Detecting Association in a Case-Control Study While Correcting for Population Stratification

David E. Reich[1*] and David B. Goldstein[2]

[1]*Whitehead Institute/MIT Center for Genome Research, Cambridge, Massachusetts*
[2]*The Galton Laboratory, Department of Biology, University College London, London, United Kingdom*

Case-control studies are subject to the problem of population stratification, which can occur in ethnically mixed populations and can lead to significant associations being detected at loci that have nothing to do with disease. Here, we describe a way to measure and correct for stratification by genotyping a moderate number of unlinked genetic markers in the same set of cases and controls in which a candidate association was found. The average of association statistics across the markers directly measures stratification. By dividing the candidate association statistic by this average, a *P*-value can be obtained that corrects for stratification. Genet. Epidemiol. 20:4–16, 2001.   © 2001 Wiley-Liss, Inc.

Key words: case-control studies; association studies; population stratification; χ^2 test; SNPs

INTRODUCTION

Case-control association studies detect non-random association between an allele and a trait and can be powerful tools for gene mapping [Risch and Merikangas, 1996]. However, when a case-control sample is ethnically mixed or is derived from a population that experienced mixture during the past few generations, non-random associations can occur even at markers completely unlinked to a disease locus [Chakraborty and Weiss, 1988; Lander and Schork, 1994]. In a population that is a mixture of African Americans and Caucasians, for instance, cases of hypertension will occur disproportionately among African Americans, who are well known to have

a higher prevalence of this disease [Kaplan, 1994]. Any alleles that occur more commonly in African Americans will tend to be associated to disease, even if they are completely unlinked to disease-causing loci.

One approach to dealing with the problem of population stratification is to match the ethnic backgrounds of patients to controls as carefully as possible, a strategy used by many epidemiologists. However, there is currently no consensus about the effectiveness of this strategy; a considerable amount of "cryptic stratification" may remain even after cases and controls are carefully matched. The transmission disequilibrium test (TDT) and related methods (SDT, sib-TDT) circumvent these uncertainties entirely by using control chromosomes from the families of affected individuals [Horvath and Laird, 1998; Spielman and Ewens, 1998; Ewens and Spielman, 1995]. However, this generates new problems because it requires the collection of DNA samples from family members, which may be much more difficult than collecting samples from unrelated cases and controls. What is clearly needed is a method for using a case-control study even in the presence of stratification and also for obtaining a quantitative measure of the extent to which a case-control sample that is designed to be stratification-free in fact meets this criterion.

If stratification exists, it should raise association statistics not only at the candidate marker, but also at unlinked markers, and thus by genotyping unlinked markers in the same set of cases and controls in which the candidate association was detected, it should be possible to adjust for stratification. Suppose that a candidate association to disease has been detected in a set of cases and controls with an unknown amount of stratification. To check whether the nominally significant association statistic really indicates the presence of a disease-causing gene at the locus, we suggest recording the association statistics observed at unlinked markers [for related approaches, see Devlin and Roeder, 1999; Schork et al., 1999]. If the statistic at the candidate marker is dramatically higher than the statistics at unlinked markers, there is a strong indication that the candidate marker is near a disease-predisposing gene. On the other hand, if the association statistic is not much different from those at unlinked markers, the association may be due to stratification alone.

TERMINOLOGY AND DESCRIPTION OF SIMULATIONS

To explore these issues in depth, we use a terminology and simulation framework that was introduced to investigate a related but quite different problem: how to detect but not quantify and correct for stratification [Pritchard and Rosenberg, 1999]. Our simulations, for simplicity, assume that the mixed population is derived from two ancestral populations, 1 and 2, and that the disease prevalences in the two populations are p_1 and p_2. For a case-control sample to be stratified, there are two requirements: first, that the two mixing populations must have been isolated from each other long enough that the populations have become differentiated, and second, there must a difference in disease prevalence between the two populations. This means that the level of stratification is affected by the disease that is being studied (stratification increases along with an increasing difference between p_1 vs. p_2), in addition to being affected by the ethnic mixture scenario in the study population. To describe how the p_1 vs. p_2 difference affects stratification, we call the probability that a diseased individual comes from population 1, f, and the probability that a healthy indi-

vidual comes from population 1, g. The variables f and g are related to p_1 and p_2 according to the equation $f(1 - g)/[g(1 - f)] = p_1(1 - p_2)/[p_2(1 - p_1)]$ [Pritchard and Rosenberg, 1999].

To reconstruct the genealogical relationships among the modern samples, we use a computer simulation based on a coalescent algorithm of Hudson [1991], which was modified as described by Reich et al. [1999]. Specifically, the ancestral population of size N is assumed to have split into the two descendant populations, both of size N, at a time $\tau = t/2N$ in the past (t is in generations). In the simulation, a fraction f of cases and a fraction g of controls are sampled from population 1, and the rest are sampled from population 2. The samples are connected together via a genealogy that is produced by the simulation, and we only use genealogies at which mutations— that is, at least one SNP—have occurred. Mutations are placed on the genealogy according to a random Poisson process at a rate of $0.2/(4N)$ per generation, which very approximately is 200 times the human mutation rate per base pair, and hence is equivalent to using markers that were identified by screening 200 base-pair DNA fragments for polymorphisms. Note that although this choice of mutation rate is arbitrary, varying the mutation rate had little effect on our analysis. If a collection of markers with allele frequencies in a specific range are desired, they can be obtained by repeating the simulation enough times until markers of the appropriate frequencies are generated (for our purposes in this paper, allele frequency is defined as the frequency of the less common allele at a marker).

To assess quantitatively the level of association at a marker, it is necessary to collect information about allele frequencies among cases and controls and record them in a contingency table (Table I). The classic χ^2 statistic with 1 degree of freedom can then be calculated:

$$\chi^2 \sim \left(\frac{n_{11}}{n_{1*}} - \frac{n_{21}}{n_{2*}}\right)^2 n_{**} \frac{n_{1*}n_{2*}}{n_{*1}n_{*2}}. \tag{1}$$

Definitions of the variables in expression 1 are provided in Table I. A high χ^2 statistic indicates that the difference in allele frequency between cases and controls, $n_{11}/n_{1*} - n_{21}/n_{2*}$, is more than could be expected by chance. For instance, a χ^2 value >3.8 is significant at the $P < 0.05$ level in the absence of stratification, and a χ^2 value >6.6 is significant at the $P < 0.01$ level. These classic χ^2 cutoffs can only be used when

TABLE I. Standardized Table for Calculation of χ^2 Values

	Allele A	Allele B	Totals
Cases	N_{11} (no. of copies of allele A among cases)	n_{12} (no. of copies of allele B among cases)	$n_{1*} = n_{11} + n_{12}$
Controls	N_{21} (no. of copies of allele A among controls)	n_{22} (no. of copies of allele B among controls)	$n_{2*} = n_{21} + n_{22}$
Totals	$n_{*1} = n_{11} + n_{21}$	$n_{*2} = n_{12} + n_{22}$	n_{**}

n_{*1} = total number of copies of allele A; n_{*2} = total number of copies of allele B; n_{1*} = total number of alleles present in cases; n_{2*} = total number of alleles present in controls; n_{**} = total number of alleles genotyped.

sample sizes are sufficiently large, in particular when the "expected values" in the four squares of Table I, defined as $n_1 \cdot n_{*1}/n_{**}$, $n_1 \cdot n_{*2}/n_{**}$, $n_2 \cdot n_{*1}/n_{**}$, $n_2 \cdot n_{*2}/n_{**}$, are all at least 5 [Rice, 1995].

A SIMPLE WAY TO DISTINGUISH TRUE FROM FALSE POSITIVES

Figure 1 illustrates how χ^2 values for true associations can be distinguished from χ^2 values for markers far away from disease-causing genes even in the presence of population stratification. The specific scenario that is envisioned is a mixture of two populations, 10% from population 1 and 90% from population 2, in which the probability of disease in population 1 is four times higher than in population 2. In our terminology, this corresponds to $g = 0.1$, $f = 0.31$, and we also assume that $\tau = 0.25$. As shown in Fig. 1, real association superimposed on population stratification can produce χ^2 statistics far higher than are expected from stratification alone; in other words, Distributions 1 and 2 are easily distinguished from each other. Even for different models of disease causality than the specific one depicted in Fig. 1 (see legend for details), the distributions should be distinguishable. For example, if a candidate marker is just linked to a variant that causes disease, rather than being causal itself as assumed in Fig. 1, we would still expect Distributions 1 and 2 to be distinguishable from each other because of linkage disequilibrium between the candidate marker and the causal variant that is inherited from the parental populations.

Suppose as an example that a χ^2 statistic of 20 is observed at a candidate marker, which would be significant at the $P < 0.000008$ level in the absence of stratification. To assess whether this is significant even in the presence of stratification, we imagine that 19 unlinked markers are drawn randomly from Distribution 1 in Fig. 1 and that all have χ^2 values <20. Thus, the candidate association statistic is in the top 5% of studied markers and the inferred significance level is $P < 0.05$. To establish a precise P-value using this method, a very large number of markers would be required. A more efficient approach is to directly use the values of the χ^2 statistics at unlinked markers to quantitatively estimate the P-value in the presence of stratification.

QUANTITATIVE ASSESSMENT OF POPULATION STRATIFICATION

To measure the stratification in a case-control sample, it is necessary to assess the degree of elevation of χ^2 values compared to the null expectation in an unstratified scenario. An obvious summary statistic is the mean χ^2 value across unlinked loci ($\bar{\chi}^2$), which is expected to be $\mu = 1$ in the absence of stratification [Rice, 1995] and should rise in the presence of stratification. Using simulations, we discovered a surprisingly simple relationship between μ and the percentiles of the χ^2 distribution in a stratified population. When μ is plotted against the 95th and 99th, and in fact every percentile of the χ^2 distribution, the scatterplot takes the form of a straight line that goes through the origin, although the relationship begins to break down for very large values of μ (Fig. 2). The slope of the line can then be determined because in the absence of stratification, $\mu = 1$ and the χ^2 percentile is its published value [e.g., Rice, 1995]:
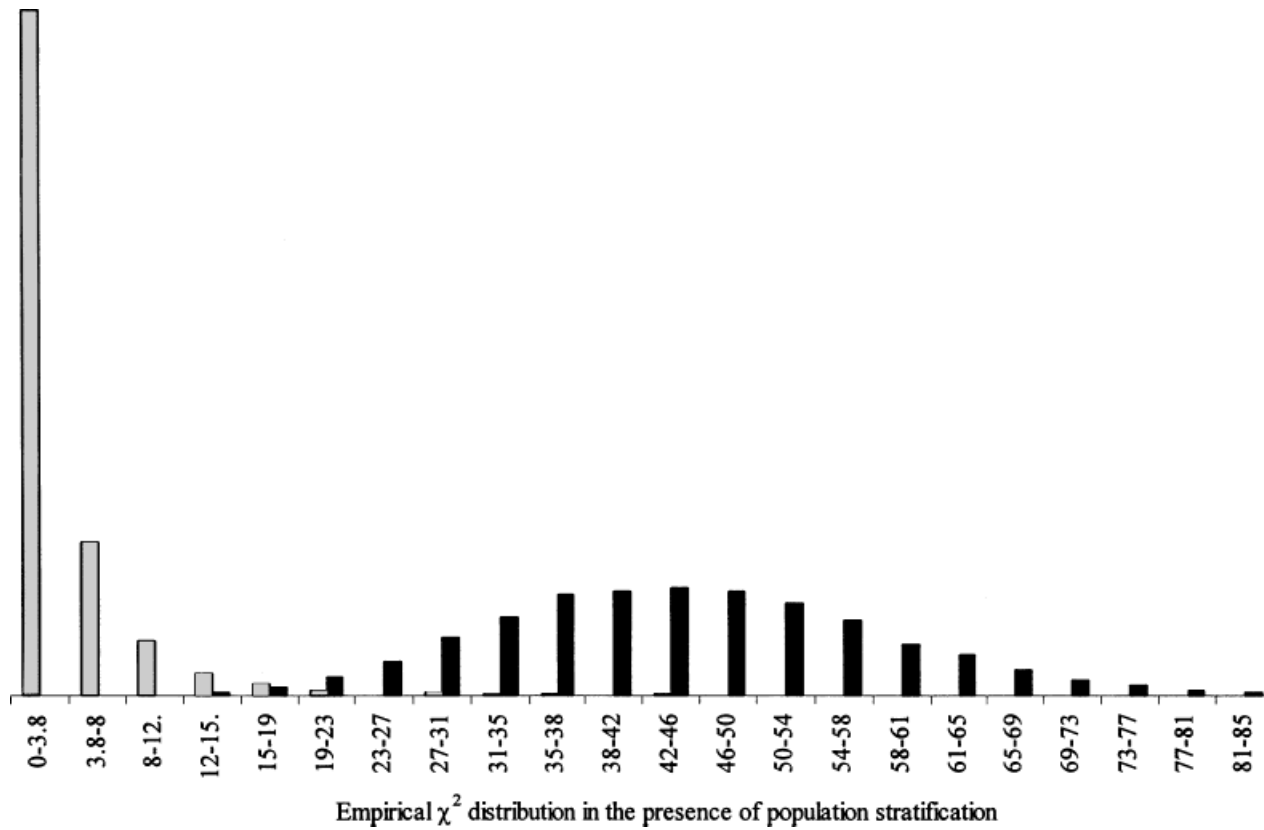
Fig. 1. Empirical distributions of χ^2 values in the presence of population stratification, calculated for the demographic model described in the text. The disease model (for Distribution 2) is such that individuals who carry two, one, or zero copies of the disease-associated allele contract disease with frequencies 16%, 4%, and 1%, respectively. Population stratification causes 27% of replicates for unlinked markers to be nominally associated at the $P < 0.05$ level ($\chi^2 > 3.8$). However, despite this elevation of χ^2 values even at the markers not associated with disease (Distribution 2), the χ^2 values for true associations, shown in Distribution 1, are easily distinguishable. Distribution 1: unlinked markers, shaded columns; Distribution 2: disease-causing polymorphism, solid columns.

$$\begin{pmatrix} \text{equivalent } \chi^2 \text{ value in the} \\ \text{absence of stratification} \end{pmatrix} = \left(\chi^2 \text{ value in the presence of stratification} \right) \div \mu \quad (2)$$

Equation (2) implies that in the presence of stratification, the χ^2 distribution changes only by being stretched out by a factor μ, but the overall shape of the distribution remains fixed. This applies not only for the 100 case and 100 control situation depicted in Fig. 2, but for a wide range of sample size configurations and even for unequal numbers of cases and controls. Devlin and Roeder [1999] obtained the same result using a purely statistical argument that is not based on simulations or on population genetic assumptions. Equation (2) thus appears to be quite general, although our simulations indicate that it breaks down for extreme cases of stratification, a result not obtained by other investigators. In particular, μ tends to underestimate the 95th percentile of the distribution for large values of μ (Fig. 2).

     To explain Equation (2) and its breakdown for large values of μ, we recall that the classical χ^2 probability distribution has a shape determined by the variance of a normally distributed variable [Rice, 1995]. The χ^2 distribution applies for our test
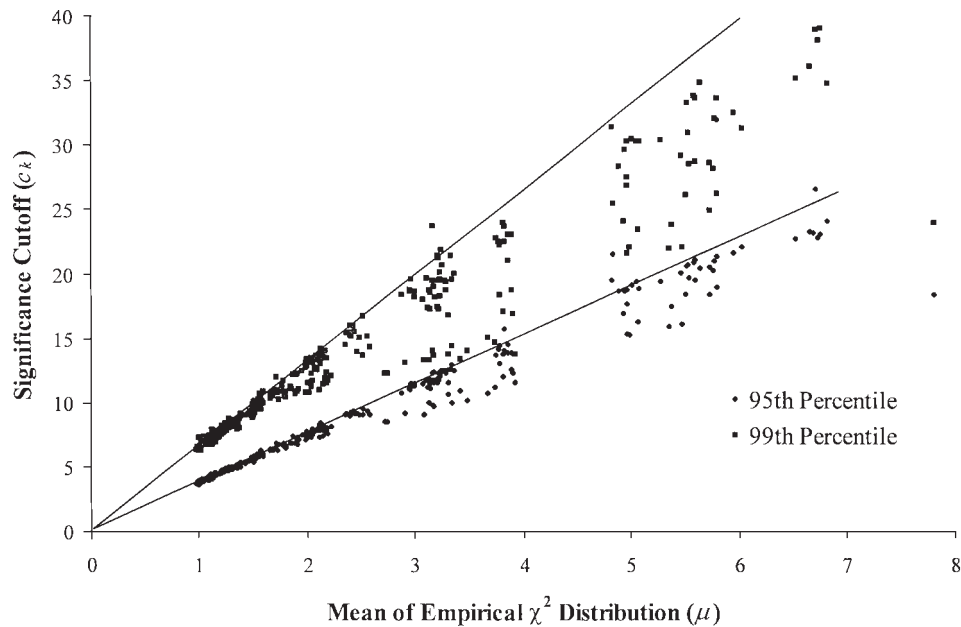

Fig. 2.   The mean of the empirical χ^2 distribution versus the 95th and 99th percentiles, obtained by coalescent computer simulations for 100 cases and 100 controls. The following parameter combinations were considered: $g = (0.05, 0.1, 0.3, \text{and } 0.4)$, $\tau = (0.025, 0.05, 0.1, 0.2 \text{ and } 0.4)$, and relative risks of disease population 1 compared to population 2 (p_1/p_2) of 2, 4, and 8. Eighty-five thousand simulations were performed for each of the 60-parameter combinations, and the results were binned into seven allele frequency ranges (15–20%, 20–25%, 25–30%, 30–35%, 35–40%, 40–45%, and 45–50%), each of which was used to calculate a mean and 95th percentile cutoff. The line fitted to the scatterplot gives a remarkably good match to the data, except when stratification is severe, in which case the mean may overestimate the cutoff.

statistic (Expression 1) because the difference in allele frequency between cases and controls, which is at the core of Expression 1, approximately follows a normal distribution in the limit of large sample size. In the face of population stratification, the statistic appears to continue to be distributed normally, with the only difference that the variance is larger due to differentiation (genetic drift) between the parental populations. The magnitude of the increase, which stretches out the χ^2 distribution without changing its shape as in fact we observe in our simulations, is μ. When stratification becomes too extreme, the filling of squares of Table I becomes non-normal, which leads to the breakdown of Equation(2) for high values of μ, and the scatter of points around the fitted line in Fig. 2. Indeed, a computer program that generates random fillings of Table I using a gamma probability distribution with fixed mean but changing variance not only regenerates the behavior described by Equation (2), but also predicts its observed breakdown for extremes of stratification.

IMPLEMENTING THE CORRECTION FOR POPULATION STRATIFICATION

For the example of 19 unlinked markers presented above, we assume that the χ^2 values, drawn randomly from Distribution 1 in Fig. 1 and presented in increasing order, are 0, 0, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.5, 0.6, 0.8, 1.1, 1.7, 3.5, 6.3, 7.3, 8.1, 9.3, 19.1, and hence that the mean association statistic is $\bar{\chi}^2 = 3.1$. We now show how to use this estimate of μ to make a direct quantitative correction for stratification. Since the observed association statistic at the candidate marker is 20, we obtain an equivalent χ^2 value in an unstratified population of $20 \times 3.1 = 6.5$, corresponding to a significance level of $P < 0.011$ using a standard χ^2 table of significance levels. However, it may be that the true μ is actually greater than the estimate of 3.1, and to calculate a conservative P-value it is necessary to use the largest μ consistent with $\bar{\chi}^2$. Table II provides the maximum factor by which the true μ could be in excess of the observed mean χ^2 value (at the 95% confidence level) for different numbers of unlinked markers used for estimating μ. A 95% upper confidence limit on μ of course does not translate to a P-value for detecting an association. If this upper limit on μ is used, however, it becomes possible to say that with 95% confidence, the resulting P-value is conservative. Note that Table II is calculated on the basis of the analytically known form of the χ^2 distribution in the absence of stratification, but since the distribution's shape is the same regardless of the level of stratification, the table is appropriate for any stratification scenario. For the present example with 19 markers, the maximum μ consistent with the data is $\bar{\chi}^2 \times 1.59 = 4.9$ at the 95% confidence level, which leads to an equivalent χ^2 value of $20 \times 4.9 = 4.1$ and a conservative significance level of $P < 0.043$. Likelihood-based methods of finding a maximal μ consistent with the data are also possible. For example, based on a likelihood analysis, Devlin and Roeder [1999] suggest taking the median of the χ^2 values, dividing it by 0.675, and squaring the result to obtain a more robust estimate of μ. We have not directly compared the two estimators of μ, but for either approach it is likely that no more than a few dozen markers will need to be genotyped to obtain a fairly accurate estimate of μ.

It is critical to emphasize the generality of the method we have introduced. The method should apply even if the number of cases and controls is not what is shown in Fig. 2 and even if there are unequal numbers of cases and controls. In addition, al-

TABLE II. Look-up Table for Obtaining an Upper Bound on μ

No. of markers	Maximum factor by which μ can exceed the mean χ^2 value at the 95% confidence level
1	3.84
2	3.01
3	2.60
4	2.37
5	2.21
6	2.10
7	2.01
8	1.94
9	1.88
10	1.83
12	1.75
14	1.69
16	1.65
18	1.61
20	1.57
24	1.52
28	1.48
32	1.45
36	1.42
40	1.40
45	1.37
50	1.35
60	1.32
65	1.30
70	1.29
80	1.27
90	1.26
100	1.24

Note: The maximum factor by which μ can exceed the mean χ^2 value at the α% confidence level, assuming that n unlinked markers have been genotyped, can be obtained using the general formula $\chi^2_{n,\alpha}/n$, where $\chi^2_{n,\alpha}$ is the α percentile of the χ^2 distribution with n degrees of freedom. Table II can be obtained, for example, by looking up values of $\chi^2_{n,0.95}$ in any standard statistical table [e.g., Rice, 1995] and dividing by n.

though our simulations did not consider mixtures of three or more populations, there is no theoretical reason to expect that the multiplicative scaling of the χ^2 distribution will not hold for these cases. Indeed, Devlin and Roeder [1999] considered the mixing of three or more populations analytically and showed that the multiplicative scaling holds equally well for these scenarios as it does for the mixing of two populations.

THE ISSUE OF ALLELE FREQUENCY MATCHING

So far we have assumed that the association statistics at unlinked markers can be compared to those at the candidate marker to calculate significance cutoffs. However, from a rigorous point of view, allele frequencies at the unlinked and candidate markers have to be identical for the χ^2 statistics to be truly comparable. The χ^2 statistics are calculated on the basis of Table I, and two fillings of the table can only be compared properly if the marginal sums in the two tables (n_{1*}, n_{2*}, n_{*1} and n_{*2}) are

the same. In practice, there are many situations for which allele frequency matching does not need to be stringent, but if there does turn out to be a strong requirement for allele frequency matching, a very large number of markers would have to be screened before identifying a few that are appropriate for comparison. To investigate when allele frequency matching is required and when it is not, we carried out a wide-ranging simulation study (Devlin and Roeder [1999] and Schork et al. [1999] did not consider this matter at all).

Figures 3 and 4 describe the effect of allele frequency on the χ^2 distribution, focusing on the 95th percentile as a surrogate for the entire distribution (for example, the 99th percentile would have shown the same behavior due to Equation [2]). Figure 3 assumes 100 cases and 100 controls and covers the 5–50% frequency range, whereas Fig. 4 assumes 250 cases and 250 controls and covers the 1–10% frequency range (the 95th percentile increases with increasing sample size, which explains why the 95th percentiles tend to be higher in Fig. 4 than in Fig. 3; results for other numbers of cases and controls can be extrapolated from these figures as described below). Each trace corresponds to a single demographic scenario, and following the trace horizontally shows how the 95th percentile of the χ^2 distribution (that is, the $P < 0.05$ cutoff) changes with allele frequency. For example, the trace indicated by an arrow in Fig. 3B ($g = 0.1$, $f = 0.31$ and $\tau = 0.2$) is flat for allele frequencies >15%, which means that allele frequency matching is not required as long as the frequencies of the candidate marker and unlinked markers are all >15%. In most real studies, in which clinicians attempt to match the ethnicities of controls to those of cases, we expect that the level of stratification will be even less than in this specific example (currently, however, there are no empirical data to support this). Hence, we believe that for most candidate markers with allele frequencies >15%, the requirement for allele frequency matching will not be stringent at all: any unlinked marker with an allele frequency >15% will be a useful match.

The type of stratification that pertains in a given case-control study is in general not known in the sense that it is unknown on which trace in Figs. 3 or 4 a particular study falls. Hence, the only way to assess whether allele frequencies need to be carefully matched is to estimate μ directly and to see which curves in Figs. 3 and 4 are consistent with μ. Suppose as an example that there are 100 cases and 100 controls and that μ is estimated as 2.4. The 95th percentile cutoff can then be estimated as $9.1 = 2.4 \times 3.8$ (Equation [2]). Percentile cutoffs around 9.1 are typical of the example highlighted in Fig. 3B, for which the trace is nearly flat, and hence allele frequency matching for this case will not be important unless the candidate marker allele frequency is below approximately 15%.

When sample sizes are different than the ones used to generate the figures, it will be necessary to make an additional adjustment to use the figures. According to our simulations, as well as the equations of Devlin and Roeder [1999], the difference between a χ^2 value in a stratified sample and its equivalent in an unstratified sample increases linearly with sample size. The linear rise of the χ^2 elevation with sample size makes sense because the χ^2 distribution measures variance, and variance—in this case, the additional variance due to stratification—also increases linearly with sample size. For example, if the 95th percentile of a χ^2 distribution is estimated as 9.8 $= 3.8 + 6$ in a sample with 300 cases and 300 controls, the 95th percentile if only 100 cases and 100 controls were sampled would be $5.8 = 3.8 + (6 \div 3)$.
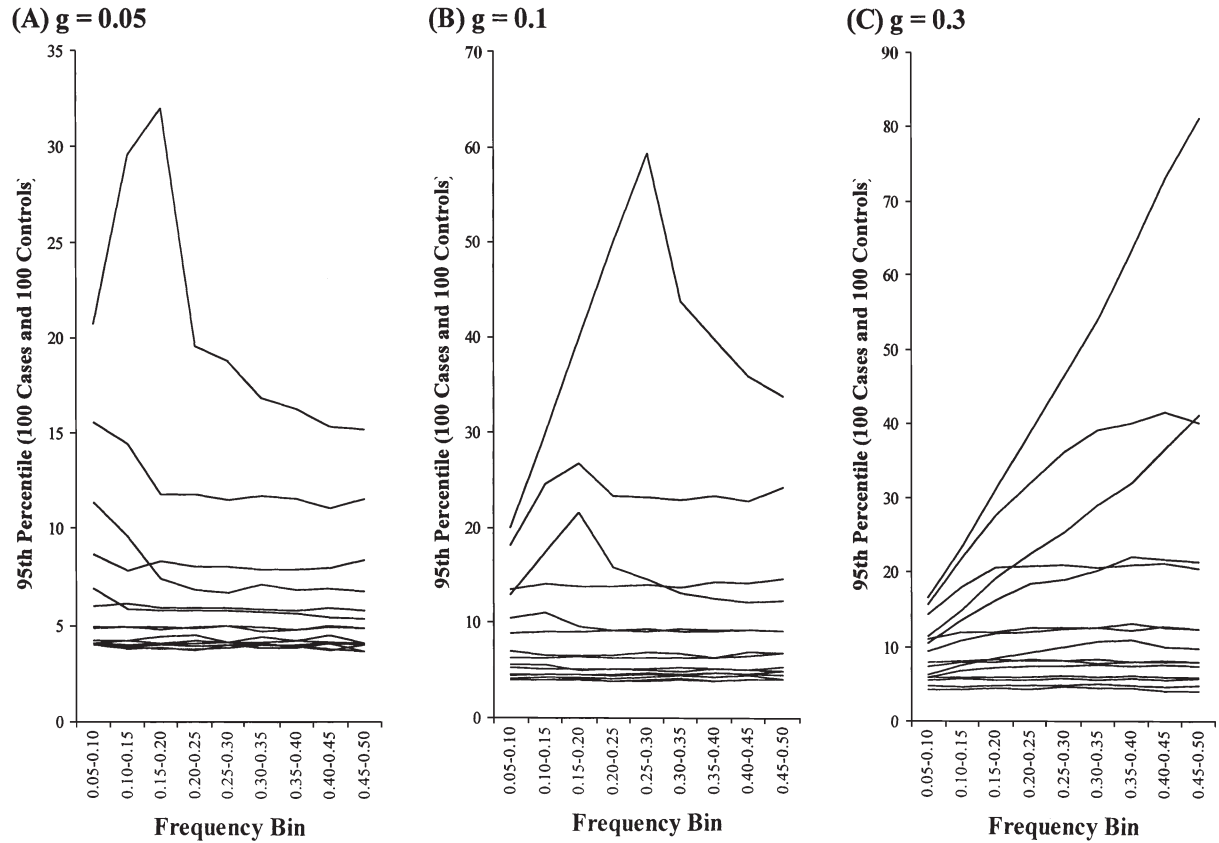
Fig. 3. Behavior of χ^2 distributions as a function of allele frequency for allele frequencies of 5–50% and for 100 cases and 100 controls. Each trace corresponds to a unique combination of parameters, and plots 95th percentile cutoffs as a function of allele frequency. Lower levels of stratification are represented by traces at the bottom of the graph, which show that the 95th percentile χ^2 cutoffs are close to 3.8, the expected value in the absence of stratification. The following parameter combinations were considered: g = (0.05, 0.1 and 0.3), τ = (0.025, 0.05, 0.1, 0.2 and 0.4), and p_1/p_2 values of 2, 4, and 8. Eighty-five thousand simulations were performed for each of the 45 parameter combinations, and the results were binned into nine allele frequency bins that were used to calculate a 95th percentile.
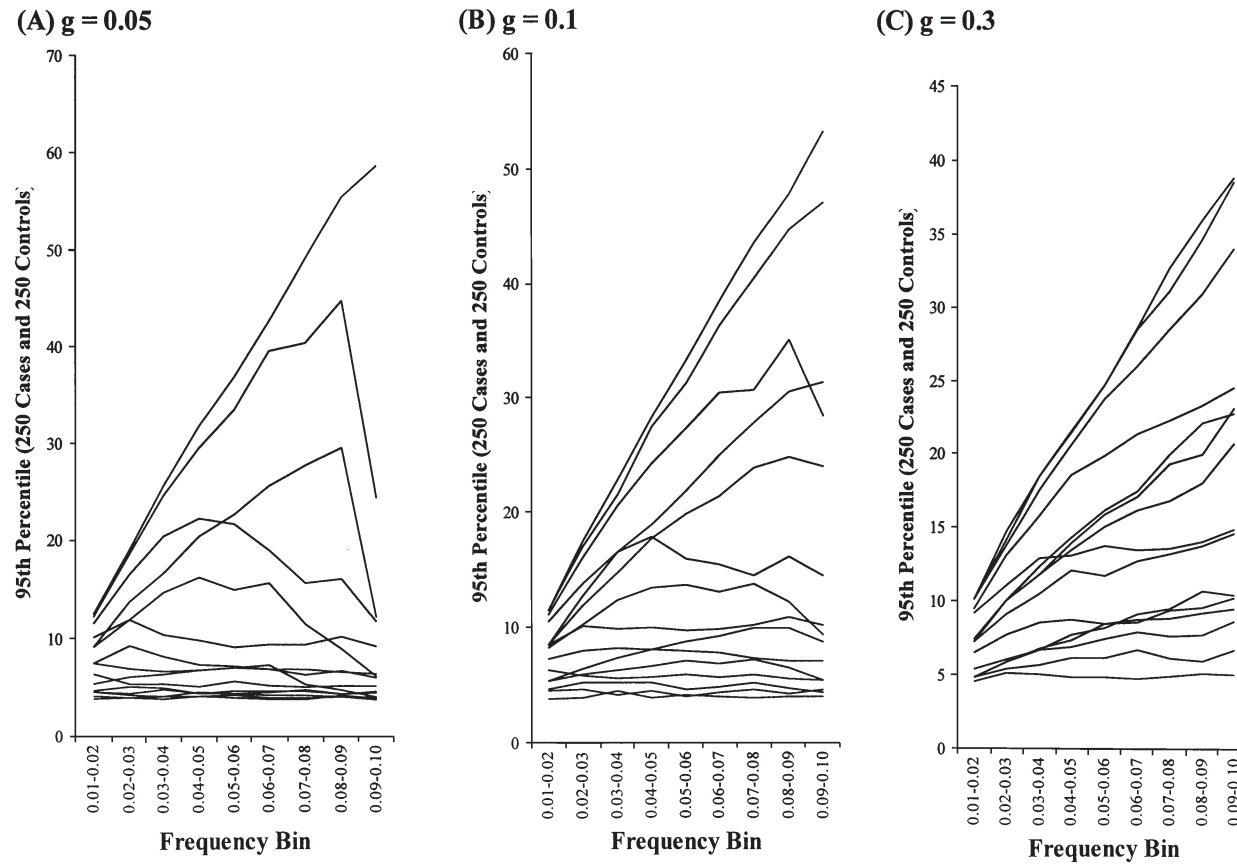
Fig. 4. Behavior of χ^2 distributions as a function of allele frequency for allele frequencies of 1–10% and for 250 cases and 250 controls. Fifty thousand simulations were performed for each stratification scenario. Otherwise, parameters are identical to those in Fig. 3. For low frequencies, the 95th percentile cutoffs tend to be highly dependent on frequency. Hence, stringent allele frequency matching is required for low frequency alleles.

To summarize, for well-designed case-control studies, in which egregious mixing of populations is avoided, the degree of stratification might be expected to be moderate, and stringent frequency-matching will not be required unless the candidate allele frequency is <15%. For many practical case-control studies, we therefore envision that no allele frequency matching will be required. When stratification is extreme, which is only likely to occur in very poorly designed case-control studies but can be assessed directly by estimating μ, it may be necessary to match allele frequencies more closely (perhaps by matching to within a 5–10% frequency window around the candidate marker). Finally, for low allele frequencies, allele frequency matching may need to be quite stringent, since 95th percentile cutoffs change relatively quickly as a function of frequency in this range, especially for allele frequencies <5% (Fig. 4).

DISCUSSION

We have described an approach for teasing out evidence of true association in the context of stratification. The method is easy to implement, and requires the genotyping of only a moderate number of unlinked markers (e.g., a few dozen markers). Using this approach, it is possible to distinguish the level of association that arises on account of population stratification alone from the quantitatively greater association that occurs when a candidate marker is actually associated to a disease-causing variant. In addition, the method makes it possible to assess empirically the degree of stratification in a case-control sample.

For most types of candidate markers, our method entails only a small amount of additional laboratory work, beyond what would be required in a traditional case-control study, to correct for stratification. Hence, the method is practicable for many reasonable scenarios. For candidate markers with low allele frequencies, our method will be more difficult to implement because of a requirement for careful allele frequency matching. Allele frequency matching should not constitute a major obstacle in the long term, however, because efforts to identify large numbers of SNPs and characterize their allele frequencies are underway. For example, as a result of the "SNP Consortium" and related projects, there will soon be so many published markers with known allele frequencies that it will be adequate to consult a published database to identify a sufficient number of frequency-matched markers.

In addition to showing how to correct for stratification in case-control studies, methods such as the one we report have other important implications. For example, it should be possible to re-evaluate published studies by genotyping at unlinked markers in the original samples, thereby determining whether the published associations are likely to be real or attributable to stratification. Our method should also help quantify how much cryptic stratification typically exists in case-control studies— until now a subject of speculation and theory rather than empirical data. It may be that epidemiologists who carefully match controls to cases usually succeed in eliminating cryptic stratification; on the other hand, it is also possible that substantial stratification persists. If the latter, then the preference by some investigators for family-based controls will have been justified in the absence of methods to correct directly for stratification. The debate about the extent of cryptic stratification should soon be settled due to this work and related reports, simply by measuring stratifica-

tion (μ) in a few real case-control samples. The important work of understanding the genetic basis of disease will then be able to move forward unfettered by worries that stratification may cause false-positive associations.

ACKNOWLEDGMENTS

REFERENCES

Chakraborty R, Weiss KM. 1988. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci U S A 85:9119–23.

Devlin B, Roeder K. 1999. Genomic control for association studies. Biometrics 55:997–1004.

Ewens WJ, Spielman RS. 1995. The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 57:455–64.

Horvath S, Laird NM. 1998. A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet 63:1886–97.

Hudson RR. 1991. Gene genealogies and the coalescent process. In: Futuyma D, Antonovics J, editors. Oxford surveys in evolutionary biology. Oxford UK: Oxford University Press.

Kaplan NM. 1994. Clinical hypertension, 6th ed. Baltimore: Williams and Wilkins.

Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. Science 265:2037–48.

Pritchard JK, Rosenberg NA. 1999. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220–8.

Reich DE, Feldman MW, Goldstein DB. 1999. Statistical properties of two tests that use multilocus data sets to detect population expansions. Mol Biol Evol 16:453–66.

Rice JA. 1995. Mathematical statistics and data analysis, 2nd ed. Belmont, CA: Duxbury Press.

Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. Science 273:1516–7.

Schork NJ, Fallin D, Xu X, Blumenfeld M, Cohen D. 1999. The future of genetic case-control studies. Am J Hum Genet 65:A86.

Spielman RS, Ewens WJ. 1998. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 62:450–8.