region, each is assayed with a different probe, so that no free haplotype information is available. Similar comments apply to the two *Pst*I sites and three *Hae*III sites of the serum albumin region. The use of some carefully constructed combination probes would obviously be advantageous for haplotype analysis.

COST CONSIDERATIONS

Quite apart from the matter of information return on the effort, we need to consider the cost of adding each successive RFLP. To add an unlinked marker, we need to deploy an additional DNA probe and usually an additional restriction enzyme; to add a tightly linked marker, it often suffices to add an additional enzyme, as, for example, with the growth hormone and serum albumin systems. Given the technical realities of Southern gel analysis, it is always cheaper to add an additional enzyme. Define the cost/trio of adding another probe as γ_p and that of adding another enzyme as γ_e, with $\gamma_p = \psi \cdot \gamma_e$ $(\psi \geq 1)$. Then, the incremental cost of adding an unlinked marker to the battery is:

$$\gamma_I = \gamma_p + \gamma_e = (\psi + 1) \cdot \gamma_e \ , \tag{10a}$$

and that of adding a tightly linked marker is:

$$\gamma_H = \gamma_e \ . \tag{10b}$$

The total cost of L markers, all unlinked, is thus:

$$\Gamma_I = L \cdot \gamma_I = L(\psi + 1) \cdot \gamma_e \ , \tag{11a}$$

while that of a set of tightly linked markers is:

$$\Gamma_H = (L + \psi) \cdot \gamma_e \ . \tag{11b}$$

The total cost of a set of L markers in $G < L$ tightly linked sets is given by:

$$\Gamma_T = G \cdot \gamma_I + (L - G) \cdot \gamma_H = (L + G \cdot \psi) \cdot \gamma_e \ . \tag{12}$$

Given an arbitrary enzyme cost/trio (γ_e), the values of G and ψ determine the distribution of cost. Relative to the cost of a complete set of tightly linked markers, the cost of a complete set of unlinked markers is $L(\psi + 1)/(L + \psi)$. The cost disadvantage of unlinked markers can be substantial if $\psi \gg 1$. It might well be advantageous to use a tightly linked set of markers even in the case where E_I(L loci) > E_H(L loci). The tradeoffs between cost and information return are best assessed with a *utility* function (U), defined as:

$$U_I = \frac{\epsilon_I(\text{L loci})}{\Gamma_I(\text{L loci})} \quad \text{and} \quad U_H = \frac{\epsilon_H(\text{L loci})}{\Gamma_H(\text{L loci})} \ . \tag{13}$$

We plot the utility criteria in figure 3 for the β-globin system, where the markers are added in the same order as for figure 2. The results are plotted for $\psi = 1$, 2, and 5, using the cost figures for a maximally useful set of unlinked markers ($P_l = \frac{1}{2} = Q_l$) as a convenient reference base. Clearly, as ψ increases, tightly linked markers become ever more attractive on a per unit cost basis. As the redundancy of additional tightly linked markers increases, however, the marginal utility of an additional marker declines, and one does best with a *small* number of tightly linked RFLPs, three to five for the genetic systems and (probe:enzyme) cost ratios (ψ) presented here.

For purposes of illustration, we assumed in figure 3 that haplotypes were detectable without additional cost. In practice, of course, either familial analysis or double-digest procedures are required, as per the previous section. For family analysis, one needs as many as four grandparents and as many sibs as can be obtained, probably an extra six people on the average. That essentially triples the laboratory cost per trio, a change that lowers the haplotypic utility curves in figure 3 considerably, although it changes neither the shapes of the curves nor any strategic decision criteria. It is difficult to specify in advance exactly how the average cost of assay per trio will change with double-digest techniques, but it will lower the haplotypic utility curves in figure 3 and will shift the peaks toward lower numbers of tightly linked markers, since phase ambiguity (and, hence, the extra cost of assay) increases with the number of markers.

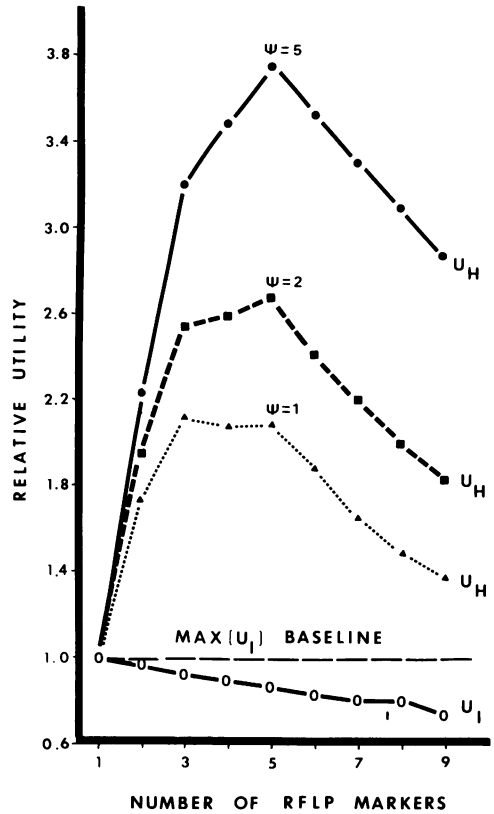

FIG. 3.—A comparison of U_H(L loci) and U_I(L loci) for the β-globin complex, using different values of the probe:enzyme cost ratio (ψ), plotted against the number of RFLP markers in decreasing order of information. All nos. are scaled relative to the maximum value of U_I(L loci) achievable.

STATISTICAL CONSIDERATIONS

Sampling

The accuracy of both X and Y are dependent on estimates of the population frequencies of the genetic markers under consideration. This fact is commonly ignored on the grounds that the sample sizes on which the estimates are based are so large that we may treat these frequencies as known [29]. For HLA haplotypes, the assumption is questionable in some instances, but it is routinely made in any case [30–34]. For DNA markers, and particularly for haplotypes, sample sizes will remain small for some time to come, and the problems of estimation will not be trivial. There are two features of interest: bias of estimation and sampling variation. We deal here only with bias, leaving considerations of sampling variation for a later communication.

We present the treatment for two-allele, codominant markers, because the treatment for multiple alleles or haplotypes is a straight-forward extension. The (X:Y) ratio assumes one of five different values for any particular marker,

depending on the genotypes of the trio:

$$\frac{1}{P}, \frac{1}{2P}, 1, \frac{1}{2Q}, \frac{1}{Q} . \tag{14}$$

In a sample of N alleles used to establish P and Q let (r) be the number of (+) types recovered and ($N - r$) the number of (−) types recovered. The usual, unbiased estimates of P and Q are given by:

$$\hat{P} = \frac{r}{N} \quad \text{and} \quad \hat{Q} = \frac{N - r}{N} . \tag{15}$$

Since there is a finite probability that either r or $N - r$ is zero, however, the expectations of the reciprocals of equation (15) are undefined.

Conditional Expectations

Observably monomorphic sites ($r = 0$ or N) are of little or no utility in paternity analysis, of course, and we really require the expectations of the reciprocals of equation (15), conditional on ($0 < r < N$):

$$E(\hat{P}^{-1} \mid 0 < r < N) = \frac{\sum_{r=1}^{N-1} \frac{N}{r} \cdot \binom{N}{r} P^r Q^{N-r}}{(1 - P^N - Q^N)} \tag{16a}$$

and

$$E(\hat{Q}^{-1} \mid 0 < r < N) = \frac{\sum_{r=1}^{N-1} \frac{N}{N-r} \cdot \binom{N}{r} P^r Q^{N-r}}{(1 - P^N - Q^N)} . \tag{16b}$$

These expectations cannot be written in closed form, but they can be computed.

Both equations (16a) and (16b) are biased, the magnitudes and directions of these biases depending upon the magnitudes of the sample size N and the parameter P (or $Q = 1 - P$). We present in figure 4 (top) the relative bias of $\hat{P}^{-1}$ (the bias divided by P^{-1}), as a function of P, for different values of N. The relative bias is large and negative for rare alleles ($P \rightarrow 0$), but positive for more frequent alleles. The only way to avoid serious biases for rare alleles is to employ large sample sizes to estimate allele frequencies, as has been done, for example, by Baird et al. [35]. Pending the availability of large sample sizes, we should choose those RFLP markers with both $N\hat{P}$ and $N\hat{Q} > 10$, say. Since it is best to have markers with $P \rightarrow \frac{1}{2} \leftarrow Q$ in order to maximize either information

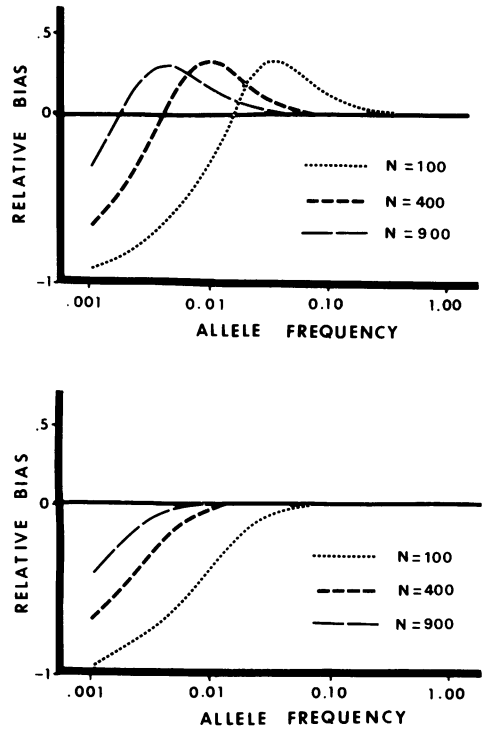

FIG. 4.—A plot of relative estimation bias against P for different values of the sample size N: (*top*) for the conditional estimate in equation (16a), and (*bottom*) for the alternate estimate in equation (17).

return or "utility," this stricture should work no real hardship with unlinked markers.

The problem comes when we use tightly linked sets of markers, where one or more haplotypes become rare. Examination of table 2 shows that 13 of the 22 recovered haplotypes for the β-globin region are singletons in our sample of $N = 65$, and four more are doubletons. Worse yet, there are $2^9 - 22 = 490$ potential haplotypes that are not present in the sample at all, but which may be present in the reference population at low frequency ($P \rightarrow 0$). The more tightly linked markers we add to the assay battery, the rarer each haplotype becomes. With limited sample size, the safest route is to use no more than a very few (perhaps two to four) markers per tightly linked set. For example, if we restrict attention to the most informative pair of markers (*Hin*dIII-$^{G}\gamma$ and *Rsa*I-3′β) for the β-globin complex, the haplotypic counts become:

$$(+\ +):9 \quad (+\ -):20 \quad (-\ +):7 \quad (-\ -):29\ .$$

Two of the numbers are less than adequate, and we would need to be cautious about the reliance we place on our (X:Y) ratios. If we add the incrementally

most useful marker (*Hgi*AI-β) to this pair, the haplotype counts become:

$$(+\ +\ +):9 \quad (+\ +\ -):16 \quad (+\ -\ +):0 \quad (-\ +\ +):7$$

$$(+\ -\ -):4 \quad (-\ +\ -):21 \quad (-\ -\ +):0 \quad (-\ -\ -):8\ ,$$

a set that has no sampled representatives for two of the haplotypes, both of which probably occur in the population at low frequency. In addition, there are less than adequate numbers of four other haplotypes. Worse yet, recall that to achieve a sample size of even $N = 65$, we lumped Italians and Greeks, a pooling of populations justified only by the necessities of illustration. Similar comments apply to the serum albumin and growth hormone regions, and a pair of tightly linked markers is about the limit of reliable statistical treatment for any of these systems, given the available sample sizes. Of course, with substantially larger sample sizes, one could probably use as many as three or four tightly linked markers.

An Alternative

Fortunately, we can devise alternative estimates of P^{-1} and Q^{-1} for rare alleles, estimates whose biases are unidirectional and smaller than those resulting from equations (16a) and (16b). Assuming a sample of r (+) alleles and $(N - r)$ (−) alleles, we define:

$$\tilde{P}^{-1} = \frac{N+1}{r+1} \quad \text{and} \quad \tilde{Q}^{-1} = \frac{N+1}{N-r+1}\ . \tag{17}$$

It can be shown (APPENDIX) that the unconditional expectations of these quantities are:

$$E(\tilde{P}^{-1}) = \frac{1}{P}[1 - Q^{N+1}] \quad \text{and} \quad E(\tilde{Q}^{-1}) = \frac{1}{Q}[1 - P^{N+1}]\ . \tag{18}$$

If P is small, the bias in $\tilde{P}^{-1}$ is large, but that in $\tilde{Q}^{-1}$ is very small. We present in figure 4 (bottom) the relative bias measure for $\tilde{P}^{-1}$, for comparison with the results in figure 4 (top). All biases are negative, and for $NP \geq 5$, the relative bias is always less than 1%.

The extension to the multiple allele or haplotype case is straightforward. The (X:Y) ratio takes only four forms [12]:

$$\frac{1}{P_j},\ \frac{1}{2P_j},\ \frac{1}{P_j + P_k} \quad \text{and} \quad 1\ ,$$

for which we suggest the estimates:

$$\frac{N+1}{r+1},\ \frac{N+1}{2r+1},\ \frac{N+1}{r+s+1} \quad \text{and} \quad 1\ ,$$

where r and s are the counts of alleles (haplotypes) j and k, respectively. While the two- and three-marker haplotypes of the β-globin complex do not yield reliable (X/Y) values from the usual frequency estimates, the values presented just above are rather better.

This alternative strategy of estimation has one other advantage. One will encounter an occasional haplotype in a test trio that has not (yet) been found in the reference panel. Adding one to both the numerator and demominator will provide an estimate, $(N + 1)/1$, negatively biased to be sure, but at least an estimate; observed zeroes need not be a problem. The resulting probability of paternity (for nonexcluded males) will then be an underestimate; the procedure is conservative in the absence of reliable reference data.

It should be obvious that with small samples there is sampling variance in the (X:Y) ratio. With small sample sizes for the reference population, we should probably place confidence intervals on our statements of the probability of paternity, explicitly quantifying our level of uncertainty. The proper derivation of those confidence limits is still a matter of active discussion ([29] and discussion following), and we shall not attempt to settle that more complicated question here. We terminate the discussion of sample size with the obvious thought that the reference population sample can be expected to grow as paternity test trios accumulate. The problem of small sample size is pressing but rectifiable.

DISCUSSION

Rare Recombinants

Up to this point, we have ignored the possibility of rare recombinants within our haplotypes. Over distances of a few kilobases, recombination should be quite rare, and we can generally afford to ignore it. More worrisome are the suspected "hot spots" for recombination, such as that postulated to occur between the *Hinc*III-3′ψβ1 and *Hgi*AI-β sites of the β-globin complex, with a recombination value thought to be as high as .001 [36]. To be completely rigorous, one should allow for recombination, as one does with the HLA system [2], although the extension to multiple markers is extremely cumbersome. If one allows for recombination among several tightly linked markers, the probability of categorical exclusion decreases dramatically, but the net result is to convert a fraction of the zero probability of paternity (PP) values (exclusions) to numbers very close to zero. There is room for personal preference in whether or not one corrects. Another approach is to avoid haplotypic sets that straddle a hot spot, a strategy that would alter the choice of tightly linked markers for the β-globin complex.

Haplotypes or Unlinked Markers

We have contrasted unlinked and tightly linked markers in several different fashions, and the results lead to a coherent set of strategic choices. Given a careful choice of tightly linked two-allele markers, detectable with a single probe, one can obtain more resolution at lower cost than with a comparable set (same array of allele frequencies) of unlinked markers. The problems of estab-

lishing linkage phase and estimating population frequencies, however, militate against the routine use of large numbers of haplotypic markers. With current technology, sets of two or three carefully chosen markers for each of several short regions are the best choice.

Replacing the Standard Battery

We come to the question of whether, in our pursuit of ever more resolution in paternity analysis, it is time for wholesale replacement of the traditional battery of antigenic and protein techniques with the more modern DNA procedures. There are a pair of considerations suggesting that such a move is premature. (1) Recent figures suggest that about 600,000 children are born out of wedlock every year in the United States [37] and only about 25,000 to 30,000 paternity tests are performed with the standard techniques [38]. We cannot handle the available case load with our standard procedures; the task of gearing up enough DNA labs to handle the load is sobering, to say the least. (2) One can obtain average exclusion probabilities on the order of 99% with the standard battery (including HLA). The commercial cost per trio ranges from $300 to $500 per trio, depending on a variety of factors (H. Gershowitz, personal communication, 1986). To achieve the same average level of resolution with RFLPs would require about the same number of markers, and the cost would be higher with today's technology.

On the other hand, as DNA technology improves and economies of scale come into play, we may want to phase in RFLP markers, gradually replacing the less informative and more expensive portions of the "standard battery." Moreover, selective use of RFLP markers to improve genetic resolution for specific mother-child pairs for whom the standard battery yields a low power of exclusion would be advantageous now, without being overly burdensome or expensive in terms of total case load. One other comment is in order if we are to add the RFLPs to the standard battery. It is important to use RFLP markers that are unlinked to and independent of any of the antigenic or protein markers that are to be used. It would be poor strategy, for example, to use the β-globin haplotypes in the same battery with the β^A and β^S markers.

Mutation Screening

Paternity analysis has traditionally been cast in a medico-legal framework, but there are also a variety of other contexts within which either a clear-cut paternity exclusion or a statement of the probability of paternity are useful. One such context is the study of spontaneous mutation rates in man [39, 40]. The standard approach is to search for an unusual genetic variant in a child that is present in neither parent, an approach described in considerable detail elsewhere [41–44]. The probability of a new mutation (the mutation rate) is not known with any precision, of course, but accumulating evidence suggests that the rate is almost certainly less than 10^{-4}/gene per generation. On those rare occasions when a "variant" is encountered, one must decide whether that variant is indeed a "mutant" or whether it represents some sort of "error."

Mutation screening is a very large-scale effort and is bound to yield occa-

sional errors; some of these may manifest as "variants." Fortunately, it is possible to reduce the rate of purely technical problems to undetectable levels with various combinations of replication, cross-checking, laboratory controls, and the like [45]. The most worrisome potential source of error is nonpaternity. We employ the standard battery of blood group antigens, red cell enzymes, serum proteins, and HLA antigens to reduce the probability of undetected nonpaternity to minimal levels, but even a probability of 10^{-4} might not be low enough for mutation work. For following up "potential mutants," each representing a very sizable scientific and financial investment, the potential resolving power of the RFLP technology has its attractions, cost notwithstanding. Paternity testing, involving (as it will) only a tiny fraction of the assayed trios, is the least expensive phase of the operation.

ACKNOWLEDGMENTS

We express our appreciation to J. Adams, A. Chakravarti, S. P. Daiger, H. Gershowitz, J. V. Neel, and a pair of anonymous reviewers for many helpful suggestions on the paper. The responsibility for any remaining errors of omission or commission is our own.

APPENDIX

PROOF OF ASSERTION (18)

The purpose of this APPENDIX is to show that the unconditional expectations of $\tilde{P}^{-1}$ and $\tilde{Q}^{-1}$ are as reported in text equations (18). Consider first $\tilde{P}^{-1}$, defined as:

$$\tilde{P}^{-1} = \frac{N+1}{r+1} , \tag{A-1}$$

where N is the allelic sample size and r the number of (+) alleles recovered. The unconditional expectation is computed as:

$$E(\tilde{P}^{-1}) = \sum_{r=0}^{N} \binom{N}{r} \frac{N+1}{r+1} P^r Q^{N-r} = \sum_{r=0}^{N} \binom{N+1}{r+1} P^r Q^{N-r} . \tag{A-2}$$

Setting $(N + 1) = M$ and expanding, we have:

$$E(\tilde{P}^{-1}) = \frac{1}{P} [\binom{M}{1} P^1 Q^{M-1} + \ldots + \binom{M}{M} P^M Q^0]$$

$$= \frac{1}{P} [1 - Q^M] = \frac{1}{P} [1 - Q^{N+1}] . \tag{A-3}$$

Similar treatment yields:

$$E(\tilde{Q}^{-1}) = \frac{1}{Q} [1 - P^M] = \frac{1}{Q} [1 - P^{N+1}] . \tag{A-4}$$

REFERENCES

1. CHAKRABORTY R, SHAW M, SCHULL WJ: Exclusion of paternity: the current state of the art. *Am J Hum Genet* 26:477–488, 1974
2. SMOUSE PE, ADAMS J: The use of closely linked genetic markers for paternity analysis, with special consideration of the HLA complex, in *Inclusion Probabilities in Parentage Testing,* edited by WALKER RH ET AL., Arlington, Va., American Association of Blood Banks, 1983, pp 397–407
3. FERRELL RE, CHAKRABORTY R: Paternity testing: a human genetic perspective, in *Advanced Family Law Course,* edited by State Bar Association Texas, Austin, Texas, Butlerworth, 1983, pp R1–R19
4. CHAKRABORTY R, RYMAN N: Use of odds of paternity computations in determining the reliability of single exclusions in paternity testing. *Hum Hered* 31:363–369, 1981
5. BALAZS I, WEXLER K, NICHOLAS L, ET AL.: The use of restriction fragment length polymorphism for the determination of paternity. *Am J Hum Genet* 36:161S, 1984
6. CHAKRAVARTI A, LI CC: The effect of linkage on paternity calculations, in *Inclusion Probabilities in Parentage Testing,* edited by WALKER RH ET AL., Arlington, Va., American Association of Blood Banks, 1983, pp 411–420
7. COOPER DN, SCHMIDTKE J: DNA restriction fragment length polymorphisms and heterogeneity in the human genome. *Hum Genet* 66:1–16, 1984
8. WIENER AS, LEDERER M, POLAYES SM: Studies in isohemagglutination. IV. On the chances of proving nonpaternity with special reference to blood groups. *J Immunol* 19:259–282, 1930
9. NEEL JV, SCHULL WJ: *Human Heredity.* Chicago, Univ. of Chicago Press, 1954
10. SALMON D, SALMON S: Blood groups and genetic markers polymorphism and probability of paternity. *Transfusion* 20:684–694, 1980
11. MORRIS J: Relationship between power of exclusion and probability of paternity, in *Inclusion Probabilities in Parentage Testing,* edited by WALKER RH ET AL., Arlington, Va., American Association of Blood Banks, 1983, pp 267–275
12. CHAKRABORTY R, FERRELL RE: Correlation of paternity index with probability of exclusion and efficiency criteria of genetic markers for paternity testing. *Forensic Sci Int* 19:113–124, 1982
13. CHAKRAVARTI H, BUETOW KH: A strategy for using multiple linked markers for genetic counseling. *Am J Hum Genet* 37:984–997, 1985
14. MURRAY JC, MILLS KA, DEMOPULOS CM, HORNUNG S, MOTULSKY AG: Linkage disequilibrium and evolutionary relationships of DNA variants (restriction enzyme fragment length polymorphisms) at the serum albumin locus. *Proc Natl Acad Sci USA* 81:3486–3490, 1984
15. CHAKRAVARTI A, PHILLIPS JA III, MELLITS KH, BUETOW KH, SEEBURG PH: Patterns of polymorphism and linkage disequilibrium suggest independent origins of the human growth hormone gene cluster. *Proc Natl Acad Sci USA* 81:6085–6089, 1984
16. SELVIN S: Probability of nonpaternity determined by multiple allele codominant systems (Letter to the Editor). *Am J Hum Genet* 32:276–278, 1980
17. WYMAN AR, WHITE R: A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA* 77:6754–6758, 1980
18. BELL GI, KARAM JH, RUTTER WJ: Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. *Proc Natl Acad Sci USA* 78:5759–5763, 1981
19. GOODBOURNE SEY, HIGGS DR, CLEGG JB, WEATHERALL DJ: Molecular basis of length polymorphism in the human ζ-globin gene complex. *Proc Natl Acad Sci USA* 80:5022–5026, 1983
20. JEFFRIES AJ, WILSON V, THAIN SL: Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67–73, 1985
21. CHAKRABORTY R, HEDRICK PW: Paternity exclusion and the paternity index for two linked loci. *Hum Hered* 33:13–23, 1983

22. NEI M: Utility and efficiency of linked marker genes for genetic counseling, in *Proceedings of the 41st Session of the International Statistical Institute,* New Delhi, 1977, pp 698–711
23. NEI M: Proportion of informative families for genetic counseling with linked marker genes. *Jpn J Genet* 24:131–142, 1979
24. CHAKRAVARTI A, NEI M: Utility and efficiency of linked marker genes for genetic counseling. II. Identification of linkage phase by offspring phenotypes. *Am J Hum Genet* 34:531–551, 1982
25. MORGAN K, HOLMES TM, SCHLAUT J, ET AL.: Genetic variability of HLA in Dariusleut Hutterites. A comparative genetic analysis of the Hutterites, the Amish, and other selected Caucasian populations. *Am J Hum Genet* 32:246–257, 1980
26. BAUR MP, NEUGEBAUER M, SIGMUND M: Genetic Analysis Workshop. III.: linkage analysis using FAP. *Genet Epidemiol* 2:201–202, 1985
27. MIGONE N, FEDER J, CANN H, ET AL.: Multiple DNA fragment length polymorphisms associated with immunoglobulin μ chain switch-like regions in man. *Proc Natl Acad Sci USA* 80:467–471, 1983
28. BECH-HANSEN NT, LINSLEY PS, COX DW: Restriction fragment length polymorphisms associated with immunoglobulin γ genes reveal linkage disequilibrium and genome organization. *Proc Natl Acad Sci USA* 80:6952–6956, 1983
29. SELVIN S: Some statistical properties of the paternity ratio, in *Inclusion Probabilities in Parentage Testing,* edited by WALKER RH ET AL., Arlington, Va., American Association of Blood Banks, 1983, pp 77–85
30. MAYR WR: Grundlagen zur Berechnung der Vaterschaftswarhrscheinlichkeit im HL-A system. *Z Immunitötsforsch* 144:18–27, 1972
31. MAYR WR, PAUSCH V: Vaterschaftsausschlusschance im HLA system. *Z Immunitötsforsch* 150:447–457, 1975
32. BAUR MP, MAYR WR, RITTNER C: Algorithm for the computation of plausibilities of paternity in the HLA system. *Z Immunitötsforsch* 152:209–219, 1976
33. MILLER WV: HLA serotyping in cases of disputed paternity, in *Paternity Testing.* Washington, D.C., American Association of Blood Banks, 1978, pp 55–67
34. LAMM LU, GÜRTLER H, HANSEN HE: The HLA system, in *Inclusion Probabilities in Parentage Testing,* edited by WALKER RH ET AL., Arlington, Va., American Association of Blood Banks, 1983, pp 381–392
35. BAIRD M, GIUSTI A, MIYAZAKE L, NICHOLAS L, WEXLER K, BALAZS I: Allele frequency distribution of highly polymorphic DNA sequences in Caucasian and black populations. *Am J Hum Genet* 36:161S, 1984
36. CHAKRAVARTI A, BUETOW KH, ANTONARAKIS SE, WALKER PA, BOEHM CD, KAZAZIAN HH: Nonuniform recombination within the human β-globin gene cluster. *Am J Hum Genet* 36:1239–1258, 1984
37. SCHUTZMAN F: Interests of the office of child support enforcement, U.S. Department of Health and Human Services, in *Inclusion Probabilities in Parentage Testing,* edited by WALKER RH ET AL., Arlington, Va., American Association of Blood Banks, 1983, pp 6–12
38. WALKER RH: Studies and surveys of the American Association of Blood Banks, in ibid., pp 25–33
39. NEEL JV: The detection of increased mutation rates in human populations. *Perspect Biol Med* 14:522–537, 1971
40. ROTHMAN ED, NEEL JV, HOPPE FM: Assigning a probability for paternity in apparent cases of mutation. *Am J Hum Genet* 33:617–628, 1981
41. NEEL JV, UEDA N, SATOH C, FERRELL RE, TANIS RJ, HAMILTON HB: The frequency in Japanese of genetic variants of 22 proteins. V. Summary and comparison with data on Caucasians from the British Isles. *Ann Hum Genet* 41:429–441, 1978
42. NEEL JV, MOHRENWEISER H, HANASH S, ET AL.: Biochemical approaches to monitoring human populations for germinal mutation rates. I. Electrophoresis, in *Utilization of Mammalian Specific Locus Studies in Hazard Evaluation and Estimation of*

Genetic Risk, edited by DESERRES FJ, SHERIDAN W, New York, Plenum Press, 1983, pp 71–93
43. MOHRENWEISER HW: Biochemical approaches to monitoring human populations for germinal mutation rates: II. Enzyme deficiency variants as a component of the estimated genetic risk, in ibid., pp 55–69
44. MOHRENWEISER HW: Enzyme-deficiency variants: frequency and potential significance in human populations. *Isozymes: Curr Top Biol Med Res* 10:51–68, 1983
45. NEEL JV, MOHRENWEISER HW, SATOH C, HAMILTON HB: A consideration of two biochemical approaches to monitoring human populations for a change in germ cell mutation rates, in *Genetic Damage in Man Caused by Environmental Agents,* edited by BORG K, New York, Academic Press, 1979, pp 29–47

Notices

Training Programs Guide
You will soon receive your copy of the new Information and Education Committee *Guide to Human Genetics Training Programs in North America.* The *Guide* will contain extensive data on 100 training programs. An index at the back of the book will list programs accredited by the American Board of Medical Genetics as of December 31, 1985. The ABMG is now in the process of granting further accreditations. An up-to-date list of accredited programs will be published in *The American Journal of Human Genetics* and will be available from the Administrative Office. Should you wish to order additional copies of the *Guide,* contact the Administrative Office. The price of additional copies is expected to be in the neighborhood of $10 each.

ASHG Position Paper on MSAFP Screening
The Society now has an official statement entitled "Guidelines for a Maternal Serum Alpha Fetoprotein Screening Program and Quality Control for Laboratories Performing Maternal Serum and Amniotic Fluid Alpha Fetoprotein Assays." This document was drafted by an ad hoc committee (Ken Garver and Jessica Davis, Co-Chairpersons) under the aegis of the ASHG Social Issues Committee and Genetic Services Committee. In accord with the newly revised Bylaws, the final form of the statement was approved by the ASHG Public Policy Committee. It has been sent to the American Academy of Pediatrics, the American College of Obstetrics and Gynecology, and the Centers for Disease Control and will soon be released to the press and published in *The American Journal of Human Genetics.* Contact the Administrative Office if you wish to have a copy.

Corrections of Errors in Membership Directory
A list of corrections is enclosed. Please moisten the sheet and affix it to your copy of the membership directory. Additional copies of the directory are available from the Administrative Office for $10 each.

Historical ASHG Data
We would like to collect data about the ASHG for archival purposes. If you have old copies of the *Journal* or old committee files you would like to donate, please contact Gerry Gurvitch in the Administrative Office. We are especially in need of abstract volumes prior to 1983.