Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2012, Article ID 680634, 12 pages
doi:10.1155/2012/680634

Research Article

A Bayesian Framework for Functional Mapping through Joint Modeling of Longitudinal and Time-to-Event Data

Kiranmoy Das,[1,2,3] Runze Li,[2] Zhongwen Huang,[4,5] Junyi Gai,[4,5] and Rongling Wu[2,3]

[1] *Department of Statistics, Temple University, Philadelphia, PA 19122, USA*
[2] *Department of Statistics, The Pennsylvania State University, University Park, PA 16802, USA*
[3] *Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033, USA*
[4] *Department of Agronomy, Henan Institute of Science and Technology, Xinxiang 453003, China*
[5] *National Center for Soybean Improvement, National Key Laboratory of Crop Genetics and Germplasm Enhancement, Soybean Research Institute, Nanjing Agricultural University, Nanjing 210095, China*

Correspondence should be addressed to Kiranmoy Das, kiranmoy.das@temple.edu

Received 20 January 2012; Revised 13 March 2012; Accepted 19 March 2012

Academic Editor: Pierre Sourdille

Copyright © 2012 Kiranmoy Das et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The most powerful and comprehensive approach of study in modern biology is to understand the whole process of development and all events of importance to development which occur in the process. As a consequence, joint modeling of developmental processes and events has become one of the most demanding tasks in statistical research. Here, we propose a joint modeling framework for functional mapping of specific quantitative trait loci (QTLs) which controls developmental processes and the timing of development and their causal correlation over time. The joint model contains two submodels, one for a developmental process, known as a longitudinal trait, and the other for a developmental event, known as the time to event, which are connected through a QTL mapping framework. A nonparametric approach is used to model the mean and covariance function of the longitudinal trait while the traditional Cox proportional hazard (PH) model is used to model the event time. The joint model is applied to map QTLs that control whole-plant vegetative biomass growth and time to first flower in soybeans. Results show that this model should be broadly useful for detecting genes controlling physiological and pathological processes and other events of interest in biomedicine.

1. Introduction

To study biology, a classic approach is dimension reduction in which a biological phenomenon or process is dissected into several discrete features over time and space. Most efforts in the past decades have been made to understand biological details of individual features and then use knowledge from each feature to draw an inference about biology as a whole. There has been increasing recognition of the limitation of this approach because it fails to detect a rule that governs the transition from one feature to next, thus leading to a significant loss of information behind the development of a biological trait. More recently, tremendous developments in statistics and computer science have enabled scientists to model and compute the dynamic behavior of a biological phenomenon and construct a comprehensive view of how a cell, tissue, or organ grows and develops across the time-space scale.

A statistical dynamic model, called functional mapping, is one of the products of such developments [1, 2]. The merit of functional mapping lies in its biological relevance to study the tempo-spatial pattern of change for the trait and further predict the physiological or pathological status of trait phenotype. Functional mapping has proven to be powerful for elucidating the dynamic genetic architecture of complex phenotypic traits by identifying when specific genes (known as quantitative trait loci or QTLs) involved turn on and turn off and how long they are expressed in a time course. With the advent of new automatic techniques that collect dynamic data in a cost-effective way, functional mapping can be anticipated to play an increasingly important role

in shedding light on the genetic control mechanisms of complex traits or diseases.

The statistical foundation of functional mapping is longitudinal data analysis or functional data analysis. There has been a considerable body of literature on statistical modeling of time-varying mean and covariance structure using various parametric, nonparametric, and semiparametric methods [3–7]. A joint mean-covariance model was proposed by Pourahmadi [8, 9], which shows some advantages over modeling the mean and covariance separately. Since the publication of the pioneering work by Laird and Ware [10], random effects model have been extensively used for longitudinal data analysis [11]. All these statistical approaches have been incorporated into functional mapping [12, 13], aiming to provide the most parsimonious estimates of QTL effects for a given data set. A Bayesian algorithm for functional mapping has been proposed recently by Liu and Wu [14].

The complexity of biology lies in the fact that no biological trait is isolated, rather every trait is affected by other traits through genes and environmental factors. For example, when a plant grows into a particular stage, reproductive behavior, such as flowering, starts to emerge as one of the important events in plant development. The time to first flower is highly associated with the amount of vegetative growth, depending on the environment where the plant is grown. Likewise, the time to recurrence of prostate cancer in humans is related with dynamic changes of prostate specific antigen level. How to jointly model longitudinal and time-to-event data within functional mapping has become an important issue for studying the common genetic basis of these processes and predicting events based on longitudinal traits.

Simultaneous modeling of longitudinal traits and time to events has been an active area in biostatistics during the past twenty years. A linear random effects model and EM estimation approach are proposed by Henderson et al. [15] for joint modeling. Guo and Carlin [16] made a comparative study between separate models and a joint model, showing that a joint model is more powerful when there is a strong correlation between the trait and the event. Wang and Taylor [17] developed a Bayesian method and MCMC algorithm for joint modeling of longitudinal and event time data and applied their algorithm on AIDS data. A review article by Tsiatis and Davidian [18] nicely summarizes the recent developments for such joint modeling.

By simply estimating the correlation between longitudinal traits and event time, Lin and Wu [19] developed a first model that connects these two aspects within functional mapping. However, they developed a likelihood-based framework where the covariance structure for the longitudinal trait was modelled by the known AR(1), and model parameters were estimated using maximum likelihood estimation. Taking advantage of event models, such as semiparametric Cox proportional hazard model, Weibull model, accelerated failure time (AFT) model, we here propose a sophisticated model for joint modeling of longitudinal trait and time to event to locate the QTLs which control the event via a dynamic trait. The detection of those QTLs that are common to these types of traits may help to prevent or accelerate the outcome by genetic approaches. Our model is constructed with a Bayesian paradigm and model parameters are estimated by the MCMC algorithm. Local polynomials are used to model the mean trajectory and generalized-linear-model- (GLM-) based approach is used to model the covariance matrix. The model is validated using a real example in which whole-plant biomass as a longitudinal trait measured at a series of discrete time points and the time to first flower as a time-to-event are jointly modeled through functional mapping. The statistical properties of the model applied to estimate QTL temporal effects in this example and its practical usefulness are investigated by simulation studies.

2. Joint Modeling Framework

2.1. Model for the Longitudinal Trait. Genetic mapping should be based on a segregating population, such as the backcross, F_2, or recombinant inbred lines (RILS), initiated with two inbred lines each carrying an alternative allele. An RIL population is generated by self-crossing the hybrids of the two inbred lines continuously for 7-8 successive generations, which leads to two homozygous genotypes for alternative alleles at each locus. Methods for other designs can be derived similarly. Suppose a backcross has n progeny which is genotyped to construct a linkage map, aiming at locating putative QTLs that trigger significant effects on a longitudinal trait and its associated event. For each progeny, the trait is measured repeatedly at T different time points and a time-to-event is also recorded. At a specific time point t, the phenotypic value of the trait for progeny i affected by a putative QTL can be expressed by a linear model as follows:

$$y_i(t) = \sum_{j=1}^{2} x_{ij} u_j(t) + r_i(t) + e_i(t), \tag{1}$$

where x_{ij} is an indicator variable for a possible QTL genotype of progeny i and defined as 1 if a particular QTL genotype j is indicated and 0 otherwise ($j = 1$ for QTL genotype QQ and 2 for genotype qq), $u_j(t)$ is the mean phenotypic value of QTL genotype j for progeny i at time t, $r_i(t)$ is the subject specific random effect, and $e_i(t)$ is the residual error assumed to follow a normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}$.

The central theme of functional mapping is to model the mean and covariance structures for the longitudinal trait efficiently. Here, we model the mean vector by polynomial function and the covariance matrix by an approach that guarantees the positive definiteness of the estimated covariance matrix. Without loss of generality, assume the response vector for progeny i, $\mathbf{y}_i = (y_i(1), \ldots, y_i(T))$, has mean 0 and covariance matrix $\mathbf{\Sigma}$. The response at time t, $y_i(t)$, can be predicted by its predecessors as the follows:

$$y_i(t) = \sum_{t'=1}^{t-1} \phi_{t,t'} y_i(t') + \epsilon_i(t), \tag{2}$$

where $\phi_{t,t'}$ is the corresponding regression coefficient, $\epsilon_i(t)$ is the prediction error for progeny i with mean $= 0$,

and $\sigma^2(t)$ is its variance. Assuming that $\epsilon_i(t)$'s are uncorrelated (Pourahmadi [8]), we get $\text{cov}(\boldsymbol{\epsilon}_i) = D$, a diagonal matrix with $\sigma^2(t)$ being the tth diagonal element, where $\boldsymbol{\epsilon}_i = (\epsilon_i(1), \ldots, \epsilon_i(T))'$ is the vector of prediction errors. Hence, the matrix representation of the above autoregression becomes

$$\boldsymbol{\epsilon}_i = \mathbf{M}\mathbf{y}_i, \tag{3}$$

where $\mathbf{M}$ is a lower triangular matrix with 1's in diagonal elements and $-\phi_{t,t'}$ in the (t, t')th position. The above equation simply gives

$$\text{cov}(\boldsymbol{\epsilon}_i) = \mathbf{M}\,\text{cov}(\mathbf{y}_i)\mathbf{M}^T = \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}^T = \mathbf{D}, \tag{4}$$

which is related to the modified Cholesky decomposition of Σ [20].

Equation (4) will be considered as the basis for modeling the covariance structure, since this guarantees the estimated covariance matrix to be positive definite. Following Pourahmadi [8, 9], we model the mean vector, unconstrained variance parameters $\log\sigma^2(t)$, and dependence parameter $\phi_{t,t'}$, using a polynomial function of a particular order, expressed as

$$u_j(t) = \beta_{j0} + \beta_{j1}t + \beta_{j2}t^2 + \cdots + \beta_{jr}t^r, \tag{5}$$

$$r_i(t) = \theta_{i0} + \theta_{i1}t + \theta_{i2}t^2 + \cdots + \theta_{im}t^m, \tag{6}$$

$$\log\sigma_t^2 = \eta_0 + \eta_1 t + \eta_2 t^2 + \cdots + \eta_g t^g, \tag{7}$$

$$\phi_{t,t'} = \delta_0 + \delta_1(t-t') + \delta_2(t-t')^2 + \cdots + \delta_h(t-t')^h, \quad (t' = 1, 2, \ldots, t-1). \tag{8}$$

The optimal (r, m, g, h) is determined from the information criteria (AIC/BIC). We note that different genotypes are assumed to have the same covariance structure but different means. Note that the above method of modeling the covariance structure for a longitudinal response is more robust than the traditional first-order autoregressive (AR(1)) or compound symmetry (CS) structure since real data might not show a parametric dependence structure. We refer to the proposed approach as GLM-based approach to estimate the covariance matrix.

Denote $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jr})$ for QTL genotype j and $\boldsymbol{\theta}_i = (\theta_{i0}, \theta_{i1}, \ldots, \theta_{im})$ for subject i. Then, the conditional mean function for progeny i carrying QTL genotype j ($j = 1$ or 2) for the given subject specific random effect ($\boldsymbol{\theta}_i$) can be expressed as

$$\boldsymbol{\mu}_{ij} = \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T + \mathbf{X}_i^{(m)}\boldsymbol{\theta}_i^T, \tag{9}$$

where

$$\mathbf{X}_i^{(r)} = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & \cdots & t_{i1}^r \\ 1 & t_{i2} & t_{i2}^2 & \cdots & t_{i2}^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{i\tau} & t_{i\tau}^2 & \cdots & t_{i\tau}^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{iT} & t_{iT}^2 & \cdots & t_{iT}^r \end{bmatrix}. \tag{10}$$

Assume the vectors of subject specific random effects $\boldsymbol{\theta}_i$ follow m-variate normal distribution with mean 0 and covariance matrix $\sigma^2 I_m$ and they are independent of the residual errors. Note that under this assumption, $\mathbf{y}_i|\boldsymbol{\theta}_i$ will follow $\text{MVN}(\mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T + \mathbf{X}_i^{(m)}\boldsymbol{\theta}_i^T, \Sigma)$, and the marginal distribution of $\mathbf{y}_i$ will be $\text{MVN}(\mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T, \Sigma + \sigma^2\mathbf{X}_i^{(m)}\mathbf{X}_i^{(m)T})$.

2.2. Model for the Event Time. We use s_i to denote the event time of progeny i. Since in the current situation, the event time is recorded for all progeny; no progeny is censored. Assuming a Cox proportional hazard model for this event, we get for progeny i,

$$\lambda_i(t) = \lambda_0(t)\exp\left(\gamma\mu_{ij}(t)\right), \tag{11}$$

where $\lambda_0(t)$ denotes the baseline hazard at time t, $\mu_{ij}(t)$ is the mean longitudinal trait at time t for given $\boldsymbol{\theta}_i$ when progeny i is of QTL genotype j and the regression coefficient γ represents the effect of the trait on the event time. The survival function for progeny i can be expressed in terms of the hazard function as $s_i(t) = \exp\left[-\int_0^t \lambda_i(u)du\right]$.

The longitudinal model described above is linked to the hazard model by γ. If $\gamma = 0$, then the event is independent of the trait, and hence we should better fit separate models for the trait and the event. However, when γ is different from zero, a joint model performs better than the separate models [17]. For simplicity, the baseline hazard is assumed to be a step function, $\lambda_0(t) = \lambda_{0k}$ over a partition of the observed time scale $[0, \max(s_i)]$ into K (possibly evenly spaced) intervals, $(t_0^\lambda = 0, t_1^\lambda, t_2^\lambda, \ldots, t_K^\lambda)$. The value of K is usually not too large, possibly smaller than 10.

2.3. Likelihood for the Joint Model. Since the QTL genotype of a progeny is unknown, we use a mixture model to describe the likelihood of the progeny in terms of its possible underlying QTL genotypes [21]. The joint likelihood of unknown parameters $\boldsymbol{\Theta}$ given the longitudinal trait $\mathbf{y} = (\mathbf{y}_i)_{i=1}^n$ and event time $\mathbf{s} = (\mathbf{s}_i)_{i=1}^n$ for all n progeny can be expressed as

$$\begin{aligned} L(\boldsymbol{\Theta} \mid \mathbf{y}, \mathbf{s}) &= \prod_{i=1}^n \left(\sum_{j=1}^2 \omega_{j|i}[\pi(\mathbf{y}_i, s_i \mid Q_i = j, \boldsymbol{\theta}_i)] \right) \\ &= \prod_{i=1}^n \left(\sum_{j=1}^2 \omega_{j|i} \left\{ [f(\mathbf{y}_i \mid Q_i = j, \boldsymbol{\theta}_i)] \right.\right. \\ &\qquad \left.\left. \times \left[\lambda_i(s_i)\exp\left(-\int_0^{s_i} \lambda_i(u)du \right) \right] \right\} \right), \end{aligned} \tag{12}$$

where $\pi(\cdot)$ denotes the joint density of the longitudinal trait and event time; $f(.)$ denotes a multivariate normal with QTL genotype-specific mean $\boldsymbol{\mu}_{ij}$ modeled as (9) and covariance matrix $\boldsymbol{\Sigma}$; hazard function $\lambda_i(s_i)$ is modeled as (11); and $\omega_{j|i}$ denotes the conditional probability of QTL genotype j given that the marker information of projeny i and Q_i is the QTL genotype for the i-th subject.

The QTL genotype is inferred from marker genotypes of the linkage map. Let $M_i = (M_{i1}, \ldots, M_{im})$ be the m-marker

genotypes for progeny i, D^* the position of the putative QTL measured by its distance from the very first marker of an ordered linkage group, and D_k the distances between marker 1 and k. Assume that the QTL is located between marker k and $k+1$. Then, the conditional probability of QTL genotype j given the genotype of these two markers that flank the QTL is expressed as

$$\omega_{j|i} = \text{Prob}\left(Q_i = j \mid D^*, M_{ik}, M_{i(k+1)}, D_k, D_{k+1}\right). \quad (13)$$

Note that, given the QTL locations D^*, D_k, and D_{k+1}, one can compute d_1, the distance of the QTL from marker k and d_2, the distance of the QTL from marker $k+1$ [22]. Using the Haldane map function, one can compute recombination fractions between marker k and QTL (r_1), between QTL and marker $k+1$ (r_2), and between markers k and $k+1$ (r) as follows:

$$r_1 = \frac{1}{2}\left(1 - e^{-2d_1}\right), \qquad r_2 = \frac{1}{2}\left(1 - e^{-2d_2}\right), \qquad r = \frac{1}{2}\left(1 - e^{-2d}\right), \quad (14)$$

where $d = D_{k+1} - D_k$ is the distance between marker k and $k+1$. Wu et al. [22] provide a procedure for deriving the conditional probabilities of QTL genotypes given marker interval genotypes for the backcross, F_2, and RIL populations, respectively.

Unknown parameters $\boldsymbol{\Theta}$ in likelihood (12) contain QTL genotype-specific parameters $\boldsymbol{\beta}_j$, σ^2, the parameters that model the variance structure and dependence structure $\boldsymbol{\eta} = (\eta_0, \eta_1, \ldots, \eta_g)$ and $\boldsymbol{\delta} = (\delta_0, \delta_1, \ldots, \delta_h)$, as shown in models (7) and (8), respectively, the effect of the longitudinal trait on the event γ, QTL position D^* and the baseline hazards λ_{0k}.

2.4. Posterior Distribution and Sampling Procedure. We derive a Bayesian approach for estimating the unknown parameters. This will first need to specify the prior distributions for the parameters and, given the data and the priors, derive the posterior distribution over all the unknown parameters. For β_j, we place a multivariate normal prior with zero mean and covariance matrix Σ_β. An inverse gamma prior with parameters (α_1,α_2) is considered for σ^2. We consider a uniform prior on γ and uniform $(0, D_m)$ prior for the parameter D^*. Independent Gamma (a, b) priors are taken for λ_{0k}, for $k = 1, \ldots, K$. Priors for $\boldsymbol{\eta}$ and $\boldsymbol{\delta}$ are taken as MVN(0,Σ_η) and MVN(0,Σ_δ), respectively.

With the above priors and likelihood function, we have the joint posterior distribution for the parameters. In this case, it is quite straightforward to get the full conditional posterior distributions. Assume that the priors are independent for different parameters. Thus, we get the posterior density of β, σ^2, $\boldsymbol{\gamma}$, D^*, $\boldsymbol{\eta}$, $\boldsymbol{\delta}$, $\boldsymbol{\lambda_0}$ as

$$\pi(\beta, \sigma^2, \gamma, D^*, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\lambda_0} \mid \mathbf{y}, \mathbf{s}) \propto \pi(\mathbf{y}, \mathbf{s} \mid \boldsymbol{\beta}, \sigma^2, \gamma, D^*, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\lambda_0}) \times \pi(\beta)\pi(\sigma^2)\pi(\gamma)\pi(D^*)\pi(\boldsymbol{\eta})\pi(\boldsymbol{\delta})\pi(\boldsymbol{\lambda_0}). \quad (15)$$

Assuming that priors for different genotypes are independent, we can express the above posterior distribution as

$$\pi(\beta, \sigma^2, \gamma, D, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\lambda_0} \mid \mathbf{y}, \mathbf{s}) \propto \pi(\mathbf{y}, \mathbf{s} \mid \boldsymbol{\beta}, \sigma^2, \gamma, D^*, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\lambda_0}) \times \left[\prod_{j=1}^{2} \pi\left(\boldsymbol{\beta}_j\right)\right] \pi(\sigma^2)\pi(\gamma)\pi(D^*)\pi(\boldsymbol{\eta})\pi(\boldsymbol{\delta})\pi(\boldsymbol{\lambda_0}). \quad (16)$$

The full conditional distributions for the model parameters, as derived in the Appendix, are used to estimate the parameters using the MCMC algorithm. Note that the full conditional distribution for β_j is expressed as a product of a normal distribution term which comes from the longitudinal trait and two other terms from the hazard model. To update β_j, therefore, we use a Metropolis-Hastings (MH) algorithm with a normal proposal density since it is a part of its posterior distribution. We also note that in the full conditionals of η and δ, normal distribution coming from the longitudinal part of the data is the main determinant. Hence for η and δ, we consider normal proposals with the current value of the parameter as the mean and covariance matrix as $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\delta$, respectively. Selection of a good proposal density for γ is a bit tricky and we follow the recommendation given by Wang and Taylor [17]. By evaluating several choices for a good proposal, we consider a normal distribution with mean as the current state of the parameter and a suitable standard deviation in such a way that the proposed density gets well mixed with the target distribution (acceptance rate between 0.25 to 0.40). Because of conjugacy, we can directly simulate from the full conditional of λ_{0k}'s.

The parameter D^* which specifies the location of the QTL is updated following the idea of Satagopan et al. [23] by using the MH algorithm. A new value of D^*, which we denote by $D^{*\text{new}}$, is generated from Uniform $(\max(0, D^* - \psi), \min(D^* + \psi, D_m))$, where ψ is the tuning parameter. Denote this proposed distribution by $q(D^*, D^{*\text{new}})$. The proposed value will be considered as the new value of the chain with probability

$$\alpha(D^*, D^{*\text{new}}) = \min\left[1, \frac{\pi(D^{*\text{new}} \mid y, s, \beta, \gamma, \eta, \delta, \lambda_0) q(D^{*\text{new}}, D^*)}{\pi(D^* \mid y, s, \beta, \gamma, \eta, \delta, \lambda_0) q(D^*, D^{*\text{new}})}\right]. \quad (17)$$

We note that $\pi(D^{*\text{new}} \mid \ldots) \propto \pi(Q \mid D^{*\text{new}}, M)\pi(D^{*\text{new}}) = \prod_{i=1}^{n} \pi(Q_i \mid D^{*\text{new}}, M_i)\pi(D^{*\text{new}})$ and, similarly, $\pi(D^* \mid \ldots) = \prod_{i=1}^{n} \pi(Q_i \mid D^*, M_i)\pi(D^*)$.

Because of the independence among n progeny, Q is updated by updating for each Q_i separately. For each progeny i, the full conditional density is in the form of a multinomial with the following cell probabilities:

$$\Pi(Q_i = j \mid y_i, s_i) = \frac{\pi(Q_i = j)\pi(y_i, s_i \mid Q_i = j)}{\sum_{j'=1}^{2} \pi(Q_i = j')\pi(y_i, s_i \mid Q_i = j')}. \quad (18)$$

We can sample the QTL genotype directly from this full conditional density at each cycle. Details of the estimation procedure can be found in Satagopan et al. [23].

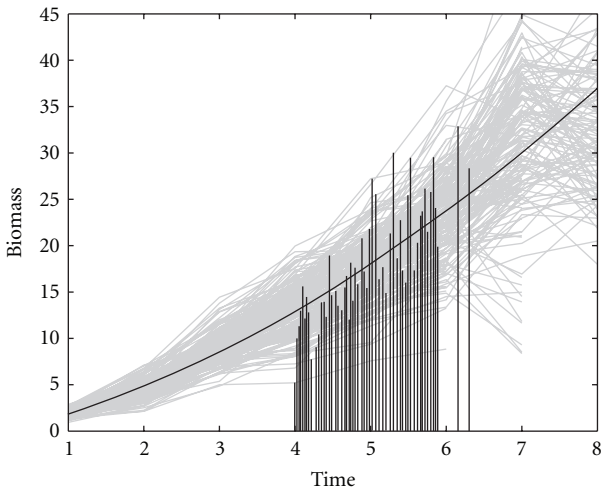

Figure 1: Whole-plant biomass growth trajectories for 184 soybean RILs. The time to first flower is indicated by a vertical line on each biomass growth curve. The black curve is the mean growth trajectory.

3. Application

3.1. Material and Analysis. The new model was applied to analyze a real data set for QTL mapping in soybeans. The mapping population contains 184 RILs derived from two cultivars, Nannong 1138-2 and Kefeng no. 1. A genetic linkage map of this population was first established by Zhang et al. [24] with 452 makers including RFLP, SSR, EST distributed among the 21 linkage group. This map was recently updated by adding some new SSR makers and dumping some unreliable markers. The new map contains 834 molecular makers covering a length of 2,308 cM in 24 linkage groups, with an average genetic distance of 2.85 cM between adjacent markers. Those markers with missing information were excluded from the analysis, leading to a total of 780 markers involved in our analysis.

The plants and their parents were grown in a sample lattice design with two replicates at Jiangpu Soybean Experiment Station, Nanjing Agricultural University, China. After 20 days of seedling emergence, plant biomass (in gms.) were measured once every 5–10 days until most plants stopped height growth. A total of 8 measurements were taken for the biomass and the time to get the first flower in that growing season was also recorded for each plant. Figure 1 shows the raw data, both the trait (biomass) and the event time for 184 plants.

Prior distributions for the model parameters were taken as follows. For genotype specific fixed effect β_j, a multivariate normal prior was used with zero mean and a diagonal covariance matrix with all diagonal elements 100. For σ^2, we took $IG(3,1)$ prior which has mean $= 0.5$ with small variance (0.25). A uniform prior $U(-3.0,-0.1)$ was taken for γ following Wang and Taylor [17]. Observed time scale for the event time data was partitioned into 5 parts; that is, we took $K = 5$ and considered independent gamma (0.04, 1.0) priors for $\lambda_{01},\ldots,\lambda_{05}$, following Wang and Taylor [17]. Uniform prior was taken for D^*. For η and δ, we considered multivariate normal priors with zero means and diagonal covariance matrices with diagonal elements 30 and 20, respectively. To investigate the effect of prior distributions on the estimation method, we did a sensitivity analysis. Considering different sets of priors, we fitted our model many times and it turned out the estimation is almost insensitive to the choice of priors. Hence the choice of our priors (even with huge variances for β, η, and δ) does not affect much the estimation of the model parameters.

We fitted our joint model as described in Section 2, by MCMC sampling. We ran chains 120,000 times. To remove the effect of the starting values, we excluded first 20,000 burn-in iterations. With the remaining 100,000 iterations, we estimated the posterior distributions and the parameters were estimated by the posterior mean and also calculated the sample standard deviations for the posterior densities. For the MH algorithm, the acceptance rates were 0.31, 0.26, 0.25, and 0.30 for σ^2, γ, η, and δ, respectively.

Since our model is complex, we perform several standard diagnostic tests to assess the convergence of the Markov chains. First, we use the method proposed by Brooks and Gelman [25]. Considering five different chains with different starting points and discarding the burn-in iterations, we computed multivariate potential scale reduction factors (MPSRF) to assess the convergence of the chains. Starting points for the model parameters were drawn from the respective priors. The computed values of this statistic get stabilized near 1 after 60,000 iterations for our model parameters, which indicates convergence of the chains.

Second, we perform Geweke test which compares the earlier part of the markov chain to the later part for assessing convergence. After deleting the burn-in iterations, from the remaining 100,000 iterations, we take out two subsequences; the first 50,000 and the last 50,000 iterations. Also consistent spectral density estimates at zero frequency are calculated to compute the z-scores. The calculated Pvalues for our model parameters are above 0.18, which indicates a small absolute z-scores assessing the convergence of our chains. A detailed discussion of this method can be found in Geweke [26].

Finally, we perform the Heidelberger and Welch test as proposed by Heidelberger and Welch [27]. This test has two parts: a stationary test and a half-width test. Our chains after deletion of the burn-in iterations, pass the stationary test. To assess whether the number of iterations is adequate to estimate the parameters accurately, we calculate relative half-width (RHW). We consider the default α value (0.05) and predetermined tolerance value is taken as 0.1. For all our model parameters, the calculated RHW is in between 0.045 and 0.081. This indicates that we have enough iterations to estimate the model parameters with 95% confidence under tolerance of 0.1.

3.2. Results. Figure 1 is the growth trajectories of whole-plant biomass over time for all RILs, in which the times to first flower for each RIL are also indicated. There are great variability found for these two traits among RILs. In the exploratory data analysis, we computed BIC values for different orders for the triplet (r,m,g,h), showing that

Table 1: BIC values for selecting the optimum (r, m, g, h) to model the mean-covariance structures for whole-plant biomass growth trajectories and the time to first flower in soybeans.

(r, m, g, h)	BIC
$(3, 3, 3, 2)$	7.18
$(3, 2, 3, 2)$	4.15
$(3, 3, 2, 2)$	4.26
$(3, 2, 2, 2)$	3.91
$(2, 2, 2, 2)$	**2.03**
$(2, 3, 2, 2)$	2.87
$(2, 2, 3, 2)$	3.01
$(2, 3, 3, 2)$	4.26
$(3, 3, 3, 3)$	4.66
$(3, 2, 3, 3)$	3.17
$(3, 3, 2, 3)$	3.93
$(3, 2, 2, 3)$	4.05
$(2, 2, 2, 3)$	3.34
$(2, 3, 2, 3)$	5.16
$(2, 2, 3, 3)$	4.88
$(2, 3, 3, 3)$	3.16

the BIC value is smallest for the order $(2, 2, 2, 2)$, that is, second order polynomials can best fit the the mean, variance, and dependence structures (Table 1). By scanning for the existence of QTLs over the genetic linkage map, we obtained the posterior distribution of the model parameters and estimate marginal posterior distributions of the QTL locations (D^*) for all 24 linkage groups (Figure 2).

We observed posterior peaks in linkage groups 1, 4, 15, 19, 20, 21, and 23. To draw inference about the existence of a putative QTL in each of these groups, we computed the Bayes Factor (BF), defined as

$$\mathrm{BF} = \frac{P(\mathbf{Y}, \mathbf{S} \mid \kappa = 0)}{P(\mathbf{Y}, \mathbf{S} \mid \kappa = 1)}, \tag{19}$$

where κ denotes the number of QTLs in that particular group. Following Jeffrey's scale, the BFS with value smaller than 1 gives strong evidence against the null hypothesis and higher than 10 gives enough evidence for the null hypothesis. Note that in this case, we are testing the existence of no QTL (null) versus the existence of one QTL (alternative) for each linkage group. Here, no QTL in a group means $\beta_1 = \beta_2$, for that group. So, in order to compute BF we run our MCMC twice, first under the null and then under the alternative. The calculated BF for the above 7 linkage groups were 0.2781, 11.274, 13.493, 10.610, 0.5913, 11.475, and 0.7953, respectively, implying the existence of QTL in groups 1, 20, and 23. Table 2 provides genotype-specific mean parameters for the QTLs located on linkage groups 1, 20, and 23, along with their 95 percent credible intervals (C.I.).

Since the nature of our model is complex and our estimation is based on MCMC, we perform posterior predictive check for the aforementioned 7 linkage groups. We simulate observations to get the posterior predictive distribution. Let

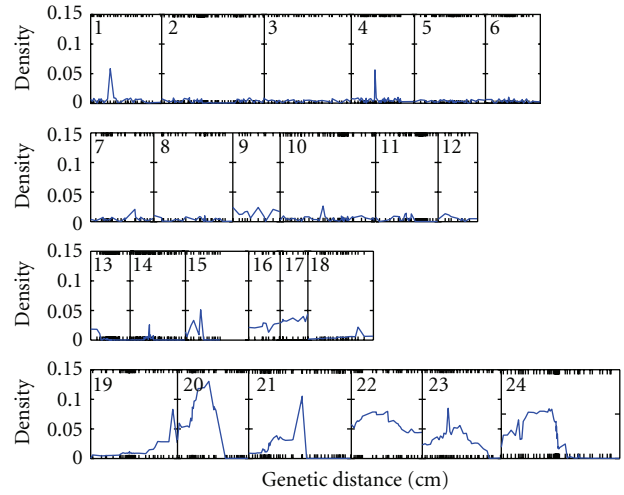

Figure 2: Marginal posterior plot for QTL locations over 24 linkage groups. Marker locations are indicated by ticks on the x-axis.

Θ be the set of all model parameters and $\mathbf{D}^{\mathrm{rep}} = (\mathbf{y}^{\mathrm{rep}}, \mathbf{s}^{\mathrm{rep}})$ be the replicated data. Then given the data $\mathbf{D} = (\mathbf{y}, \mathbf{s})$, the posterior predictive distribution of $\mathbf{D}^{\mathrm{rep}}$ is given by $p(\mathbf{D}^{\mathrm{rep}} \mid \mathbf{D}) = \int p(\mathbf{D}^{\mathrm{rep}} \mid \Theta) p(\Theta \mid \mathbf{D}) d\Theta$.

One can simulate from the posterior predictive distribution using the following two steps. First from the posterior distributions of the model parameters, simulate m values (vectors) of Θ. Next for each value of Θ, simulate a value (vector) $\mathbf{D}^{\mathrm{rep}}$ from the likelihood. The m values (vectors) of $\mathbf{D}^{\mathrm{rep}}$ drawn in this way will essentially come from posterior predictive distribution $p(\mathbf{D}^{\mathrm{rep}} \mid \mathbf{D})$.

We simulate 100 draws (m=100) from the posterior predictive distribution and then apply proposed joint analysis and estimate the model parameters using MCMC as described earlier. We compute BF for each of those 7 linkage groups by considering the problem of testing the existence of no QTL (null) versus the existence of one QTL (alternative). Table 3 shows the average BF (with estimated SE) for each linkage group and the existence of QTL in groups 1, 20, and 23 is quite evident.

Heritability (broad-sense) for the traits is estimated from the data, as the proportion of phenotypic variance attributable to genetic variance. The estimated heritability in our case is 32.6%. Also we compute the percentage of variance explained by three identified QTLs. It turns out the QTLs identified in linkage groups 1, 20, and 23 explain 6.8%, 14.3% and 11.4% of the total variance, respectively.

We show the marginal posterior plots with 95% credible intervals for the parameter γ in Figure 3. Note that for all three groups, the estimates and the confidence intervals are in the negative part which indicates a negative relationship between the trait and the event time. Biologically this is sensible since the plants with higher body mass will take less time to have the first flower compared to the plants with lower body masses.

Figure 4 illustrates genotypic differences in whole-plant biomass trajectory and the time to first flower for the three

Table 2: Estimates of the parameters that describe genotype-specific biomass growth trajectories and QTL locations on linkage groups 1, 20, and 23, with 95% credible intervals.

Parameter	Group 1		Group 20		Group 23	
	estimate	C.I.	estimate	C.I.	estimate	C.I.
β_{10}	−0.4762	(−0.5081, −0.4442)	−1.1524	(−1.1641, −1.1405)	0.4190	(0.3897, 0.4484)
β_{11}	3.3214	(3.3023, 3.3404)	2.6829	(2.6463, 2.7194)	0.5971	(0.5566, 0.6377)
β_{12}	0.1548	(0.1363, 0.1731)	0.2295	(0.2020, 0.2570)	0.5438	(0.5176, 0.5701)
β_{20}	−5.4762	(−5.4788, −5.4734)	−3.1004	(−3.0231, −2.9768)	−2.3571	(−2.3701, −2.3442)
β_{21}	8.4464	(8.4381, 8.4546)	5.3250	(5.3159, 5.3340)	4.4660	(4.4294, 4.5028)
β_{22}	−0.4702	(−0.5011, −0.4392)	−0.0250	(−0.0618, 0.0118)	0.0410	(0.0390, 0.0432)
Marker Interval	Sat-356–B30T		GNE097b–A199H		LC4-4T–Sat-280	
D^*	30.810	(29.1934, 31.7150)	49.600	(48.7515, 50.0726)	39.472	(38.8143, 40.1863)

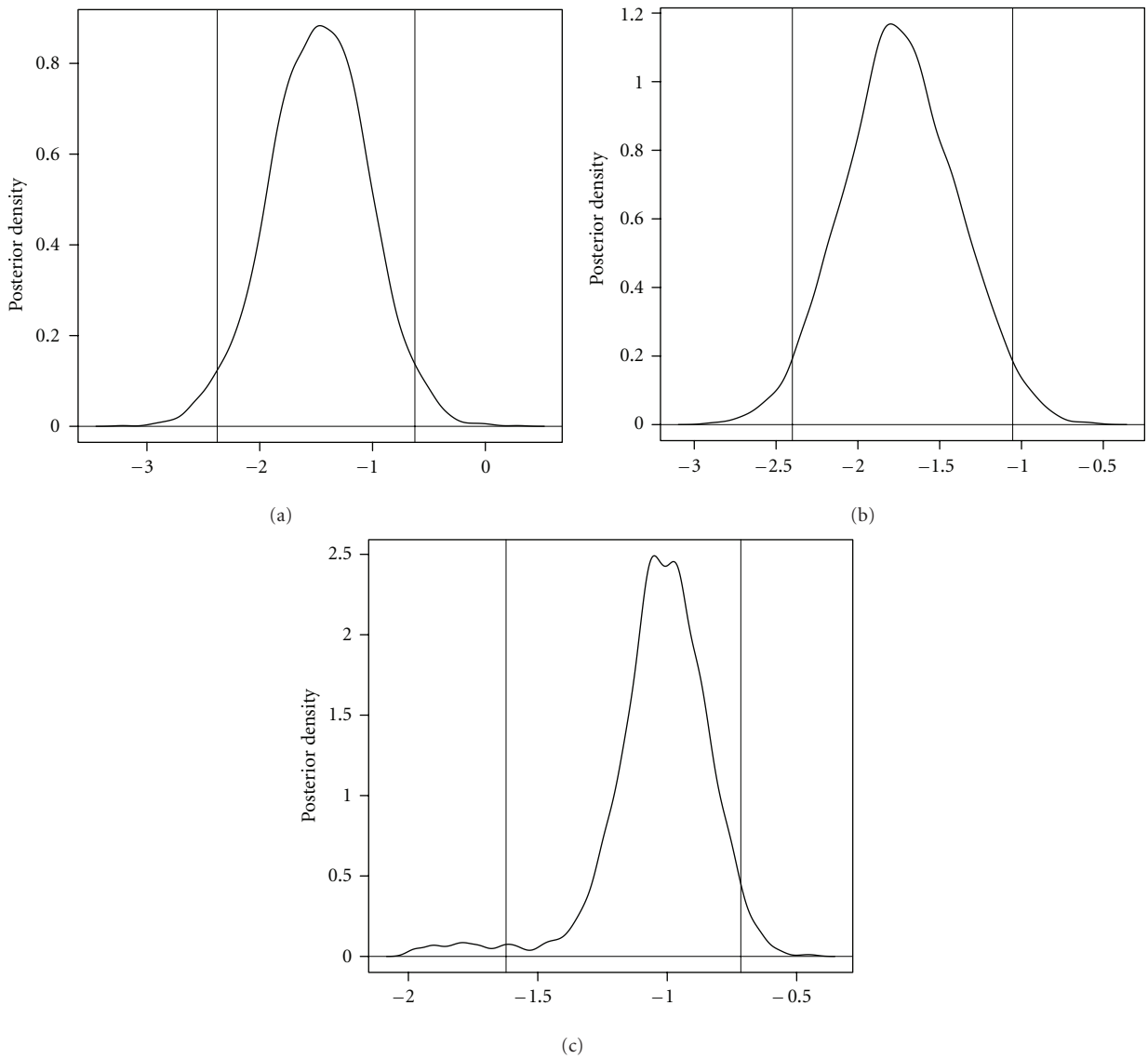

Figure 3: Marginal posterior plot for γ for linkage groups 1 (a), 20 (b), and 23 (c).

Table 3: Posterior predictive check for 7 linkage groups.

Linkage group	BF (from actual data)	BF with SE (from posterior predictive distribution)
1	0.2781	0.2659 (0.15)
4	11.274	12.086 (1.76)
15	13.493	12.962 (2.17)
19	10.610	11.138 (1.56)
20	0.5913	0.6281 (0.73)
21	11.475	12.183 (1.18)
23	0.7953	0.7682 (0.89)

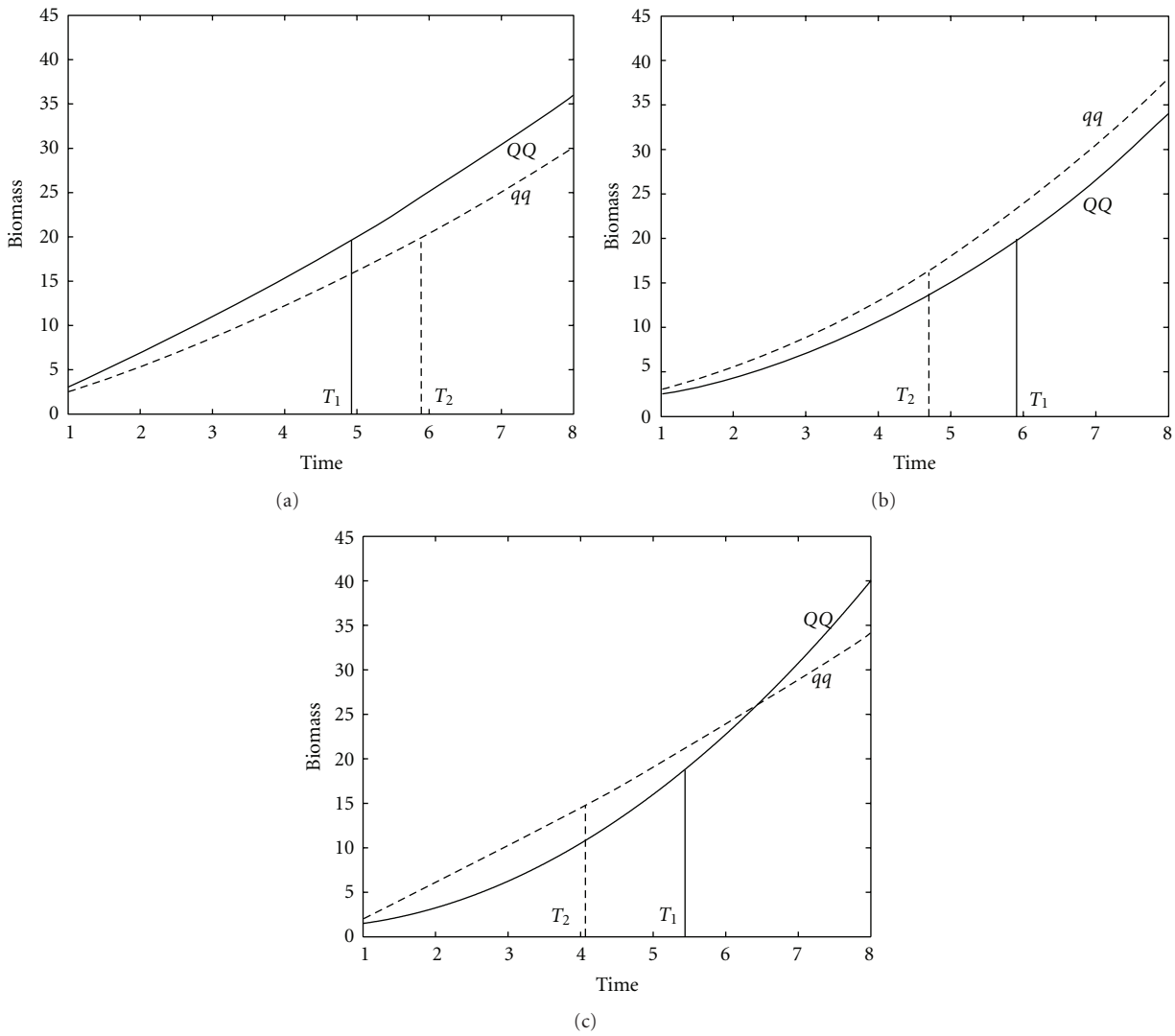

Figure 4: Whole-plant biomass growth trajectories and times to first flower (T_1 and T_2) for two different genotypes at each of the QTLs detected on linkage groups 1 (a), 20 (b), and 23 (c). Genotypes *QQ* inherit two alleles from parent Nannong 1138-2, whereas genotype *qq* inherits two alleles from parent Kefeng no. 1.

QTLs detected. At the QTL on linkage group 1, the allele (*Q*) inherited from parent Nannong 1138-2 leads to increased biomass growth and earlier flowering than the allele (*q*) from parent Kefeng no. 1. The inverse pattern is observed for the QTL on linkage group 20. The QTL on linkage group 23 alters its direction of genetic effect. Affected by the first two QTLs, poor vegetative biomass growth in plants stimulates early flowering (Figures 4(a) and 4(b)). Yet, the QTL on linkage group 23 makes the fast-growing genotype to flower earlier than the slow-growing genotype (Figure 4(c)).

Table 4: Simulation results for genotypic-mean parameters and QTL locations under different covariance structures, AR(1)-, CS- and GLM-based approach.

Parameter		AR(1)		CS		GLM-based approach	
	Actual value	Estimate	MCSE	Estimate	MCSE	Estimate	MCSE
β_{10}	−1.1524	−1.1872	0.0871	−1.0982	0.1094	−1.1667	0.0361
β_{11}	2.6829	2.5391	0.0502	2.6103	0.0495	2.7001	0.0103
β_{12}	0.2295	0.2288	0.0805	0.2301	0.0302	0.2291	0.0113
β_{20}	−3.1004	−3.1204	0.0307	−3.093	0.1025	−3.1255	0.1011
β_{21}	5.3250	5.3140	0.0291	5.2998	0.1130	5.3433	0.0405
β_{22}	−0.0250	−0.0241	0.1302	−0.0257	0.1035	−0.0244	0.0603
D^*	43.00	39.32	2.3561	40.61	1.5694	42.48	1.1572

4. Simulation Study

We performed simulation studies to study the statistical properties of the joint model. We assumed an RIL design of 200 progeny and simulated 11 evenly spaced markers on a linkage group of length 100 cM. A QTL is located at 43 cM from the very first marker of the linkage group. To reflect a practical problem, we used parameter estimates of the soybean QTL detected in linkage group 20 as true values to simulate the data, allowing the covariance structure. Time-dependent phenotypic values were assumed to follow a multivariate normal distribution and the event times were taken the same as the soybean data. To make a comparison, we analyzed the simulated data using our nonparametric GLM-based covariance structure and the traditional AR(1) and CS covariance structures.

The prior distributions for the model parameters were taken in the same way as discussed in Section 3.1. A uniform prior on (0,100) was considered for D^*. For each situation, we ran Markov chains 120,000 iterations and initial 20,000 burn-in iterations were discarded. Model parameters were estimated from the posterior distributions on the basis of remaining 100,000 iterations. The computed BF was 0.649 giving a strong evidence against the null hypothesis. Table 4 shows the means of the Bayesian estimates of model parameters with their respective Monte Carlo standard errors, (MCSE). It can be seen that the estimates are quite close to the actual values with a reasonably small standard errors which justifies the accuracy and precision of our estimation procedure. However, our GLM-based approach provides better estimation of parameters than AR(1) and CS-based approaches.

Figure 5 elucidates the marginal posterior plot for the QTL location under three different covariance structures. It is found that both AR(1)- and CS-based models provide the peaks at wrong locations, whereas GLM-based nonparametric covariance structure locates QTL more accurately in which case the length of the credible interval is narrower than those obtained from the former two structures. This provides numerical evidence that the proposed GLM-based model has better precision of QTL localization.

We perform further simulation studies to assess the reliability of BF in our data application. For each of the linkage groups 1, 20, and 23, we simulate data under the "null" model. As mentioned earlier, under the "null" model, $\beta_1 = \beta_2$. For group 1, we consider the null model $\beta_1 = \beta_2 = (-0.4762,\ 3,3214,\ 0.1548)$, for group 20 it is $\beta_1 = \beta_2 = (-1.1524,\ 2.6829,\ 0.2295)$, and for group 23, our null model is given as $\beta_1 = \beta_2 = (0.4190,\ 0.5971,\ 0.5438)$. The computed BFS for these three groups are 34.55, 46.79, and 41.86, respectively, suggesting strong evidence for the null.

5. Discussion

Tools to reveal the secret of life should reflect the dynamic nature of life. More recently, a series of statistical models have been developed to map quantitative trait loci (QTLs) that control the dynamic process of a complex trait [1, 2, 22, 28]. These so-called functional mapping models integrate mathematical aspects of biological processes into a statistical framework derived to map complex trait QTLs and have proved to be useful for detecting and identifying genes and genetic interactions involved in quantitative genetic variation for plant height, plant rooting ability, and animal body mass. Functional mapping is also flexible to incorporate complex biological phenomena, such as genotype-environment interactions and allometric scaling providing powerful means for addressing biological questions of fundamental importance.

In this paper, we develop a new version of functional mapping that can map QTLs for developmental events affected by organismic growth trajectories in time. This version is benefited from recent statistical developments for joint modeling of longitudinal traits and event time [16–18, 29]. In the presence of strong correlation between a longitudinal trait and event, a joint model performs better than submodels separately for a single trait. In the joint modeling framework for these two types of traits, we applied a GLM-based approach to model the covariance structure and local polynomials for the mean curves. Bayesian estimation method using the MCMC algorithm was used since it is computationally much simpler than a likelihood-based approach. Simulation results show the effectiveness of GLM-based covariance model compared to traditional parametric compound symmetry or autoregressive structure.

Our joint model, embedded within functional mapping, promotes the study of testing how QTLs pleiotropically

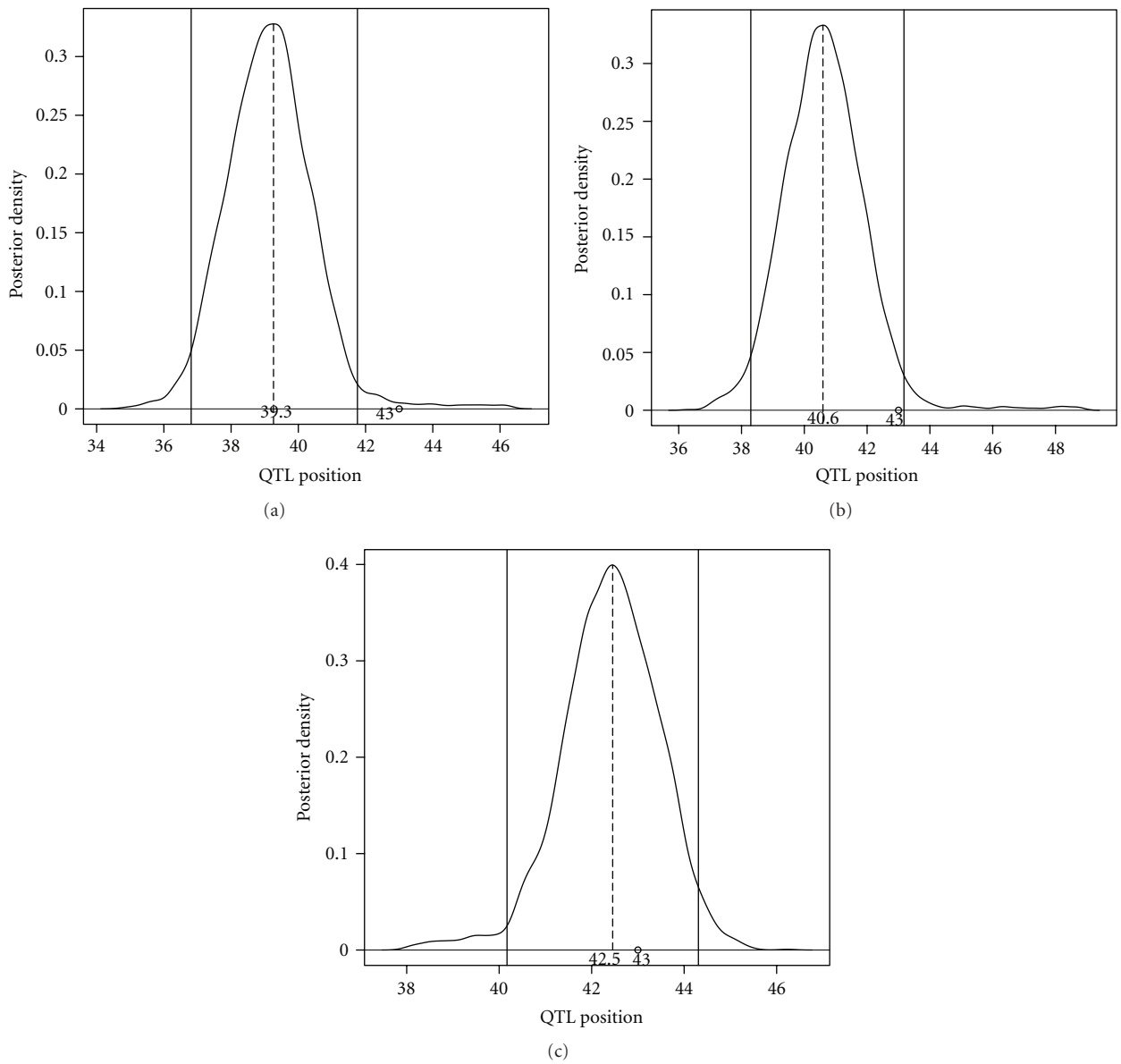

Figure 5: The Bayesian estimate of QTL location (indicated by dash vertical lines) from simulation studies under different covariance structures, AR(1) (a), CS (b), and nonparametric (c). The true QTL location is at 43 cM from the very first marker of the linkage group.

affect different biological processes and how one trait is predicted by other traits through genetic information. The application of the new model to soybean mapping data does not only validate its usefulness and utilization, but also gains new insight into the genetic and developmental regulation of trait correlations in plants. There is no doubt that the new model can be modified to study the genetic associations between HIV dynamics and the time to death as well as prostate specific antigen change and the time to recurrence of prostate. However, there is much room for modifying this model. First, to clearly describe our idea, we assume one QTL at a time for trait control. Epistatic interactions between multiple QTLs may play an important role in trait development as well as in correlations between longitudinal traits and events. Second, from a dynamic systems perspective, we need to model dynamic correlations among multiple longitudinal traits and multiple events. Third, with the availability of efficient genotyping techniques, our model should accommodate a high-dimension model selection scheme to identify significant genetic variants from a flood of marker data.

Appendix

Denote $\beta_{-j} = (\beta_{j'}\colon j' = 1, 2; j' \neq j)$, $S(t) = \exp(-\int_0^t \lambda(u)du)$, $h(t) = (1, t, t^2, \ldots, t^r)$, and $g(t) = (1, t, t^2, \ldots, t^m)$. Let m_j be the number of progeny that carries QTL genotype j. We

derive the full conditional distributions for the unknown parameters as

$$\begin{aligned}
&\pi\left(\boldsymbol{\beta}_j \mid \mathbf{y}, \mathbf{s}, \sigma^2, \gamma, D, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\lambda}_0, \boldsymbol{\beta}_{-j}\right) \\
&\propto \pi\left(\boldsymbol{\beta}_j\right)\pi(\mathbf{y}, \mathbf{s} \mid \boldsymbol{\beta}, \sigma^2, \gamma, D, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\lambda}_0) \\
&\propto \pi\left(\boldsymbol{\beta}_j\right)\pi(\mathbf{s} \mid \mathbf{y}, \boldsymbol{\beta}, \sigma^2, \gamma, D, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\lambda}_0) \\
&\times \pi(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \gamma, D, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\lambda}_0) \\
&\propto \exp\left[-\frac{1}{2}\boldsymbol{\beta}_j\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}_j^T - \frac{1}{2}\sum_{i=1}^{m_j}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right)^T \right. \\
&\qquad \left. \times\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right)\right] \times \left[\exp\left(\gamma\sum_{i=1}^{m_j} h_i(s_i)\boldsymbol{\beta}_j^T\right)\right] \\
&\quad \times \left[\prod_{i=1}^{m_j} S_i(s_i)\right] \\
&= \exp\left[-\frac{1}{2}\boldsymbol{\beta}_j\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}_j^T - \frac{1}{2}\sum_{i=1}^{m_j}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right)^T \right. \\
&\qquad \left. \times\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right)\right] \times \left[\exp\left(\gamma\sum_{i=1}^{m_j} h_i(s_i)\boldsymbol{\beta}_j^T\right)\right] \\
&\quad \times \left[\prod_{i=1}^{m_j} S_i(s_i)\right] \\
&\propto \exp\left[\boldsymbol{\beta}_j\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}_j^T + \sum_{i=1}^{m_j}\left(\boldsymbol{\beta}_j\mathbf{X}_i^{(r)T}\boldsymbol{\Sigma}^{-1}\mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right) \right. \\
&\qquad \left. -2\sum_{i=1}^{m_j}\left(\mathbf{y}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right)\right]\left[\exp\left(\gamma\sum_{i=1}^{m_j} h_i(s_i)\boldsymbol{\beta}_j^T\right)\right] \\
&\quad \times \left[\prod_{i=1}^{m_j} S_i(s_i)\right] \\
&= \exp\left[\boldsymbol{\beta}_j\left(\Sigma_\beta^{-1} + \sum_{i=1}^{m_j}\mathbf{X}_i^{(r)T}\boldsymbol{\Sigma}^{-1}\mathbf{X}_i^{(r)}\right)\boldsymbol{\beta}_j^T - 2\sum_{i=1}^{m_j}\left(\mathbf{y}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right)\right] \\
&\quad \times \left[\exp\left(\gamma\sum_{i=1}^{m_j} h_i(s_i)\boldsymbol{\beta}_j^T\right)\right]\left[\prod_{i=1}^{m_j} S_i(s_i)\right].
\end{aligned} \tag{A.1}$$

Hence, we have

$$\begin{aligned}
\pi\left(\boldsymbol{\beta}_j \mid .\right) &: \mathrm{MVN}\left[\beta_j, M_\beta, V_\beta\right] \\
&\times \left[\exp\left(\gamma\sum_{i=1}^{m_j} h_i(s_i)\boldsymbol{\beta}_j^T\right)\right]\left[\prod_{i=1}^{m_j} S_i(s_i)\right],
\end{aligned} \tag{A.2}$$

where

$$\begin{aligned}
M_\beta &= \left[\left(\boldsymbol{\Sigma}_\beta^{-1} + \sum_{i=1}^{m_j}\mathbf{X}_i^{(r)T}\boldsymbol{\Sigma}^{-1}\mathbf{X}_i^{(r)}\right)^{-1}\left(\sum_{i=1}^{m_j}\mathbf{X}_i^{(r)T}\boldsymbol{\Sigma}^{-1}\mathbf{y}_i\right)\right], \\
V_\beta &= \left(\boldsymbol{\Sigma}_\beta^{-1} + \sum_{i=1}^{m_j}\mathbf{X}_i^{(r)T}\boldsymbol{\Sigma}^{-1}\mathbf{X}_i^{(r)}\right)^{-1}.
\end{aligned} \tag{A.3}$$

Similarly, the full conditional distributions for the other parameters can be derived as follows:

$$\begin{aligned}
\pi(\gamma \mid \cdot) &\propto \left[\exp\left(\gamma\sum_{j=1}^{2}\sum_{i=1}^{m_j}\left(h_i(s_i)\boldsymbol{\beta}_j^T + g_i(s_i)\boldsymbol{\theta}_i^T\right)\right)\right] \\
&\quad \times \left[\prod_{i=1}^{n} S_i(s_i)\right]. \\
\pi(\eta \mid \cdot) &\propto \left[\prod_{i=1}^{n}\det(\boldsymbol{\Sigma})^{-(1/2)}\right]\mathrm{f} \\
&\quad \times \exp\left[-\frac{1}{2}\eta'\boldsymbol{\Sigma}_\eta^{-1}\eta - \frac{1}{2}\sum_{j=1}^{2}\sum_{i=1}^{m_j}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T \right.\right. \\
&\qquad \left.\left. \times - \mathbf{X}_i^{(m)}\boldsymbol{\theta}_i^T\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T - \mathbf{X}_i^{(m)}\boldsymbol{\theta}_i^T\right)\right]. \\
\pi(\boldsymbol{\delta} \mid \cdot) &\propto \left[\prod_{i=1}^{n}\det(\boldsymbol{\Sigma})^{-(1/2)}\right] \\
&\quad \times \exp\left[-\frac{1}{2}\delta'\boldsymbol{\Sigma}_\delta^{-1}\delta - \frac{1}{2}\sum_{j=1}^{2}\sum_{i=1}^{m_j}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T \right.\right. \\
&\qquad \left.\left. \times - \mathbf{X}_i^{(m)}\boldsymbol{\theta}_i^T\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T - \mathbf{X}_i^{(m)}\boldsymbol{\theta}_i^T\right)\right]. \\
\pi(\sigma^2 \mid \cdot) &\propto \left[\prod_{i=1}^{n}\det\left(\boldsymbol{\Sigma} + \sigma^2\mathbf{X}_i^{(m)}\mathbf{X}_i^{(m)T}\right)^{-(1/2)}\right] \\
&\quad \times (\sigma^2)^{-\alpha_1-1}\exp\left(-\frac{\alpha_2}{\sigma^2}\right) \\
&\quad \times \exp\left[-\frac{1}{2}\sum_{j=1}^{2}\sum_{i=1}^{m_j}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right)^T \right. \\
&\qquad \left. \times\left(\boldsymbol{\Sigma} + \sigma^2\mathbf{X}_i^{(m)}\mathbf{X}_i^{(m)T}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i^{(r)}\boldsymbol{\beta}_j^T\right)\right]. \\
\pi(\lambda_{0k} \mid \cdot) &\propto \mathrm{Gamma}\left[d_k + 1, \phi_{\lambda_{0k}}^{-1}\right]\pi(\lambda_{0k}) \\
&= \mathrm{Gamma}\left(a + d_k + 1, \left[\phi_{\lambda_{0k}} + \frac{1}{b}\right]^{-1}\right),
\end{aligned} \tag{A.4}$$

where d_k is the number of events in the interval $(t_{k-1}^\lambda, t_k^\lambda]$ and

$$\begin{aligned}
\phi_{\lambda_{0k}} &= \sum_{i:s_i \geq t_k^\lambda}\int_{t_{k-1}^\lambda}^{t_k^\lambda}\exp\left[\gamma\left(h(t)\beta_j^T + g(t)\theta_i^T\right)\right]dt \\
&\quad + \sum_{i:t_{k-1}^\lambda < s_i \leq t_k^\lambda}\int_{t_{k-1}^\lambda}^{s_i}\exp\left[\gamma\left(h(t)\beta_j^T + g(t)\theta_i^T\right)\right]dt.
\end{aligned} \tag{A.5}$$

Acknowledgments

This work is partially supported by NSF/IOS-0923975, Changjiang Scholars Award, "Thousand-Person Plan" Award, the China National Key Basic Research Program (2006CB1017, 2009CB1184, 2010CB125906), the China National Hightech R&D Program (2006AA100104), the

Natural Science Foundation of China (30671266), and the China MOE 111 Project (B08025). R.Li was supported by an NIDA Grant P50-DA10075 and an NNSF of China Grant 11028103. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA, or the NIH.

References

[1] C. X. Ma, G. Casella, and R. Wu, "Functional mapping of quantitative trait loci underlying the character process: a theoretical framework," *Genetics*, vol. 161, no. 4, pp. 1751–1762, 2002.

[2] R. Wu and M. Lin, "Opinion: functional mapping—how to map and study the genetic architecture of dynamic complex traits," *Nature Reviews Genetics*, vol. 7, no. 3, pp. 229–237, 2006.

[3] J. Fan and R. Li, "New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis," *Journal of the American Statistical Association*, vol. 99, no. 467, pp. 710–723, 2004.

[4] J. Fan, T. Huang, and R. Li, "Analysis of longitudinal data with semiparametric estimation of covariance function," *Journal of the American Statistical Association*, vol. 102, no. 478, pp. 632–641, 2007.

[5] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York, NY, USA, 2nd edition, 2005.

[6] F. Yao, H. G. Müller, and J. L. Wang, "Functional data analysis for sparse longitudinal data," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 577–590, 2005.

[7] F. Yao, H. G. Muller, and J. L. Wang, "Functional linear regression analysis for longitudinal data," *Annals of Statistics*, vol. 33, no. 6, pp. 2873–2903, 2005.

[8] M. Pourahmadi, "Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation," *Biometrika*, vol. 86, no. 3, pp. 677–690, 1999.

[9] M. Pourahmadi, "Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix," *Biometrika*, vol. 87, no. 2, pp. 425–435, 2000.

[10] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, vol. 38, no. 4, pp. 963–974, 1982.

[11] M. Davidian and D. M. Giltinan, "Nonlinear models for repeated measurement data: an overview and update," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 8, no. 4, pp. 387–419, 2003.

[12] J. S. Yap, J. Fan, and R. Wu, "Nonparametric modeling of longitudinal covariance structure in functional mapping of quantitative trait loci," *Biometrics*, vol. 65, no. 4, pp. 1068–1077, 2009.

[13] K. Das, J. Li, Z. Wang et al., "A dynamic model for genome-wide association studies," *Human Genetics*, vol. 129, no. 6, pp. 629–639, 2011.

[14] T. Liu and R. L. Wu, "A bayesian algorithm for functional mapping of dynamic complex traits," *Algorithms*, vol. 2, pp. 667–691, 2009.

[15] R. Henderson, P. Diggle, and A. Dobson, "Joint modelling of longitudinal measurements and event time data," *Biostatistics*, vol. 1, pp. 465–480, 2000.

[16] X. Guo and B. P. Carlin, "Separate and joint modeling of longitudinal and event time data using standard computer packages," *American Statistician*, vol. 58, no. 1, pp. 16–24, 2004.

[17] Y. Wang and J. M. G. Taylor, "Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 895–905, 2001.

[18] A. A. Tsiatis and M. Davidian, "Joint modeling of longitudinal and time-to-event data: an overview," *Statistica Sinica*, vol. 14, no. 3, pp. 809–834, 2004.

[19] M. Lin and R. Wu, "A joint model for nonparametric functional mapping of longitudinal trajectory and time-to-event," *BMC Bioinformatics*, vol. 7, article 138, 2006.

[20] H. J. Newton, *IMESLAB: A Time Series Analysis Laboratory*, T. Wadsworth and Brooks/Coles, Pacific Grove, Calif, USA, 1988.

[21] E. S. Lander and S. Botstein, "Mapping mendelian factors underlying quantitative traits using RFLP linkage maps," *Genetics*, vol. 121, no. 1, p. 185, 1989.

[22] R. Wu, W. Hou, Y. Cui et al., "Modeling the genetic architecture of complex traits with molecular markers," *Recent Patents on Nanotechnology*, vol. 1, no. 1, pp. 41–49, 2007.

[23] J. M. Satagopan, B. S. Yandell, M. A. Newton, and T. C. Osborn, "A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo," *Genetics*, vol. 144, no. 2, pp. 805–816, 1996.

[24] W. K. Zhang, Y. J. Wang, G. Z. Luo et al., "QTL mapping of ten agronomic traits on the soybean (Glycine max L. Merr.) genetic map and their association with EST markers," *Theoretical and Applied Genetics*, vol. 108, no. 6, pp. 1131–1139, 2004.

[25] S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 434–455, 1998.

[26] J. Geweke, "Evaluating the accuracy of sampling-based approaches to calculating posterior moments," in *Bayesian Statistics*, vol. 4, Clarendon Press, Oxford, UK, 1992.

[27] P. Heidelberger and P. D. Welch, "Simulation run length control in the presence of an initial transient," *Operations Research*, vol. 31, no. 6, pp. 1109–1144, 1983.

[28] R. Wu, C. X. Ma, M. Lin, Z. Wang, and G. Casella, "Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model," *Biometrics*, vol. 60, no. 3, pp. 729–738, 2004.

[29] D. Rizopoulos, "Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data," *Biometrics*, vol. 67, no. 3, pp. 819–829, 2011.