Am J Hum Genet 35:1048–1052, 1983

Brief Communication

The Probability of Exclusion Based on the *HLA* Locus

ARAVINDA CHAKRAVARTI[1] AND CHING CHUN LI

It has been estimated that in 1978 approximately 543,900 illegitimate births occurred in the United States [1]. A large number of these cases appear before courts of law each year to adjudicate questions of paternity. It has also been shown that by using 57 blood-group and enzyme marker systems more than 95% of falsely accused males in these paternity suits may be excluded [2]. The probability with which a marker locus will exclude a falsely accused male depends on the degree of polymorphism at that locus. Since the closely linked *HLA* loci that reside on human chromosome 6 constitute the most polymorphic marker system known [3], these loci are of great utility for excluding paternity among falsely accused males. For evaluating the usefulness of the *HLA* system one needs to compute the average probability of exclusion for each *HLA* locus. In other words, we need to compute the average probability with which a falsely accused male will be exonerated on the basis of *HLA* testing of a mother-child-putative father trio. Several authors [4–6] have given expressions for computing this probability. Unfortunately, these formulas are incorrect because these authors consider each *HLA* type as an allele dominant over the blank allele but fail to consider the fact that expressed alleles are codominant to each other. This paper describes a simple method for evaluating the probability of exclusion for *HLA* loci.

Any single *HLA* locus, such as *HLA-A*, can be considered to be a generalized *ABO*-like system because there are several codominant alleles and a single null/blank type recessive to all others. Let us suppose that at this locus there are k codominant alleles $A_1, A_2, \ldots, A_k$ and a null allele A_x, with frequencies $p_1, p_2, \ldots, p_k$, and r, respectively, where

$$\sum_{i=1}^{k} p_i + r = 1 .$$

For evaluating the average probability of exclusion $P(\mathrm{E})$, we first compute the probability of exclusion for each possible mother-child (MC) combination $P_{\mathrm{E|MC}}$.

Received September 30, 1982; revised December 27, 1982.

[1] Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261.

The $P(E)$ value is then obtained by weighting the $P_{E|MC}$ values by the probability of observing the specific mother-child combination, P_{MC}, so that

$$P(E) = \sum_{MC} P_{E|MC} \cdot P_{MC} , \tag{1}$$

where the summation is over all mother-child pairs.

As an example, consider the case when both mother and child have genotype A_iA_j $(i < j)$. Then, it is clear that any male who does not possess either the A_i or the A_j allele will be excluded. Therefore, in a random-mating population, $P_{E|MC} = (1 - p_i - p_j)^2$. On the other hand, P_{MC} can be easily obtained using the ITO method of Li and Sacks [7] and is $p_ip_j \cdot (p_i + p_j)$. Thus, for this particular mother-child combination, the contribution to $P(E)$ is $p_ip_j(p_i + p_j)(1 - p_i - p_j)^2$. By summing this quantity for all i and j $(i < j)$, we obtain the third term in equation (2). Table 1 provides these values for all possible mother-child combinations. Using the values in table 1 and equation (1) we obtain after simplification:

$$\begin{aligned} P(E) = {} & \sum_{i=1}^{k} p_i(1 - p_i)^4 + 2r^2 \sum_{i<j}^{k} p_ip_j \\ & + \sum_{i<j}^{k} p_ip_j(p_i + p_j)(1 - p_i - p_j)^2 \\ & + \sum_{i<j}^{k} p_ip_j(p_i + r)[(1 - r - p_i)^2 - h + p_i^2] \\ & + \sum_{i=1}^{k} p_i(p_i^2 + 3p_ir + r^2)[(1 - r - p_i)^2 - h + p_i^2] , \end{aligned} \tag{2}$$

where

$$h = \sum_{i=1}^{k} p_i^2 .$$

Previous formulations [4–6] used only the term

$$\sum_{i=1}^{k} p_i(1 - p_i)^4$$

and is thus an underestimate. By setting $k = 2$, one obtains the probability of exclusion for the *ABO* locus as:

$$P(E) = p_1(1 - p_1)^4 + p_2(1 - p_2)^4 + 2p_1p_2r^2 + p_1p_2(p_1 + p_2)r^2 , \tag{3}$$

which is identical to Wiener et al.'s formula [8].

TABLE 1

THEORETICAL PROBABILITY OF EXCLUSION FOR AN *HLA* LOCUS

MOTHER'S PHENOTYPE		CHILD'S PHENOTYPE						
		$A_iA_j(i<j)$	$A_iA_\alpha(i<\alpha)$	$A_jA_\beta(j<\beta)$	A_i	A_j	A_α	A_x
A_iA_j for $i<j$; $i,j=$ 1, 2, . . . , k	(P_{MC})	$p_ip_j(p_i+p_j)$	$p_ip_jp_\alpha$	$p_ip_jp_\beta$	$p_ip_j(p_i+r)$	$p_ip_j(p_j+r)$	. . .	. . .
	($P_{E\|MC}$)	$(1-p_i-p_j)^2$	$(1-p_\alpha)^2$	$(1-p_\beta)^2$	$(1-r-p_i)^2+p_i^2-h$	$(1-r-p_j)^2+p_j^2-h$	. . .	. . .
A_i for $i=$ 1, 2, . . . , k	(P_{MC})	. . .	$p_ip_\alpha(p_i+r)$	. . .	$p_i(p_i^2+3p_ir+r^2)$	. . .	rp_ip_α	p_ir^2
	($P_{E\|MC}$)	. . .	$(1-p_\alpha)^2$	. . .	$(1-r-p_i)^2+p_i^2-h$	. . .	$(1-p_\alpha)^2$	$\sum_{\alpha\neq\beta} p_\alpha p_\beta$
A_x	(P_{MC})	. . .	. . .	. . .	. . .	. . .	$p_\alpha r^2$	r^3
	($P_{E\|MC}$)	. . .	. . .	. . .	. . .	. . .	$(1-p_\alpha)^2$	$\sum_{\alpha\neq\beta} p_\alpha p_\beta$

NOTE: Most of the exclusion probabilities are self-evident. A few words for the mother-child pair $A_iA_j - A_i$ may be helpful. Let the mother be A_1A_2 and child A_1. The biological father must have contributed A_1 or A_x. Hence, all males of genotypes lacking A_1 or A_x may be excluded, yielding the probability $(1-p_1-r)$.[2] However, the nonexcluded genotypes A_2A_x, A_3A_x, . . . are indistinguishable from A_2A_2, A_3A_3, . . . (which should be but are not excluded due to dominance). Thus, the actual exclusion probability decreases to $(1-p_1-r)^2-p_2^2-p_3^2-\ldots=(1-p_1-r)^2+p_1^2-h$, where

$$h=\sum_{i=1}^{k} p_i^2 .$$

TABLE 2

PROBABILITY OF EXCLUSION USING *HLA* LOCI

POPULATION	*HLA* LOCUS			
	A	*B*	*C*	*DRw*
European Caucasoid	.727	.856	.465	.660
African black	.740	.725	.370	.507
Japanese	.508	.462	.289	.452

The probability of exclusion when all alleles are equally frequent can be easily calculated. Thus, when $p_1 = p_2 \ldots = p_k = r = 1/(k + 1)$, we can show that:

$$P(\mathrm{E}) = k[k^4 + (k - 1)(3k^2 - 2k - 4)]/(k + 1)^5 . \quad (4)$$

Therefore, if $k = 2$, 5, or 10, $P(\mathrm{E})$ takes the values .165, .559, and .775. On the other hand, in the case where all the $k + 1$ alleles are codominant, the corresponding values are .370, .660 and .813, respectively [9]. The occurrence of blank alleles, therefore, reduces the exclusion probability significantly unless the number of alleles is large.

Equation (2) can be used to compute the probability of exclusion for various populations using observed frequencies of alleles at the *HLA-A*, *B*, *C*, or *DRw* locus. Such frequencies are given in Bodmer and Bodmer [3], and the $P(\mathrm{E})$ values are presented in table 2. These results should be interpreted with caution because *HLA* testing in non-Caucasoids and using *HLA-C* or *DRw* has not been extensive so that the frequency data may be unreliable. However, the data from the Caucasians demonstrate that *HLA-B* is the most useful locus, followed by the *A*, *DRw*, and *C* loci. Another larger sample of Caucasians studied in this country [10] shows the probability of exclusion for *HLA-A* and *B* to be .687 and .811, respectively.

For most cases tested using *HLA*, only locus *A* and *B* are used. It is thus of interest to calculate the total probability of exclusion using both these loci. Then, the total probability of exclusion in Caucasoids will be (from table 2): $P(\mathrm{E}; HLA\text{-}A, B) = 1 - [1 - P(\mathrm{E}; HLA\text{-}A)] \cdot [1 - P(\mathrm{E}; HLA - B)] = .961$, provided *A* and *B* are assumed to be independent. This value overestimates the actual probability because of considerable nonrandom associations between *HLA-A* and *B* [3]. Usually this will mean that an exclusion/nonexclusion at the *A* locus will imply the same at locus *B*. However, we can compute $P(\mathrm{E})$ under complete linkage disequilibrium when each *A* allele is associated with a particular *B* allele; the probability of exclusion will be .856. The actual probability lies somewhere in the range .856 – .961; we take the average .909 as a representative value.

Besides providing a method for computing $P(\mathrm{E})$, the computations above show that for closely linked loci such as the *HLA* markers there is considerable redundancy in information about paternity exclusion [11]. The above results suggest that *HLA-B* is quite informative by itself, at least among Caucasians. In view of this, would it not be more prudent to type paternity cases with *HLA-B* and another highly polymorphic unlinked locus than to type cases with both *HLA-A* and *B*?

The discovery of DNA markers will surely be excellent candidates for such analyses [12].

ACKNOWLEDGMENT

We thank Mr. Kenneth Buetow for his help in the numerical calculations.

REFERENCES

1. KEITH RE: Resolution of paternity disputes by analysis of the blood. *Fam Law Reptr* 8:4001–4008, 1981
2. CHAKRABORTY R, SHAW M, SCHULL WJ: Exclusion of paternity: current state of the art. *Am J Hum Genet* 26:477–488, 1974
3. BODMER WF, BODMER JG: Evolution and function of the HLA system. *Br Med Bull* 34:309–316, 1978
4. MAYR WR: Die genetik des HL-A systems. Populations-und Familienuntersuchungen unter besonder Beruksichtigung der Paternitatsserologie. *Humangenetik* 12:195–243, 1971
5. JEANNET M, HASSIG A, BERNHEIM J: Use of the HL-A antigen system in disputed paternity cases. *Vox Sang* 23:197–200, 1972
6. LEE CL, HENRY JB: Laboratory evaluation of disputed parentage, in *Clinical Diagnosis and Management by Laboratory Methods*, edited by HENRY JB, NELSON DA, WASHINGTON JA, MCLENDON WW, STATLAND BE, TOMAR RH, Philadelphia, W. B. Saunders, 1979, pp 1507–1548
7. LI CC, SACKS L: The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. *Biometrics* 10:347–360, 1954
8. WIENER AS, LEDERER M, POLAYES HS: Studies in isohemagglutination: IV. On the chances of proving nonpaternity with special reference to blood groups. *J Immunol* 19:259–282, 1930
9. SELVIN S: Probability of nonpaternity determined by multiple allele codominant systems (Letter to the Editor). *Am J Hum Genet* 32:276–278, 1980
10. DAUSSET J, COLOMBANI J, EDS: *Histocompatibility Testing 1972*. Copenhagen, Munksgaard, 1973, p 235
11. CHAKRAVARTI A, LI CC: The effect of linkage on paternity calculations, in *Inclusion Probabilities in Parentage Testing*, edited by WALKER RH, Washington, D.C., American Association of Blood Banks, 1983
12. BOTSTEIN D, WHITE RL, SKOLNICK M, DAVIS RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331, 1980