Acknowledgment

This work was supported by the Canadian Genome Analysis and Technology Program and the Medical Research Council of Canada.

References

Barnavi E (ed) (1992) A historical atlas of the Jewish people. Schocken Books, New York

Beinart H (1992) Atlas of medieval Jewish history. Simon & Schuster, New York

Blumenfeld A, Slaugenhaupt SA, Axelrod PB, Lucente DE, Maayan Ch, Liebert ChB, Ozelius LJ, et al (1993) Localization of the gene for familial dysautonomia on chromosome 9 and definition of DNA markers for genetic diagnosis. Nat Genet 4:160–163

Bouchard G, De Braekeleer M (eds) (1990) Histoire d'un génome. Presses de l'Université du Québec, Québec

Davies N (1984) God's playground: a history of Poland. Columbia University Press, New York

De Braekeleer M (1991) Hereditary disorders in Saguenay-Lac-St-Jean (Quebec, Canada). Hum Hered 41:141–146

Diamond JM, Rotter JI (1987) Observing the founder effect in human evolution. Nature 329:105–106

Gauvreau D, Guérien M, Hamel M (1990) De Charlevoix au Saguenay: mesure et caractéristiques du mouvement migratoire avant 1911. In: Bouchard G, De Braekeleer M (eds) Histoire d'un génome. Presses de l'Université du Québec, Québec, pp 145–162

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

Jennings HS (1917) The numerical results of diverse systems of breeding, with the respect to two pairs of characters, linked or independent, with special relation to the effect of linkage. Genetics 2:97–154

Kaplan NL, Hill WG, Weir BS (1995) Likelihood methods for locating disease genes in nonequilibrium populations. Am J Hum Genet 56:18–32

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, et al (1989) Identification of the cystic fybrosis gene: genetic analysis. Science 245: 1073–1080

Labuda M, Labuda D, Korab-Laskowska M, Cole DEC, Ziętkiewicz E, Weissenbach J, Popowska E, et al (1996) Linkage disequilibrium analysis in young populations: pseudo–vitamin D–deficiency rickets and the founder effect in French Canadians. Am J Hum Genet 59:633–643

Luria SE, Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511

Motulsky AG (1995) Jewish diseases and origins. Nat Genet 9:99–101

Neuhausen SL, Mazoyer S, Friedman L, Stratton M, Offit K, Caligo A, Tomlinson G, et al (1996) Haplotype and phenotype analysis of six recurrent BRCA1 mutations in 61 families: results of an international study. Am J Hum Genet 58: 271–280

Risch N, de Leon D, Fahn S, Bressman S, Ozelius L, Breakefield X, Kramer P, et al (1995*a*) ITD in Ashkenazi Jews—genetic drift or selection? Nat Genet 11:14–15

Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, et al (1995*b*) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nat Genet 9:152–159

Roberts DF (1968) Genetic effects of population size reduction. Nature 220:1084–1088

Robbins RB (1917) Some applications of mathematics to breeding problems III. Genetics 3:375–389

Thompson EA, Neel JV (1997) Allelic disequilibirium and allele frequency distribution as a function of social and demographic history. Am J Hum Genet 60:197–204

Tonin P, Weber B, Offit K, Couch F, Rebbeck TR, Neuhausen S, Godwin AK, et al (1996) Frequency of recurrent BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer families. Nat Med 2:1179–1183

Zoosman-Diskin A (1995) ITD in Ashkenazi Jews—genetic drift or selection? Nat Genet 11:13–14

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 97–166

Address for correspondence and reprints: Dr. Damian Labuda, Centre de Cancérologie Charles Bruneau, Centre de Recherche, Hôpital Sainte-Justine, Département de Pédiatrie, Université de Montréal, 3175 Côte Sainte-Catherine, Montréal, Québec, H3T 1C5 Canada. E-mail: Labuda@ere.umontreal.ca

0002-9297/97/6103-0036$02.00

Am. J. Hum. Genet. 61:771–774, 1997

Approximate Variance of the Standardized Measure of Gametic Disequilibrium *D′*

To the Editor:

The detection of nonrandom association of alleles at different loci (i.e., gametic disequilibrium) is one of the first steps in the study of the multilocus systems, as well as a powerful tool for analysis of marker order and location of human disease genes. The magnitude of gametic disequilibrium can be estimated from a sample of haplotypes from the population, by means of a variety of disequilibrium coefficients (Hedrick et al. 1978; Hedrick 1987). Although no measure of disequilibrium is completely independent of allele frequency (Lewontin 1988), the standardized disequilibrium coefficient *D′*, proposed by Lewontin (1964), is very useful for comparison of disequilibrium between loci with different allelic frequencies, because its range is frequency independent (Hedrick 1987, 1988; Lewontin 1988; Zapata and Alvarez 1992, 1993, 1997*a;* Zapata and Visedo 1995). It is surprising that, despite the extensive usage of *D′* since 1964, its variance has not been investigated, although Hedrick and Thomson (1986) and Hedrick (1987) examined the distribution of *D′*, for sam-

ples taken from populations under neutral equilibrium, by computer simulation. Consequently, estimates of D' between genetic markers are always provided without their corresponding SEs, which makes it difficult to assess whether differences in disequilibrium intensity are occurring over pairs of loci. Therefore, the main utility of D'—that is, to compare the intensity of disequilibrium for locus pairs with different allele frequencies—is quite inoperative with this practice. Disequilibrium studies usually are restricted to the testing of the null hypothesis of gametic equilibrium, by means of different statistical tests (see Zapata and Alvarez 1997*b*), although recently a procedure based on bootstrap resampling methods has been used for testing of disequilibrium homogeneity from D' values (Zapata and Alvarez 1997*a*). In the particular context of studies of gametic disequilibrium between disease-locus alleles and alleles at adjacent marker loci (i.e., disequilibrium mapping), Lewontin's (1964) D' disequilibrium measure is, in many circumstances, essentially identical to the robust formulation of the population attributable risk for which the sampling variance is known (Devlin and Risch 1995; Devlin et al. 1996). Here, we give the expression to obtain the approximate sampling variance of D', $V(\hat{D}')$, for the case of two loci, each with two alleles, and we check it by computer simulation.

Consider two polymorphic loci, A and B, each having two alleles, A_1 and A_2 and B_1 and B_2, respectively. Suppose that n random haplotypes are sampled from the population; let the relative frequencies of the four possible haplotypes, A_1B_1, A_1B_2, A_2B_1, and A_2B_2, be x_1, x_2, x_3, and x_4, respectively, and let p, q, u, and v be the allelic frequencies of A_1, A_2, B_1, and B_2, respectively.

The measure of the magnitude of the nonrandom association between alleles at the two loci, by means of the disequilibrium coefficient, D, of Lewontin and Kojima (1960), is given by $D = x_1 - pu$. D' is the ratio of D to its theoretical maximum value, D_{max}, given the allele frequencies and the sign of D: $D' = D/D_{max}$, where D_{max} is $\min(pu, qv)$ when $D < 0$ or $\min(pv, qu)$ when $D > 0$ (Lewontin 1964).

An important problem in obtaining $V(\hat{D}')$ is that D_{max} is not a continuous function, because it is undefined for $D = 0$. Therefore, a strategy to obtain $V(\hat{D}')$ is to work with the truncated distribution of D', by the breaking of the set of permissible values of D' into two regions, $D' < 0$ and $D' > 0$. Since D' is the ratio of two random variables, $V(\hat{D}')$ for a gametic random sample taken from a population can be approximated by Taylor's series expansion (Kendall and Stuart 1977, p. 247):

$$V(\hat{D}') = V(D/D_{max}) \approx \frac{[E^2(D_{max})][V(D)] + [E^2(D)][V(D_{max})] - 2[E(D)][E(D_{max})][\mathrm{cov}(D, D_{max})]}{E^4(D_{max})},$$

where $E(\hat{D}) = D(n-1)/n$, and, in a large sample, $V(\hat{D})$ is approximated by (see Hill 1974; Brown 1975)

$$V(\hat{D}) \approx \frac{[pquv + D(q-p)(v-u) - D^2]}{n}.$$

In addition, from the first central product moment, $E(\hat{D}_{max}) = D_{max}(n-1)/n$.

The variance of D_{max}, $V(\hat{D}_{max})$, also can be approximated by Taylor's series expansion. When it is taken into account that $\mathrm{cov}(\hat{p}\hat{u})$ and $\mathrm{cov}(\hat{q}\hat{v})$ are D/n and $\mathrm{cov}(\hat{p}\hat{v})$ and $\mathrm{cov}(\hat{q}\hat{u})$ are $-D/n$ (see Hill 1974; Brown 1975),

$$V(\hat{D}_{max}) \approx \frac{D_{max}}{n}(pa + qb - 2|D|),$$

where $a = v$ and $b = u$, if $D' < 0$, and where $a = u$ and $b = v$, if $D' > 0$.

The covariance between D and D_{max}, $\mathrm{cov}(\hat{D}, \hat{D}_{max})$, can be obtained from the expression $\pm 2[\mathrm{cov}(D, D_{max})] = V(D \pm D_{max}) - V(D) - V(D_{max})$. Then,

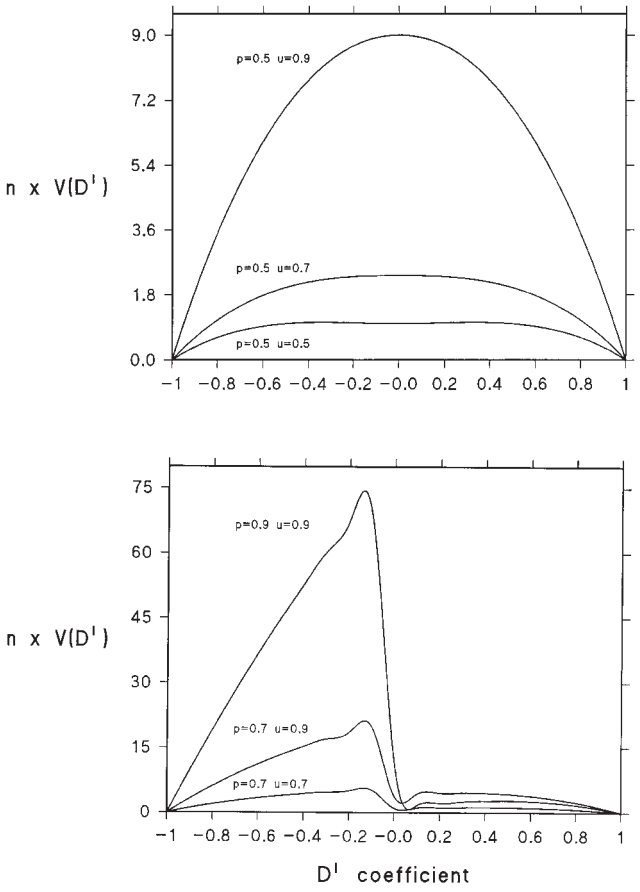

Figure 1 Product of $V(\hat{D}')$ and n, as a function of D', for different marginal p and u.

$$\text{cov}(\hat{D}, \hat{D}_{\max}) = \frac{1}{2n}\,[\text{I}(D')][x_i(1 - x_i) - n\text{V}(D) - D_{\max}(pa + qb - 2|D|)]\,,$$

where $\text{I}(D') = 1$ if $D' < 0$, $\text{I}(D') = -1$ if $D' > 0$, and x_i is x_1, x_2, x_3, and x_4 when $D_{\max}$ is *pu, pv, qu,* and *qv,* respectively.

Therefore, in a large sample, $\text{V}(\hat{D}')$ can be approximated by

$$\text{V}(\hat{D}') \approx \left[\frac{1}{n(D_{\max})^2}\right]\{(1 - |D'|)[n\text{V}(D) - |D'|D_{\max}(pa + qb - 2|D|)] + |D'|x_i(1 - x_i)\}\,.$$

When $D' = \pm 1$, then $\text{V}(\hat{D}') = 0$.

In figure 1, the product of $\text{V}(\hat{D}')$ and n is plotted against D', for six different values of p and u, for the two loci. The product of $\text{V}(\hat{D}')$ and n is given, since the sampling variances are approximately proportional to $1/n$. Given that $D_{\max}$ is undefined for $D' = 0$, we have included in figure 1 estimates of $\text{V}(\hat{D}')$ close to 0, by using very small positive and negative D' disequilibrium values ($D' = \pm .05$). It can be seen that $\text{V}(\hat{D}')$ exhibits a complex behavior as a function of D' and of p and u. In general, $\text{V}(\hat{D}')$ increases either as the magnitude of D' decreases or if the marginal allele frequencies tend to be extreme. However, $\text{V}(\hat{D}')$ does not increase monotonously for D' values around $D' = 0$, and it varies greatly with small increases in D', when allele frequencies are extreme. When both p and u differ from .5, then $\text{V}(\hat{D}')$ can vary dramatically with the sign of D', since $\text{V}(\hat{D}')$ is remarkably higher for $D' < 0$ than for $D' > 0$. This is a consequence of the asymmetry of $D_{\max}$, when p and u are different from .5 (Chakraborty 1984; Zapata and Alvarez 1993). Here, we have denoted the two gametic classes for which the expected value is more extreme than the coupling values. Otherwise, the opposite picture arises, and $\text{V}(\hat{D}')$ is less for $D' < 0$ than for $D' > 0$. This dependence of the variance on the sign of D' explains why the statistical power for the detection of negative disequilibrium in natural populations is very low, compared with that for the detection of positive disequilibrium (see Brown 1975; Zapata and Alvarez 1993).

Monte Carlo simulation was used to check the theoretical $\text{V}(\hat{D}')$. We constructed populations that were assumed to have two alleles at each locus, with different magnitudes of disequilibrium (D' varied between −.8 and +.8, in steps of .2, excluding $D' = 0$), combinations of allele frequencies at the two loci (p = .5, .7, and .9; and u = .5, .7, and .9), and haplotype sample sizes (n

Table 1

Ratio between the Asymptotic V(*D*′) and the Estimated Variance from Computer Simulation, for Different Values of *D*′, *p* and *u* for the Loci, and *n*

		Ratio, for D' =							
p at n =	u	−.8	−.6	−.4	−.2	.2	.4	.6	.8
n = 50:									
.5	.5	1.16	1.06	.92	.66	.54	.87	1.25	1.62
	.7	1.27	.88	1.08	1.42	.97	.85	.95	1.20
	.9	.86	.95	.75	.87	.74	.94	.52	.61
.7	.7	1.17	.96	1.01	1.16	.45	.88	1.15	1.13
	.9	1.39	1.22	1.57	1.37	.46	.61	.53	1.00
.9	.9	7.36	3.92	4.03	3.56	.11	.31	.28	1.84
n = 100:									
.5	.5	1.60	1.07	.93	.70	.87	1.03	.98	1.53
	.7	.93	1.11	1.26	1.04	.99	1.11	.97	1.05
	.9	1.05	.82	.79	.90	.77	.78	.84	.79
.7	.7	.96	1.00	.95	1.18	.84	1.58	1.08	.92
	.9	.98	1.19	1.15	1.55	.49	.65	1.08	1.06
.9	.9	1.71	1.88	2.19	2.54	.21	.95	1.01	1.67
n = 1,000:									
.5	.5	1.50	1.19	1.04	.96	.95	1.12	1.31	1.49
	.7	.99	1.07	1.07	1.00	1.01	1.01	1.00	1.03
	.9	.87	.96	1.09	1.01	.99	.93	1.02	.94
.7	.7	1.08	1.03	.98	.98	.94	1.06	1.23	1.45
	.9	1.00	1.05	1.02	1.09	.97	1.06	1.06	.90
.9	.9	.97	.98	1.13	1.46	.99	1.00	1.14	1.29

= 50, 100, and 1,000). Each population was sampled, and D' was calculated. This sampling was repeated 1,000 ×, and the mean and variance for D' were calculated. Table 1 shows the ratio between the theoretical values of $V(\hat{D}')$ and the variances in the computer simulation. It is clear that, in general, the approximation of the theoretical $V(\hat{D}')$ is quite satisfactory and that the ratio approximates to 1 quite well, even for samples as small as $n = 100$. As expected from asymptotic theory, most of the significant differences between the two variances, detected by the F_{max}-statistic test (Sokal and Rohlf 1995, p. 397), occur for $n = 50$, especially for extreme allele frequencies. From the results, use of $V(\hat{D}')$ for experimental sample sizes equal to or higher than $n = 100$ can be recommended.

CARLOS ZAPATA,[1] GONZALO ALVAREZ,[1] AND CARMEN CAROLLO[2]

[1]*Departmento de Biología Fundamental and* [2]*Departamento de Estadística e I. O., Universidad de Santiago, Santiago de Compostela, Spain*

Acknowledgments

This work was supported by Xunta de Galicia (of Spain) grant XUGA-20002B95 (to C.Z).

References

Brown AHD (1975) Sample sizes required to detect linkage disequilibrium between two or three loci. Theor Popul Biol 8:184–201

Chakraborty R (1984) Detection of nonrandom association of alleles from the distribution of the number of heterozygous loci in a sample. Genetics 108:719–731

Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29:311–322

Devlin B, Risch N, Roeder K (1996) Disequilibrium mapping: composite likelihood for pairwise disequilibrium. Genomics 36:1–16

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

——— (1988) Inference of recombinational hotspots using gametic disequilibrium values. Heredity 60:435–438

Hedrick PW, Jain S, Holden L (1978) Multilocus systems in evolution. Evol Biol 11:101–184

Hedrick PW, Thomson 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

Kendall M, Stuart A (1977) The advanced theory of statistics. Vol 1. Charles Griffin, London

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49: 49–67

——— (1988) On measures of gametic disequilibrium. Genetics 120:849–852

Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. Evolution 14:458–472

Sokal RR, Rohlf FJ (1995) Biometry. WH Freeman, New York

Zapata C, Alvarez G (1992) The detection of gametic disequilibrium between allozyme loci in natural populations of *Drosophila.* Evolution 46:1900–1917

——— (1993) On the detection of nonrandom associations between DNA polymorphisms in natural populations of *Drosophila.* Mol Biol Evol 10:823–841

——— (1997*a*) Testing for homogeneity of gametic disequilibrium among populations. Evolution 51:606–607

——— (1997*b*) On Fisher's exact test for detecting gametic disequilibrium between DNA polymorphisms. Ann Hum Genet 61:71–77

Zapata C, Visedo G (1995) Gametic disequilibrium and physical distance. Am J Hum Genet 57:190–191

Address for correspondence and reprints: Dr. Carlos Zapata, Departamento de Biología Fundamental, Area de Genética, Facultad de Biología, Universidad de Santiago, Santiago de Compostela, Spain. E-mail: bfcazaba@usc.es

0002-9297/97/6103-0037$02.00

Am. J. Hum. Genet. 61:774–778, 1997

Transmission/Disequilibrium Tests for Multiallelic Loci

To the Editor:

Kaplan et al. (1997) address the interesting question of how the biallelic transmission/disequilibrium test (TDT) should be extended to multiallele loci. Four recently proposed test statistics were described, and their properties were investigated by simulation studies. Here, I would like to point out some defects of the Monte Carlo–T_m test and the χ^2-$T_{m\text{het}}$ test that were not revealed by these simulation studies.

All four test statistics are based on the square contingency table of the counts of allele transmission, as set out in table 1. The cell count n_{ij} is the number of parents

Table 1

Counts of Allele Transmission and Nontransmission

Transmitted Allele	Nontransmitted Allele				
	1	2	..	m	Total
1	n_{11}	n_{12}	..	n_{1m}	$n_{1.}$
2	n_{21}	n_{22}	..	n_{2m}	$n_{2.}$
.	..	..	..	..	
m	n_{m1}	n_{m2}	..	n_{mm}	$n_{m.}$
Total	$n_{.1}$	$n_{.2}$	..	$n_{.m}$	$n_{..}$