Copyright © 2008 by the Genetics Society of America
DOI: 10.1534/genetics.107.083162

An Improved Method for Quantitative Trait Loci Detection and Identification of Within-Line Segregation in F_2 Intercross Designs

Lars Rönnegård,[1] Francois Besnier and Örjan Carlborg[2]

Linnaeus Centre for Bioinformatics, Uppsala University, 75124 Uppsala, Sweden

Manuscript received October 9, 2007
Accepted for publication February 9, 2008

ABSTRACT

We present a new flexible, simple, and powerful genome-scan method (flexible intercross analysis, FIA) for detecting quantitative trait loci (QTL) in experimental line crosses. The method is based on a pure random-effects model that simultaneously models between- and within-line QTL variation for single as well as epistatic QTL. It utilizes the score statistic and thereby facilitates computationally efficient significance testing based on empirical significance thresholds obtained by means of permutations. The properties of the method are explored using simulations and analyses of experimental data. The simulations showed that the power of FIA was as good as, or better than, Haley–Knott regression and that FIA was rather insensitive to the level of allelic fixation in the founders, especially for pedigrees with few founders. A chromosome scan was conducted for a meat quality trait in an F_2 intercross in pigs where a mutation in the halothane (Ryanodine receptor, RYR1) gene with a large effect on meat quality was known to segregate in one founder line. FIA obtained significant support for the halothane-associated QTL and identified the base generation allele with the mutated allele. A genome scan was also performed in a previously analyzed chicken F_2 intercross. In the chicken intercross analysis, four previously detected QTL were confirmed at a 5% genomewide significance level, and FIA gave strong evidence ($P < 0.01$) for two of these QTL to be segregating within the founder lines. FIA was also extended to account for epistasis and using simulations we show that the method provides good estimates of epistatic QTL variance even for segregating QTL. Extensions of FIA and its applications on other intercross populations including backcrosses, advanced intercross lines, and heterogeneous stocks are also discussed.

THE detection of quantitative trait loci (QTL) in domestic animals has been greatly enhanced by the design of experimental crosses between highly divergent lines (Andersson 2001). The animals of the founder lines are usually taken from two different breeds with large phenotypic differences, such as European Wild Boar and Large White domestic pigs (Knott *et al.* 1998), Jungle Fowl and White Leghorn chicken (Kerje *et al.* 2003), and selected mouse lines (Brockmann *et al.* 1998). Here large genetic differences between the breeds are expected for the studied trait, and the power of detecting QTL is high even for moderately sized experimental crosses. There is often a substantial within-breed heritability for the studied trait, but still the commonly used regression model to detect QTL assumes a biallelic QTL that is fixed within each of the two founder lines (Haley *et al.* 1994). If the QTL alleles are not fixed, however, the regression model will underestimate the QTL allele effect and the power to detect the QTL decreases (Perez-Enciso and Varona 2000). Furthermore, an increased understanding of the magnitude of the genetic variation within lines is crucial for further studies to identify the causative genes underlying QTL.

Several approaches have been adopted to account for within-line QTL variation in line crosses of outbred lines. Following Goddard (1992), Wang *et al.* (1998) suggested the use of a multibreed model with a fixed breed effect and a random QTL effect. For analysis of F_2 intercrosses, Knott *et al.* (1996) developed a nested within-half-sib family model that does not assume fixation of QTL alleles in the founder lines, and the number of alleles is constrained only by the number of families. This is a model with fixed effects and the number of estimated parameters increases with the number of half-sib families. Furthermore, the genotypic information of the dams is not included in the model and the sires are assumed to be unrelated.

Perez-Enciso and Varona (2000) developed a mixed-QTL model that accounts for line differences and within-line variation of QTL effects. In this model, which is similar to the model developed by Wang *et al.* (1998), a fixed line effect is estimated together with a random within-line QTL variance. The model is a combination of the regression model (Haley *et al.* 1994) and the QTL variance component (VC) model (Fernando and Grossman 1989; Goldgar 1990). A drawback of the

[2]*Present address:* Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden.

[1]*Corresponding author:* Linnaeus Centre for Bioinformatics, SE-75124 Uppsala, Sweden. E-mail: lars.ronnegard@lcb.uu.se

model is, however, the difficulty to compare estimates in different genomic locations as the total QTL variance is a combination of fixed and random effects. The method of Perez-Enciso and Varona (2000) maximizes the likelihood with a nonderivative algorithm and is rather slow. Their model was further developed in Perez-Enciso *et al.* (2001) to account for dominance effects.

Meuwissen and Goddard (2000) developed a model for fine mapping in outbred populations. They model a QTL as a random effect and introduce nonzero correlations between the founder alleles given by estimated linkage disequilibrium within founder haplotypes. We develop this approach further by, instead of estimating the founder correlations from external information about the population history, estimating these correlations directly from the marker–pedigree–phenotype data from the experimental cross.

On the basis of this model, we develop a new flexible genome-scan method for QTL detection in experimental intercrosses. The method is based on a general VC model that utilizes the score statistic to detect QTL. The VC framework relaxes the assumptions of previous fixed-effects models (*e.g.*, Haley *et al.* 1994) and ensures simple interpretation and testing of the estimated parameters. The method can be extended to account for epistasis for both segregating and fixed QTL. The use of the score statistic makes the analysis fast and robust, which enables empirical significance testing. This new method is referred to as flexible intercross analysis (FIA).

METHODS

Statistical modeling: First we present the single-locus VC model, where all base QTL alleles are assumed to be uncorrelated (Fernando and Grossman 1989; Goldgar 1990). Thereafter, we develop the single-locus model to account for within-founder-line correlations. Our model can be used to identify within-line variation in a three-stage process:

Step 1: Perform a genome scan utilizing the score statistic to test for QTL that are either fixed or segregating within founder lines.

Step 2: Estimate the within-line correlation (ρ) at significant QTL locations and test H_0: $\rho = 1$ *vs.* H_1: $\rho < 1$.

Step 3: If H_0 can be rejected, different correlations will be estimated within lines and different allele effects will be identified within lines by analyzing the least-squares estimates of the base generation allele effects.

The traditional single-locus VC model for general pedigrees with n phenotyped individuals is given by

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{v} + \mathbf{a} + \mathbf{e} \tag{1a}$$

(*e.g.*, Lynch and Walsh 1998, Chap. 16), where $\mathbf{y}$ is the vector of individual phenotypes (length n), β is the vector of fixed effects and $\mathbf{X}$ is the corresponding design matrix, $\mathbf{v}$ is a vector of random individual QTL effects (length n) in position τ, $\mathbf{a}$ is a vector of random polygenic effects (length n), and $\mathbf{e}$ is a vector of residual effects (length n). The variance–covariance matrix of $\mathbf{y}$, assuming independent allelic effects in the base generation, is:

$$V(\mathbf{y}) = \Pi\sigma^2_v + \mathbf{A}\sigma^2_a + \mathbf{I}_n\sigma^2_e, \tag{1b}$$

where Π is the genotype identity-by-descent (IBD) matrix (size $n \times n$) calculated in position τ, σ^2_v is the corresponding genotype QTL variance, $\mathbf{A}$ is the additive relationship matrix, σ^2_a is the variance of polygenic effects, $\mathbf{I}_n$ is the identity matrix of size $n \times n$, and σ^2_e is the residual variance.

An alternative, and equivalent, presentation of model (1) is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{v}^* + \mathbf{a} + \mathbf{e} \tag{2}$$

(Rönnegård and Carlborg 2007), where $\mathbf{v}^*$ is the vector of m independent normal-distributed base generation QTL alleles with variance $\frac{1}{2}\sigma^2_v$, $\mathbf{Z}$ is an incidence matrix of size $n \times m$ relating individuals with the QTL alleles in the base generation, and $\mathbf{e}$ is the residual vector with variance σ^2_e. In the calculations of the IBD matrix in model (1), it is implicitly assumed that the QTL allele effects are uncorrelated in the base generation, so $V(\mathbf{v}^*) = \frac{1}{2}\sigma^2_v\mathbf{I}_m$, where $\mathbf{I}_m$ is the identity matrix of size $m \times m$. We also have the identity $\Pi\sigma^2_v = \mathbf{Z}\mathbf{Z}'\frac{1}{2}\sigma^2_v$.

The single-locus VC QTL model for experimental line crosses: In a line cross, the QTL allele effects will be correlated within the base generation lines *A* and *B*. This correlation can be included in model (2). We then get a mixed linear model $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{v}^* + \mathbf{a} + \mathbf{e}$, where $\mathbf{v}^*$ is a random effect with m levels, where $m = m_A + m_B$ and m_A is twice the number of individuals in line *A* and m_B is twice the number of individuals in line *B*. In our model, there is an unknown correlation ρ between the first m_A levels and the last m_B levels of $\mathbf{v}^*$, such that $\mathbf{v}^* \sim \text{MVN}(0, \mathbf{G})$ and, *e.g.*, for one founder in line *A* and three founders in line *B* ($m_A = 2$ and $m_B = 6$),

$$\mathbf{G} = \frac{1}{2}\begin{pmatrix} \sigma^2_v & \sigma^2_c & & & & & & \\ \sigma^2_c & \sigma^2_v & & & & & & \\ & & \sigma^2_v & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_c \\ & & \sigma^2_c & \sigma^2_v & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_c \\ & & \sigma^2_c & \sigma^2_c & \sigma^2_v & \sigma^2_c & \sigma^2_c & \sigma^2_c \\ & & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_v & \sigma^2_c & \sigma^2_c \\ & & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_v & \sigma^2_c \\ & & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_c & \sigma^2_v \end{pmatrix} \tag{3a}$$

or in alternative notation

$$\mathbf{G} = \frac{1}{2}\sigma_v^2 \begin{pmatrix} 1 & \rho & & & & & & \\ \rho & 1 & & & & & & \\ & & 1 & \rho & \rho & \rho & \rho & \rho \\ & & \rho & 1 & \rho & \rho & \rho & \rho \\ & & \rho & \rho & 1 & \rho & \rho & \rho \\ & & \rho & \rho & \rho & 1 & \rho & \rho \\ & & \rho & \rho & \rho & \rho & 1 & \rho \\ & & \rho & \rho & \rho & \rho & \rho & 1 \end{pmatrix}. \quad (3b)$$

Hence, $\mathbf{y} \sim \text{MVN}(\mathbf{X}\beta, \mathbf{V})$, where

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{A}\sigma_a^2 + \mathbf{I}_n\sigma_e^2. \quad (4a)$$

Using the gametic IBD matrix formulation (*e.g.*, Pong-Wong *et al.* 2001), we have a gametic IBD matrix $\mathbf{W}$ for the base individuals, with $\frac{1}{2}\sigma_v^2\mathbf{W} = \mathbf{G}$ and $\mathbf{V} = \frac{1}{2}\sigma_v^2\mathbf{KWK}' + \mathbf{A}\sigma_a^2 + \mathbf{I}_n\sigma_e^2$, where $\mathbf{K} = \mathbf{I}_n \otimes (1, 1)$ ($\otimes$ denotes the Kronecker product).

The key to estimating σ_v^2 and σ_c^2 (and thereby also ρ) is to formulate the variance–covariance matrix of $\mathbf{y}$ as

$$\mathbf{V} = \Pi\sigma_v^2 + (\Pi_\mathbf{J} - \Pi)\sigma_c^2 + \mathbf{A}\sigma_a^2 + \mathbf{I}_n\sigma_e^2, \quad (4b)$$

where Π and $\Pi_\mathbf{J}$ are the IBD matrices calculated for base-generation structures with alleles independent and fixed within lines, respectively. From (2), it is possible to show that (4a) and (4b) are equivalent since

$$\begin{aligned} \mathbf{ZGZ}' &= \mathbf{Z}\left(\mathbf{I}_m\frac{1}{2}\sigma_v^2 + (\mathbf{J}_{m_A,m_B} - \mathbf{I}_m)\frac{1}{2}\sigma_c^2\right)\mathbf{Z}' \\ &= \mathbf{ZI}_m\mathbf{Z}'\frac{1}{2}\sigma_v^2 + (\mathbf{ZJ}_{m_A,m_B}\mathbf{Z}' - \mathbf{ZI}_m\mathbf{Z}')\frac{1}{2}\sigma_c^2 \\ &= \mathbf{ZZ}'\frac{1}{2}\sigma_v^2 + (\mathbf{ZJ}_{m_A,m_B}\mathbf{Z}' - \mathbf{ZZ}')\frac{1}{2}\sigma_c^2 \\ &= \Pi\sigma_v^2 + (\Pi_\mathbf{J} - \Pi)\sigma_c^2, \end{aligned}$$

Where $\mathbf{J}_{m_A,m_B}$ is a block diagonal matrix with ones in blocks of size $m_A \times m_A$ and $m_B \times m_B$.

Epistatic VC QTL model: Including epistasis is an essential part of QTL modeling (Carlborg and Haley 2004). FIA accounts for epistasis by an extension of model (4b) as

$$\mathbf{V} = \Pi_1\sigma_v^2 + (\Pi_\mathbf{J} - \Pi_1)\sigma_c^2 + \Pi_1 \circ \Pi_2\sigma_{12}^2 + \mathbf{I}_n\sigma_e^2, \quad (5)$$

where Π_1 and Π_2 are the IBD matrices for the two QTL, assuming an outbred base generation, and the Hadamard product $\circ$ was used for the epistatic variance component (Mitchell *et al.* 1997; Rönnegård *et al.* 2008).

Genome scan with the score statistic: The score statistic is calculated as $S = \mathbf{D}'\mathbf{F}^{-1}\mathbf{D}$ (Cox and Hinkley 1974; Tang and Siegmund 2001; Putter *et al.* 2002), where $\mathbf{D}$ is the gradient vector for the log-likelihood function under the alternative hypothesis calculated at the parameter values given by the null hypothesis, and $\mathbf{F}$ is Fisher's information matrix (see Lynch and Walsh 1998).

We base our analysis on the residual log-likelihood function (Patterson and Thompson 1971):

$$\begin{aligned} L = &-\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &-\frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|. \end{aligned}$$

Let $\theta = \begin{pmatrix}\sigma_v^2 & \sigma_c^2 & \sigma_a^2 & \sigma_e^2\end{pmatrix}$ be the parameter vector for the VC part of the model. We then have the following partial derivatives:

$$\begin{aligned} \frac{\partial\mathbf{V}}{\partial\theta_1} &= \frac{\partial\mathbf{V}}{\partial\sigma_v^2} = \Pi_\mathbf{I} \\ \frac{\partial\mathbf{V}}{\partial\theta_2} &= \frac{\partial\mathbf{V}}{\partial\sigma_c^2} = \Pi_\mathbf{J} - \Pi_\mathbf{I} \\ \frac{\partial\mathbf{V}}{\partial\theta_3} &= \frac{\partial\mathbf{V}}{\partial\sigma_a^2} = \mathbf{A} \\ \frac{\partial\mathbf{V}}{\partial\theta_4} &= \frac{\partial\mathbf{V}}{\partial\sigma_e^2} = \mathbf{I}_n. \end{aligned}$$

The elements of the gradient $\mathbf{D}$ under the null hypothesis of $\sigma_v^2 = 0$ were calculated as

$$\frac{\partial L}{\partial\theta_i} = -\frac{1}{2}\text{tr}\left(\frac{\partial\mathbf{V}}{\partial\theta_i}\mathbf{P}_0\right) + \frac{1}{2}\mathbf{y}'\mathbf{P}_0\frac{\partial\mathbf{V}}{\partial\theta_i}\mathbf{P}_0\mathbf{y}$$

(Lynch and Walsh 1998), where $\mathbf{P}_0 = \mathbf{V}_0^{-1} - \mathbf{V}_0^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_0^{-1}$, and $\mathbf{V}_0 = \mathbf{A}\tilde{\sigma}_a^2 + \mathbf{I}_n\tilde{\sigma}_e^2$ with $\tilde{\sigma}_a^2$ and $\tilde{\sigma}_e^2$ being the estimated polygenic and residual variance under the null hypothesis. The elements of $\mathbf{F}$ are calculated as

$$\mathbf{F}_{ij} = \frac{1}{2}\text{tr}\left(\frac{\partial\mathbf{V}}{\partial\theta_i}\mathbf{P}_0\frac{\partial\mathbf{V}}{\partial\theta_j}\mathbf{P}_0\right).$$

Note that the value of $\mathbf{P}_0$ does not change between positions along the genome and that the computational requirements for the score statistic are low.

Parameter estimation: The parameters in the model are estimated using average information–restricted maximum likelihood (AI–REML) (Johnson and Thompson 1995). The algorithm was programmed in R (R Development Core Team 2004) and set to converge when the log likelihood differed $<10^{-4}$ between iterations. The derivatives of the log-likelihood function used in the algorithm are

$$\frac{\partial L}{\partial\theta_i} = -\frac{1}{2}\text{tr}\left(\frac{\partial\mathbf{V}}{\partial\theta_i}\mathbf{P}\right) + \frac{1}{2}\mathbf{y}'\mathbf{P}\frac{\partial\mathbf{V}}{\partial\theta_i}\mathbf{P}\mathbf{y},$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$ with $\mathbf{V}$ given in (4b).

This estimation enables testing of whether the QTL allele effects are identical within founder lines or not by testing H_0: $\sigma_v^2 = \sigma_c^2$ *vs.* H_1: $\sigma_v^2 > \sigma_c^2$ using a likelihood-ratio (LR) test.

Identification of segregating alleles within founder lines: When a QTL with $\rho < 1$ is detected, we test within which line and individual that the alleles are segregating. First, we estimate separate correlations within each line, so that [following the example in (3)] we have

$$\mathbf{G} = \frac{1}{2}\sigma_v^2 \begin{pmatrix} 1 & \rho_A & & & & & & \\ \rho_A & 1 & & & & & & \\ & & 1 & \rho_B & \rho_B & \rho_B & \rho_B & \rho_B \\ & & \rho_B & 1 & \rho_B & \rho_B & \rho_B & \rho_B \\ & & \rho_B & \rho_B & 1 & \rho_B & \rho_B & \rho_B \\ & & \rho_B & \rho_B & \rho_B & 1 & \rho_B & \rho_B \\ & & \rho_B & \rho_B & \rho_B & \rho_B & 1 & \rho_B \\ & & \rho_B & \rho_B & \rho_B & \rho_B & \rho_B & 1 \end{pmatrix},$$

where ρ_A is the correlation within founder line *A* and ρ_B is the correlation within founder line *B*. Furthermore, the individual base generation allele effects can be estimated. The least-squares estimates are obtained by constructing a design matrix $\mathbf{X}_m$ from the first m columns of the gametic IBD matrix **W**. The base generation allele effects β_m are then estimated from the linear model $\mathbf{y} = \mathbf{X}\beta + \mathbf{X}_m\beta_m + \mathbf{e}$.

Calculation of significance thresholds: The significance thresholds for the genome scan are calculated by means of permutation testing (Churchill and Doerge 1994; Good 2005). An important assumption of a permutation test is that the observations are exchangeable under the null hypothesis (Davison and Hinkley 1997, p. 143; Anderson and Ter Braak 2002) and we guarantee this by permuting residuals estimated from the null model $\mathbf{y} = \mathbf{X}\beta + \mathbf{a} + \mathbf{e}$. Replicates of the phenotypic data were simulated with $\mathbf{y} = \mathbf{X}\tilde{\beta} + \tilde{\mathbf{a}} + \check{\mathbf{e}}$, where $\check{\mathbf{e}}$ is the vector of permuted residuals, and $\tilde{\beta}$ and $\tilde{\mathbf{a}}$ are the estimated fixed and random polygenic effects obtained from the null model. (In our analyses of the European Wild Boar × Large White F_2 cross and the Red Jungle Fowl × White Leghorn F_2 cross no polygenic effects were detected and we therefore used $\tilde{\mathbf{a}} = 0$). For each replicate, the score statistic was calculated at every tested position (5 cM apart) along the genome using (4). The empirical distribution of the maximum score value from each replicate was used to obtain significance thresholds.

IBD-matrix estimation: Several different IBD estimation programs are available. We use a deterministic algorithm based on Pong-Wong *et al.* (2001) to calculate the IBD matrices in (4b). The original algorithm by Pong-Wong *et al.* assumes that founder marker phases are unknown and does not enable general base generation structures. We implemented the original algorithm to account for known phases of marker genotypes and consequently allow user-specified base generation structures. The most likely phases of the marker genotypes in the base generation are estimated using a genetic algorithm that uses both marker and pedigree information as input. Resulting marker genotype phases were subsequently used as input in the IBD estimation algorithm (F. Besnier, B.-W. Kim and Ö. Carlbourg, unpublished data).

Simulation setup for a single QTL: We used simulations to study the distribution of ρ and variation of LR for different levels of fixation within lines. In the results, if not otherwise stated, $\mathrm{LR} = -2(l_0 - l_1)$ with l_1 equaling the likelihood of model (4) and l_0 equaling the likelihood of the same model with $\sigma_v^2 = 0$. Furthermore, we wished to compare the power of the Haley–Knott regression with additive effects (Haley *et al.* 1994) to our more flexible method.

In the analyses of VC estimates and comparisons of LRs to the ones obtained from Haley–Knott regression, 100 replicates were simulated for each of the four cases described below with four founders and 800 F_2 individuals, and a 20% QTL at a fully informative marker. The setup for these simulations gives expected values of $\rho =$ 1.00, 0.68, 0.20, and 0.00 for cases 1–4, respectively.

In the power analyses, four parameters were varied to thoroughly evaluate the differences between FIA and Haley–Knott regression: number of founders in base generation, number of F_2 individuals, marker information, and level of fixation within founder lines. The methods were compared by their power to detect a QTL for a given position at a 5% significance level.

Base generation size: Two types of pedigrees were simulated with random mating: one with a small base generation (no. of founders = 4) and one with a large base generation (no. of founders = 50). These are referred to as *small base* and *large base*, respectively. The structure for the small base pedigree was designed to mimic the pedigree of a Red Jungle Fowl × White Leghorn F_2 cross (Kerje *et al.* 2003) with one jungle fowl male mated to three leghorn females in the base generation. The large base generation structure pedigree consisted of 25 individuals in each line and resembles an F_2 intercross between two divergently selected bodyweight chicken lines described in Jacobsson *et al.* (2005). The two simulated base generation sizes were chosen as most experimental intercrosses are expected to have base generation sizes intermediate to these.

Number of F_2 individuals: The number of F_2 individuals simulated was either 200 or 800. A biallelic QTL was simulated with a difference between the two QTL allele effects of $\sqrt{2h^2}$, where h^2 is the QTL heritability in an outbred population. For the pedigrees with 200 F_2 individuals, h^2 was set to 0.05, whereas h^2 was set to 0.02 for the pedigrees with 800 F_2 individuals. The phenotype of an F_2 individual i was simulated with $y_i = A_{1_i} + A_{2_i} + e_i$, where A_{1_i} is the QTL allele effect on the paternally inherited chromosome and A_{2_i} is the QTL allele effect on the maternally inherited chromosome, and e_i is an i.i.d. normally distributed residual effect with a variance of $1 - h^2$.

Marker information: A QTL was simulated either at a fully informative marker or in the middle of a 40-cM

TABLE 1

Proportion of *A* and *B* QTL alleles in the two founder lines of a line cross for the four simulated scenarios ranging from a fixed QTL (case 1) to a completely segregating QTL (case 4)

		Case 1[a]	Case 2[a]	Case 3[b]	Case 4[a]
Line *A*	*A* alleles	1	1	3/4	1/2
	B alleles	0	0	1/4	1/2
Line *B*	*A* alleles	0	1/6	1/4	1/2
	B alleles	1	5/6	3/4	1/2

[a] For cases 1, 2, and 4, the small base pedigree consisted of one founder in line *A* and three founders in line *B*, which was the base generation structure in the Red Jungle Fowl × White Leghorn F_2 intercross reported by Kerje *et al.* (2003).

[b] For case 3, the small base pedigree consisted of two founders in line *A* and two founders in line *B*.

marker interval flanked by two fully informative markers.

Level of fixation within founder lines: Four different cases (Table 1) were studied by varying the fixation level within lines of *A* and *B* alleles for a biallelic QTL. The theoretical values of ρ for each of the four cases are derived in the appendix.

Simulation setup for a QTL with epistatic interaction: Using simulations, we evaluated how well our model estimates epistatic variance components and how the existence of epistasis influences the estimates of other variance components in the presence of QTL segregation. A small base generation and 800 F_2 individuals were simulated for cases 1–4 (Table 1). A biallelic QTL was simulated with a difference between the two QTL allele effects of 3.162. An additional unlinked QTL was simulated with no main effect. The base alleles of this QTL were assumed to be unique and to be interacting with the former biallelic QTL. For every possible pairwise combination of alleles between the two QTL a value for the epistatic interaction was drawn from $N(0, \sigma^2_{12})$, where σ^2_{12} is the epistatic variance. The residual variance was $95 - \sigma^2_{12}$, which gives a heritability for the main QTL effect of 5% in an outbred population. One hundred replicates were simulated for each of the four cases and for $\sigma^2_{12} = 0$, 5, or 10, and a fully informative marker was assumed.

Analyses of experimental data: *European Wild Boar × Large White F_2 cross:* To compare the properties of FIA to Haley–Knott regression when applied to real data, we analyzed data from a European Wild Boar × Large White cross, where the causal mutation underlying the studied trait has been detected (Lundström *et al.* 1995) and is known to be segregating in the founders. In this cross, two European Wild Boars were mated to eight Large White sows, producing 191 F_2 offspring with measured genotypes and phenotypes. Twenty-two markers were genotyped on the studied chromosome 6 at 0.0, 8.6, 36.6, 49.7, 50.5, 62.9, 79.2, 80.4, 83.7, 84.1, 84.8, 90.6, 95.4, 100.7, 101.9, 115.9, 116.7, 119.0, 120.2, 124.0, 127.0, and 170.9 cM. In our analysis, we examined a meat quality trait (reflectance value, EEL, scored over a cross-section of the longissimus dorsi muscle), which is known to be affected by the halothane gene (Lundström *et al.* 1995) located at 80.4 cM. One of the founder boars is known to be heterozygous (Hal^N/Hal^n), whereas all other founders are homozygous for the wild-type allele (Hal^N/Hal^N). We performed a QTL scan on chromosome 6 using Haley–Knott regression and FIA and also explored whether FIA could identify which founder allele carried the mutated halothane allele (Hal^n). Following Knott *et al.* (1998), we included sex, litter, and slaughter weight as fixed effects in our analysis. The data are described in detail in Knott *et al.* (1998) and Andersson-Eklund *et al.* (1998).

Red Jungle Fowl × White Leghorn F_2 cross: In a Red Jungle Fowl × White Leghorn F_2 cross, we performed a full-genome scan using FIA and Haley–Knott regression. In this pedigree, one Red Jungle Fowl male was mated to three White Leghorn females, producing 756 F_2 offspring with measured genotypes and phenotypes. We used an updated marker map to those reported in Kerje *et al.* (2003), including 439 markers (L. Andersson, personal communication) covering chromosomes 1–28. We analyzed body weight at 200 days of age for which Kerje *et al.* (2003) found two 1% genomewide significant QTL on chromosome 1 (growth 1 at 68 cM and growth 2 at 420 cM) and 5% genomewide significant QTL on chromosome 5 (growth 8 at 21 cM) and chromosome 27 (growth 13 at 20 cM). Two additional 5% genomewide significant QTL were also found in this study, which had large and significant dominance effects. In the original publication, it was noted that there was a within-line QTL variation for growth 2 on the basis of indications from a heterogeneity test among the four largest F_1 families, but it was not investigated how this influenced the results or within which founder line the QTL was segregating. We repeated the genome scan to explore how many of the previously detected QTL were segregating. Following Kerje *et al.* (2003), we included sex and batch as fixed effects in the model. The data are described in detail in Kerje *et al.* (2003).

RESULTS

Simulations of a single QTL: *Estimates of within-line correlations and comparisons to LRs from Haley–Knott regression:* We compared the LRs from our VC model using REML with those obtained from Haley–Knott regression. LRs from Haley–Knott regression decrease rapidly when the level of segregation within lines increases (Table 2), whereas LRs from our model differed only marginally between the evaluated cases. The estimated within-line correlation (ρ) drops from one to zero as the level of fixation decreases in cases 1–4 (Table 2).

TABLE 2

Likelihood ratios from Haley–Knott regression with additive effects and FIA together with the estimated within-line correlation (ρ) for four simulated scenarios

Simulated cases[a]	Simulated ρ	FIA		Haley–Knott likelihood ratio
		Estimated ρ	Likelihood ratio	
Case 1	1.00	0.99 (0.001)[b]	168.1 (2.5)[b]	174.3 (2.5)[b]
Case 2	0.68	0.71 (0.009)	160.6 (2.1)	130.5 (2.3)
Case 3	0.20	0.21 (0.011)	150.4 (2.4)	44.3 (1.9)
Case 4	0.00	0.0004 (10^{-5})	152.9 (2.5)	2.3 (0.5)

[a] The simulated cases 1–4 cover scenarios from total fixation to complete segregation within lines. One hundred replicates of an F_2 pedigree with four founders and 800 F_2 individuals with a strong QTL effect having a 20% heritability were simulated.

[b] Standard errors are within parentheses.

Power analysis: There were no substantial differences in power between FIA and Haley–Knott regression when the QTL were fixed within lines (case 1 in Figure 1). The power of FIA was up to 10 times higher, however, when there was segregation within lines (Figure 1, case 4). For case 4, the difference between Haley–Knott regression and FIA was largest when the base generation was small and the number of F_2 individuals was high (Figure 1a), *i.e.*, when there were a large number of copies of each base generation allele in the F_2. The difference in power was highest when the QTL was located at the marker.

With a large base generation, the power of both Haley–Knott regression and FIA decreases significantly. For example, in case 4, Haley–Knott regression had no power at all to detect the segregating QTL (detection equal to the type I error rate of 5%). FIA, however, still had some power (18%) to detect the segregating QTL near a fully informative marker in the 800 F_2 pedigree. Decreasing the number of copies of each base generation allele in the F_2 increases the uncertainty of the estimated base generation structure, especially when the base is fully outbred or close to outbred. Therefore a decline in the power of FIA going from case 1 to 4, as shown in Figure 1, c and d, is expected.

Simulations of a QTL with epistatic interaction: Both the main and the epistatic QTL variance were estimated satisfactorily with FIA (Table 3), although the average epistatic variance tended to be slightly overestimated for $\sigma^2_{12} = 0$ and $\sigma^2_{12} = 5$ because the distribution of estimates was highly skewed. Most estimates were, however, very close to the true value with the medians for $\sigma^2_{12} = 0$ being <0.02. The estimated residual variances were all close to the simulated ones with no bias detected in any of the simulated cases (results not shown). Thus, FIA gives accurate estimates of the QTL variances also when there are substantial epistatic effects. The estimated within-line correlation was, however, confounded with the epistatic variance, giving an increased $\hat{\rho}$ for cases 2 and 3 when $\sigma^2_{12} = 5$ and $\sigma^2_{12} = 10$. We may therefore expect that the estimated degree of segregation within lines is conservative in the presence of epistasis.

Analyses of experimental data: *European wild boar × Large White F_2 cross:* In the scan of chromosome 6 for the meat quality trait EEL (reflectance score), FIA gave a maximum score value of 135.9 (LR = 33.3 from REML) and the Haley–Knott regression resulted in LR = 3.5. Both maxima were located at 80 cM (Figure 2). The LR for a segregating QTL *vs.* a fixed QTL was 26.0 ($P < 0.01$) with an estimated within-line correlation of $\hat{\rho} = 0.0$. The 1% chromosomewise significance threshold was obtained using 10,000 permutations. The estimates for different within-line correlations were $\hat{\rho}_A = 0.0$ and $\hat{\rho}_B = 0.79$, thus indicating that the QTL is segregating within the wild boars but not within the domestic pigs. The estimates of the allelic effects clearly show that only one of the wild boar alleles (allele no. 3) carries the halothane mutation (Figure 3).

Red Jungle Fowl × White Leghorn F_2 intercross: Six genomewide significant QTL were detected with both FIA and Haley–Knott regression (Table 4). FIA showed that at least two of the six QTL segregate within the founder lines. The segregating QTL are located on chromosomes 1 and 27. The REML estimates (chromosome 1, 488 cM: $\hat{\rho}_A = 0.87$, $\hat{\rho}_B = 0.0$; chromosome 27, 21 cM: $\hat{\rho}_A = 1.0$, $\hat{\rho}_B = 0.0$) and the estimated allele effects (Figures 4 and 5) indicate that both QTL are segregating among the White Leghorn base individuals but not within the Red Jungle Fowl male. Moreover, for both of the two segregating QTL, it was possible to separate the allele effects into three distinct groups (Figures 4 and 5), indicating that these QTL are triallelic.

An additional QTL on chromosome 5 was segregating according to the test for fixation ($P = 0.01$, Table 4), but the REML estimates for the model with different within-line correlations did not converge. The estimates of the base allele effects (Figure 6) were significantly different within the Jungle Fowl founder, whereas the differences within the White Leghorn founders were relatively small.

The QTL heritabilities (calculated as $\sigma^2_v/\tilde{\sigma}^2_e$, with $\tilde{\sigma}^2_e = 29614.4$ estimated as the residual variance without any

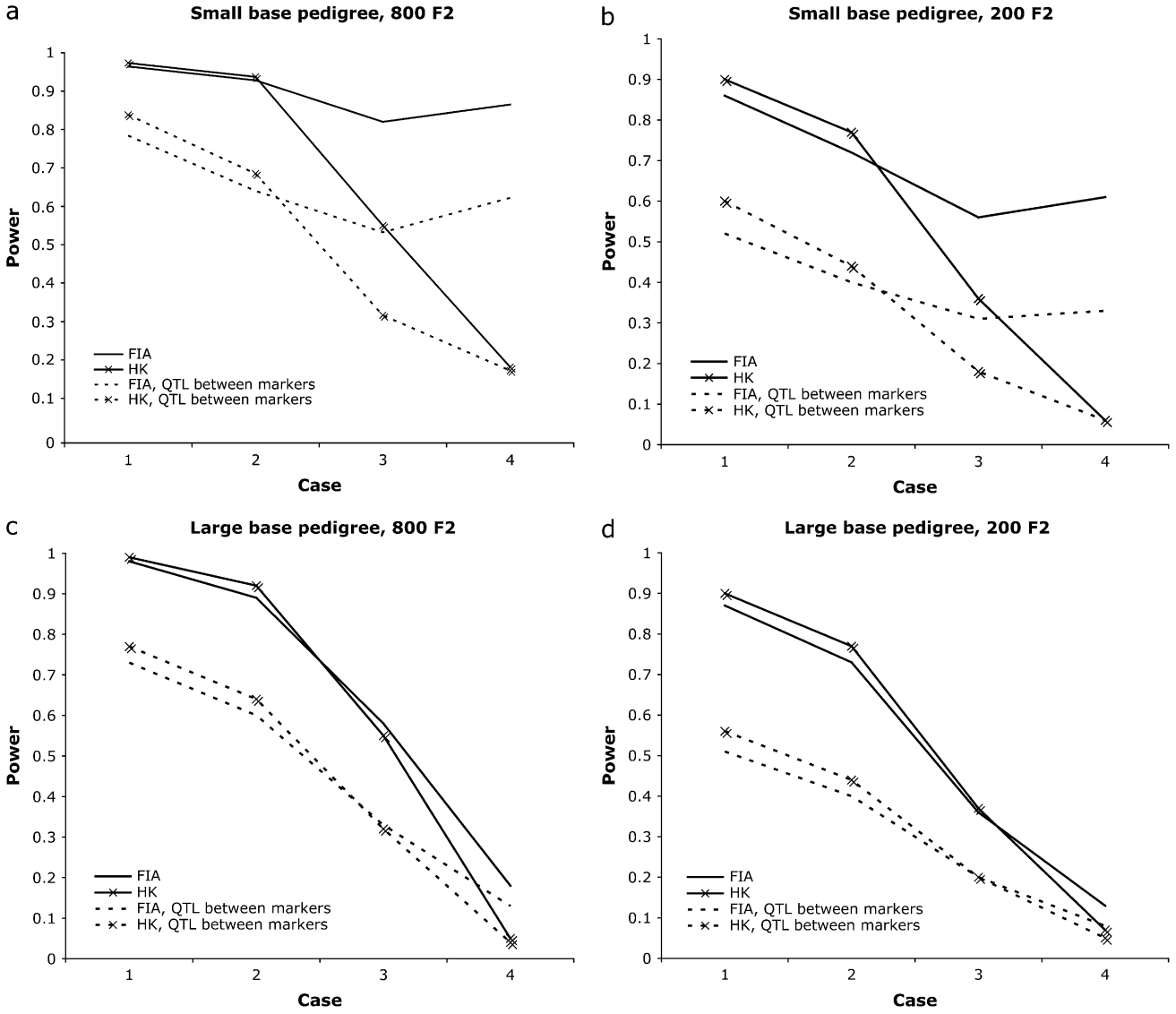

FIGURE 1.—Power to detect a QTL with Haley–Knott regression (HK) and FIA for the four simulated cases presented in Table 1, ranging from total fixation (case 1) to complete segregation within lines (case 4). Four different types of F_2 pedigrees were simulated: (a) small base (4 founders) and 800 F_2 individuals, (b) small base and 200 F_2 individuals, (c) large base (50 founders) and 800 F_2 individuals, and (d) large base and 200 F_2 individuals. For the pedigrees with 800 F_2 individuals, 1000 replicates were simulated with a 2% QTL heritability. For the pedigrees with 200 F_2 individuals, 5000 replicates were simulated with a 5% QTL heritability. The QTL was simulated either at a fully informative marker (solid lines) or between two fully informative markers 40 cM apart (dashed lines).

random effect in the model) were between 0.04 and 0.10. The 5% genomewide significance threshold was calculated from 1000 replicates of permuted data.

DISCUSSION

We have developed a new, general and powerful method for QTL detection in line-cross experiments, which does not rely on the assumption of fixation within founder lines. The method is based on VC theory and includes a parameter of within-line correlation to model segregation of QTL within the founder lines. It uses the score statistic for significance testing, which enables fast and robust genome scans based on empirical significance thresholds. The new method, FIA, is powerful in detecting fixed and segregating QTL and provides good estimates of epistasis among QTL. Our simulations showed that the reduction in power, compared to Haley–Knott regression, was only marginal for fixed QTL, whereas the gain in power was substantial for segregating QTL. Since the power to detect fixed QTL is high in F_2 line crosses, the advantage of getting a major increase in the power to detect segregating QTL supersedes the minor reduction in the power to detect fixed QTL, and we

TABLE 3

Estimated within-line correlation (ρ) and variance components (variance of main QTL effect σ^2_v, epistatic variance σ^2_{12}) for four simulated scenarios

	Simulated values		$\sigma^2_{12}=0$			$\sigma^2_{12}=5$			$\sigma^2_{12}=10$		
	ρ	σ^2_v	$\hat{\rho}$	$\hat{\sigma}^2_v$	$\hat{\sigma}^2_{12}$	$\hat{\rho}$	$\hat{\sigma}^2_v$	$\hat{\sigma}^2_{12}$	$\hat{\rho}$	$\hat{\sigma}^2_v$	$\hat{\sigma}^2_{12}$
Case 1	1.00	10.0	0.98	10.63	0.40*	1.00	10.86	5.45	1.00	*13.73*	9.71
Case 2	0.68	8.80	0.66	8.74	0.38*	0.80	8.38	*5.80*	0.88	8.94	10.63
Case 3	0.20	6.25	0.21	6.29	0.63*	0.34	5.78	*5.87*	0.46	5.15	*11.35*
Case 4	0.00	4.44	0.00	4.62	0.44*	0.02	3.70	*5.83*	0.02	4.02	10.28

The simulated cases 1–4 cover scenarios from total fixation to complete segregation within lines. One hundred replicates of an F_2 pedigree with four founders and 800 F_2 individuals with a QTL effect having a 5% heritability in an outbred base population were simulated. Bias with $P<0.05$ is in italics, $P<0.01$ is underlined, and * indicates biased estimates ($P<0.01$) but with median value <0.02.

therefore recommend FIA as a standard QTL method in F_2 intercross designs.

A substantial gain in power, compared to Haley–Knott regression, was also shown in real data (Figure 2) from a European Wild Boar × Large White F_2 cross, where the functional halothane gene affecting meat quality as well as the mutated base generation allele could be identified using FIA but not with Haley–Knott regression. In the analyses of body weight in the Red Jungle Fowl × White Leghorn F_2 cross, there was a tendency for the estimated within-line correlation to decrease with the level of the QTL effect (Table 4). The reason for this decrease is not known, but it might be due to chance, since the uncertainty of $\hat{\rho}$ increases as the QTL effect decreases, or has an underlying genetic explanation as it is more likely that QTL with large effects have been fixed during selection for increased body weight.

The method was given for an F_2 pedigree, but the extension to backcrosses, deeper pedigrees (*e.g.*, advanced intercross lines; Darvasi and Soller 1995), and several founder lines (*e.g.*, heterogeneous stocks; Talbot *et al.* 1999) is relatively straightforward since it is a VC-based method. The only principal modification needed is to construct suitable IBD matrices and, in the case of

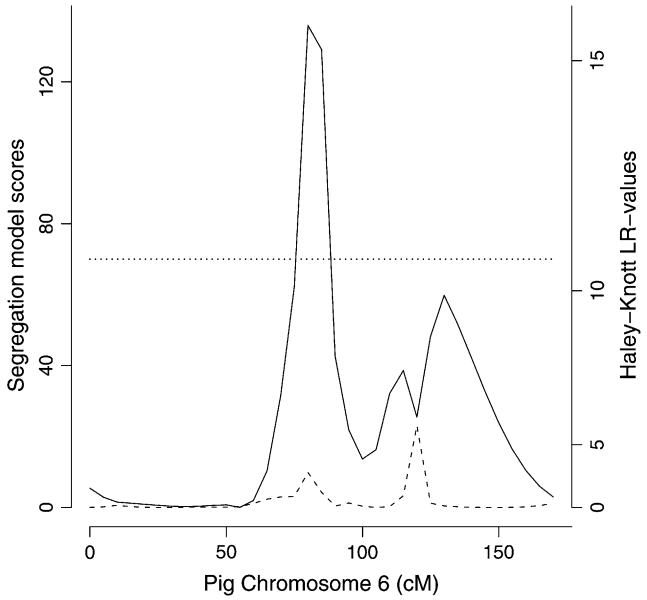

Figure 2.—Scan of pig chromosome 6 in a Wild Boar × Large White intercross for the reflectance (EEL) score. The curves given are the score statistic from the FIA model (solid line) and the Haley–Knott LR values (dashed line). The scale of the right-hand axis has been transformed such that the heights of the two curves correspond to approximately the same significance level. The dotted line is the 1% chromosomewise significance level.

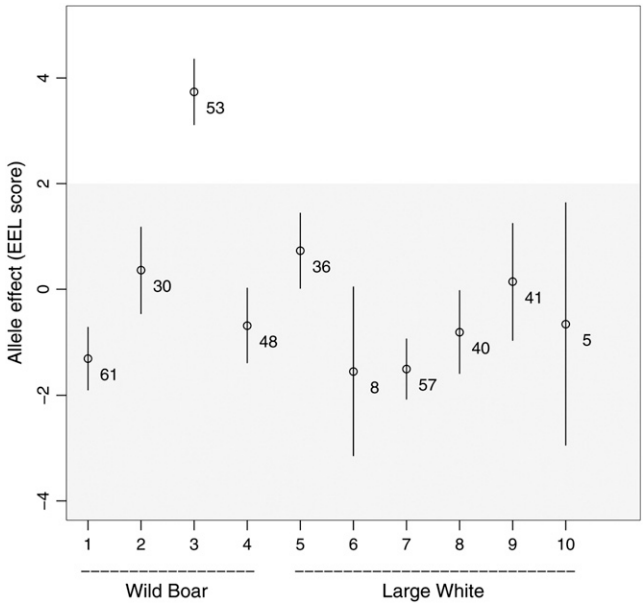

Figure 3.—Estimated base generation allele effects (±1 SE) of the reflectance (EEL) score from the Wild Boar × Large White intercross. Eight alleles from the domestic line were not transmitted to the F_2 and only 10 of the total 20 base generation alleles had two or more copies in the F_2 generation. The numbers indicate the expected number of copies in the F_2 generation. Estimated effects for alleles with two or less copies had large standard errors and are not included. Allele no. 3 was known to be the only carrier of the halothane mutation. The *Hal*^{N}-type alleles are shown in the shaded region and the *Hal*^{n}-type is shown in the open region.

TABLE 4

Genomewide significant QTL for body weight at 200 days of age using Haley–Knott regression with additive effects and FIA, and estimated levels of fixation within founder lines, from an F_2 Red Jungle Fowl × White Leghorn intercross

Haley–Knott				FIA						
Chr.	Pos. (cM)	Locus name[a]	LR: H_0: no QTL	Chr.	Pos. (cM)	Score	LR: H_0: no QTL	σ_v^2	ρ	*P*-value:[b] H_0: fixed QTL
1	102	Growth 1	249.6[c]	1	102	22,312[c]	235.0	1,755	1.00	1.00
1	488	Growth 2	69.2[c]	1	488	2,222[c]	73.0	2,448	0.17	<0.01
5	32	Growth 8	14.9[d]	5	32	99.6[d]	18.0	2,903	0.86	0.01
6	30	*New*	13.6[d]	6	30	85.8[d]	11.4	1,300	0.80	0.10
27	20	Growth 13	12.8[d]	27	21	89.1[d]	23.6	2,919	0.00	<0.01
28	35	*New*	13.8[d]	28	35	101.6[d]	12.4	1,219	0.00	0.05

Chr., chromosome; Pos., position.

[a] Locus name as assigned in Kerje *et al.* (2003). The QTL on chromosomes 6 and 28 were not detected in Kerje *et al.* (2003) where a less dense marker map was used.

[b] Calculated from the likelihood-ratio test statistic $\frac{1}{2}\chi_0^2 : \frac{1}{2}\chi_1^2$ distributed under the null hypothesis (Self and Liang 1987).

[c] 1% genomewide significance.

[d] 5% genomewide significance.

several founder lines, allow for several within-line correlations. An analysis of heterogeneous stocks, for instance, could enable estimation of correlations between founder line QTL effects and thereby improve QTL detection. Previous studies indicate that similarities between the founder lines vary between genome positions (Talbot *et al.* 1999). We might therefore expect to get different correlation estimates at different QTL positions. An analysis of this kind should improve our understanding of how the QTL were generated in the founder lines.

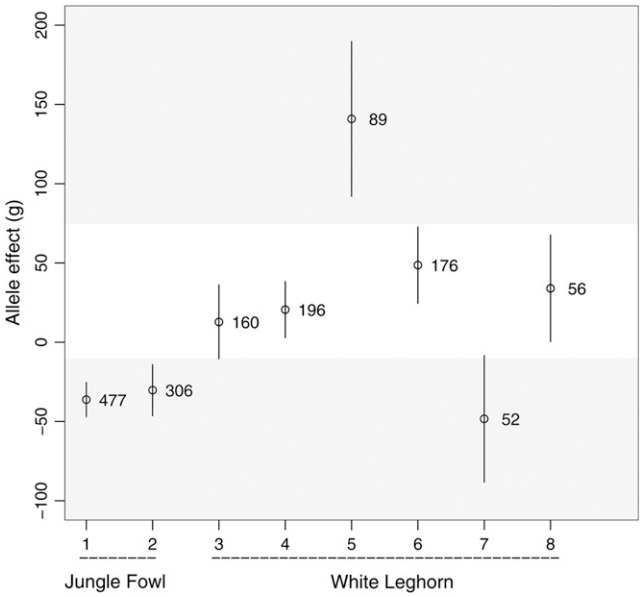

Figure 4.—Estimated base generation allele effects for growth 2 (Table 4) on chromosome 1 (±1 SE) in the Red Jungle Fowl × White Leghorn F_2 intercross. The studied trait was body weight (grams) at 200 days of age. The numbers indicate the expected number of copies of the founder alleles in the F_2 generation. The shaded regions show the clustering of the suggested triallelic QTL.

Following the modeling of Meuwissen and Goddard (2000), we included both the between- and the within-line effects as random. There are several advantages of using a pure random QTL-effects model. Including a potential QTL as a fixed effect in a linear model does not account for the extra sampling variation due to uncertain transmission of alleles through the pedigree (Xu 1998a; Feenstra *et al.* 2006), whereas a VC approach does

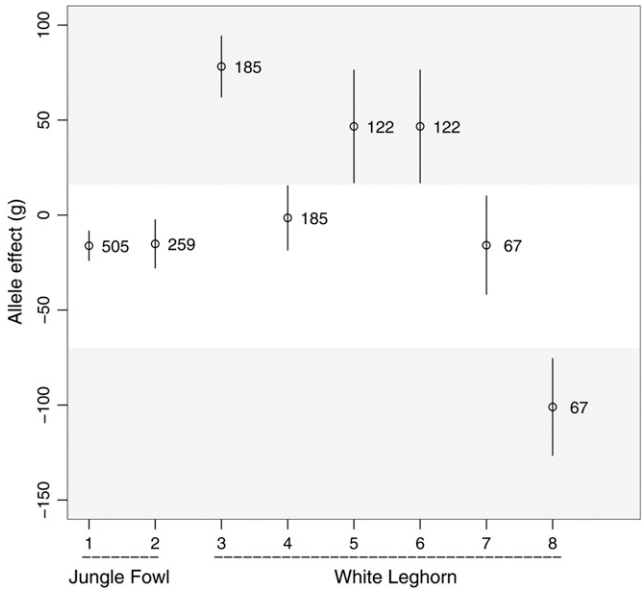

Figure 5.—Estimated base generation allele effects for growth 13 (Table 4) on chromosome 27 (±1 SE) in the Red Jungle Fowl × White Leghorn F_2 intercross. The studied trait was body weight (grams) at 200 days of age. The numbers indicate the expected number of copies of the founder alleles in the F_2 generation. There was no marker information to separate alleles 5 and 6, and they were therefore merged in the analysis to estimate a common effect for both alleles. The shaded regions show the clustering of the suggested triallelic QTL.

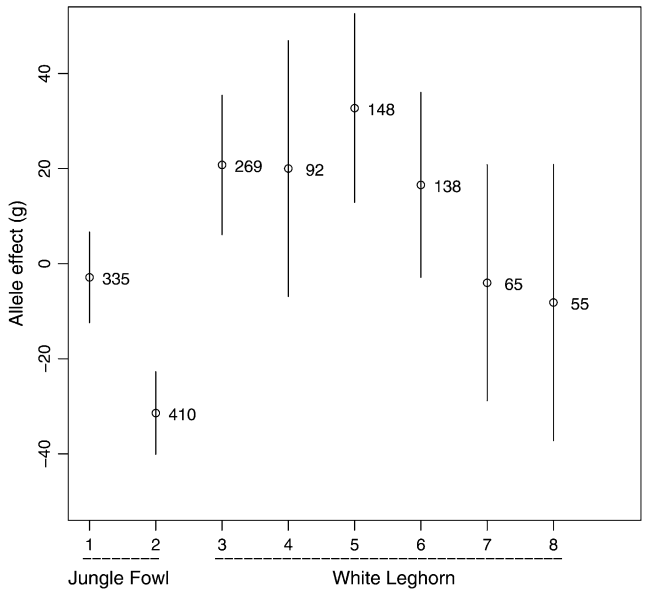

Figure 6.—Estimated base generation allele effects for growth 8 (Table 4) on chromosome 5 (±1 SE) in the Red Jungle Fowl × White Leghorn F_2 intercross. The studied trait was body weight (grams) at 200 days of age. The numbers indicate the expected number of copies of the founder alleles in the F_2 generation.

(Rönnegård and Carlborg 2007). Moreover, Xu (1998b) pointed out that the interpretation in terms of heritabilities and average gene substitution effects is conceptually more logical in VC modeling than in models where the family-specific QTL effects are included as fixed effects. Furthermore, when the parameters of interest are modeled as a combination of fixed and random, model selection using a residual likelihood is not straightforward (Welham and Thompson 1997) since the residual likelihood is based on the residual values after fitting the fixed effects. Hence, standard derivative-based REML estimation algorithms cannot be used, and the method of Perez-Enciso and Varona (2000), therefore, maximizes the likelihood with a relatively slow nonderivative algorithm. The method of Perez-Enciso and Varona (2000) may be powerful but is too slow for practical use in large-scale QTL studies where empirical thresholds are desirable.

The nested within-half-sib family model developed by Knott *et al.* (1996) allows for QTL segregating within lines and is simple and fast. Its use in F_2 intercrosses is, however, limited to pedigrees with a low degree of relationship between the F_1 individuals. Moreover, the power decreases with the number of F_1 males, because each male is assumed to have fixed and independent QTL effects.

In our model, we did not include dominance. Dominance is possible to implement in a VC model assuming an outbred base population (*e.g.*, Xu 1996). Including dominance in our segregation model should be feasible by incorporating the two dominance IBD matrices (Xu 1996) obtained by letting the base generation alleles be all independent in the first matrix and fixed within lines in the second matrix. Dominance was not included here to keep the model and the analyses simple. FIA did not detect the two QTL with large dominance effects published in Kerje *et al.* (2003) and our intention is to develop the model to include dominance in the future.

Our analysis of additive-by-additive epistasis showed that both the variance component of the main QTL effect and the epistatic effect could be adequately estimated. The within-line correlation was, however, overestimated when epistasis was included. This bias might be explained by the fact that we did not include the within-line correlation in the interaction part of the model. We plan to include within-line correlation in the modeling of epistasis, which will require a REML algorithm that allows a nonlinear VC model since the parameter ρ is included both in the variance component of the main effect and the epistatic effect.

We used a common correlation within lines in the genome-scan model. A possible extension would be to use a model with different correlations within the two founder lines. We do, however, recommend that the number of parameters (and the degrees of freedom used) should be as few as possible, not to decrease the power to detect fixed QTL. For the same reason, we do not expect that different QTL variances in the two founder lines would improve the model.

The REML estimates for different within-line correlations did not converge for the QTL on chromosome 5. Convergence problems in REML are common when many VCs are included and our model with different within-line correlations includes three different VCs for the QTL effect plus the residual variance. The observation that the QTL is likely to be segregating only within the jungle fowl founder line (Figure 6), consisting of a single individual and only two alleles, can also be a reason why the model did not converge. One of the great advantages of using the score statistic in the genome scan is that problems of convergence are completely avoided, since the calculation of the score statistic is noniterative.

The computational requirements with FIA are higher than with Haley–Knott regression, but the computation time is still low. For the simulated pedigree with 800 F_2 individuals, the computation time at one position for Haley–Knott regression was <1 sec whereas the calculations of our score statistic took <10 sec (on a standard laptop computer 1.33 GHz PowerPC G4). Including epistasis (Equation 5) increased the computation time to <20 sec. Our implementation of the score statistic in R was not numerically optimized and we may therefore expect the computation time to be reduced even further in the future.

When generating F_2 pedigrees with large base generations for QTL analyses it is plausible that several base generation alleles will be represented only with few or no copies in the F_2 generation. Hence, a substantial pro-

portion of the genetic information about segregation among the founders cannot be utilized. Our results show that the power to detect QTL segregating within lines increases when there are few F_0 and many F_2 individuals, since there will be more copies of each base generation allele among the F_2 individuals. Increasing the number of founders will, however, increase the chance of having alleles with different effects if they are not fixed within lines. Our simulations were not designed to recommend an optimum number of founders to be used in F_2 intercrosses, but the results showed a clear increase in power to detect QTL segregating within lines when the ratio of F_2 to F_0 individuals was increased from 16 to 50 (Figure 1). On the basis of these results, a rough guideline would be to use <20 founders for an intercross of 1000 F_2 individuals.

A major advantage of our model is that it is quite general and easy to implement. We therefore expect that the method will have large practical importance for future QTL analyses of line crosses. The information that the method gives about which QTL that are likely to segregate within lines, the number of alleles having different effects, and which founders that are likely to carry these will be useful for fine mapping of QTL and detection of functional genes. The ability to obtain this new information in a single analysis with a method that has the same, or higher, power as the most frequently used method in all evaluated scenarios indicates that it will be the method of choice in the future. For more complex models with, *e.g.*, dominance and linked QTL, further investigation is required.

LITERATURE CITED

Anderson, M. J., and C. F. J. Ter Braak, 2002 Permutation tests for multi-factorial analysis of variance. J. Stat. Comput. Simul. **73:** 85–113.

Andersson, L., 2001 Genetic dissection of phenotypic diversity in farm animals. Nat. Rev. Genet. **2:** 1–11.

Andersson-Eklund, L., L. Marklund, K. Lundström, C. S. Haley, K. Andersson *et al.*, 1998 Mapping quantitative trait loci for carcass and meat quality traits in a Wild Boar x Large White intercross. J. Anim. Sci. **76:** 694–700.

Brockmann, G. A., C. S. Haley, U. Renne, S. A. Knott and M. Schwerin, 1998 Quantitative trait loci affecting body weight and fatness from a mouse line selected for extreme high growth. Genetics **150:** 369–381.

Carlborg, Ö., and C. S. Haley, 2004 Epistasis: Too often neglected in complex trait studies? Nat. Rev. Genet. **5:** 618–625.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

Cox, D., and C. Hinkley, 1974 *Theoretical Statistics.* Chapman & Hall, London.

Darvasi, A., and M. Soller, 1995 Advanced intercross lines, an experimental population for fine genetic mapping. Genetics **141:** 1199–1207.

Davison, A. C., and D. V. Hinkley, 1997 *Bootstrap Methods and Their Application.* Cambridge University Press, Cambridge, UK/London.

Feenstra, B., I. M. Skovgaard and K. W. Broman, 2006 Mapping quantitative trait loci by an extension of the Haley–Knott regression method using estimating equations. Genetics **173:** 2269–2282.

Fernando, R. L., and M. Grossman, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467–477.

Goddard, M. E., 1992 A mixed model for analyses of data on multiple genetic markers. Theor. Appl. Genet. **83:** 878–886.

Goldgar, D. E., 1990 Multipoint analysis of human quantitative genetic variation. Am. J. Hum. Genet. **47:** 957–967.

Good, P., 2005 *Permutation, Parametric, and Bootstrap Tests of Hypothesis*, Ed. 3. Springer, New York.

Haley, C. S., S. A. Knott and J. M. Elsen, 1994 Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics **136:** 1195–1207.

Henderson, C. R., 1953 Estimation of variance and covariance components. Biometrics **9:** 226–252.

Jacobsson, L., H. B. Park, P. Wahlberg, R. Fredriksson, M. Perez-Enciso *et al.*, 2005 Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. Genet. Res. **86:** 115–125.

Johnson, D. L., and R. Thompson, 1995 Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. J. Dairy Sci. **78:** 449–456.

Kerje, S., Ö. Carlborg, L. Jacobsson, K. Schütz, C. Hartmann *et al.*, 2003 The twofold difference in adult size between the red junglefowl and White Leghorn chickens is largely explained by a limited number of QTLs. Anim. Genet. **34:** 264–274.

Knott, S. A., J. M. Elsen and C. S. Haley, 1996 Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. Theor. Appl. Genet. **93:** 71–80.

Knott, S. A., L. Marklund, C. S. Haley, K. Andersson, D. Davies *et al.*, 1998 Multiple marker mapping of quantitative trait loci in a cross between outbred Wild Boar and Large White pigs. Genetics **149:** 1069–1080.

Lundström, K., A. Karlsson, J. Håkansson, I. Hansson, M. Johansson *et al.*, 1995 Production, carcass and meat quality traits of F2-crosses between European wild pigs and domestic pigs including halothane gene carriers. Anim. Sci. **61:** 325–331.

Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.

Meuwissen, T. H. E., and M. E. Goddard, 2000 Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics **155:** 421–430.

Mitchell, B. D., S. Ghosh, J. L. Schneider, G. Birznieks and J. Blangero, 1997 Power of variance component linkage analysis to detect epistasis. Genet. Epidemiol. **14:** 1017–1022.

Patterson, H. D., and R. Thompson, 1971 Recovery of inter-block information when block sizes are unequal. Biometrika **58:** 545–554.

Perez-Enciso, M., and L. Varona, 2000 Quantitative trait loci mapping in F2 crosses between outbred lines. Genetics **155:** 391–405.

Perez-Enciso, M., R. L. Fernando, J. P. Bidanel and P. Le Roy, 2001 Quantitative trait locus analysis in crosses between outbred lines with dominance and inbreeding. Genetics **159:** 413–422.

Pong-Wong, R., A. W. George, J. A. Woolliams and C. S. Haley, 2001 A simple and rapid method for calculating identity-by-descent matrices using multiple markers. Genet. Sel. Evol. **33:** 453–471.

Putter, H., L. A. Sandkuijl and J. C. van Houwelingen, 2002 Score test for detecting linkage to quantitative traits. Genet. Epidemiol. **22:** 345–355.

R Development Core Team, 2004 *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna. http://www.r-project.org.

Rönnegård, L., and Ö. Carlborg, 2007 Separation of base allele and sampling term effects gives new insights in variance component QTL analysis. BMC Genet. **8:** 1.

Rönnegård, L., R. Pong-Wong and Ö. Carlborg, 2008 Defining the assumptions underlying modelling of epistatic QTL using variance component methods. J. Hered. (in press).

Self, S. G., and K.-Y. Liang, 1987 Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Am. Stat. Assoc. **82:** 605–610.

Talbot, C. J., A. Nicod, S. S. Cherny, D. W. Fulker, A. C. Collins *et al.*, 1999 High resolution mapping of quantitative trait loci in outbred mice. Nat. Genet. **21:** 305–308.

Tang, H.-K., and D. Siegmund, 2001 Mapping quantitative trait loci in oligogenic models. Biostatistics **2:** 147–162.

Wang, T., R. L. Fernando and M. Grossman, 1998 Genetic evaluation by best linear unbiased prediction using marker and trait information in a multibreed population. Genetics **148:** 507–515.
Welham, S. J., and R. Thompson, 1997 Likelihood ratio tests for fixed model terms using residual maximum likelihood. J. R. Stat. Soc. Ser. B (Methodol.) **59:** 701–714.
Xu, S., 1996 Computation of the full likelihood function for estimating variance at a quantitative trait locus. Genetics **144:** 1951–1960.
Xu, S., 1998a Further investigation on the regression method of mapping quantitative trait loci. Heredity **80:** 364–373.
Xu, S., 1998b Mapping quantitative trait loci using multiple families of line crosses. Genetics **148:** 517–524.

Communicating editor: L. McIntyre

APPENDIX: DERIVING THE THEORETICAL VALUES OF THE VARIANCE COMPONENTS FOR CASES 1–4

The theoretical values of the variance components for cases 1–4 can be derived using Henderson's (1953) method 1. It is important to keep in mind that the estimated variances in a VC model are measures of the variances in the population that the founders were taken from. If the m founders are outbred, the $2m$ QTL alleles are independently sampled from a common metapopulation, whereas if the two founder lines are fixed then only two QTL alleles have been drawn from the common metapopulation of alleles. Hence, a VC QTL model estimates the QTL variance of this metapopulation.

Let y be the vector of allele effects in the founders and a the simulated QTL allele effect. Then for the small base population with four founders we have for cases 1, 2, 3, and 4, respectively,

$$y = \begin{pmatrix} 0 \\ 0 \\ a \\ a \\ a \\ a \\ a \\ a \end{pmatrix}, y = \begin{pmatrix} 0 \\ 0 \\ a \\ a \\ a \\ 0 \\ a \\ a \end{pmatrix}, y = \begin{pmatrix} 0 \\ 0 \\ 0 \\ a \\ a \\ a \\ 0 \\ a \end{pmatrix}, y = \begin{pmatrix} 0 \\ a \\ 0 \\ a \\ 0 \\ a \\ 0 \\ a \end{pmatrix}.$$

Let $\mathbf{Z}$ be the incidence matrix relating the founder allele effects to lines. Then for cases 1, 2, and 4 we have

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

and for case 3

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

For simplicity, we assume that the only fixed effect we have is the population mean, which is the case in our simulations also. Let θ be the parameter vector

$$\theta = \begin{pmatrix} \mu^2 \\ \sigma_b^2 \\ \sigma_w^2 \end{pmatrix},$$

where μ is the population mean from which the QTL alleles have been sampled, σ_b^2 is the variance between lines and σ_w^2 is the within-line variance. The genotypic QTL variance is then given by $2\left(\sigma_b^2 + \sigma_w^2\right)$ and the within-line correlation is given by

$$\frac{\sigma_b^2}{\left(\sigma_b^2 + \sigma_w^2\right)}.$$

The theoretical value for the genotypic variance is easy to obtain intuitively for case 1. In case 1 the random effects are known to be $-\frac{1}{2}a$ and $\frac{1}{2}a$ given y above, and since they have been sampled from a common population of allelic effects the estimated variance of this population is $(1/(N-1))\sum_{i=1}^{N}\left(\frac{1}{2}a\right)^2$ with $N = 2$, which is equal to $\frac{1}{2}a^2$, and the genotypic variance is twice the allelic variance. Hence, the genotypic variance is a^2 for case 1.

For cases 2–4, it is more difficult to derive the expectation of θ intuitively. The expectation of θ may then obtained from Henderson's method 1 as

$$\theta = \mathbf{A}^{-1}\mathbf{w},$$

where

$$\mathbf{w} = \begin{pmatrix} y'y \\ \frac{1}{n_0}y'\mathbf{J}y \\ y'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'y \end{pmatrix},$$

where $n_0 = 8$ and $\mathbf{J}$ is an $n_0 \times n_0$ matrix of ones,

$$\mathbf{A} = \begin{pmatrix} n_0 & n_0 & n_0 \\ n_0 & \mathrm{tr}(\mathbf{Q}_2\mathbf{Z}\mathbf{Z}') & \mathrm{tr}(\mathbf{Q}_2) \\ n_0 & \mathrm{tr}(\mathbf{Q}_3\mathbf{Z}\mathbf{Z}') & \mathrm{tr}(\mathbf{Q}_3) \end{pmatrix},$$

where $\mathbf{Q}_2 = (1/n_0)\mathbf{J}$, and $\mathbf{Q}_3 = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.