

# Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees

Laura Almasy and John Blangero

Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio

## Summary

Multipoint linkage analysis of quantitative-trait loci (QTLs) has previously been restricted to sibships and small pedigrees. In this article, we show how variance-component linkage methods can be used in pedigrees of arbitrary size and complexity, and we develop a general framework for multipoint identity-by-descent (IBD) probability calculations. We extend the sib-pair multipoint mapping approach of Fulker et al. to general relative pairs. This multipoint IBD method uses the proportion of alleles shared identical by descent at genotyped loci to estimate IBD sharing at arbitrary points along a chromosome for each relative pair. We have derived correlations in IBD sharing as a function of chromosomal distance for relative pairs in general pedigrees and provide a simple framework whereby these correlations can be easily obtained for any relative pair related by a single line of descent or by multiple independent lines of descent. Once calculated, the multipoint relative-pair IBDs can be utilized in variance-component linkage analysis, which considers the likelihood of the entire pedigree jointly. Examples are given that use simulated data, demonstrating both the accuracy of QTL localization and the increase in power provided by multipoint analysis with 5-, 10-, and 20-cM marker maps. The general pedigree variance component and IBD estimation methods have been implemented in the SOLAR (Sequential Oligogenic Linkage Analysis Routines) computer package.

## Introduction

Methods of linkage analysis that exploit identity-by-descent (IBD) allele sharing between pairs of relatives are

widely used in the genetic analysis of complex traits as these methods generally require few assumptions about the genetic model underlying expression of the trait. There are a limited range of IBD allele-sharing methods that can be used for quantitative-trait linkage analysis. The best known of these is the sib-pair approach of Haseman and Elston (1972). Recently, variance-component linkage analysis methods, which are more powerful than relative pair-based approaches and have the added advantage of providing reasonable estimates of the magnitude of effect of the detected locus, have been developed (Goldgar 1990; Schork 1993; Amos 1994; Blangero and Almasy 1997). These variance-component methods have been extended to accommodate general pedigrees of arbitrary size and complexity (Comuzzie et al. 1997) and to allow analyses that include genotype  $\times$  environment interaction (Blangero 1993; Towne et al. 1997), epistasis (Stern et al. 1996; Mitchell et al. 1997), threshold models for discrete traits (Duggirala et al. 1997), and pleiotropy (Almasy et al. 1997c), as well as multivariate and oligogenic analyses (Schork 1993; Almasy et al. 1997c; Blangero and Almasy 1997; Williams et al. 1997).

Multipoint linkage analysis increases the power to detect true linkages and decreases the false-positive rate. When linkage is detected, multipoint analysis also allows support or confidence intervals to be determined for the location of a gene. To date, practical application of multipoint IBD methods has been confined to sibships or small pedigrees (Fulker et al. 1995; Kruglyak and Lander 1995; Kruglyak et al. 1996; Todorov et al. 1997), although there have been some recent promising developments utilizing computer-intensive Monte Carlo-based techniques (Sobel and Lange 1996; Heath 1997; Heath et al. 1997) in large pedigrees.

The development of variance-component linkage methodologies for use in extended families has created a need for a multipoint IBD method suitable for use in such pedigrees. In general, the computational burden for exact multipoint calculations is considerable even in nuclear families and is prohibitive in large pedigrees. To alleviate this problem, Fulker et al. (1995) developed a multipoint approximation for sib pairs that uses a linear function of IBD values at genotyped markers to estimate IBD sharing at arbitrary chromosomal locations. The Fulker method is based on the evaluation of average

Received January 27, 1998; accepted for publication March 13, 1998; electronically published April 17, 1998.

Address for correspondence and reprints: Dr. Laura Almasy, Department of Genetics, Southwest Foundation for Biomedical Research, 7620 Northwest Loop 410, P.O. Box 760549, San Antonio, TX 78245-0549. E-mail: almasy@darwin.sfbr.org

© 1998 by The American Society of Human Genetics. All rights reserved.  
0002-9297/98/6205-0026\$02.00

number of alleles shared IBD for a pair of siblings and, although much less computationally expensive, has been shown to be as effective as maximum-likelihood estimation of the exact multipoint IBD distribution (Fulker and Cherny 1996). In this article, we extend this simple approach to allow multipoint analysis in pedigrees of unlimited size and complexity. After presenting a general variance-components framework for oligogenic quantitative-trait linkage analysis in arbitrary pedigrees, we derive a series of functions for the correlation between loci in IBD sharing as a function of chromosomal distance in relative pairs found in extended families, including pairs as distant as third cousins (seventh-degree relatives) and relatives related through multiple lines of descent, such as double-first cousins and double-second cousins. We then demonstrate the power and accuracy of the method by using simulation techniques.

**Method**

*Variance-Component Linkage Analysis in General Pedigrees*

The pedigree-based variance-component linkage method uses an extension of the strategy developed by Amos (1994) to estimate the genetic variance attributable to the region around a specific genetic marker. Goldgar (1990) and Schork (1993) have proposed similar variance-component models. This approach is based on specifying the expected genetic covariances between arbitrary relatives as a function of the IBD relationships at a quantitative-trait locus (QTL). The modeling framework used in variance-component analysis is remarkably general (Lange et al. 1976; Hopper and Mathews 1982), although it is also parsimonious with regard to the number of parameters that are required to be estimated relative to that needed in penetrance model-based linkage analysis. Also, unlike most penetrance model-free linkage analysis methods, the variance-component method can be used both for localization of QTLs and for obtaining good estimates of the relative importance of the QTL in determining phenotypic variance in the population (Amos et al. 1996; Blangero and Almasy 1997; Williams et al. 1997).

Let the quantitative phenotype,  $y$ , be written as a linear function of the  $n$  QTLs that influence it:

$$y = \mu + \sum_{i=1}^n \gamma_i + e, \tag{1}$$

where  $\mu$  is the grand mean,  $\gamma_i$  is the effect of the  $i$ th QTL, and  $e$  represents a random environmental deviation. Assume  $\gamma_i$  and  $e$  are uncorrelated random variables with expectation 0 so that the variance of  $y$  is  $\sigma_y^2 = \sum_{i=1}^n \sigma_{\gamma_i}^2 + \sigma_e^2$ . We also allow for both additive and dom-

inance effects, and therefore  $\sigma_{\gamma_i}^2 = \sigma_{ai}^2 + \sigma_{di}^2$ , where  $\sigma_{ai}^2$  is the additive genetic variance due to the  $i$ th locus and  $\sigma_{di}^2$  is the dominance variance. If we assume two allelic variants,  $Q$  and  $q$  with frequencies of  $p_Q$  and  $(1 - p_Q)$  at a given QTL, the genotype-specific means are given by  $\mu_{QQ} = \mu + a$ ,  $\mu_{Qq} = \mu + d$ , and  $\mu_{qq} = \mu - a$  and the QTL-specific genetic variances are given by  $\sigma_{ai}^2 = 2p_Q(1 - p_Q)[a + (1 - 2p_Q)d]^2$  and  $\sigma_{di}^2 = [2p_Q(1 - p_Q)d]^2$ .

For such a simple random effects model, we can easily obtain the expected phenotypic covariance between the trait values of any pair of relatives as

$$\text{Cov}(y_1, y_2) = \sum_{i=1}^n [(k_{1i}/2 + k_{2i})\sigma_{ai}^2 + k_{2i}\sigma_{di}^2], \tag{2}$$

where  $\text{Cov}(y_1, y_2) = E[(y_1 - \mu)(y_2 - \mu)]$ , and the  $k$  terms represent the  $k$  coefficients of Cotterman (1940) with  $k_{ji}$  being the  $i$ th QTL-specific probability of the pair of relatives sharing  $j$  alleles IBD. Similarly, the expected phenotypic correlation between any pair of relatives is given by

$$\rho(y_1, y_2) = \sum_{i=1}^n [(k_{1i}/2 + k_{2i})h_{ai}^2 + k_{2i}d_i^2], \tag{3}$$

where  $h_{ai}^2$  is the proportion of the total phenotypic variance due to the additive genetic contribution of the  $i$ th QTL, and  $d_i^2$  is the proportion due to the dominance effect. In the classical quantitative genetic variance-component model, we do not have information on specific QTLs but utilize the expectation of the  $k$  probabilities over the genome to obtain the following approximation:

$$\text{Cov}(y_1, y_2) \approx 2\phi\sigma_a^2 + \delta_r\sigma_d^2, \tag{4}$$

where  $\sigma_a^2 = \sum_{i=1}^n \sigma_{ai}^2$  is the total additive genetic variance,  $\sigma_d^2 = \sum_{i=1}^n \sigma_{di}^2$  is the total dominance genetic variance,  $\phi = \frac{1}{2}E[(k_{1i}/2 + k_{2i})]$  is the expected kinship coefficient over the genome with  $2\phi = R$  giving the expected coefficient of relationship, and  $\delta_r = E[k_{2i}]$  is the expected probability of sharing 2 alleles IBD. Because we are generally interested in the examination of one or a few QTLs at a time, we exploit the above approximation to reduce the number of parameters that need to be considered. For example, if we are focusing on the analysis of the  $i$ th QTL in equation (1), we can absorb the effects of all of the remaining QTLs in residual components of covariance. Employing these residual covariance terms, the expected phenotypic covariance between relatives is well approximated by

$$\text{Cov}(y_1, y_2) = \pi_i\sigma_{ai}^2 + k_{2i}\sigma_{di}^2 + 2\phi\sigma_g^2 + \delta_r\sigma_d^2, \tag{5}$$

where  $\pi_i = (k_{1i}/2 + k_{2i})$  is the coefficient of relationship

or the probability of a random allele being IBD at the  $i$ th QTL,  $\sigma_g^2$  represents the residual additive genetic variance, and  $\sigma_d^2$  now represents the residual dominance genetic variance. The  $\pi$  and  $k_2$  coefficients and their expectations effectively structure the expected phenotypic covariances and are the basis for much of quantitative-trait linkage analysis such as the sib-pair difference method of Haseman and Elston (1972). For any given chromosomal location,  $\pi$  and  $k_2$  can be estimated from genetic marker data and information on the genetic map.

Given the simple model for phenotypic variation described above, it is possible to use data from pedigree structures of arbitrary complexity to make inferences regarding the localization and effect sizes of QTLs. For the simple additive model in which  $n$  QTLs and an unknown number of residual polygenes influence a trait, the covariance matrix for a pedigree can be written

$$\Omega = \sum_{i=1}^n \hat{\Pi}_i \sigma_{ai}^2 + 2\Phi \sigma_g^2 + \mathbf{I} \sigma_e^2, \quad (6)$$

where  $\hat{\Pi}_i$  is the matrix whose elements  $(\pi_{ijl})$  provide the predicted proportion of genes that individuals  $j$  and  $l$  share IBD at a QTL that is linked to a genetic marker locus,  $\Phi$  is the kinship matrix, and  $\mathbf{I}$  is an identity matrix.  $\hat{\Pi}_i$  is a function of the estimated IBD matrix for a genetic marker itself ( $\hat{\Pi}_m$ ) and a matrix of correlations  $\mathbf{B}$  between the proportions of genes IBD at the marker and at the QTL

$$\hat{\Pi}_i = 2\Phi + \mathbf{B}(r, \theta) \odot (\hat{\Pi}_m - 2\Phi), \quad (7)$$

where  $\theta$  is the recombination frequency between marker locus  $m$  and QTL  $i$ , and the elements  $b_{ij} = \rho(\pi_i, \pi_m | r, \theta)$  are the correlations between the IBD probabilities, where  $r$  denotes  $r$ th type of kinship relationship. Equation (7) is a matrix generalization of the results provided by Amos (1994). The  $\rho$ -functions provide the autocorrelation functions between IBD probabilities as a function of genetic distance, and they also allow prediction of the  $\hat{\Pi}$  matrix at any chromosomal location given  $\hat{\Pi}$  estimates at correlated locations (e.g., when  $\theta < .5$ ). Derivation of the  $\rho$  functions for arbitrary pedigree relationships is provided below.

By assuming multivariate normality as a working model within pedigrees, the likelihood of any pedigree can be easily written and numerical procedures can be used to estimate the variance-component parameters. For the model in equation (6), the ln-likelihood of a pedigree of  $t$  individuals with phenotypic vector  $\mathbf{y}$  is given by

$$\begin{aligned} & \ln L(\mu, \sigma_{ai}^2, \sigma_g^2, \sigma_e^2, \beta | \mathbf{y}, \mathbf{X}) \\ &= -\frac{t}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega| - \frac{1}{2} \mathbf{\Delta}' \Omega^{-1} \mathbf{\Delta}, \quad (8) \end{aligned}$$

where  $\mu$  is the grand trait mean,  $\mathbf{\Delta} = (\mathbf{y} - \mu - \mathbf{X}\beta)$ ,  $\mathbf{X}$  is a matrix of covariates, and  $\beta$  is the matrix of regression coefficients associated with these covariates. Likelihood estimation assuming multivariate normality can be shown to yield consistent parameter estimates even when the distributional assumptions are violated (Beatty et al. 1985; Amos 1994). By performing an extensive series of simulations, we have confirmed the consistency of variance-component estimates of genetic effect size (Blangero and Almasy 1997; J. T. Williams and J. Blangero, unpublished data).

Using the variance-component model, we can test the null hypothesis that the additive genetic variance due to the  $i$ th QTL equals zero (no linkage) by comparing the likelihood of this restricted model with that of a model in which the variance due to the  $i$ th QTL is estimated. The difference between the two  $\log_{10}$  likelihoods produces a LOD score that is the equivalent of the classical LOD score of linkage analysis. Twice the difference in  $\log_e$  likelihoods of these two models yields a test statistic that is asymptotically distributed as a  $\frac{1}{2}:\frac{1}{2}$  mixture of a  $\chi^2$  variable and a point mass at zero (Self and Liang 1987). When multiple QTLs are jointly considered, the resulting likelihood-ratio test statistic has a more complex asymptotic distribution that continues to be a mixture of  $\chi^2$  distributions.

This basic model has been extended to incorporate a number of more complex genetic models by allowing for additional sources of genetic and nongenetic variance. In multilocus models, an additive  $\times$  additive component of epistatic variance can be estimated by use of the Hadamard product of  $\Pi$  matrices for each locus as the coefficient matrix that structures the expected covariances among pedigree members (Mitchell et al. 1997). Dominance  $\times$  dominance, additive  $\times$  dominance, and dominance  $\times$  additive variance components also can be specified by Hadamard products of appropriate  $\Pi$  and  $\mathbf{K}_2$  coefficient matrices. For example, allowing for additive  $\times$  additive interactions between two QTLs leads to the following equation for the phenotypic covariance matrix of a pedigree:

$$\begin{aligned} \Omega &= \hat{\Pi}_1 \sigma_{a1}^2 + \hat{\Pi}_2 \sigma_{a2}^2 + (\hat{\Pi}_1 \odot \hat{\Pi}_2) \sigma_{a1 \times a2}^2 \\ &+ \sum_{i=3}^n \hat{\Pi}_i \sigma_{ai}^2 + 2\Phi \sigma_g^2 + \mathbf{I} \sigma_e^2. \quad (9) \end{aligned}$$

A household or shared environment effect can be added by an additional variance component with a coefficient matrix ( $\mathbf{H}$ ) whose elements are 1 if the relative pair in

question shares the environmental exposure or 0 otherwise. This simple incorporation of a shared household component leads to the following model for the phenotypic covariance matrix

$$\mathbf{\Omega} = \sum_{i=1}^n \hat{\mathbf{\Pi}}_i \sigma_{ai}^2 + 2\mathbf{\Phi} \sigma_g^2 + \mathbf{H} \sigma_h^2 + \mathbf{I} \sigma_e^2 . \quad (10)$$

The general pedigree variance-component linkage method, with all of the above described extensions, has been implemented in a computer analysis package called Sequential Oligogenic Linkage Analysis Routines (SOLAR), which employs the computer programs FISHER and SEARCH (Lange et al. 1988) for likelihood optimization in quantitative-trait analysis. For any of the complex genetic models described above, SOLAR can also incorporate covariate effects as well as multivariate quantitative traits (Almasy et al. 1997c); discrete traits, by use of a threshold model (Duggirala et al. 1997); mixed discrete/quantitative-trait analyses; and genotype  $\times$  environment interaction (Towne et al. 1997).

*Estimation of the IBD Probability Matrix for a Genetic Marker*

In the above formulation, all of the information regarding linkage is a function of the estimated  $\hat{\mathbf{\Pi}}_i$  matrices. For a given genetic marker, a number of methods have been proposed to calculate this IBD probability matrix (Amos et al. 1990; Curtis and Sham 1994; Whittemore and Halpern 1994). One simple and effective approach is to perform pairwise likelihood-based estimation of the elements of a  $\hat{\mathbf{\Pi}}_i$  matrix by calculating the posterior probability of genotypes at a completely linked pseudomarker at which there is an extremely rare allele (i.e., an allele frequency less than the expected mutation rate). With this approach, the  $\pi$  for each pair of individuals is evaluated by randomly assigning the rare homozygous pseudomarker genotype to one of the individuals and then calculating the likelihoods of seeing the three possible pseudomarker genotypes in the other individual conditional on the marker information in the complete pedigree. From the resulting posterior probabilities, it is simple to calculate the three locus-specific  $k$  coefficients for any marker and then to calculate the  $\pi$  estimate. This method is relatively rapid for simple pedigrees but can become tedious in complex pedigrees, especially ones with multiple inbreeding loops. Any software that can calculate two-point pedigree likelihoods can be used for calculating IBD probabilities.

A second alternative for pedigrees of arbitrary size and complexity is to calculate an estimate of all the elements of the  $\mathbf{\Pi}_i$  matrix jointly by Monte Carlo techniques. When there is no missing genetic marker information in a pedigree, exact IBD probabilities can rapidly be cal-

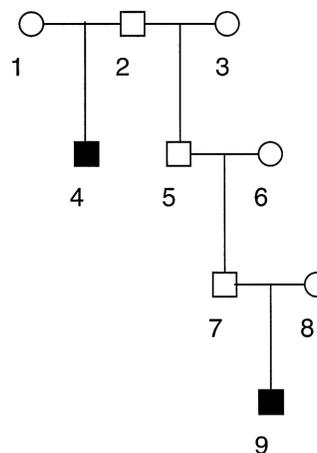
culated by use of the algorithm of Davis et al. (1996). Therefore, Monte Carlo methods can be used to impute marker genotypes for individuals not typed in a pedigree conditional on all other marker and pedigree information. Once the marker genotype vector is filled in by such a process, the exact maximum likelihood estimate of  $\pi$  can be obtained immediately. The results of many such imputations can be averaged by use of the likelihood of the imputed marker genotype vector as a weighting factor. There are many possible variations of such a Monte Carlo approach, but all methods require substantial computing for large pedigrees.

We use both of these approaches in our computer program, SOLAR, and have noticed few differences between them across a wide range of practical applications including extensive computer simulations. A practical benefit of both approaches is the independence of major aspects of the calculations, which renders the estimation problem infinitely scalable with regard to parallel computation.

Although it is comparatively straightforward to obtain an estimate of the  $\mathbf{\Pi}$  matrix for any genetic marker, exact calculation of multipoint IBD probabilities given a number of genetic markers is formidable except for relatively small and simple pedigrees. Since it is well known that exploitation of multipoint information can dramatically improve the power to detect QTLs, fast and accurate approximate methods would be of great benefit. In the next section, we outline our approach to obtaining such approximate multipoint IBD probabilities for any chromosomal location.

*Derivation of IBD Correlation Formulas for Multipoint Analysis*

Given the simplicity and accuracy of the Fulker method (Fulker et al. 1995) for approximating multi-



**Figure 1** Half-grand-avuncular pair (blackened symbols)

**Table 1**

**Possible Two-Locus Combinations of  $\pi$  for Relative Pairs Able to Share Only One Allele IBD**

SECOND LOCUS	FIRST LOCUS		TOTAL $p$
	$\pi_1 = 0, i = 0$	$\pi_1 = \frac{1}{2}, i = 1$	
$\pi_2 = 0, j = 0$	$P_{00}$	$P_{10}$	$1 - 2E(\pi)$
$\pi_2 = \frac{1}{2}, j = 1$	$P_{01}$	$P_{11}$	$2E(\pi)$
Total $p$	$1 - 2E(\pi)$	$2E(\pi)$	

point calculations for sib pairs, we decided to generalize this approach to arbitrary pedigree relationships. Such a general average sharing method requires that we formulate all possible  $\rho(\pi_i, \pi_j | r, \theta)$  functions (i.e., IBD probability autocorrelation functions), which can be used to provide the expected correlation in IBD between genotyped marker loci and any chromosomal location with a known position relative to these marker loci. The correlation in the proportion of alleles shared IBD by a relative pair over some chromosomal distance can be expressed with a simple formula:

$$\rho(\pi_1, \pi_2 | r, \theta) = \frac{\text{Cov}(\pi_1, \pi_2)}{\sigma(\pi_1)\sigma(\pi_2)}, \tag{11}$$

where  $\text{Cov}(\pi_1, \pi_2)$  is the covariance in IBD allele sharing between locus 1 (the genotyped marker) and locus 2 (the arbitrary chromosomal location at which IBD sharing is being estimated), and  $\sigma(\pi_1)$  and  $\sigma(\pi_2)$  are the expected standard deviations in IBD allele sharing at the two loci. These standard deviations depend on the degree of relationship between the relative pair under consideration and will be the same for the two loci. Thus, the denominator reduces to the expected variance in the proportion of alleles shared IBD for the type of relative pair,  $\text{Var}(\pi)$ . This variance in IBD sharing can be calculated by use of the formula  $\text{Var}(\pi) = E(\pi^2) - E(\pi)^2$  where  $E(\pi) = 2\phi$  is the expected IBD sharing for the relative pair.

The covariance is a simple function involving each possible value of  $\pi$  at locus 1 and locus 2, adjusted by  $E(\pi)$  and weighted by the probability of observing the two-locus combination of  $\pi$ ,  $p_{ij}$ :

$$\text{Cov}(\pi_1, \pi_2) = \sum_{ij} p_{ij}[\pi_1 - E(\pi)][\pi_2 - E(\pi)]. \tag{12}$$

For unilineal relative pairs, the number of alleles shared IBD ( $i$  and  $j$ ) at locus 1 and locus 2 will take the values 0 and 1, with the resulting IBD probabilities  $\pi_1$  and  $\pi_2$  being  $i/2$  and  $j/2$ , which yields four possible two-locus combinations of  $\pi$ . For bilineally related pairs able to share two alleles IBD, such as siblings,  $i$  and  $j$  may also be 2, resulting in nine possible combinations. If inbreed-

ing is present,  $i$  and  $j$  may equal 4 when both members of a pair are autozygous for the same ancestral allele. This results in 16 potential two-locus combinations of allele sharing.

The process of obtaining the  $\rho$  function for any class of unilineal relationship is straightforward and may best be described by example.

*An Example: Half-Grand-Avuncular Pairs*

A half-grand-avuncular pair (fig. 1) are fourth-degree relatives for whom  $\text{Var}(\pi) = 7/256$  and  $E(\pi) = 1/16$ . They may share 0 or 1 alleles IBD, yielding  $\pi$  values of 0 and  $\frac{1}{2}$ . To obtain the covariance for this relative pair, we will need the probabilities of observing the four possible two-locus combination of  $\pi$  (table 1).

These probabilities can be determined by calculating the probability of  $\pi_2$  equaling 0 or  $\frac{1}{2}$ , given  $\pi_1$  and all possible patterns of recombination between the two loci. For example, let  $p_{11}$  be the probability that 1 allele is shared IBD at the second locus ( $\pi_2 = \frac{1}{2}$ ), given that 1 allele is shared at the first locus ( $\pi_1 = \frac{1}{2}$ ) and taking into account all possible patterns of recombination. In figure 1, individuals 4 and 9 represent a half-grand-avuncular pair. The probability that they share 1 allele IBD at the first locus ( $\pi_1 = \frac{1}{2}$ ) is  $\frac{1}{8}$ . Any allele shared IBD by 4 and 9 is necessarily also shared by intervening individuals 5 and 7. For 4 and 9 to share an allele at the second locus, transmission from 2, the father of the half-sibs, to his sons 4 and 5 must be either both nonrecombinant with probability  $(1 - \theta)^2$ , or both recombinant with probability  $\theta^2$ . In addition, transmissions from 5 to his son 7 and from 7 to 9 must both be nonrecombinant with probability  $(1 - \theta)^2$ . Thus,  $p_{11} = \frac{1}{8}[\theta^2 + (1 - \theta)^2](1 - \theta)^2$ .

For pairs related by a single line of descent,  $p_{11}$  can be calculated simply from one of four formulas provided in table 2. These formulas use the degree of relationship between the members of the pair and differ by whether the pair is related through a direct line of descent (grandparental relationships), a half-sibling pair (half-avuncular and half-cousin relationships), a full sibling pair descending through only one sib (full avuncular rela-

**Table 2**

**Formulas for  $p_{11}$  in Relative Pairs Related by a Single Line of Descent**

Type of Relative Pair	$p_{11}$
Direct descent	$(1/2^{(d-1)})(1 - \theta)^{(d-1)}$
Half-avuncular or half-cousin	$(1/2^{(d-1)})(1 - \theta)^{(d-2)}[\theta^2 + (1 - \theta)^2]$
Full avuncular	$1/2^d(1 - \theta)^{(d-2)}(2 - 5\theta + 8\theta^2 - 4\theta^3)$
Full cousin	$1/2^d(1 - \theta)^{(d-3)}(2 - 8\theta + 15\theta^2 - 12\theta^3 + 4\theta^4)$

NOTE.— $d$  represents the degree of relationship.

**Table 3**  
Correlation Coefficients for IBD Allele Sharing in Various Types of Relative Pairs

Relation	Correlation in Proportion of Alleles Shared IBD	E( $\pi$ )	Var( $\pi$ )
Sibs	$1 - 4\theta + 4\theta^2$	$\frac{1}{2}$	$\frac{1}{8}$
Half-sibs	$1 - 4\theta + 4\theta^2$	$\frac{1}{4}$	$\frac{1}{16}$
Avuncular	$1 - 5\theta + 8\theta^2 - 4\theta^3$	$\frac{1}{4}$	$\frac{1}{16}$
Grandparent	$1 - 2\theta$	$\frac{1}{4}$	$\frac{1}{16}$
First cousin	$1 - \frac{16}{3}\theta + 10\theta^2 - 8\theta^3 + \frac{8}{3}\theta^4$	$\frac{1}{8}$	$\frac{3}{64}$
Half-avuncular	$1 - 4\theta + \frac{16}{3}\theta^2 - \frac{8}{3}\theta^3$	$\frac{1}{8}$	$\frac{3}{64}$
Grand-avuncular	$1 - \frac{14}{3}\theta + \frac{26}{3}\theta^2 - 8\theta^3 + \frac{8}{3}\theta^4$	$\frac{1}{8}$	$\frac{3}{64}$
Great-grandparent	$1 - \frac{8}{3}\theta + \frac{4}{3}\theta^2$	$\frac{1}{8}$	$\frac{3}{64}$
Half-first cousin	$1 - \frac{32}{7}\theta + 8\theta^2 - \frac{48}{7}\theta^3 + \frac{16}{7}\theta^4$	$\frac{1}{16}$	$\frac{7}{256}$
First cousin, once removed	$1 - \frac{40}{7}\theta + \frac{92}{7}\theta^2 - \frac{108}{7}\theta^3 + \frac{64}{7}\theta^4 - \frac{16}{7}\theta^5$	$\frac{1}{16}$	$\frac{7}{256}$
Half-grand-avuncular	$1 - \frac{32}{7}\theta + 8\theta^2 - \frac{48}{7}\theta^3 + \frac{16}{7}\theta^4$	$\frac{1}{16}$	$\frac{7}{256}$
Great-grand-avuncular	$1 - \frac{36}{7}\theta + \frac{80}{7}\theta^2 - \frac{100}{7}\theta^3 + \frac{64}{7}\theta^4 - \frac{16}{7}\theta^5$	$\frac{1}{16}$	$\frac{7}{256}$
Great-great-grandparent	$1 - \frac{24}{7}\theta + \frac{24}{7}\theta^2 - \frac{8}{7}\theta^3$	$\frac{1}{16}$	$\frac{7}{256}$
Second cousin	$1 - \frac{32}{5}\theta + \frac{88}{5}\theta^2 - \frac{80}{3}\theta^3 + \frac{344}{15}\theta^4 - \frac{32}{3}\theta^5 + \frac{32}{15}\theta^6$	$\frac{1}{32}$	$\frac{15}{1024}$
Half-cousin, once removed	$1 - \frac{16}{3}\theta + \frac{176}{15}\theta^2 - \frac{208}{15}\theta^3 + \frac{128}{15}\theta^4 - \frac{32}{15}\theta^5$	$\frac{1}{32}$	$\frac{15}{1024}$
First cousin, twice removed	$1 - \frac{32}{5}\theta + \frac{88}{5}\theta^2 - \frac{80}{3}\theta^3 + \frac{344}{15}\theta^4 - \frac{32}{3}\theta^5 + \frac{32}{15}\theta^6$	$\frac{1}{32}$	$\frac{15}{1024}$
Half-second cousin	$1 - \frac{192}{31}\theta + \frac{512}{31}\theta^2 - \frac{768}{31}\theta^3 + \frac{672}{31}\theta^4 - \frac{320}{31}\theta^5 + \frac{64}{31}\theta^6$	$\frac{1}{64}$	$\frac{31}{4096}$
Second cousin, once removed	$1 - \frac{224}{31}\theta + \frac{720}{31}\theta^2 - \frac{1328}{31}\theta^3 + 48\theta^4 - \frac{1008}{31}\theta^5 + \frac{384}{31}\theta^6 - \frac{64}{31}\theta^7$	$\frac{1}{64}$	$\frac{31}{4096}$
Third cousin	$1 - \frac{512}{63}\theta + \frac{1888}{63}\theta^2 - \frac{4096}{63}\theta^3 + \frac{5632}{63}\theta^4 - \frac{1664}{21}\theta^5 + \frac{928}{21}\theta^6 - \frac{128}{9}\theta^7 + \frac{128}{63}\theta^8$	$\frac{1}{128}$	$\frac{63}{16384}$

tionships), or a full sibling pair descending through both sibs (full-cousin relationships).

The probabilities for the remaining two-locus sharing states may be similarly derived from the possible patterns of recombination, or for pairs that can share only 0 or 1 alleles IBD, they can be obtained by subtracting from the marginal totals for single-locus sharing of 0 or 1 allele IBD (table 1). For the half-grand-avuncular pair described above,

$$p_{01} = p_{10} = \frac{1}{8} - p_{11} = \frac{1}{8} - \frac{1}{8}[\theta^2 + (1 - \theta)^2(1 - \theta)^2]$$

and

$$p_{00} = \frac{7}{8} - p_{01} = \frac{3}{4} + p_{11} = \frac{3}{4} + \frac{1}{8}[\theta^2 + (1 - \theta)^2(1 - \theta)^2].$$

When these values are used, the covariance for a half-grand-avuncular pair is

$$\begin{aligned} \text{Cov}(\pi_1, \pi_2) &= p_{00}\left(0 - \frac{1}{16}\right)^2 \\ &+ 2p_{01}\left(0 - \frac{1}{16}\right)\left(\frac{1}{2} - \frac{1}{16}\right) \\ &+ p_{11}\left(\frac{1}{2} - \frac{1}{16}\right)^2, \end{aligned} \tag{13}$$

and, after standardization and gathering of terms, the correlation is given by

$$\begin{aligned} \rho(\pi_1, \pi_2 | \text{half - grand - avuncular}, \theta) &= \frac{\text{Cov}(\pi_1, \pi_2)}{7/256} \\ &= 1 - \frac{32}{7}\theta + 8\theta^2 - \frac{48}{7}\theta^3 + \frac{16}{7}\theta^4. \end{aligned} \tag{14}$$

Table 3 shows E( $\pi$ ), Var( $\pi$ ), and the correlation between IBD probabilities for other unilineal classes of relative pairs. These relationships are the most common ob-

**Table 4**

**Correlation Coefficients for IBD Allele Sharing in Relative Pairs with Multiple or Compound Relationships**

Relation	Correlation in Proportion of Alleles Shared IBD	E( $\pi$ )	Var( $\pi$ )
Double–first cousin	$1 - \frac{16}{3}\theta + 10\theta^2 - 8\theta^3 + \frac{8}{3}\theta^4$	$\frac{1}{4}$	$\frac{3}{32}$
Double–first cousin, once removed	$1 - \frac{143}{18}\theta + 32\theta^2 - \frac{731}{9}\theta^3 + \frac{1226}{9}\theta^4 - \frac{5557}{36}\theta^5 + \frac{1058}{9}\theta^6 - 58\theta^7 + \frac{152}{9}\theta^8 - \frac{20}{9}\theta^9$	$\frac{1}{8}$	$\frac{3}{64}$
Double–second cousin (fig. 2a)	$1 - \frac{176}{21}\theta + \frac{785}{21}\theta^2 - \frac{2320}{21}\theta^3 + \frac{674}{3}\theta^4 - \frac{2248}{7}\theta^5 + \frac{4553}{14}\theta^6 - \frac{4828}{21}\theta^7 + \frac{2284}{21}\theta^8 - \frac{656}{21}\theta^9 + \frac{88}{21}\theta^{10}$	$\frac{1}{16}$	$\frac{7}{256}$
Double–second cousin (fig. 2b)	$1 - \frac{32}{5}\theta + \frac{88}{5}\theta^2 - \frac{80}{3}\theta^3 + \frac{344}{15}\theta^4 - \frac{32}{3}\theta^5 + \frac{32}{15}\theta^6$	$\frac{1}{16}$	$\frac{15}{512}$
Double–second cousin (fig. 2c)	$1 - \frac{46}{7}\theta + \frac{130}{7}\theta^2 - \frac{208}{7}\theta^3 + 28\theta^4 - \frac{104}{7}\theta^5 + \frac{24}{7}\theta^6$	$\frac{1}{16}$	$\frac{7}{256}$
First cousin and second cousin	$1 - \frac{352}{63}\theta + \frac{248}{21}\theta^2 - \frac{112}{9}\theta^3 + \frac{472}{63}\theta^4 - \frac{160}{63}\theta^5 + \frac{32}{63}\theta^6$	$\frac{10}{64}$	$\frac{63}{1024}$
Half-sib and first cousin	$1 - \frac{32}{7}\theta + \frac{46}{7}\theta^2 - \frac{24}{7}\theta^3 + \frac{8}{7}\theta^4$	$\frac{3}{8}$	$\frac{7}{64}$
Half-sib and half-avuncular	$1 - 4\theta + \frac{32}{7}\theta^2 - \frac{8}{7}\theta^3$	$\frac{3}{8}$	$\frac{7}{64}$
Double–half-first cousin	$1 - \frac{32}{7}\theta + 8\theta^2 - \frac{48}{7}\theta^3 + \frac{16}{7}\theta^4$	$\frac{1}{8}$	$\frac{7}{128}$
Double–half-avuncular	$1 - 4\theta + \frac{16}{3}\theta^2 - \frac{8}{3}\theta^3$	$\frac{1}{4}$	$\frac{3}{32}$
Half-sib and half–first cousin	$1 - \frac{96}{23}\theta + \frac{120}{23}\theta^2 - \frac{48}{23}\theta^3 + \frac{16}{23}\theta^4$	$\frac{5}{16}$	$\frac{23}{256}$

served in human extended family studies. Table 4 provides the same information for a variety of relative pairs related by multiple lines of descent. Note that while some of these pairs have the same E( $\pi$ ) as pairs in table 3, the variances may be different, since some pairs in table 4 can share both alleles IBD ( $\pi = 1$ ). This leads to nine possible IBD allele sharing states, rather than four, and complicates the calculation of the probability of each sharing state. However, when a pair is related through two independent lines of descent, the elements of the  $2 \times 2$  matrices of sharing probabilities for each independent relationship can be multiplied to form a  $3 \times 3$  matrix of sharing-state probabilities for the compound relationship (table 5). The first formula shown for double–second cousins (table 4) applies only to pairs related through double–first cousins (fig. 2a). Double–second cousins also occur when two sets of first cousins marry (fig. 2b) or when one person’s parent is cousin to both of the other person’s parents (fig. 2c). Each of these double–second cousin pairs have different correlation formulas since the possible  $\pi$  values are not the same (for the pair in fig. 2b,  $\pi$  may equal 1, while in figs. 2a and 2c it cannot) and the possible patterns of recombination also differ. The IBD sharing probability matrix for the double–second cousins in figure 2b can be calculated by multiplying the elements from the basic sharing probability matrix for second cousins as described above. However, the probabilities of the sharing states for the double–second cousins in figures 2a and 2c cannot make use of the formulas above, since the two lines of descent pass through the same individual(s) and are not independent. Thus, the two-locus sharing probabilities

for these types of second cousins were derived by examining the possible patterns of recombination as described for the half–grand-avuncular pair.

*k*<sub>2</sub>-Correlation Functions for Incorporating Dominance

Extension of the above results to allow for dominance effects via the location-specific *k*<sub>2</sub> probabilities requires that we formulate the possible  $\rho(k_{2_i}, k_{2_j} | r, \theta)$  functions (i.e., the *k*<sub>2</sub>-autocorrelation functions). These can be expressed as

$$\rho(k_{2_i}, k_{2_j} | r, \theta) = \frac{\sum_{s=0}^1 \sum_{t=0}^1 f_{(2 \cdot s)(2 \cdot t)} (s - \delta_{7r})(t - \delta_{7r})}{\text{Var}(\delta_{7r})}, \quad (15)$$

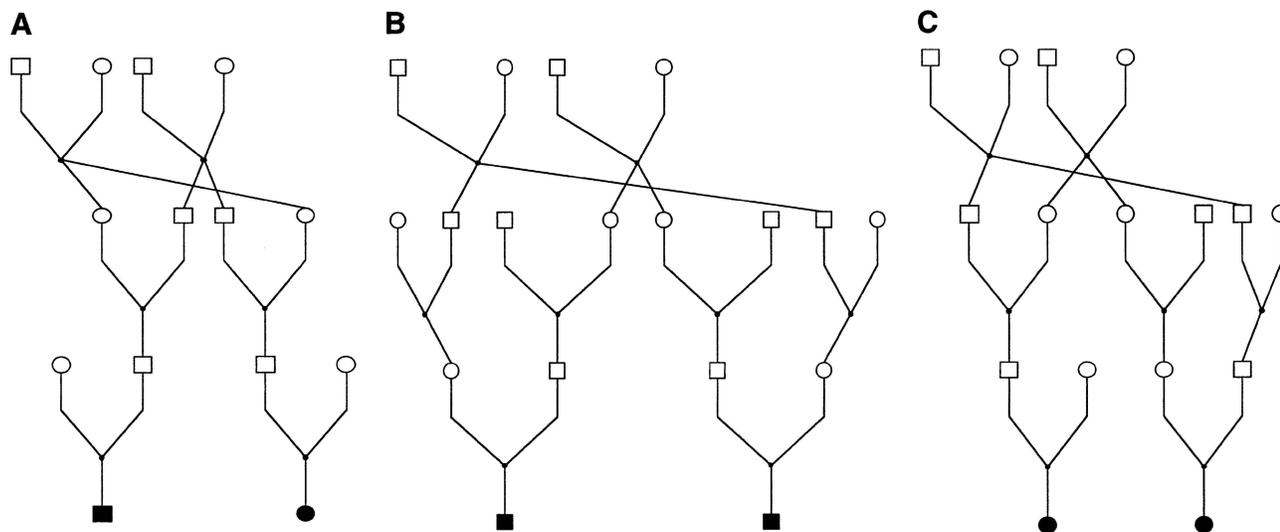
where the summations over *s* and *t* are performed over the possible values (i.e., 0 and 1) of *k*<sub>2</sub> so that the necessary probabilities are limited to *f*<sub>22</sub>, *f*<sub>02</sub>, *f*<sub>20</sub>, and *f*<sub>00</sub>. The probabilities designated by *f* can be obtained from those derived above for the  $\pi$ -autocorrelations for bilinear rel-

**Table 5**

**Probabilities of Two-Locus Combinations of  $\pi$  for Bilinear Relative Pairs**

SECOND LOCUS	FIRST LOCUS		
	$\pi_1 = 0, i = 0$	$\pi_1 = \frac{1}{2}, i = 1$	$\pi_1 = 1, i = 2$
$\pi_2 = 0, j = 0$	$x_{00}y_{00}$	$x_{10}y_{00} + x_{00}y_{10}$	$x_{10}y_{10}$
$\pi_2 = \frac{1}{2}, j = 1$	$x_{01}y_{00} + x_{00}y_{01}$	$x_{11}y_{00} + 2x_{10}y_{01} + x_{00}y_{11}$	$x_{11}y_{10} + x_{10}y_{11}$
$\pi_2 = 1, j = 2$	$x_{01}y_{01}$	$x_{11}y_{01} + x_{01}y_{11}$	$x_{11}y_{11}$

NOTE.—*x* and *y* represent the two locus-sharing probabilities for the two independent lines of relationship.



**Figure 2** Three types of double-second cousins (blackened symbols)

atives. Specifically,  $f_{22} = p_{22}$ ,  $f_{02} = f_{20} = \delta_{7r} - f_{22}$ , and  $f_{00} = 1 - 2\delta_{7r} + f_{22}$ .

Table 6 provides most of the required  $k_2$ -autocorrelation functions that are encountered in studies of extended human families. In general, for a given relationship class, we find that  $\rho(\pi_i, \pi_j | r, \theta) > \rho(k_{2i}, k_{2j} | r, \theta)$ . In other words, the correlation between  $k_2$  values decays more rapidly with genetic distance than does that for the  $\pi$  values. For example, comparing the appropriate correlation functions for sib pairs, we find that  $\rho(\pi_i, \pi_j | \text{sibling}, \theta) - \rho(k_{2i}, k_{2j} | \text{sibling}, \theta) = \frac{1}{3}(4\theta - 20\theta^2 + 32\theta^3 - 16\theta^4)$ , which is  $>0$  for all  $\theta > 0$ . Therefore, the

incorporation of dominance effects into a variance-component model will be most useful when the QTL is comparatively close to a genetic marker.

*Estimation of  $\mathbf{\Pi}$  and  $\mathbf{K}_2$  Matrices by Use of Multipoint Information*

Given the  $\pi$ - and  $k_2$ -correlation functions provided in tables 3, 4, and 6, it is possible to estimate the  $\mathbf{\Pi}$  matrix at any chromosomal location conditional on all of the available genetic marker information and the map locations of the markers. A Haldane mapping function is

**Table 6**

**Correlation Coefficients for  $k_2$  as a Function of  $\theta$  and Relationship**

Relation	$\rho(k_{2i}, k_{2j}   r, \theta)$	$E(k_2)$	$Var(k_2)$
Siblings	$1 - \frac{16}{3}\theta + \frac{32}{3}\theta^2 - \frac{32}{3}\theta^3 + \frac{16}{3}\theta^4$	$\frac{1}{4}$	$\frac{3}{16}$
Double-first cousin	$1 - \frac{128}{15}\theta + \frac{496}{15}\theta^2 - \frac{384}{5}\theta^3 + \frac{1732}{15}\theta^4 - \frac{1696}{15}\theta^5 + \frac{352}{5}\theta^6 - \frac{128}{5}\theta^7 + \frac{64}{15}\theta^8$	$\frac{1}{16}$	$\frac{15}{256}$
Double-second cousin (fig. 2b)	$1 - \frac{1024}{85}\theta + \frac{5888}{85}\theta^2 - \frac{63488}{255}\theta^3 + \frac{157504}{255}\theta^4 - \frac{282368}{255}\theta^5 + \frac{373376}{255}\theta^6 - \frac{365824}{255}\theta^7 + \frac{87744}{85}\theta^8 - \frac{27136}{51}\theta^9 + \frac{15872}{85}\theta^{10} - \frac{2048}{51}\theta^{11} + \frac{1024}{255}\theta^{12}$	$\frac{1}{256}$	$\frac{255}{65536}$
First cousin and second cousin	$1 - \frac{640}{63}\theta + \frac{1024}{21}\theta^2 - \frac{9088}{63}\theta^3 + \frac{18128}{63}\theta^4 - \frac{8416}{21}\theta^5 + \frac{8420}{21}\theta^6 - \frac{16768}{63}\theta^7 + \frac{7552}{63}\theta^8 - \frac{2048}{63}\theta^9 + \frac{256}{63}\theta^{10}$	$\frac{1}{64}$	$\frac{63}{4096}$
Half-sib and first cousin	$1 - \frac{48}{7}\theta + 20\theta^2 - \frac{232}{7}\theta^3 + \frac{232}{7}\theta^4 - \frac{128}{7}\theta^5 + \frac{32}{7}\theta^6$	$\frac{1}{8}$	$\frac{7}{64}$
Half-sib and half-avuncular	$1 - \frac{40}{7}\theta + \frac{96}{7}\theta^2 - \frac{128}{7}\theta^3 + \frac{96}{7}\theta^4 - \frac{32}{7}\theta^5$	$\frac{1}{8}$	$\frac{7}{64}$
Double-half-first cousin	$1 - \frac{512}{63}\theta + \frac{640}{21}\theta^2 - \frac{4352}{63}\theta^3 + \frac{6464}{63}\theta^4 - \frac{6400}{63}\theta^5 + \frac{4096}{63}\theta^6 - \frac{512}{21}\theta^7 + \frac{256}{63}\theta^8$	$\frac{1}{64}$	$\frac{63}{4096}$
Double-half-avuncular	$1 - \frac{32}{5}\theta + \frac{272}{15}\theta^2 - \frac{448}{15}\theta^3 + \frac{448}{15}\theta^4 - \frac{256}{15}\theta^5 + \frac{64}{15}\theta^6$	$\frac{1}{16}$	$\frac{15}{256}$
Half-sib and half-first cousin	$1 - \frac{32}{5}\theta + \frac{272}{15}\theta^2 - \frac{448}{15}\theta^3 + \frac{448}{15}\theta^4 - \frac{256}{15}\theta^5 + \frac{64}{15}\theta^6$	$\frac{1}{16}$	$\frac{15}{256}$

**Table 7****Phenotyped Relative Pairs Informative for Linkage in the Simulated Pedigrees**

Degree (Coefficient) of Relationship and Relationship Type	No. of Pairs
First ( $\frac{1}{2}$ ):	
Sibs	771
Parent-offspring	801
Second ( $\frac{1}{4}$ ):	
Avuncular	1,485
Grandparent-grandchild	151
Half-sibs	26
Third ( $\frac{1}{8}$ ):	
Cousins	2,761
Grand-avuncular	497
Half-avuncular	64
Fourth ( $\frac{1}{16}$ ):	
Cousins once removed	3,051
Half-cousins	27
Great-grand-avuncular	19
Half-grand-avuncular	13
Fifth ( $\frac{1}{32}$ ):	
Cousins twice removed	423
Second cousins	169
	<hr/> 10,258 <hr/>

employed to relate genetic distances to  $\theta$ . To estimate IBD probabilities at any chromosomal location, we have chosen to generalize the regression-based averaging method of Fulker et al. (1995) to arbitrary relationships. Basically, for any pair of individuals of relationship  $r$ , we find the vector of regression coefficients ( $\beta_{r\ell}$ ) on the available estimated marker-specific  $\hat{\pi}_z$  vector that predict  $\pi_\ell$ , where the subscripts now refer to chromosomal locations in centimorgans. This is done by the standard regression method in which

$$\beta_{r\ell} = \mathbf{V}(\hat{\pi}_z)^{-1} \text{Cov}(\hat{\pi}_z, \pi_\ell), \quad (16)$$

where  $\beta_{r\ell}$  is a vector of  $n$  regression coefficients (assuming that we have typed  $n$  markers on the chromosome),  $\mathbf{V}(\hat{\pi}_z)$  is the  $n \times n$  covariance matrix of the marker IBD probabilities, and  $\text{Cov}(\hat{\pi}_z, \pi_\ell)$  is a vector of the expected covariances between the marker IBD probabilities and those at the chromosomal location  $\ell$ . As shown by Fulker et al. (1995), the elements of  $\mathbf{V}(\hat{\pi})$  are determined by the genetic distances between the markers, the  $\rho(\hat{\pi}_i, \hat{\pi}_j | r, \theta)$  functions derived above, and the empirical variances of the  $\hat{\pi}_i$ . Likewise, the elements of the vector  $\text{Cov}(\hat{\pi}_z, \pi_\ell)$  are given by the product of  $\rho(\hat{\pi}_\ell, \hat{\pi}_i | r, \theta)$  values and the empirical variances of the marker  $\hat{\pi}_i$ .

Once obtained, the  $\beta_{r\ell}$  vector is used to estimate  $\pi_\ell$  for the  $ij$ th pair of relatives by

$$\hat{\pi}_{\ell ij} = 2\phi_r + \beta'_{r\ell}(\hat{\pi} - \bar{\pi}), \quad (17)$$

where the symbol  $\hat{\pi}$  without a subscript indicates the

vector of marker IBD probabilities, and  $\bar{\pi}$  is its empirical mean vector. Subject to constraints on the acceptable parameter space that are  $r$  dependent, equation (16) can be used to estimate each pairwise element of the  $\hat{\mathbf{\Pi}}_\ell$  matrix, which is then used to structure the expected phenotypic covariances between relatives as shown in equation (6). The similarity of equation (7) and equation (17) is also apparent, since equation (7) is the matrix prediction equation when there is only a single marker.

A similar approach can be employed to obtain multipoint estimates of  $k_{2\ell}$  by substituting the appropriate expectations,  $k_2$ -autocorrelations, empirical variances, and means in equations (16) and (17).

### Simulations

To evaluate the utility of this multipoint variance-component method for detecting QTLs, we performed a series of computer simulations to assess its properties and accuracy. In the first set of simulations, six quantitative traits and genotype data were simulated for 200 replicates of a data set containing 1,497 total individuals, 1,000 phenotyped, based approximately on the pedigree structure of the San Antonio Family Heart Study. These are extended pedigrees, including all available first-, second-, and third-degree relatives of a proband and the proband's spouse as well as the married-in parents of any descendants. Pedigree size ranges from 37 to 128 individuals ( $\bar{x} = 65$ ); thus, multipoint quantitative-trait linkage analysis of these pedigrees would not be possible with any previously published method. The number of relative pairs with both members phenotyped is shown in table 7 for each type of relative pair present in these pedigrees. Although the SOLAR general pedigree variance-component linkage analysis uses IBD allele sharing between these relative pairs, it should be noted that it is not a relative-pair method as likelihoods are maximized over entire families considered jointly. The number of relative pairs of various types is shown in order to illustrate the depth and complexity of these pedigrees.

Fully informative markers were simulated at a position of 33 cM on a 100-cM chromosome. The alleles of this fully informative marker were grouped together into "high" and "low" bins in various ways to obtain biallelic QTLs whose most common allele took one of three possible generating values, 0.5, 0.7, or 0.9. Two generating values of the additive effect parameter  $a = \frac{1}{2}(\mu - \mu_{qq})$  were considered that produced either a 2- or 2.5-SD difference between the contrasting QTL genotypes. For these simulations, dominance effects were not included. Using the six sets of generating parameters, we simulated six quantitative traits in which the relative variance due to the QTL (i.e., the heritability due to the QTL) ranged from .15 (where  $p_Q = .9$  and  $a = 1$ ) to .44 (where  $p_Q = .5$  and  $a = 1.25$ ). With CHRSM (Speer et al.

**Table 8**

**Percentage of Simulation Replicates with a Maximum LOD Score  $\geq 3.0$  and Mean Maximum LOD Score**

<i>h</i> <sup>2</sup>	DUE TO QTL	ALLELE FREQUENCY	DISPLACEMENT	FULLY INFORMATIVE MARKER AT QTL	PERCENTAGE OF REPLICATES WITH MAXIMUM LOD $\geq 3.0$ (MEAN MAXIMUM LOD)					
					5-cM MAP		10-cM MAP		20-cM MAP	
					Two-point	Multipoint	Two-point	Multipoint	Two-point	Multipoint
.44	.5	2.5	99.5 (13.14)	98.5 (6.45)	98.9 (7.63)	96.0 (6.05)	97.2 (6.86)	72.0 (4.24)	84.5 (5.05)	
.40	.5	2.0	97.5 (10.44)	81.5 (5.11)	87.5 (5.99)	78.5 (4.81)	82.5 (5.32)	52.0 (3.31)	63.0 (3.88)	
.33	.7	2.5	95.5 (7.30)	68.0 (3.89)	75.0 (4.39)	57.0 (3.59)	67.0 (4.01)	35.0 (2.62)	47.5 (2.99)	
.30	.7	2.0	86.0 (6.00)	49.5 (3.27)	58.5 (3.61)	40.0 (2.95)	51.0 (3.31)	21.5 (2.14)	35.5 (2.49)	
.22	.9	2.5	54.0 (3.71)	32.0 (2.45)	34.5 (2.52)	24.5 (2.17)	28.0 (2.26)	10.5 (1.58)	16.5 (1.72)	
.15	.9	2.0	27.0 (2.10)	8.0 (1.60)	11.5 (1.59)	6.0 (1.39)	10.0 (1.45)	2.5 (1.05)	3.5 (1.12)	

NOTE.—Simulated QTLs were biallelic and accounted for 15%–44% of the trait variance. The second and third columns provide the frequency of the more common QTL allele and the displacement between homozygote means in standard deviation units, respectively.

1992; Terwilliger et al. 1993), marker loci were simulated every 5 cM, based on allele number and frequency patterns drawn from a commercially available screening set. For each of the six independent traits/generating models, two-point LOD scores were assessed at each of the marker loci and at the fully informative marker underlying the trait. Multipoint analysis was performed with 5-, 10-, and 20-cM maps drawn from the 21 simulated markers, with IBD sharing estimated every 2 cM for every relative pair. Both two-point and multipoint linkage analyses were performed by use of the variance-component linkage methods described above and implemented in SOLAR.

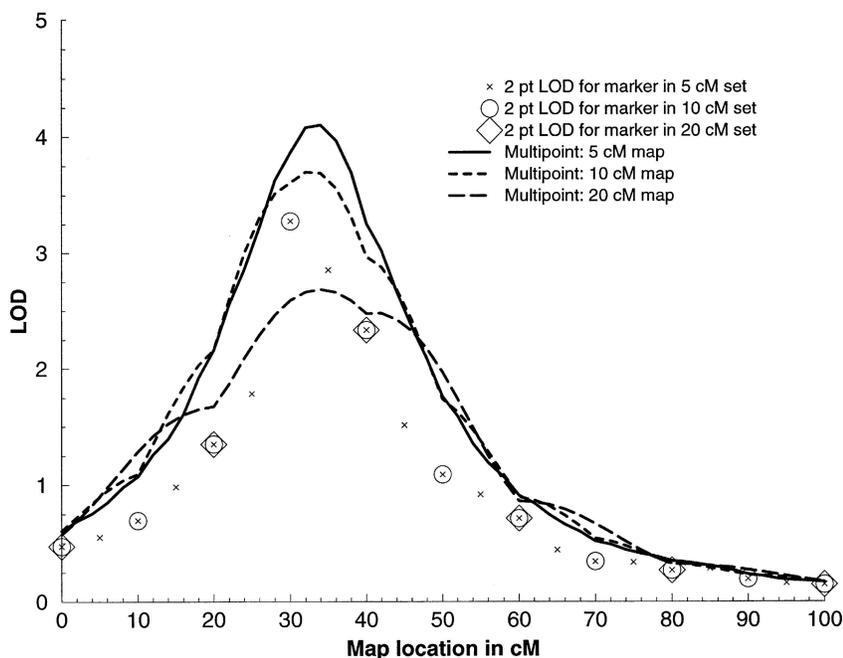
Table 8 provides the mean maximum LOD scores and the percentage of LOD scores  $>3.0$  obtained for each generating model. The fourth column of table 8 shows the mean LOD obtained when the fully informative marker directly on the QTL location was used. This value reflects the maximum LOD scores obtainable in these pedigrees under ideal conditions of marker placement and heterozygosity and serves as a gold standard against which to compare the other linkage analyses. For all three densities of marker maps, multipoint variance-component analysis, as compared to the best two-point variance-component result, improved both the mean maximum LOD score and the percentage of maximum LOD scores  $>3.0$ . For example, with a 5-cM map, the mean LOD for multipoint analysis was an average of 0.5 LOD units higher for the multipoint analysis over the best two-point LOD when considered across all generating values. In addition, the percentages of maximum LOD scores  $>3.0$ , which have standard errors ranging from 0.1 to 1.8, are improved under all six generating models. Table 8 also shows that a substantial amount of linkage information is unavailable even at the 5-cM density, which can be seen by the difference in the mean LOD scores when the fully informative marker at the QTL is compared to the mean multipoint LOD (13.14 vs. 7.63). Because we have arbitrarily placed the QTL

at the midpoint of the 5-cM interval, simply adding another marker within the interval would substantially improve the LOD.

The increase in power with both multipoint variance-component analysis and denser marker maps as well as the accuracy of multipoint localization of the QTL are illustrated in figure 3, which compares the LOD profiles, averaged over the 200 simulations, for one of the simulated traits. Even with a sparse map with an intermarker distance of 20 cM, multipoint analysis provided a noticeable improvement in LOD score over the two-point analyses, as well as an unbiased estimate of QTL location.

For all of the generating models, the multipoint point analysis produced excellent estimates of QTL location. For example, for the model in which the QTL heritability was 0.44, the estimated locations were  $33.21 \pm 0.33$ ,  $33.03 \pm 0.53$ , and  $34.31 \pm 0.59$  for the 5-, 10-, and 20-cM scans, respectively. Similarly for the model in which the QTL heritability was 0.30, the estimated locations were  $32.67 \pm 0.58$ ,  $32.15 \pm 0.65$ , and  $34.82 \pm 0.98$ . In all cases, the estimated chromosomal location was not significantly different from the generating value. Additional evidence that our multipoint variance-component procedure yields unbiased estimates of QTL location is provided elsewhere (Almasy et al. 1997c; Dugirala et al. 1997; Towne et al. 1997; Williams et al. 1997; J. T. Williams and J. Blangero, unpublished data).

The six sets of generating parameters used in these simulations are effectively single major gene models in which there are two QTL alleles acting in a simple codominant manner. This straightforward model does not take advantage of the strengths of the variance-component linkage method. The existence of a single major gene inherently violates the assumption of multivariate normality on which the variance-component linkage method is based. However, it has been demonstrated that the method is robust to violations of this assumption (Beaty et al. 1985). In addition, the use of a biallelic



**Figure 3** Two-point and multipoint LOD score profiles for 5-, 10-, and 20-cM marker maps averaged over 200 simulations for a QTL at 33 cM and with an additive genetic heritability of .33.

QTL is somewhat limiting, since the variance-component methodology is capable of exploiting the greater information content in a multiallelic QTL system.

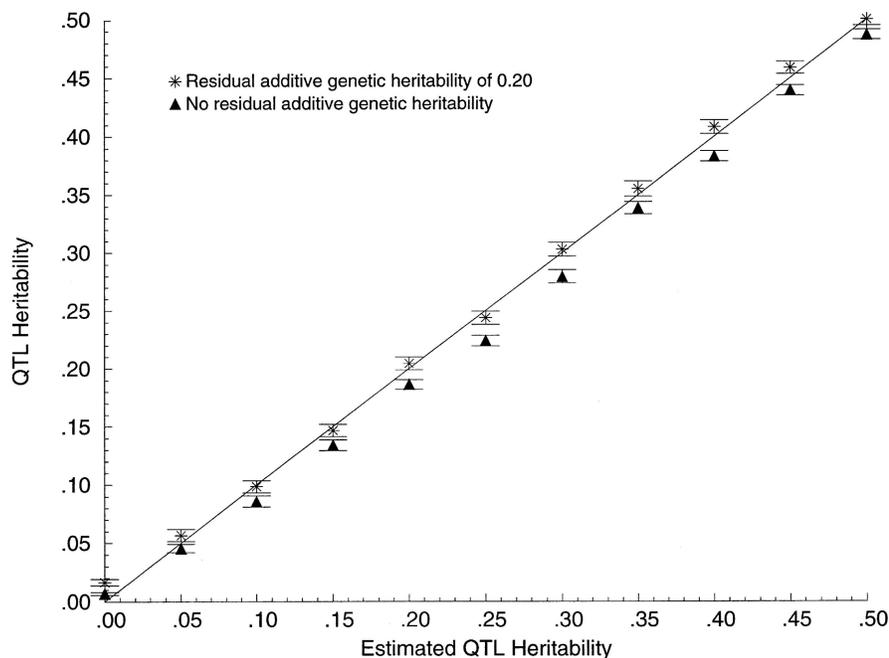
In order to test the accuracy of our estimates of genetic effect size, we performed a second set of simulations in which, given a QTL allele frequency of  $p_Q = .5$ , we chose  $a$  to produce a series of generating models in which the additive genetic heritability due to the QTL ( $h_q^2$ ) varied from 0.05 to 0.50 in increments of 0.05 units. In this simulation, we also allowed for a residual genetic heritability of 0.20. For each generating model, 100 replicates were assessed and quantitative-trait linkage analysis was performed on each. Figure 4 shows a plot of the expected  $h_q^2$  and the mean of the maximum likelihood estimates of  $h_q^2$  at the expected QTL location. Figure 4 clearly shows that the variance-component procedure yields outstanding estimates of genetic effect size. These simulations were also performed with a QTL allele frequency of  $p_Q = .9$  with similar results (not shown).

## Discussion

This powerful variance-components method makes it possible to perform multipoint linkage analysis with quantitative-trait data in pedigrees of arbitrary size and complexity. Such an analysis would previously have required either fragmentation of any large pedigrees into smaller subsets, resulting in a reduction in power to detect linkage, or the application of one of the new com-

puter intensive Monte Carlo-based parametric linkage methods (e.g., the method of Heath 1997). The multipoint IBD estimation method presented in this article has already been utilized successfully in variance-component linkage analyses of simulated data from Genetic Analysis Workshop 10 (Almasy et al. 1997c) as well as such quantitative traits as serum leptin (Comuzzie et al. 1997), and HDL-cholesterol levels (Almasy et al. 1997b) in the extended pedigrees of the San Antonio Family Heart Study and event-related brain potentials in the Collaborative Study on the Genetics of Alcoholism (Almasy et al. 1997a; Porjesz et al. 1997; Begleiter et al., in press).

The IBD estimation procedure is quite efficient and compares favorably to other multipoint methods suitable for use in pedigrees. In contrast to the Elston-Stewart algorithm (1971), in which computation increases exponentially with the number of markers, or the Lander-Green Hidden Markov Model (Lander and Green 1987; Kruglyak et al. 1996), in which computation increases exponentially with the number of non-founders in a pedigree, because the suggested multipoint algorithms are linear functions of previously computed IBDs, processing time increases only linearly for additional individuals or additional loci. For an input file containing IBD information on 16 genotyped marker loci for 20,854 relative pairs, SOLAR, running on a Sun workstation, required only 1 min 10 s to estimate the



**Figure 4** Plot of expected vs. estimated additive genetic heritability due to the QTL. Bars indicate  $\pm 1$  standard error.

IBD matrix at an arbitrary chromosomal location. Such computational speed makes it feasible to estimate multipoint IBD matrices every 1 cM along an entire chromosome, even for very large data sets with many genotyped markers. SOLAR was recently used to analyze complex pedigree data from Genetic Analysis Workshop 10 (Almasy et al. 1997c), with >1,000 genotyped individuals and as many as 50 markers on a chromosome. Similarly, we are employing this method on a large complex baboon pedigree (with a pedigree size of 750 animals) and an extremely large pedigree of individuals from an isolated human population (with a pedigree size of 1,200 individuals). An additional benefit of this approach is that, once a marker data set is deemed final, IBD calculations need be performed only once and the resulting matrices stored for all future analyses. This feature is particularly useful in large studies of complex disease where many different phenotypes have been measured and each needs to be processed through genome-wide linkage analysis.

IBD correlation formulas have previously been derived by a number of authors for limited classes of relative pairs. Amos (1988) derived the IBD correlations for half-sibling, grandparent-grandchild, avuncular, and first-cousin pairs by methods similar to those described above. Feingold (1993) and colleagues (Feingold et al. 1993) employed a different strategy, using a Markov approximation to derive these same four formulas for use in affected relative pair-based linkage analysis using

IBD status. These authors and Lander and Kruglyak (1995) were primarily interested in the  $\pi$ -autocorrelations in order to assess the importance of correlated test statistics in genome scanning. In this regard, it is useful to point out that the Lander and Kruglyak crossover rate parameter, which is central to their method for evaluating genome-wide significance levels, is given by  $-\frac{1}{2} \lim_{\theta \rightarrow 0} [d\rho(\pi_i, \pi_j | r, \theta) / d\theta]$ . Thus, the crossover rate parameter is an approximate measure of how rapidly the  $\pi$ -autocorrelations decay with genetic distance and can be obtained for any pairwise relationship simply as half the absolute value of the coefficient associated with  $\theta^1$  in the appropriate  $\rho$ -function. For example, from table 4, we can immediately determine that the crossover rate parameter for third cousins is  $\frac{1}{2} (512/63) = 256/63$ . Our results in tables 3 and 4 can be used to extend observations on the behavior of correlated test statistics for linkage methods based on extended pairwise relationships.

The present study extends the IBD correlation formulas to many other classes of relative pairs and provides a simple framework for deriving similar formulas for any relative pair related by a single line of descent or by multiple independent lines of descent. Simulations suggest that multipoint variance-component linkage analyses with IBDs calculated based on these correlations recover an unbiased estimate of the location of a gene and provide increased power to detect linkage even with intermarker distances as widely spaced as 20 cM.

These multipoint IBD estimates remove an impediment to making full use of the recent expansions of variance-component linkage methodology, improving the power to examine a wide variety of complex genetic models for both quantitative and discrete traits in general pedigrees.

Applications of quantitative-trait linkage analysis are increasing rapidly. Because of the superior information content of quantitative traits, genetic analysis of quantitative risk factors serves as a powerful tool for elucidating the genetic mechanisms influencing common diseases. Numerous strategies and sampling designs are being formulated, and each has its own strengths and weaknesses. It is well known that, in many situations, extended pedigrees will dramatically outperform smaller family units such as sib pairs, sibships, or nuclear families with regard to the power to detect and accurately localize QTLs (Wijsman and Amos 1997). Unfortunately, although quantitative data have often been collected in more extended kindreds, the lack of adequate linkage tools has generally led to such rich data sets being leached of their potential linkage information by truncation to smaller familial units. Recently, direct comparisons of pedigree-based and nuclear family-based samples consisting of the same number of phenotyped individuals in the same distribution of sibship sizes has underscored the loss of power resulting from fragmentation of a large pedigree-based sample (Duggirala et al. 1997; Towne et al. 1997; Williams et al. 1997). However, with the advent of the multipoint variance-component linkage method, the superior power of extended pedigrees can now be routinely and fully exploited for the localization of QTLs.

The SOLAR software, which incorporates the pedigree-based variance-component and multipoint IBD methods described here, is freely available to interested investigators in a compiled version. SOLAR can be obtained through the Southwest Foundation for Biomedical Research (<http://www.sfbr.org>).

## Acknowledgments

This research was supported by NIH grants GM18897, HL45522, GM31575, and DK44297. Pedigree drawings were produced with the program Pedigree/Draw (Mamelka et al. 1988). The authors gratefully acknowledge the expert assistance of T. Dyer in simulation of data analyzed in this article.

## References

- Almasy L, Blangero J, Porjesz B, Begleiter H, and COGA Collaborators (1997a) Genetic analysis of the N100 event-related brain potential. *Am J Med Genet* 74:595–596
- Almasy L, Blangero J, Rainwater DL, VandeBerg JL, Mahaney MC, Stern MP, MacCluer JW, et al (1997b) Two major genes influence levels of unesterified cholesterol in an HDL subfraction of HDL<sub>2a</sub>. *Atherosclerosis* 134:76
- Almasy L, Dyer TD, Blangero J (1997c) Bivariate quantitative trait linkage analysis: pleiotropy versus coincident linkages. *Genet Epidemiol* 14:953–958
- Amos CI (1988) Robust methods for the detection of genetic linkage for data from extended families and pedigrees. PhD thesis, Louisiana State University, New Orleans
- (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 54:535–543
- Amos CI, Dawson DV, Elston RC (1990) The probabilistic determination of identity-by-descent sharing for pairs of relatives from pedigrees. *Am J Hum Genet* 47:842–853.
- Amos CI, Zhu DK, Boerwinkle E (1996) Assessing genetic linkage and association with robust components of variance approaches. *Ann Hum Genet* 60:143–160
- Beatty TH, Self SG, Liang KY, Connolly MA, Chase GA, Kwitrovich PO (1985) Use of robust variance components models to analyse triglyceride data in families. *Ann Hum Genet* 49:315–328
- Begleiter H, Porjesz B, Reich T, Edenberg H, Goate A, Blangero J, Almasy L, et al. Quantitative trait linkage analysis of human event-related brain potentials: P3 voltage. *Electroenceph Clin Neurophysiol* (in press)
- Blangero J (1993) Statistical genetic approaches to human adaptability. *Hum Biol* 65:941–966
- Blangero J, Almasy L (1997) Multipoint oligogenic linkage analysis of quantitative traits. *Genet Epidemiol* 14:959–964
- Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, Stern MP, et al (1997) A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat Genet* 15:273–275
- Cotterman, CW (1940) A Calculus for Stastico-genetics. PhD thesis, Ohio State University, Columbus
- Curtis D, Sham PC (1994) Using risk calculation to implement an extended relative pair analysis. *Ann Hum Genet* 58:151–162
- Davis S, Schroeder M, Goldin LR, Weeks DE (1996) Non-parametric simulation-based statistics for detecting linkage in general pedigrees. *Am J Hum Genet* 58:867–880
- Duggirala R, Williams JT, Williams-Blangero S, Blangero J (1997) A variance component approach to dichotomous trait linkage analysis using a threshold model. *Genet Epidemiol* 14:987–992
- Elston, RC, Stewart, J (1971) A general model for the genetic analysis of pedigree data. *Hum Hered* 21:523–542
- Feingold E (1993) Markov processes for modeling and analyzing a new genetic mapping method. *J Appl Prob* 30:766–779
- Feingold E, Brown PO, Siegmund D (1993) Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Am J Hum Genet* 53:234–251
- Fulker DW, Cherny SS (1996) An improved multipoint sib-pair analysis of quantitative traits. *Behav Genet* 26:527–532
- Fulker DW, Cherny SS, Cardon LR (1995) Multipoint interval mapping of quantitative trait loci, using sib pairs. *Am J Hum Genet* 56:1224–1233
- Goldgar DE (1990) Multipoint analysis of human quantitative genetic variation. *Am J Hum Genet* 47:957–967
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3–19

- Heath SC (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 61:748–760
- Heath SC, Snow GL, Thompson EA, Tseng C, Wijisman EM (1997) MCMC segregation and linkage analysis. *Genet Epidemiol* 14:1011–1016
- Hopper JL, Mathews JD (1982) Extensions to multivariate normal models for pedigree analysis. *Ann Hum Genet* 46:373–383
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347–1363
- Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439–454
- Lander ES, Green P (1987) Construction of multilocus genetic maps in humans. *Proc Natl Acad Sci USA* 84:2363–2367
- Lander ES, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247
- Lange K, Weeks D, Boehnke M (1988) Programs for pedigree analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol* 5:471–472
- Lange K, Westlake J, Spence MA (1976) Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann Hum Genet* 39:485–491
- Mamelka PM, Dyke B, MacCluer JW (1988) Pedigree/Draw for the Apple Macintosh. Southwest Foundation for Biomedical Research, San Antonio
- Mitchell BD, Ghosh S, Schneider JL, Birznieks G, Blangero J (1997) Power of variance component linkage analysis to detect epistasis. *Genet Epidemiol* 14:1017–1022
- Porjesz, B, Begleiter, H, Blangero, J, Almasy, L, Reich, T, COGA Collaborators (1997) QTL analysis of visual P3 component of the event-related brain potential in humans. *Am J Med Genet* 74:573
- Schork NJ (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *Am J Hum Genet* 53:1306–1319
- Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J Am Stat Assoc* 82:605–610
- Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 58:1323–1337
- Speer M, Terwilliger JD, Ott J (1992) A chromosome-based method for rapid computer simulation. *Am J Hum Genet Suppl* 51:A202
- Stern M, Duggirala R, Mitchell B, Reinhart JL, Shivakumar S, Shipman PA, Uresandi OC, et al (1996) Evidence for linkage of regions on chromosomes 6 and 11 to plasma glucose concentrations in Mexican Americans. *Genome Res* 6:724–734
- Terwilliger JD, Speer M, Ott J (1993) Chromosome-based method for rapid computer simulation in human genetic linkage analysis. *Genet Epidemiol* 10:217–224
- Todorov AA, Siegmund KD, Gu C, Borecki IB, Elston RC (1997) Probabilities of identity-by-descent patterns in sibships when the parents are not genotyped. *Genet Epidemiol* 14:909–913
- Towne B, Siervogel RM, Blangero J (1997) Effects of genotype-by-sex interaction on quantitative trait linkage analysis. *Genet Epidemiol* 14:1053–1058
- Whittemore AS, Halpern J (1994) Probability of gene identity by descent: computation and applications. *Biometrics* 50:109–117
- Wijisman EM, Amos CI (1997) Genetic analysis of simulated oligogenic traits in nuclear families and extended pedigrees: summary of GAW10 contributions. *Genet Epidemiol* 14:719–735
- Williams JT, Duggirala R, Blangero J (1997) Statistical properties of a variance-components method for quantitative trait linkage analysis in nuclear families and extended pedigrees. *Genet Epidemiol* 14:1065–1070