Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model

Yongcai Mao, Nicole R. London, Li Ma, Daniel Dvorkin, and Yang Da
Department of Animal Science, University of Minnesota, Saint Paul, Minnesota

Submitted 31 May 2006; accepted in final form 28 August 2006

Mao Y, London NR, Ma L, Dvorkin D, Da Y. Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model. Physiol Genomics 28: 46–52, 2006. First published August 29, 2006; doi:10.1152/physiolgenomics.00096.2006.—Epistasis effects (gene interactions) have been increasingly recognized as important genetic factors underlying complex traits. The existence of a large number of single nucleotide polymorphisms (SNPs) provides opportunities and challenges to screen DNA variations affecting complex traits using a candidate gene analysis. In this article, four types of epistasis effects of two candidate gene SNPs with Hardy-Weinberg disequilibrium (HWD) and linkage disequilibrium (LD) are considered: additive × additive, additive × dominance, dominance × additive, and dominance × dominance. The Kempthorne genetic model was chosen for its appealing genetic interpretations of the epistasis effects. The method in this study consists of extension of Kempthorne’s definitions of 35 individual genetic effects to allow HWD and LD, genetic contrasts of the 35 extended individual genetic effects to define the 4 epistasis effects, and a linear model method for testing epistasis effects. Formulas to predict statistical power (as a function of contrast heritability, sample size, and type I error) and sample size (as a function of contrast heritability, type I error, and type II error) for detecting each epistasis effect were derived, and the theoretical predictions agreed well with simulation studies. The accuracy in estimating each epistasis effect and rates of false positives in the absence of all or three epistasis effects were evaluated using simulations. The method for epistasis testing can be a useful tool to understand the exact mode of epistasis, to assemble genome-wide SNPs into an epistasis network, and to assemble all SNP effects affecting a phenotype using pairwise epistasis tests.

single nucleotide polymorphism; quantitative trait; linkage disequilibrium; Hardy-Weinberg disequilibrium

The significance of epistasis in complex traits has been well recognized (3, 12, 16, 17). It has been hypothesized that epistasis is ubiquitous in determining susceptibility to common human diseases. Observations supporting this hypothesis include deviations from Mendelian ratios, molecular interactions in gene regulation and biochemical and metabolic systems, nonrepeaatable positive effects of single polymorphisms, and gene interaction effects commonly found when properly investigated (16). Large epistasis networks found in yeast (19, 21) also point to the ubiquitous nature of epistasis. The presence of a large number of single nucleotide polymorphisms (SNPs) provides opportunities and challenges to identify DNA variations associated with complex traits using SNPs as candidate genes. A number of studies on SNP-disease association have been reported (2, 7, 8, 20, 24), but most current candidate gene studies focus on single gene effects. Ongoing studies of genotyping up to 500,000 SNPs on 2,000 individuals have been reported (1). Such large-scale SNP studies would be capable of detecting certain epistasis effects.1

For two bi-allelic genes unaffected by the environment, a total of 512 disease models are possible, but the number of nonredundant models is in the range of 50–102 (12). For two bi-allelic genes such as two SNP loci with quantitative measures, epistasis effects are typically modeled by a linear partition of the nine possible genotypic values in quantitative genetics. Fisher (6) partitioned the nine genotypic values into single gene effects (additive and dominance effects) and epistasis effects assuming Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE). Also assuming HWE and LE, Cockerham (5) and Kempthorne (10) partitioned Fisher’s epistasis effect into four components, additive × additive, additive × dominance, dominance × additive, and dominance × dominance epistasis effect with the genetic interpretation of allele × allele, allele × genotype, genotype × allele, and genotype × genotype interactions, respectively. This partitioning can be used as a tool for identifying the exact mode of a gene interaction effect. Although these four types of epistasis effects could be further partitioned, such a more detailed partitioning may not have practical value, because nine genotypic values allow only nine independent parameters (common mean and 8 genetic effects) to be estimated and eight independent genetic effects to be tested using linear model methods. A statistical model that includes the main effect of each locus and the interaction between the two loci can be used to detect the gene interaction between the two loci, but this type of epistasis detection does not offer a mechanism to test each of the four epistasis effects under the Cockerham and Kempthorne models. Cockerham’s model uses orthogonal contrasts of genotypic values, whereas Kempthorne’s model uses the deviation of “genetic combination” from the sum of individual genetic factors. For example, an additive × additive effect (combination effect of 2 alleles, each from a different locus) is a deviation of the genotypic mean from the two single allelic means. This is clearly a measure for the gene combination

---

1 The 2nd International Symposium on Animal Functional Genomics was held May 16–19, 2006, at Michigan State University in East Lansing, MI, and was organized by Jeanne Burton of Michigan State University and Guillerme J. M. Rosa of University of Wisconsin, Madison, WI (see meeting report by Drs. Burton and Rosa, Physiol Genomics 28: 1-4, 2006).
effect and hence has a more straightforward genetic interpretation of the epistasis effect. A genetic model (4) without the derivations of Cockerham and Kempthorne could also be used for testing SNP epistasis effects, but this model yields the same epistasis results as the Cockerham and Kempthorne models under the assumptions of HWE, LE, and equal allele frequencies (see Supplemental Materials; online version of article contains Supplemental Materials). With LD, an epistasis model defining each epistasis effect as a weighted average of genotypic values has been described (9). This model could easily be extended to account for HWD. Among the above epistasis models, the Kempthorne model has more straightforward genetic interpretations of epistasis effects and can be easily extended to allow HWD and LD.

In this article, we extend the Kempthorne model for genetic effects of quantitative traits to allow HWD and LD, define contrasts of the extended Kempthorne genetic effects for testing the four epistasis effects of two SNP loci, develop a linear model procedure for testing each epistasis effect, and present analytical and simulation results on statistical power and sample size required for detecting each type of epistasis effect. The accuracy in estimating each epistasis effect, the rate of false-positive results in the absence of the true epistasis effects, and the rate of potential confounding between the four types of epistasis effects are evaluated using simulations. The methods in this article should provide a useful tool for genome-wide pairwise testing of SNP epistasis where HWD and LD may exist, and for the assembly of total SNP effects or an epistasis network of a phenotype.

**MATERIALS AND METHODS**

The development of SNP epistasis testing methods in this article consists of three steps: extend the Kempthorne model for genetic effects to allow HWD and LD, construct genetic and statistical contrast for testing each epistasis effect, and evaluate the performance of the resulting method for testing epistasis with HWD and LD. Two bi-allelic linked loci, locus 1 with A and a alleles and locus 2 with B and b, are assumed to affect a complex trait of quantitative nature. Each locus is allowed to have HWD. For these two bi-allelic loci, nine genotypes are possible. Let \( g_{ijkl} \) and \( p_{ijkl} \) denote the genotypic value and genotypic probability of individuals possessing genotype \( ij \) at locus 1 and \( kl \) at locus 2, where \( i, j, k, \) and \( l \) denote alleles A and a of locus 1, and \( k \) and \( l \) denote alleles B and b of locus 2 (Table 1). Under the Kempthorne model, each genetic effect is defined as a deviation of genetic combination effect from the sum of individual genetic factors in the genetic combination. Such definitions of gene effects require the calculation of 27 means of genotypic values: 1 population mean and 26 marginal means, which include 4 allelic means, 6 single locus genotypic means, 4 gametic means, 6 allele-genotype means, and 6 genotype-allele means.

To extend the Kempthorne definitions of genetic effects to allow HWD and LD, the calculation of the population mean requires the use of the observed genotypic frequencies (\( p_{ijkl} \)), and the calculations of the 26 marginal means require the use of a series of conditional probabilities based on the observed genotypic frequencies rather than using allele frequencies under HWE and LE, as shown in the Supplemental Material. Let \( \mu_i \) = the mean genotypic value in the population, \( \mu_{ijk} \) = the marginal mean of genotypic values for individuals with alleles \( i = A, a; k = B, b \), \( \mu_{ijl} \) = the marginal mean of genotypic values for individuals with allele \( k = B \), \( b \), \( \mu_{ijkl} \) = the marginal mean of genotypic values for individuals with genotype \( ij \) at the first locus \( ij = AA, Aa, aa \), \( \mu_{ijk} \) = the marginal mean of genotypic values for individuals with allele \( i = a \) at locus 1 and allele \( k = b \) at locus 2 (i = A, a; k = B, b), \( \mu_{ijkl} \) = the marginal mean of genotypic values for individuals with genotype \( ij \) at locus 1 and genotype \( kl \) at locus 2 (i = A, a; k = B, b), and \( \mu_{ijkl} \) = the marginal mean of genotypic values for individuals with genotype \( ij \) at locus 1 and allele \( k = b \) at locus 2 (i = AA, Aa; a = k = B, b). With the use of 27 means, typical expressions of 35 possible individual genetic effects of two bi-allelic loci using Kempthorne’s definitions can be expressed as follows:

\[
\begin{align*}
& a_i = \mu_i - \mu, \quad i = A, a \quad (1.1) \\
& a_k = \mu_k - \mu, \quad k = B, b \quad (1.2) \\
& d_{ij} = \mu_{ij} - \mu_i - a_i - a_j, \quad ij = AA, Aa, aa \quad (1.3) \\
& d_{ik} = \mu_{ik} - \mu_i - a_i - a_k, \quad ik = BB, Bb, bb \quad (1.4) \\
& (aa)_i = \mu_{ik} - \mu - a_i - a_k, \quad ik = AB, Ab, aB, ab \quad (1.5) \\
& (da)_{ik} = \mu_{jkl} - \mu_i - a_i - a_k - d_{ij} - (aa)_i - (aa)_k, \\
& i = A, a; k = BB, Bb, bb \quad (1.6) \\
& (da)_{ik} = \mu_{jkl} - \mu_i - a_i - a_k - d_{ij} - (aa)_i - (aa)_k, \\
& ij = AA, Aa, aa; k = B, b \quad (1.7) \\
& (dd)_{ijkl} = g_{ijkl} - \mu_i - a_i - a_j - a_k - a_l - d_{ij} - d_{kl} - (aa)_i - (aa)_k - (aa)_l - (dd)_{ijkl}, \\
& ij = AA, Aa, aa; k = BB, Bb, bb \quad (1.8)
\end{align*}
\]

where \( a_i \) = additive effect of allele \( i \) at locus 1 (i = A, a), \( a_k \) = additive effect of allele \( k \) at locus 2 (k = B, b), \( d_{ij} \) = dominance effect of genotype \( ij \) at locus 1, \( d_{ik} \) = dominance effect of genotype \( kl \) at locus 2, \( (aa)_i \) = additive \times additive epistasis effect of genotypes with alleles \( ik \), \( (dd)_{ijkl} \) = dominance \times dominance epistasis effect of genotypes with alleles \( ikl \), \( (dd)_{ijkl} \) = dominance \times additive epistasis effect of genotypes with alleles \( ikl \), and \( (dd)_{ijkl} \) = dominance \times dominance epistasis effect of \( ijl \) genotypes. The mathematical expressions of genetic effects of Eqs. 1.1–1.8 do not change with or without HWD and LD. The only change required to allow HWD and LD is in the calculations of the 27 means. With the definitions of Eqs. 1.1–1.8, the Kempthorne model of a genotypic value for two loci is

\[
g_{ijkl} = \mu + a_i + a_j + a_k + a_l + d_{ij} + d_{kl} + (aa)_i + (aa)_j + (aa)_k + (dd)_{ijl} + (dd)_{ijkl} \quad (2)
\]

With the extended Kempthorne genetic effects, genetic contrasts for testing the eight individual genetic effects of two loci allowing HWD and LD can be defined. The resulting contrasts for epistasis testing will be evaluated for its performance, including statistical power, rate of false positives, and accuracy in estimating each epistasis effect. Analytical formulas will be derived to predict the statistical power and sample size requirement for detecting epistasis effects, and the theoretical results will be compared with simulation results. Two types of false-positive results will be considered, namely false positives in the absence of any epistasis effects and false positives in the presence of one type of epistasis effect. The former is to evaluate whether the tests have more false positives than expected for random reasons, and the latter is to evaluate whether any epistasis effect has more false positives due to confounding with other epistasis effects. The rates of these two types of false-positive results will be evaluated using simulations. The accuracy in estimating each epistasis effect will be

---

Table 1. Genotypic values \( (g_{ijkl}) \) and frequencies \( (p_{ijkl}) \) for 2 bi-allelic loci

<table>
<thead>
<tr>
<th>Genotype</th>
<th>BB</th>
<th>Bb</th>
<th>bb</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>AA</td>
<td>Aa</td>
<td>aa</td>
</tr>
<tr>
<td>Aa</td>
<td>Aa</td>
<td>AA</td>
<td>aA</td>
</tr>
<tr>
<td>aa</td>
<td>aa</td>
<td>aa</td>
<td>aa</td>
</tr>
</tbody>
</table>

**Physiol Genomics • VOL 28 • www.physiolgenomics.org**
measured by the mean square error (MSE) from simulation results, where MSE is the sum of the variance and squared bias of the epistasis estimates.

RESULTS

Contrasts of gene effects allowing HWD and LD. On the basis of the definitions of genetic effects by Eqs. 1.1–1.8 with the extension to allow HWD and LD, eight genetic contrasts can be defined to test for four single gene effects and four epistasis effects as follows

\[ a_1 = a_A - a_a = k_A^i \beta = s_i g \]  
\[ a_2 = a_B - a_b = k_B^i \beta = s_i g \]  
\[ d_1 = d_{AA} - \frac{1}{2} (d_A + d_{aa}) = k_A \beta = s_i g \]  
\[ d_2 = d_{BB} - \frac{1}{2} (d_B + d_{bb}) = k_B \beta = s_i g \]  
\[ i_{aa} = (aa)_{AB} - (aa)_{Ab} - [(aa)_{ab} - (aa)_{ab}] = k_{ad} \beta = s_i g \]  
\[ i_{ad} = \frac{1}{2} [(ad)_{ABB} + (ad)_{Abb}] \]  
\[ i_{da} = \frac{1}{2} [(da)_{AAB} + (da)_{aab}] \]  
\[ i_{dd} = \frac{1}{2} [(dd)_{ABB} + (dd)_{abb}] \]  
\[ i_{aa} = (aa)_{AB} - (aa)_{Ab} - [(aa)_{ab} - (aa)_{ab}] = k_{ad} \beta = s_i g \]  
\[ i_{ad} = \frac{1}{2} [(ad)_{ABB} + (ad)_{Abb}] \]  
\[ i_{da} = \frac{1}{2} [(da)_{AAB} + (da)_{aab}] \]  
\[ i_{dd} = \frac{1}{2} [(dd)_{ABB} + (dd)_{abb}] \]  

where \( \alpha_i \) = additive effect (gene substitution effect) of locus \( i \) (\( i = 1,2, \ldots, 9 \)), \( d_i \) = dominance effect of locus \( i \) (\( i = 1,2, \ldots, 9 \)), \( i_{aa} \) = additive × additive effect, \( i_{ad} \) = additive × dominance effect, \( i_{da} \) = dominance × additive effect, \( i_{dd} \) = dominance × dominance effect, \( k_i \) = contrast vector of \( B \) (\( i = 2, \ldots, 9 \)), \( \beta \) = column vector of the population mean and the 35 genetic effects represented by Eqs. 1.1–1.8, \( g = (g_AAB, g_{AAb}, g_AaB, g_Aab, g_{ABB}, g_{AbB}, g_{aBB}, g_{aab}, g_{ABB}, g_{AbB}, g_{aBB}, g_{aab}) \), and \( s_i \) is a function of conditional probabilities (Supplemental Material). Under the null hypothesis of no gene combination effect, each of the four epistasis contrasts of Eqs. 3.5–3.8 is expected to be zero. Eqs. 3.1–3.8 are orthogonal comparisons of the 35 individual genetic effects represented by Eqs. 1.1–1.8. Therefore, the test for each genetic effect is independent of the other genetic effects.

Linear model method for epistasis testing. To test the significance of each genetic contrast, a statistical contrast needs to be used in place of the genetic contrast. Each statistical contrast is obtained by replacing \( g \) with the estimate of \( g (\hat{g}) \) using the following procedure. A phenotypic value is modeled as the summation of a genotypic value \((g_{ijkl})\) of the two genes underlying the complex trait and a random residual \((e)\) following \(N(0, \sigma_e^2)\) distribution, i.e., phenotypic observation \( m \) with \( g_{ijkl} \) genotypic value is modeled as \( y_{ijkl} = g_{ijkl} + e_{ijkl} \), or \( y = Xg + e \) in matrix notations, where \( y \) is the column vector of phenotypic values, \( X \) is the design matrix, and \( e \) is the vector of random residuals. Note that the above linear model can be extended to include nongenetic effects such as age and gender so that the genetic effects can be tested without the influence of those nongenetic effects. This flexibility of accounting for nongenetic effects is an advantage of linear model methods over frequency-based methods in genetic analysis. The normal equations in matrix notations are \( X^T X \hat{g} = X^T y \), and the least squares estimate of \( g \) is given by \( \hat{g} = (X^T X)^{-1} X^T y \). Then, from Eqs. 3.5–3.8, the four marker contrasts for testing epistasis effects can be expressed as

\[ \hat{L}_i = s_i^2 \hat{g}, \quad (x,i) = (\alpha_i, 2), \ldots, (dd, 9) \]  

The epistasis estimator represented by Eq. 4 is unbiased, i.e.,

\[ E(L_{aa}) = i_{aa}, \quad E(L_{ad}) = i_{ad}, \quad E(L_{da}) = i_{da}, \quad E(L_{dd}) = i_{dd} \]  

The null and alternative hypotheses for testing each epistasis effect are \( H_0: i_{ad} = 0 \) and \( H_{1d}: i_{ad} \neq 0 \), respectively, where \( x = aa, ad, da, dd \). It is readily seen from Eq. 5 that \( H_0 \) is equivalent to the hypothesis \( L_x = 0 \). The test statistic for testing \( H_0 \) is

\[ T_x = \frac{L_x}{\hat{\sigma}_L} = \frac{s_i^2 \hat{g}}{\sigma_L} \]  

where \( s_i^2 = (y - X\hat{g})'(y - X\hat{g})/(n - k) \). Under \( H_0 \), this statistic has a Student’s \( t \)-distribution with degrees of freedom \( n - k \) (18), where \( n \) is the number of observations and \( k \) is the rank of \( X \). To test the overall epistasis effects is to test the general null hypothesis that no epistasis effect exists, i.e., \( H_0^g \): \( i_{aa} = 0, i_{ad} = 0, i_{da} = 0, \) and \( i_{dd} = 0 \). The alternative hypothesis is that at least one form of epistasis effect exists. It is easily seen from Eq. 5 that \( H_0^g \) is equivalent to the hypothesis \( S_g \hat{g} = 0 \), where \( S_g = (s_1, s_2, \ldots, s_9)' \). The \( F \)-test statistic is

\[ F = \frac{(n - k)(S_g \hat{g})'(S_g(X'X)^{-1}S_g \hat{g})}{4(y'y - \hat{g}X'y)} \]  

Under the null hypothesis \( H_0^g \): \( S_g \hat{g} = 0 \), the \( F \)-statistic given by Eq. 7 follows an \( F \)-distribution with degrees of freedom 4 and \( n - k \) (18). This \( F \)-test can also be used to test all the genetic effects, including single gene effects or a combination of the genetic effects by modifying Eq. 7.

Statistical power and sample size. Analytic formulas for predicting statistical power and sample size determination for a given set of statistical and genetic parameters were derived (Eqs. C5–C18 in Supplemental Material) using an approach similar to that used in Refs. 13, 14, and 15. Table 2 shows the sample size requirement for each selected level of epistasis effect assuming a 95% statistical power and a 5% type I error rate. The results show that detecting \( d \times d \) effect requires a considerably larger sample size than detecting the other epistasis effects. The reason is that the \( d \times d \) variance is small even when the size of the \( d \times d \) effect is the same as that of \( a \times a \).
SNP EPISTASIS EFFECTS

Table 2. Sample size required to achieve 95% power with a type I error of 5%

<table>
<thead>
<tr>
<th>Contrast Heritability ((a^2 = a_1^2 = a_2^2 = \ldots = a_{n}^2))</th>
<th>(a \times a)</th>
<th>(a \times d)</th>
<th>(d \times a)</th>
<th>(d \times d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_{aa}^2)</td>
<td>(H_{ad}^2)</td>
<td>(H_{dd}^2)</td>
<td>(H_{aa}^2)</td>
<td>(H_{ad}^2)</td>
</tr>
<tr>
<td>0.0013</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0005</td>
<td>11,545</td>
</tr>
<tr>
<td>0.0050</td>
<td>0.0059</td>
<td>0.0054</td>
<td>0.0020</td>
<td>2,776</td>
</tr>
<tr>
<td>0.0100</td>
<td>0.0119</td>
<td>0.0108</td>
<td>0.0040</td>
<td>1,302</td>
</tr>
<tr>
<td>0.0150</td>
<td>0.0178</td>
<td>0.0162</td>
<td>0.0060</td>
<td>810</td>
</tr>
<tr>
<td>0.0200</td>
<td>0.0238</td>
<td>0.0216</td>
<td>0.0080</td>
<td>365</td>
</tr>
<tr>
<td>0.0250</td>
<td>0.0297</td>
<td>0.0270</td>
<td>0.0100</td>
<td>417</td>
</tr>
<tr>
<td>0.0300</td>
<td>0.0357</td>
<td>0.0324</td>
<td>0.0120</td>
<td>319</td>
</tr>
<tr>
<td>0.0350</td>
<td>0.0416</td>
<td>0.0378</td>
<td>0.0140</td>
<td>249</td>
</tr>
<tr>
<td>0.0400</td>
<td>0.0476</td>
<td>0.0432</td>
<td>0.0160</td>
<td>196</td>
</tr>
<tr>
<td>0.0450</td>
<td>0.0535</td>
<td>0.0486</td>
<td>0.0180</td>
<td>155</td>
</tr>
<tr>
<td>0.0500</td>
<td>0.0594</td>
<td>0.0540</td>
<td>0.0200</td>
<td>122</td>
</tr>
</tbody>
</table>

*Indicates that all 8 effects are assumed to be of the same size in defining each heritability. \(H_{aa}^2\) values of 0.00123–0.0500 correspond to 0.10–0.6304 of \(\text{IAA}\), i.e., the size of \(\text{IAA}\) is about 0.10–0.6304 phenotypic standard deviations.

Under the assumptions of HWE, LE, and equal allele frequencies, \(\sigma_{IA}^2 = 2\sigma_{DA}^2 = 2\sigma_{DA}^2 = 4\sigma_{DD}^2\) when \(\text{IAA} = \text{DA} = \text{DD} = \text{IAA} = \text{IAA} = \text{IAA}\) based on the results of Cockerham (5) and Kempthorne (10), where \(\sigma_{IA} = a \times a\) variance, \(\sigma_{AD} = a \times d\) variance, \(\sigma_{DA} = d \times a\) variance, and \(\sigma_{DD} = d \times d\) variance. Consequently, the contrast heritability of \(\text{IAA}\), which is the fraction of the phenotypic variance accounted for by the \(d \times d\) contrast variance, is \(-0.25\) of the contrast heritability of \(\text{IAA}\). For the same contrast heritability, the sample size required is the same for detecting each of the four epistasis effects. The sample size numbers in Table 2 are merely illustrative, and the actual sample size for a given set of data and threshold values of statistical power and type I error could vary significantly. In general, statistical power is an increasing function of the effect size, sample size, and type I error, while sample size is a decreasing function of type I error, type II error, and effect size.

Evaluation of statistical power, accuracy of estimating epistasis effects, and rates of false positives, using simulations. The simulated phenotypic value of each individual is obtained as the summation of the individual quantitative trait locus (QTL) genotypic value and a random residual following a normal distribution. Each simulation generated 10,000 replicates. Statistical powers observed from the simulated data agreed well with the predicted powers (Figs. 1 and 2). When more than one type of epistasis effect exists, the joint testing of the overall epistasis effect had the highest statistical power (predicted QTL and observed QTL in Figs. 1 and 2). Among the four types of epistasis effects involving two loci, the effect with largest contrast heritability (or contrast variance) is the easiest to detect. However, predicting the epistasis effect with the largest variance or contrast variance is not as straightforward as under the assumptions of HWE and LE, where the rank of statistical power is in the order of \(a \times a\), \(a \times d\) (or \(d \times a\)), and \(d \times d\), because the \(a \times a\) effect has the largest contrast variance or heritability, i.e., \(H_{aa}^2 = 2H_{ad}^2 = 2H_{aa}^2 = 4H_{dd}^2\). Under HWD and LD, this order may not hold because of the existence of covariances between the eight genetic effects (Eq. C3 in Supplemental Materials and Ref. 9). For example, when the \(d \times d\) contrast heritability is \(-2\%\) of the phenotypic variance (corresponding to \(5\%\) \(a \times a\) contrast heritability; Table 2), the statistical power is nearly as good as for the other three epistasis effects (Fig. 2). In terms of accuracy of estimating epistasis effects measured by MSE, estimates for \(a \times a\) and \(a \times d\) (or \(d \times a\)) had similar accuracy, while estimates for \(d \times d\) had the worst accuracy (Fig. 3). Again, the poor accuracy in estimating the \(d \times d\) effect was due to the fact that \(d \times d\) had the smallest variance under the assumption of equal effect sizes for all epistasis effects. As sample size increases, all estimates tend to have similar accuracies (Fig. 3).

To evaluate the rate of false positives in the absence of any epistasis effect, the phenotypic value was defined as the random residual following \(\mathcal{N}(0,1)\) distribution, so that the phenotypic value has no genetic effect. With 10,000 repeats, the observed rate of false positives for any of the four epistasis effects was virtually identical to the cutoff type I error used as the nominal significance level. This shows that the method of testing does not have inflated type I errors over those for.
random reasons expected by the nominal significance levels. To evaluate the rate of false positives due to confounding among the four types of epistasis effects in the presence of one type of epistasis effect, the phenotypic value was defined as the sum of a genotypic value and a random residual following $N(0,1)$ distribution, where the genotypic value contains only one type of epistasis effect and is obtained by Eq. C2 in Supplemental Materials. With 10,000 repeats, the observed rate of false positives for any of the three epistasis effects not present in the phenotypic value was virtually identical to the cutoff type I error used as the nominal significance level. This demonstrated that the method of testing did not have inflated type I errors due to confounding among the eight genetic effects over those expected for random reasons, as expected by Eqs. 3.1–3.8.

Comparison with methods requiring HWE and LE. The method based on the extended Kempthorne model was compared with two methods that have been used in human epistasis testing (17): Cockerham’s model (5) under the assumptions of HWE and LE and the genetic model of Cheverud (4). Theoretical analysis showed that the Cheverud model yields identical epistasis results as the Cockerham model under HWE and LE with equal allele frequencies (Supplemental Materials). For this reason, the Cheverud model will be considered as a method requiring HWE and LE in the following discussions. Both theoretical analysis and data simulation showed that the extended Kempthorne model had identical results for testing epistasis effects under HWE and LE with equal allele frequencies as the other two methods requiring HWE and LE (Supplemental Materials). However, application of methods requiring HWE and LE to data with HWD and LD generally had three problems: decreased statistical power, decreased accuracy in parameter estimation (except for $a \times a$ effect), and confounding between epistasis effects (or increased rate of false positives). Comparison of Fig. 1 with Fig. 4 shows that the two methods requiring HWE and LE had considerably lower statistical power than our method based on the extended Kempthorne model for detecting each of the $a \times a$, $d \times a$, $d \times d$ epistasis effects when applied to data with HWD and LD. In terms of accuracy in estimating epistasis effects, the two methods requiring HWE and LE had larger MSEs (lower accuracies) than our method for $a \times d$, $d \times a$, and $d \times d$ epistasis effects, particularly when the sample size was small (Figs. 3 and 5). For detecting $a \times a$ effect, the methods requiring HWE and LE in fact had smaller MSEs. Confounding or increased rate of false positives is another problem when applying methods requiring HWE and LE to data with HWD and LD. Unlike other statistical problems such as power and accuracy problems that tend to diminish as sample size increases, the problem of confounding or false positives of the methods requiring HWE and LE becomes more serious as sample size increases (Figs. 6 and 7). The most serious confounding was that between $d \times d$ and $a \times a$ effects (Fig. 6), where $d \times d$ is the only genetic effect present in the phenotypic value but a high frequency of significant $a \times a$ effect was also observed. Confounding between $a \times d$ and $d \times a$ effects was also observed (Fig. 7) but was less serious than those in Fig. 6.

Computer implementation. A computer program, epiSNP, has been developed to implement the epistasis testing method in this article using a two-step least squares analysis similar to that of Wolfinger et al. (23). The first step corrects the phenotypes for any fixed effects (including confounding variables) using a two-step least squares analysis, and the second step performs the least squares analysis for each type of epistasis effect. The results of the first step are then used as covariates for the second step. The program epiSNP is available from the authors.

Fig. 2. Observed (dotted lines) and predicted (solid lines) statistical power as a function of $H_a^2$ using the new method for detecting epistasis effects, allowing Hardy-Weinberg and linkage disequilibria ($\alpha = 0.0034$, $n = 500$).

Fig. 3. Mean square error (MSE) of the epistasis estimations as a function of the population size for the same data as in Figs. 1 and 2 ($\alpha = 0.0034$, $H_a^2 = 0.025$).

Fig. 4. Observed statistical power as a function of the population size using methods requiring Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) for the same data as in Fig. 1 ($\alpha = 0.0034$, $H_a^2 = 0.025$). Decreased statistical power was observed for every type of epistasis effect.
Typic values to remove nongenetic systematic effects such as gender, treatment, and living conditions, and the second step conducts a significance test for SNP effects using the corrected phenotypic values as the observations. The two-step analysis estimates nongenetic systematic effect only once and offers considerable computational efficiency when the number of SNPs is large. The computer program is freely available at [http://animalgene.umn.edu](http://animalgene.umn.edu).

**DISCUSSION**

The pairwise epistasis testing method in this article can be used to assemble interactive SNPs over all autosomes and to assemble all SNP effects affecting a phenotype. The X chromosome can be included for pairwise analysis, but only female individuals can be used. This is because each female has two copies of the X chromosome, so that the analysis of the X chromosome in females is the same as that for autosomes.

Similarly, the Z chromosome in birds can be included, but only male individuals can be used in the analysis. The final SNP assembly may contain single gene effects and epistasis effects of SNPs affecting the trait. This pairwise strategy could be used to identify higher order epistasis effects that can be placed in one “epistasis network” analogous to establishing a linkage group, i.e., if locus A interacts with locus B and locus B interacts with locus C, then A, B, and C are within one epistasis network. The pairwise analysis also can be used to assemble all SNP effects as the total SNP effect on a phenotype and is more practical than a higher order analysis. We caution that pairwise analysis does not offer the exact estimate of any higher order epistasis effect. Difficulties associated with a higher order analysis include mathematical complexity and computational difficulty of the method as well as sample size requirement. As the number of loci increases, the mathematical tediousness of the methodology is expected to grow rapidly but should still be manageable for a small number of loci. Potential computation difficulty is an issue that should be noted. The total number of tests is \( n(n - 1)/2 \), which is about 125 billion for 500,000 SNPs. Testing three loci at a time will increase the number of tests to more than 20,000 trillion. This magnitude of increase in the number of tests could easily reach the point where the numerically intensive computations exceed the capabilities of the most advanced computing facilities. Another challenge of higher order analysis is in the sample size requirement. The sample size required to add each locus in the higher order epistasis analysis may need to be increased substantially to ensure that each subclass has a sufficient number of observations.

**ACKNOWLEDGMENTS**

We thank an anonymous reviewer for the thorough reading of the manuscript and thoughtful suggestions that improved the manuscript.

**GRANTS**

This research was supported in part by funding from the National Research Initiative Competitive Grants Program/United States Department of Agriculture (grant no. 03275) and the Agriculture Experimental Station of the University of Minnesota (project MN-16-043).
REFERENCES