Genet. Res., Camb. (2010), **92**, *pp.* 13–23. © Cambridge University Press 2010
doi:10.1017/S0016672310000029

Mapping interacting QTL for count phenotypes using hierarchical Poisson and binomial models: an application to reproductive traits in mice

JUN LI[1], RICHARD REYNOLDS[1], DANIEL POMP[3], DAVID B. ALLISON[1,2] AND NENGJUN YI[1,2]*

[1] *Department of Biostatistics, Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA*
[2] *Clinical Nutrition Research Center, University of Alabama at Birmingham, Birmingham, AL 35294, USA*
[3] *Departments of Genetics, Nutrition, Cell and Molecular Physiology, University of North Carolina, Chapel Hill, NC 27599, USA*

(*Received 28 September 2009 and in revised form 20 December 2009; first published online 4 March 2010*)

Summary

We proposed hierarchical Poisson and binomial models for mapping multiple interacting quantitative trait loci (QTLs) for count traits in experimental crosses. We applied our methods to two counted reproductive traits, live fetuses (*LF*) and dead fetuses (*DF*) at 17 days gestation, in an F_2 female mouse population. We treated observed number of *corpora lutea* (ovulation rate) as the baseline and the total trials in our Poisson and binomial models, respectively. We detected more than 10 QTLs for *LF* and *DF*, most having epistatic and pleiotropic effects. The epistatic effects were larger, involved more QTLs, and explained a larger proportion of phenotypic variance than the main effects. Our analyses revealed a complex network of multiple interacting QTLs for the reproductive traits, and increase our understanding of the genetic architecture of reproductive characters. The proposed statistical models and methods provide valuable tools for detecting multiple interacting QTLs for complex count phenotypes.

1. Introduction

Reproductive traits of animals determine the efficiency of food production, and hence are economically important traits for genetic improvement in the livestock industry (Rocha *et al.*, 2004*b*). However, since the reproductive traits usually have a low heritability, their improvement has met with limited success (Hansen *et al.*, 1983; Avalos & Smith, 1987; Bennett & Leymaster, 1989; Lamberson, 1990). Fortunately, along with the recent advances of genomic technologies and the collective efforts of statistical genetics, methodologies have been introduced to identify quantitative trait loci (QTLs) regulating reproductive traits, and ultimately enhance genetic processes by using QTL-complemented breeding and selection strategies in animal food production (Rathje *et al.*, 1997; Rohrer *et al.*, 1999; Wilkie *et al.*, 1999; Cassady *et al.*, 2001; King *et al.*, 2003; Holl *et al.*, 2004; Rocha *et al.*, 2004*b*).

* Corresponding author: Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294-0022, USA. Email: nyi@ms.soph.uab.edu

Most QTL mapping studies for counted reproductive traits have used interval mapping or composite interval mapping, which focus on detecting marginal (main) effects of QTL (Kirkpatrick *et al.*, 1998; Spearow *et al.*, 1999; Rocha *et al.*, 2004*b*; Sato *et al.*, 2006; Ren *et al.*, 2009). Epistatic interaction effects (epistasis) have not been thoroughly examined, even though the genetic architecture of most quantitative traits in organisms from plants to animals including humans is quite complex, including inputs from multiple QTLs and epistasis (Doebley & Stec, 1991; Cheverund & Routman, 1995; Lynch & Walsh, 1998; Price & Courtois, 1999; Sugiyama *et al.*, 2001; Wade, 2001; Kopp *et al.*, 2003; Carlborg & Haley, 2004). Furthermore, some studies have shown that epistatically acting QTLs play a larger role than do the marginal effect QTLs in genetic modulation and evolution of quantitative traits (Yu *et al.*, 1997; Yi *et al.*, 2004*a*, *b*, 2006; Moore, 2005; Valdar *et al.*, 2006). Therefore, to better understand the genetic architecture of counted reproductive traits, statistical analyses are required to simultaneously identify main effects and epistatic interactions of multiple QTLs.

Yi & Banerjee (2009) proposed a unified methodology for mapping multiple interacting QTLs based upon the hierarchical generalized linear model framework. The key to the approach is the use of a continuous prior distribution on coefficients that favours sparseness in the fitted models and facilitates computation. The method of Yi & Banerjee (2009) can analyse various continuous and discrete phenotypes, and can fit a large number of effects, including covariates, main effects of numerous loci and epistatic and gene–environment interactions. Yi & Banerjee (2009) illustrated the hierarchical generalized linear model framework with normal and probit models for continuous and binary traits. In this study, we developed hierarchical Poisson and binomial models for mapping multiple interacting QTLs for complex count traits in experimental crosses based on the approach of Yi & Banerjee (2009).

We applied our methods to data collected from a QTL mapping study of female reproductive traits in mice (Rocha *et al*., 2004*a*, *b*). The mouse sample was derived from parental lines having undergone long-term selection for high and low body weight gain. The previous analysis of these mouse data adopted the composite interval mapping method and detected 15 QTLs on five chromosomes for six reproductive traits. However, epistatic effects on the traits were not evaluated. In addition, two discrete counted reproductive traits, live fetuses (*LF*) and dead fetuses (*DF*), were treated as continuous traits. Improved models for these two count phenotypes should be Poisson models with observed number of *corpora lutea* (ovulation rate (OR)) treated as the baseline or *offset*, or binomial models treating *LF* and *DF* as the numbers of 'success' out of the corresponding number of *corpora lutea* (trials). The observed number of OR provides us the information about the upper bound of *LF* and *DF*. Therefore, to comprehensively and properly explore the genetic architecture of these reproductive traits in mice, we reanalysed the data by using the hierarchical Poisson and the binomial regression models. The aims of the present study were to estimate the epistatic effects of interacting QTLs on reproductive traits, and to detect possible additional QTLs for these traits.

2. Materials and methods

The mouse population, phenotypes and marker genotyping have been presented in detail in Rocha *et al*. (2004*a*, *b*), but are briefly described here.

(i) *Mouse lines and crosses*

A total of 439 F_2 female mice were bred from two long-term selected lines, the high-growth M16i line and the low-body-weight L6 line. The former was derived from an outbred population, whereas the latter was derived from a cross of four inbred lines. M16i females were intercrossed with L6 males, and the resulting F_1 mice were *inter se* mated (no full-sib pairings) in two consecutive replicates encompassing a total of 64 full-sib F_2 families. These mice were reared at 21 °C in a 12 : 12 h light : dark cycle and 55 % relative humidity. Food and water were supplied *ad libitum*. All the animals were handled according to the Institutional Animal Care and Use Committee Guidelines.

(ii) *Phenotypes and environmental effects*

To measure the reproductive traits of female mice, 10-week-old F_2 females were exposed to unrelated F_1 males (B6C3F1/J) until a copulatory plug was detected. Pregnant females were subsequently euthanized at day 17 of gestation to obtain three counted reproductive phenotypes: *LF*, *DF* and OR (also called *corpora lutea*) (Fig. 1). Body weights at 10 weeks of age (*WK10*) were also measured, which are significantly correlated with the reproductive phenotypes (Rocha *et al*., 2004*b*). Thus, our models adjusted for the covariates *WK10* as well as the family indicator as environmental factors.

(iii) *Genotyping*

A total of 63 fully informative microsatellite markers spanning the 19 autosomes were genotyped (Rocha *et al*., 2004*a*, *b*). Marker genotypes were determined by PCR and agarose gel electrophoresis protocols. Segregation distortion was evaluated by Chi-square test. Detection and correction of genotyping errors were conducted with MAPMAKER. A linkage map was generated with MAPMAKER/EXP and QTL analysis was carried out after marker distances were estimated. The marker linkage map covered 1257·8 cM (Kosambi) with an average spacing of 30 cM. A total of 21 (0·04 %) marker genotypes were missing.

(iv) *Statistical analyses*

(a) *Hierarchical Poisson and binomial models*

Poisson and binomial models were used to detect QTL for the numbers of *LF* and *DF*. The Poisson model treated the number of *LF* and *DF* as counted responses relative to the number of *OR*, and the binomial model treated the number of *LF* and *DF* as the events of 'success' out of the corresponding 'trials' (the number of *OR*). Under each of these models, two approaches were applied: the first considering only main effects of QTL, and the second considering not only main effects but also their epistatic effects. All the models included *WK10* and the family indicator as continuous and categorical covariates, respectively.

https://doi.org/10.1017/S0016672310000029 Published online by Cambridge University Press

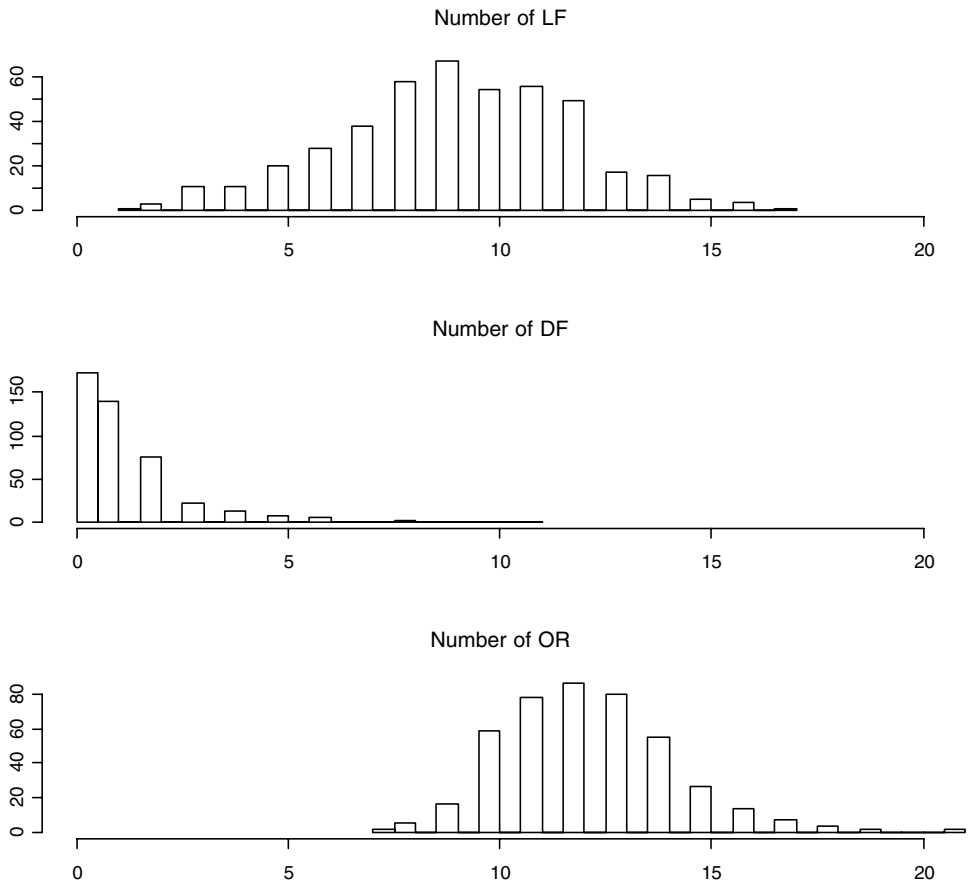

Fig. 1. Histograms of numbers of *LF*, *DF* and *OR* (also called *corpora lutea*).

Denoting the number of *LF* or *DF* by y_i and the number of *OR* by T_i, for mouse $i=1, \ldots, n$, the Poisson epistatic QTL model can be expressed as

$$y_i \sim \text{Poisson}(T_i\theta_i),$$

$$\log(\theta_i) = (\beta_{\mathbf{0}} + \boldsymbol{X}_E\boldsymbol{\beta}_E + \boldsymbol{X}_G\boldsymbol{\beta}_G + \boldsymbol{X}_{GG}\boldsymbol{\beta}_{GG})_i \triangleq (\boldsymbol{X}\boldsymbol{\beta})_i, \quad (1)$$

and the binomial epistatic QTL model can take the form

$$y_i \sim \text{Bin}(T_i, p_i),$$

$$\log\left(\frac{p_i}{1-p_i}\right) = (\beta_{\mathbf{0}} + \boldsymbol{X}_E\boldsymbol{\beta}_E + \boldsymbol{X}_G\boldsymbol{\beta}_G + \boldsymbol{X}_{GG}\boldsymbol{\beta}_{GG})_i \triangleq (\boldsymbol{X}\boldsymbol{\beta})_i, \quad (2)$$

where Poisson($T_i\theta_i$) presents Poisson distribution with mean $T_i\theta_i$, and Bin(T_i,p_i) represents binomial distribution with size T_i and probability p_i; $\boldsymbol{X}_{\mathrm{E}}$, $\boldsymbol{X}_{\mathrm{G}}$ and $\boldsymbol{X}_{\mathrm{GG}}$ are design matrices for the environmental effects, main effects (additive and dominance) and epistatic effects (additive–additive, additive–dominance, dominance–additive and dominance–dominance), respectively; $\beta_{\mathbf{0}}$ is the intercept, and $\boldsymbol{\beta}_{\mathrm{E}}$, $\boldsymbol{\beta}_{\mathrm{G}}$ and $\boldsymbol{\beta}_{\mathrm{GG}}$ are vectors of the environmental, main and epistatic effects, respectively. In the Poisson model, θ_i is the underlying rate per unit of *OR* for the *i*th mouse, and the logarithm of θ_i is expressed as a linear predictor of the environmental effects, main and epistatic effects of potential QTL. In the binomial model, p_i is the probability of each *OR* being live or dead for the *i*th mouse, and the logit transformation is used to relate p_i to the environmental effects, main and epistatic effects of potential QTLs. The Poisson and binomial models are different but both reasonable for the present data analyses. It would be expected that these two models may each enable us to detect some different QTLs.

The genetic variables were coded by using a modified Cockerham genetic model (Yi & Banerjee, 2009). The dominance variable was coded as $-0{\cdot}5$ and $0{\cdot}5$ for homozygotes and heterozygote of each QTL, respectively. To achieve the same scale as that of the dominance contrast, the additive variable was coded as $-2^{-0.5}$, 0 and $2^{-0.5}$ for the three genotypes of F_2 population, rather than the original Cockerham codes of -1, 0 and 1. The epistasis $\boldsymbol{X}_{\mathrm{GG}}$ was constructed by multiplying two corresponding main-effect variables. For each environmental variable, the raw values were transformed to have a mean of 0 and a standard deviation of $0{\cdot}5$ (Gelman *et al.*, 2008; Yi & Banerjee, 2009). This transformation standardized all the environmental effects to the scale of all the genetic main

https://doi.org/10.1017/S0016672310000029 Published online by Cambridge University Press

effects described above. For loci with missing genotypic values, the codes of contrasts were replaced by their conditional expectations given the observed marker data (Haley & Knott, 1992). Although the proposed models can include loci between the observed markers, we describe our methods by considering observed markers as potential QTLs (e.g. Xu, 2007; Yi & Banerjee, 2009).

A total of 63 observed markers lead to a total of 7938 effects, including 63 additive effects, 63 dominance effects and 7812 epistatic effects. Most of these effects are expected to be zero or at least negligible. To incorporate this notion into the model and to make the large-scale model identifiable, an independent Student-t prior $t_{\nu_j}(0,\ s_j^2)$ was assumed on coefficients β_j for $j=1, \ldots, J$ (Yi & Banerjee, 2009). For the intercept β_0, a weakly informative prior with $\nu_0=1$ and $s_0=10$ was used, and for main and epistatic effects, $\nu_j=0{\cdot}01$ and $s_j=10^{-4}$ were assigned. These shrinkage priors give each coefficient a high probability of being near zero while still allowing for occasionally large effects (Yi & Xu, 2008; Yi & Banerjee, 2009). To facilitate the estimation of parameters, the t distribution $t_{\nu_j}(0,\ s_j^2)$ was expressed as a mixture of normal distributions with mean 0 and variance distributed as scaled inverse-χ^2:

$$\beta_j|\tau_j^2 \sim N(0,\tau_j^2), \quad \tau_j^2 \sim \text{Inv}-\chi^2(\nu_j, s_j^2). \tag{3}$$

(b) *Model fit and search algorithm*

The computational algorithm of Yi & Banerjee (2009) as implemented in the freely available R package qtlbim (Yandell *et al.*, 2007) was used to fit the Poisson and binomial models by estimating the posterior modes of the coefficients. Yi & Banerjee (2009) developed a procedure to fit generalized linear models with the Student-t prior by incorporating an expectation–maximum (EM) algorithm into the usual iteratively weighted least squares (IWLS). The IWLS algorithm approximates the generalized linear model by a weighted normal linear regression (Gelman *et al.*, 2003). At each iteration, one calculates pseudo-datum z_i and pseudo-variance σ_i^2 for each observation i based on the current estimates of parameters, approximates the generalized linear model likelihood $p(y_i|\boldsymbol{X_i\beta},\phi)$ by the normal likelihood $N(z_i|\boldsymbol{X_i\beta},\sigma_i^2)$, and then updates the parameters β_j by a weighted linear regression. The EM algorithm of Yi & Banerjee (2009) treats variances τ_j^2 as missing data and replaces them by their posterior means at each E-step. Given the variances τ_j^2, the parameters β_j are estimated by including the priors of $\beta_j|\tau_j^2 \sim N(0,\tau_j^2)$ as additional 'data points' in the normal likelihood $N(z_i|\boldsymbol{X_i\beta},\sigma_i^2)$. Briefly, the algorithm starts with initial values for each τ_j^2 and β_j, and then proceeds as follows: (1) based on the current values of β_j, calculate pseudo-datum z_i and pseudo-variance σ_i^2; (2) E-step: replace each variance τ_j^2 by its conditional posterior expectation; (3) M-step: perform the weighted least square regression based on the normal likelihood approximation to obtain estimates $\hat{\beta}_j$ and (4) repeat steps 1–3 until convergence.

The hierarchical model can simultaneously handle many correlated variables (Xu, 2007; Yi & Banerjee, 2009). However, fitting a model with thousands of variables requires large memory and intensive computation. Thus, the model search strategy proposed by Yi & Banerjee (2009) was used to build a parsimonious model. The approach sets two threshold values t_1 (say 10^{-10}) and t_2 (say 0·005) to control the effect size and the P-values, respectively, and then proceeds to identify significant effects as follows: (1) searching for main effects: for each chromosome c ($c=1, 2, \ldots, 19$), simultaneously add all possible main effects for markers on chromosome c into the current model, fit the model, and then delete main effects satisfying $|\hat{\beta}_j| < t_1$; (2) searching for epistatic effects among the included main effects: simultaneously add all possible interactions among the included main effects into the current model, fit the model, and then delete epistatic effects satisfying $|\hat{\beta}_j| < t_1$; (3) searching for epistatic effects between the excluded and the included main effects: for chromosome c ($c=1, 2, \ldots, 19$), simultaneously add all possible interactions between the remaining main effects (not included in the current model) on chromosome c and the included main effects into the current model, fit the model, and then delete epistatic effects $|\hat{\beta}_j| < t_1$; (4) removing main and epistatic effects with P-values larger than t_2 from the current model.

After these steps, a final model would be obtained with both the preset environmental effects and the genetic effects that are associated with the phenotype. The estimates of these effects and the corresponding P-values indicate how strongly they influence the phenotype. We present estimate, standard error and P-value for each effect included in the final model, and calculated two measures of model fit and comparison, the deviance and the Akaike information criterion (AIC).

3. Results

(i) *LF*

(a) *Poisson models*

Using the main-effect model, five main effects were detected (the left panel of Fig. 2). Among them, two additive effects were located on chromosomes 2 and 10, and three dominance effects on chromosomes 1, 2 and 10, respectively. The strongest main effect on *LF* was on chromosome 2, at position 65·3 cM, and it

https://doi.org/10.1017/S0016672310000029 Published online by Cambridge University Press

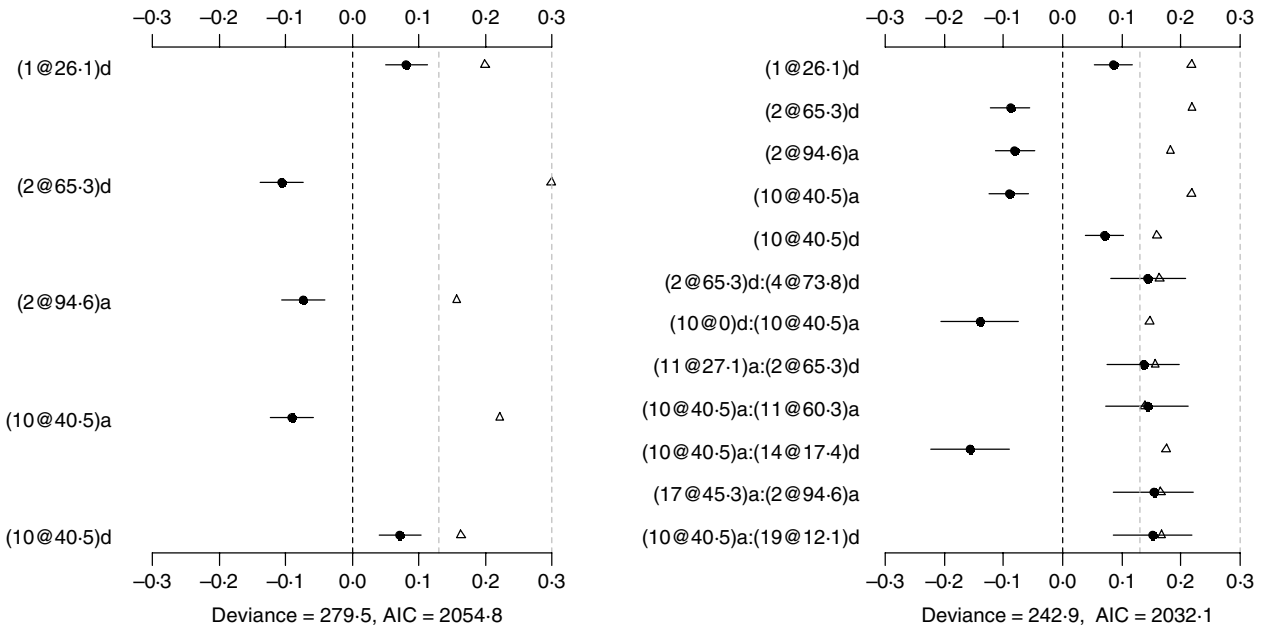

Fig. 2. Poisson main-effect (left) and epistatic (right) models for *LF*: estimated effects with ± 1 standard errors (dots and short lines), *P*-values (rescaled as $-\log_{10} p/10$) (triangles), and deviance and AIC. The notation for additive effect (C@h)a dominance effect (C@h)d, indicates chromosome C and position h cM. The term $X_1 : X_2$ represents interaction between X_1 and X_2. The two grey lines indicate the *P*-values of 0·05 and 0·001, respectively.

accounted for the highest proportion of phenotypic variance (about 3%). The *P*-values for the main effects were all less than 0·05.

The Poisson epistatic model detected five main effects and seven epistatic effects (the right panel of Fig. 2). The five main effects were identical to those of the main effect model. Among the seven epistatic effects, there were two additive–additive, three additive–dominance, one dominance–additive and one dominance–dominance effects on *LF* located on some pairs of the seven chromosomes (2, 4, 10, 11, 14, 17 and 19). The epistatic effects involved three QTLs identified in the main effect model and seven novel QTLs, suggesting that the epistasis model can identify QTLs with weak main effects but strong epistasis. The strongest epistatic effect on *LF* was observed between chromosome 10 and chromosome 14. The absolute values of estimated epistatic effects ranged from 0·13 to 0·16, which were about two times greater than those of the main effects, and the total proportion of phenotypic variances explained by the epistatic effects was approximately 9%, which was higher than that of the main effects. This revealed evidence that QTL affect *LF* mainly through their epistatic effects and the epistasis plays a more important role than does the main effect in controlling the genetic variation of *LF*. The *P*-values for the main and epistatic effects were all less than 0·05.

For comparison purposes, the Poisson main-effect model for *LF* was refitted without model search (i.e. always including all of 126 additive and dominance effects in the model). Figure 3 shows that the main effects detected in this saturated model coincided with those in the parsimonious model (the left panel of Fig. 2). The majority of genetic effects were shrunk to zero, indicating that the proposed model with shrinkage prior can capture the notion that most genetic effects influencing *LF* are very weak. We also tried to simultaneously fit all the main effects and epistatic interactions, but the attempt failed because of requirement of large memory. This demonstrated that a model search strategy is necessary to build a parsimonious model by seeking significant genetic effects when the number of effects is huge.

(b) *Binomial models*

The binomial main-effect model identified seven main effects (the left panel of Fig. 4). Among these main effects, four additive effects were located on chromosomes 2, 6 and 9, and three dominance effects on chromosomes 1, 2 and 10, respectively. The strongest main effect on *LF* was on chromosome 9, at position 0 cM, and it contributed the highest proportion (about 6%) to the phenotypic variance. The *P*-values for the main effects were all less than 0·0005.

Under the epistatic model, five main effects and 11 epistatic effects were identified (the left panel of Fig. 4). Among the five main effects, four effects overlapped those in the binomial main-effect model. The strongest main effect on *LF* was on chromosome 2, at position 65·3 cM, and it explained the highest proportion (about 5%) of phenotypic variance.

https://doi.org/10.1017/S0016672310000029 Published online by Cambridge University Press

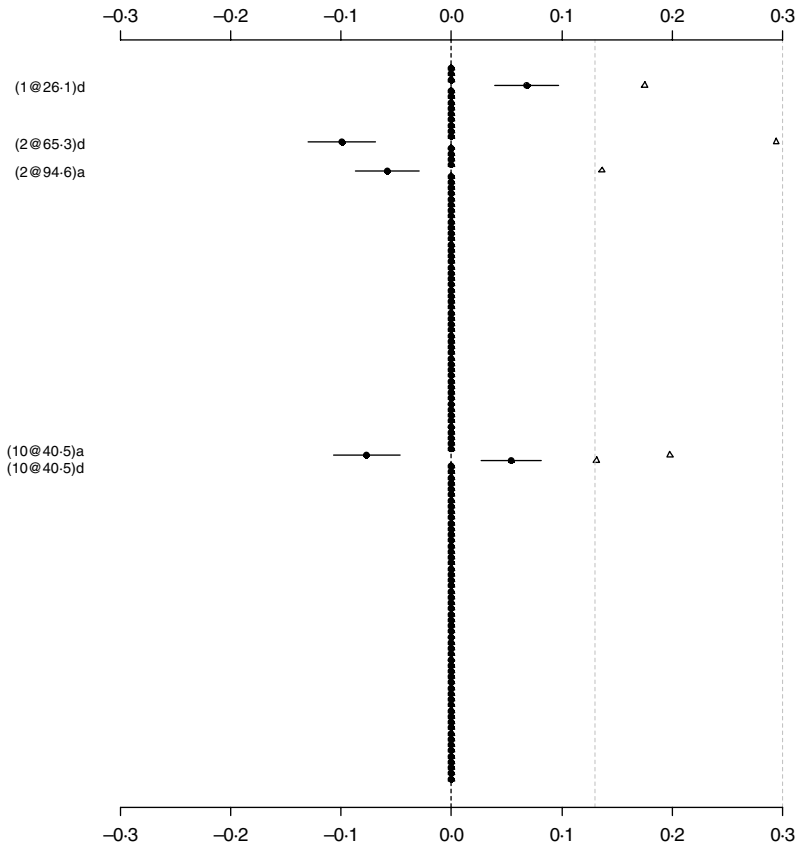

Fig. 3. Poisson main-effect model without model search for *LF*: estimated effects with ± 1 standard errors (dots and short lines), and *P*-values (rescaled as $-\log_{10} p/10$) (triangles). The notation for additive effect (C@h)a dominance effect (C@h)d, indicates chromosome C and position h cM. The term $X_1 : X_2$ represents interaction between X_1 and X_2. The two grey lines indicate the *P*-values of 0·05 and 0·001, respectively.

Among the 11 epistatic effects, there were one additive–additive, two additive–dominance, five dominance–additive and three dominance–dominance effects. The epistatic effects involved three main effects that were identified and 16 effects that were not identified in the main-effect model. This shows that the model can identify QTLs which have a weak main effect but a strong interaction with other QTLs. The strongest epistatic effect was observed between chromosome 13 and chromosome 16. The absolute values of estimated epistatic effects ranged from 0·4 to 0·9, two times greater than those of main effects, and the total proportion of phenotypic variance explained by the epistatic effects was approximately 23%, about two times greater than that of main effects. All the *P*-values for the main and epistatic effects in the model were less than 0·001.

(ii) *DF*

(a) *Poisson models*

The Poisson main-effect model detected nine main effects on *DF* (the left panel of Fig. 5). Among these main effects, six additive effects were on chromosomes 1, 3, 6, 8, 10 and 11, and three dominance effects on chromosomes 1, 5 and 14, respectively. The strongest main effect on *DF* was on chromosome 6, at position 0 cM, and it contributed the highest proportion (about 2%) to the overall phenotypic variance. The *P*-values for the main effects were all less than 0·05.

Using the Poisson epistatic model, seven main effects and nine epistatic effects were identified (the right panel of Fig. 5). Among the seven main effects, five were identical to those identified in the main-effect model. The strongest main effect on *DF* was the same as that in the main-effect model. Among the nine epistatic effects, there were three additive–additive, two additive–dominance, three dominance–additive and one dominance–dominance effects. The epistatic effects involved only one main effect that was identified and 15 effects that were not identified in the main-effect model. The strongest epistatic effect on *DF* was observed between chromosome 2 and chromosome 13, accounting for the highest proportion of phenotypic variance (about 3%). The absolute values of estimated epistatic effects ranged from 0·5 to 0·9, which were about three times greater than those of main effects, and the total proportion of phenotypic

https://doi.org/10.1017/S0016672310000029 Published online by Cambridge University Press

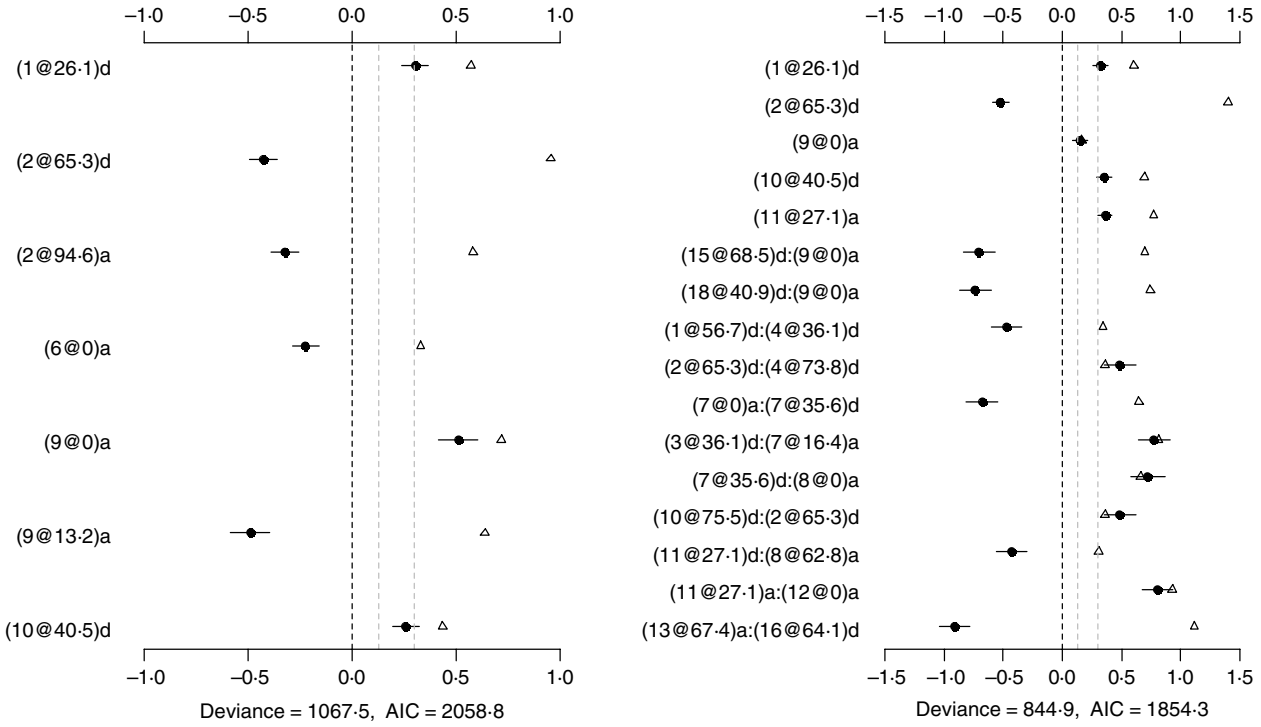

Fig. 4. Binomial main-effect (left) and epistatic (right) models for *LF*: estimated effects with ± 1 standard errors (dots and short lines), *P*-values (rescaled as $-\log_{10} p/10$) (triangles), and deviance and AIC. The notation for additive effect (C@h)a dominance effect (C@h)d, indicates chromosome C and position h cM. The term $X_1 : X_2$ represents interaction between X_1 and X_2. The two grey lines indicate the *P*-values of 0·05 and 0·001, respectively.

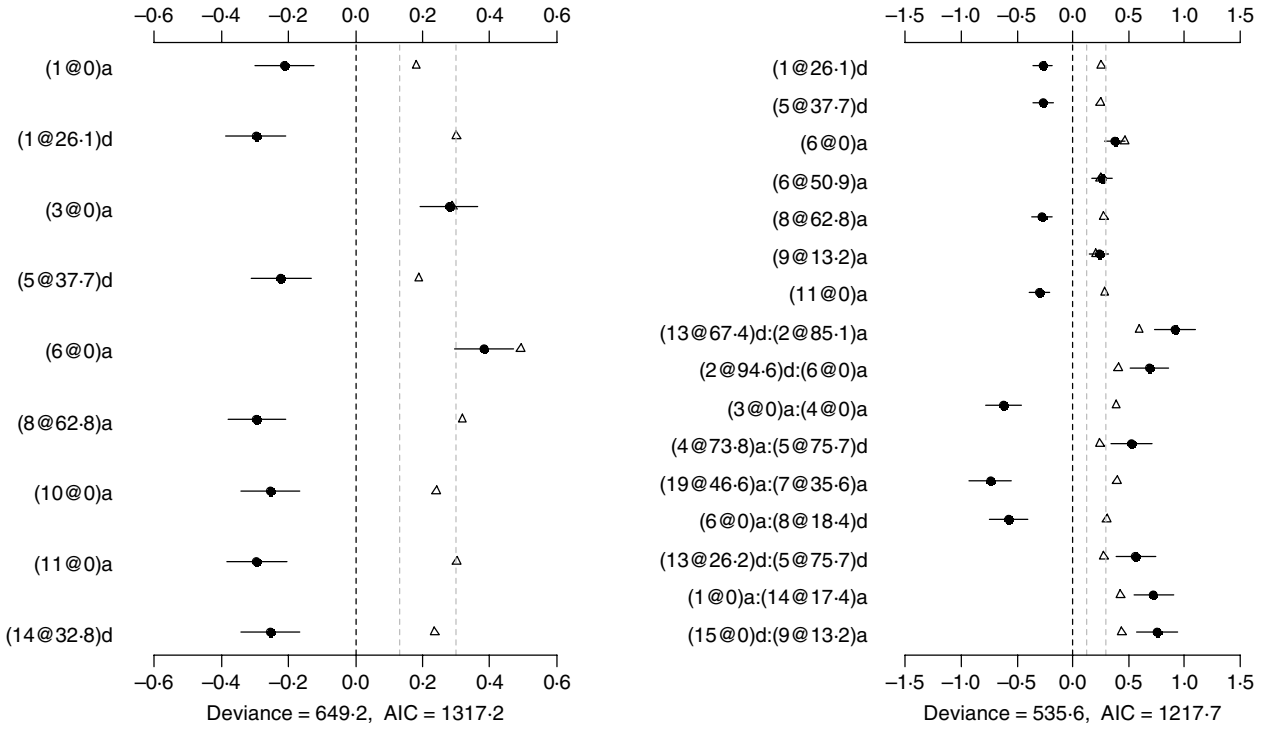

Fig. 5. Poisson main-effect (left) and epistatic (right) models for *DF*: estimated effects with ± 1 standard errors (dots and short lines), *P*-values (rescaled as $-\log_{10} p/10$) (triangles), and deviance and AIC. The notation for additive effect (C@h)a dominance effect (C@h)d, indicates chromosome C and position h cM. The term $X_1 : X_2$ represents interaction between X_1 and X_2. The two grey lines indicate the *P*-values of 0·05 and 0·001, respectively.

https://doi.org/10.1017/S0016672310000029 Published online by Cambridge University Press

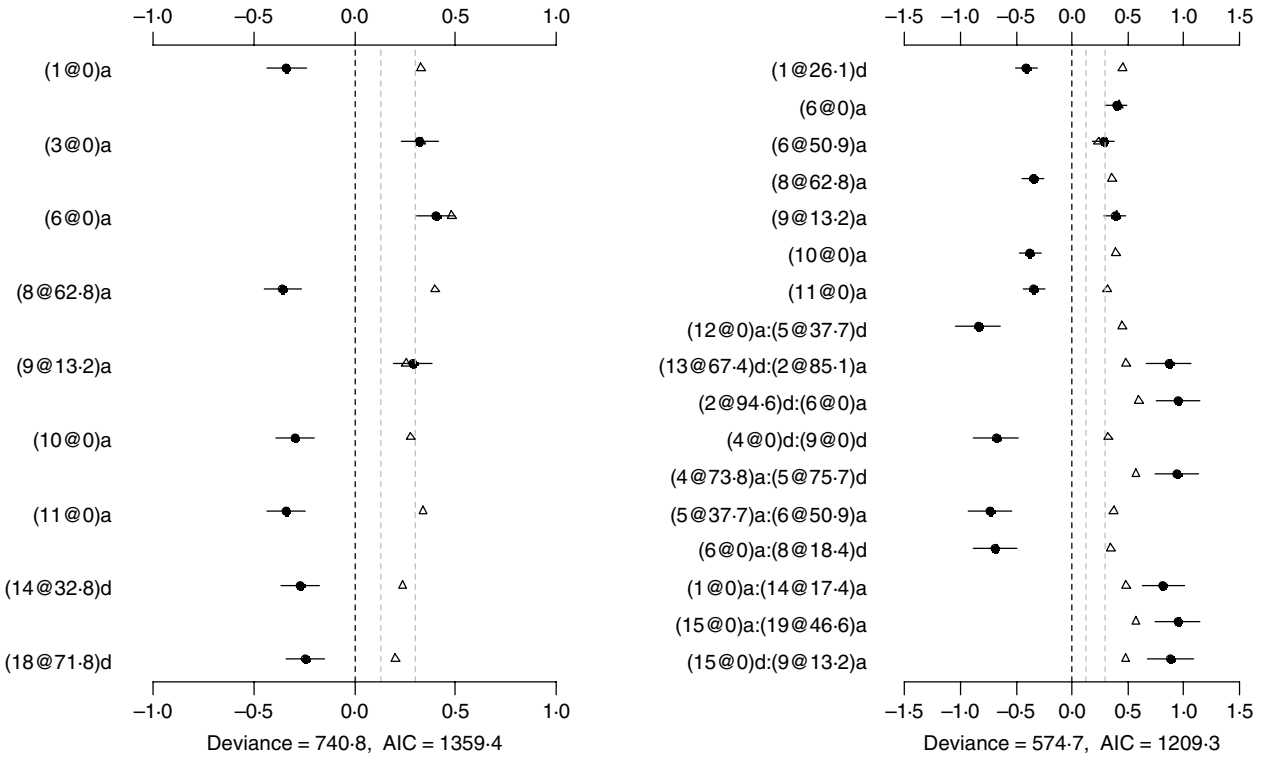

Fig. 6. Binomial main-effect (left) and epistatic (right) models for *DF*: estimated effects with ±1 standard errors (dots and short lines), *P*-values (rescaled as $-\log_{10} p/10$) (triangles), and deviance and AIC. The notation for additive effect (C@h)a dominance effect (C@h)d, indicates chromosome C and position h cM. The term $X_1 : X_2$ represents interaction between X_1 and X_2. The two grey lines indicate the *P*-values of 0·05 and 0·001, respectively.

variances explained by the epistatic effects was approximately 15%, about two times greater than that of main effects. The *P*-values for the main and epistatic effects were less than 0·001.

(b) *Binomial models*

The binomial main-effect model identified nine main effects (the left panel of Fig. 6). Seven additive effects were located on chromosomes 1, 3, 6, 8, 9, 10 and 11, and two dominance effects on chromosomes 14 and 18, respectively. The strongest main effect on *DF* was on chromosome 6, at position 0 cM, and it contributed the highest proportion (about 2%) to the phenotypic variance. The *P*-values for the main effects were all less than 0·01.

The epistatic model detected seven main effects and ten epistatic effects (the right panel of Fig. 6). Among the seven main effects, five coincided with those identified in the binomial main-effect model. The strongest main effect on *DF* was located on chromosome 1, at position 26·1 cM, and it accounted for the highest proportion of phenotypic variance (about 2%). There were three additive–additive, three additive–dominance, three dominance–additive and one dominance–dominance effects on *DF* Among these epistatic effects, six were also identified by the preceding Poisson epistatic model for *DF*. The epistatic effects involved three main effects that were identified and 15 effects that were not identified in the main-effect model. The strongest epistatic effect on *DF* was between chromosome 2 and chromosome 6. The absolute values of estimated epistatic effects ranged from 0·6 to 1·0, which were about two times greater than those of main effects, and the total proportion of phenotypic variances explained by the epistatic effects was approximately 20%, two times greater than that of main effects. All the *P*-values for the main and epistatic effects in the model were less than 0·001.

(iii) *Model comparison*

We used two summary measures, the deviance and the AIC, to compare different models. The deviance, defined as −2 times the log-likelihood, is a statistical summary of model fit; lower deviance means better fit to data. The AIC, defined as deviance +2 (number of predictors), measures the predictive power; a model is estimated to reduce out-of-sample prediction error if the AIC decreases. In all the analyses, the epistatic model had lower deviance and AIC than the corresponding main-effect model (Figs 2–6). This indicated that inclusion of the significant epistatic interactions

https://doi.org/10.1017/S0016672310000029 Published online by Cambridge University Press

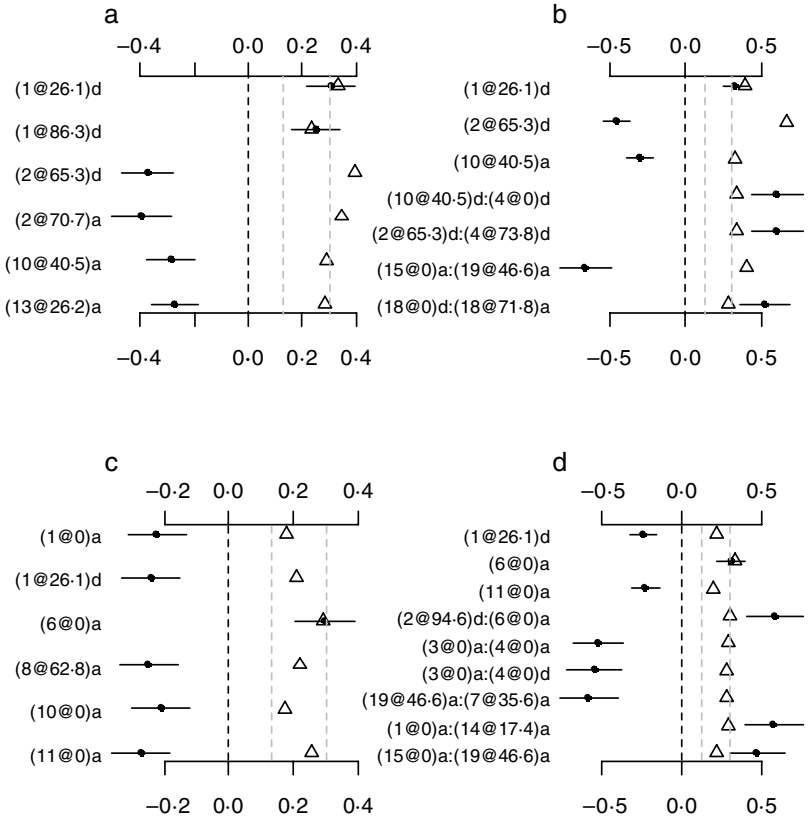

Fig. 7. Treating discrete *LF* and *DF* as normally continuous traits. a, main-effect model for *LF*; b, epistatic model for *LF*; c, main-effect model for *DF*; d, epistatic model for *DF*; estimated effects with ± 1 standard errors (dots and short lines), *P*-values (rescaled as $-\log_{10} p/10$) (triangles), and deviance and AIC. The notation for additive effect (C@h)a dominance effect (C@h)d, indicates chromosome C and position h cM. The term X_1:X_2 represents interaction between X_1 and X_2. The two grey lines indicate the *P*-values of 0·05 and 0·001, respectively.

improved the fit of the model to data and reduced out-of-sample prediction error.

(iv) *Treating discrete* LF *and* DF *as normally continuous traits*

To investigate whether or not count phenotypes can be analysed by methods for continuous traits, we performed Bayesian methods by treating the count phenotypes *LF* and *DF* or the ratios *LF*/*OR* and *DF*/*OR* as normally continuous traits. We also tried to use the Box–Cox transformation to these traits and obtained similar results displayed here. Our analyses used the same priors and the model search method as in the analyses of our Poisson and Binomial models. Figure 7 displays the estimates, standard errors and *P*-values for the effects detected in the final models. Most of the effects detected by the normal models were also detected in the Poisson and the binomial models. However, the normal models missed several strong effects that were detected by both the Poisson and the Binomial models. The results indicated that the proposed methods were more appropriate for analysing the count phenotypes. This is expected from the general framework of generalized linear models (e.g. Gelman *et al.*, 2003).

4. Discussion

We have developed hierarchical Poisson and binomial models for mapping multiple interacting QTLs for count phenotypes based upon the unified generalized linear model framework of Yi & Banerjee (2009). Our method can fit a large number of effects, including covariates, main effects of numerous loci, and epistatic and gene–environment interactions, and can accommodate the correlation among the variables. Many complex traits, including reproductive phenotypes in mice, were measured as Poisson or binomial data. However, statistical methods for mapping multiple interacting QTLs for such phenotypes have not been fully developed previously.

To better characterize the genetic architecture of reproductive traits in mice, we applied our methods to two counted reproductive traits, *LF* and *DF*, in an F_2 female mouse population. Since the observed number of OR provides important information about the variation of *LF* and *DF*, the numbers of *LF* and *DF*

https://doi.org/10.1017/S0016672310000029 Published online by Cambridge University Press

were modelled using the Poisson model with *OR* as an offset, and as the events of 'success' out of total 'trials' (*OR*) using the binomial model. As described earlier, the Poisson and binomial models capture different properties of the trait, and allow us to detect some different QTLs. In contrast, since the Poisson distribution can be derived as a limiting case to the binomial distribution as the number of trials goes to infinity and the expected number of successes remains fixed, the two models can lead to similar results (Casella & Berger, 2001). In the present study, however, since the sample size ($n=439$) is not large enough and the proportions of *LF* and *DF* (0·75 and 0·10, respectively) are not small, it is not surprising that the QTLs detected in the Poisson models are not in full agreement with those detected in the binomial models. Furthermore, since the proportion of *DF* is lower than that of *LF*, there is more coincidence between two models for *DF* than that for *LF*.

More than 10 QTLs involved in the main and epistatic effects were identified for *LF* and *DF*, respectively, which exhibit a complex pattern of genetic influence on *LF* and *DF*. Most QTLs show a very weak main effect but a strong epistatic effect on *LF* and *DF*. Compared to the initial study (Rocha *et al*., 2004*b*), in which only 3 and 1 QTLs were identified for *LF* and *DF*, respectively, and no epistasis was evaluated, the current study not only identified additional QTLs but also provided new information about the genetic architecture of *LF* and *DF* through the epistasis. These results also demonstrated that the models incorporating analysis of epistasis can identify QTLs, which might have a weak main effect but a strong epistatic effect with other QTLs. Moreover, based on the number of QTLs involved in the epistatic effects, the absolute values of estimated epistatic effects, and the total proportion of phenotypic variances accounted for by the epistatic effects, one conclusion can be drawn that the epistasis plays a more crucial role than does the main effect in regulating the genetic variation of *LF* and *DF*. Given the importance of epistasis in the genetic architecture of complex traits, appropriate statistical analyses should accommodate epistatic effects (Manolio *et al*., 2009).

Several main and epistatic effects are shared by *LF* and *DF*, which suggests that the pleiotropy plays an important role in the reproductive traits in this particular context of F_2 mice. Among all of the chromosomes on which some QTLs were detected for *LF* and *DF* in the present study, the most active one is chromosome 2, on which about seven QTLs involved in the main and epistatic effects were detected. Furthermore, the QTL on chromosome 2 has the strongest main or epistatic effects and contribute the highest proportion to the overall phenotypic variance in some fitted models, which suggests that chromosome 2 has potentially biological relevance to *LF* and *DF*. The result is consistent with that in the initial study (Rocha *et al*., 2004*b*). Other frequently involved chromosomes include chromosomes 1, 6, 9 and 10.

This work was supported by National Institutes of Health (NIH) Grants R01 GM069430 to N. Y., and P30DK056336. The opinions expressed herein are those of the authors and do not necessarily represent those of the NIH or any other organization with which the authors are affiliated.

References

Avalos, E. & Smith, C. (1987). Genetic improvement of litter size in pigs. *Animal Production* **44**, 153–164.

Bennett, G. L. & Leymaster, K. A. (1989). Integration of ovulation rate, potential embryonic viability and uterine capacity into a model of litter size in swine. *Journal of Animal Science* **67**, 1230–1241.

Carlborg, Ö. & Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics* **5**, 618–625.

Casella, G. & Berger, R. L. (2001). *Statistical Inference*, 2nd edn. Belmont, CA: Duxbury Press.

Cassady, J. P., Johnson, R. K., Pomp, D., Rohrer, G. A., Van Vleck, L. D., Spiegel, E. K. & Gilson, K. M. (2001). Identification of quantitative trait loci affecting reproduction in pigs. *Journal of Animal Science* **79**, 623–633.

Cheverund, J. M. & Routman, E. J. (1995). Epistasis and its contribution to genetic variance components. *Genetics* **139**, 1455–1461.

Doebley, J. & Stec, A. (1991). Genetic analysis of the morphological differences between maize and teosinte. *Genetics* **129**, 285–295.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd edn. London: Chapman & Hall.

Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* **2**, 1360–1383.

Haley, C. S. & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Hansen, L. B., Freeman, A. E. & Berger, P. J. (1983). Yield and fertility relationships in dairy cattle. *Journal of Dairy Science* **66**, 293–305.

Holl, J. W., Cassady, J. P., Pomp, D. & Johnson, R. K. (2004). A genome scan for quantitative trait loci and imprinted regions affecting reproduction in pigs. *Journal of Animal Science* **82**, 3421–3429.

King, A. H., Jiang, Z., Gibson, J. P., Haley, C. S. & Archibald, A. L. (2003). Mapping quantitative trait loci affecting female reproductive traits on porcine chromosome 8. *Biology of Reproduction* **68**, 2172–2179.

Kirkpatrick, B. W., Mengelt, A., Schulman, N. & Martin, I. C. (1998). Identification of quantitative trait loci for prolificacy and growth in mice. *Mammalian Genome* **9**, 97–102.

Kopp, A., Graze, R. M., Xu, S., Carroll, S. B. & Nuzhdin, S. V. (2003). Quantitative trait loci responsible for variation in sexually dimorphic traits in *Drosophila melanogaster*. *Genetics* **163**, 771–787.

Lamberson, W. R. (1990). Genetic parameters for reproductive traits. In *Genetics of the Pig* (North Central Regional Research Project NC-103 Report)

(ed. L. D. Young), pp. 70–76. Lincoln: USMARC, Clay Center, NE, and University of Nebraska.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, CA: Sinauer Associates.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A. & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.

Moore, J. H. (2005). A global view of epistasis. *Nature Genetics* **37**, 13–14.

Price, A. H. & Courtois, B. (1999). Mapping QTLs associated with drought resistance in rice: progress, problems and prospects. *Plant Growth Regulation* **29**, 123–133.

Rathje, T. A., Rohrer, G. A. & Johnson, R. K. (1997). Evidence for quantitative trait loci affecting ovulation rate in pigs. *Journal of Animal Science* **75**, 1486–1494.

Ren, D. R., Ren, J., Xing, Y. Y., Guo, Y. M., Wu, Y. B., Yang, G. C., Mao, H. R. & Huang, L.-S. (2009). A genome scan for quantitative trait loci affecting male reproductive traits in a White Duroc × Chinese Erhualian resource population. *Journal of Animal Science* **87**, 17–23.

Rocha, J. L., Eisen, E. J., Van Vleck, L. D. & Pomp, D. (2004a). A large-sample QTL study in mice: I. Growth. *Mammalian Genome* **15**, 83–99.

Rocha, J. L., Eisen, E. J., Siewerdt, F., Van Vleck, L. D. & Pomp, D. (2004b). A large-sample QTL study in mice: III. Reproduction. *Mammalian Genome* **15**, 878–886.

Rohrer, G. A., Ford, J. J., Wise, T. H., Vallet, J. L. & Christenson, R. K. (1999). Identification of quantitative trait loci affecting female reproduction traits in a multigeneration Meishan–White composite swine population. *Journal of Animal Science* **77**, 1385–1391.

Sato, S., Atsuji, K., Saito, N., Okitsu, M., Sato, S., Komatsuda, A., Mitsuhashi, T., Nirasawa, K., Hayashi, T., Sugimoto, Y. & Kobayashi, E. (2006). Identification of quantitative trait loci affecting corpora lutea and number of teats in a Meishan × Duroc F2 resource population. *Journal of Animal Science* **84**, 2895–2901.

Spearow, J. L., Nutson, P. A., Mailliard, W. S., Porter, M. & Barkley, M. (1999). Mapping genes that control hormone-induced ovulation rate in mice. *Biology of Reproduction* **61**, 857–872.

Sugiyama, F., Churchill, G. A., Higgins, D. C., Johns, C., Makaritsis, K. P., Gavras, H. & Paigen, B. (2001). Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics* **71**, 70–77.

Valdar, W., Solberg, L. C., Gauguier, D., Cookson, W. O., Rawlins, J. N. P., Mott, R. & Flint, J. (2006). Genetic and environmental effects on complex traits in mice. *Genetics* **174**, 959–984.

Wade, M. J. (2001). Epistasis, complex traits, and mapping genes. *Genetica* **112–113**, 59–69.

Wilkie, P. J., Paszek, A. A., Beattie, C. W., Alexander, L. J., Wheeler, M. B. & Schook, L. B. (1999). A genomic scan of porcine reproductive traits reveals possible quantitative trait loci (QTLs) for number of corpora lutea. *Mammalian Genome* **10**, 573–578.

Xu, S. (2007). An empirical Bayes method for estimating epistatic effects of quantitative trait loci. *Biometrics* **63**, 513–521.

Yandell, B. S., Mehta, T., Banerjee, S., Shriner, D., Venkataraman, R., Moon, J. Y., Neely, W. W., Wu, H., von Smith, R. & Yi, N. (2007). R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. *Bioinformatics* **23**, 641–643.

Yi, N., Diamont, A., Chiu, S., Kim, K., Allison, D. B., Fisler, J. S. & Warden, C. H. (2004a). Characterization of epistasis influencing complex spontaneous obesity in the B5B model. *Genetics* **167**, 399–409.

Yi, N., Chiu, S., Allison, D. B., Fisler, J. S. & Warden, C. H. (2004b). Epistatic interaction between two non-structural loci on chromosomes 7 and 3 influences hepatic lipase activity in B5B mice. *Journal of Lipid Research* **45**, 2063–2070.

Yi, N., Zinniel, D. K., Kim, K., Eisen, E. J., Bartolucci, A., Allison, D. B. & Pomp, D. (2006). Bayesian analysis of multiple epistatic QTL models for body weight and body composition in mice. *Genetical Research* **87**, 45–60.

Yi, N. & Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping. *Genetics* **179**, 1045–1055.

Yi, N. & Banerjee, S. (2009). Hierarchical generalized linear models for multiple quantitative trait locus mapping. *Genetics* **181**, 1101–1113.

Yu, S. B., Li, J. X., Xu, C. G., Tan, Y. F., Gao, Y. J., Li, X. H., Zhang, Q. & Saghai Maroof, M. A. (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proceedings of the National Academy of Sciences of the USA* **94**, 9226–9231.