

## ARTICLE

# Pedigree linkage disequilibrium mapping of quantitative trait loci

Ruzong Fan<sup>\*,1,2</sup>, Christie Spinka<sup>1</sup>, Lei Jin<sup>1</sup> and JeeSun Jung<sup>1</sup>

<sup>1</sup>Department of Statistics, Texas A&M University, 447 Blocker Building, 3143 TAMUS, College Station, TX 77843, USA; <sup>2</sup>Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Sigmund Freud Strasse 25, D-53105 Bonn, Germany

In this paper, we propose to use pedigrees of any size and any types of relatives in joint high-resolution linkage disequilibrium (LD) and linkage mapping of quantitative trait loci (QTL) by variance component models. Two or multiple markers can be simultaneously used in modeling association with the trait locus, instead of using one marker a time in the analysis. The proposed method can provide a unified result by using two or multiple markers in the modeling. This may avoid the complications of different results obtained from the separate analysis of marker by marker. The models simultaneously incorporate both linkage and LD information. The measures of LD are modeled by mean coefficients, and linkage information is modeled by variance–covariance matrix. Using analytical formulas to calculate the regression coefficients, the genetic effects are shown to be decomposed into additive and dominance components. The noncentrality parameter approximations of test statistics of LD are provided to make power calculations. Power and type I error rates are explored to investigate the merit of the proposed method by both the analytical formulas and simulations. Comparing with the association between-family and association within-family ('AbAw') approach of Fulker and Abecasis *et al*, it is evident that the method proposed in this article is more powerful. The method is applied to investigate the relation between polymorphisms in the angiotensin 1-converting enzyme (ACE) genes and circulating ACE levels, with a better result than that of the 'AbAw' approach. Moreover, two markers I/D and 4656(CT)3/2 can fully interpret association with the trait locus at a 0.01 significance level, which provides a unique result for the ACE data.

*European Journal of Human Genetics* (2005) 13, 216–231. doi:10.1038/sj.ejhg.5201301

Published online 13 October 2004

**Keywords:** QTL; LD mapping; IBD; variance component models

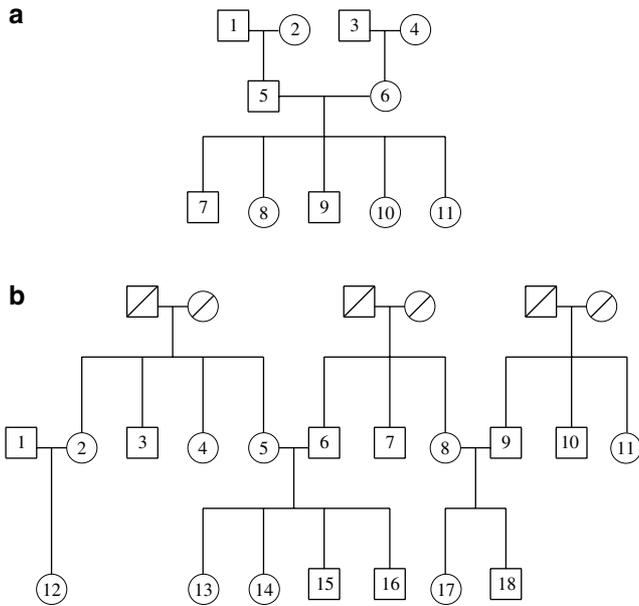
## Introduction

To map quantitative trait loci (QTL) of complex diseases, linkage disequilibrium (LD or association) regression analysis using population data can be performed.<sup>1</sup> However, population substructures may affect the results and produce false positives. To remedy this, variance

component models are proposed to perform joint LD and linkage mapping of QTL using both population and pedigree data.<sup>2–12</sup> However, most research limits on models that utilize small nuclear families in research. In principle, the method of George *et al*<sup>13</sup> can be used to analyze general pedigrees in joint linkage and LD mapping, but the examples of George *et al*<sup>13</sup> are limited to nuclear families. To our knowledge, Abecasis *et al*,<sup>3</sup> Göring and Terwillinger,<sup>10</sup> and Martin *et al*<sup>11</sup> are the only literature that present joint linkage and LD mapping methods using extended multi-generation pedigrees for QTL mapping.

\*Correspondence: Dr R Fan, Department of Statistics, Texas A&M University, 447 Blocker Building, 3143 TAMUS, College Station, TX 77843, USA. Tel: +1 979 845 3156/3141; Fax: +1 979 845 3144; E-mail: rfan@stat.tamu.edu

Received 2 April 2004; revised 17 August 2004; accepted 27 August 2004



**Figure 1** Pedigrees used in power calculations and comparison, which are taken from Figure 1 of Abecasis *et al.*<sup>3</sup> The number in the box or circle is individual ID.

In our previous research of joint linkage and LD mapping (or association study), either nuclear families or sibships were combined with population data to construct variance component models to map QTL for complex diseases.<sup>1,7,8</sup> This article generalizes our previous work to extended multigeneration pedigrees of any sizes and any types of relatives. Intuitively, large pedigrees contain more linkage and LD information. Therefore, it is important to develop models that may accommodate any types of data, including population data, sibships, nuclear families and multi-generation large pedigrees for a combined analysis.

The arrangement of the paper is as follows. First, variance component models are constructed based on multigeneration pedigree data following our previous research. The models incorporate both linkage and LD information in a unified analysis. The linkage information is modeled in variance-covariance matrix, while the association parameters are modeled in mean coefficients. Moreover, genetic effects are decomposed into orthogonal summation of additive and dominance effects. Type I error rates are calculated for five test cases to show the robustness of the proposed approach. To explore the effectiveness of the proposed methods, comparison with the association between-family and association within-family ('AbAw') approach of Abecasis and Faulker *et al* is investigated.<sup>2-4,9,12</sup> Two pedigrees of Figure 1 in Abecasis *et al.*<sup>3</sup> are taken as example. The comparison is based on both analytical formulae of noncentrality parameter approximations of test statistics and simulation results, using the same values of parameters of Abecasis *et al.*<sup>3</sup> The simulation

program PEDSIMUL kindly provided by Dr Abecasis is used to generate simulated datasets. As a practical example, the method is applied to investigate the relation between polymorphisms in the angiotensin-1-converting enzyme (ACE) genes and circulating ACE levels.<sup>14,15</sup>

## Methods

Consider a quantitative trait locus  $Q$  which has two alleles  $Q_1$  and  $Q_2$  with allele frequencies  $q_1$  and  $q_2$ , respectively. Assume that two markers  $A$  and  $B$  are typed in a chromosome region of the trait locus  $Q$ . 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. Suppose that the data are composed of  $I$  families. Let us list the log-likelihood of the  $I$  families by  $L_1, \dots, L_I$ . The overall log-likelihood is  $L = \sum_{i=1}^I L_i$ . In the  $i$ th family, let  $n_i$  be the total number of individuals who are listed as  $j = 1, 2, \dots, n_i$ , each individual  $j$  is preceded by all his/her ancestors. Let us denote the quantitative traits of  $i$ th family by a vector  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$ , genotypes at marker  $A$  by a vector  $(A_{i1}, A_{i2}, \dots, A_{in_i})^T$ , and genotypes at marker  $B$  by a vector  $(B_{i1}, B_{i2}, \dots, B_{in_i})^T$ . The log-likelihood is defined by

$$L_i = -\frac{n_i}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma_i| - \frac{1}{2} (y_i - X_i \mu)^T \Sigma_i^{-1} (y_i - X_i \mu),$$

under the assumption of multivariate normality. The notations of the log-likelihood are defined as follows. The mean component  $X_i \mu$  is from the following regression equation in a similar way as that of our previous work:<sup>1,7,8</sup>

$$y_{ij} = \beta + x_{Aij} \alpha_A + x_{Bij} \alpha_B + z_{Aij} \delta_A + z_{Bij} \delta_B + G_{ij} + e_{ij}, \quad (1)$$

where  $\beta$  is the overall mean,  $G_{ij}$  is the polygenic effect,  $e_{ij}$  is the error term. Assume that  $G_{ij}$  is normal  $N(0, \sigma_G^2)$ , and  $e_{ij}$  is normal  $N(0, \sigma_e^2)$ . Moreover,  $G_{ij}$  and  $e_{ij}$  are independent.  $x_{Aij}$ ,  $x_{Bij}$ ,  $z_{Aij}$  and  $z_{Bij}$  are dummy variables defined by

$$x_{Aij} = \begin{cases} 2P_a & \text{if } A_{ij} = AA \\ P_a - P_A & \text{if } A_{ij} = Aa \\ -2P_A & \text{if } A_{ij} = aa \end{cases}, \quad z_{Aij} = \begin{cases} -P_a^2 & \text{if } A_{ij} = AA \\ P_a P_A & \text{if } A_{ij} = Aa \\ -P_A^2 & \text{if } A_{ij} = aa \end{cases}$$

$$x_{Bij} = \begin{cases} 2P_b & \text{if } B_{ij} = BB \\ P_b - P_B & \text{if } B_{ij} = Bb \\ -2P_B & \text{if } B_{ij} = bb \end{cases}, \quad z_{Bij} = \begin{cases} -P_b^2 & \text{if } B_{ij} = BB \\ P_b P_B & \text{if } B_{ij} = Bb \\ -P_B^2 & \text{if } B_{ij} = bb \end{cases} \quad (2)$$

$\alpha_A$ ,  $\alpha_B$ ,  $\delta_A$  and  $\delta_B$  are regression coefficients of the dummy variables  $x_{Aij}$ ,  $x_{Bij}$ ,  $z_{Aij}$  and  $z_{Bij}$ .

The model matrix  $X_i$  is defined by

$$X_i = \begin{pmatrix} 1 & x_{Ai1} & x_{Bi1} & z_{Ai1} & z_{Bi1} \\ 1 & x_{Ai2} & x_{Bi2} & z_{Ai2} & z_{Bi2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{Ain_i} & x_{Bin_i} & z_{Ain_i} & z_{Bin_i} \end{pmatrix} = \begin{pmatrix} X_{i1}^T \\ X_{i2}^T \\ \vdots \\ X_{in_i}^T \end{pmatrix}$$

and  $\mu = (\beta, \alpha_A, \alpha_B, \delta_A, \delta_B)^T$  is a vector of regression coefficients. An intuitive rationale of regression (1) is

provided on pages 608–609 in Fan and Xiong.<sup>1</sup> In regression (1), it is assumed that there are no covariates, which may be included as that of Fan and Xiong.<sup>1,8</sup>

$\Sigma_i$  is a  $n_i \times n_i$  variance–covariance matrix defined as

$$\Sigma_i = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n_i} \\ \rho_{12} & 1 & \cdots & \rho_{2n_i} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{1n_i} & \rho_{2n_i} & \cdots & 1 \end{pmatrix} \sigma^2$$

where  $\sigma^2 = \sigma_g^2 + \sigma_G^2 + \sigma_e^2$ ,  $\sigma_g^2$  is variance explained by the putative QTL Q,  $\sigma_G^2$  is polygenic variance, and  $\sigma_e^2$  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 dominance components.  $\rho_{jk} = (\pi_{jkQ}\sigma_{ga}^2 + \Delta_{jkQ}\sigma_{gd}^2 + 2\Phi_{jk}\sigma_{Ga}^2 + \Delta_{7jk}\mu_{Gd}^2)/\sigma^2$  is correlation between the  $j$ th individual and the  $k$ th individual of the family, where  $\pi_{jkQ}$  is the proportion of alleles shared identical by descent (IBD) at QTL Q by the  $j$ th and the  $k$ th individuals,  $\Delta_{jkQ}$  is the probability that both alleles at QTL Q shared by the  $j$ th and the  $k$ th individuals are IBD,  $\Phi_{jk}$  is kinship coefficient of individuals  $j$  and  $k$ , and  $\Delta_{7jk}$  is the probability that both alleles shared by the  $j$ th and the  $k$ th individuals are IBD at any locus.  $\pi_{jkQ}$  and  $\Delta_{jkQ}$  are usually estimated by marker information.<sup>16–18</sup> The recombination fractions between the genotyped markers and the unobserved QTL are contained in the estimations of  $\pi_{jkQ}$  and  $\Delta_{jkQ}$ .<sup>7,19–22</sup> Hence, linkage information is modeled in variance–covariance matrix.

Let  $\mu_{ij}$  be the effect of genotype  $Q_iQ_j$ ,  $i, j = 1, 2$ ,  $\mu_{12} = \mu_{21}$ , that is  $\mu_{ij}$  is the mean displacement in phenotype trait for a given genotype  $Q_iQ_j$ . Denote the population effect mean by  $\mu = \mu_{11}q_1^2 + 2\mu_{12}q_1q_2 + \mu_{22}q_2^2$ , that is,  $\mu$  is the aggregate effect of the QTL on the trait mean in the population. Define  $\alpha_Q = q_1\mu_{11} + (q_2 - q_1)\mu_{12} - q_2\mu_{22}$ ,  $\delta_Q = 2\mu_{12} - \mu_{11} - \mu_{22}$ . If  $\mu_{11} = a$ ,  $\mu_{12} = d$ , and  $\mu_{22} = -a$  as in the traditional quantitative genetics,<sup>23</sup>  $\alpha_Q = a + (q_2 - q_1)d$  is the average allele substitution effect, and  $\delta_Q = 2d$  characterizes the dominance 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 dominance variance  $\sigma_{gd}^2 = (q_1q_2)^2\delta_Q^2$ .

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

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

Such as in Appendix B, Fan and Xiong,<sup>1</sup> we can show that the coefficients of regression equation (1) are given by

$$\begin{pmatrix} \alpha_A \\ \alpha_B \end{pmatrix} = V_A^{-1} \begin{pmatrix} 2D_{AQ} \\ 2D_{QB} \end{pmatrix} \alpha_Q \quad \text{and} \quad \begin{pmatrix} \delta_A \\ \delta_B \end{pmatrix} = V_D^{-1} \begin{pmatrix} D_{AQ}^2 \\ D_{QB}^2 \end{pmatrix} \delta_Q. \quad (3)$$

If only one marker A is used in the analysis, Eq. (3) can be replaced by  $\alpha_A = D_{AQ}\alpha_Q/(P_aP_A)$ ,  $\delta_A = D_{AQ}^2\delta_Q/(P_a^2P_A^2)$ . When the marker and QTL are the same and allele A is  $Q_1$  and allele a is  $Q_2$ ,  $D_{AQ} = q_1q_2 = P_aP_A$ , this implies that  $\alpha_A = \alpha_Q$  and  $\delta_A = \delta_Q$ , that is, all genetic substitution effect and dominance effect are encompassed in the allele association parameters. Hence, the estimates of the residual additive genetic variance  $\sigma_{ga}^2$  and  $\sigma_{gd}^2$  should be close to 0 in data analysis.

For a relative pair (1,2) of individuals 1 and 2. Table 1 gives the conditional probability  $P(G_1, G_2|C)$  given their allele IBD sharing status. Here  $G_i$  is genotype of relative  $i$ , and C is one event of  $(IBD = k)$ ,  $k = 0, 1, 2$ . For example,  $P(AA, AA|IBD = 0) = P_A^4$ ,  $P(AA, AA|IBD = 1) = P_A^3$  and  $P(AA, AA|IBD = 2) = P_A^2$ . Denote indicator functions

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

$$x_{Bi} = \begin{cases} 2P_b & \text{if } B_i = BB \\ P_b - P_B & \text{if } B_i = Bb \\ -2P_B & \text{if } B_i = bb \end{cases}, \quad z_{Bi} = \begin{cases} -P_b^2 & \text{if } B_i = BB \\ P_bP_B & \text{if } B_i = Bb \\ -P_B^2 & \text{if } B_i = bb \end{cases}, \quad i = 1, 2. \quad (4)$$

Utilizing the conditional probabilities of Table 1, the conditional expectation of product of the indicator functions can be calculated and the results are listed in Table 2. For instance,  $E(x_{A1}x_{A2}|IBD = 2) = 4P_a^2P_A^2 + (P_a - P_A)^2 2P_aP_A + 4P_b^2P_B^2 = 2P_aP_A$ . Based on Table 2, the expectation of product of indicator functions of (4) can be calculated for various types of relative pair. If individuals 1 and 2 are the same person,  $P(IBD = 2) = 1$  and then

$$E[(1, x_{A1}, x_{B1}, z_{A1}, z_{B1})^T (1, x_{A2}, x_{B2}, z_{A2}, z_{B2})] = \begin{pmatrix} 1 & O & O \\ O & V_A & O \\ O & O & V_D \end{pmatrix}, \quad (5)$$

where O are 0 matrices. If individuals 1 and 2 are common type of relatives, Appendix A provides expectations of product of indicator functions (4).

From above discussions, it is clear that the linkage is modeled in the variance–covariance matrix, while the association parameters are modeled in the mean coefficients. Hence, the model simultaneously takes linkage and association into account. The individual test of either linkage or association can be performed as follows: compare the full model in which all parameters are estimated and sub-models in which some parameters are fixed as constant.

In practice, one may first test linkage to detect a broad region of a trait locus. That is, the coefficients of  $\alpha_A$ ,  $\alpha_B$ ,  $\delta_A$

**Table 1** Conditional probability  $P(G_1, G_2|C)$  of a relative pair (1, 2) given their allele IBD sharing status

Conditional probability	Allele IBD sharing status $C$		
	$IBD = 0$	$IBD = 1$	$IBD = 2$
$P(AA, AA C)$	$P_A^4$	$P_A^3$	$P_A^2$
$P(AA, Aa C)$	$2P_aP_A^3$	$P_aP_A^2$	0
$P(AA, aa C)$	$P_a^2P_A^2$	0	0
$P(Aa, AA C)$	$2P_aP_A^3$	$P_aP_A^2$	0
$P(Aa, Aa C)$	$4P_a^2P_A^2$	$P_a^2P_A + P_A^2P_a$	$2P_aP_A$
$P(Aa, aa C)$	$2P_a^3P_A$	$P_a^2P_A$	0
$P(aa, AA C)$	$P_a^2P_A^2$	0	0
$P(aa, Aa C)$	$2P_a^3P_A$	$P_a^2P_A$	0
$P(aa, aa C)$	$P_a^4$	$P_a^2$	$P_a^2$
$P(AA, BB C)$	$P_A^2P_B^2$	$P_AP_BP(AB)$	$P(AB)^2$
$P(AA, Bb C)$	$2P_a^2P_bP_B$	$P_AP_bP(AB) + P_AP_BP(Ab)$	$2P(AB)P(Ab)$
$P(AA, bb C)$	$P_A^2P_b^2$	$P_AP_bP(AB)$	$P(AB)^2$
$P(Aa, BB C)$	$2P_aP_AP_B^2$	$P_AP_BP(aB) + P_aP_BP(AB)$	$2P(AB)P(aB)$
$P(Aa, Bb C)$	$4P_a^2P_AP_bP_B$	$P_AP_BP(ab) + P_aP_bP(aB) + P_aP_BP(Ab) + P_aP_bP(AB)$	$2P(AB)P(ab) + 2P(Ab)P(aB)$
$P(Aa, bb C)$	$2P_a^2P_AP_b^2$	$P_AP_bP(ab) + P_aP_bP(AB)$	$2P(AB)P(ab)$
$P(aa, BB C)$	$P_a^2P_B^2$	$P_aP_BP(aB)$	$P(aB)^2$
$P(aa, Bb C)$	$2P_a^2P_bP_B$	$P_aP_bP(aB) + P_aP_BP(ab)$	$2P(aB)P(ab)$
$P(aa, bb C)$	$P_a^2P_b^2$	$P_aP_bP(ab)$	$P(ab)^2$

Here  $G_i$  is genotype of relative  $i$ , and  $C$  is one event of ( $IBD = k$ ),  $k = 0, 1, 2$ .

**Table 2** Conditional expectation of a relative pair (1, 2) given their allele IBD sharing status

Conditional expectation	Allele IBD sharing status		
	$IBD = 0$	$IBD = 1$	$IBD = 2$
$E(X_{A1}X_{A2} C)$	0	$P_aP_A$	$2P_aP_A$
$E(X_{A1}Z_{A2} C)$	0	0	0
$E(X_{A1}X_{B2} C)$	0	$D_{AB}$	$2D_{AB}$
$E(X_{A1}Z_{B2} C)$	0	0	0
$E(Z_{A1}X_{A2} C)$	0	0	0
$E(Z_{A1}Z_{A2} C)$	0	0	$P_a^2P_A^2$
$E(Z_{A1}X_{B2} C)$	0	0	0
$E(Z_{A1}Z_{B2} C)$	0	0	$D_{AB}^2$

and  $\delta_B$  are fixed to be 0; and test the existence of linkage. In the presence of linkage, the association test can be performed for fine mapping of the QTL. If both additive and dominance variances  $\sigma_{ga}^2$  and  $\sigma_{gd}^2$  are significantly larger than 0, the genetic effects are not equal to 0 (ie,  $\alpha_Q \neq 0$  and  $\delta_Q \neq 0$ ). This implies that a null hypothesis  $H_{AB,ad}: \alpha_A = \alpha_B = \delta_A = \delta_B = 0$  is equivalent to  $D_{AQ} = D_{QB}$ , that is, no association between the QTL  $Q$  and the markers. Hence,  $H_{AB,ad}$  can be used to test association. Notice that there are four parameters under the null hypothesis  $H_{AB,ad}$ , but only two independent parameters  $D_{AQ}$  and  $D_{QB}$  are under test. Hence, at least two of the regression coefficients of  $\alpha_A, \alpha_B, \delta_A$ , and  $\delta_B$  must be 0 in data analysis. Suppose only additive variance  $\sigma_{ga}^2$  is significantly larger than 0, but

dominance variance  $\sigma_{gd}^2$  is not significantly larger than 0 (ie,  $\alpha_Q \neq 0$  and  $\delta_Q = 0$ ). Regression (1) can be simplified by excluding the dominance effects from the analysis, that is, setting  $\delta_A = \delta_B = 0$ . A null hypothesis  $H_{AB,a}: \alpha_A = \alpha_B = 0$  can be used to test association between the QTL  $Q$  and the markers.

In regression (1), two markers  $A$  and  $B$  are used in the analysis. If multiple markers are available, one may extend it to fit the data. For example, assume one more marker  $C$  is typed. Then regression (1) can be extended to

$$y_{ij} = \beta + x_{Aij}\alpha_A + x_{Bij}\alpha_B + x_{Cij}\alpha_C + Z_{Aij}\delta_A + Z_{Bij}\delta_B + Z_{Cij}\delta_C + G_{ij} + e_{ij}$$

where  $x_{Cij}$  and  $Z_{Cij}$  are defined accordingly in the same way of (2), and  $\alpha_C$  and  $\delta_C$  are their regression coefficients.

Evidence of association can be tested by likelihood ratio test (LRT) procedure. For instance, let  $L_{ad}$  be the log-likelihood under the alternative hypothesis of  $H_{AB,ad}$ , and  $L_0$  be the log-likelihood under the null hypothesis  $H_{AB,ad}$ . Then, the quantity  $2[L_{ad} - L_0]$  is asymptotically distributed as  $\chi^2$ . Notice again that there are only two measures of LD,  $D_{AQ}$  and  $D_{QB}$ , under the null hypothesis  $H_{AB,ad}$ . In data analysis, the number of coefficients  $\alpha_A, \alpha_B, \delta_A, \delta_B$ , which are significantly different from 0, should be less than or equal to 2. This number is the degrees of freedom of the likelihood ratio test  $2[L_{ad} - L_0]$ . For large sample data, the likelihood ratio test is accurate based on the statistical theory. In the following, we will develop a  $F$ -test procedure based on linear model theory.<sup>24</sup>

**F-tests and noncentrality parameter approximations**

For the convenience of explanation, let us consider  $I$  families given in graph A or graph B of Figure 1 (Figure 1 in Abecasis *et al*<sup>3</sup>). Let  $N$  be the total number of individuals, that is,  $N = nl$ , where  $n = 11$  for graph A and  $n = 18$  for graph B of Figure 1, respectively. For each individual in Figure 1, an ID is assigned. For the grand parents of graph B, the genotypes are unavailable and so no IDs are assigned. Denote the total trait values  $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_I^T)^T$ , the total variance-covariance matrix by  $\Sigma = \text{diag}(\Sigma_1, \dots, \Sigma_I)$ , and the model matrix by  $X = X_1^T, \dots, X_I^T)^T$ . Assume that  $\Sigma_1 = \dots = \Sigma_I$ . Let  $\hat{\alpha}_A, \hat{\alpha}_B, \hat{\delta}_A, \hat{\delta}_B, \hat{\Sigma}_i, \hat{\Sigma}$  be the maximum likelihood estimators of  $\alpha_A, \alpha_B, \delta_A, \delta_B, \Sigma_i, \Sigma$ . The estimate of  $\mu$  is  $\hat{\mu} = [X^T \hat{\Sigma}^{-1} X]^{-1} X^T \hat{\Sigma}^{-1} y = [\sum_{i=1}^I X_i^T \hat{\Sigma}_i^{-1} X_i]^{-1} \sum_{i=1}^I X_i^T \hat{\Sigma}_i^{-1} y_i$ . Let  $H$  be a  $q \times 5$  test matrix of rank  $q$ . By Graybill<sup>24</sup> (Chapter 6), the test statistic of a hypothesis  $H\mu = 0$  is noncentral  $F(q, N-5)$  defined by

$$F = \frac{(H\hat{\mu})^T [H(X^T \hat{\Sigma}^{-1} X)^{-1} H^T]^{-1} (H\hat{\mu})}{y^T [\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} X (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1}] y} \frac{N-5}{q}$$

The noncentrality parameter of the test statistic  $F$  can be calculated by  $\lambda = (H\mu)^T [H(X^T \Sigma^{-1} X)^{-1} H^T]^{-1} H\mu = (H\mu)^T [H [\sum_{i=1}^I X_i^T \Sigma_i^{-1} X_i]^{-1} H^T]^{-1} (H\mu)$ . Denote  $\Sigma_i^{-1} = \frac{1}{\sigma^2} \gamma_{B1 \times n}$ . In Appendix B, an approximation of the noncentrality parameter is provided by

$$\lambda \approx \frac{I}{\sigma^2} (H\mu)^T \left[ H \begin{pmatrix} 1/\sum_k \sum_l \gamma_{kl} & O & O \\ O & V_A^{-1}/b_1 & O \\ O & O & V_D^{-1} b_2 \end{pmatrix} H^T \right]^{-1} \times (H\mu),$$

where  $b_1$  and  $b_2$  are constants given by equations (B.1) and (B.2) in Appendix B for graph A and graph B of Figure 1, respectively. Notations  $O$  are 0 matrices.

To test the null hypothesis  $H_{AB,a}: \alpha_A = \alpha_B = 0$ , the test matrix

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

Let us denote the corresponding  $F$ -test statistic by  $F_{AB,a}$ . The noncentrality parameter is

$$\begin{aligned} \lambda_{AB,a} &\approx \frac{b_1 I}{\sigma^2} (\alpha_A, \alpha_B) V_A \begin{pmatrix} \alpha_A \\ \alpha_B \end{pmatrix} \\ &= \frac{b_1 I}{\sigma^2} 2\alpha_Q^2 [P_b P_B D_{AQ}^2 - 2D_{AQ} D_{AB} D_{QB}] \\ &\quad + P_a P_A D_{QB}^2 / (P_a P_A P_b P_B - D_{AB}^2) \\ &= \frac{b_1 I}{\sigma^2} \sigma_{ga}^2 [P_b P_B D_{AQ}^2 - 2D_{AQ} D_{AB} D_{QB}] \\ &\quad + P_a P_A D_{QB}^2 / [q_1 q_2 (P_a P_A P_b P_B - D_{AB}^2)] \end{aligned}$$

To test the null hypothesis  $H_{AB,d}: \delta_A = \delta_B = 0$ , the test matrix

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

Denote the corresponding  $F$ -test statistic by  $F_{AB,d}$ . The noncentrality parameter is

$$\begin{aligned} \lambda_{AB,d} &\approx \frac{b_2 I}{\sigma^2} (\delta_A, \delta_B) V_D \begin{pmatrix} \delta_A \\ \delta_B \end{pmatrix} \\ &= \frac{b_2 I}{\sigma^2} \delta_Q^2 [P_b^2 P_B^2 D_{AQ}^4 - 2D_{AQ}^2 D_{AB}^2 D_{QB}^2 \\ &\quad + P_a^2 P_A^2 D_{QB}^4] / (P_a^2 P_A^2 P_b^2 P_B^2 - D_{AB}^4) \\ &= \frac{b_2 I}{\sigma^2} \sigma_{gd}^2 [P_b^2 P_B^2 D_{AQ}^4 - 2D_{AQ}^2 D_{AB}^2 D_{QB}^2 \\ &\quad + P_a^2 P_A^2 D_{QB}^4] / [q_1^2 q_2^2 (P_a^2 P_A^2 P_b^2 P_B^2 - D_{AB}^4)]. \end{aligned}$$

To test the null hypothesis  $H_{AB,ad}: \alpha_A = \alpha_B = \delta_A = \delta_B = 0$ , the test matrix

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

Denote the corresponding  $F$ -test statistic by  $F_{AB,ad}$ . The noncentrality parameter is  $\lambda_{AB,ad} = \lambda_{AB,a} + \lambda_{AB,d}$ . Assume that only one marker  $A$  is used in the analysis. The test statistic of a hypothesis  $H\mu = 0$  is noncentral  $F(q, N-3)$  defined by

$$F = \frac{(H\hat{\mu})^T [H(X^T \hat{\Sigma}^{-1} X)^{-1} H^T]^{-1} (H\hat{\mu})}{y^T [\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} X (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1}] y} \frac{N-3}{q},$$

where  $\mu = (\beta, \alpha_A, \delta_A)^T$ . The noncentrality parameter is  $\lambda_{A,ad} \approx \frac{b_1 I}{\sigma^2} \sigma_{ga}^2 D_{AQ}^2 / (P_a P_a q_1 q_2) + \frac{b_2 I}{\sigma^2} \sigma_{gd}^2 D_{AQ}^4 / (P_a^2 P_a^2 q_1^2 q_2^2)$  for the null hypothesis  $H_{A,ad}: \alpha_A = \delta_A = 0$ . Correspondingly, we denote the  $F$ -test statistic by  $F_{A,ad}$ . Similarly,  $\lambda_{A,a} \approx (b_1 I / \sigma^2) \sigma_{ga}^2 D_{AQ}^2 / (P_a P_a q_1 q_2)$  is the noncentrality parameter of the test statistic  $F_{A,a}$  for the null hypothesis  $H_{A,a}: \alpha_A = 0$ . The noncentrality parameter of the test statistic  $F_{A,d}$  for the null hypothesis  $H_{A,d}: \delta_A = 0$  is  $\lambda_{A,d} \approx (b_2 I / \sigma^2) \sigma_{gd}^2 D_{AQ}^4 / (P_a^2 P_a^2 q_1^2 q_2^2)$ .

**Type I error rates and power comparison**

Before making comparison, let us briefly describe the ‘AbAw’ method and comments on the difference of the ‘AbAw’ approach with ours. Such as our proposed method, the random effect parameters ( $\sigma_g^2, \sigma_G^2$  and  $\sigma_e^2$ ) are modeled in the variance-covariance matrices in the ‘AbAw’ approach. However, the ‘AbAw’ approach decomposes the genetic association into effects of between-family members and within-family members.<sup>2-4,9,12</sup> The effects of between family members and within-family members are modeled in two coefficients  $\beta_b$  and  $\beta_w$ . This is primarily the difference between the ‘AbAw’ approach and our method.

Suppose only one marker *A* is used in the analysis. Ignoring the dominance effect, our model only has one additive coefficient  $\alpha_A$  but the ‘AbAw’ still has two regression coefficients  $\beta_b$  and  $\beta_w$ . Under the assumption of multivariate normality, the ‘AbAw’  $\chi^2_{qtl}$  is defined as the likelihood ratio test of hypothesis  $\beta_w = 0$  against no constraints on the coefficient. In addition, the ‘AbAw’ approach uses only one marker and it is not clear how to extend the ‘AbAw’ approach to use more than one markers in analysis. We use two biallelic markers, and it is easy to extend our approach to use multiple biallelic markers in analysis. For sibship data, Fan and Jung<sup>7</sup> compared our method with the ‘AbAw’ approach and found that our method is advantageous. In the following, we are going to calculate the type I error rates to show the robustness of the proposed approach. Then, we will make power comparison for general extended pedigrees.

### Type I error rates

To investigate the robustness of the proposed approach, we calculate type I error rates at a 0.05 significance level for five test cases of Table 3, which is the same as those of Table 1 of Abecasis *et al.*<sup>3</sup> Trait values are constructed from a normal distribution with mean 100, and total variance  $\sigma^2 = 1$  at the trait locus *Q* except for the *Admixture* case. Here  $\sigma^2 = \sigma_{ga}^2 + \sigma_{Ga}^2 + \sigma_e^2$  is the summation of the additive variance of major gene effect  $\sigma_{ga}^2$ , the variance of additive polygenic effect  $\sigma_{Ga}^2$ , and the error variance  $\sigma_e^2$ . A biallelic marker *A* is simulated at a recombination fraction 0 between QTL *Q* and locus *A*. In each case, linkage equilibrium is assumed between QTL *Q* and marker *A*, that is,  $D_{AQ} = 0$ . In the test case of *Admixture*, population

admixture is generated by mixing families equally draw from one of the two subpopulations *C* and *D*. In both subpopulations *C* and *D*, no major gene effect or familial effect is assumed, that is,  $\sigma_G^2 = \sigma_G^2 = 0$ . However, the trait mean of subpopulation *C* is fixed as 1 and the variance is fixed as 1, and the marker allele frequency  $P_A$  is taken as 0.7 in subpopulation *C*. The trait mean of subpopulation *D* is fixed as 0 and the variance is fixed as 1, and the marker allele frequency  $P_A$  is taken as 0.3 in subpopulation *D*. Therefore, the total variance in the mixing population is  $\sigma^2 = 1.25$ . The admixture contributed to  $(1-0)^2 / [4 \times 1.25] = 0.20$  of the total variance. The related parameters of the other four test cases are given in the legend of Table 3.

To calculate the type I error rates, 1000 data sets are simulated for each test case. Each data set contains 50 pedigrees of either graph *A* or graph *B* of Figure 1, respectively. Using the data sets, we fit the following model

$$y_{ij} = \beta + x_{Aij}\alpha_A + G_{ij} + e_{ij}$$

where  $G_{ij}$  is normal  $N(0, \sigma_{Ga}^2)$ ,  $y_{ij}$  is normal  $N(\beta + x_{Aij}\alpha_A, \sigma^2)$ , and  $\sigma^2 = \sigma_{ga}^2 + \sigma_{Ga}^2 + \sigma_e^2$ . The null hypothesis is  $H_{A,a}: \alpha_A = 0$ . Since the QTL *Q* is in linkage equilibrium with marker *A*, an empirical test statistic, which is larger than the cutting point at a 0.05 significance level, is treated as a false positive. Based on either likelihood ratio test or *F*-test, type I error rates are calculated as the proportions of the 1000 simulation data sets, which give significant result at the 0.05 significant level based on  $F_{A,a}$  and likelihood ratio test statistic, respectively.

The results of Table 3 shows that the type I error rates are around the nominal level 0.05. Hence, the model is reasonably robust. For the large three-generation pedigree of Graph *B* of Figure 1, the type I error rates are very close to the nominal 0.05 significance level. Hence, the model

**Table 3** Type I error rates (%) at a 0.05 significance level

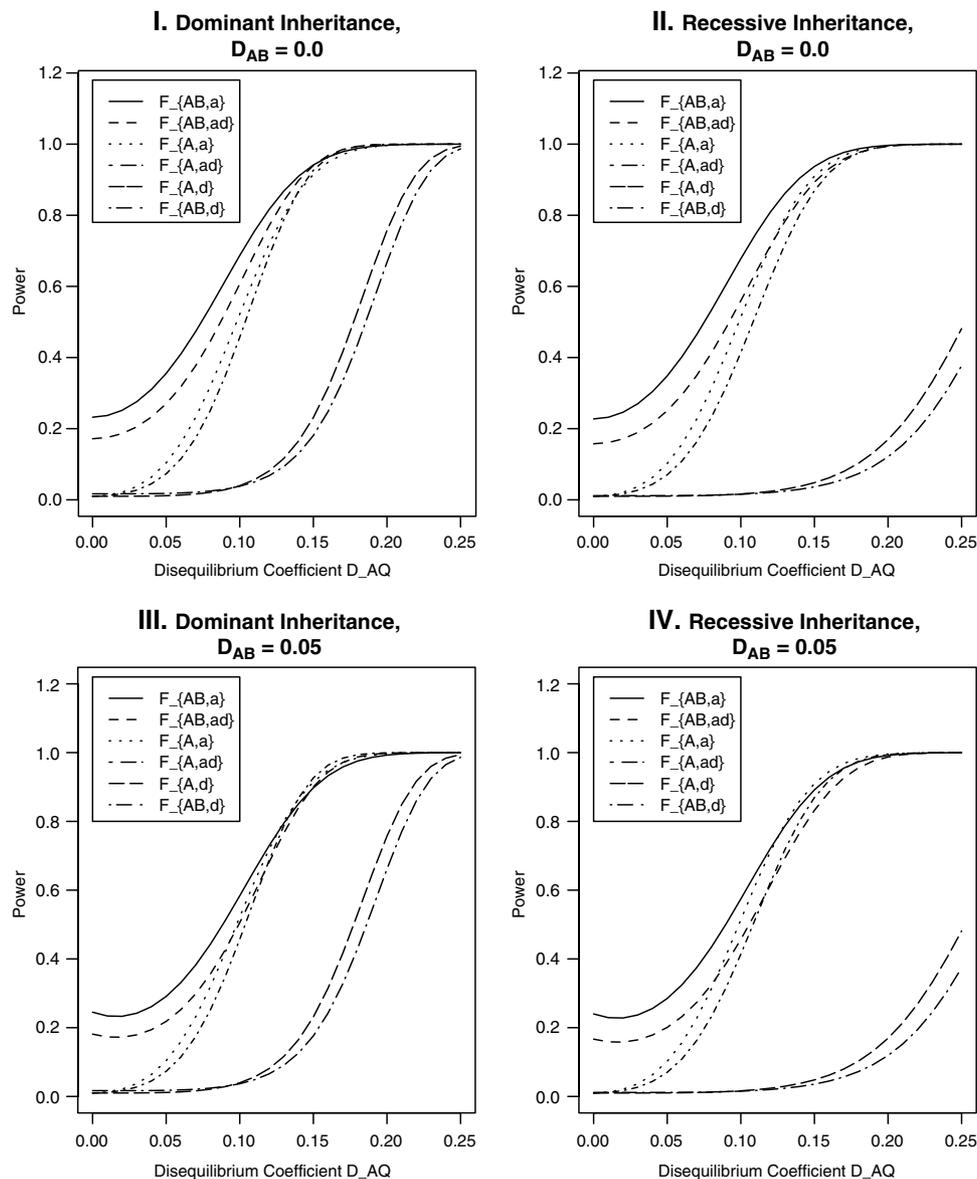
Pedigree type	Test case	Error rates	
		LRT	$\hat{F}_{1,a}$
Small 3-generation pedigree	Null	4.7	4.9
	Familiality	4.6	4.8
	Admixture	5.9	5.9
	Linkage	5.3	5.5
	Composite	5.6	5.7
Large 3-generation pedigree	Null	5.0	5.1
	Familiality	5.1	5.1
	Admixture	5.0	5.2
	Linkage	4.8	4.9
	Composite	4.7	4.9

The total variance is fixed as  $\sigma^2 = 1.0$  (see text for explanation of *Admixture* case). *Null*: no major gene effect or familial effect  $\sigma_g^2 = \sigma_G^2 = 0$ ; *Familiality*: large familial effect  $\sigma_G^2 = \sigma_{Ga}^2 = 0.5$ , but no major gene effect  $\sigma_g^2 = 0$ ; *Admixture*: no major gene effect or familial effect  $\sigma_g^2 = \sigma_G^2 = 0$ , but with population admixture; *Linkage*: large linkage effect  $\sigma_g^2 = \sigma_{ga}^2 = 0.5$ ,  $\theta_{M1Q} = 0$ , but no familial effect  $\sigma_G^2 = 0$ ; *Composite*: large linkage effect  $\sigma_g^2 = \sigma_{ga}^2 = 0.2$ ,  $\theta_{M1Q} = 0$ , and large familial effect  $\sigma_G^2 = \sigma_{Ga}^2 = 0.3$ . There is no linkage disequilibrium between QTL and marker *A* ( $D_{AQ} = 0$ ).

**Table 4** Power (%) of 50 pedigrees (Figure 1) for varying levels of LD between trait locus and marker *A* at 0.01 significance level

<i>D'</i>	0.00	0.25	0.50	0.75	1.00
<i>Graph A: small 3-generation pedigree</i>					
AbAw's $\chi^2_{qtl}$	0.6	7.8	33.5	76.5	98.4
$F_{A,a}$	1.0	18.2	77.7	99.3	100
$\hat{F}_{A,a}$	1.0	15.2	75.7	99.2	100
LRT	0.9	14.3	74.6	99.2	100
<i>Graph B: large 3-generation pedigree</i>					
AbAw's $\chi^2_{qtl}$	0.6	16.4	73.3	99.1	100
$F_{A,a}$	1.0	39.3	97.9	100	100
$\hat{F}_{A,a}$	1.2	29.8	95.0	100	100
LRT	1.2	29.7	94.9	100	100

The parameters are given by  $\sigma_{ga}^2 = 0.1$ ,  $\sigma_{gd}^2 = 0$ ,  $\sigma_{Ga}^2 = 0.5$ ,  $\sigma_{Gd}^2 = 0$ ,  $\sigma_e^2 = 0.4$ ,  $q_1 = P_A = 0.5$ , which are the same as those of Table 2 of Abecasis *et al.*<sup>3</sup> The power of AbAw's  $\chi^2_{qtl}$  is taken from Table 2 of Abecasis *et al.* In addition,  $D' = D_{AQ}/D_{max}$  and  $D_{max} = \min(P_A, q_1) - P_A q_1$ .



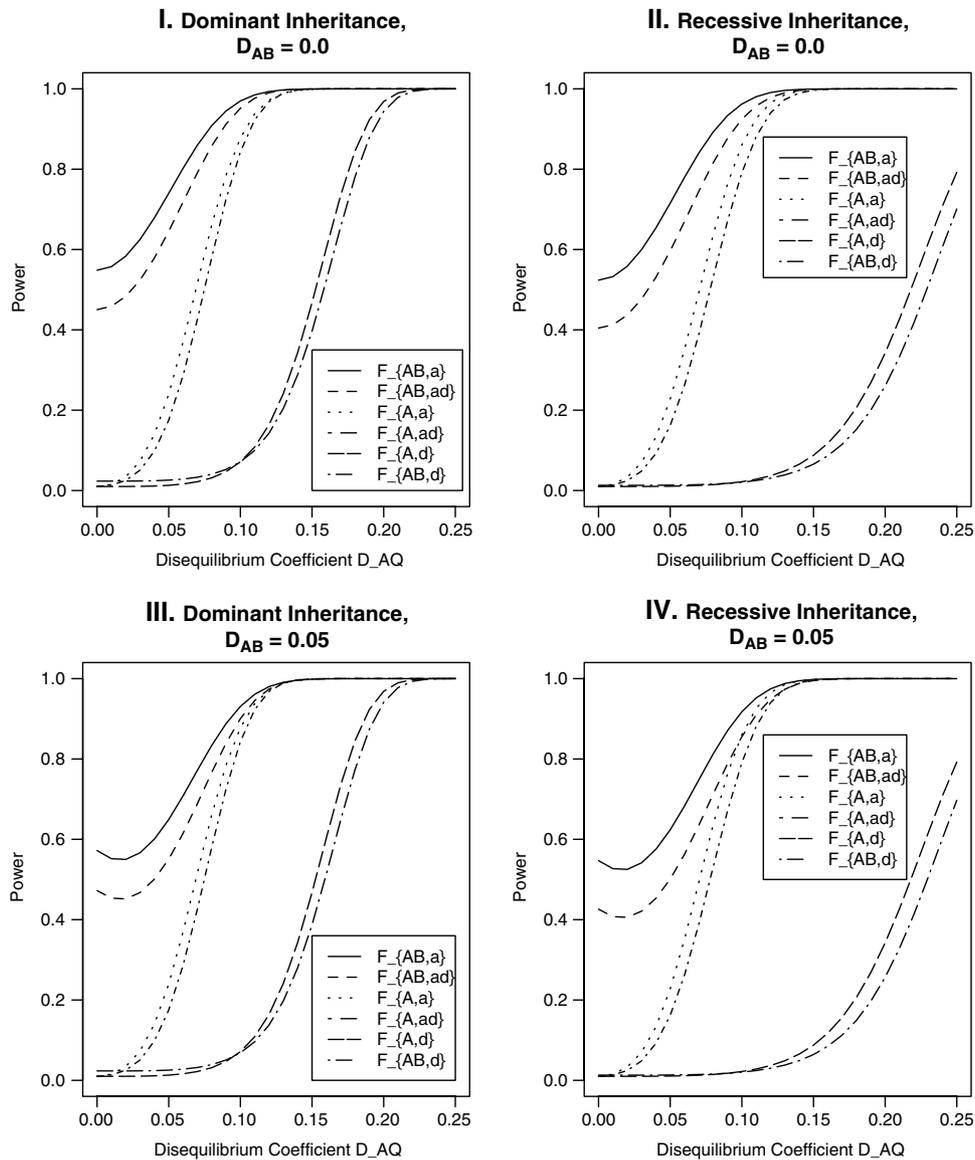
**Figure 2** Power of 30 small three-generation pedigrees (Figure 1, graph a) of test statistics  $F_{AB,a}$ ,  $F_{AB,ad}$ ,  $F_{A,a}$ ,  $F_{A,ad}$ ,  $F_{A,d}$  and  $F_{AB,d}$  against LD coefficient  $D_{AQ}$  at 0.01 significance level, when  $q_1 = P_A = P_B = 0.50$ ,  $D_{QB} = 0.08$ ,  $\sigma_{G_a}^2 = 0.10$ ,  $\sigma_{G_d}^2 = 0.05$  and  $h^2 = 0.15$ , for a dominant mode of inheritance  $a = d = 1.0$ , and a recessive mode of inheritance  $a = 1.0$ ,  $d = -0.5$ , respectively.

works well under large sample case, as expected (notice that the total number of individuals  $N = 550$  for the 50 small three-generation pedigrees of Graph A, and  $N = 900$  for the 50 large three-generation pedigrees of Graph B). Moreover, the results of  $F$ -test  $\hat{F}_{A,a}$  are very close to those of likelihood ratio test statistic. This is, theoretically, guaranteed under large sample by the property of the two tests (Chapter 6, Graybill<sup>24</sup>).

#### Comparison with the 'AbAw' approach

Table 4 shows the power of 50 duplicates of pedigree A (and pedigree B) of Figure 1, respectively, for varying levels of LD

between trait locus and marker A at 0.01 significance level. The parameters are exactly the same as those of Abecasis *et al.*<sup>3</sup> Besides, the power of AbAw's  $\chi_{qti}^2$  is taken from Table 2 of Abecasis *et al.*<sup>3</sup> The power of  $F_{A,a}$  is calculated based on the noncentrality parameter approximation  $\lambda_{A,a}$ ; the power of  $\hat{F}_{A,a}$  and  $LRT$  are calculated as the proportions of 1000 simulation data sets, which give significant result at the 0.01 significance level based on  $F_{A,a}$  and likelihood ratio test statistic, respectively. It is obvious from Table 4 that  $F_{A,a}$  and likelihood ratio test based on our model have higher power than that of AbAw's  $\chi_{qti}^2$ . Since other test statistics in Table 2 of Abecasis *et al.*<sup>3</sup> is less powerful than



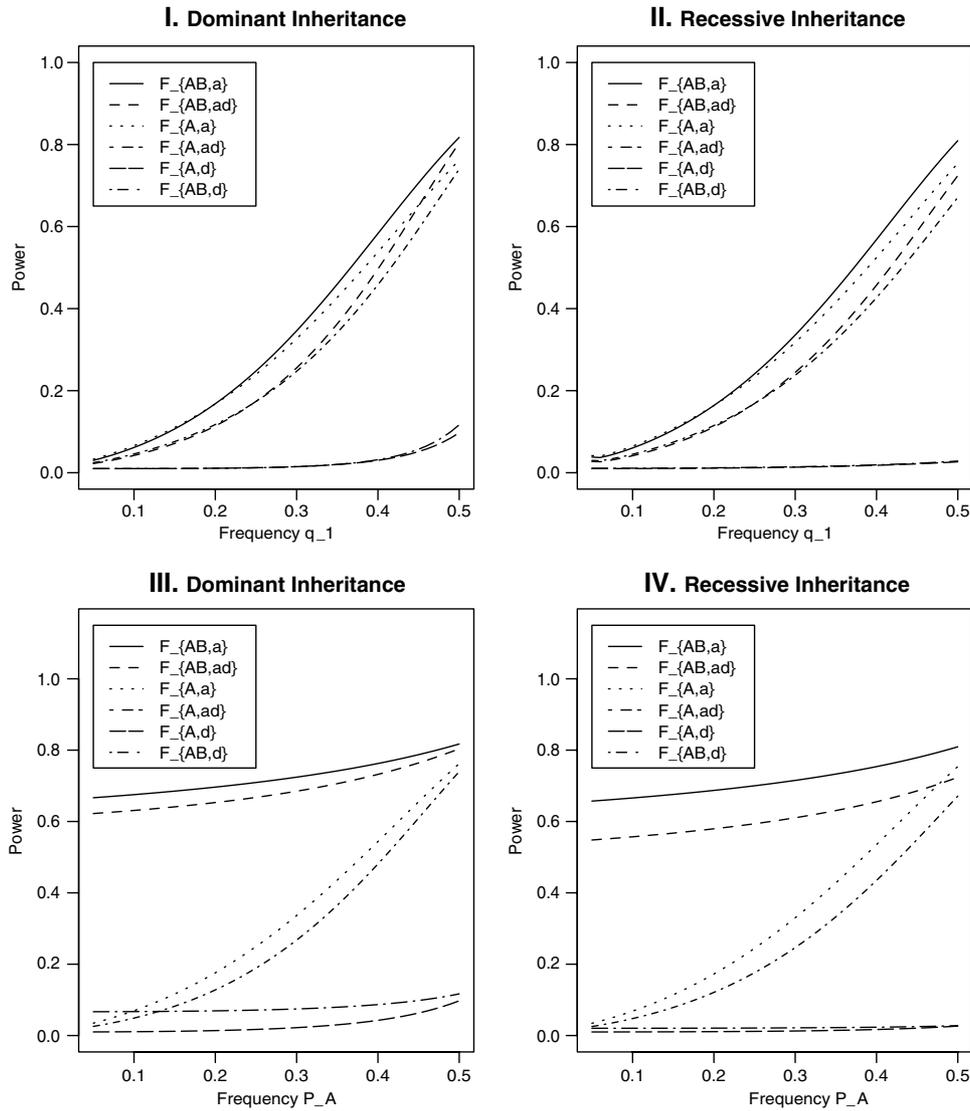
**Figure 3** Power of 30 large three-generation pedigrees (Figure 1, graph b) of test statistics  $F_{AB,a}$ ,  $F_{AB,ad}$ ,  $F_{A,a}$ ,  $F_{A,ad}$ ,  $F_{A,d}$ , and  $F_{AB,d}$  against LD coefficient  $D_{AQ}$  at 0.01 significance level, when  $q_1 = P_A = P_B = 0.50$ ,  $D_{QB} = 0.08$ ,  $\sigma_{Ca}^2 = 0.10$ ,  $\sigma_{Cd}^2 = 0.05$  and  $h^2 = 0.15$ , for a dominant mode of inheritance  $a = d = 1.0$ , and a recessive mode of inheritance  $a = 1.0$ ,  $d = -0.5$ , respectively.

the AbAw's  $\chi_{qtl}^2$ , the method proposed in the current article is advantageous over the AbAw approach. As the type I error rates in Table 3, the results of  $F$ -test  $\hat{F}_{A,a}$  are very close to those of likelihood ratio test statistics.

### Power comparison

Let us denote the heritability by  $h^2$ , which is  $h^2 = \sigma_{ga}^2 / \sigma^2$ .<sup>23</sup> To make power comparison, we consider 30 duplicates of small and large 3-generation pedigrees of graphs A and B of Figure 1, respectively. Two modes of inheritance are taken as example: a dominant mode of inheritance  $a = d = 1.0$ , and a recessive mode of inheritance  $a = 1.0$ ,  $d = -0.5$ , respectively. Figures 2 and 3 show the power

curves of test statistics  $F_{AB,a}$ ,  $F_{AB,ad}$ ,  $F_{A,a}$ ,  $F_{A,ad}$ ,  $F_{A,d}$ , and  $F_{AB,d}$  against LD coefficient  $D_{AQ}$  at 0.01 significance level, for the 30 duplicates of graph A and graph B of Figure 1, respectively. The other parameters are given in the legend of the Figures. It can be seen that  $F_{AB,ad}$  generally has higher power than that of  $F_{A,ad}$  and  $F_{AB,a}$  generally has higher power than that of  $F_{A,a}$ . Hence, analysis using two markers are advantageous. Owing to the increase of degrees of freedom of test statistics,  $F_{AB,ad}$  generally has lower power than that of  $F_{AB,a}$  and  $F_{A,ad}$  generally has lower power than that of  $F_{A,a}$ . Besides, the power of  $F_{AB,d}$  and  $F_{A,d}$  is low unless the LD between marker A and QTL Q is very strong. Since pedigree of Graph B of Figure 1 is larger than



**Figure 4** Power of 30 small three-generation pedigrees (Figure 1, graph a) of test statistics  $F_{AB,a}$ ,  $F_{AB,ad}$ ,  $F_{A,a}$ ,  $F_{A,ad}$ ,  $F_{A,d}$  and  $F_{AB,d}$  against trait frequency  $q_1$  or marker allele frequency  $P_A$  at 0.01 significance level, when  $P_A=0.5$  (graphs I and II),  $q_1=0.5$  (graphs III and IV),  $P_B=0.50$ ,  $\sigma_{Ca}^2=0.10$ ,  $\sigma_{Cd}^2=0.05$  and  $h^2=0.15$ , for a dominant mode of inheritance  $a=d=1.0$ , and a recessive mode of inheritance  $a=1.0, d=-0.5$ , respectively. The linkage disequilibrium coefficients are  $D_{AB}=(\min(P_A, P_B)-P_AP_B)/2$ ,  $D_{AQ}=(\min(P_A, q_1)-P_Aq_1)/2$  and  $D_{QB}=(\min(P_B, q_1)-P_Bq_1)/2$ .

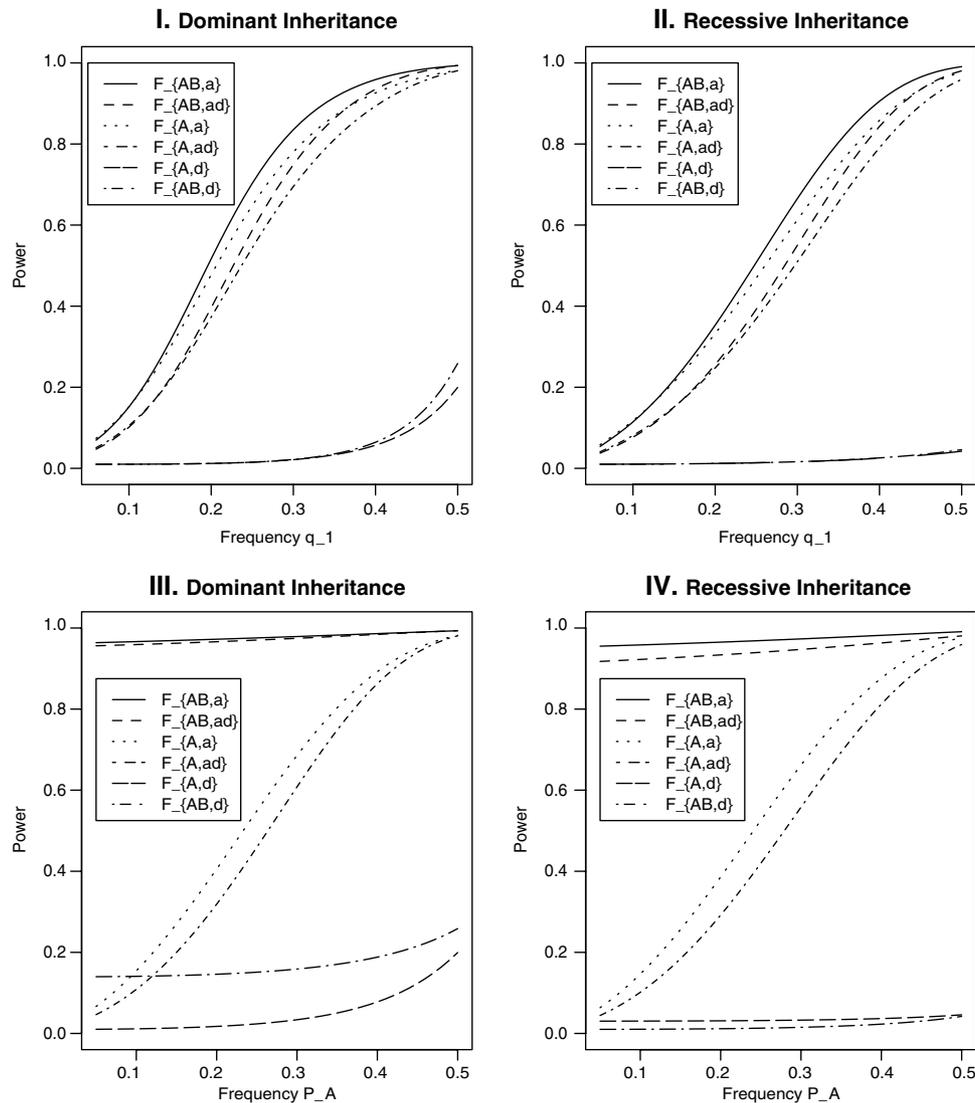
that of Graph A, the power of test statistics for Figure 3 is higher than that of corresponding statistics for Figure 2. Thus, large pedigrees contain more LD information and the models proposed rightly use the information.

Figures 4 and 5 show power curves of test statistics  $F_{AB,a}$ ,  $F_{AB,ad}$ ,  $F_{A,a}$ ,  $F_{A,ad}$ ,  $F_{A,d}$  and  $F_{AB,d}$  against trait frequency  $q_1$  or marker allele frequency  $P_A$  at 0.01 significance level, for the 30 duplicates of graph A and graph B of Figure 1, respectively. In addition to the merits shown in Figures 2 and 3, the power of test statistics  $F_{AB,ad}$  and  $F_{AB,a}$  heavily depends on the frequency  $q_1$  of QTL (graphs I and II of each figure), but not very much on the marker allele frequency  $P_A$

(graphs III and IV of each figure). Figures 6 and 7 show power curves of test statistics  $F_{AB,a}$ ,  $F_{AB,ad}$ ,  $F_{A,a}$ ,  $F_{A,ad}$ ,  $F_{A,d}$  and  $F_{AB,d}$  against heritability  $h^2$  at 0.01 significance level, for the 30 duplicates of graph A and graph B of Figure 1, respectively. It can be seen that the power of  $F_{AB,ad}$ ,  $F_{AB,a}$ ,  $F_{A,ad}$  and  $F_{A,a}$  can be high as the heritability  $h^2$  is larger than 0.1. The power of  $F_{AB,d}$  and  $F_{A,d}$  is generally low.

**An example**

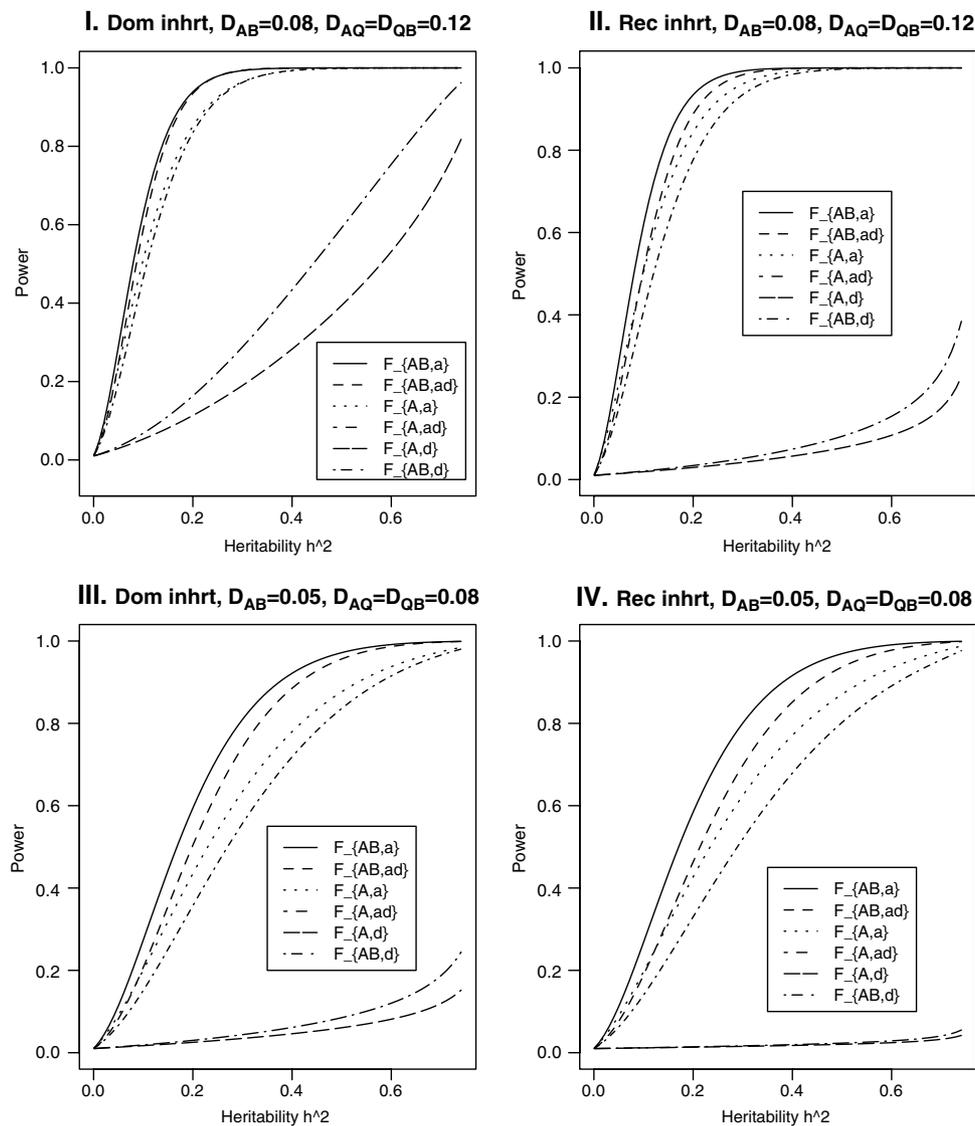
The proposed method is applied to analyze ACE data.<sup>14,15</sup> The data consist of 83 extended families with between four



**Figure 5** Power of 30 large three-generation pedigrees (Figure 1, graph b) of test statistics  $F_{AB,ar}$ ,  $F_{AB,adr}$ ,  $F_{A,ar}$ ,  $F_{A,adr}$ ,  $F_{A,dr}$  and  $F_{AB,dr}$  against trait frequency  $q_1$  or marker allele frequency  $P_A$  at 0.01 significance level, when  $P_A=0.5$  (graphs I and II),  $q_1=0.5$  (graphs III and IV),  $P_B=0.50$ ,  $\sigma_{G_a}^2=0.10$ ,  $\sigma_{G_d}^2=0.05$  and  $h^2=0.15$ , for a dominant mode of inheritance  $a=d=1.0$ , and a recessive mode of inheritance  $a=d=-0.5$ , respectively. The linkage disequilibrium coefficients are  $D_{AB}=(\min(P_A, P_B)-P_AP_B)/2$ ,  $D_{AQ}=(\min(P_A, q_1)-P_Aq_1)/2$  and  $D_{QB}=(\min(P_B, q_1)-P_Bq_1)/2$ .

and 18 members, and the pedigrees range in size from two to three generations. Circulating ACE levels were measured for 405 individuals. In total, 10 biallelic polymorphisms in the ACE gene were genotyped. Multipoint IBD at each marker are calculated by Merlin.<sup>25</sup> Software QTD-2.4.3 is used to analyze the data. There are missing genotype information at markers, and so the total number of individuals is different from marker to marker. For instance, there are four individuals whose genotype information is missing at marker I/D and the total number  $N=401$ ; at marker G2350A, on the other hand, there are more missing genotype data, and  $N=365$ .

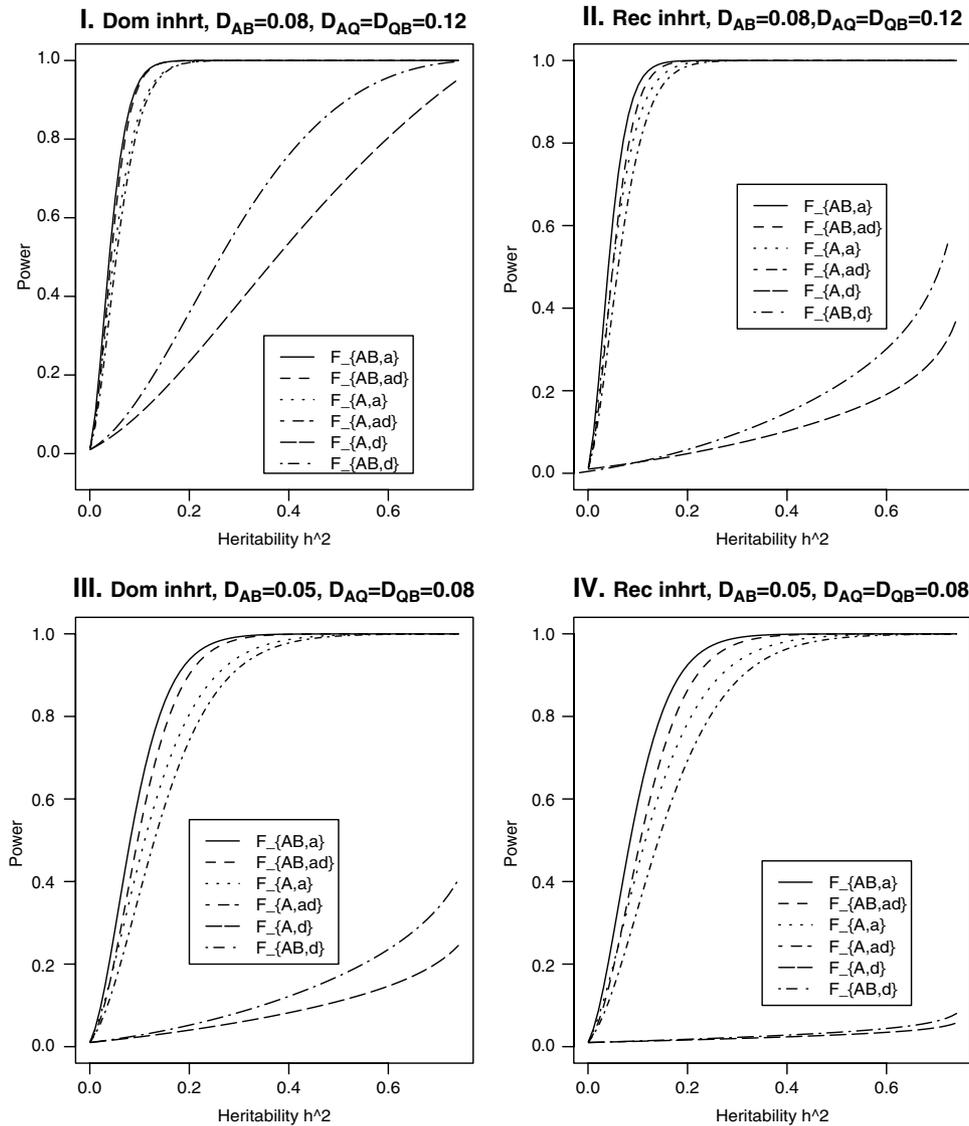
Variance component linkage analysis shows that additive variances are significantly larger than 0, but dominance variances are not significantly larger than 0.<sup>3</sup> Hence, dominance effects can be excluded from regression equation. The total variance is modeled as  $\sigma^2=\sigma_{ga}^2+\sigma_{G_a}^2+\sigma_e^2$ , and related variance-covariance structure can be constructed. Table 5 shows LD analysis of the ACE gene by individual marker. To make comparison with the 'AbAw' approach, results of AbAw's lod are taken from Table 4 of Abecasis *et al.*<sup>3</sup> After fitting the proposed models in this article, *lod* is calculated by  $LRT/(2\ln 10)$ , where  $LRT=2(L_1-L_0)$ ,  $L_1$  is the log-likelihood under  $y_{ij}=\beta+$



**Figure 6** Power of 30 small three-generation pedigrees (Figure 1, graph a) of test statistics  $F_{AB,ar}$ ,  $F_{AB,adr}$ ,  $F_{A,ar}$ ,  $F_{A,adr}$ ,  $F_{A,dr}$  and  $F_{AB,d}$  against heritability  $h^2$  at 0.01 significance level, when  $q_1 = P_A = P_B = 0.50$ ,  $\sigma_{Ga}^2 = 0.10$ ,  $\sigma_{Gd}^2 = 0.05$ , for a dominant mode of inheritance  $a = d = 1.0$ , and a recessive mode of inheritance  $a = 1.0$ ,  $d = -0.5$ , respectively. Abbreviations: Dom is dominant, Inhrt is inheritance, Rec is recessive.

$x_{Aij}\alpha_A + G_{ij} + e_{ij}$ , and  $L_0$  is the log-likelihood under  $y_{ij} = \beta + G_{ij} + e_{ij}$ . The lod scores calculated by the proposed method in this article is generally higher than those of the 'AbAw' approach, which is consistent with the results of Table 4. Hence, the proposed method is advantageous. Moreover, the results of Table 5 confirm the finding that the association is strongest around the G2215A, I/D, G2350A and 4656(CT)3/2 polymorphisms (lod = 27.01, 27.37, 28.01 and 27.93; and the 'AbAw' lod = 14.91, 14.91, 15.76, 14.40 and 14.22, respectively;<sup>3</sup>). The estimations of  $\sigma_{ga}^2$  are 0.05 and 0.03 at markers I/D and G2350A, respectively. In Table 5, the additional linkage effects are presented. The additional linkage lod is calculated

by  $2(L_1 - L_2)/(2 \ln 10)$ , where  $L_2$  is the log-likelihood  $y_{ij} = \beta + x_{Aij}\alpha_A + G_{ij} + e_{ij}$  assuming the total variance  $\sigma^2 = \sigma_{Ga}^2 + \sigma_e^2$  and related variance-covariance structure. For markers I/D and G2350A, the additional linkage lods are 0.09 and 0.03 (the related  $P$ -values are 0.76 and 0.85, respectively). The residual QTL-specific variances of 0.05 and 0.03 for the two ACE markers of largest effect are not significantly different from zero. Little additional linkage effects are present after modeling the association effects. Therefore, these markers are likely in complete LD with the trait alleles, and the genetic variance attributable to the QTL is encompassed in the association coefficient  $\alpha_A$ .<sup>2,3</sup>



**Figure 7** Power of 30 large three-generation pedigrees (Figure 1, graph b) of test statistics  $F_{AB,a}$ ,  $F_{AB,ad}$ ,  $F_{A,a}$ ,  $F_{A,ad}$ ,  $F_{A,d}$ , and  $F_{AB,d}$  against heritability  $h^2$  at 0.01 significance level, when  $q_1 = P_A = P_B = 0.50$ ,  $\sigma_{G_a}^2 = 0.10$ ,  $\sigma_{G_d}^2 = 0.05$ , for a dominant mode of inheritance  $a = d = 1.0$ , and a recessive mode of inheritance  $a = 1.0$ ,  $d = -0.5$ , respectively. Abbreviations: Dom is dominant, Inhrt is inheritance, Rec is recessive.

Notice that there are 10 individual results in Table 5, and it is unclear if there is any relation among the models/markers from results in Table 5. Using the proposed models, we explore LD analysis based on two or more markers. Table 6 shows LD analysis of the ACE gene by two markers. Regressions are given by: (1)  $y_{ij} = \beta + G_{ij} + e_{ij}$ ; (2)  $y_{ij} = \beta + x_{Aij}\alpha_A + G_{ij} + e_{ij}$ ; (3)  $y_{ij} = \beta + x_{Aij}\alpha_A + x_{Bij}\alpha_B + G_{ij} + e_{ij}$ . We first investigate markers I/D and G2350A that show strongest association in individual marker analysis of Table 5. It turns out that marker I/D is correlated with marker G2350A in the following sense: if  $A = I/D$  and  $B = G2350A$ ,  $x_{Aij}$  are correlated to each other. This phenomenon motivates us to further investigate the relation of these two markers.

Utilizing data from <http://www.well.ox.ac.uk/~mfarrall/oxhap-freq.html>, we calculate the frequency of the four haplotypes of these two markers as follows:  $P(IA) = 0.478875$ ,  $P(IG) = 0.002817$ ,  $P(DG) = 0.515494$ ,  $P(DA) = 0.002817$  (note here  $A$  is an allele at marker G2350A). This shows that allele I at marker I/D is almost always present with allele A at marker G2350A, and allele D at marker I/D is almost always present with allele G at marker G2350A. Besides, the measure of LD is 0.246878 between the markers I/D and G2350A. The two markers are almost in complete LD with each other.

Since there are more missing data at marker G2350A than that at marker I/D, we choose I/D as baseline marker

**Table 5** LD analysis of the ACE gene by individual marker

Marker A	N	AbAw's lod	lod	$\hat{\sigma}_{ga}^2$	$\hat{\alpha}_A$	Additional linkage lod
T-5491C	391	9.86	13.91	0.35	-0.695	2.29
A-5466C	392	9.04	14.06	0.29	-0.707	1.62
T-3892C	400	12.49	18.27	0.24	-0.746	1.39
A-240T	401	10.81	13.35	0.32	-0.676	1.72
T-93C	392	10.93	13.00	0.34	-0.678	1.73
T-1237C	377	11.52	20.59	0.20	-0.770	1.13
G2215A	372	14.91	27.01	0.07	0.878	0.17
I/D	401	15.76	27.37	0.05	-0.864	0.09
G2350A	365	14.40	28.01	0.03	-0.883	0.03
4656(CT)3/2	390	14.22	27.93	0.13	-0.893	0.53

The *AbAw's lod* is taken from Table 4 of Abecasis *et al.*<sup>3</sup> Under a total variance  $\sigma^2 = \sigma_{ga}^2 + \sigma_{Ga}^2 + \sigma_e^2$  and related variance-covariance structure, *lod* is calculated by  $LRT/(2 \ln 10)$ , where  $LRT = 2(L_1 - L_0)$ ,  $L_1$  is the log-likelihood under  $y_{ij} = \beta + x_{Aij}\alpha_A + G_{ij} + e_{ij}$ , and  $L_0$  is the log-likelihood under  $y_{ij} = \beta + G_{ij} + e_{ij}$ . Additional linkage lod is calculated by  $2(L_1 - L_2)/(2 \ln 10)$ , where  $L_2$  is the log-likelihood under  $y_{ij} = \beta + x_{Aij}\alpha_A + G_{ij} + e_{ij}$ , assuming the total variance  $\sigma^2 = \sigma_{Ga}^2 + \sigma_e^2$  and related variance-covariance structure.

**Table 6** LD analysis of the ACE gene by two markers

Markers	Regression	N	df	log-L	$H_0$	LRT	lod	P-value
A = I/D B = 4656(CT)3/2	(1)	387	383	-150.79	$\alpha_A = 0$ $\alpha_B = 0$	122.16 7.94	26.53 1.72	<0.001 0.005
	(2)		382	-89.71				
	(3)		381	-85.74				
A = I/D B = T-5491C	(1)	393	389	-150.23	$\alpha_A = 0$ $\alpha_B = 0$	130.86 5.18	28.42 1.12	<0.001 0.023
	(2)		388	-84.80				
	(3)		387	-82.21				
A = I/D B = T-5466C	(1)	390	386	-151.64	$\alpha_A = 0$ $\alpha_B = 0$	127.50 5.38	27.69 1.17	<0.001 0.020
	(2)		385	-87.89				
	(3)		384	-85.20				
A = I/D B = T-3892C	(1)	398	394	-154.99	$\alpha_A = 0$ $\alpha_B = 0$	126.10 5.60	27.38 1.22	<0.001 0.018
	(2)		393	-91.94				
	(3)		392	-89.14				
A = I/D B = A-240T	(1)	398	394	-155.81	$\alpha_A = 0$ $\alpha_B = 0$	125.22 4.22	27.19 1.22	<0.001 0.040
	(2)		393	-93.20				
	(3)		392	-91.09				
A = I/D B = T-93C	(1)	389	385	-153.53	$\alpha_A = 0$ $\alpha_B = 0$	120.94 4.10	26.26 0.89	<0.001 0.043
	(2)		384	-93.06				
	(3)		383	-91.01				

Regressions are given by: (1)  $y_{ij} = \beta + G_{ij} + e_{ij}$ ; (2)  $y_{ij} = \beta + x_{Aij}\alpha_A + G_{ij} + e_{ij}$ ; (3)  $y_{ij} = \beta + x_{Aij}\alpha_A + x_{Bij}\alpha_B + G_{ij} + e_{ij}$ . Log-L is the abbreviation of log-likelihood.  $lod = LRT/(2 \ln 10)$  and  $LRT = 2(L_1 - L_0)$ , where  $L_0$  is the log-likelihood under the null hypothesis  $H_0$ , and  $L_1$  is the log-likelihood under the alternative hypothesis. For example,  $122.16 = 2(-89.71 + 150.79)$  and  $7.94 = 2(-85.74 + 89.71)$ , in the case of  $A = I/D$ ,  $B = 4656(CT)3/2$ .

A. Six markers, 4656(CT)3/2, T-5491C, A-5466C, T-3892C, A-240T and T-93C, show significant association at a 0.05 significance level, in addition to the association of marker I/D (see the *P*-value of test  $H_0: \alpha_B = 0$  in Table 6). In these six markers, 4656(CT)3/2 shows strongest additional association with  $LRT = 7.94$ ,  $lod = 1.72$ ,  $P$ -value = 0.005 (Table 6). Hence, we choose I/D and 4656(CT)3/2 as markers *A* and *B* for the basis of further investigation. Table 7 shows results by one more marker in addition to markers I/D and 4656(CT)3/2. Four markers,

T-5491C, A-5466C, T-3892C and A-240T, show significant association at a 0.05 significance level, in addition to the association of markers I/D and 4656(CT)3/2 (see the *P*-value of test  $H_0: \alpha_C = 0$  in Table 7). Further investigation shows that there is no significant evidence to analyze the data using more than three markers in the regression equations.

Notice that if the significance level is set at 0.01 instead of 0.05, then only markers I/D and 4656(CT)3/2 are associated with the trait locus (Table 6). The regression,

**Table 7** LD analysis of the ACE gene by multiple markers

Markers	Regression	N	Df	log-L	H <sub>0</sub>	LRT	lod	P-value
A = I/D	(1)	379	375	-144.90				
B = 4656(CT)3/2	(2)		374	-81.40	$\alpha_A = 0$	127.00	27.58	< 0.001
C = T-5491C	(3)		373	-77.78	$\alpha_B = 0$	7.24	1.57	0.007
	(4)		372	-75.21	$\alpha_C = 0$	5.14	1.12	0.023
A = I/D	(1)	376	372	-146.90				
B = 4656(CT)3/2	(2)		371	-84.66	$\alpha_A = 0$	123.46	26.81	< 0.001
C = A-5466C	(3)		370	-80.58	$\alpha_B = 0$	8.16	1.77	0.004
	(4)		369	-77.92	$\alpha_C = 0$	5.32	1.16	0.021
A = I/D	(1)	384	380	-149.72				
B = 4656(CT)3/2	(2)		379	-88.65	$\alpha_A = 0$	122.14	26.58	< 0.001
C = T-3892C	(3)		378	-84.24	$\alpha_B = 0$	8.90	1.93	0.003
	(4)		377	-82.19	$\alpha_C = 0$	4.02	0.87	0.045
A = I/D	(1)	384	380	-150.54				
B = 4656(CT)3/2	(2)		379	-89.87	$\alpha_A = 0$	121.34	26.35	< 0.001
C = A-240T	(3)		378	-75.89	$\alpha_B = 0$	7.96	1.73	0.005
	(4)		377	-83.90	$\alpha_C = 0$	3.98	0.86	0.046

Regressions are given by: (1)  $y_{ij} = \beta + G_{ij} + e_{ij}$ ; (2)  $y_{ij} = \beta + x_{Aij}\alpha_A + G_{ij} + e_{ij}$ ; (3)  $y_{ij} = \beta + x_{Aij}\alpha_A + x_{Bij}\alpha_B + G_{ij} + e_{ij}$ ; (4)  $y_{ij} = \beta + x_{Aij}\alpha_A + x_{Bij}\alpha_B + x_{Cij}\alpha_C + G_{ij} + e_{ij}$ . Log-L is the abbreviation of log-likelihood.  $lod = LRT / (2 \ln 10)$  and  $LRT = 2(L_1 - L_0)$ , where  $L_0$  is the log-likelihood under the null hypothesis  $H_0$ , and  $L_1$  is the log-likelihood under the alternative hypothesis.

$y_{ij} = \beta + x_{Aij}\alpha_A + x_{Bij}\alpha_B + G_{ij} + e_{ij}$ , A = I/D, B = 4656(CT)3/2, can fully interpret association at the 0.01 significance level, which provides a unique result.

## Discussion

This article focuses on building models that may utilize extended multi generation large pedigrees for LD mapping of QTL. Based on the same logic developed in our previous work,<sup>7,8</sup> variance component models are constructed for high resolution joint linkage and LD mapping of QTL. By type I error rate evaluation, it is found that the proposed models have correct type I errors and the methods are robust. The methods proposed are compared with the 'AbAw' approach. It is found that our models have higher power. This is consistent with our previous findings based on sibpair data.<sup>7</sup> Based on power comparison, it is shown that large pedigrees may contain more LD information than small pedigrees. Moreover, the proposed method is applied to ACE data with a better result than that of the 'AbAw' approach. In addition, it is found that markers I/D and 4656(CT)3/2 can fully describe the association with the trait locus at a 99% significance level for the ACE data. This result is very meaningful. For instance, all 10 biallelic markers of ACE data show association with the trait locus. However, it is markers I/D and 4656(CT)3/2 which are really necessary in describing the association at the 99% significance level. Marker G2350A is correlated to marker I/D, as a matter of fact!

One possible explanation that the proposed method is more powerful than the 'AbAw' approach is as follows. The

proposed method decomposes the association into the summation of additive and dominance effects. If dominance effect is not significantly larger than 0, then only additive effect is modeled such as the ACE data. The 'AbAw' approach, on the other hand, decomposes the association into the summation of between-family (b) and within-family (w) components. In Table 4 of Abecasis *et al*,<sup>3</sup> the 'AbAw' lod is reported to be the within-family effect, which is less powerful than the proposed method. By decomposing the association effect into association between-family and association within-family in the 'AbAw' approach, the information is diluted and so the approach is less powerful. This is, in the opinion of the authors of the current article, is not a criticism for the 'AbAw' approach, since it may be a good method in certain circumstances. It is our hope that the current article may stimulate more interest in exploring the most appropriate method in LD mapping of complex diseases.

Population may contain LD information, and pedigree data may contain both linkage and LD information. In mapping complex diseases, it is difficult to get replicate linkage findings using pedigree data. It is interesting to combine the same pedigree data with population data for fine association study. In general, any type of pedigree data can be combined with population data for a unified analysis of linkage and LD mapping of QTL, based on the models developed in the current paper and our previous research.<sup>7,8</sup> Linkage analysis can localize genetic traits in a broad region, and is less sensitive to population structures. LD mapping, on the other hand, is appropriate for fine high-resolution mapping and sensitive to population admixtures. As the first step, one may perform linkage

analysis using pedigree data to get prior linkage evidences based on a sparse genetic map.<sup>26,27</sup> Then, a combined analysis using both population data and pedigree data can take advantage of both linkage and LD mapping for high resolution of genetic traits. That is, it may provide high resolution of LD mapping, and be more likely to avoid the false positives by using the prior linkage evidences based on a dense genetic map.<sup>28</sup>

Up to now, most research focus on using biallelic markers in LD mapping of QTL. It would be interesting in building models using multiallelic markers such as macrosatellites or haplotype blocks. One problem in using multiallelic markers in analysis is that the number of parameters can be big, which may lead to high degrees of freedom of test statistics. Hence, selection of significant markers and relevant alleles is an important issue. In addition, genotyping errors can heavily impact the result of a study. These issues deserve more in depth investigations.

### Acknowledgements

We thank two reviewers, and Dr Gert-Jan B von 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. We are grateful to Dr Abecasis for kindly providing the simulation program PEDSIMUL to generate simulated datasets.

### References

- Fan R, Xiong M: High resolution mapping of quantitative trait loci by linkage disequilibrium analysis. *Eur J Hum Genet* 2002; **10**: 607–615.
- 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.
- Abecasis GR, Cookson WOC, Cardon LR: Pedigree tests of linkage disequilibrium. *Eur J Hum Genet* 2000; **8**: 545–551.
- 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.
- Allison DB, Bonnie T, St Jean P, Elston RC, Infante MC, Schork NJ: Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. *Am J Hum Genet* 1998; **63**: 1190–1201.
- Almasy L, Williams JT, Dyer TD, Blangero J: Quantitative trait locus detection using combined linkage/disequilibrium analysis. *Genet Epidemiol* 1999; **17** (Suppl 1): S31–S36.
- Fan R, Jung J: High resolution joint linkage disequilibrium and linkage mapping of quantitative trait loci based on sibship data. *Hum Hered* 2003; **56**: 166–187.
- Fan R, Xiong M: Combined high resolution linkage and association mapping of quantitative trait loci. *Eur J Hum Genet* 2003; **11**: 125–137.
- 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.
- Göring HHH, Terwillinger JD: Linkage analysis in the presence of error IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am J Hum Genet* 2000; **66**: 1310–1327.
- Martin ER, Monks SA, Warren LL, Kaplan NL: A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000; **67**: 146–154.
- 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.
- George V, Tiwari HK, Zhu XF, Elston RC: A test of transmission/disequilibrium for quantitative traits in pedigree data, by multiple regression. *Am J Hum Genet* 1999; **65**: 236–245.
- Farrall M, Keavney B, MckKenzie CA *et al.*: Fine mapping of an ancestral recombination break-point in DCP1. *Nat Genet* 1999; **23**: 270–271.
- Keavney B, MckKenzie CA, Connell JM *et al.*: Measured haplotype analysis of the angiotensin-1 converting enzyme gene. *Hum Mol Genet* 1998; **7**: 1745–1751.
- Amos CI: Robust variance-components approach for assessing linkage in pedigrees. *Am J Hum Genet* 1994; **54**: 534–543.
- Amos CI, Elston RC: Robust methods for the detection of genetic linkage for quantitative data from pedigrees. *Genet Epidemiol* 1989; **6**: 349–360.
- Amos CI, Elston RC, Wilson AF, Bailey-Wilson JE: A more powerful robust sib-pair test of linkage for quantitative traits. *Genet Epidemiol* 1989; **6**: 435–449.
- Fulker DW, Cherny SS, Cardon LR: Multiple interval mapping of quantitative trait loci, using sib-pairs. *Am J Hum Genet* 1995; **56**: 1224–1233.
- Almasy L, Blangero J: Multipoint quantitative trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**: 1198–1211.
- Goldgar DE, Oniki RS: Comparison of a multipoint identity-by-descent method with parametric multipoint linkage analysis for mapping quantitative traits. *Amer J Hum Genet* 1992; **50**: 598–606.
- Pratt SC, Daly M, Kruglyak L: Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 2000; **66**: 1153–1157.
- Falconer DS, Mackay TFC: *Introduction to quantitative genetics*, 4th edn. London: Longman, 1996.
- Graybill FA: *Theory and application of the linear model*. Belmont, CA: Wadsworth Publishing Company, Inc., 1976.
- Abecasis GR, Cherny SS, Cookson WOC, Cardon LR: Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- Broman KW, Murray JC, Sheffied VC, White RL, Weber JL: Comprehensive human genetic map: individual and sex-specific variation in recombination. *Am J Hum Genet* 1998; **63**: 861–869.
- Kong A, Gudbjartsson DF, Sainz J *et al.*: A high resolution recombination map of the human genome. *Nat Genet* 2002; **31**: 241–247.
- 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.

### Appendix A

In this appendix, we provides expectations of product of indicator functions (4) for a few common type of relatives. If individuals 1 and 2 are heterozygous sibs  $P(IBD=0)=P(IBD=2)=1/4$ ,  $P(IBD=1)=1/2$  and then

$$E[(1, x_{A1}, x_{B1}, z_{A1}, z_{B1})^T (1, x_{A2}, x_{B2}, z_{A2}, z_{B2})] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & V_A/2 & 0 \\ 0 & 0 & V_D/4 \end{pmatrix}. \quad (A.1)$$

If individuals 1 and 2 are a parent–offspring pair,  $P(IBD=0)=P(IBD=2)=0$ ,  $P(IBD=1)=1$  and then

$$E[(1, x_{A1}, x_{B1}, z_{A1}, z_{B1})^\tau (1, x_{A2}, x_{B2}, z_{A2}, z_{B2})] = \begin{pmatrix} 1 & O & O \\ O & V_A/4 & O \\ O & O & O \end{pmatrix}. \quad (A.2)$$

If individuals 1 and 2 are a grandparent/grandchild pair (or uncle/niece or aunt/nephew pair),  $P(IBD=0)=1/2$ ,  $P(IBD=2)=0$ ,  $P(IBD=1)=1/2$  and then

$$E[(1, x_{A1}, x_{B1}, z_{A1}, z_{B1})^\tau (1, x_{A2}, x_{B2}, z_{A2}, z_{B2})] = \begin{pmatrix} 1 & O & O \\ O & V_A/4 & O \\ O & O & O \end{pmatrix}. \quad (A.3)$$

If individuals 1 and 2 are first cousins,  $P(IBD=0)=3/4$ ,  $P(IBD=2)=0$ ,  $P(IBD=1)=1/4$  and then

$$E[(1, x_{A1}, x_{B1}, z_{A1}, z_{B1})^\tau (1, x_{A2}, x_{B2}, z_{A2}, z_{B2})] = \begin{pmatrix} 1 & O & O \\ O & V_A/8 & O \\ O & O & O \end{pmatrix}. \quad (A.4)$$

In general, consider two individuals 1 and 2 who are noninbred relatives. Let  $\Phi_{12}$  be their kinship coefficient, and  $\Delta_{712}$  be the probability that both alleles shared by the two individuals are IBD at any locus. Then

$$E[(1, x_{A1}, x_{B1}, z_{A1}, z_{B1})^\tau (1, x_{A2}, x_{B2}, z_{A2}, z_{B2})] = \begin{pmatrix} 1 & O & O \\ O & 2\Phi_{12}V_A & O \\ O & O & \Delta_{712}V_D \end{pmatrix}. \quad (A.5)$$

### Appendix B

Based on Equations (5) and (A.1), (A.2), (A.3) and (A.4), an approximation of noncentrality parameter of test statistics  $F$  can be obtained as follows: Let us denote  $\Sigma_i^{-1} = (1/\sigma^2)(\gamma_{kl})_{n \times n}$   $n=11$  for graph A and  $n=18$  for graph B of Figure 1, respectively. When  $I$  is sufficiently large, the laws of large numbers implies that

$$\frac{1}{I} \sum_{i=1}^I X_i^\tau \Sigma_i^{-1} X_i \approx \frac{1}{\sigma^2} \begin{pmatrix} \sum_k \sum_l \gamma_{kl} & O & O \\ O & b_1 V_A & O \\ O & O & b_2 V_D \end{pmatrix}$$

Here,  $b_1$  and  $b_2$  are constants as follows for pedigrees in graph A of Figure 1

$$b_1 = \sum_{k=1}^{11} \gamma_{kk} + [\gamma_{15} + (\gamma_{17} + \dots + \gamma_{1,11})/2] + [\gamma_{25} + (\gamma_{27} + \dots + \gamma_{2,11})/2] + [\gamma_{36} + (\gamma_{37} + \dots + \gamma_{3,11})/2] + [\gamma_{46} + (\gamma_{47} + \dots + \gamma_{4,11})/2] + [\gamma_{57} + \dots + \gamma_{5,11}] + [\gamma_{67} + \dots + \gamma_{6,11}] \quad (B.1)$$

$$+ \sum_{k=7}^{11} \sum_{l=k+1}^{11} \gamma_{kl}$$

$$b_2 = \sum_{k=1}^{11} \gamma_{kk} + \sum_{k=7}^{11} \sum_{l=k+1}^{11} \gamma_{kl}/2$$

For pedigrees in graph B of Figure 1, constants  $b_1$  and  $b_2$  are given by

$$b_1 = \sum_{k=1}^{18} \gamma_{kk} + \gamma_{1,12} + [\gamma_{2,12} + (\gamma_{2,13} + \dots + \gamma_{2,16})/2] + [\gamma_{3,12} + \dots + \gamma_{3,16})/2] + [\gamma_{4,12} + \dots + \gamma_{4,16})/2] + [\gamma_{5,12}/2 + (\gamma_{5,13} + \dots + \gamma_{5,16})] + [(\gamma_{6,13} + \dots + \gamma_{6,16}) + (\gamma_{6,17} + \gamma_{6,18})/2] + [\gamma_{7,13} + \dots + \gamma_{7,18})/2] + [(\gamma_{8,13} + \dots + \gamma_{8,16})/2 + (\gamma_{8,17} + \gamma_{8,18})] + (\gamma_{9,17} + \gamma_{9,18}) + (\gamma_{10,17} + \gamma_{10,18})/2 + [(\gamma_{11,17} + \dots + \gamma_{11,18})/2 + (\gamma_{12,13} + \dots + \gamma_{12,16})/4] + (\gamma_{13,14} + \gamma_{13,15} + \gamma_{13,16}) + (\gamma_{14,15} + \gamma_{14,16}) + \gamma_{15,16} + (\gamma_{13,17} + \dots + \gamma_{16,17})/4 + [\gamma_{13,18} + \dots + \gamma_{16,18})/4 + \gamma_{17,18}$$

$$b_2 = \sum_{k=1}^{18} \gamma_{kk} + [(\gamma_{13,14} + \gamma_{13,15} + \gamma_{13,16}) + (\gamma_{14,15} + \gamma_{14,16}) + \gamma_{15,16}]/2 + \gamma_{17,18}/2 \quad (B.2)$$

Therefore, the noncentrality parameter can be approximated by

$$\lambda = (H\mu)^\tau \left[ H \left[ \sum_{i=1}^I X_i^\tau \Sigma_i^{-1} X_i \right]^{-1} H^\tau \right]^{-1} (H\mu)$$

$$\approx \frac{I}{\sigma^2} (H\mu)^\tau \left[ H \begin{pmatrix} 1/\sum_k \sum_l \gamma_{kl} & O & O \\ O & V_A^{-1}/b_1 & O \\ O & O & V_D^{-1}b_2 \end{pmatrix} H^\tau \right]^{-1} (H\mu)$$