Genetic Structure and the Search for Genotype-Phenotype Relationships: An Example from Disequilibrium in the Apo B Gene Region

Kim E. Zerba,* Anna M. Kessling,† Jean Davignon‡ and Charles F. Sing*

*Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109, †Department of Biochemistry and Molecular Genetics, St. Mary's Hospital Medical School, London W2 1PG, United Kingdom, and ‡Hyperlipidemia and Atherosclerosis Research Group, Clinical Research Institute of Montreal, Montreal, Quebec H2W 1R7, Canada

Manuscript received January 11, 1991
Accepted for publication June 8, 1991

ABSTRACT

We analyzed allelic associations (disequilibria) for four restriction fragment length polymorphisms (RFLPs) in the region of the 43-kb Apo B gene in a sample of 233 unrelated individuals from Montreal, Canada, sampled for health. This total sample (T) included 160 individuals of known French Canadian (FC) ancestry. We present a rigorous application of current methodology to these samples, including estimation of type II error probabilities and correlations between markers for estimates of disequilibria. We then consider the utility of these estimates of allelic disequilibria for the interpretation of genotype-phenotype relations. Significant deviations from Hardy-Weinberg equilibrium were not predicted by proximity to other markers in disequilibrium. We found significant quadri-allelic disequilibrium for two marker pairs despite absence of significant deviations from Hardy-Weinberg equilibrium for either marker or tri-allelic disequilibrium, respectively. Altogether these results underscore the complexity of the genotypic structure of the data. A combination of nonevolutionary factors, including sampling for health, small sample size and data exclusion due to methodological constraints of not successfully typing all members of the sample for every RFLP, is a likely explanation for this complexity. These types of factors are common to many RFLP studies. Patterns of composite di-allelic disequilibrium indicated that some RFLP allele pairs may have a longer shared evolutionary history than others and that disequilibrium is not predicted by distance between RFLPs. Type II error probabilities were generally much higher than those for type I errors. Correlations between marker pairs for disequilibria were generally not high. We show from a review of 14 published studies of association between the XbaI RFLP and variation in a total of 15 lipid traits that deviations from Hardy-Weinberg equilibrium can cause substantial differences in the estimation of variability associated with phenotypic differences among marker genotypes relative to Hardy-Weinberg conditions.

MARKERS such as restriction fragment length polymorphisms (RFLPs) are in widespread use to identify DNA sequences containing mutations responsible for quantitative phenotypic variation in a population. When multiple markers of a gene region are available, they are commonly used individually in separate analyses of variance (ANOVAs) to detect associations of marker genotypes with phenotypic variation. Inferences are then made about an unknown functional mutation with a phenotypic effect, which is nonrandomly associated with marker alleles, considering all single-marker analyses simultaneously. This approach implicitly assumes the marker allele frequencies are independent. Inferences from this analytical strategy ignore the evolutionary history shared by the markers.

Evolutionary history events, including admixture, finite population size, selection, migration and mutation, determine the genetic structure defined by frequency associations among marker alleles on the same (gametic allele associations) and different homologous chromosomes (nongametic allele associations). Sampling design factors, including nonrandom sampling, small sample size and data exclusion due to methodological constraints, represent an additional, very important and largely unrecognized class of determinants of such allele frequency associations. Whichever the cause(s), nonrandom marker allele frequency associations (allelic disequilibria) result in two major problems for ANOVA studies of marker-phenotype associations.

First, the presence of gametic allele associations means that inferences from analyses of markers are interdependent (TEMPLETON, BOERWINKLE and SING 1987; TEMPLETON et al. 1988). In addition, since evolutionary history events are infrequent and unpredictable in time and in chromosomes they affect, a predictable relationship is not expected between the degree of gametic allele association and physical distance. This unpredictability applies to associations among marker alleles and between alleles of markers and the mutation(s) responsible for a phenotypic effect. Consequently, specificity of information from individual markers about the location of a mutation
with a phenotypic effect is not possible. Where there is gametic allelic disequilibrium, inferences from multiple single-marker analyses are limited to showing whether or not the region being marked is involved with the phenotype (Kessling et al. 1991).

Second, nongametic allelic disequilibrium will affect inferences about association of markers with phenotypic variation. Phenotypic variation associated with marker allelic variation is dependent on three factors related to frequency and scale: (1) the frequency distribution of marker genotypes; (2) the association between a mutation with a phenotypic effect and marker alleles; and (3) the impact of the mutation on the phenotype (scale). Inferences about associations of single markers with phenotypic variation are often made assuming that genotypes are in Hardy-Weinberg proportions in the population. If it is assumed that the "truth" is represented by genotypes in Hardy-Weinberg proportions, then deviation of marker genotypes from Hardy-Weinberg expected values results in differences in the estimation of phenotypic variation associated with a single marker relative to these conditions.

The first step in dealing with these issues is to understand the genetic structure of marker variation. The strength of inferences about genetic structure becomes especially important because this information is critical in the proper design of statistical analyses to locate mutations with phenotypic effects. Power to detect disequilibrium given sample sizes used in most studies may be low (Ward and Sing 1970; Thompson et al. 1988). Thus, the complement to power, i.e., the type II error probability of not detecting disequilibrium when it is present, may be high. Type II errors may also play an important role in explaining discordant inferences about genetic structure of a region, estimated by different studies using the same set of markers. In addition, there is the concern regarding correlated type I and type II errors of inference resulting from estimating many allelic associations from the same set of data.

This report focuses on an analysis of the genetic structure of marker alleles in the Apolipoprotein B (Apo B) gene region in humans. The DNA sequence encoding the Apo B gene is on the short arm of chromosome 2, 43 kb in length and includes 29 exons (Blackhart et al. 1986). Apo B plays a central role in assembly of triglyceride-rich lipoproteins (Leighton et al. 1990), is a major component of low density lipoproteins (LDL) and the ligand for the LDL receptor (Brown and Goldstein 1986). LDL is a major transporter of cholesterol in plasma. It is well established that elevated LDL cholesterol levels convey increased risk of coronary artery disease (CAD). Thus apolipoprotein B is a candidate for a role in the pathogenesis of CAD (Sniderman and Silberberg 1990).

We evaluate the genetic structure defined by allelic associations at each and between pairs of four RFLPs in the Apo B region. We present a rigorous application of current methodology to this data set including estimation of type II error probabilities and correlations between markers for estimates of allelic disequilibria. We then consider the utility of these estimates of genetic structure for the interpretation of marker genotype-phenotype relationships. We show from a review of 14 published studies of association between the Apo B XbaI RFLP and variation in a total of 15 lipid traits that deviations from Hardy-Weinberg equilibrium can cause substantial differences in the estimation of variability associated with phenotypic scale differences among marker genotypes relative to Hardy-Weinberg conditions.

MATERIALS AND METHODS

Sample: Unrelated white-collar workers, aged 20 to 59 years from the Hydro-Quebec power company of Montreal, Canada, were sampled according to health status by the Hyperlipidemia and Atherosclerosis Research Group at the Clinical Research Institute of Montreal to represent a population free of clinical manifestations of disease, including cardiovascular disease, hypertension, chronic disorders requiring medication, previously diagnosed thyroid dysfunction, diabetes, gout, renal dysfunction or hyperlipidemia (Lussier-Cacan, unpublished data). There were 233 individuals with typings for all four Apo B RFLPs; 146 male and 87 female. This was the total (T) sample used for this study. Of these, 160 individuals were of known French Canadian (FC) ancestry (Kessling et al. 1991). These two samples were part of an overall sample of 347 individuals on which RFLP typings were attempted. The difference between this larger sample and the T sample used in this study is the result of data exclusion due to methodological constraints. Individuals were not included if they could not be typed for an RFLP because independent assessments by three investigators did not agree on the typing from reading the autoradiograms. In common with other RFLP studies, typings were repeated where DNA was sufficient but rejected where DNA was insufficient for a repeat of the assay.

Molecular genetic analyses: Restriction endonuclease analyses defining the four markers in this region for the sample have been described (Kessling et al. 1991). The Apo B gene region and location of RFLPs is depicted in Figure 1. Two of the four polymorphisms are near the 5' end with the others near the 3' portion of the Apo B region. The HincII and PvuII polymorphisms are in intron 4, 171 bp 3' to exon 4 and within an Alu sequence 523 bp 5' to exon 5 (Huang, Ripps and Breslow 1990), respectively. Both are restriction site polymorphisms resulting from single base substitutions. The XbaI polymorphism is in exon 26 and changes the third base of codon 2488, but does not alter an amino acid (Berg et al. 1986). The EcoRI polymorphism is

![Figure 1.—The Apo B gene region and location of RFLPs. Filled bars represent exons; open bars represent introns.](image-url)
in exon 29. The presence of the EcoRI variable restriction site alters glutamic acid to lysine (SHOULders et al. 1985) and is associated with the Ag antisera type t/z (MA et al. 1987).

**Statistical analyses:** In the following expressions we employ the notation of WEIR and Cockerham (1989). The most frequent alleles of a pair of RFLPs are designated A and B, respectively. Genotype frequency estimates were the observed 12 single-marker and 54 two-marker genotype frequencies. For each pair of marker loci, nine two-marker genotypes were recognized; double heterozygotes were not distinguished. Allele frequencies were estimated by gene counting. Variances of allele frequencies were determined according to WEIR (1990). Using an approach and methodology outlined by WEIR and Cockerham (1989), for each RFLP, we estimated deviations of the three genotype frequencies from those expected at Hardy-Weinberg equilibrium \((D_A)\). For pairs of RFLPs we estimated composite di-allelic \((D_{AB})\), tri-allelic \((D_{AB}, D_{AB})\) and quadri-allelic \((D_{AB})\) disequilibria. We also estimated variances for each of these disequilibria.

Bounds of all these disequilibrium estimates are affected by allele frequencies (HEDRICK 1987; LEWONTIN 1988). Disequilibrium estimates and their variances other than \(D_A\) are affected by other disequilibria estimates (WEIR and Cockerham 1989). We therefore determined the limits of each of the observed disequilibria and the percentage of each divided by its limit. Bounds for \(D_{AB}\) were determined according to WEIR and BROOKS (1986). Since \(D_{AB}\) is a composite measure of two disequilibria having the same limits, we determined the bounds for \(D_{AB}\) as twice values for \(D_{AB}\) defined by WEIR and Cockerham (1989). Bounds for tri-allelic disequilibria were determined according to WEIR and Cockerham (1989). Bounds for quadri-allelic disequilibria have not been derived and hence were not estimated.

For each disequilibrium estimate, we tested the single degree of freedom null hypothesis that the disequilibrium parameter \(\delta = 0\) by

\[
\chi^2 = \frac{\delta^2}{\text{var}(\delta)} \tag{1}
\]

where \(\delta\) represents the disequilibrium estimated and \(\text{var}(\delta)\) represents the variance of \(\delta\) estimated under the null hypothesis with the \(\delta\) set to zero and all other parameters at the respective point estimates. \(\chi^2\) in (1) is approximately distributed as chi-square with one degree of freedom. We use the conventional statistical significance level \(\alpha = 0.05\) for single tests throughout.

We used computer simulation to determine the type II error probabilities of not detecting disequilibrium of the size determined by the observed allele and genotype frequencies and sample sizes. We constructed T and FC populations with the relative genotype and allele frequencies and hence the disequilibria observed in our samples. These observed disequilibria served as the alternatives to the null hypothesis of disequilibrium equal to zero for the type II error simulations. Then, for each disequilibrium, 10,000 replicate random samples were drawn from each of these populations with replacement. Samples were of size 233 or 160 depending on whether we were considering the T sample or FC subsample. For each replicate sample we calculated the disequilibrium coefficient of interest and tested the null hypothesis using the strategy described above. Type II error probability was determined as the proportion of the simulations in which the chi-square was less than or equal to the critical value 3.84 for a single degree of freedom test at \(\alpha = 0.05\). Altogether, 28 type II error probability simulations were performed for the estimated disequilibria in our samples.

**Figure 2.** Impact of deviations from Hardy-Weinberg equilibrium on estimation of sum of squares variability associated with marker genotypes. Relationship in figure based on review of 14 published association studies of a total of 15 lipid traits with the Apo B XbaI RFLP as marker (Law et al. 1986; COCOZZA et al. 1987; Talmud et al. 1987; Alto-Setala et al. 1988, 1989; Dunnig et al. 1988; Leren et al. 1988; Monsalve et al. 1988; Demant et al. 1988; Houlston et al. 1988; Darnfors et al. 1989; Myant et al. 1989; Wulkund et al. 1989; Paulweber et al. 1990).

Since all of the tests for disequilibria were performed on the same data set, we were concerned with the potential for correlated type I (SOKAL and ROLFL 1981) and correlated type II errors of inference. We evaluated such potential correlated errors of inference using bootstrap techniques (Efron 1982; WEIR and Brooks 1986) to estimate the correlation between a pair of markers or two pairs of markers for the respective allelic disequilibria. A substantial positive correlation would support the potential for correlated errors of inference. We focused on the correlations between RFLPs for \(D_A\) in the T sample and FC subsample and between different pairs of RFLPs for composite measures of di-allelic disequilibrium in the FC subsample.

We estimated the effects of adjusting critical \(\alpha\) levels for multiple tests on type II error rates. We estimated the overall probability of at least one type II error for families of tests of \(D_A\) in the T and FC subsamples and \(D_{AB}\) in the FC subsample. For each family of tests we simultaneously estimated all respective disequilibria for each of the 10,000 bootstrap replicate samples. Type II error probability was determined as the proportion of simulations in which the chi-square was less than or equal to the critical value for any of the disequilibria. For each family of tests the simulations were run, using an individual test \(\alpha = 0.05\) and an experiment-wide \(\alpha = 0.05\).

We reviewed the results of 14 published studies of association between variability in a total of 15 lipid traits and the Apo B XbaI RFLP (references under Figure 2). For each study we estimated the \(D_A\). We then estimated the sum of squares variability associated with the XbaI RFLP using

\[
\text{SSR} = \sum f_g (\bar{Y}_g - \bar{Y})^2, \tag{2}
\]

where \(f_g\) is the frequency of the \(g\)th genotype, \(\bar{Y}_g\) the \(g\)th genotype mean and \(\bar{Y}\) the grand mean of the sample. We estimated SSR in two ways: (1) using the observed \(f_g\) (SSRobs), and (2) using Hardy-Weinberg equilibrium expected \(f_g\) (SSRexp). We estimated the percent difference in estimating SSR relative to genotype proportions under Hardy-Weinberg conditions by

\[
((\text{SSR}_{\text{obs}} - \text{SSR}_{\text{exp}})/\text{SSR}_{\text{exp}}) \times 100. \tag{3}
\]
RESULTS

Common allele frequencies and variances for each RFLP for the T sample and FC subsample are in Table 1. RFLPs are listed in order 5' to 3'. Relative frequencies of the common allele ranged from 0.5 for XbaI to 0.9 for PvuII. They were not significantly different between the two samples. Variance of allele frequencies were larger in the FC subsample, reflecting its smaller sample size.

Estimates of $D_A$, corresponding variances, percent of maximum possible $D_A$, $\chi^2$ tests and type II error probabilities are also presented in Table 1. Estimates of $D_A$ were significant for: the HincII and XbaI RFLPs in the T sample, but not in the FC subsample. In contrast, the estimate of $D_A$ was not significant for PvuII RFLP in the T sample, but was significant in the FC subsample. As expected, variances of all $D_A$ estimates were larger in the smaller FC subsample. None of the estimates of $D_A$ were at or near bounds set by allele frequencies. Type II error probability over all tests ranged from 0.43 to 0.89. For the three tests where significant $D_A$ was detected in the actual samples, the type II error probability was 8.5 to 10.0 times higher (0.43–0.50) than the type I error probability given by the nominal level of significance $\alpha = 0.05$ for a single test for the same three estimates.

Estimates of two-marker disequilibria, variances, percent of maximum possible disequilibrium, $\chi^2$ tests and type II error probabilities are in Table 2. The variances from the T sample and FC subsample were the same for 22 of the 24 tests. However, because of differences in significant $D_A$ between the two samples, we present estimates only of two-marker disequilibria in the FC subsample. Composite measures of di-allelic disequilibria were significant for four of six pairs of markers; $\Delta_{AB}$ for the HincII RFLP paired with the two RFLPs at the other end of the region (XbaI and EcoRI) was not significant in either case. We note that $\Delta_{AB}$ for the PvuII RFLP adjacent to the HincII RFLP was significant for all pairings with the other three markers. None of these composite estimates was at the bounds set by allele frequencies. None of the tri-allelic measures was significant and 3/12 were at the maximum possible values. HincII-XbaI and XbaI-EcoRI were the only two of the six marker pairs which gave evidence for significant quadri-allelic disequilibrium. Type II error probabilities over all tests ranged from 0.0 to 0.9989. We note that for the three non-significant estimates of tri-allelic disequilibria at bounds (100% maximum) set by allele frequencies, type II error probability was very high. In addition, two of the three bounded tri-allelic disequilibria are from the PvuII-EcoRI RFLP pair. The estimate of quadri-allelic disequilibrium depends, in part, on these tri-allelic measures. Type II error probability for PvuII-EcoRI quadri-allelic disequilibrium was the highest of all the tests (0.9989). Furthermore, for the six two-marker tests in which disequilibrium was detected, type II error probabilities ranged from 4.5 to 10.4 times higher (0.22–0.52; three tests) to over 100 times lower (0.0005–0; two tests) than the fixed $\alpha = 0.05$ probability of type I error for each of the same six tests.

Bootstrap estimates of correlations between the RFLPs for estimates of $D_A$ in the T sample and FC subsample are in Table 3. Estimated variances are on the diagonal of Table 3, with T above FC. These bootstrap variances were nearly the same as those estimated in Table 1 from the theoretical equations. Correlations for the T sample are in the upper diagonal and for the FC subsample in the lower diagonal of Table 3. All correlations were positive and low ($r \leq 0.27$). These results indicate the probability of correlated errors of inferences about $D_A$ is low for this data set.

Bootstrap estimates of correlations between the estimates of two-marker composite measures of di-allelic disequilibria for the FC subsample are in Table 4. The bootstrap estimated variances were again nearly the same as those estimated from the theoretical equations in Table 2. The correlation structure among the $\Delta_{AB}$ was more complex than among the $D_A$ with not
Disequilibrium in the Apo B Gene

TABLE 2

Two-marker disequilibria in French Canadians

<table>
<thead>
<tr>
<th>RFLP pair</th>
<th>Disequilibrium</th>
<th>Variance × 10^{-4}</th>
<th>Percent maximum</th>
<th>x²</th>
<th>Pr (x²)</th>
<th>Pr (type II error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HincII-PvuII</td>
<td>ΔAB</td>
<td>0.05</td>
<td>1.9</td>
<td>38.1</td>
<td>19.54</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>-0.006</td>
<td>0.2</td>
<td>50.2</td>
<td>1.80</td>
<td>0.1796</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>0.006</td>
<td>0.3</td>
<td>50.3</td>
<td>1.37</td>
<td>0.2423</td>
</tr>
<tr>
<td></td>
<td>Δ_{AABB}</td>
<td>0.002</td>
<td>&lt;0.1</td>
<td>0.33</td>
<td>0.5641</td>
<td>0.9381</td>
</tr>
<tr>
<td>HincII-XbaI</td>
<td>ΔAB</td>
<td>-0.02</td>
<td>4.0</td>
<td>7.7</td>
<td>1.29</td>
<td>0.2570</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>-0.001</td>
<td>0.4</td>
<td>2.4</td>
<td>0.05</td>
<td>0.8264</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>-0.006</td>
<td>0.4</td>
<td>6.8</td>
<td>0.70</td>
<td>0.4042</td>
</tr>
<tr>
<td></td>
<td>Δ_{AABB}</td>
<td>0.01</td>
<td>0.2</td>
<td>6.55</td>
<td>0.0105</td>
<td>0.2893</td>
</tr>
<tr>
<td>HincII-EcoRI</td>
<td>ΔAB</td>
<td>0.02</td>
<td>2.6</td>
<td>8.6</td>
<td>1.82</td>
<td>0.1778</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>0.002</td>
<td>0.3</td>
<td>7.9</td>
<td>0.12</td>
<td>0.7545</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>-0.006</td>
<td>0.3</td>
<td>20.5</td>
<td>1.79</td>
<td>0.1808</td>
</tr>
<tr>
<td></td>
<td>Δ_{AABB}</td>
<td>0.002</td>
<td>0.1</td>
<td>0.24</td>
<td>0.6235</td>
<td>0.9384</td>
</tr>
<tr>
<td>PvuII-XbaI</td>
<td>ΔAB</td>
<td>-0.03</td>
<td>1.6</td>
<td>34.3</td>
<td>6.4</td>
<td>0.0114</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>0.006</td>
<td>0.2</td>
<td>65.7</td>
<td>1.9</td>
<td>0.1680</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>-0.001</td>
<td>0.2</td>
<td>3.4</td>
<td>0.1</td>
<td>0.8118</td>
</tr>
<tr>
<td></td>
<td>Δ_{AABB}</td>
<td>0.002</td>
<td>&lt;0.1</td>
<td>1.1</td>
<td>0.2950</td>
<td>0.7798</td>
</tr>
<tr>
<td>PvuII-EcoRI</td>
<td>ΔAB</td>
<td>-0.03</td>
<td>0.5</td>
<td>81.6</td>
<td>8.0</td>
<td>0.0047</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>0.002</td>
<td>&lt;0.1</td>
<td>100.0</td>
<td>0.5</td>
<td>0.4677</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>&lt;0.001</td>
<td>&lt;0.1</td>
<td>100.0</td>
<td>&lt;0.1</td>
<td>0.9977</td>
</tr>
<tr>
<td></td>
<td>Δ_{AABB}</td>
<td>-0.0004</td>
<td>&lt;0.1</td>
<td>0.2</td>
<td>0.6714</td>
<td>0.9989</td>
</tr>
<tr>
<td>XbaI-EcoRI</td>
<td>ΔAB</td>
<td>-0.07</td>
<td>2.5</td>
<td>43.6</td>
<td>21.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>0.002</td>
<td>0.3</td>
<td>100.0</td>
<td>0.1</td>
<td>0.7625</td>
</tr>
<tr>
<td></td>
<td>D_{AB}</td>
<td>0.008</td>
<td>0.2</td>
<td>72.7</td>
<td>2.3</td>
<td>0.1285</td>
</tr>
<tr>
<td></td>
<td>Δ_{AABB}</td>
<td>0.005</td>
<td>&lt;0.1</td>
<td>4.07</td>
<td>0.0437</td>
<td>0.5196</td>
</tr>
</tbody>
</table>

TABLE 3

Bootstrap variances of and correlations between RFLPs for Da in the total sample and French Canadian subsample

<table>
<thead>
<tr>
<th></th>
<th>HincII</th>
<th>PvuII</th>
<th>XbaI</th>
<th>EcoRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HincII</td>
<td>2.1</td>
<td>0.12</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>PvuII</td>
<td>2.9</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>XbaI</td>
<td>0.21</td>
<td>0.27</td>
<td>1.8</td>
<td>3.8</td>
</tr>
<tr>
<td>EcoRI</td>
<td>0.06</td>
<td>0.01</td>
<td>0.27</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Variances (× 10^{-7}) are in italics on the diagonal with T above FC. Correlations for the T and FC samples are given in the upper and lower diagonals, respectively. Bootstrap samples were 3000 for each of the T and FC samples.

Estimates of at least one type II error in a family of tests were 0.9819, 0.9976 and 0.9948 for Da in the T and FC subsamples and ΔAB in the FC subsample, respectively. Adjusting critical α levels for multiple tests inflated these estimates of type II error rates to 0.9988, 0.9999 and 0.9942, respectively.

Assuming the “truth” is represented by genotypes in Hardy-Weinberg proportions, the relationship between percent difference in estimation of SSR associated with the Apo B XbaI RFLP relative to Hardy-Weinberg conditions for 14 published studies is shown in Figure 2 (references under Figure 2). The relationship is linear and predictable. Deviation of individual study points from linearity are due to differences among studies in the relative position of the heterozygote mean between the homozygotes and differences in allele frequencies. When the mean for the heterozygote falls outside of the range determined by the homozygotes the percent sum of squares difference relationship with Da is still predictable, but more complex and nonlinear (K. E. Zerba, unpublished results). We included results only from studies for traits where the heterozygote mean was estimated as between the homozygotes. As evident from Figure 2,
there can be substantial difference in the estimation of trait variability associated with the Apo B XbaI RFLP relative to Hardy-Weinberg conditions, ranging from as much as 34% higher to 22% lower.

**DISCUSSION**

The pattern of significant $D_A$ among the four RFLPs in each of the T and FC samples illustrates that $D_A$ for a marker is not predicted by proximity to other markers in disequilibrium. This is a remarkable result given the high degree of composite di-allelic disequilibrium among most pairs of markers in this region. The differences between the T sample and FC subsample in significant estimates of $D_A$ were also surprising given the RFLP allele frequencies were not different between the two samples. These differences in $D_A$ may be partly explained by differences in genotypic structure between the samples of French Canadians and other populations of origin that comprise the T sample. Smaller sample size of the FC subsample may also explain some of the differences. Moreover, we found significant estimates of quadri-allelic disequilibrium for two marker pairs in the FC subsample, despite none of the single markers from these pairs and neither pair showing evidence for significant estimates of $D_A$ or tri-allelic disequilibrium, respectively. We suggest that chance sampling associated with small sample size and data exclusion due to methodological constraints at the RFLP typing stage of analysis probably have very important roles in estimates of $D_A$ and quadri-allelic disequilibrium. For these obvious reasons inferences about the genetic structure of the original sample of healthy individuals from Hydro-Quebec cannot be made from this study. Sampling design factors including samples consisting of multiple strata with respect to population of origin or race, nonrandom sampling for disease or health, small sample size and data exclusion due to methodological constraints at the RFLP typing level are features common to many RFLP studies. These results underscore the complex genotypic structure that may occur in any particular set of data and force reconsideration of the meaning of such concepts as Hardy-Weinberg equilibrium as conventionally considered for a gene region as a whole. The implications of such complex genotypic structure for construction of meaningful population genetic models offers an extraordinary challenge.

Many studies of genetic structure have considered only gametic di-allelic (linkage) disequilibrium (Chakravarti et al. 1984; but see Chakravarti et al. 1986; Weir and Hill 1986), ignoring the potential for other types of nonrandom allelic associations (Sinnock and Sing 1972; Weir 1979, Weir and Cockermham 1989) that may influence the variance and hence significance of measures of linkage disequilibrium. Furthermore, haplotype frequencies are often estimated from non-family data assuming Hardy-Weinberg equilibrium, but not considering the possible presence of $D_A$. Weir and Brooks (1986) and Haviland et al. (1991) present the only other published applications of this more comprehensive approach of characterizing disequilibrium to human marker data sets.

Patterns of composite di-allelic disequilibria we observed in the FC subsample suggests that certain alleles for these RFLPs are often packaged together on individual chromosomes. The PvuII RFLP was in significant disequilibrium with all three other RFLPs. However, even though PvuII was in significant disequilibrium with the adjacent HindII RFLP, there was nonsignificant disequilibrium estimated when HindII was paired with the other two markers at the other end of the region. These results indicate HindII may have had a longer shared evolutionary history with XbaI and EcoRI than with PvuII. However, a role for small sample size and data exclusion due to methodological constraints in determining these patterns is also a possibility.

Jenner et al. (1988) found no evidence for disequilibrium between XbaI and EcoRI. However, not detecting significant disequilibrium does not imply it is absent in the population (Cox, Bell and Xiang 1988). There are at least four explanations for differences between our study and that of Jenner et al. (1988). First, the samples are from different populations (French Canadian vs. persons living in England) with possibly different evolutionary histories. This is supported by differences between the respective stud-

---

**TABLE 4**

Bootstrap variances of and correlations between RFLP pairs for $\Delta AB$ in French Canadians

<table>
<thead>
<tr>
<th></th>
<th>HindII-PvuII</th>
<th>HindII-XbaI</th>
<th>HindII-EcoRI</th>
<th>PvuII-XbaI</th>
<th>PvuII-EcoRI</th>
<th>XbaI-EcoRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HindII-PvuII</td>
<td>1.9</td>
<td>-0.40</td>
<td>-0.28</td>
<td>-0.66</td>
<td>-0.45</td>
<td>0.24</td>
</tr>
<tr>
<td>HindII-XbaI</td>
<td>4.0</td>
<td></td>
<td>-0.39</td>
<td>0.42</td>
<td>0.18</td>
<td>0.01</td>
</tr>
<tr>
<td>HindII-EcoRI</td>
<td>2.6</td>
<td>0.15</td>
<td>0.08</td>
<td>-0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PvuII-XbaI</td>
<td>1.6</td>
<td>0.20</td>
<td>0.5</td>
<td>-0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PvuII-EcoRI</td>
<td></td>
<td>0.5</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XbaI-EcoRI</td>
<td></td>
<td></td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variances ($\times 10^{-5}$) are in italics on the diagonal. Bootstrap sample was 3000.
EcoRI: did not consider this possibility. Fourth, the JENNER probability of making at least one type I error for the equilibrium. Third, disequilibria can be at maximal possible values and not significant. The JENNER I1 error probability with regard to detecting disequilibrium because of unknown allelic phase relationships, which could have affected not only allele frequencies, but also the estimate of disequilibrium.

Our results also indicate no direct relationship between physical distance between markers and degree of composite di-allelic disequilibrium. In theory, increasing distance between markers is accompanied by increased probability of recombination over time and subsequent decreases in gametic disequilibrium relationships with time. This relationship does not appear to hold in short DNA regions where the force of recombination on haplotype organization may be similar to or less than the force of mutation (LITT and JORDE 1986; THOMPSON et al. 1988; BØRRESEN, MØLLER and BERG 1988).

Most studies of genetic structure have focused on fixed a levels of 0.05 for individual tests. Other studies have recognized that many tests are being repeated on the same set of data with the potential for correlated type I errors and thus lower individual test a levels such that an experimentwise a level 0.05 is attained (HEGELE, PLAETKE and LALOUEL 1990). In contrast, WARD and SING (1970) showed huge sample sizes are needed to detect significant $D_A$ at levels normally observed in most studies of human populations. THOMPSON et al. (1988) also showed, for linkage disequilibrium statistics, that large sample sizes are needed to detect disequilibrium, especially for negative linkage disequilibrium. We considered the very important complement to power, i.e., the type II error probability, for each disequilibrium estimate from the samples considered. Our computer simulation results showed, for most individual tests that detected presence of significant $D_A$ and two-marker disequilibrium (given the observed sample sizes and allele frequencies), the type II error probability can be much greater than the fixed type I error, a, probability of 0.05.

We can compare estimates of the overall probability of making at least one type I error vs. that of a type II error for a family of tests of disequilibrium. A standard probability argument is used to estimate an overall type I error probability $a'$ for a family of $k$ tests,

$$a' = 1 - (1 - a)^k.$$  \hfill (4)

For example, given the conventional $a = 0.05$, the probability of making at least one type I error for the four tests of $D_A$ in the T or FC samples is $a' = 0.19$. For the six tests of $\Delta_{AB}$ in the FC subsample $a' = 0.26$.

A similar argument can be made regarding the probability of making at least one type II error for a family of $k$ tests ($b'$). Our simulations indicated the type II error probabilities for these same three families of tests were quite high ($b' > 0.9$). These arguments assume the tests are independent, which of course they are not, given the bootstrap estimates of correlation structure among the tests presented in Tables 3 and 4. Furthermore, the type II error probabilities are a posteriori estimates, whereas type I error probability $a$ is determined a priori. Our arguments do, however, provide crude estimates of the relative differences in probabilities of the two types of errors for sets of tests.

Extending this argument, to achieve an experimentwise $a' = 0.05$, the individual test $a$ is lowered. Our results showed, however, that using an experimentwise $a' = 0.05$ for families of tests to lower the type I error rate simultaneously inflates already high type II error rates. Given these results and the bootstrap estimates of relatively low probabilities of both correlated type I and type II errors for this data set, we must conclude, for disequilibrium analyses of this data set and others like it, that using experimentwise type I error rates results in unacceptable experimentwise type II error rates. Using an experimentwise $a' = 0.05$ for the family of tests of $\Delta_{AB}$ in Table 2 would mean that the value for PvuII-XbaI is now not significant. It is true that this value may not represent real disequilibrium and could be significant by chance alone. It is also true, however, that any of the nonsignificant values for $\Delta_{AB}$ in Table 2 may represent real disequilibrium and could be not significant by chance alone. This point is especially important for inferences about measures of composite di-allelic disequilibrium in short DNA regions where we expect such disequilibrium (LITT and JORDE 1986). Also, given the potential for involvement of nonevolutionary sampling design factors in other types of disequilibrium for many studies, it is reasonable to consider a less conservative alternative to adjusting critical values for multiple tests. In this manner, any significant disequilibrium would require explanation. An argument could certainly be made for even increasing the $a$ level above 0.05 for individual tests to achieve a better balance between the two types of potential errors.

Since physical distance between markers in this region does not predict the magnitude or significance of disequilibria among marker alleles, we conclude the physical distance relationships between markers and unknown functional mutations will be likewise unpredictable from the degree of associations inferred about individual markers and phenotypic variation. This unpredictability is a natural consequence of the unpredictable evolutionary history shared by alleles. Furthermore, nonindependence of RFLPs dictates...
markers should be used as multisite haplotypes or genotypes in ANOVA studies to identify the chromosomes likely to carry the unknown functional mutations with significant phenotypic effects. Using multisite haplotypes or genotypes automatically incorporates the evolutionary history shared by marker and unknown functional mutation alleles. And lastly, it is clear from this study that genetic structure cannot be ignored. Our simple example of $D_s$ from published results indicates nonrandom allele frequency associations will affect genotypic frequencies and hence inferences about genetic variance of quantitative traits if we assume in our inferences that such disequilibria do not exist. Work in progress examines the role of the more complex two-marker allelic associations on such inferences. Our argument assumes that completely random allelic associations are the "truth." If evolutionary factors are the sole explanation for nonrandom allelic associations are the "truth." If, however, nonevolutionary sampling factors are involved such as those we have described in this study, then the phenotypic variation associated with a marker is estimated with bias and will remain a problem in the search for genotype-phenotype relationships.


Communicating editor: B. S. WEIR