Published in final edited form as:

Stat Appl Genet Mol Biol. ; 11(6): . doi:10.1515/1544-6115.1675.

A Maximum Likelihood Approach to Functional Mapping of Longitudinal Binary Traits

Chenguang Wang^{*}, Hongying Li[†], Zhong Wang[‡], Yaqun Wang[§], Ningtao Wang[¶], Zuoheng Wang^{||}, and Rongling Wu^{**}

^{*}Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21205

[†]Biostatistics and Bioinformatics Group, Moores Cancer Center, University of California at San Diego, La Jolla, CA 92093

[‡]Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033

[§]Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033

[¶]Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033

^{||}Division of Biostatistics, Yale University, New Haven, CT 06510

^{**}Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033

Abstract

Despite their importance in biology and biomedicine, genetic mapping of binary traits that change over time has not been well explored. In this article, we develop a statistical model for mapping quantitative trait loci (QTLs) that govern longitudinal responses of binary traits. The model is constructed within the maximum likelihood framework by which the association between binary responses is modeled in terms of conditional log odds-ratios. With this parameterization, the maximum likelihood estimates (MLEs) of marginal mean parameters are robust to the misspecification of time dependence. We implement an iterative procedures to obtain the MLEs of QTL genotype-specific parameters that define longitudinal binary responses. The usefulness of the model was validated by analyzing a real example in rice. Simulation studies were performed to investigate the statistical properties of the model, showing that the model has power to identify and map specific QTLs responsible for the temporal pattern of binary traits.

Keywords

binary trait; dynamic trait; functional mapping; maximum likelihood estimate

1 Introduction

Many traits important to biology and biomedicine arise in a dichotomous or binary form. For example, the fusiform rust disease in loblolly pine is described as presence or absence of the formation of galls. Despite only two states (yes or no) existing for a binary trait, this type of trait may involve a polygenic, complex component. Genetic mapping based on DNA-based molecular markers has proven to be powerful for detecting and mapping specific genes, known as quantitative trait loci (QTLs), that control a complex trait. Since Lander and

Botstein's (1989) interval mapping was published, there has been a tremendous growth in developing various statistical models for QTL mapping (see Wu et al. 2007; Broman and Sen 2009). Xu and group are among the first to develop a series of statistical approaches for mapping binary traits within the maximum likelihood or Bayesian contexts (Xu and Atchley 1996; Yi and Xu 1999a, 1999b, 2000; Xu et al. 2003, 2005). Manichaikul and Broman (2009) implemented selective genotyping approaches into QTL mapping for binary traits. Visscher et al. (1996) used stochastic stimulation to investigate the behavior of binary trait mapping in backcross and F_2 populations. Deng et al. (2006) developed finite logistic regression mixture models for interval mapping of binary traits. Kadarmideen et al. (2001) extended Xu and Atchley's (1996) approach to consider a binary mapping method for outbred lines.

Although there has been a considerable body of literature for binary mapping, a significant lack in the area is the development of dynamic models for mapping QTLs involved in longitudinal processes of binary traits. Longitudinal binary response data are common in many fields from plant and animal breeding to biomedical research. For example, the presence or absence of mastitis is longitudinally recorded in the course of lactation to breed healthy and productive cows (Rekaya et al. 2003; Hinrichs et al. 2011). To test the efficacy of hypnotic drugs in controlling insomnia, a sleep disorder that pervades the human population, a longitudinal clinical trial is designed in which the time of how quickly patients fall asleep after being treated with the drugs is viewed as a binary response, recorded repeatedly in a time course (Ghahroodi et al. 2010). For longitudinal continuous responses, a statistical model, called *functional mapping*, has been developed to map their genetic architecture (Ma et al. 2002; Wu and Lin 2006; Li and Wu 2010). Functional mapping uses mathematical functions that describe particular biological processes, such as growth trajectories, to quantify the temporal-spatial pattern of genetic control based on genotype-specific mathematical parameters. It has been extended to takes advantage of statistical modeling of longitudinal mean-covariance structures using various parametric, nonparametric, or semiparametric approaches, thereby equipped with power to test biologically meaningful hypotheses with a minimum set of parameters. It has been widely used to map QTLs that govern body mass growth in mice (Wu et al. 2005) and human (Li et al. 2009) and, recently, integrated with field experimental designs to detect QTLs controlling plant stem height and root growth trajectories in rice (Zhao et al. 2004), soybean (Li et al. 2010; Wu et al. 2011) and poplar (Zhang et al. 2009).

In this article, we propose a statistical model for mapping QTLs for longitudinal binary traits in an experimental cross. Genetic mapping of binary traits considered at a single time point has been shown to be more difficult than that of continuous counterparts (Hackett and Weller 1995; Lange and Whittaker 2001; Manichaikul and Broman 2009). Mapping longitudinal changes of binary traits, therefore, presents a greater challenge than continuous-trait mapping described in previous functional mapping models. In longitudinal studies, we measure repeated observations of a response variable and a set of covariates on the same individual over time so that the responses variables are usually correlated. This dependence must be accounted for in order to make correct inference. To incorporate longitudinal binary traits into a mixture model for functional mapping, we used Fitzmaurice and Laid's (1993) approach for modeling the association between binary responses in terms of conditional log odds-ratios. We show that functional mapping incorporated by Fitzmaurice and Laid's model provides great power to map dynamic QTLs for binary traits. We implement the maximum likelihood approach and EM algorithm to estimate and test QTL genotype-specific parameters that define binary dynamics. Simulation studies were used to investigate the statistical properties of the model and validate its usefulness.

2 Methods

2.1 Experimental design

We consider a population derived from a cross between two parental inbred lines P_1 and P_2 , differing substantially in a longitudinal binary trait of interest. Let markers $\mathbf{M}$, with alleles M and m , and $\mathbf{N}$, with alleles N and n , denote two flanking markers for an interval where a putative QTL is tested. A cross between two parents P_1 and P_2 is performed to produce an F_1 population. The F_1 progeny are all heterozygotes with the same genotype MN/mn . The F_1 individual is crossed back to P_1 or P_2 to produce a backcross population. There are four possible marker genotypes in the backcross population. Consider an unobserved QTL $\mathbf{Q}$ with alleles Q and q that is located in the interval flanked by markers $\mathbf{M}$ and $\mathbf{N}$. The distribution of unobserved QTL genotypes can be inferred from the observed flanking marker genotypes according to the recombination frequencies between them (Wu et al. 2007). The conditional probabilities of the QTL genotypes given marker genotypes are given in Table 1.

2.2 Genetic model

Two genotypes Qq , qq of the QTL in the backcross population each have a $\frac{1}{2}$ frequencies. The genetic model for a QTL

$$G = \begin{bmatrix} G_1 \\ G_0 \end{bmatrix} = \begin{bmatrix} \mu + \frac{1}{2}a \\ \mu - \frac{1}{2}a \end{bmatrix} \quad (1)$$

was proposed to specify the relation between a genotypic value G and genetic parameters, overall mean (μ) and additive effect (a). G_1 and G_0 denote the genotypic values of genotypes Qq (1) and qq (0), respectively. If two QTLs are considered for epistatic modeling, the genetic model is written as

$$G = \begin{bmatrix} G_{11} \\ G_{10} \\ G_{01} \\ G_{00} \end{bmatrix} = \begin{bmatrix} \mu + \frac{1}{2}a_1 + \frac{1}{2}a_2 + \frac{1}{4}i_{aa} \\ \mu + \frac{1}{2}a_1 - \frac{1}{2}a_2 - \frac{1}{4}i_{aa} \\ \mu - \frac{1}{2}a_1 + \frac{1}{2}a_2 - \frac{1}{4}i_{aa} \\ \mu - \frac{1}{2}a_1 - \frac{1}{2}a_2 + \frac{1}{4}i_{aa} \end{bmatrix} \quad (2)$$

where a_1 and a_2 are the additive effects of the first and second QTL, respectively, and i_{aa} is the additive $\times$ additive epistatic effect between the two QTLs.

2.3 Data structure

The data of QTL mapping consists of two parts: phenotypic data and marker information. Often, we also observe covariate information for each progeny; for example, sex or age. We assume that binary response for each of N individuals is observed at T time points. Then, we generate a $T \times 1$ vector $Y_i = (Y_{i1}, \dots, Y_{iT})^T$, where binary random variable $Y_{it} = 1$ if progeny i has response 1, i.e., presence, at time t , and 0 otherwise. Also, marker information M_i for progeny i is observed for L loci on a linkage group. Let $M_i = (M_{i1}, \dots, M_{iL})$, where $M_{il} = 1$ if the marker is heterozygous at locus l and 0 if the marker is homozygous at locus l . Each progeny has a $J \times 1$ covariate vector X_{it} at time t , and we let $X_i = (X_{i1}, \dots, X_{iT})^T$ represent the $T \times J$ matrix of covariates for progeny i . Thus, the data for progeny i includes marker and phenotypic observations in (Y_i, M_i, X_i) .

2.4 Multivariate model for binary responses

Much of this part is derived from Fitzmaurice and Laird's (1993) work. First, we describe the statistical model for longitudinal binary responses $Y = (Y_1, \dots, Y_T)^T$. Let $X = (X_1, \dots, X_T)^T$ denote the matrix of covariate for response Y , where X_t is a $J \times 1$ covariate vector corresponding to response Y_t . The marginal distribution of Y_t is binary, expressed as

$$f(Y_t|X_t) = \exp[Y_t \varepsilon_t - \log\{1 + \exp(\varepsilon_t)\}] \quad (3)$$

where a logistic link $\varepsilon_t = \log\{\mu_t/(1-\mu_t)\} = X_t^T \beta$ is assumed, with $\mu_t = \mu_t(\beta) = E(Y_t) = \Pr(Y_t = 1|X_t, \beta)$ being the probability of presence at time t and β being a $J \times 1$ vector of parameters. The logit link function is a natural choice for binary responses, although any link function could be used. We use $\mu(\beta)$ to denote the vector of marginal probabilities of presence, $\mu(\beta) = E(Y) = (\mu_1, \dots, \mu_T)^T$.

Next, following Fitzmaurice and Laird (1993), we use the form of the joint distribution of Y as follows:

$$f(Y|\Psi, \Omega) = \exp\{\Psi^T Y + \Omega^T W - A(\Psi, \Omega)\} \quad (4)$$

where $W = (Y_1 Y_2, \dots, Y_{T-1} Y_T, \dots, Y_1 Y_2 \dots Y_T)^T$ is a vector of two- and higher-way cross-products of Y , $\Psi = (\psi_1, \dots, \psi_T)^T$, $\Omega = (\omega_{12}, \dots, \omega_{(T-1)T}, \dots, \omega_{12\dots T})^T$ are vectors of canonical parameters, and $A(\Psi, \Omega)$ is a normalizing constant, $\exp\{A(\Psi, \Omega)\} = \sum \exp(\Psi^T Y + \Omega^T W)$, with the summation being over all 2^T possible values of Y . Note that μ is a function of both Ψ and Ω . Parameters Ψ and Ω can be straightforwardly interpreted in terms of conditional probabilities. For example,

$$\begin{aligned} \psi_r &= \text{logit}\{\Pr(Y_r=1|Y_s=0, s \neq r)\}, \quad r=1, \dots, T. \\ \omega_{rs} &= \log \text{OR}(Y_r, Y_s|Y_t=0, t \neq r, s), \quad r < s=1, \dots, T. \end{aligned} \quad (5)$$

and

$$\omega_{123} = \log \text{OR}(Y_1, Y_2|Y_3=1, Y_s=0, s>3) - \log \text{OR}(Y_1, Y_2|Y_3=0, Y_s=0, s>3)$$

where

$$\text{OR}(\nu, \eta) = \frac{\Pr(\nu=\eta=1)\Pr(\nu=\eta=0)}{\Pr(\nu=1, \eta=0)\Pr(\nu=0, \eta=1)}$$

is the odds ratio.

We assume that Ω is a function of a $K \times 1$ parameter vector $a = (a_1, \dots, a_K)^T$. In principle, we could use any dependence link function. Yet, a natural choice is a linear link function, $\Omega = Za$, where Z is a design matrix.

The form of the joint distribution above may model varying degrees of dependence among Y_t . If $\Omega = 0$, the independence model results. If $\Omega = (\omega_{12}, \dots, \omega_{(T-1)T}, \dots, \omega_{12\dots T})^T$, we have a saturated model for the association parameters. Between these extremes, parsimonious models for the time dependence can be considered. For instances, we can obtain a quadratic

exponential family or pairwise model by fixing a three- and higher-way association parameters of Ω to zero. In Fitzmaurice and Larid (1993), the expression for the derivative of the log-likelihood with respect to β and α is

$$\begin{pmatrix} \partial L / \partial \beta \\ \partial L / \partial \alpha \end{pmatrix} = \begin{pmatrix} X^T \Delta V_{11}^{-1} (Y - \mu) \\ Z^T (W - E(W) - V_{21} V_{11}^{-1} (Y - \mu)) \end{pmatrix} \quad (6)$$

where $V_{11} = \text{cov}(Y)$, $V_{21} = \text{cov}(W, Y)$, and $\Delta = \text{diag}\{\text{var}(Y_t)\}$ is a $(T \times T)$ diagonal matrix.

2.5 QTL mapping

Let us first consider a QTL bracketed by two markers, with conditional probabilities of QTL genotypes given in Table 1. We assume no interference in crossing over in the testing interval. Multiple QTLs with epistasis can be considered in a similar way.

The joint distribution of $(Y_i, Q_i | M_i, X_i, Z_i)$ for progeny i with QTL genotype Q_i is expressed as

$$\begin{aligned} f(Y_i, Q_i | M_i, X_i, Z_i; \Theta_i) &= \prod_{j=0}^1 [p_{ij} f(Y_i | X_i, Z_i; \Psi_{ij}, \Omega_{ij})]^{I(Q_i=j)} \\ &= \prod_{j=0}^1 [p_{ij} \exp(\Psi_{ij}^T Y_i + \Omega_{ij}^T W_i - A(\Psi_{ij}, \Omega_{ij}))]^{I(Q_i=j)} \end{aligned}$$

where p_{ij} is the conditional probability of QTL genotype j ($j = 1$ for Qq and 0 for qq) given the marker genotype of progeny i , $\Omega_{ij} = Z_i a_j$, and Ψ_{ij} is a function of both β_j and α_j . All parameters are arrayed in $\Theta_i = (p_{ij}, \beta_j, \alpha_j)$.

In practice, QTL genotype Q_i for progeny i is unobservable. The joint distribution for $(Y_i | M_i, X_i, Z_i)$ is expressed as

$$\begin{aligned} f(Y_i | M_i, X_i, Z_i; \Theta_i) &= \sum_{j=0}^1 [f(Y_i | Q_i=j, X_i, Z_i) Pr(Q_i=j | M_i)] \\ &= \sum_{j=0}^1 [p_{ij} f(Y_i | X_i, Z_i; \Psi_{ij}, \Omega_{ij})] \\ &= \sum_{j=0}^1 [p_{ij} \exp(\Psi_{ij}^T Y_i + \Omega_{ij}^T W_i - A(\Psi_{ij}, \Omega_{ij}))] \end{aligned}$$

This is a mixture model of two possible multivariate binary densities with different parameters.

Finally, the observed log-likelihood is

$$L(\Theta | Y, M, X, Z) = \sum_{i=1}^N \left\{ \log \left[\sum_{j=0}^1 p_{ij} \exp(\Psi_{ij}^T Y_i + \Omega_{ij}^T W_i - A(\Psi_{ij}, \Omega_{ij})) \right] \right\}. \quad (7)$$

2.6 Hypothesis testing

In QTL mapping, we test whether there is a QTL that controls a longitudinal binary trait at a given position within a marker interval. This can be performed by the hypotheses

$$\begin{cases} H_0: & \beta_1 = \beta_0, \alpha_1 = \alpha_0 \\ H_1: & \text{At least one of the equalities above does not hold} \end{cases} \quad (8)$$

We use the likelihood ratio test (LRT) for the existence of a QTL. The LRT statistics is

$-2\log \left[\frac{\sup_{\Theta_0} L(\Theta_0|Y, M, X, Z)}{\sup_{\Theta} L(\Theta|Y, M, X, Z)} \right]$, where Θ_0 and Θ are the parameter spaces under the H_0 and H_1 , respectively. The threshold value to reject the null hypothesis cannot be simply chosen from a χ^2 distribution because of the violation of regularity conditions of asymptotic theory under H_0 . Instead, we use permutation tests (Churchill and Doerge 1994) to get the critical value.

2.7 Maximum likelihood estimation

For a pre-specified testing position, the conditional probability parameters p_{ij} can be determined from Table 1. Thus, only β_j and α_j need to be estimated. This can be done by obtaining the derivative of the log-likelihood of progeny i with respect to β_j and α_j and the Fisher information matrix (see Appendix A for a detail). Parameter estimates are obtained under the H_1 and H_0 , respectively.

2.8 Epistatic Modeling

Genetic interactions between different QTLs, called epistasis, are thought to play an important role in trait control (Wu et al. 2005). However, there is no study yet that reports the detection of epistasis for longitudinal binary traits. Here we show how to map QTLs for longitudinal changes of binary traits. Model (2) shows the epistatic interactions between two QTLs in a backcross design. By incorporating it into the mixture likelihood (7), we have

$$L(\Theta|Y, M, X, Z) = \sum_{i=1}^N \left\{ \log \left[\sum_{j_1=0}^1 \sum_{j_2=0}^1 p_{ij_1j_2} \exp(\Psi_{ij_1j_2}^T Y_i + \Omega_{ij_1j_2}^T W_i - A(\Psi_{ij_1j_2}, \Omega_{ij_1j_2})) \right] \right\}, \quad (9)$$

where $j_1 = 1, 0$ and $j_2 = 1, 0$ are the genotypes of the first and second QTL, both of which can be inferred from marker information using conditional probability, $p_{ij_1j_2}$; and unknown vector Θ contains $(p_{ij_1j_2}, \alpha_{ij_1j_2}, \beta_{ij_1j_2})$ which are used to specify two-QTL genotypic values. Based on the relations (2), we can solve the time-dependent changes of additive effects at each QTL (α_{1t} and α_{2t}) and additive $\times$ additive epistatic effect (i_{aat}).

3 Simulation Study

The statistical properties of the new mapping model are examined through simulation studies. Consider a random sample composed of $N = 400$ progeny from a backcross population. We assume only one QTL on a chromosome with 11 equally spaced markers. The interval between two adjacent markers is 20 cM. The assumed QTL is located in the second interval (25 cM from the left end). For each progeny, measurements are taken at 3 different time points, although the model allows an arbitrary number of time points to be analyzed.

The model for the marginal probability of Y_i given X_i , $Q_i = j$ and β_j is

$$Pr(Y_{it}=1|X_{it}, Q_i=j)=\frac{\exp(\beta_{j0}+\beta_{j1}X_{it}+\beta_{j2}(t-1))}{1+\exp(\beta_{j0}+\beta_{j1}X_{it}+\beta_{j2}(t-1))} \quad t=1, 2, 3. \quad (10)$$

where X_{it} is obtained from the distribution $\text{bin}(1, 0.5)$. The time-dependence of the model is characterized in terms of conditional log odds-ratios, Ω_j . The true marginal parameters of the model is determined on the basis of heritability size. In a backcross population, genotypes Qq and qq with an equal frequency are valued as G_1 and G_0 , respectively (1). The genetic variation of a trait explained by the QTL is calculated as $\sigma_g^2 = \frac{G_1^2 + G_0^2}{2} - (\frac{G_1 + G_0}{2})^2$. Suppose the binary response corresponds to a latent variable, referred to as the liability, which is considered to be continuous and normally distributed. According to Xu and Atchley (1996), the liability is described as

$$Z_{it}|(Q_i=j)=\frac{1}{c}(\beta_{j0}+\beta_{j1}X_{it}+\beta_{j2}(t-1))+e_{it},$$

where Z_{it} is the liability of progeny i at time t , e_{it} is the residual with a distribution of $N(0, 1)$, and $c=\pi/\sqrt{3}$. It is assumed that the binary response of progeny i at time t is 1 if Z_{it} is above 0 and 0 otherwise. In this case,

$G_1=\frac{1}{c}(\beta_{10}+(\beta_{11}-\beta_{01})X_{it}+(\beta_{12}-\beta_{02})(t-1))$, $G_0=\frac{1}{c}\beta_{00}$. By assuming a heritability 0.1 or 0.4 and residual (e_{it}) variance 1, we can determine the values of the marginal parameters ($\beta_{00}, \beta_{01}, \beta_{02}, \beta_{10}, \beta_{11}, \beta_{12}$) and the association parameters $\Omega_{ji} = (\omega_{ji12}, \omega_{ji13}, \omega_{ji23}, \omega_{ji123})^T = (a, a, a, 0)^T (j=0, 1; i=1, \dots, N)$.

The simulated longitudinal binary data were analyzed by the new model, with results given in Figure 1 and Table 2. First, as shown by the maximum LRT, the position of the QTL can be very well estimated even for a small heritability (0.1) (Fig. 1). When the sample size increases from 100 to 400, the QTL can be accurately mapped to a correct position. The parameters, ($\beta_{10}, \beta_{11}, a_1$) for QTL genotype Qq and ($\beta_{00}, \beta_{01}, a_0$) for QTL genotype qq , that define longitudinal trends of the binary trait can be reasonably well estimated. To precisely estimate these parameters, a sample size 400 is suggested when the heritability is modest, say 0.1. It seems that sample size 100 works if the heritability is high, say 0.4. High heritability in plant and animal genetics can be obtained through controlled experiments that minimize the noise. Simulation studies also indicate that the power of detecting a significant QTL is adequately high, 0.80 or high, for a modest heritability (say 0.1) with a small sample size (say 100).

4 Worked Example

We used a simple example to demonstrate the usefulness of our new mapping model. A doubled-haploid (DH) population of 123 lines was derived from semi-dwarf IR64 and tall Azucena in which a genetic linkage map was constructed to cover 12 rice chromosomes (Huang et al. 1997). The DH population was planted in a randomized complete design in the field. Plant height was measured every 10 days from 10 days after transplanted into the field to the date of plant heading. Zhao et al. (2004) used functional mapping to map QTL $\times$ environment interactions for plant height trajectories in the same population. By categorizing DH lines into two groups based on whether plant height is beyond (1) or below the median (0), one obtains a binary response for plant height at multiple time points.

By scanning the rice genome, we identified a QTL located at markers *adh1* on chromosome 11, which is tested to be significant from permutation tests at the chromosome level (Figure

2). Based on Xu and Atchley's (1996) liability, we calculated and found 1% of the total phenotypic variance, explained by this QTL, for plant height treated as a binary traits. In Zhao et al.'s (2004) analysis, chromosome 11 was also found to harbor a significant QTL that affect environment-dependent variation in plant growth trajectories. Although our categorization of plant height into a binary response is to aim to produce a longitudinal binary trait for testing our model, such a treatment seems not to be unreasonable given the DH progeny derived from one tall parent Azucena and other semi-dwarf parent IR64. A single gene, *semi-dwarf-1* (*sd-1*), has been characterized to control rice plant height through gibberellins signaling pathway (Sasaki et al. 2002). The association between the alcohol dehydrogenase marker, *adh1*, and plant height was also detected in a different mapping population, in which Sripongpangkul et al. (2000) identified the genomic region linked to *adh1* that carries QTLs responsible for plant elongation in rice.

5 Discussion

Dynamic changes of complex traits including binary traits are a ubiquitous phenomenon in biology and biomedicine (Rekaya et al. 2003; Hinrichs et al. 2011; Ghahroodi et al. 2010). However, the genetic architecture of dynamic binary traits expressed in a time course has been poorly understood, thus limiting our inference about their developmental regulation in relation to important yield traits in agriculture or disease traits in medicine. In this article, we propose to shed light on the genetic control of dynamic binary traits by developing a statistical model for functional mapping on longitudinal binary response. Previous studies have focused on the genetic mapping of binary traits collected at single time points (Xu and Atchley 1996; Visscher et al. 1996; Yi and Xu 1999a, b, 2000; Kadarmideen et al. 2001; Xu et al. 2003, 2005; Deng et al. 2006; Manichaikul and Broman 2009), or functional mapping of dynamic traits that continuously vary (Ma et al. 2002; Wu and Lin 2006; Li and Wu 2010). The model presented here unifies these two aspects to better address an important genetic issue.

This unification involves a special consideration for modeling the correlation structure of repeated binary measurements, rather than is a simple summation of binary mapping and functional mapping. By making use of Fitzmaurice and Laird's (1993) model, we integrated the association between binary responses in terms of conditional log odds-ratios within the functional mapping framework, allowing the test and characterization of higher-order associations. This parametrization produces the maximum likelihood estimates of the marginal mean parameters that are robust to misspecification of the time dependence, as theoretically proven in Fitzmaurice and Laird (1993). Our simulation studies show that the mapping model by integrating Fitzmaurice and Laird's parameterization provides good accuracy and precision for the estimation of QTL locations and effects on dynamic patterns of binary response.

Statistical modeling and analysis of longitudinal binary responses have received considerable attention owing to the importance of these variables in addressing biological questions (Fitzmaurice and Laird 1993; Lin et al. 2004; Fitzmaurice et al. 2006). The model developed here to map QTLs for longitudinal features of binary traits is a starting point from which new models of various levels of complexity can be developed to handle problems that are closer to reality. Birmingham and Fitzmaurice (2004) discussed a case in which binary responses are measured repeatedly with some subjects who drop out from the trial with a mechanism depending on unobserved responses. Lin et al. (2004) provided an analysis model for longitudinal data with irregular, outcome-dependent follow-up. Many authors made efforts to model multiple longitudinal variables that include either binary responses or a mix of binary and continuous responses (Zeger and Liang 1986; Sammel et al. 1997; Skrondal and Rabe-Hesketh 2004). Although our approach is based on maximum likelihood

approaches, other approaches, like regression models (Haley and Knott 1992), can also be developed to map QTLs for longitudinal binary response. With all these statistical developments, we will be in an excellent position to ask and address fundamental questions about the genetic control of biological processes that arise as presence or absence over time and space.

Acknowledgments

This work is partially supported by NSF/IOS-0923975 and NIH/UL1RR0330184. We thank Dr. Jun Zhu at Zhejiang University for providing his rice data for us to test our binary mapping model.

References

- Birmingham J, Fitzmaurice GM. A pattern-mixture model for longitudinal binary responses with nonignorable nonresponse. *Biometrics*. 2002; 58:989–996. [PubMed: 12495154]
- Broman, KW.; Sen, S. A Guide to QTL Mapping with R/qlt. Springer; New York: 2009.
- Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics*. 1994; 138:963–971. [PubMed: 7851788]
- Deming WE, Stephan FF. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann Math Stat*. 1940; 11:427–444.
- Deng WP, Chen HF, Li ZH. A logistic regression mixture model for interval mapping of genetic trait loci affecting binary phenotypes. *Genetics*. 2006; 172:1349–1358. [PubMed: 16272416]
- Fitzmaurice GM, Laird NM. A likelihood-based method for analysing longitudinal binary responses. *Biometrika*. 1993; 80:141–151.
- Fitzmaurice GM, Lipsitz SR, Ibrahim JG, Gelber R, Lipshultz S. Estimation in regression models for longitudinal binary data with outcome-dependent follow-up. *Biostatistics*. 2006; 7:469–485. [PubMed: 16428260]
- Ghahroodi ZR, Ganjali M, Kazemi I. Models for longitudinal analysis of binary response data for identifying the effects of different treatments on insomnia. *Appl Math Sci*. 2010; 4:3067–3082.
- Hackett CA, Weller JI. Genetic mapping of quantitative trait loci for traits with ordinal distributions. *Biometrics*. 1995; 51:1252–1263. [PubMed: 8589221]
- Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*. 1992; 69:315–324. [PubMed: 16718932]
- Hinrichs D, Bennewitz J, Stamer E, Junge W, Kalm E, Thaller G. Genetic analysis of mastitis data with different models. *J Dairy Sci*. 2011; 94:471–478. [PubMed: 21183058]
- Huang N, Parco A, Mew T, Magpantay G, McCouch S, Guiderdoni E, Xu J, Subudhi P, Angeles ER, Khush GS. RFLP mapping of isozymes, RAPD and QTLs for grain shape, brown planthopper resistance in a doubled haploid rice population. *Mol Breed*. 1997; 3:105–113.
- Kadarmideen HN, Janss LLG, Dekkers JCM. Generalized marker regression and interval QTL mapping methods for binary traits in half-sib family designs. *J Anim Breed Genet*. 2001; 118:297–309.
- Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*. 1989; 121:185–199. [PubMed: 2563713]
- Lange C, Whittaker JC. Mapping quantitative trait loci using generalized estimating equations. *Genetics*. 2001; 159:1325–1337. [PubMed: 11729173]
- Li N, Das K, Wu RL. Functional mapping of human growth trajectories. *J Theor Biol*. 2009; 261:33–42. [PubMed: 19632241]
- Li Q, Huang ZW, Xu M, Wang CG, Gai JY, Huang YJ, Pang XM, Wu RL. Functional mapping of genotype-environment interactions for soybean growth by a semiparametric approach. *Plant Methods*. 2010; 6:13. [PubMed: 20525184]
- Lin H, Scharfstein DO, Rosenheck RA. Analysis of longitudinal data with irregular, outcome-dependent follow-up. *J Roy Stat Soc Ser B*. 2004; 66:791–813.
- Manichaikul A, Broman KW. Binary trait mapping in experimental crosses with selective genotyping. *Genetics*. 2009; 182:863–874. [PubMed: 19416941]

- Rekaya R, Gianola D, Shook G. Longitudinal random effects models for genetic analysis of binary data with application to mastitis in dairy cattle. *Genet Sel Evol.* 2003; 35:457–468. [PubMed: 12939200]
- Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M. Green revolution: A mutant gibberellin-synthesis gene in rice. New insight into the rice variant that helped to avert famine over thirty years ago. *Nature.* 2002; 416:701–702. [PubMed: 11961544]
- Sammel MD, Ryan LM, Legler JM. Latent variable models for mixed discrete and continuous outcomes. *J Roy Stat Soc Ser B.* 1997; 59:667–678.
- Sripongpangkul K, Posa GBT, Senadhira DW, Brar D, Huang N, Khush GS, Li ZK. Genes/QTLs affecting flood tolerance in rice. *Theor Appl Genet.* 2000; 101:1074–1081.
- Skrondal, K.; Rabe-Hesketh, S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Chapman & Hall/CRC; Boca Raton, FL: 2004.
- Visscher PM, Haley CS, Knott SA. Mapping QTLs for binary traits in backcross and F2 populations. *Genet Res.* 1996; 68:55–63.
- Wu, RL.; Ma, CX.; Casella, G. *Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL.* Springer-Verlag; New York: 2007.
- Wu RL, Ma CX, Hou W, Corva P, Medrano JF. Functional mapping of quantitative trait loci that interact with the *hg* gene to regulate growth trajectories in mice. *Genetics.* 2005; 171:239–249. [PubMed: 15965258]
- Wu RL, Cao JG, Huang ZW, Wang Z, Gai JY, Vallejos CE. Systems mapping: How to improve the genetic mapping of complex traits through design principles of biological systems. *BMC Sys Biol.* 2011; 5:84.
- Xu S, Atchley WR. Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics.* 1996; 143:1417–1424. [PubMed: 8807312]
- Xu C, Li Z, Xu S. Joint mapping of quantitative trait loci for multiple binary characters. *Genetics.* 2005; 169:1045–1059. [PubMed: 15489509]
- Xu S, Yi N, Burke D, Galecki A, Miller RA. An EM algorithm for mapping binary disease loci: application to fibrosarcoma in a four-way cross mouse family. *Genetical Research.* 2003; 82:127–138. [PubMed: 14768897]
- Yi N, Xu S. Mapping quantitative trait loci for complex binary traits in outbred populations. *Heredity.* 1999a; 82:668–676. [PubMed: 10383689]
- Yi N, Xu S. A random model approach to mapping quantitative trait loci for complex binary traits in outbred populations. *Genetics.* 1999b; 153:1029–1040. [PubMed: 10511576]
- Yi N, Xu S. Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics.* 2000; 155:1391–1403. [PubMed: 10880497]
- Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986; 42:121–130. [PubMed: 3719049]
- Zhang B, Tong CF, Yin TM, Zhang XY, Zhuge Q, Huang MR, Wang MX, Wu RL. Detection of quantitative trait loci influencing growth trajectories of adventitious roots in *Populus* using functional mapping. *Tree Genet Genom.* 2009; 5:539–552.
- Zhao W, Zhu J, Gallo-Meagher M, Wu RL. A unified statistical model for functional mapping of genotype environment interactions for ontogenetic development. *Genetics.* 2004; 168:1751–1762. [PubMed: 15579721]

Appendix A

A1. Maximum likelihood estimation under the H_1

The derivative of the observed log-likelihood (L^o) of progeny i with respect to β_j and α_j is expressed as

$$\begin{pmatrix} \partial L_i^o / \partial \beta_1 \\ \partial L_i^o / \partial \alpha_1 \\ \partial L_i^o / \partial \beta_0 \\ \partial L_i^o / \partial \alpha_0 \end{pmatrix} = \mathcal{E} \begin{pmatrix} \partial L_i^f / \partial \beta_1 \\ \partial L_i^f / \partial \alpha_1 \\ \partial L_i^f / \partial \beta_0 \\ \partial L_i^f / \partial \alpha_0 \end{pmatrix} \\ = \begin{pmatrix} \mathcal{E}(Q_i) X_i^T \Delta_{1i} V_{1i11}^{-1} (Y_i - \mu_{1i}) \\ \mathcal{E}(Q_i) Z_i^T \{W_i - \nu_{1i} - V_{1i21} V_{1i11}^{-1} (Y_i - \mu_{1i})\} \\ (1 - \mathcal{E}(Q_i)) X_i^T \Delta_{0i} V_{0i11}^{-1} (Y_i - \mu_{0i}) \\ (1 - \mathcal{E}(Q_i)) Z_i^T \{W_i - \nu_{0i} - V_{0i21} V_{0i11}^{-1} (Y_i - \mu_{0i})\} \end{pmatrix}$$

where $\mathcal{E}$ denotes taking expectation with respect to Q_i given Y_i , i.e.,

$\mathcal{E}(Q_i) = E(Q_i | Y_i) = \frac{p_{i1} \exp(\Psi_{i1}^T Y_i + \Omega_{i1}^T W_i - A(\Psi_{i1}, \Omega_{i1}))}{\sum_{j=0}^1 [p_{ij} \exp(\Psi_{ij}^T Y_i + \Omega_{ij}^T W_i - A(\Psi_{ij}, \Omega_{ij}))]}$, $L_i^o = L_i(\theta | Y_i)$ is the observed log-likelihood for progeny i and $L_i^f = L_i(\theta | Y_i, Q_i)$ is the full log-likelihood for this progeny.

Next, we derive the Fisher information matrix under the full log-likelihood function. Let

$$L^f = \sum_{i=1}^N L_i^f,$$

$$\begin{aligned} E \begin{pmatrix} \partial L^f / \partial \beta_1 \\ \partial L^f / \partial \alpha_1 \\ \partial L^f / \partial \beta_0 \\ \partial L^f / \partial \alpha_0 \end{pmatrix} \begin{pmatrix} \partial L^f / \partial \beta_1 \\ \partial L^f / \partial \alpha_1 \\ \partial L^f / \partial \beta_0 \\ \partial L^f / \partial \alpha_0 \end{pmatrix}^T &= \sum_{i=1}^N E \begin{pmatrix} \partial L_i^f / \partial \beta_1 \\ \partial L_i^f / \partial \alpha_1 \\ \partial L_i^f / \partial \beta_0 \\ \partial L_i^f / \partial \alpha_0 \end{pmatrix} \begin{pmatrix} \partial L_i^f / \partial \beta_1 \\ \partial L_i^f / \partial \alpha_1 \\ \partial L_i^f / \partial \beta_0 \\ \partial L_i^f / \partial \alpha_0 \end{pmatrix}^T \\ &= \text{Diagonal} \left[\sum p_{i1} X_i^T \Delta_{1i} V_{1i11}^{-1} \Delta_{1i} X_i, \sum p_{i1} Z_i^T (V_{1i22} - V_{1i21} V_{1i11}^{-1} V_{1i21}^T) Z_i, \right. \\ &\quad \left. \sum p_{i0} X_i^T \Delta_{0i} V_{0i11}^{-1} \Delta_{0i} X_i, \sum p_{i0} Z_i^T (V_{0i22} - V_{0i21} V_{0i11}^{-1} V_{0i21}^T) Z_i \right]. \end{aligned}$$

The MLEs, $(\hat{\beta}_1, \hat{\alpha}_1, \hat{\beta}_0, \hat{\alpha}_0)$, can be obtained using the following general Fisher scoring algorithm:

$$\begin{aligned} \hat{\beta}_1^{(\tau+1)} &= \hat{\beta}_1^{(\tau)} + \left(\sum_{i=1}^N p_{i1} X_i^T \Delta_{1i}^{(\tau)} V_{1i11}^{-1(\tau)} \Delta_{1i}^{(\tau)} X_i \right)^{-1} \times \left\{ \sum_{i=1}^N \mathcal{E}(Q_i)^{(\tau)} X_i^T \Delta_{1i}^{(\tau)} V_{1i11}^{-1} (Y_i - \mu_{1i}^{(\tau)}) \right\}, \\ \hat{\alpha}_1^{(\tau+1)} &= \hat{\alpha}_1^{(\tau)} + \left\{ \sum_{i=1}^N p_{i1} Z_i^T (V_{1i22}^{(\tau)} - V_{1i21}^{(\tau)} V_{1i11}^{-1(\tau)} V_{1i21}^{(\tau)}) Z_i \right\}^{-1} \times \left[\sum_{i=1}^N \mathcal{E}(Q_i)^{(\tau)} Z_i^T \{W_i - \nu_{1i}^{(\tau)} - V_{1i21}^{(\tau)} V_{1i11}^{-1(\tau)} (Y_i - \mu_{1i}^{(\tau)})\} \right], \\ \hat{\beta}_0^{(\tau+1)} &= \hat{\beta}_0^{(\tau)} + \left(\sum_{i=1}^N p_{i0} X_i^T \Delta_{0i}^{(\tau)} V_{0i11}^{-1(\tau)} \Delta_{0i}^{(\tau)} X_i \right)^{-1} \times \left\{ \sum_{i=1}^N (1 - \mathcal{E}(Q_i))^{(\tau)} X_i^T \Delta_{0i}^{(\tau)} V_{0i11}^{-1} (Y_i - \mu_{0i}^{(\tau)}) \right\}, \\ \hat{\alpha}_0^{(\tau+1)} &= \hat{\alpha}_0^{(\tau)} + \left\{ \sum_{i=1}^N p_{i0} Z_i^T (V_{0i22}^{(\tau)} - V_{0i21}^{(\tau)} V_{0i11}^{-1(\tau)} V_{0i21}^{(\tau)}) Z_i \right\}^{-1} \times \left[\sum_{i=1}^N (1 - \mathcal{E}(Q_i))^{(\tau)} Z_i^T \{W_i - \nu_{0i}^{(\tau)} - V_{0i21}^{(\tau)} V_{0i11}^{-1(\tau)} (Y_i - \mu_{0i}^{(\tau)})\} \right]. \end{aligned}$$

Note that, in the iteration procedure, the Fisher information matrix we used is not the Fisher information matrix under the observed log-likelihood but the one under the full data log-likelihood. That is why we called this algorithm the general Fisher scoring algorithm (see Appendix B). The reason we used the full log-likelihood Fisher information matrix is that it is easier and faster to compute than the observed log-likelihood Fisher information matrix.

Evaluation of the scoring equations above requires estimation of the joint probabilities for each progeny. That is, for any μ_{ji} and Ω_{ji} , ν_{ji} , V_{ji11} , V_{ji21} , and V_{ji22} depend on the two-way

and higher-way marginal probabilities. In general, there is no closed form expression representing the joint probabilities as a function of μ_{ji} and Ω_{ji} . But we can use an iterative proportional fitting (IPF) algorithm to estimate the two-way and higher-way marginal probabilities (see Appendix C).

This algorithm does not provide us with an estimate of the asymptotic variance-covariance matrix of $(\hat{\beta}_1, \hat{a}_1, \hat{\beta}_0, \hat{a}_0)$. Yet, we can calculate the Fisher information matrix under the observed log-likelihood using the MLEs obtained from the algorithm. The covariance can be approximated by the inverse of the Fisher information matrix. The sample empirical covariance matrix of the individual scores in any correctly specified model is a consistent estimator of the Fisher information and involves only the first derivatives. Thus, a consistent estimator of the asymptotic variance-covariance matrix of $(\hat{\beta}_1, \hat{a}_1, \hat{\beta}_0, \hat{a}_0)$ is

$$\begin{aligned} \text{var} \begin{pmatrix} \beta_1 \\ \alpha_1 \\ \beta_0 \\ \alpha_0 \end{pmatrix} &= \left[\sum_{i=1}^N \left\{ \begin{pmatrix} \frac{\partial L_i^o}{\partial \beta_1} \\ \frac{\partial L_i^o}{\partial \alpha_1} \\ \frac{\partial L_i^o}{\partial \beta_0} \\ \frac{\partial L_i^o}{\partial \alpha_0} \end{pmatrix} \begin{pmatrix} \frac{\partial L_i^o}{\partial \beta_1} \\ \frac{\partial L_i^o}{\partial \alpha_1} \\ \frac{\partial L_i^o}{\partial \beta_0} \\ \frac{\partial L_i^o}{\partial \alpha_0} \end{pmatrix}^T \right\} \right]^{-1} \\ &= \left[\sum_{i=1}^N \left\{ \mathcal{E} \begin{pmatrix} \frac{\partial L_i^f}{\partial \beta_1} \\ \frac{\partial L_i^f}{\partial \alpha_1} \\ \frac{\partial L_i^f}{\partial \beta_0} \\ \frac{\partial L_i^f}{\partial \alpha_0} \end{pmatrix} \mathcal{E} \begin{pmatrix} \frac{\partial L_i^f}{\partial \beta_1} \\ \frac{\partial L_i^f}{\partial \alpha_1} \\ \frac{\partial L_i^f}{\partial \beta_0} \\ \frac{\partial L_i^f}{\partial \alpha_0} \end{pmatrix}^T \right\} \right]^{-1} \end{aligned}$$

A2. Maximum likelihood estimation under the H_0

Under the H_0 , there is only one component for the data so that there is only one group of parameters (β, a) . We can just let $p_{i1} = 1, \forall i \in 1, \dots, N$, and use the same algorithm as above.

Appendix B

To obtain the MLEs of Θ in observed log-likelihood function $L^o(\Theta|Y)$, a classical Fisher scoring algorithm is iterated as follows:

$$\Theta^{(\tau+1)} = \Theta^{(\tau)} + \left\{ V^{(-1)} \frac{\partial L^o}{\partial \Theta} \right\} \Big|_{\Theta = \Theta^{(\tau)}}$$

where $V = \left[E \left(\frac{\partial L^o}{\partial \Theta} \right) \left(\frac{\partial L^o}{\partial \Theta} \right)^T \right]$.

Actually, replacing V with any V_1 where $V_1 > V$, the algorithm still can work but will need more step to converge. Since $\text{var}(X) = E(XX^T) - E(X)E(X)^T > 0$, we know that this is always true $E(XX^T) > E(X)E(X)^T$. Let $L^f(\Theta|Y, Q)$ denote the full log-likelihood function. If $X = \frac{\partial L^f}{\partial \Theta}$, we have $E \left[\left(\frac{\partial L^f}{\partial \Theta} \right) \left(\frac{\partial L^f}{\partial \Theta} \right)^T \mid Y \right] > E \left(\frac{\partial L^f}{\partial \Theta} \mid Y \right) E \left(\frac{\partial L^f}{\partial \Theta} \mid Y \right)^T$. Taking expectation with respect to Y for both side and applying $\frac{\partial L^o}{\partial \Theta} = E \left(\frac{\partial L^f}{\partial \Theta} \mid Y \right)$, we get $E \left[\left(\frac{\partial L^f}{\partial \Theta} \right) \left(\frac{\partial L^f}{\partial \Theta} \right)^T \right] > E \left[\left(\frac{\partial L^o}{\partial \Theta} \right) \left(\frac{\partial L^o}{\partial \Theta} \right)^T \right]$, a general Fisher scoring algorithm is iterated as follows:

$$\Theta^{(\tau+1)} = \Theta^{(\tau)} + \left\{ V^{(-1)} \frac{\partial L^o}{\partial \Theta} \right\} \Big|_{\Theta = \Theta^{(\tau)}}$$

$$\text{where } V = \left[E \left(\frac{\partial L^f}{\partial \Theta} \right) \left(\frac{\partial L^f}{\partial \Theta} \right)^T \right].$$

Appendix C

The iterative proportional fitting procedure was originally developed by Deming and Stephan (1940) for adjusting the counts in an $r \times c$ contingency table to satisfy marginal totals derived from another source. The classical procedure takes a start table, and multiplies the elements of the table by appropriate scaling factors, sequentially adjusting them until they satisfy the desired set of margins. This iterative procedure can be applied to multidimensional tables and is always convergent.

Thus, given Ω_i , and set $\Psi_i = 0$, we get a start 2^T table, $S_i(\Omega_i)$, which has the specified conditional log odds-ratios. Then we fit the first-order margins μ_i to $S_i(\Omega_i)$. The procedure iterates like this

$$\begin{aligned} Pr(Y_{i1}=j_1, \dots, Y_{i(t-1)}=j_{(t-1)}, Y_{it}=1, Y_{i(t+1)}=j_{(t+1)}, \dots, Y_{iT}=j_{iT})^{(k+1)} &= \\ \frac{Pr(Y_{i1}=j_1, \dots, Y_{i(t-1)}=j_{(t-1)}, Y_{it}=1, Y_{i(t+1)}=j_{(t+1)}, \dots, Y_{iT}=j_{iT})^{(k)} \mu_{it}}{\sum_{j_1, \dots, j_T} Pr(Y_{i1}=j_1, \dots, Y_{i(t-1)}=j_{(t-1)}, Y_{it}=1, Y_{i(t+1)}=j_{(t+1)}, \dots, Y_{iT}=j_{iT})^{(k)}} \\ Pr(Y_{i1}=j_1, \dots, Y_{i(t-1)}=j_{(t-1)}, Y_{it}=0, Y_{i(t+1)}=j_{(t+1)}, \dots, Y_{iT}=j_{iT})^{(k+1)} &= \\ \frac{Pr(Y_{i1}=j_1, \dots, Y_{i(t-1)}=j_{(t-1)}, Y_{it}=0, Y_{i(t+1)}=j_{(t+1)}, \dots, Y_{iT}=j_{iT})^{(k)} (1-\mu_{it})}{\sum_{j_1, \dots, j_T} Pr(Y_{i1}=j_1, \dots, Y_{i(t-1)}=j_{(t-1)}, Y_{it}=0, Y_{i(t+1)}=j_{(t+1)}, \dots, Y_{iT}=j_{iT})^{(k)}} \quad t=1, \dots, T. \end{aligned}$$

This procedure proportionally adjusts the cells of $S_i(\Omega_i)$ until they satisfy the set of margins defined by μ_i . Finally, it will give a 2^T table of cell probabilities, $m_i\{\Omega_i, \mu_i\}$, with margins satisfying μ_i and with conditional log odds-ratios, Ω_i . Using these cell probabilities, we can then calculate updated estimates of v_i , V_{i11} , V_{i21} , and V_{i22} .

Figure 1. LRT profiles for the search of a QTL over the linkage group under different sample sizes (n) and heritabilities (H^2). The 5% critical values are obtained from 200 permutation tests. The arrowed vertical lines indicate the estimated QTL position (the correct position is 25 cM from the left end).

Figure 2.

Log-likelihood ratio profile of QTL detection for plant height through rice chromosome 11. The position of QTL is indicated by the arrow. The horizontal line is the 5% significance chromosome-wise threshold obtained from permutation tests.

Table 1

Conditional probabilities of a putative QTL given the flanking marker genotypes for a backcross population

Marker genotype	Expected frequency	QTL genotype	
		Qq	qq
MN/mn	$(1 - r)/2$	$\frac{(1-r_{MQ})(1-r_{QN})}{(1-r)}$	$\frac{r_{MQ}r_{QN}}{(1-r)}$
Mn/mn	$r/2$	$\frac{(1-r_{MQ})r_{QN}}{r}$	$\frac{r_{MQ}(1-r_{QN})}{r}$
mN/mn	$r/2$	$\frac{r_{MQ}(1-r_{QN})}{r}$	$\frac{(1-r_{MQ})r_{QN}}{r}$
mn/mn	$(1 - r)/2$	$\frac{r_{MQ}r_{QN}}{(1-r)}$	$\frac{(1-r_{MQ})(1-r_{QN})}{(1-r)}$

r_{MQ} , r_{QN} are the recombination fractions between the left marker **M** and the putative QTL, the putative QTL and the right marker **N**, respectively. The recombination fraction between the two flanking markers is r .

MLEs of the parameters that QTL genotype-specific longitudinal binary responses. The root mean square errors (RMSE) of the MLEs calculated from 200 simulation replicates are also given.

Table 2

Model	β_{10}	β_{11}	β_1	β_{00}	β_{01}	α_0	QTL	maxLRT	5% Cut Point
$H^2 = 0.1$									
True	-1	0.15	0.5	0.1	0.18	0.5	25	-	-
MLE($N = 100$)	-0.9231	0.2021	0.4153	0.0149	0.1604	0.6178	20	14.80	12.17
RMSE	0.1851	0.0757	0.1274	0.1919	0.0601	0.0835			
MLE($N = 400$)	-0.9214	0.1177	0.5035	-0.0854	0.2048	0.5212	23	49.58	12.01
RMSE	0.0969	0.0361	0.0578	0.0938	0.0331	0.0478			
$H^2 = 0.4$									
True	-2	0.3	0.5	0.2	0.36	0.5	25	-	-
MLE($N = 100$)	-2.0869	0.3689	0.6332	-0.1854	0.4367	0.1647	25	53.25	12.81
RMSE	0.2563	0.0977	0.1741	0.2598	0.1159	0.1662			
MLE($N = 400$)	1.9836	0.2786	0.3367	0.2586	0.3396	0.5124	25	237.76	14.44
RMSE	0.1146	0.0511	0.1292	0.1376	0.0471	0.0569			