Pages 297–303, Genetic Analysis of Common Diseases:
Applications to Predictive Factors in Coronary Disease

POTENTIAL INFORMATION IN FAMILY STUDIES OF LINKAGE*

Ranajit Chakraborty, Paul A. Fuerst and
Robert E. Ferrell

Center for Demographic and Population Genetics
University of Texas Health Science Center
P.O. Box 20334, Houston, Texas 77025

The search for major genetic factors in certain disease states through linkage and association has been discussed several times during this meeting. When family data are used for this purpose it is often necessary to evaluate the expected number of mating pairs that are potentially informative with regard to a given set of genetic markers. The question of whether marginally polymorphic genetic markers are useful for linkage studies also arises occasionally in this connection. In this note, therefore, our purpose is to propose a criterion to indicate whether a mating pair randomly drawn from a population is potentially informative for such a linkage study. Specifically, we call a mating pair informative (potentially) if their genotypes are such that they can, in theory, produce phenotypically segregating offspring. Given this criteria, we may then determine (1) what proportion of mating pairs will be phenotypically segregating for k (= 0,1,2,...) number of genetic markers, (2) the cumulative probability distribution of segregation, and (3) the average number of genetic systems which would show phenotypic segregation in a random pair using the currently available genetic markers which can be routinely typed using human serum and erythrocytes. We first present the theoretical background for deriving the above distribution. Using the current gene frequencies in Caucasians and Japanese populations we then compute the various measures to illustrate the extent of information provided by genetic

*Research supported by U.S. Public Health Service grants GM 19513 and CA 19311 from the National Institutes of Health.

markers mapped on various human chromosomes. We recognize that the above criterion have been considered by Renwick (1969) and Elston and Lange (1975) to evaluate the prior probability of autosomal linkage.

THEORETICAL BACKGROUND

A mating pair is defined here to be potentially informative if it is capable of producing offspring of at least two different phenotypes. For a given marker we may then compute the probability that a randomly drawn pair is informative from the knowledge of gene frequencies at that locus, assuming random mating. When all genotypes at such a locus are phenotypically distinguishable, this probability is simply given by $1 - (\Sigma p_i^2)^2$, where p_i is the gene frequency. This is equivalent to saying that in such cases at least one member of such a mating pair has to be a heterozygote. If the locus is such that not all genotypes are phenotypically distinguishable (as in the case of several immunological markers) we have to subtract from the above quantity the probability of obtaining a mating pair which produces offspring which segregate genotypically but not phenotypically. At a two allelic locus where one allele (a) is recessive (with frequency q in the population) this probability is $4p^3q$ (representing the probability of AA × Aa matings). In Table 1 we provide the formulae for computing the probability of (potentially) phenotypic segregation for some specific systems which are used for subsequent computations.

Distribution of Phenotypically Segregating Loci

Let p_i denote the probability that a random mating pair will show (potentially) phenotypic segregation with respect to the i-th genetic marker ($i = 1,2,\ldots,n$; n being the number of genetic markers studied). If X denotes a random variable representing the number of systems with respect to which the random pair shows segregation, the probability distribution of X (taking integral values from 0 to n) can be obtained following Chakraborty and Schull (1976). The same authors also derived an algorithm to compute the cumulative probability distribution of X from which we may evaluate the expected proportion of mating pairs that would be informative with respect to k or more ($k \leq n$) systems.

The mean and variance of X can also be evaluated following Chakraborty and Schull (1976) as $E(X) = \Sigma p_i$ and $V(X) = \Sigma p_i(1 - p_i)$, respectively.

Table 1. Probabilities of phenotypic segregation for some typical genetic loci

Locus	Probability of potential phenotypic segregation
Autosomal codominant	$1 - J^2$
Autosomal two allelic with dominance	$4pq^2$
A_1A_2BO	$1 - J^2 - 4(p_1^3r + p_2^3 + q^3r + p_1^3p^2 + p_1^2p_2^2)$
Rh	$1 - J^2 - 4(R_z^2 + R_2^2 + R_1^2 + R_0^2)(r_yR_z + r''R_2 + r'R_1 + rR_0)$

J = homozygosity at the locus, obtained as sum of squares of all gene frequencies; allele frequency notations as used in Mourant et al. (1976)

NUMERICAL EXAMPLE

To provide some estimate of the usefulness of the genetic markers which may be typed using human serum and erythrocytes we obtained the gene frequency data from Caucasians and Japanese populations surveying the literature (sources are given in the appendix). Table 2 presents the probabilities of phenotypic segregation for 39 genetic systems which are variable in at least one of these two populations. We also obtain the maximum probabilities of phenotypic segregation (given in column 5 of Table 2) which are theoretically attainable for each system. Using these probabilities we then obtain the distribution of the number of genetic markers with respect to which a mating pair randomly drawn from a population would show phenotypic segregation. It is surprising that in Caucasians about 96% of mating pairs should be potentially informative for 11 to 20

of these 39 systems whereas in Japanese populations about 95% are expected to be informative for 9 to 17 of these systems.

Table 2. Probability that a random mating will be informative at each of 39 genetic loci

Chromosomal assignment*	Marker system	Probability of phenotypic segregation			Reference**
		Caucasian	Japanese	Maximum	
1	6PGD	0.0732	0.2925	0.7500	1,2 (26)
	Rh***	0.8613	0.7458	0.9745	3 (2049,1939)
	UMPK	0.1715	0.1907	0.7500	4,5
	PGM-1	0.6011	0.6120	0.8889	1,2 (262)
	Fy	0.7434	0.3819	0.7500	3 (2337,2067)
	PepC	0.0430	0.0000	0.7500	1,6
2	ACP-1	0.7686	0.5327	0.7500	1,2 (29)
	MNSs	0.9076	0.8138	0.9375	3 (549,959)
3	GALT	0.3429	0.0392	0.8889	19
4	Gc	0.6645	0.5787	0.0000	6,7 (5)
6	Pg	0.4280	0.0000	0.7520	7 (47,48)
	Bf	0.6634	0.4960	0.9833	8,9
	GLO-1	0.7441	no data	0.7500	10
	P	0.4834	0.4725	0.5926	3 (1320,856)
9	ABO	0.7256	0.7744	0.8694	3 (272)
	AK-1	0.1511	0.0000	0.7500	1,2 (26)
10	GOT	0.0000	0.0582	0.7500	7 (8)
13	ESD	0.3607	0.6976	0.7500	11
16	Hp	0.7335	0.6119	0.7500	7 (12,54)
	PepD	0.0315	0.0080	0.7500	1,5
20	ADA	0.1907	0.1023	0.7500	1,2 (26)
Not assigned	Or	0.7093	0.6330	0.7500	3 (24)
	Tf	0.0237	0.0393	0.9600	12
	αAT	0.2254	0.1766	0.8933	6,13
	E_2	0.1837	0.1575	0.7500	6,14
	Ag	0.5829	0.6341	0.7500	6,15
	E_1	0.0844	0.0392	0.8889	7 (33)
	Inv	0.2182	0.4668	0.7500	6,7 (5)
	A2M	0.0080	0.4867	0.7500	16
	α-GLU	0.1165	0.0915	0.7500	17

Table 2. Continued

Chromosomal assignment	Marker system	Probability of phenotypic segregation Caucasian	Japanese	Maximum	Reference
Not assigned	GPT	0.7499	0.7188	0.7500	7 (7)
	Crp1	0.0237	0.0000	0.7500	2(10, 7(51,56)
	C'3	0.6094	0.0276	0.8889	7 (1,3)
	Lu	0.1198	0.0000	0.5926	3 (1320,2068)
	Kell	0.1443	0.0000	0.7500	3 (1320,2067)
	Se	0.4921	0.4704	0.5926	3 (1280,1339)
	Kidd	0.4800	0.5874	0.5926	3 (2534), 4
	Diego	0.0000	0.0798	0.5926	1 (1695,3162)
	Gm	0.7582	0.8558	0.9600	6,16

*Chromosomal assignments and abbreviations are taken from McKusick and Ruddle (18). **Primary references are given in the Appendix. Numbers in parentheses refer to citations given in these original papers. ***Antisera used: ABO (A_1AB), Rh (CcDEe), MNSs (M,N,S,s), Fy (Fy^a,Fy^b), P (P_1), Kell (K,k), Lu (Lu^a), Kidd (Jk^a), Diego (Di^a), GM (1,2,3,5,13,14,21), Inv (Inv^1), A2M ($A2M^1$)

Figure 1 plots these distributions. Since probabilities vary widely, these are plotted on logarithmic scale. Figure 2 indicates the cumulative distribution which indicates the proportion of mating pairs which should be potentially informative for k (≤39) or more systems. For example, more than 97% of Caucasian mating pairs are expected to be potentially informative for 11 or more systems whereas about 85% of Japanese mating pairs would show phenotypic segregation in their offspring for 11 or more systems. The mean and variance of the number of informative systems in Caucasian mating pairs are 15.2 and 5.9 whereas the Japanese figures are 12.9 and 5.3. Figures 1 and 2 also indicate theoretically maximum values of such probabilities based on these 39 systems (indicated by dotted lines). It is worth mentioning that even if a locus is marginally polymorphic in a population, for linkage studies this locus may still provide considerable information. For example, the combined gene frequency for all rare alleles at the peptidase C locus in Caucasians is only 0.011, but still 4.3 of every 100 matings would be expected to be informative at this locus. This would not be a trivial number of informative matings in large linkage studies.

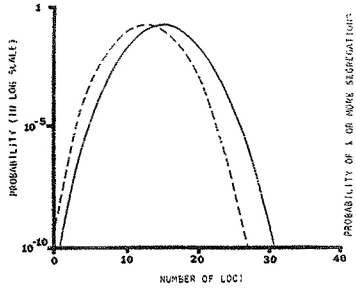

FIGURE 1. THE DISTRIBUTION OF POTENTIALLY INFORMATIVE SYSTEMS AS DERIVED FROM THE GENE FREQUENCIES AT 39 LOCI IN CAUCASIANS (SOLID LINE) AND JAPANESE (DASHED LINE) POPULATIONS. THE THEORETICALLY MAXIMUM ATTAINABLE PROBABILITIES ARE INDICATED BY A DOTTED LINE.

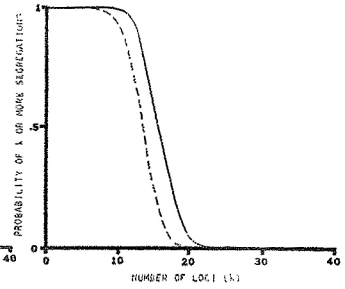

FIGURE 2. CUMULATIVE PROBABILITY DISTRIBUTIONS FOR THE NUMBER OF POTENTIALLY INFORMATIVE SYSTEMS IN CAUCASIAN (SOLID LINE) AND JAPANESE (DASHED LINE) POPULATIONS. THEORETICAL MAXIMUM PROBABILITIES ARE INDICATED BY A DOTTED LINE.

Several other points are worth mentioning. Consideration of chromosomal assignments indicate that only 10 of the 23 chromosomes are currently marked by informative loci. Using the Paris Conference (1972) standards these chromosomes are estimated to make up approximately 55% of the human genome. Obviously because the available markers are not uniformly distributed over these chromosomes, the proportion of the genome marked by known mapped markers will be somewhat smaller than this figure. On the other hand, 18 of the 39 markers which we list are currently unassigned, and most have additionally not been shown to be linked with other unassigned loci. This suggests that a substantial proportion of the genome may in fact be within a detectable distance from one of the markers listed in Table 2. Whether a linkage can in fact be detected will of course depend upon the map distance between the marker and the "disease" locus and the sample size of families, dependent upon the proportion of potentially informative families as shown in Table 2.

Acknowledgments

The authors are grateful to Dr. R. H. Ward for making critical comments on an early draft of this paper.

REFERENCES

Chakraborty R, Schull WJ (1976). A note on the distribution of the number of exclusions to be expected in paternity testing. Am J Hum Genet 28:615.

Elston RC, Lange K (1975) The prior probability of autosomal linkage. Ann Hum Genet 38:341.

Mourant AE, Kopec AC, Domaniewska-Sobczak K (1976). "The distribution of the Human Blood Groups and other Polymorphisms." New York: Oxford University Press.

Paris Conference (1972). Standardization of human cytogenetics. Birth Defects Original Articles Series 8:7.

Renwick JH (1969). Progress in mapping human autosomes. Brit Med Bull 25:65.

APPENDIX: REFERENCES TO GENE FREQUENCY DATA (Table 2)

1. Harris H et al (1973) Ann Hum Genet 37:237. 2. Neel JV et al (1978) Ann Hum Genet 41:429. 3. Mourant AE et al (1976) "The Distribution of the Human Blood Groups." 4. Giblett E et al (1974) Ann J Hum Genet 26:627. 5. Harada S et al (1975) Humangenetik 29:255. 6. Ishimoto I (1975) In "Anthropological and Genetic Studies on the Japanese." 7. Chakraborty R et al (1974) Am J Hum Genet 26:477. 9. Alper CA et al (1972) J Exp Med 135:68. 9. Harai S (1976) Jap J Hum Genet 21:177. 10. Parr CW et al (1977) Biochem Genet 15:109. 11. Papiha SS et al (1977) Hum Hered 27:424. 12. Kirk RL (1968) Acta Genet Med Gemel 17:613. 13. Cook PJL (1975) Ann Hum Genet 38:275. 14. Whittaker M et al (1976) Hum Hered 26:380. 15. Butler RE et al (1974) Vox Sang 26:485. 16. Johnson WE et al (1977) Clin Immunol Immunopathol 7:97. 17. Swallow DM et al (1975) Ann Hum Genet 38:391. 18. McKusick VA et al (1977) Science 196:390. 19. Ng WG et al (1973) Ann Hum Genet 37:1.