European Journal of Human Genetics (2002) 11, 125–137
© 2002 Nature Publishing Group All rights reserved 1018–4813/02 $25.00

www.nature.com/ejhg

ARTICLE

Combined high resolution linkage and association mapping of quantitative trait loci

Ruzong Fan*[1,2] and Momiao Xiong[3]

[1]Department of Statistics, The Texas A&M University, 447 Blocker Building, College Station, Texas 77843-3143, USA; [2]Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Sigmund Freud Strasse 25, D-53105 Bonn, Germany; [3]Human Genetics Center, University of Texas – Houston, P.O. Box 20334, Houston, Texas 77225, USA

In this paper, we investigate variance component models of both linkage analysis and high resolution linkage disequilibrium (LD) mapping for quantitative trait loci (QTL). The models are based on both family pedigree and population data. We consider likelihoods which utilize flanking marker information, and carry out an analysis of model building and parameter estimations. The likelihoods jointly include recombination fractions, LD coefficients, the average allele substitution effect and allele dominant effect as parameters. Hence, the model simultaneously takes care of the linkage, LD or association and the effects of the putative trait locus. The models clearly demonstrate that linkage analysis and LD mapping are complementary, not exclusive, methods for QTL mapping. By power calculations and comparisons, we show the advantages of the proposed method: (1) population data can provide information for LD mapping, and family pedigree data can provide information for both linkage analysis and LD mapping; (2) using family pedigree data and a sparse marker map, one may investigate the prior suggestive linkage between trait locus and markers to obtain low resolution of the trait loci, because linkage analysis can locate a broad candidate region; (3) with the prior knowledge of suggestive linkage from linkage analysis, both population and family pedigree data can be used simultaneously in high resolution LD mapping based on a dense marker map, since LD mapping can increase the resolution for candidate regions; (4) models of high resolution LD mappings using two flanking markers have higher power than that of models of using only one marker in the analysis; (5) excluding the dominant variance from the analysis when it does exist would lose power; (6) by performing linkage interval mappings, one may get higher power than by using only one marker in the analysis.

European Journal of Human Genetics (2002) **11,** 125–137. doi:10.1038/sj.ejhg.5200941

Keywords: variance component models; linkage interval mapping; linkage disequilibrium mapping; QTL

Introduction

Twenty years ago, variations in human DNA were recognized as genetic markers in linkage study.[1] After that, the advances in molecular biology and computational technology have led to mapping several human inherited disease genes. Using restriction fragment length polymorphism (RFLP) markers and polymorphic microsatellite loci, linkage analysis and positional cloning have been used successfully in mapping the chromosome locations of Mendelian disease genes. The success mainly depends on one premise that the disease genes of Mendelian traits have a large effect on the phenotypes.[2] In fact, there is usually a one-to-one correspondence between disease gene genotypes and the disease phenotype for Mendelian traits. Moreover, the correlations

*Correspondence: Dr Ruzong Fan; Department of Statistics, The Texas A&M University, 447 Blocker Building, College Station, Texas 77843-3143, USA. Tel: 979-845-3156; Fax: 979-845-3144; E-mail: rfan@stat.tamu.edu

Received 4 June 2002; revised 14 October 2002; accepted 21 November 2002

between genotypes and phenotype of Mendelian traits are strong. Given sufficient family data, Mendelian traits can be mapped with high probability by linkage analysis.

With the encouragement of successful mapping Mendelian trait genes, there has been growing interests and endeavors in the study of complex traits such as asthma and diabetes. For complex diseases, the inheritance patterns and phenotype definitions as with genetic etiology are much more complex. The trait/affection status is usually a continuous variable.[3] The mapping of complex disease genes is much harder. Novel statistical methods such as both linkage analysis and linkage disequilibrium (LD) mapping or association study are needed in dissecting complex traits. As very dense marker maps such as single nucleotide polymorphism (SNP) are available,[4] both linkage analysis and association study are utilized simultaneously for mapping complex disease loci.[5,6] Almasy *et al*[7] and Fulker *et al*[8] proposed to use combined linkage and association analysis for quantitative trait loci (QTL). Sham *et al*[9] studied the power of linkage versus association analysis of quantitative traits by analytically calculating non-centrality parameters of test statistics. Abecasis *et al*[10–12] proposed test statistics of association studies for quantitative traits in nuclear families, general pedigrees, and selected samples. Cardon[13] studied a sib-pair regression model of LD for quantitative traits. All these researches concentrated on family data which include sib-pairs, and used only one marker in analysis. In Fan and Xiong,[14] we proposed a linear regression method of high resolution mapping of quantitative trait loci by LD mapping analysis. The method is based on population data. Using two flanking markers, the regression models have higher power than that of models using only one marker.[14]

It is well-known that family pedigree data can be used in both linkage analysis and association study, and population data can be used in association study. Hence, it is necessary to consider a method to combine both population data and family pedigree data in the analysis. In this paper, we propose to perform both linkage analysis and high resolution LD mapping for QTL based on combined family and population data. Linkage interval mapping is based on family data, and LD mapping is based on both family pedigree and population data. Based on variance component models, we construct likelihood to analyse family and population data in Section of Models. Then, we discuss the parameter estimations and regression coefficients. The linkage information, i.e., recombination fractions, is contained in the variance-covariance matrix, and the association information, i.e., the LD coefficients, is contained in the mean parameters or the regression coefficients. We calculate the non-centrality parameters for association study and linkage analysis, respectively. Using the non-centrality parameters, we perform power calculations and comparisons. The technical details to calculate the regression coefficients, parameters, non-centrality parameters are left in the Appendixes.

Models

Consider a quantitative trait locus Q which has two alleles Q_1 and Q_2. Suppose that the allele frequencies of Q_1 and Q_2 are q_1 and q_2, respectively. Assume that two markers A and B flank the trait locus Q in an order of AQB. Marker A has two alleles A and a with frequencies P_A and P_a, respectively. Marker B has two alleles B and b with frequencies P_B and P_b, respectively. For a nuclear family of k children and two parents, let us denote their quantitative traits by a vector y=$(y_f, y_m, y_1, \cdots, y_k)^\tau$, genotypes at marker A by a vector $(A_f,A_m,A_1, \cdots, A_k)^\tau$, and genotypes at marker B by a vector $(B_f,B_m,B_1,\cdots, B_k)^\tau$. Here y_f is the trait value of the father, A_f is the genotype of the father at marker A, and B_f is the genotype of the father at marker B. Other notations are defined, similarly, for the mother with subscript m and for the i-th child with subscript i. The log-likelihood is defined by $L = \quad \frac{k+2}{2}\log(2\pi) \quad \frac{1}{2}\log|\Sigma| \quad \frac{1}{2}(\mathbf{y} \quad X\mu)^\tau\Sigma^{\ \ 1}$ $(\mathbf{y} \quad X\mu)$. The notations of the log-likelihood are defined as follows. For the mean component $X\mu$, we consider the following regression equation such as model (1) in Fan and Xiong[14]

$$y_i = \beta + \omega_i\gamma + x_{Ai}\alpha_A + x_{Bi}\alpha_B + z_{Ai}\delta_A + z_{Bi}\delta_B + G_i + e_i, \qquad (1)$$

where β is overall mean, w_i is a row vector of covariates such as gender and age, γ is a column vector of regression coefficients for the covariates w_i, G_i is polygenic effect, e_i is error term. Assume that G_i is normal $N(0, \sigma_G^2)$, and e_i is normal $N(0, \sigma_e^2)$. Moreover, G_i and e_i are independent. x_{Ai}, x_{Bi}, z_{Ai} and z_{Bi} are dummy variables defined by

$$x_{Ai} = \begin{cases} 2P_a & \text{if} \quad A_i = AA \\ P_a \quad P_A & \text{if} \quad A_i = Aa, \\ \quad 2P_A & \text{if} \quad A_i = aa \end{cases} \qquad z_{Ai} = \begin{cases} P_a^2 & \text{if} \quad A_i = AA \\ P_aP_A & \text{if} \quad A_i = Aa, \\ P_A^2 & \text{if} \quad A_i = aa \end{cases}$$

$$x_{Bi} = \begin{cases} 2P_b & \text{if} \quad B_i = BB \\ P_b \quad P_B & \text{if} \quad B_i = Bb, \\ \quad 2P_B & \text{if} \quad B_i = bb \end{cases} \qquad z_{Bi} = \begin{cases} P_b^2 & \text{if} \quad B_i = BB \\ P_bP_B & \text{if} \quad B_i = Bb. \\ P_B^2 & \text{if} \quad B_i = bb \end{cases}$$

α_A, α_B, δ_A, and δ_B are regression coefficients of the dummy variables x_{Ai}, x_{Bi}, z_{Ai} and z_{Bi}.

The model matrix X is defined by

$$X = \left\{ \begin{matrix} 1 & w_f & x_{Af} & x_{Bf} & z_{Af} & z_{Bf} \\ 1 & w_m & x_{Am} & x_{Bm} & z_{Am} & z_{Bm} \\ 1 & w_1 & x_{A1} & x_{B1} & z_{A1} & z_{B1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & w_k & x_{Ak} & x_{Bk} & z_{Ak} & z_{Bk} \end{matrix} \right\} = \begin{pmatrix} X_f^\tau \\ X_m^\tau \\ X_1^\tau \\ \vdots \\ X_k^\tau \end{pmatrix},$$

and μ=$(\beta, \gamma^\tau, \alpha_A, \alpha_B, \delta_A, \delta_B)^\tau$ is a vector of regression coefficients. Σ is a $(k+2) \times (k+2)$

variance-covariance matrix defined as

$$\Sigma = \begin{pmatrix} 1 & 0 & \rho_0 & \rho_0 & \cdots & \rho_0 \\ 0 & 1 & \rho_0 & \rho_0 & \cdots & \rho_0 \\ \rho_0 & \rho_0 & 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_0 & \rho_0 & \rho_{21} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_0 & \rho_0 & \rho_{k1} & \rho_{k2} & \cdots & 1 \end{pmatrix} \sigma^2, \text{ where } \sigma^2 =$$

$\sigma_g^2 + \sigma_G^2 + \sigma_e^2, \sigma_g^2$ is variance explained by the putative QTL Q, σ_G^2 is polygenic variance, and σ_e^2 is error variance. The genetic variances $\sigma_g^2 = \sigma_{ga}^2 + \sigma_{gd}^2$ and $\sigma_G^2 = \sigma_{Ga}^2 + \sigma_{Gd}^2$ are decomposed into additive and dominant components. $\rho_0 = (\sigma_{ga}^2 + \sigma_{Ga}^2)/(2\sigma^2)$ is correlation between parents and children, $\rho_{ij} = (\pi_{ijQ}\sigma_{ga}^2 + \Delta_{ijQ}\sigma_{gd}^2 + \sigma_{Ga}^2/2 + \sigma_{Gd}^2/4)/\sigma^2$ is correlation between the *i*-th child and the *j*-th child, π_{ijQ} is the proportion of alleles shared identical by descent (IBD) at QTL Q by the *i*-th child and the *j*-th child, and Δ_{ijQ} is the probability that both alleles at QTL Q shared by the *i*-th child and the *j*-th child are IBD.

For population data, an intuitive rationale of regression model (1) is given in Fan and Xiong[14]. In general, one can construct a variance-covariance matrix for any type of pedigree in a similar way as above. Assume that there are two independent sub-samples of data: (1) population data: n independent individuals; (2) family data: $I-n$ $(I>n)$ independent families. Let us list the log-likelihood of the n independent individuals by $L_1, \cdots, L_n$, and the likelihood of the $I-n$ families by $L_{n+1}, \cdots, L_I$. Then the overall log-likelihood is $L = \sum_{i=1}^{I} L_i$. The unknown parameters are $\mu = (\beta, \gamma, \alpha_A, \alpha_B, \delta_A, \delta_B)^\tau, \sigma_{ga}^2, \sigma_{gd}^2, \sigma_{Ga}^2, \sigma_{Gd}^2$, and σ_e^2. Using the likelihood ratio tests, one may test statistical significance of the parameters of interest.

Parameter estimations and regression coefficients

Regression coefficients Let μ_{ij} be the effect of genotype Q_iQ_j, i, j=1, 2, μ_{12} =μ_{21}. Denote the population effect mean by $\mu = \mu_{11}q_1^2 + 2\mu_{12}q_1q_2 + \mu_{22}q_2^2$ and define α_Q=$q_1\mu_{11}$ +$(q_2-q_1)\mu_{12}-q_2\mu_{22}$, δ_Q=$2\mu_{12}-\mu_{11}-\mu_{22}$. If μ_{11}=a, μ_{12}=d, and μ_{22}=$-a$ as in the traditional quantitative genetics,[15] α_Q= a+$(q_2-q_1)d$ is the average allele substitution effect, and δ_Q =2d characterizes the dominant effect. In general, one may define a=$\mu_{11}-(\mu_{11}+\mu_{22})/2$ and d=$\mu_{12}-(\mu_{11}+\mu_{22})/2$. It is well known that the additive variance $\sigma_{ga}^2 = 2q_1q_2\alpha_Q^2$ and the dominant variance $\sigma_{gd}^2 = (q_1q_2)^2\delta_Q^2$. A true random effect model describing the trait value is y_i=β+$w_i\gamma$+g_i+G_i+e_i, where

$$g_i = \begin{cases} \mu_{11} & \text{for genotype} \quad Q_1Q_1 \\ \mu_{12} & \text{for genotype} \quad Q_1Q_2 \, . \\ \mu_{22} & \text{for genotype} \quad Q_2Q_2 \end{cases}$$

Denote LD coefficient between trait locus Q and marker A by D_{AQ}=$P(AQ_1)-q_1P_A$, LD coefficient between trait locus Q and marker B by D_{QB}=$P(BQ_1)-q_1P_B$, and LD coefficient between marker A and marker B by D_{AB}=$P(AB)-P_AP_B$. Let the additive and dominant variance–covariance matrices be

$$V_A = \begin{pmatrix} 2P_aP_A & 2D_{AB} \\ 2D_{AB} & 2P_bP_B \end{pmatrix}, \text{ and } V_D = \begin{pmatrix} P_a^2P_A^2 & D_{AB}^2 \\ D_{AB}^2 & P_b^2P_B^2 \end{pmatrix}. \qquad (2)$$

Moreover, let us denote three ratios $D_{AB}^2/(P_aP_AP_bP_B) = R_{AB}^2, D_{AQ}^2/(P_aP_Aq_1q_2) = R_{AQ}^2$, and $D_{QB}^2/(q_1q_2P_bP_B) = R_{QB}^2$. As in Appendix B,[14] we can show that the coefficients of regression equation (1) are given by

$$\begin{pmatrix} \alpha_A \\ \alpha_B \end{pmatrix} = V_A^{\ 1} \begin{pmatrix} 2D_{AQ} \\ 2D_{QB} \end{pmatrix} \alpha_Q = \begin{pmatrix} \frac{R_{AQ} \quad R_{AB}R_{QB}}{\sqrt{P_AP_a}} \\ \frac{R_{QB} \quad R_{AB}R_{AQ}}{\sqrt{P_BP_b}} \end{pmatrix} \frac{\sqrt{q_1q_2}\alpha_Q}{1 \quad R_{AB}^2}, \qquad (3)$$

$$\begin{pmatrix} \delta_A \\ \delta_B \end{pmatrix} = V_D^{\ 1} \begin{pmatrix} D_{AQ}^2 \\ D_{QB}^2 \end{pmatrix} \delta_Q = \begin{pmatrix} \frac{R_{AQ}^2 \quad R_{AB}^2R_{QB}^2}{P_AP_a} \\ \frac{R_{QB}^2 \quad R_{AB}^2R_{AQ}^2}{P_BP_b} \end{pmatrix} \frac{q_1q_2\delta_Q}{1 \quad R_{AB}^4}. \qquad (4)$$

Parameters of variance–covariances Denote the recombination fraction between trait locus Q and marker A by θ_{AQ}, the recombination fraction between trait locus Q and marker B by θ_{QB}, and the recombination fraction between marker A and marker B by θ_{AB}. Fulker and Cardon[16] proposed to estimate the proportion π_{ijQ} of allele IBD at putative QTL Q for a sib-pair i and j by $\hat{\pi}_{ijQ} = \mathrm{E}(\pi_{ijQ}|\pi_{ijA}, \pi_{ijB}) = \alpha_\pi + \beta_{\pi A}\pi_{ijA} + \beta_{\pi B}\pi_{ijB}$ where π_{ijA} and π_{ijB} are the IBD proportions of alleles shared at the marker A and marker B, respectively. The coefficients α_π, $\beta_{\pi A}$ and $\beta_{\pi B}$ are given by

$$\beta_{\pi A} = \frac{(1 \quad 2\theta_{AQ})^2 \quad (1 \quad 2\theta_{AB})^2(1 \quad 2\theta_{QB})^2}{1 \quad (1 \quad 2\theta_{AB})^4},$$

$$\beta_{\pi B} = \frac{(1 \quad 2\theta_{QB})^2 \quad (1 \quad 2\theta_{AB})^2(1 \quad 2\theta_{AQ})^2}{1 \quad (1 \quad 2\theta_{AB})^4},$$

$$\alpha_\pi = \frac{1 \quad \beta_{\pi A} \quad \beta_{\pi B}}{2}.$$

Let Δ_{ijA}, Δ_{ijB} be the probability of sharing two alleles IBD at markers A and B for a pair of sibs, respectively. In Fan,[17] we proposed to estimate Δ_{ijQ} by equation $\hat{\Delta}_{ijQ} = \alpha + \beta_A\pi_{ijA} + \beta_B\pi_{ijB} + r_A\Delta_{ijA} + r_B\Delta_{ijB}$. Under the assumption of no interference, the coefficients are as follows (Fan[17]):

$$r_A = \frac{(1 \quad 2\theta_{AQ})^4 \quad (1 \quad 2\theta_{AB})^4(1 \quad 2\theta_{AB})^4}{1 \quad (1 \quad 2\theta_{AB})^8},$$

$$r_B = \frac{(1 \quad 2\theta_{QB})^4 \quad (1 \quad 2\theta_{AQ})^4(1 \quad 2\theta_{AB})^4}{1 \quad (1 \quad 2\theta_{AB})^8},$$

$$\beta_A = \beta_{\pi A} \quad r_A, \beta_B = \beta_{\pi B} \quad r_B,$$

$$\alpha = \frac{(1 \quad \psi_A)^2(1 \quad \psi_B)^2}{[\psi_A\psi_B + (1 \quad \psi_A)(1 \quad \psi_B)]^2},$$

where $\psi_A = \theta_{AQ}^2 + (1 \quad \theta_{AQ})^2$ and $\psi_B = \theta_{QB}^2 + (1 \quad \theta_{QB})^2$. Assuming that the positions of marker A and marker B are known, θ_{AB} can be calculated through Haldane's map function. Then only one of θ_{AQ} and θ_{QB} is unknown since the other can be calculated through Trow's formula.[18] For general relatives i and j, Almasy and Blangero[19] proposed an algorithm to calculate the proportion π_{ijQ} of allele IBD at putative QTL Q, and the expected probability Δ_{ijQ} that both alleles at QTL Q are IBD. In Fan,[17] we derived formulas to calculate the covariances of trait values for a few types of relatives directly without performing matrix operations.

Association and linkage studies From equations (3) and (4), we can see that the coefficients of LD (i.e., D_{AQ} and D_{QB}) and gene effects (i.e., α_Q and δ_Q) are contained in the regression coefficients. Moreover, we show in the above paragraph that the linkage parameters (i.e., recombination fractions θ_{AQ}, θ_{QB} and θ_{AB}) are contained in the variance-covariance matrix. Assume that markers A and B are in LD with the trait locus Q, i.e., $D_{AQ}\neq 0, D_{QB}\neq 0$. We may simultaneously test LD of marker A and marker B with trait locus Q, the gene substitution and dominant effects by testing $\alpha_A=\alpha_B=\delta_A=\delta_B=0$. From equation (3), we may test LD of markers A and B with the trait locus Q and the gene substitution effect α_Q by testing $\alpha_A=\alpha_B=0$. From equation (4), we may test LD of markers A and B with the trait locus Q and the dominant effect by testing $\delta_A=\delta_B=0$.

To test linkage, one may use the likelihood ratio test of the log-likelihood L. Under the null hypothesis of no linkage between the major trait locus Q and the markers, $\theta_{AQ}=\theta_{QB}=1/2$. Under the alternative hypothesis of linkage, $\theta_{AQ}\neq 1/2$ or $\theta_{QB}\neq 1/2$. By comparing the difference of maximum log-likelihoods under the alternative and null hypotheses, we may use χ^2 statistic to test the linkage. We will derive analytical formulas to explore the linkage interval mapping by the nuclear families in a similar way to Sham *et al*[9] according to statistical theory of likelihood ratio tests.[20]

Non-centrality parameters of association study

Assume that there are no covariates. Then $\mu=(\beta, \alpha_A, \alpha_B, \delta_A, \delta_B)^\tau$. Consider the overall log-likelihood $L=\sum_{i=1}^{I} L_i$, where L_i is the log-likelihood of trait values y_i of the i-th family or individual. Let Σ_i be the variance-covariance matrix of y_i, and X_i be its model matrix. Denote the total trait values $y=(y_1^\tau,\cdots,y_1^{\hat{\tau}})^\tau$, the total variance–covariance matrix by $\Sigma=diag(\Sigma_1,\cdots,\Sigma_I)$, and the model matrix $X=(X_1^\tau,\cdots,X_I^\tau)^\tau$. Let $\beta,\hat{\alpha}_A,\hat{\alpha}_B,\hat{\delta}_A,\hat{\delta}_B,\hat{\Sigma}_i,\hat{\Sigma}$ be the maximum likelihood estimators of β, α_A, α_B, δ_A, δ_B, Σ_i, Σ. The estimate of μ is $\hat{\mu}=[X^\tau\hat{\Sigma}^{\ 1}X]^{\ 1}X^\tau\hat{\Sigma}^{\ 1}\vec{y}=[\sum_{i=1}^{I}X_i^\tau\hat{\Sigma}_i^{\ 1}X_i]^{\ 1}\sum_{i=1}^{I}X_i^\tau\hat{\Sigma}_i^{\ 1}\vec{y}_i$. Let H be a $q\times 5$ test matrix of rank q. Suppose that the total number of individuals is N. By Graybill,[21] Chapter 6, the test statistic of a hypothesis $H\mu=0$ is non-central $F(q, N-5)$ defined by

$$F=\frac{(H\hat{\mu})^\tau[H(X^\tau\hat{\Sigma}^{\ 1}X)^{\ 1}H^\tau]^{\ 1}(H\hat{\mu})}{Y^\tau[\hat{\Sigma}^{\ 1}\quad\hat{\Sigma}^{\ 1}X(X^\tau\hat{\Sigma}^{\ 1}X)^{\ 1}X^\tau\hat{\Sigma}^{\ 1}]Y}\,\frac{N\quad 5}{q}.$$

The non-centrality parameter of the test statistic F can be calculated by $\lambda=(H\mu)^\tau[H[X^\tau\Sigma^{\ 1}X]^{\ 1}H^\tau]^{\ 1}H\mu$. If the data are composed of n individuals of a population, Fan and Xiong[14] worked out the non-centrality parameters to test if there are allele substitution and/or dominant effects and LDs between the markers and the major gene locus. In the following, we discuss a situation that the data are composed of both individual population data and family data.

Suppose that there are n individuals of a population, and n is sufficiently large. For each y_i of the n individuals, $\Sigma_i=\sigma^2$ and $X_i=(1\ x_{Ai}\ x_{Bi}\ z_{Ai}\ z_{Bi})$, $i=1, 2,\cdots, n$. From formulas in Fan and Xiong,[14] Appendix A and Appendix B, we may show that

$$\frac{1}{n}\sum_{i=1}^{n}X_i^\tau\Sigma_i^{\ 1}X_i=\frac{1}{n\sigma^2}\sum_{i=1}^{n}X_i^\tau X_i\approx\frac{1}{\sigma^2}diag(1, V_A, V_D), \qquad (5)$$

where V_A and V_D are additive and dominant variance-covariance matrices given in (2).

Secondly, suppose that there are m trio families, and m is sufficiently large. A trio family is composed of both parents and a single child. Notice that the means of x_{Ai}, x_{Bi}, z_{Ai} and z_{Bi} are 0. Let $K_f=(x_{Af}\ x_{Bf}\ z_{Af}\ z_{Bf})$ and $K_m=(x_{Am}\ x_{Bm}\ z_{Am}\ z_{Bm})$. We show in Appendix A that the covariance matrix between parents and their offspring is

$$\mathrm{E}\ K_f^\tau K_1=\mathrm{E}\ K_m^\tau K_1=\begin{pmatrix} V_A/2 & O_2\\ O_2 & O_2\end{pmatrix}, \qquad (6)$$

where $K_1=(x_{A1}\ x_{B1}\ z_{A1}\ z_{B1})$ and O_2 is zero 2×2 matrix. For each of the trio families, the variance–covariance Σ_i is a 3×3 matrix whose inverse is

$$\Sigma_i^{\ 1}=\frac{1}{(1\quad 2\rho_0^2)\sigma^2}\begin{pmatrix} 1 & \rho_0^2 & \rho_0^2 & \rho_0\\ \rho_0^2 & 1 & \rho_0^2 & \rho_0\\ & \rho_0 & \rho_0 & 1\end{pmatrix}. \qquad (7)$$

Using equations (5), (6), and (7), we show in Appendix B

$$\frac{1}{m}\sum_{i=n+1}^{n+m}X_i^\tau\Sigma_i^{\ 1}X_i\approx\frac{1}{(1\quad 2\rho_0^2)\sigma^2}\begin{pmatrix} 3\quad 4\rho_0 & 0 & 0\\ 0 & (3\quad 2\rho_0\quad 2\rho_0^2)V_A & 0\\ 0 & 0 & (3\quad 2\rho_0^2)V_D\end{pmatrix}. \qquad (8)$$

Thirdly, suppose that there are k nuclear families each of them has both parents and two offspring, and the correlation of the two offspring is ρ_{12}. Assume that k is sufficiently large. For each family, the variance–covariance Σ_i is a 4×4 matrix whose inverse is

$$\Sigma_i^{\ 1}=\frac{1}{\sigma^2}\begin{pmatrix} 1+2C\rho_0 & 2C\rho_0 & C & C\\ 2C\rho_0 & 1+2C\rho_0 & C & C\\ C & C & \frac{C(1\ \ 2\rho_0^2)}{\rho_0(1\ \ \rho_{12})} & \frac{C(\rho_{12}\ \ 2\rho_0^2)}{\rho_0(1\ \ \rho_{12})}\\ C & C & \frac{C(\rho_{12}\ \ 2\rho_0^2)}{\rho_0(1\ \ \rho_{12})} & \frac{C(1\ \ 2\rho_0^2)}{\rho_0(1\ \ \rho_{12})}\end{pmatrix}, \qquad (9)$$

where $C=\rho_0(1\quad\rho_{12})/[(1\quad 2\rho_0^2)^2\quad(\rho_{12}\quad 2\rho_0^2)^2]$. In Appendix C, we show that the covariance matrix between two offspring is

$$\mathrm{E}(x_{A1}x_{B1}z_{A1}z_{B1})^{\tau}(x_{A2}\ x_{B2}z_{A2}z_{B2}) = \begin{pmatrix} V_A/2 & O_2 \\ O_2 & V_D/4 \end{pmatrix}. \quad (10)$$

Using equations (5), (6), (9) and (10), we show in Appendix D that

$$\frac{1}{k}\sum_{i=n+m+1}^{n+m+k} X_i^{\tau}\Sigma_i^{\ \ 1}X_i \approx diag(d_{11}, d_{22}V_A, d_{44}V_D)/\sigma^2, \quad (11)$$

where the constants are given by $d_{11} = 2[1 + 4C\rho_0\quad 4C + C/\rho_0], d_{22} = 2 + 4C(\rho_0\quad 1) + C(2\quad \rho_{12}\quad 2\rho_0^2)/[\rho_0(1\quad \rho_{12})]$, $d_{44} = 2(1+2C\rho_0) + C[4(1\quad 2\rho_0^2)\quad (\rho_{12}\quad 2\rho_0^2)]/[2\rho_0(1\quad \rho_{12})]$. Combine the n individuals, m trio families, and k families with two offspring. Define $a_1 = n + m(1\quad 2\rho_0^2)^{\ \ 1}(3\quad 4\rho_0) + kd_{11},\ a_2 = n + m(1\quad 2\rho_0^2)^{\ \ 1}(3\quad 2\rho_0\quad 2\rho_0^2) + kd_{22}, a_3 = n + m(1\quad 2\rho_0^2)^{\ \ 1}(3\quad 2\rho_0^2) + kd_{44}$. Then equations (5), (8) and (11) lead to

$$\sum_{i=1}^{n+m+k} X_i^{\tau}\Sigma_i^{\ \ 1}X_i \approx diag(a_1, a_2V_A, a_3V_D)/\sigma^2. \quad (12)$$

To test if there are additive and dominant effects, we may test the hypothesis $H_{AB,ad}$: α_A=α_B=δ_A=δ_B=0. Then the test matrix H is defined by

$$H = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Let us denote the corresponding F-test statistic by $F_{AB,ad}$, and the non-centrality parameter by $\lambda_{AB,ad}$. Then we have from (3), (4), and (12) that

$$\begin{aligned}\lambda_{AB,ad} &\approx \frac{1}{\sigma^2}\left[a_2(\alpha_A\alpha_B)V_A\begin{pmatrix}\alpha_A\\ \alpha_B\end{pmatrix} + a_3(\delta_A\delta_B)V_D\begin{pmatrix}\delta_A\\ \delta_B\end{pmatrix}\right] \\ &= \frac{1}{\sigma^2}\Big[2a_2\alpha_Q^2[P_bP_BD_{AQ}^2\quad 2D_{AQ}D_{AB}D_{QB} + P_aP_AD_{QB}^2]/ \\ &\quad (P_aP_AP_bP_B\quad D_{AB}^2) + a_3\delta_Q^2[P_b^2P_B^2D_{AQ}^4\quad 2D_{AQ}^2D_{AB}^2D_{QB}^2 \\ &\quad + P_a^2P_A^2D_{QB}^4]/(P_a^2P_A^2P_b^2P_B^2\quad D_{AB}^4)\Big] \\ &= \frac{1}{\sigma^2}\Big[a_2\sigma_{ga}^2[R_{AQ}^2\quad 2R_{AQ}R_{AB}R_{QB} + R_{QB}^2](1\quad R_{AB}^2) \\ &\quad + a_3\sigma_{gd}^2[R_{AQ}^4\quad 2R_{AQ}^2R_{AB}^2R_{QB}^2 + R_{QB}^4]/(1\quad R_{AB}^4)\Big].\end{aligned}$$

Assume that the two markers A and B are in linkage equilibrium, then D_{AB}=0. Moreover, assume that the trait locus Q is in LD with marker A but not with marker B, then D_{QB}=0 and D_{AQ}≠0. Then $\lambda_{AB,ad} \approx [1/\sigma^2]\left[a_2\sigma_{ga}^2R_{AQ}^2 + a_3\sigma_{gd}^2R_{AQ}^4\right]$, which only involves marker A and can be written as $\lambda_{A,ad}$. Correspondingly, we denote the F-test statistic by $F_{A,ad}$. Similarly, $\lambda_{A,a} \approx [a_2/\sigma^2]\sigma_{ga}^2R_{AQ}^2$ is the non-centrality parameter of a test statistic $F_{A,a}$. To test the other hypotheses, we may get the non-centrality parameters in a similar way by taking appropriate test matrices H. To test if there is dominant effect, we may test the hypothesis $H_{AB,d}$: δ_A=δ_B=0. The non-centrality parameter is $\lambda_{AB,d} \approx \frac{a_3}{\sigma^2}\sigma_{gd}^2[R_{AQ}^4\quad 2R_{AQ}^2R_{AB}^2R_{QB}^2 + R_{QB}^4]/(1\quad R_{AB}^4)$. To test if there is an additive or substitution effect, we may test the hypothesis $H_{AB,a} : \alpha_A = \alpha_B = 0$. The non-centrality parameter is $\lambda_{A,Ba} \approx \frac{a_2}{\sigma^2}\sigma_{ga}^2[R_{AQ}^2\quad 2R_{AQ}R_{AB}R_{QB} + R_{QB}^2]/(1\quad R_{AB}^2)$. The corresponding F-test statistic is denoted by $F_{AB,a}$.

Non-centrality parameters of linkage studies

Consider a nuclear family with k children and both parents. Under the null hypothesis of no linkage between the trait locus and markers, the correlation of each sib-pair is

$$\rho_N = \frac{\sigma_{ga}^2}{2\sigma^2} + \frac{\sigma_{gd}^2}{4\sigma^2} + \frac{\sigma_{Ga}^2}{2\sigma^2} + \frac{\sigma_{Gd}^2}{4\sigma^2}.$$

The expected log-likelihood is $\mathrm{E}(2L_{Null}) = \quad (k+2)\left[\log(2\pi\sigma^2)+1\right]\quad \log\left[\ (1\quad 2\rho_0^2)+(k\quad 1)(\rho_N\quad 2\rho_0^2)\right)(1\quad \rho_N)^{k\ \ 1}\right]$. Under the alternative hypothesis of linkage between the trait locus and marker A, the correlation between a sib-pair is $C_i = \mathrm{Cov}(y_1, y_2|\pi_A = i/2)/\sigma^2 = (\sigma_{ga}^2 + \sigma_{gd}^2)P(\pi_Q = 1|\pi_A = i/2)/\sigma^2 + \frac{\sigma_{ga}^2}{2}P(\pi_Q = 1/2|\pi_A = i/2)/\sigma^2 + [\sigma_{Ga}^2/2 + \sigma_{Gd}^2/4]/\sigma^2, i = 0, 1, 2.$ From Haseman and Elston,[22] Table IV, we have

$$\begin{aligned} C_2 &= \left[(\sigma_{ga}^2 + \sigma_{gd}^2)\psi_A^2 + \sigma_{ga}^2\psi_A(1\quad \psi_A) + \sigma_{Ga}^2/2 + \sigma_{Gd}^2/4\right]/\sigma^2 \\ C_1 &= \Big[(\sigma_{ga}^2 + \sigma_{gd}^2)\psi_A(1\quad \psi_A) + \sigma_{ga}^2[1\quad 2\psi_A(1\quad \psi_A)]/2 \\ &\quad + \sigma_{Ga}^2/2 + \sigma_{Gd}^2/4\Big]/\sigma^2 \\ C_0 &= \left[(\sigma_{ga}^2 + \sigma_{gd}^2)(1\quad \psi_A)^2 + \sigma_{ga}^2\psi_A(1\quad \psi_A) + \sigma_{Ga}^2/2 + \sigma_{Gd}^2/4\right]/\sigma^2. \end{aligned} \quad (13)$$

The expected log-likelihood under the alternative hypothesis of linkage is

$$\begin{aligned}\mathrm{E}(2L_{random,A}) = &\quad (k+2)[\log(2\pi\sigma^2) + 1] \\ & \sum_{\pi_{12A}}\cdots\sum_{\pi_{k\ \ 1,kA}} P(\pi_{12A})\cdots P(\pi_{k\ \ 1,kA}).\end{aligned}$$

$$\log\det\begin{pmatrix} 1 & 0 & \rho_0 & \rho_0 & \cdots & \rho_0 \\ 0 & 1 & \rho_0 & \rho_0 & \cdots & \rho_0 \\ \rho_0 & \rho_0 & 1 & C_{2\pi_{12A}} & \cdots & C_{2\pi_{1kA}} \\ \rho_0 & \rho_0 & C_{2\pi_{21A}} & 1 & \cdots & C_{2\pi_{2kA}} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_0 & \rho_0 & C_{2\pi_{k1A}} & C_{2\pi_{k2A}} & \cdots & 1 \end{pmatrix},$$

where $P(\pi_{ijA}$=0)=$P(\pi_{ijA}$=1)=1/4 and $P(\pi_{ijA}$=1/2)=1/2. From Stuart and Ord,[20] the non-centrality parameter for linkage of the nuclear family is equal to $\lambda_{linkage,A}$=E($2L_{random,A}$)−E($2L_{Null}$). If k=2, it can be shown that $\lambda_{linkage,A} = \log[1\quad 4\rho_0^2\quad \rho_N^2 + 4\rho_0^2\rho_N]\quad \sum_{i=0}^{2} P(\pi_{12A} = i/2)\log[1\quad 4\rho_0^2\quad C_i^2 + 4\rho_0^2C_i]$.

Under the alternative hypothesis of linkage between the trait locus and markers A and B, the correlation between a sib-pair is given by for i, j=0, 1, 2

$$\begin{aligned} C_{ij} &= \mathrm{Cov}(y_1, y_2 | \pi_{12A} = i/2, \pi_{12B} = j/2)/\sigma^2 \\ &= \Big[(\sigma^2_{ga} + \sigma^2_{gd}) P(\pi_{12Q} = 1 | \pi_{12A} = i/2, \pi_{12B} = j/2) \\ &\quad + \frac{\sigma^2_{ga}}{2} P(\pi_{12Q} = 1/2 | \pi_{12A} = i/2, \pi_{12B} = j/2) \\ &\quad + \sigma^2_{Ga}/2 + \sigma^2_{Gd}/4\Big]/\sigma^2. \end{aligned} \tag{14}$$

To calculate C_{ij} , we need to calculate the joint distribution of π_{12A}, π_{12Q} and π_{12B} of a sib-pair under the alternative hypothesis of linkage. Assume that there is no interference for disjoint regions of the chromosome. Then we have

$$\begin{aligned} & P(\pi_{12A} = i_A, \pi_{12Q} = i_Q, \pi_{12B} = i_B) \\ &= P(\pi_{12A} = i_A, \pi_{12Q} = i_Q) P(\pi_{12B} = i_B | \pi_{12A} = i_A, \pi_{12Q} = i_Q) \\ &= P(\pi_{12A} = i_A | \pi_{12Q} = i_Q) P(\pi_{12Q} = i_Q) P(\pi_{12B} = i_B | \pi_{12Q} = i_Q). \end{aligned} \tag{15}$$

From Haseman and Elston,[22] Table IV, we may construct the joint distribution of π_{12Q}, π_{12A} and $\pi_{12\mathrm{B}}$ by relation (15), and the results are presented in Table 3 of Fan.[17] Based on the results, we can calculate C_{ij}, i, j=0, 1, 2, which are given in Appendix D of Fan.[17] The expected log-likelihood under the alternative hypothesis of linkage is $\mathrm{E}(2L_{random,AB}) = -(k+2)[\log(2\pi\sigma^2)+1] - \Sigma_{\pi_{12A}} \ \Sigma_{\pi_{12B}} \cdots \Sigma_{\pi_{k-1,kA}} \ \Sigma_{\pi_{k-1,kB}} \ P(\pi_{12A})P(\pi_{12B}) \cdots P(\pi_{k-1,kA}) \ P(\pi_{k-1,kB})$

$$\log\det \begin{pmatrix} 1 & 0 & \rho_0 & \rho_0 & \cdots & \rho_0 \\ 0 & 1 & \rho_0 & \rho_0 & \cdots & \rho_0 \\ \rho_0 & \rho_0 & 1 & C_{2\pi_{12A}, 2\pi_{12B}} & \cdots & C_{2\pi_{1kA}, 2\pi_{1kB}} \\ \rho_0 & \rho_0 & C_{2\pi_{21A}, 2\pi_{21B}} & 1 & \cdots & C_{2\pi_{2kA}, 2\pi_{2kB}} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_0 & \rho_0 & C_{2\pi_{k1A}, 2\pi_{k1B}} & C_{2\pi_{k2A}, 2\pi_{k2B}} & \cdots & 1 \end{pmatrix},$$

where $P(\pi_{ijB}=0)=P(\pi_{ijB}=1)=1/4$ and $P(\pi_{ijB}=1/2)=1/2$ such as those for marker A. From Stuart and Ord[20], the non-centrality parameter for linkage of the nuclear family is equal to $\lambda_{linkage,AB} = \mathrm{E}(2L_{random,AB}) - \mathrm{E}(2L_{Null})$. If k=2, it can be shown that $\lambda_{linkage,AB} = \log[1 \quad 4\rho_0^2 \quad \rho_N^2 + 4\rho_0^2\rho_N] \quad \sum_{i,j=0}^{2} P(\pi_{12A} = i/2) P(\pi_{12B} = j/2) \log[1 \quad 4\rho_0^2 \quad C_{ij}^2 + 4\rho_0^2 C_{ij}]$.

Power calculation and comparison

Let us denote heritability by h^2 which is defined by $h^2 = \sigma^2_{ga}/\sigma^2$. In the power calculations, we take the additive polygenic variance $\sigma^2_{Ga} = 0.10$, polygenic dominant variance $\sigma^2_{Gd} = 0.05$, the equal allele frequencies P_A=q_1=P_B=0.5 at the two markers A and B, and the QTL Q. Moreover, suppose that μ_{11}=a, μ_{12}=μ_{21}=d and μ_{22}=$-a$.

Suppose that the map distance λ_{AB} between marker A and marker B is known. Under the assumption of no interference, we may calculate the recombination fraction by Haldane's map function θ_{AB}=[1$-$exp($-$2λ_{AB})]/2. Similarly, we may calculate the recombination fractions θ_{AQ} and θ_{QB} by the map distances λ_{AQ} and λ_{QB}. Assume that marker A and marker B are in linkage equilibrium, i.e., D_{AB}=0, the genetic distances λ_{AB}=5 cM, λ_{AQ}=λ_{QB}=2.5 cM, and the heritability h^2=0.25. Suppose we have a sample with n=100 individuals, m=30 trio families, and k=20 nuclear families with two offspring. Assume that the IBD proportions shared by the two offspring in the k=20 families at both markers A and B are π_A= π_B=0.5, and the probability of sharing two alleles IBD at markers A and B are Δ_A=Δ_B=0.5. Figure 1 shows the power of the test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ against the disequilibrium coefficient D_{AQ} when D_{QB}=0.15 for a mode of dominant inheritance with a=d=1.0, and a mode of recessive inheritance with a=1.0, d=$-$0.5, respectively. Several features are interesting in the two graphs of Figure 1. First, the power of $F_{AB,ad}$ and $F_{AB,a}$ are higher than that of $F_{A,ad}$ and $F_{A,a}$. Hence, the regression mapping which uses two markers A and B has its advantage

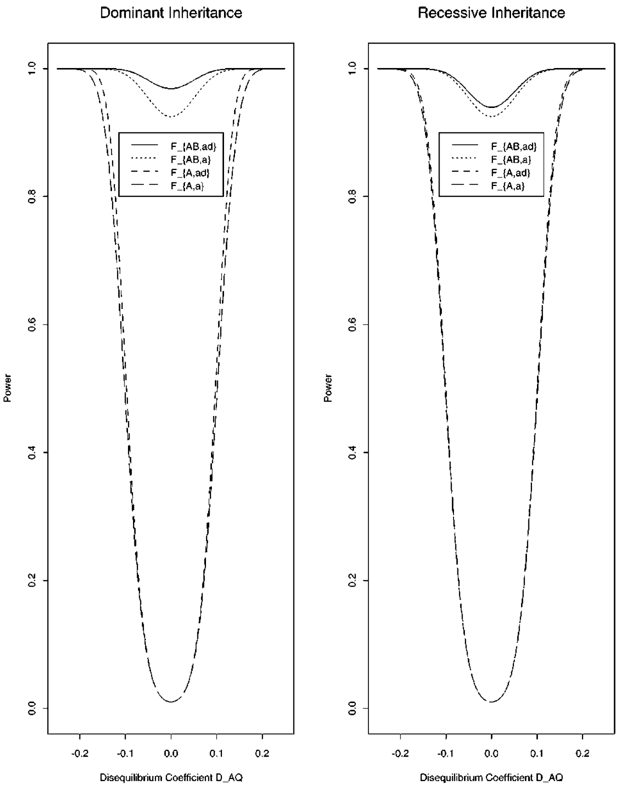

Figure 1 Power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ against disequilibrium coefficient D_{AQ} at 0.01 significant level, when q_1=P_A=P_B=0.50, D_{AB}=0.0, D_{QB}=0.15, h^2=0.25, n=100, m=30, k=20, π_A=π_B=Δ_A=Δ_B=0.5, λ_{AB}=5 cM, λ_{AQ}=λ_{QB}=2.5 cM, $\sigma^2_{Ga} = 0.10$, $\sigma^2_{Gd} = 0.05$ for a mode of dominant inheritance a=d=1.0, and a mode of recessive inheritance a=1.0, d=$-$0.5, respectively.

over the one marker mapping which only uses one marker A or B. Second, the statistic $F_{AB,ad}$ has higher power than that of $F_{AB,a}$, and the statistic $F_{A,ad}$ has higher power than that of $F_{A,a}$. Thus, excluding the dominant variance from the analysis when it does exist would lose power. Third, as expected, when D_{AQ}=0 the power to detect LD using only marker A is minimal. More interestingly, when D_{AQ}=0.15 the power is still higher using the flanking two markers than using only marker A.

Figure 2 shows the power of the test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ against the heritability h^2 when D_{AB}=0.10 and D_{AQ}=D_{QB}=0.15 for a mode of dominant inheritance with a=d=1.0, and a mode of recessive inheritance with a=1.0, d=−0.5, respectively. The other parameters are the same as those of Figure 1. Among the features observed in Figure 1, the power is reasonably high when the heritability h^2 is bigger than 0.15. To compare the power of population based and family based methods, Figure 3 shows the power of the test statistics $F_{AB,ad}$ and $F_{AB,a}$ for a mode of dominant inheritance with a=d=1.0, and a mode of recessive inheritance with a=1.0, d=−0.5, respectively. For Figure 3, population data contain n=252 individuals, but no family data (m=k=0). For dominant inheritance of Figure 3, the data contain m=84 trio families (n=k=0). For recessive inheritance of Figure 3, the data contain k=63 nuclear families each has two offspring (n=m=0). Notice that m=84 or k=63 family data contain 252 individuals, and thus the number of individuals is the same as that of the population data. We can see that population based method is more powerful than the family based method for the same number of individuals.

In a population, the LD can exist due to mutations at the trait locus. In the absence of tight linkage between the trait locus and a marker, the recombination between the marker locus and the trait locus can rapidly dissipate the disequilibrium from generation to generation. Denote the frequency of haplotype AQ at the generation when the mutations occur by $P(AQ)(0)$. Then LD coefficient is $D_{AQ}(0)$= $P(AQ)(0)-q_1P_A$ for the generation when the mutations occur. For the following generations, the disequilibrium

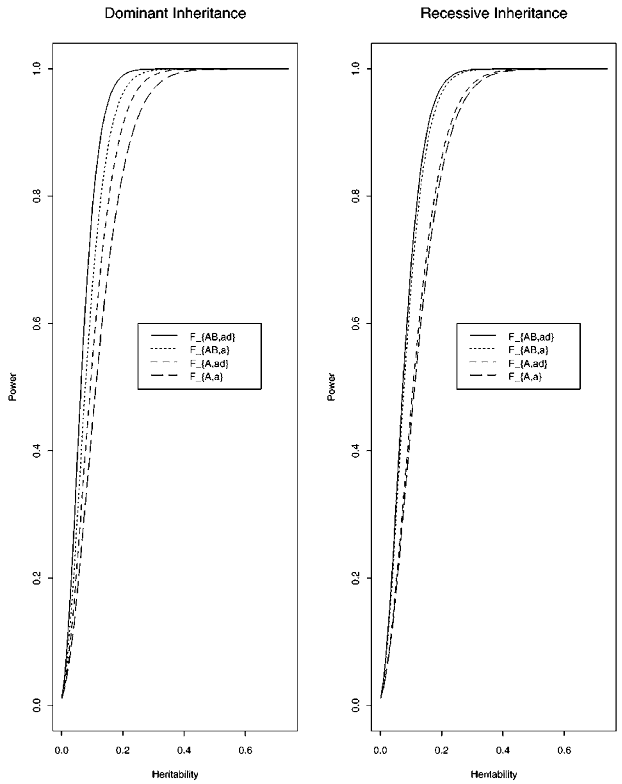

Figure 2 Power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ against heritability h^2 at 0.01 significant level, when q_1=P_A=P_B= 0.50, D_{AB}=0.10, D_{AQ}=D_{QB}=0.15, n=100, m=30, k=20, π_A=π_B= Δ_A=Δ_B=0.5, λ_{AB}=5 cM, λ_{AQ}=λ_{QB}=2.5 cM, $\sigma^2_{Ga}=0.10$, $\sigma^2_{Gd}=0.05$ for a mode of dominant inheritance a=d=1.0, and a mode of recessive inheritance a=1.0, d=−0.5, respectively.

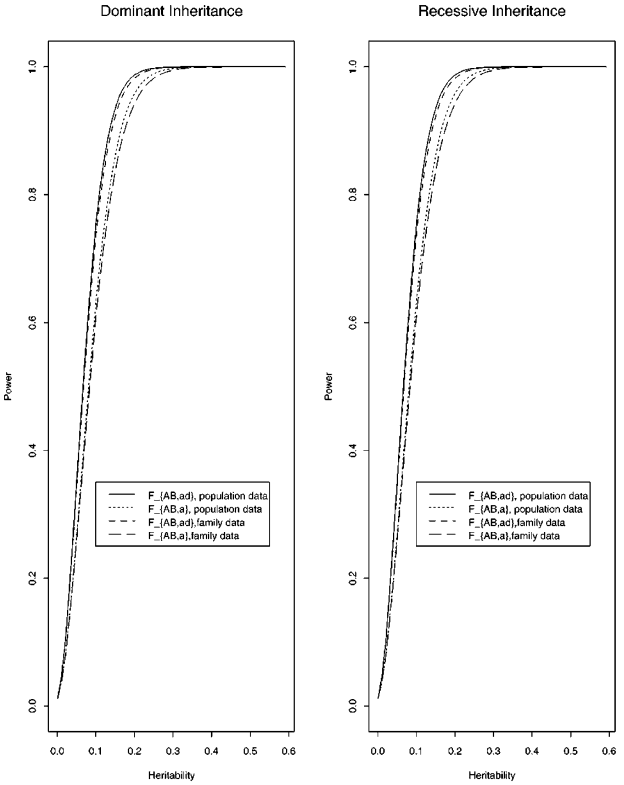

Figure 3 Power of test statistics $F_{AB,ad}$ and $F_{AB,a}$ against heritability h^2 at 0.01 significant level for a mode of dominant inheritance a=d=1.0, and a mode of recessive inheritance a=1.0, d=−0.5, respectively. For population data n=252, m=k=0; for dominant family data n=k=0, m=84; for recessive family data n=m=0, k=63. Other parameters are the same as those of Figure 2.

coefficient is reduced by a factor $1-\theta_{AQ}$ in each generation.[23] Suppose that the mutation is already T generation old. Then the LD coefficient is $D_{AQ}(T)=D_{AQ}(0)(1-\theta_{AQ})^T$. Similarly, the other LD coefficients are $D_{AB}(T)=D_{AB}(0)(1-\theta_{AB})^T$ and $D_{QB}(T)=D_{QB}(0)(1-\theta_{QB})^{\mathrm{T}}$.

Assume that the map distance between marker A and marker B is λ_{AB}=5 cM, and the other parameters are given by $D_{AB}(0)=0.20, D_{AQ}(0)=D_{QB}(0)=0.25$, h^2=0.25, λ_{AB}=5 cM, n=100, m=30, k=20, T=30, $\pi_A=\pi_B$=0.5, $\Delta_A=\Delta_B$=0.25. Figure 4 shows the power of the test statistics $F_{AB,ad}$, $F_{AB.a}$, $F_{A.ad}$, and $F_{A,a}$ against the recombination fraction θ_{AQ} for a mode of dominant inheritance with a=d=1.0, and a mode of recessive inheritance with a=1.0, d=−0.5, respectively. We can see that the power curves of $F_{AB,ad}$ and $F_{AB,a}$ are very high, although the power curves of $F_{A.ad}$ and $F_{A,a}$ decrease very rapidly as the recombination fraction θ_{AQ} increases. Hence, high resolution LD mappings have advantage to do fine gene mappings, and appropriate for the dense marker maps such as single nucleotide polymorphisms on human genome. To investigate the effect of the age of the mutation on the power, Figure 5 shows the power curves against the position of markers. In the Figure, the QTL locates at 15 cM which is flanked by two markers A and B. One marker is one the right-hand side of the QTL, and the other is on the left-hand side with equal distance to the QTL. The power decreases quickly when the age of the mutation increases. For a mutation which is 30 generations old, one should expect very low power if the markers locate 5 cM away from the QTL.

To explore the linkage interval mapping, we take a sample of k=250 nuclear families each has two offspring. Multiplying $\lambda_{linkage,A}$ and $\lambda_{linkage,AB}$ by k, we may calculate the non-centrality parameters for the linkage mapping using marker A and the linkage interval mapping using markers A and B. Moreover, assume that the genetic distances are λ_{AB}=30 cM, and $\lambda_{AQ}=\lambda_{QB}$=15 cM, i.e., the QTL Q is right in the middle between markers A and B. Figure 6 gives power curves of linkage interval mapping by markers A and B, and linkage mapping by marker A against heritability h^2 for a mode of dominant inheritance with

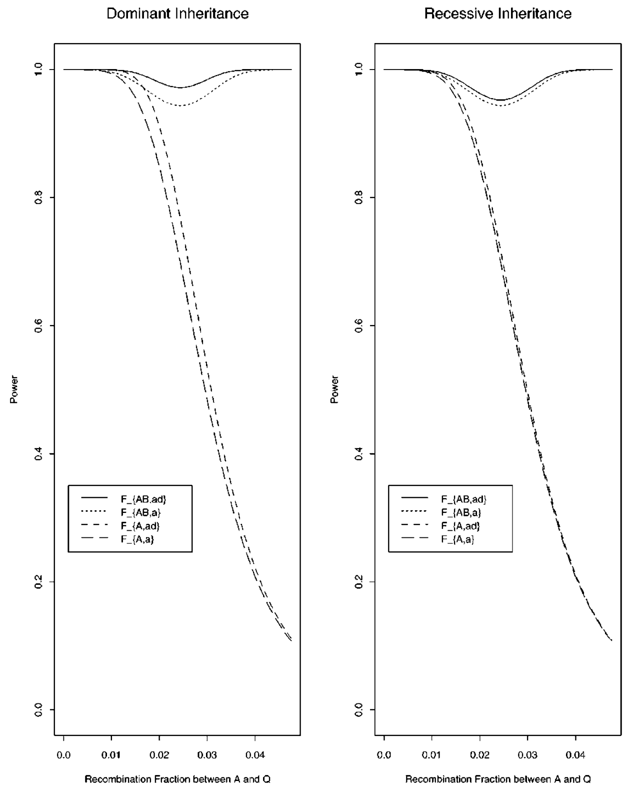

Figure 4 Power curves of the test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ against the recombination fraction θ_{AQ} at 0.01 significant level, when $q_1=P_A=P_B$=0.50, $D_{AB}(0)$=0.20, $D_{AQ}(0)=D_{QB}(0)$= 0.25, h^2=0.25, λ_{AB}=5 cM, T=30, n=100, m=30, k=20, $\pi_A=\pi_B$=0.5, $\Delta_A=\Delta_B$=0.25, $\sigma^2_{Ga}=0.10$, $\sigma^2_{Gd}=0.05$ for a mode of dominant inheritance a=d=1.0, and a mode of recessive inheritance a=1.0, d=−0.5, respectively.

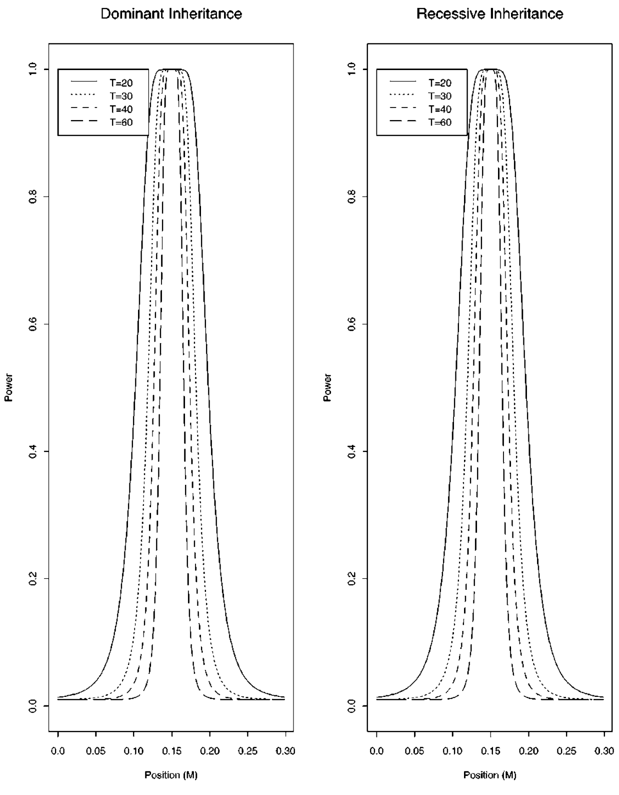

Figure 5 Power curves of the test statistics $F_{AB,ad}$ against the position of markers at 0.01 significant level for a mode of dominant inheritance a=d=1.0, and a mode of recessive inheritance a=1.0, d=−0.5, respectively. The QTL locates at 15 cM which is flanked by two markers A and B. Here the mutation age T=20, 30, 40, 60, and the other parameters are the same as those in Figure 4.

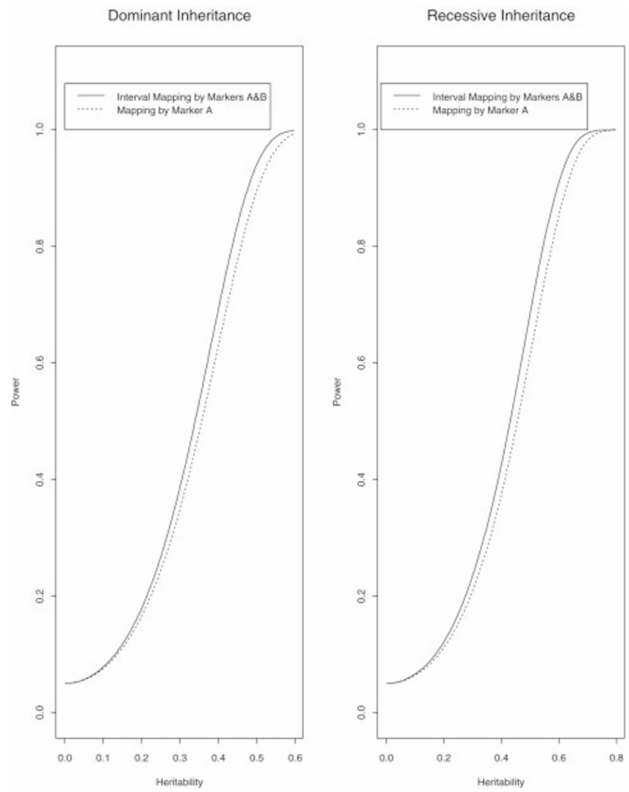

Figure 6 Power curves of the linkage interval mapping by markers *A* and *B*, and linkage mapping by marker *A* against the heritability h^2, when q_1=P_A=P_B=0.50, λ_{AB}=30 cM, λ_{AQ}=λ_{QB}= 15 cM, *k*=250, $\sigma^2_{Ga} = 0.10$, $\sigma^2_{Gd} = 0.05$, at 0.05 significant level for a mode of dominant inheritance *a*=*d*=1.0, and a mode of recessive inheritance *a*=1.0, *d*=−0.5, respectively.

a=*d*=1.0, and a mode of recessive inheritance with *a*=1.0, *d*=−0.5, respectively. It is clear that the power of interval linkage mapping using both markers *A* and *B* is higher than that of linkage mapping using only one marker *A*.

Discussion

In this paper, we investigate variance component models of both high resolution LD mapping and linkage analysis for QTL. The models are based on family pedigree and population data. We consider likelihoods which utilizes flanking marker information. The likelihoods jointly include recombination fractions, LD coefficients, the average allele substitution effect and allele dominant effect as parameters. The linkage parameters are contained in the variance-covariance matrix. The parameters of LD and gene effects are contained in the regression coefficients.[8,9,11,12] The model simultaneously takes care of the linkage, LD and the effects of the putative trait locus *Q*, and hence clearly demonstrates that linkage analysis and LD mapping are complimentary, not exclusive, methods for QTL mapping. The family data which have at least two offspring contain information for both linkage and association, and population data and trio family data which have two parents and only one offspring contain information for association. By combining the family and population data in the analysis, one may expect to get better results than that by analysing them separately.

Linkage analysis can localize genetic trait loci in broad chromosome regions of a few cM (<10 cM), and is less sensitive to population admixture than LD mapping. In practice, one may carry out linkage analysis as a first step to obtain prior suggestive linkage based on a sparse marker map. By performing linkage interval mappings, one may get higher power than that of using only one marker. With prior linkage in hand, LD mapping can be used to get high resolution of the genetic trait loci based on a dense marker map. We have shown that models of high resolution LD mappings using two flanking markers have higher power than that of models of using only one marker. Hence, high resolution LD mappings have the advantages to do fine gene mappings, and appropriate for the dense marker maps such as SNPs on human genome. Performing both LD mapping and linkage analysis has potential to avoid false positives due to population history or environmental effects. In the meantime, it takes the advantage of high resolution of LD mapping.

The power of association study depends on the existence of LD between trait locus and markers. In the absence of LD, the power of LD mappings is very low. To increase the probability of detecting LD, one may need to carry out suitable design for a genetic study.[24] It is well known that the level of LD is heavily affected by population stratification. On the one hand, the family based methods are less likely influenced by population stratification than those of population data based methods. On the other hand, a family based association study is less powerful than that of population based study for the same number of individuals. Combining the family and population data, one may expect more information, and take the advantage of population data and family data. More investigation is needed to explore the population stratification effect on high resolution LD mapping of QTL, and to develop robust methods to identify association between multiple markers and QTL in the presence of population stratification.

To our knowledge, there is not much research on statistical analysis about high resolution LD mapping of QTL. Using only one bi-allelic marker, the statistical analysis of LD mapping has been studied by a few colleagues.[8–13] Relatively, multipoint linkage mapping has been studied more intensively.[16,19,25] It is our hope that the current research may shed more light on the high resolution association study, and stimulate more interests to utilize the advantage of LD mapping in fine resolution of genetic studies. In the Section of power calculation and comparison, we mainly explore a set of scenarios of LD mapping. For several sets of parameters, we compare the power of four test statistics for LD mapping. Moreover, we compare the power of LD mapping of using population data and family data. We also investigate the effect of mutation age on the power. For linkage mapping, we only include one figure to make power comparison of linkage interval mapping using two markers with linkage mapping using only one marker.[9] This reflects the need for more research on high resolution LD mapping of QTL, since the research on linkage interval/multipoint mapping is more mature.

In this paper, we treat LD as a fixed effect since only two markers are considered. In general, inference about the LD structure in the population are desirable, and LD should be modeled as a random effect when multiple markers/haplotypes are used in analysis, which would need more investigation. We assume that the data of all family members are available. For some late-onset diseases, the data for the parents or former family members may no longer be available. In principle, one can use similar methods as the ones proposed in this paper to perform high resolution LD mapping for sib-pair data of late-onset diseases. This is an area which is of importance and needs more research. Due to the length of this paper, we do not pursue these issues in depth, and they will be explored in other projects.

Acknowledgements

We thank two reviewers, and Dr Gert-Jan B van Ommen for their helpful comments to improve the paper. R Fan was supported partially by a research fellowship from the Alexander von Humboldt Foundation, Germany, and an International Research Travel Assistance Grant of the Texas A&M University. M Xiong was supported by NIH grant R01-GM56515, and MH59518. Ms JS Jung kindly writes the SAS micro, which is available from R Fan upon request.

References

1 Botstein D, White RL, Skolnick MH, Davis RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980; **32**: 314–331.
2 Morton NE: Sequential tests for the detection of linkage. *Am J Hum Genet* 1955; **7**: 277–318.
3 Morton NE: Significance levels in complex inheritance. *Am J Hum Genet* 1998; **62**: 690–697.
4 The International SNP Map Working Group: A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; **409**: 928–933.
5 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.
6 Abecasis GR, Cherny SS, Cookson WOC, Cardon LR: Merlin – rapid analysis of dense genetic maps using sparse gene flow tress. *Nature Genetics* 2002; **30**: 97–101.
7 Almasy L, Williams JT, Dyer TD, Blangero J: Quantitative trait locus detection using combined linkage/disequilibrium analysis. *Genetic Epidemiology* 1999; **17** (Suppl 1): S31–S36.
8 Fulker DW, Cherny SS, Sham PC, Hewitt JK: Combined linkage and association sib-pair analysis for quantitative traits. *Am J Hum Genet* 1999; **64**: 259–267.
9 Sham PC, Cherny SS, Purcell S, Hewitt JK: Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* 2000; **66**: 1616–1630.
10 Abecasis GR, Cardon LR, Cookson WOC: A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000; **66**: 279–292.
11 Abecasis GR, Cookson WOC, Cardon LR: Pedigree tests of linkage disequilibrium. *Eur J Hum Genet* 2000; **8**: 545–551.
12 Abecasis GR, Cookson WOC, Cardon LR: The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am J Hum Genet* 2001; **68**: 1463–1474.
13 Cardon LR: A sib-pair regression model of linkage disequilibrium for quantitative traits. *Hum Hered* 2000; **50**: 350–358.
14 Fan R, Xiong M: High resolution mapping of quantitative trait loci by linkage disequilibrium analysis. *Eur J Hum Gen* 2002; **10**: 607–615.
15 Falconer DS, Mackay TFC: *Introduction to Quantitative Genetics*. London: Longman, 1996, 4th edn.
16 Fulker DW, Cardon LR: A sib-pair approach to interval mapping of quantitative trait loci. *Am J Hum Genet* 1994; **54**: 1092–1103.
17 Fan R: Interval mapping of quantitative trait loci. 2002 http://stat.tamu.edu/~rfan/paper.html/interval_mapping.pdf
18 Lange K: *Mathematical and Statistical Methods for Genetic Analysis*. New York: Springer-Verlag, 1997.
19 Almasy L, Blangero J: Multipoint quantitative trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**: 1198–1211.
20 Stuart A, Ord JK: *Kendall's. Advanced Theory of Statistics, Vol. 2: Classical Inference and Relationships*. Oxford, 1991, 5th edn.
21 Graybill FA: *Theory and Application of the Linear Model*. California: Pacific Grove, 1976.
22 Haseman JK, Elston RC: The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics* 1972; **2**: 3–19.
23 Hartl DL, Clark AG: *Principles of Population Genetics*. 2nd edn. Sinauer, 1989.
24 Boehnke M, Langefeld CD: Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am J Hum Genet* 1998; **62**: 950–961.
25 Pratt SC, Daly M, Kruglyak: Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 2000; **66**: 1153–1157.

Appendix A

In this Appendix, we show equation (6). Actually, we have

$$
\begin{aligned}
\mathrm{E}[x_{Af}x_{A1}] &= 2P_a\mathrm{E}[x_{A1}, A_f = AA] + (P_a \quad P_A)\mathrm{E}[x_{A1}, A_f = Aa]\\
&\quad 2P_A\mathrm{E}[x_{A1}, A_f = aa]\\
&= 2P_a\Big[2P_aP_A + (P_a \quad P_A)P_a \quad 2P_A\cdot 0\Big]P_A^2\\
&\quad 2P_A\Big[2P_a\cdot 0 + (P_a \quad P_A)P_A \quad 2P_AP_a\Big]P_a^2\\
&\quad + (P_a \quad P_A)\Big[2P_aP_A + (P_a \quad P_A)(P_a + P_A) \quad 2P_AP_a\Big]\\
&\quad (2P_AP_a/2)\\
&= 2P_aP_aP_A^2 + (P_a \quad P_A)(P_a \quad P_A)P_aP_A\\
&\quad 2P_A(\quad P_AP_a^2) = P_aP_A\\
\mathrm{E}[x_{Af}x_{B1}] &= 2P_a\mathrm{E}[x_{B1}, A_f = AA] + (P_a \quad P_A)\mathrm{E}[x_{B1}, A_f = Aa]\\
&\quad 2P_A\mathrm{E}[x_{B1}, A_f = aa]\\
&= 2P_a\Big[2P_b\cdot P(AB)P_A\cdot P_B + (P_b \quad P_B)\cdot\\
&\quad \Big(P(AB)P_A\cdot P_b + P(Ab)P_A\cdot P_B\Big) \quad 2P_B\cdot P(Ab)P_A\cdot P_b\Big]\\
&\quad + (P_a \quad P_A)\Big[2P_b\cdot\Big(P(AB)P_a\cdot P_B + P(aB)P_A\cdot P_B\Big)\\
&\quad + (P_b \quad P_B)\cdot\Big(P(AB)P_a\cdot P_b + P(Ab)P_a\cdot P_B\\
&\quad + P(aB)P_A\cdot P_b + P(ab)P_A\cdot P_B\Big)\\
&\quad 2P_B\cdot\Big(P(Ab)P_a\cdot P_b + P(ab)P_A\cdot P_b\Big)\Big]\\
&\quad 2P_A\Big[2P_b\cdot P(aB)P_a\cdot P_B\\
&\quad + (P_b \quad P_B)\cdot\Big(P(aB)P_a\cdot P_b + P(ab)P_a\cdot P_B\Big)\\
&\quad 2P_B\cdot P(ab)P_a\cdot P_b\Big]\\
&= 2P_a[P(AB)P_AP_b \quad P(Ab)P_AP_B]\\
&\quad + (P_a \quad P_A)[P(AB)P_aP_b + P(aB)P_AP_b\\
&\quad P(Ab)P_aP_B \quad P(ab)P_AP_B] \quad 2P_A[P(aB)P_aP_b\\
&\quad P(ab)P_aP_B]\\
&= P(AB)P_aP_b \quad P(aB)P_AP_b \quad P(Ab)P_aP_B\\
&\quad + P(ab)P_AP_B = D_{AB}\\
\mathrm{E}[x_{Af}z_{A1}] &= 2P_a\mathrm{E}[z_{A1}, A_f = AA] + (P_a \quad P_A)\mathrm{E}[z_{A1}, A_f = Aa]\\
&\quad 2P_A\mathrm{E}[z_{A1}, A_f = aa]\\
&= 2P_a\Big[\quad P_a^2P_A + (P_aP_A)P_a \quad P_A^2\cdot 0\Big]P_A^2\\
&\quad 2P_A\Big[\quad P_a^2\cdot 0 + P_aP_AP_A \quad P_A^2P_a\Big]P_a^2\\
&\quad + (P_a \quad P_A)\Big[\quad P_a^2P_A + P_aP_A(P_a + P_A)\\
&\quad P_A^2P_a\Big]2P_AP_a/2 = 0\\
\mathrm{E}[x_{Af}z_{B1}] &= 2P_a\mathrm{E}[z_{B1}, A_f = AA] + (P_a \quad P_A)\mathrm{E}[z_{B1}, A_f = Aa]\\
&\quad 2P_A\mathrm{E}[z_{B1}, A_f = aa]\\
&= 2P_a\Big[\quad P_b^2\cdot P(AB)P_A\cdot P_B + P_bP_B\cdot\\
&\quad \Big(P(AB)P_A\cdot P_b + P(Ab)P_A\cdot P_B\Big) \quad P_B^2\cdot P(Ab)P_A\cdot P_b\Big]\\
&\quad + (P_a \quad P_A)\Big[\quad P_b^2\cdot\Big(P(AB)P_a\cdot P_B + P(aB)P_A\cdot P_B\Big)\\
&\quad + P_bP_B\cdot\Big(P(AB)P_a\cdot P_b + P(Ab)P_a\cdot P_B + P(aB)P_A\cdot P_b\\
&\quad + P(ab)P_A\cdot P_B\Big) \quad P_B^2\cdot\Big(P(Ab)P_a\cdot P_b + P(ab)P_A\cdot P_b\Big)\Big]\\
&\quad 2P_A\Big[\quad P_b^2\cdot P(aB)P_a\cdot P_B\\
&\quad + P_bP_B\cdot\Big(P(aB)P_a\cdot P_b + P(ab)P_a\cdot P_B\Big)\\
&\quad P_B^2\cdot P(ab)P_a\cdot P_b\Big] = 0.
\end{aligned}
$$

Similarly, we may show the other terms in equation (6).

Appendix B

By equations (6), (7), and large number theory, we can show the approximation (8). For instance, the approximation for element on the second row and the second column is

$$
\begin{aligned}
&\frac{1}{m}\sum_{i=m+1}^{n+m}(x_{Afi} \quad x_{Ami} \quad x_{A1i})\Sigma_i^{\ 1}(x_{Afi} \quad x_{Ami} \quad x_{A1i})^{\tau}\\
&= \frac{1}{(1 \quad 2\rho_0^2)\sigma^2}\frac{1}{m}\sum_{i=n+1}^{n+m}\Big[\Big((1 \quad \rho_0^2)x_{Afi} + \rho_0^2x_{Ami} \quad \rho_0x_{A1i}\Big)x_{Afi}\\
&\quad + \Big(\rho_0^2x_{Afi} + (1 \quad \rho_0^2)x_{Ami} \quad \rho_0x_{A1i}\Big)x_{Ami}\\
&\quad + \Big(\quad \rho_0x_{Afi} \quad \rho_0x_{Ami} + x_{A1i}\Big)x_{A1i}\Big]\\
&\approx \frac{1}{(1 \quad 2\rho_0^2)\sigma^2}\Big[2(1 \quad \rho_0^2)2P_aP_A \quad 4\rho_0P_aP_A + 2P_aP_A\Big]\\
&= \frac{2P_aP_A(3 \quad 2\rho_0 \quad 2\rho_0^2)}{(1 \quad 2\rho_0^2)\sigma^2}.
\end{aligned}
$$

For the element on the forth row and the forth column, we have

$$
\begin{aligned}
&\frac{1}{m}\sum_{i=m+1}^{n+m}(z_{Afi} \quad z_{Ami} \quad z_{A1i})\Sigma_i^{\ 1}\ z_{Afi} \quad z_{Ami} \quad z_{A1i}\ )^{\tau}\\
&= \frac{1}{(1 \quad 2\rho_0^2)\sigma^2}\frac{1}{m}\sum_{i=n+1}^{n+m}\Big[\Big((1 \quad \rho_0^2)z_{Afi} + \rho_0^2z_{Ami} \quad \rho_0z_{A1i}\Big)z_{Afi}\\
&\quad + \Big(\rho_0^2z_{Afi} + (1 \quad \rho_0^2)z_{Ami} \quad \rho_0z_{A1i}\Big)z_{Ami}\\
&\quad + \Big(\quad \rho_0z_{Afi} \quad \rho_0z_{Ami} + z_{A1i}\Big)z_{A1i}\Big] \approx \frac{P_a^2P_A^2(3 \quad 2\rho_0^2)}{(1 \quad 2\rho_0^2)\sigma^2}.
\end{aligned}
$$

Appendix C

To prove equation (10), we first have

$$\begin{aligned}
\mathrm{E}(x_{A1}x_{A2}) &= P_A^4\mathrm{E}[x_{A1}x_{A2}|A_f = AA, A_m = AA]\\
&\quad + 2P_A^2(2P_AP_a)\mathrm{E}[x_{A1}x_{A2}|A_f = AA, A_m = Aa]\\
&\quad + 2P_A^2P_a^2\mathrm{E}[x_{A1}x_{A2}|A_f = AA, A_m = aa]\\
&\quad + (2P_AP_a)^2\mathrm{E}[x_{A1}x_{A2}|A_f = Aa, A_m = Aa]\\
&\quad + 2P_a^2(2P_AP_a)\mathrm{E}[x_{A1}x_{A2}|A_f = Aa, A_m = aa]\\
&\quad + P_a^4\mathrm{E}[x_{A1}x_{A2}|A_f = aa, A_m = aa]\\
&= P_A^4(4P_a^2) + 4P_A^3P_a\Big[2P_a/2 + (P_a \quad P_A)/2\Big]^2\\
&\quad + 2P_A^2P_a^2[P_a \quad P_A]^2 + (2P_AP_a)^2\cdot\\
&\quad \Big[2P_a/4 + (P_a \quad P_A)/2 \quad 2P_A/4\Big]^2\\
&\quad + 4P_a^3P_A\Big[(P_a \quad P_A)/2 \quad 2P_A/2\Big]^2 + P_a^4(4P_A^2)\\
&= P_AP_a\\
\mathrm{E}(x_{A1}z_{A2}) &= P_A^4\mathrm{E}[x_{A1}z_{A2}|A_f = AA, A_m = AA]\\
&\quad + 2P_A^2(2P_AP_a)\mathrm{E}[x_{A1}z_{A2}|A_f = AA, A_m = Aa]\\
&\quad + 2P_A^2P_a^2\mathrm{E}[x_{A1}z_{A2}|A_f = AA, A_m = aa]\\
&\quad + (2P_AP_a)^2\mathrm{E}[x_{A1}z_{A2}|A_f = Aa, A_m = Aa]\\
&\quad + 2P_a^2(2P_AP_a)\mathrm{E}[x_{A1}z_{A2}|A_f = Aa, A_m = aa]\\
&\quad + P_a^4\mathrm{E}[x_{A1}z_{A2}|A_f = aa, A_m = aa]\\
&= P_A^4(\quad 2P_a^3) + 4P_A^3P_a[2P_a/2 + (P_a \quad P_A)/2]\\
&\quad [\quad P_a^2/2 + P_aP_A/2] + 2P_A^2P_a^2[P_a \quad P_A]P_aP_A\\
&\quad + (2P_AP_a)^2\Big[2P_a/4 + (P_a \quad P_A)/2 \quad 2P_A/4\Big]\\
&\quad \Big[\quad P_a^2/4 + P_aP_A/2 \quad P_A^2/4\Big]\\
&\quad + 4P_a^3P_A\Big[(P_a \quad P_A)/2 \quad 2P_A/2\Big]\Big[P_aP_A/2 \quad P_A^2/2\Big]\\
&\quad + P_a^4(2P_A^3) = 0
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(z_{A1}z_{A2}) &= P_A^4\mathrm{E}[z_{A1}z_{A2}|A_f = AA, A_m = AA]\\
&\quad + 2P_A^2(2P_AP_a)\mathrm{E}[z_{A1}z_{A2}|A_f = AA, A_m = Aa]\\
&\quad + 2P_A^2P_a^2\mathrm{E}[z_{A1}z_{A2}|A_f = AA, A_m = aa]\\
&\quad + (2P_AP_a)^2\mathrm{E}[z_{A1}z_{A2}|A_f = Aa, A_m = Aa]\\
&\quad + 2P_a^2(2P_AP_a)\mathrm{E}[z_{A1}z_{A2}|A_f = Aa, A_m = aa]\\
&\quad + P_a^4\mathrm{E}[z_{A1}z_{A2}|A_f = aa, A_m = aa]\\
&= P_A^4(\quad P_a^2)^2 + 4P_A^3P_a\Big[\quad P_a^2/2 + P_aP_A/2\Big]^2\\
&\quad + 2P_A^2P_a^2[P_aP_A]^2\\
&\quad + (2P_AP_a)^2\Big[\quad P_a^2/4 + P_aP_A/2 \quad P_A^2/4\Big]^2\\
&\quad + 4P_a^3P_A\Big[P_aP_A/2 \quad P_A^2/2\Big]^2 + P_a^4(\quad P_A^2)^2\\
&= P_a^2P_A^2/4.
\end{aligned}$$

Similarly, we may show that $\mathrm{E}(x_{B1}x_{B2}) = P_bP_B$, $\mathrm{E}\,(x_{B1}z_{B2}) = 0$, $\mathrm{E}(z_{B1}z_{B2}) = P_b^2P_B^2/4$. To show the other terms of (10), we first calculate the joint probabilities $P(A_1,B_2)$, in which the first offspring's genotype is A_1 at marker A and the second offspring's genotype is B_2 at marker B, $A_1 \in \{AA, Aa, aa\}$, $B_2 \in \{BB, Bb, bb\}$.We need to consider nine possible phases $\{AA, Aa, aa\} \times \{BB, Bb, bb\}$ for each parent. At the first glance, one needs to consider 9×9 possible matings to calculate $P(A_1,B_2)$. However, many matings do not lead to specific genotypes (A_1,B_2) of a sib pair. This eliminates many terms and reduces the amount of calculations. For instance, a mating of $(A_f = AA,\ B_f = BB) \times (A_m = AA,\ B_m = BB)$ only results offspring with genotype (AA,BB). Then, we have

$$\begin{aligned}
&P(A_1 = AA, B_2 = BB)\\
&= \sum_{A_f,B_f} P(A_f, B_f) \sum_{A_m,B_m} P(A_m, B_m)P[A_1 = AA, B_2\\
&\quad = BB|(A_f, B_f), (A_m, B_m)]\\
&= P(AB)^2\Big[P(AB)^2 + 2\cdot 2P(AB)P(Ab)/2 + 2\cdot 2P(AB)P(aB)/2\\
&\quad + 2\cdot[2P(AB)P(ab) + 2P(Ab)P(aB)]/4\Big]\\
&\quad + 2P(AB)P(Ab)\Big[2P(AB)P(Ab)/4\\
&\quad + 2\cdot 2P(AB)P(aB)/4 + 2\cdot[2P(AB)P(ab)\\
&\quad + 2P(Ab)P(aB)]\cdot 1/2\cdot 1/4\Big]\\
&\quad + 2P(AB)P(aB)\Big[2P(AB)P(aB)/4 + 2\cdot[2P(AB)P(ab)\\
&\quad + 2P(Ab)P(aB)]\cdot 1/2\cdot 1/4\Big]\\
&\quad + [2P(AB)P(ab) + 2P(Ab)P(aB)]^2\cdot 1/4\cdot 1/4\\
&= (P(AB) \quad P_AP_B)^2/4 + P_AP_BP(AB) = \Delta_{AB}^2/4 + P_AP_BP(AB).
\end{aligned} \tag{16}$$

Symmetrically, we may get the following three terms

$$\begin{aligned}
P(A_1 = AA, B_2 = bb) &= \Delta_{AB}^2/4 + P_AP_bP(Ab),\\
P(A_1 = aa, B_2 = BB) &= \Delta_{AB}^2/4 + P_aP_BP(aB),\\
P(A_1 = aa, B_2 = bb) &= \Delta_{AB}^2/4 + P_aP_bP(ab).
\end{aligned} \tag{17}$$

Note that $P(A_1{=}AA,\ B_2{=}Bb){=}P(A_1{=}AA){-}P(A_1{=}AA,\ B_2{=}BB$ or $bb)$. Hence,

$$P(A_1 = AA, B_2 = Bb) = \quad \Delta_{AB}^2/2 + P(AB)P_AP_b + P(Ab)P_AP_B. \tag{18}$$

Similarly, we may calculate the following three terms

$$\begin{aligned}
P(A_1 = aa, B_2 = Bb) &= \quad \Delta_{AB}^2/2 + P(aB)P_aP_b + P(ab)P_aP_B\\
P(A_1 = Aa, B_2 = BB) &= \quad \Delta_{AB}^2/2 + P(AB)P_aP_B + P(aB)P_AP_B\\
P(A_1 = Aa, B_2 = bb) &= \quad \Delta_{AB}^2/2 + P(Ab)P_aP_b + P(ab)P_AP_b.
\end{aligned} \tag{19}$$

Finally, we can calculate the following term using equation $\Sigma_{A1}\ \Sigma_{B2}\ P(A_1,B_2)=1$

$$\begin{aligned}
&P(A_1 = Aa, B_2 = Bb)\\
&= \Delta_{AB}^2 + P(AB)P_aP_b + P(Ab)P_aP_B + P(aB)P_AP_b + P(ab)P_AP_B.
\end{aligned} \tag{20}$$

Using equations (16), (17), (18), (19) and (20), we may calculate

$$\begin{aligned}
\mathrm{E}[x_{A1}x_{B2}] = 2P_a\Big[&2P_bP[A_1 = AA, B_2 = BB] \\
&+ (P_b \quad P_B)P[A_1 = AA, B_2 = Bb] \\
&\quad 2P_BP[A_1 = AA, B_2 = bb]\Big] \\
&+ (P_a \quad P_A)\Big[2P_bP[A_1 = Aa, B_2 = BB] \\
&+ (P_b \quad P_B)P[A_1 = Aa, B_2 = Bb] \\
&\quad 2P_BP[A_1 = Aa, B_2 = bb]\Big] \\
&\quad 2P_A\Big[2P_bP[A_1 = aa, B_2 = BB] \\
&+ (P_b \quad P_B)P[A_1 = aa, B_2 = Bb] \\
&\quad 2P_BP[A_1 = aa, B_2 = bb]\Big] = D_{AB}.
\end{aligned}$$

Similarly, we may get $\mathrm{E}[x_{A1}z_{B2}] = 0$, $\mathrm{E}[z_{A1}z_{B2}] = D_{AB}^2/4$. By symmetric property, we may calculate the remaining terms in (10).

Appendix D

Let $\sum_i{}^{1}$ be the matrix given by (9). To show the approximation of (11), we notice that d_{11} can be calculated by $d_{11} = \sigma^2(1 \quad 1 \quad 1 \quad 1)\ \sum_i{}^{1}(1 \quad 1 \quad 1 \quad 1)^\tau$. The element on the second row and the second column of approximation (11) can be calculated by

$$\begin{aligned}
&\frac{1}{k}\sum_{i=n+m+1}^{n+m+k}(x_{Afi} \quad x_{Ami} \quad x_{A1i} \quad x_{A2i})\Sigma_i{}^{1}(x_{Afi} \quad x_{Ami} \quad x_{A1i} \quad x_{A2i})^\tau \\
&= \frac{1}{\sigma^2}\frac{1}{k}\sum_{i=n+m+1}^{n+m+k}\Big[\Big((1+2C\rho_0)x_{Afi} + 2C\rho_0x_{Ami} \quad Cx_{A1i} \quad Cx_{A2i}\Big)x_{Afi} \\
&\quad + \Big(2C\rho_0x_{Afi} + (1+2C\rho_0)x_{Ami} \quad Cx_{A1i} \quad Cx_{A2i}\Big)x_{Ami} \\
&\quad + \Big(\quad Cx_{Afi} \quad Cx_{Ami} + \frac{C(1 \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}x_{A1i} \\
&\qquad \frac{C(\rho_{12} \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}x_{A2i}\Big)x_{A1i} \\
&\quad + \Big(\quad Cx_{Afi} \quad Cx_{Ami} \quad \frac{C(\rho_{12} \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}x_{A1i} \\
&\qquad + \frac{C(1 \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}x_{A2i}\Big)x_{A2i}\Big]
\end{aligned}$$

$$\begin{aligned}
&\approx \frac{1}{\sigma^2}\Big[\Big((1+2C\rho_0)2P_aP_A + 0 \quad CP_aP_A \quad CP_aP_A\Big) \\
&\quad + \Big(0 + (1+2C\rho_0)2P_aP_A \quad CP_aP_A \quad CP_aP_A\Big) \\
&\quad + \Big(\quad CP_aP_A \quad CP_aP_A + \frac{C(1 \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}2P_aP_A \\
&\qquad \frac{C(\rho_{12} \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}P_aP_A\Big) \\
&\quad + \Big(\quad CP_aP_A \quad CP_aP_A \quad \frac{C(\rho_{12} \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}P_aP_A \\
&\qquad + \frac{C(1 \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}2P_aP_A\Big)\Big] = 2P_aP_Ad_{22}/\sigma^2.
\end{aligned}$$

Similarly, the element on the forth row and the forth column of approximation (11) is

$$\begin{aligned}
&\frac{1}{k}\sum_{i=n+m+1}^{n+m+k}(z_{Afi} \quad z_{Ami} \quad z_{A1i} \quad z_{A2i})\Sigma_i{}^{1}(z_{Afi} \quad z_{Ami} \quad z_{A1i} \quad z_{A2i})^\tau \\
&= \frac{1}{\sigma^2}\frac{1}{k}\sum_{i=n+m+1}^{n+m+k}\Big[\Big((1+2C\rho_0)z_{Afi} + 2C\rho_0z_{Ami} \quad Cz_{A1i} \quad Cz_{A2i}\Big)z_{Afi} \\
&\quad + \Big(2C\rho_0z_{Afi} + (1+2C\rho_0)z_{Ami} \quad Cz_{A1i} \quad Cz_{A2i}\Big)z_{Ami} \\
&\quad + \Big(\quad Cz_{Afi} \quad Cz_{Ami} + \frac{C(1 \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}z_{A1i} \\
&\qquad \frac{C(\rho_{12} \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}z_{A2i}\Big)z_{A1i} \\
&\quad + \Big(\quad Cz_{Afi} \quad Cz_{Ami} \quad \frac{C(\rho_{12} \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}z_{A1i} \\
&\qquad + \frac{C(1 \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}z_{A2i}\Big)z_{A2i}\Big] \\
&\approx \frac{2}{\sigma^2}\Big[(1+2C\rho_0)P_a^2P_A^2 + \Big(\frac{C(1 \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}P_a^2P_A^2 \\
&\qquad \frac{C(\rho_{12} \quad 2\rho_0^2)}{\rho_0(1 \quad \rho_{12})}P_a^2P_A^2/4\Big)\Big] = P_a^2P_A^2d_{44}/\sigma^2.
\end{aligned}$$

The other terms of approximation (11) can be calculated in a similar manner.