Theor Appl Genet (2011) 123:483–492
DOI 10.1007/s00122-011-1601-0

ORIGINAL PAPER

Bayesian functional mapping of dynamic quantitative traits

Runqing Yang · Jiahan Li · Xin Wang · Xiaojing Zhou

Received: 16 April 2010 / Accepted: 26 April 2011 / Published online: 15 May 2011

Abstract Without consideration of other linked QTLs responsible for dynamic trait, original functional mapping based on a single QTL model is not optimal for analyzing multiple dynamic trait loci. Despite that composite functional mapping incorporates the effects of genetic background outside the tested QTL in mapping model, the arbitrary choice of background markers also impact on the power of QTL detection. In this study, we proposed Bayesian functional mapping strategy that can simultaneously identify multiple QTL controlling developmental patterns of dynamic traits over the genome. Our proposed method fits the change of each QTL effect with the time by Legendre polynomial and takes the residual covariance structure into account using the first autoregressive equation. Also, Bayesian shrinkage estimation was employed to estimate the model parameters. Especially, we specify the gamma distribution as the prior for the first-order autoregressive coefficient, which will guarantee the convergence of Bayesian sampling. Simulations showed that the proposed method could accurately estimate the QTL parameters and had a greater statistical power of QTL detection than the composite functional mapping. A real data analysis of leaf age growth in rice is used for the demonstration of our method. It shows that our Bayesian functional mapping can detect more QTLs as compared to composite functional mapping.

Communicated by F. van Eeuwijk.

R. Yang (✉)
College of Animal Science and Veterinary Medicine, Heilongjiang Bayi Agricultural University, 163319 Daqing, People's Republic of China
e-mail: runqingyang@sjtu.edu.cn

R. Yang · J. Li · X. Wang
School of Agriculture and Biology, Shanghai Jiaotong University, 200240 Shanghai, People's Republic of China

X. Zhou
Department of Mathematics, Heilongjiang Bayi Agricultural University, 163319 Daqing, People's Republic of China

Introduction

Dynamic traits, the phenotypes of which change with time or the quantitative factors, are often encountered in the plant, animal and human research fields. From the point of view of classical quantitative genetics, researchers found that not only the dynamic point, but also the dynamic pattern is inherited (Kirkpatrick and Heckman 1989; Kirkpatrick et al. 1990; Schaeffer 2004). As a result, genetic analysis of dynamic traits can be conducted by fitting a dynamic pattern to the phenotypic values across time points and analyzing the fitted parameters of the dynamic trajectory under the theory of multivariate analysis. Subsequently, random regression model (Henderson 1982; Schaeffer 2004) has been developed to dissect the dynamic phenotype into different functions that describe the effects of various genetic and environmental factors. Theoretically speaking, these methods are also suitable for locating dynamic trait loci by regarding the QTL effects as whole genetic effects in linecross (Yang et al. 2006, 2007; Yang and Xu 2007) or a part of them in pedigree population (Macgregor et al. 2005). Similar to this idea, Wu and his colleagues (Ma et al. 2002; Wu et al. 2002, 2003, 2004b) have developed a functional mapping strategy.

The central idea of functional mapping is that the mean vectors of QTL genotypes in a time interval are modeled by a biologically meaningful mathematical equation, whereas the residual covariance matrix is modeled in terms of its

time series autocorrelation structure (Ma et al. 2002). From a biological viewpoint, functional mapping is advantageous because it incorporates fundamental principles behind biological processes or networks into a QTL mapping framework. Several attempts have been made to embed various mathematical and differential equations in the mapping framework to unravel dynamic changes of genetic control over a biological or medical trait given its developmental nature. For example, the Richards growth equation has been used to model growth trajectories for various biological entities and bi-exponential curves to approximate HIV-1 dynamics (Perelson et al. 1996). Emax models are proposed to fit drug response curve over a range of concentrations (Giraldo 2003), and power curve is derived to quantify the power required by a bird to fly at a particular speed (Tobalske et al. 2003). The parameters in these equations have particular biological or clinical means. Thus, the results from functional mapping are biologically more interpretable and meaningful than those from general approaches without biological incorporation. Also, through individual or joint testing for the curves parameters in functional mapping, many biologically interesting questions can be answered and addressed at the interplay between genetic actions and developmental patterns (Wu et al. 2004a). Functional mapping attempts to estimate mathematical parameters that model the time-dependent vector of QTL effects and residual covariance matrix rather than estimate all elements in the vector and matrix and, therefore, it has the statistical property of strong robustness due to estimating fewer parameters (Ma et al. 2002). As functional mapping has been constructed within the context of simple interval mapping, it has not considered the situation of separating multiple linked QTL for a dynamic trait. So far, Yang et al. (2007) have presented a composite functional mapping framework, which capitalizes on the strengths of functional mapping and composite interval mapping (Zeng 1994) to take multiple QTL into account. Their method used a flexible nonparametric approach based on Legendre polynomials to model the time-dependent genetic background.

The functional mapping based on either interval mapping or composite interval mapping is built mainly on a single QTL model (Jansen and Stam 1994; Lander and Botstein 1989; Zeng 1994) and therefore is not optimal to detect multiple QTLs governing dynamic traits. Although composite functional mapping took into account the effects of genetic background outside the QTL interval, the arbitrary choice of background marker will affect statistical power of QTL detection. Moreover, functional mapping has not been reported under the Bayesian framework. In this study, we implement Legendre polynomial to model the change of each QTL effect with time, and the first autoregressive equation to construct residual covariance matrix to propose a multiple QTL method to map for dynamic traits. Bayesian shrinkage estimation method was adopted to estimate the parameters in the model. Both computer simulation and analysis of a real data set were used to demonstrate our method.

Methods

Genetic model for functional mapping

In the mapping population including only two segregating genotypes at any locus, such as a backcross (BC), double-haploid lines (DHLs) or recombinant inbred lines (RILs), phenotypes of repeated measurements in time interval $[t_0, t_m]$ are collected and molecular markers with known linkage map are genotyped for n individuals. Assume that there are q quantitative trait loci responsible for the changing trajectory of dynamic traits; the phenotypic value $y_i(t)$ of individual i measured at time t can be then described by the following multiple QTL model

$$y_i(t) = \mu(t) + \sum_{j=1}^{q} x_{ij}\beta_j(t) + \varepsilon_i(t) \quad (1)$$

where, $\mu(t)$ is the population mean at time t and $\beta_j(t)$ for $j = 1, 2, \ldots, q$ is the effect of the jth QTL at time point t. Variable x_{ij} is a genotype indicator variable for individual i at locus j and defined as 1 for one genotype and -1 for the other genotype, and ε_i is a random environmental error distributed as $N(0, \sigma^2(t))$.

Utilizing the flexibility and linear additivity of polynomials, we choose the Legendre polynomial of p orders to fit changing trajectories of the population mean and effect of each QTL. Let $\psi(t)$ be the basis of the Legendre polynomial (Yang et al. 2006) and stipulate that $\mu(t) = \psi(t)\mu$ and $\beta_j(t) = \psi(t)\beta_j$, model (1) is then rewritten as

$$y_i(t) = \psi(t)\mu + \sum_{j=1}^{q} x_{ij}\psi(t)\beta_j + \varepsilon_i \quad (2)$$

where, each one of μ and β_j is a vector of $p + 1$ dimensions.

Further, we denote $y_i = [\, y_i(t_0) \quad y_i(t_1) \quad \ldots \quad y_i(t_m)]^T$ as $(m + 1) \times 1$ column vector for the repeated measurements of the dynamic traits and define $\psi = [\, \psi^T(t_0) \quad \psi^T(t_1) \quad \cdots \quad \psi^T(t_m)\,]$ as $(p + 1) \times (m + 1)$ matrix. In matrix notation, model (2) becomes

$$y_i = \psi^{\mathrm{T}}\mu + \sum_{j=1}^{q} x_{ij}\psi^T\beta_j + \varepsilon_i \quad (3)$$

where $\varepsilon_i = [\, \varepsilon_{i0} \quad \cdots \quad \varepsilon_{im}\,]^T$ is a $(m + 1) \times 1$ vector for the residual errors with a multivariate normal distribution

$N(0, \sum_e)$. Following the idea of functional mapping, a series of stationary and non-stationary time series models can be used to structure the residual covariance $\sum_e$; the simplest choice is the first-order auto-regressive model with

$$\Sigma_e = \begin{bmatrix} 1 & \rho^{t_1-t_0} & \cdots & \rho^{t_m-t_0} \\ \rho^{t_1-t_0} & 1 & \cdots & \rho^{t_{m-1}-t_0} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{t_m-t_0} & \rho^{t_{m-1}-t_0} & \cdots & 1 \end{bmatrix} \sigma_0^2 = \Sigma_\rho \sigma_0^2, \quad (4)$$

where ρ is the first-order auto-regressive coefficient. In the residual covariance structure, obviously, the residual variance is the same for the trait at each time interval and the covariance between different measurements decays in a proportion of ρ ($0 < \rho < 1$) purely with time interval.

Bayesian shrinkage estimation of QTL parameters

Based on Bayes' theorem, given the observed data (Y) the joint posterior distribution of the parameters (B) can be calculated by

$$p(B|Y) \propto p(Y|B)p(B)$$

In mapping analysis of dynamic traits, the observed data are phenotypes $y = \{y_i\}$ for $i = 1, 2, \cdots, n$ and marker information (M), and the parameters include population mean μ, QTL regression effects $\beta = \{\beta_j\}$, QTL positions $\lambda = \{\lambda_j\}$, prior covariance matrices of QTL regression effects $\Sigma = \{\Sigma_j\}$ for $j = 1, 2, \cdots, q$ and residual covariance matrices Σ_e.

The two observed data are conditionally independent (Sen and Churchill 2001; Wang et al. 2005) so that

$$p(Y|B) = p(y|B) \cdot p(M|B)$$

where $p(y|B)$ is the likelihood, expressed by

$$p(y|B) = \prod_{i=1}^{n} p(y_i|B) \propto |\Sigma_e|^{-n/2} \times \exp\left[\sum_{i=1}^{n}\left(y_i - \psi^T\mu - \sum_{j=1}^{q} x_{ij}\psi^T\beta_j\right)^T \times \Sigma_e^{-1}\left(y_i - \psi^T\mu - \sum_{j=1}^{q} x_{ij}\psi^T\beta_j\right)\right]$$

which arises from the model (1) and corresponding distribution assumptions. $p(M|B) =$ constant if the markers are not missing, but for the missing markers

$$p(M_{miss}|\lambda_i) = \prod_{i=1}^{q} \frac{p(M_{miss}M_{flanking}|\lambda_i)}{p(M_{flanking}|\lambda_i)}$$

which is derived from a Markov model under the assumption of no segregation interference.

$p(B)$ contains the prior distribution for each unknown parameter that needs to be specified beforehand.

Specification of priors

In the Bayesian shrinkage analysis, the number of QTLs, q, is treated as a constant. Generally, at most one QTL is allowed within each marker interval. As a result, the maximum number of QTL is equivalent to the number of marker intervals. Empirically, the length of interval where the QTL may be located should be less than 40 cM under moderate density of markers in Bayesian shrinkage estimation (see Wang et al. 2005; Yang et al. 2007 for justification). Thus, prior $q \in [h_1, h_2]$, where h_1 equals the length of genome divided by 40 cM, and h_2 is the number of marker intervals in the entire genome. For the fixed effects vector μ, there is little knowledge about the priors and thus $p(\mu) \propto$ constant. According to the Bayesian shrinkage estimation of multiple traits (Xu 2007), the prior for each QTL regression effect can be considered as $\beta_j \sim N_{p+1}(0, \Sigma_j)$ with $\Sigma_j \sim IW[v_\beta, (v_\beta\Gamma_j)^{-1}]$ for $j = 1, 2, \cdots, q$, where, v_β is a prior belief and Γ_j are $(p + 1) \times (p + 1)$ positive definite scale matrix. Especially, we adopt a gamma (b, a) distribution with small positive values of a and b ($b \ll a$) as a prior distribution for ρ, denoted as $p(\rho) \propto \rho^{a-1}e^{-b\rho}$. As usual, σ_0^2 is assigned to inverted Chi-square distribution $\mathrm{IC}\left(v_e, \frac{1}{v_e S_e}\right)$ with v_e and S_e being hyperparameters and $p(\lambda_j) = \frac{1}{d_j}$ for $j = 1, 2, \cdots, q$, where d_j is the distance between the two neighboring QTLs.

Conditional posterior distributions and MCMC sampling

To implement Bayesian estimation via the Markov Chain Monte Carlo (MCMC), all of the conditional posterior distributions of the parameters need to be derived from the above joint posterior distribution by the derivative for each parameter. The marginal posterior distribution of μ, given all other parameters, is multivariate normal with $(n\psi\Sigma_\rho^{-1}\psi^T)^{-1}\psi\Sigma_\rho^{-1}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{q} x_{ij}\psi^T\beta_j)$ and covariance matrix $(n\psi\Sigma_\rho^{-1}\psi^T)^{-1}\sigma_0^2$.

Like μ, the conditional posterior distribution of β_j is also normal, whose mean is $\left[\sum_{i=1}^{n} x_{ij}^2\psi\Sigma_\rho^{-1}\psi^T + \Sigma_j^{-1}\sigma_0^2\right]^{-1}\psi\Sigma_\rho^{-1}\sum_{i=1}^{n} x_{ij}(y_i - \mu - \sum_{k\neq j}^{q} x_{ik}\beta_k)$ and covariance matrix is $\left[\sum_{i=1}^{n} x_{ij}^2\psi\Sigma_\rho^{-1}\psi^T + \Sigma_j^{-1}\sigma_0^2\right]^{-1}\sigma_0^2$ for $j = 1, 2, \cdots, q$.

The conditional posterior distribution of the covariance matrix Σ_j of the random effects is inverse-Wishart with parameters $h_\beta = v_\beta + p + 1$ and $h_\beta\Gamma_j + (\beta_j^T\beta_j)^{-1}$.

For the residual variance σ_0^2, the corresponding marginal posterior distribution is a scaled inverse Chi-square with parameters $n(m+1)+v_e$ and $h_e S_e + (\sum_{i=1}^{n} e_i^T e_i)^{-1}$, where $h_e = v_e + n(m+1)$ and $e_i = y_i - \psi^T \mu - \sum_{j=1}^{q} x_{ij}\psi^T \beta_j$.

As the conditional posterior distribution of autoregressive coefficient does not have a closed form, we embed a Metropolis–Hastings acceptance or rejection sampling step in the MCMC scheme to obtain draws for ρ. Let ρ be the current value of autoregressive coefficient. We sample a proposed autoregressive coefficient ρ^* with the uniform distribution from the numerical interval $[\rho - d, \rho + d]$, where d is a small positive number (tuning parameter). Then, the acceptance probability of ρ^* is

$$\alpha = \frac{p(\rho^*)\prod_{i=1}^{n} p(y_i|\rho^*, B_{-\rho})}{p(\rho)\prod_{i=1}^{n} p(y_i|\rho, B_{-\rho})},$$

where $B_{-\rho}$ represents unknown parameters except for ρ.

Consider that the genotype of QTL closely depends on the QTL position. We adopt Metropolis–Hastings algorithm to sample QTL position and relative genotype with one locus at a time jointly. Each locus is drawn from a variable interval, the boundaries of which are the positions of two adjacent QTLs (Wang et al. 2005; Zhang and Xu 2005).

Missing marker genotypes were generated randomly in each iteration on the basis of the probability inferred from the nearest non-missing flanking markers and the phenotype (Wang et al. 2005). The probabilities from the markers are treated as the prior probabilities. After incorporation of the marker (QTL) effects through the phenotype, the probabilities become the posterior probabilities. These posterior probabilities are used to generate the missing marker genotypes.

Given the number of QTLs and the conditional posterior distributions stated above, the sampling steps implementing the MCMC algorithm are summarized as follows:

(1) Initialize all variables with some legal values or values sampled from their prior distributions.
(2) Update the population mean, genetic regression effects for each QTL, covariance matrix for each QTL, residual variance, autoregressive coefficient, QTL positions and the genotypes for each QTL.
(3) Impute the genotypes of missing markers.
(4) Repeat steps (2)–(3) until the Markov chain reaches a given length.

Post-MCMC analysis

The posterior sample, i.e., the realized values of all unknown parameters, drawn from the joint posterior distribution can be used to infer the existence of QTL for dynamic traits and to estimate QTL positions and regression effects. In general, the marginal posterior distribution of QTL position is a visual criterion for the existence of QTL (Sillanpää and Arjas 1998, 1999; Wang et al. 2005; Yi and Xu 2000a, b). It can be illustrated by using the QTL intensity profile of the number of hits by sampling values for position parameter in a short segment (called 'bin' and say a 1- or 2-cM segment) against the genome position. If a marker interval contains a true QTL, we expect that the QTL intensity profile within the interval will show a peak. Otherwise, the profile should be flat. To test the significance of QTL, however, we propose to depict the Wald test statistic profile denoted by $T^2(\lambda) = \beta^T(\lambda)V^{-1}(\lambda)\beta(\lambda)$ where $\beta(\lambda)$ is the average regression effect of QTL for the bin located at position λ and $V(\lambda)$ is the corresponding sample covariance matrix for the QTL regression effect at position λ. Under the null hypothesis, i.e., there is no QTL at position λ, $T^2(\lambda)$ follows a Chi-square distribution with $(d+1)$ degrees of freedom. The test statistic $U_k(\lambda) = \sqrt{T_k^2(\lambda)}$, which is a univariate version of the Wald test statistic, may be further regarded as a univariate test statistic profile for the kth regression effect ($k = 0, 1, \cdots, p$). When there is no QTL affecting the kth regression effect at position λ, $U_k(\lambda)$ is a random variable of standard normal distribution. The calculation of these statistics and significance tests has been described by Yang et al. (2007) in detail.

Simulation

We simulated a chromosome segment of 600-cM long, covered by 61 evenly placed markers (10 cM per marker interval) in a BC population. Assume that the growth pattern of the dynamic trait is controlled by 10 QTL, whose positions and effects are listed in Table 1. The Legendre polynomial of three orders was used to describe the changes

Table 1 QTL parameters used in the simulation experiment

QTL	Position	β_0	β_1	β_2	β_3
1	23	0.00	1.65	2.52	1.20
2	56	2.34	2.08	1.37	1.18
3	148	2.55	1.36	−2.02	−1.27
4	193	1.05	−2.57	1.24	−1.10
5	267	1.12	1.68	1.25	0.00
6	332	2.94	0.00	−1.68	1.72
7	338	1.82	−0.80	−1.20	−0.80
8	476	1.28	−0.50	−1.17	1.32
9	522	2.00	−1.25	0.00	−1.18
10	574	1.75	1.31	−1.45	1.17

of population mean and QTL genetic effects. The population mean was assumed to be $\mu = [45 \quad 44 \quad -1 \quad -7]^T$. Also, the residual variance and the autoregressive coefficient were assumed to be $\sigma^2 = 2.0$ and $\rho = 0.8$, respectively. The dynamic phenotypes were generated at the same eight time points as in real data using model (3) from $n =$ 110 and 150 individuals, respectively. The simulated data were analyzed using composite functional mapping and Bayesian functional mapping method, respectively. The simulations were replicated 50 times to evaluate statistical power of detecting QTL with the two mapping methods.

In composite functional mapping, the number of cofactors was assigned as 4 as proposed in Yang et al. (2007). In Bayesian shrinkage estimation, we took the number of QTLs to 30, although 10 QTLs were simulated in each replicated simulation. The actual values for the hyperparameters used in the data analysis are $v_\beta = 1$, $\Gamma_0 = 0.5I$ with I being the identity matrix, $v_e = 1$ and $S_e = 0.5$. Theoretically, the initial values of all variables should be sampled from their prior distributions to ensure the convergence of Bayesian sampling. Also, Bayesian sampling for linear model system can easily reach convergence as long as the initial values of regression coefficients were set to zero and the covariance matrices were set to identity ones. The MCMC was run for 30,000 cycles after the burn-in of 6,000 cycles. One observation in every 60 cycles was saved to reduce serial correlation and then 5,000 observations were used for the post-MCMC analysis.

The QTL locations and effects estimated with Bayesian functional mapping are listed in Table 2. In Table 2, the QTL parameters are calculated by averaging posterior estimates from those simulations in which significant QTL is detected. It shows that Bayesian functional mapping was able to estimate the QTL parameters well. As expected, the precision and accuracy of QTL parameter estimation increase with the increasing sample size. In contrast, composite functional mapping estimate the QTL parameters with relatively low precision and accuracy (see Table 3). Especially, composite functional mapping does not have the ability to separate two closely linked QTL (the sixth and seventh QTLs) and often detects some false-positive QTLs in each replicate simulation. In general, Bayesian functional mapping provides considerably higher statistical power of QTL detection than composite functional mapping.

We also conducted significance test for every regression coefficient using U test statistic (results not shown) and found that, under sample size of 150, Bayesian functional mapping was able to accurately infer the regression coefficient whose true value was 0 by providing a non-significant statistic. Therefore, it shows that no variable selection procedure is required for Bayesian shrinkage estimation because the coefficients corresponding to the non-significant one are shrunken to be near zero.

Table 2 QTL parameters estimated using composite functional mapping for the simulated data

n	110							150						
QTL	Position	β_0	β_1	β_2	β_3	Power		Position	β_0	β_1	β_2	β_3	Power	
1	20.0 (2.98)	−0.10 (0.61)	1.51 (0.77)	2.40 (1.04)	1.23 (0.64)	54%		24.9 (2.16)	−0.07 (0.42)	1.50 (0.65)	2.47 (0.97)	1.19 (0.53)	72%	
2	52.3 (4.54)	2.40 (1.08)	2.19 (0.91)	1.46 (0.75)	1.06 (0.71)	64%		54.1 (3.91)	2.31 (0.89)	2.14 (0.83)	1.42 (0.57)	1.13 (0.60)	80%	
3	154.6 (7.64)	2.43 (0.94)	1.36 (0.71)	−2.08 (0.97)	−1.20 (0.70)	62%		150.8 (6.31)	2.50 (0.86)	1.40 (0.65)	−2.07 (0.79)	−1.24 (0.62)	82%	
4	187.1 (9.22)	0.91 (0.49)	−2.64 (1.13)	1.20 (0.67)	−0.95 (0.53)	48%		190.6 (8.32)	1.00 (0.54)	−2.64 (0.90)	1.20 (0.50)	−1.00 (0.43)	62%	
5	261.9 (11.47)	1.23 (0.75)	1.79 (0.79)	1.30 (0.71)	−0.10 (0.55)	60%		263.4 (9.73)	1.17 (0.64)	1.76 (0.66)	1.31 (0.67)	−0.06 (0.43)	80%	
6	326.9 (14.18)	2.49 (0.89)	−0.12 (0.51)	−1.60 (0.77)	1.79 (0.86)	58%		338.8 (12.98)	2.68 (0.74)	−0.09 (0.49)	−1.63 (0.68)	1.76 (0.70)	78%	
7	–	–	–	–	–	–		–	–	–	–	–	–	
8	472.3 (16.28)	1.15 (0.74)	−0.38 (0.50)	−1.00 (0.51)	1.27 (0.62)	50%		474.1 (14.98)	1.20 (0.69)	−0.43 (0.46)	−1.11 (0.50)	1.29 (0.55)	70%	
9	518.2 (17.99)	1.84 (0.88)	−1.11 (0.71)	0.09 (0.44)	−1.26 (0.66)	50%		520.6 (15.77)	1.92 (0.63)	−1.13 (0.52)	0.05 (0.34)	−1.20 (0.56)	68%	
10	570.0 (19.72)	1.53 (0.79)	1.20 (0.70)	−1.53 (0.70)	1.22 (0.64)	52%		571.9 (16.43)	1.68 (0.66)	1.26 (0.62)	−1.50 (0.61)	1.19 (0.51)	72%	

Numbers in parentheses are the posterior standard deviations

Table 3 QTL parameters estimated using Bayesian functional mapping for the simulated data

n	110						150					
QTL	Position	β_0	β_1	β_2	β_3	Power	Position	β_0	β_1	β_2	β_3	Power
1	21.2 (2.05)	−0.07 (0.49)	1.50 (0.57)	2.45 (0.61)	1.26 (0.44)	86%	22.6 (1.46)	−0.05 (0.21)	1.53 (0.33)	2.48 (0.36)	1.24 (0.30)	100%
2	55.0 (3.82)	2.31 (0.51)	2.14 (0.43)	1.41 (0.35)	1.09 (0.31)	92%	56.2 (3.01)	2.33 (0.30)	2.10 (0.31)	1.40 (0.28)	1.14 (0.28)	100%
3	150.4 (6.44)	2.48 (0.54)	1.34 (0.45)	−2.00 (0.47)	−1.24 (0.40)	88%	148.2 (4.33)	2.51 (0.36)	1.36 (0.27)	−2.02 (0.30)	−1.27 (0.28)	100%
4	195.2 (7.83)	0.95 (0.29)	−2.62 (0.67)	1.27 (0.34)	−0.99 (0.30)	84%	194.0 (5.12)	1.03 (0.24)	−2.58 (0.33)	1.25 (0.29)	−1.04 (0.21)	100%
5	265.4 (9.46)	1.19 (0.45)	1.66 (0.45)	1.18 (0.41)	0.04 (0.35)	74%	267.2 (7.03)	1.15 (0.27)	1.69 (0.33)	1.20 (0.24)	0.02 (0.22)	98%
6	334.8 (11.28)	2.58 (0.56)	−0.10 (0.44)	−1.65 (0.47)	1.62 (0.56)	86%	333.2 (8.93)	2.84 (0.34)	−0.04 (0.21)	−1.68 (0.22)	1.70 (0.30)	100%
7	340.8 (11.51)	1.74 (0.44)	−0.62 (0.33)	−0.98 (0.36)	−1.00 (0.40)	82%	338.2 (9.01)	1.80 (0.31)	−0.72 (0.23)	−1.10 (0.28)	−0.90 (0.24)	100%
8	474.2 (14.22)	1.18 (0.40)	−0.44 (0.37)	−1.05 (0.41)	1.24 (0.42)	84%	476.8 (12.98)	1.24 (0.29)	−0.48 (0.25)	−1.15 (0.26)	1.30 (0.24)	100%
9	520.8 (16.87)	1.87 (0.48)	−1.15 (0.42)	−0.06 (0.34)	−1.21 (0.36)	82%	522.2 (13.48)	2.03 (0.33)	−1.20 (0.28)	−0.04 (0.24)	−1.19 (0.26)	100%
10	572.8 (18.10)	1.58 (0.39)	1.28 (0.42)	−1.50 (0.40)	1.14 (0.34)	88%	574.4 (15.02)	1.70 (0.30)	1.30 (0.22)	−1.43 (0.20)	1.16 (0.21)	100%

Numbers in parentheses are the posterior standard deviations

The estimated population mean $\hat{\mu} = [46.1\,(3.19)\quad 44.9\,(3.11) -0.83\,(0.31) -7.80\,(0.96)]^T$ under sample size of 110 and $\hat{\mu} = [45.4\,(2.85)\quad 44.1\,(2.31) -0.96\,(0.21)\quad -7.12\,(0.82)]^T$ under sample size of 150, autoregressive coefficients $\rho = 0.813$ (0.188) and 0.802 (0.112), and residual variance $\hat{\sigma}^2 = 1.835\,(0.212)$ and 1.903 (0.197) under sample size of 110 and 150, respectively, where the numbers in parentheses are the standard deviations.

Example

The proposed method is applied to reanalyze QTL governing the developmental pattern of the leaf age in rice. The rice population is a doubled-haploid (DH) population derived from the crosses of an indica rice variety Gui-630 and a japonica rice variety Taiwanjing. This DH population was grown with replicates in a field trial. For each plant, the number of developed leaves on the main stem was counted, and the length of the developing leaf was measured every 3–7 days from day 30 after sowing until the full development of the leaf. The time points of measurements $t = (5\quad 8\quad 13\quad 18\quad 21\quad 26\quad 32\quad 39)$. These measured data were used to estimate the leaf age of a plant (y) (Zhou et al. 2001). A total of 111 DH lines were estimated for leaf ages and were genotyped for 175 RFLP markers. A linkage map was constructed using those observed markers in the DH population, covering a total length of 1,225 cM with average space of 7 cM (Weng et al. 2000).

We select the two-order Legendre polynomial to fit changes of population mean and genetic effects with growth time, based on the changing law of phenotypes of trait. Prior to Bayesian sampling, we take the number of QTLs to 70, so that each QTL covers a range of 17.50 cM. The MCMC sampling scheme is the same as that in the simulation study. The analysis took ~4 h on an Inter Core 2 PC with a 2.0-GHz processor and 4.00 GB RAM.

Figure 1a, b shows the profile of QTL intensity and Wald test statistic. In Fig. 1a, 11 more clear peaks arise on chromosomes 1, 3, 5, 7, 10, 11 and 12. Except for the first peak on chromosome 11, the Wald test statistic values at the other ten peaks exceed 7.82 (see Fig. 1b), the critical value of Chi-square statistic at the degree of freedom of 3 and significant level of 5%. Compared to the composite functional mapping, our Bayesian functional mapping detects more QTLs. The detected QTLs with our method overlap with those of the composite functional mapping (Yang et al. 2006), except for QTL on chromosome 6.

In the significance test of individual regression effect using U statistic, as seen from Fig. 2, the intercepts of Legendre polynomials are significant for all detectable QTLs; except for the second, seventh and ninth QTLs; the

first-order regression coefficients are significant for the other seven detectable QTLs, while there are few significant quadric regression coefficients for the detected QTLs, except for the second, sixth, seventh and eighth QTLs.

Estimates for positions and regression effects of QTL detected with Bayesian functional mapping (BFM) along with QTL positions from composite functional mapping (CFM) are shown in Table 4. Using these estimated regression effects, we depict the changes in genetic effects of ten detectable QTL with measurement time in Fig. 3. These curves represent the three patterns: convex for the first and eighth QTLs, concave for the second, seventh and ninth QTLs and linear ones for the rest, suggesting that the genetic effects of the detectable QTLs on leaf ages vary in various directions during the developmental process. Yang and Xu (2007) only found seven of ten QTLs by using Bayesian shrinkage analysis with polynomial residuals, but there was no difference in genetic change of detectable QTL between the two mapping methods.

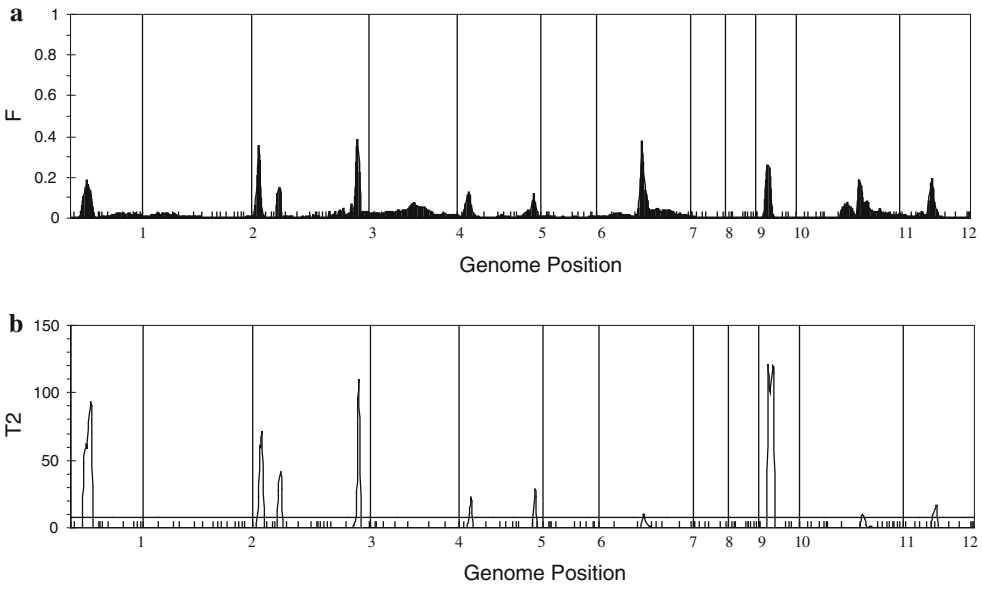

Fig. 1 Profiles of QTL parameters drawn from the real data analysis: **a** QTL intensity profile $F(\lambda)$ and **b** Wald test statistic profile $T^2(\lambda)$. The reference line on the W plot (plot b) is the critical value for the Wald test statistic ($\chi^2_{3,1-0.05} = 7.82$)

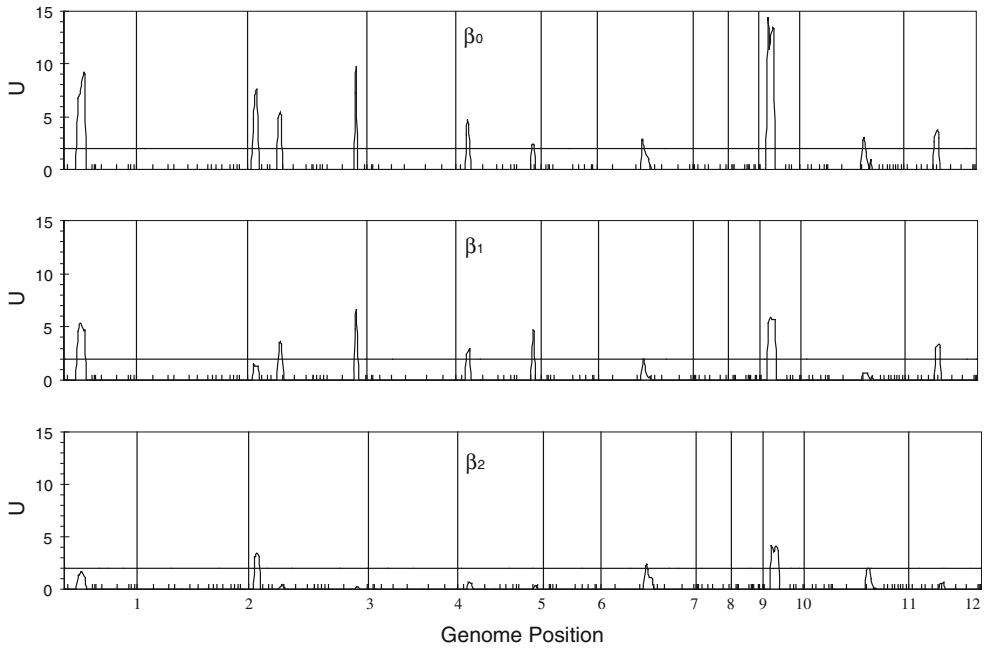

Fig. 2 U test statistic profile for each of the three components of the QTL effect $\beta(\lambda) = [\beta_0(\lambda)\beta_1(\lambda)\beta_2(\lambda)\beta_3(\lambda)]^T$ from the real data analysis, where $\beta_k(\lambda)$ is the component for order k of the polynomial. The reference line on each plot is the critical value for the U test ($U_{1-0.05} = 1.96$)

Table 4 Estimates for QTL parameters obtained from real data analysis

QTL No	BFM					CFM
	Chr-position	Marker interval	β_0	β_1	β_2	Chr-position
1	1–22	C178-RZ776	0.196 (0.027)	0.083 (0.015)	−0.014 (0.009)	1–19
2	3–10	C6C-C814A	−0.323 (0.047)	−0.022 (0.016)	0.034 (0.011)	3–12
3	3–38	RZ993-C825B	0.115 (0.021)	0.050 (0.014)	0.003 (0.010)	3–37
4	3–144	C1032A-RZ142	0.306 (0.032)	0.116 (0.018)	−0.003 (0.013)	3–141
5	5–26	R830-Y11L	−0.153 (0.032)	−0.058 (0.021)	0.006 (0.011)	–
6	5–104	C246-C246	−0.038 (0.016)	−0.073 (0.016)	−0.004 (0.011)	–
7	6–62	R276-R2071	–	–	–	6–62
8	7–61	R1394-C451	−0.399 (0.142)	−0.086 (0.044)	0.035 (0.014)	–
9	10–16	RZ892-RZ561	0.444 (0.036)	0.093 (0.016)	−0.036 (0.010)	10–24
10	11–85	G257-R2869	−0.243 (0.078)	−0.013 (0.021)	0.034 (0.017)	11–118
11	12–44	RG9-CDO344	−0.145 (0.039)	−0.069 (0.020)	−0.006 (0.010)	–

Numbers in parentheses are the posterior standard deviations

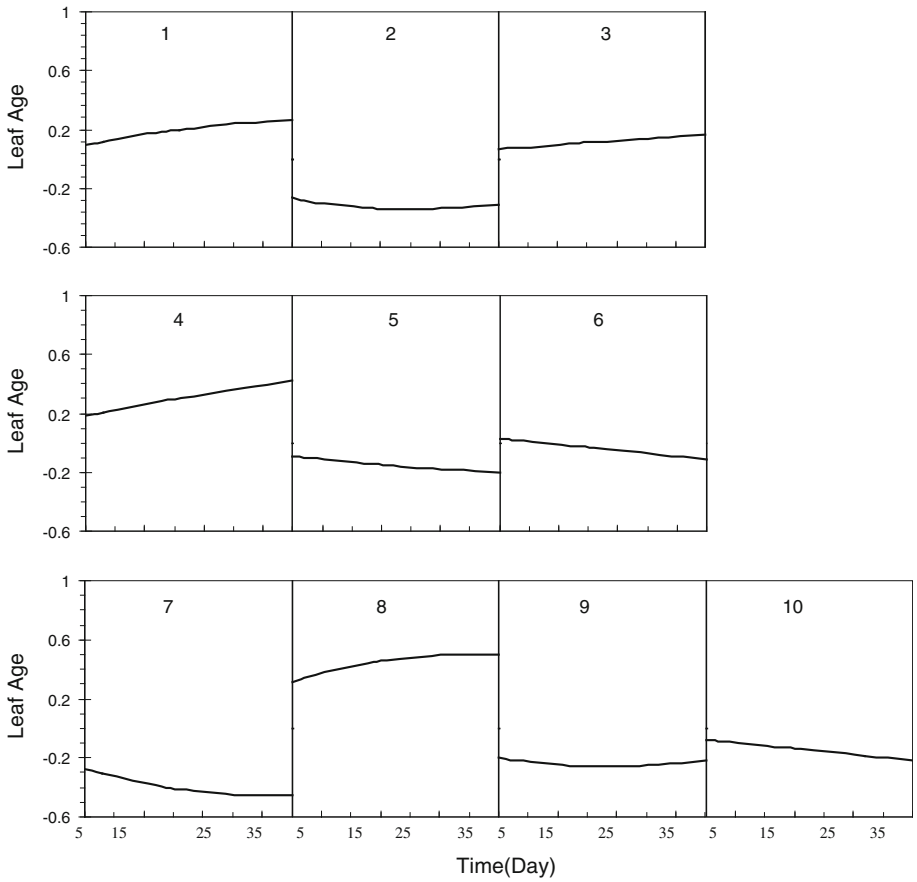

Fig. 3 The change in genetic effects of the detected QTL with time. Each panel corresponds to different QTL

Discussion

Through fitting the change of each QTL effect with the time by Legendre polynomial and modeling the residual covariance structure by first autoregressive equation, we extend the functional mapping under the framework of maximum likelihood for a single QTL model to Bayesian functional mapping that can detect multiple dynamic trait loci simultaneously. We choose Gamma distribution as the appropriate prior for the first order autoregressive coefficient. By our approach, the convergence of sampling for autoregressive coefficient can be easily achieved and the implementation of sampling for all model parameters can be estimated in the MCMC algorithm.

Although these functional mapping strategies have emerged as a powerful tool for mapping dynamic trait loci by using nonlinear biologically meaningful mathematical model, fitting changes of QTL genotype effects may limit their extension to multiple QTL model due to the non-additivity of nonlinear models. Moreover, some biologically mathematical models may not be very meaningful in interpreting most of the dynamic traits. The additivity, orthogonality and flexibility of the Legendre polynomial allow us to construct a linear model that describes time-dependent genetic effects on multiple QTL of dynamic trait with various shapes. In addition to the order of polynomial, the population mean can be determined according to the shape of phenotypic trajectories of dynamic traits; the orders of polynomials for each QTL effect in model (1) are unknown and different. To determine the order of each polynomial, we may simply choose the highest possible order for all QTL effects and handle the excessive regression coefficients with the shrinkage mechanism, which can shrink higher-order coefficients to zero.

In the functional mapping, the residual covariance matrix can be structured by stationary and non-stationary parametric models with a single or more parameters (Ma et al. 2002; Zhao et al. 2005) and nonparametric models based on the Legendre polynomial (Gao and Yang 2006; Yang et al. 2006). In the study, we choose the AR (1) to fit the residuals in multiple dynamic trait loci models. The reasons are as follows: (1) The AR (1) model, as one of the simplest time series models, provides a simple closed expression for calculating the determinant and inverse of the matrix for any number of time points measured. (2) The residuals of the AR (1) model may be sufficient because the residuals are simpler in the multiple QTL than in a single QTL. Although Legendre polynomials have higher goodness of fit to the residuals by choosing appropriate orders, too many parameters to be estimated and lack of biological meaning may limit its application. (3) Most importantly, the prior use of gamma distribution may ensure a quicker convergence for the sampling of the autoregressive coefficient with Metropolis-Hasting algorithm. We found that the specified uniform distribution prior to autoregressive coefficient and the sampling method proposed by Gianola et al. (2003) do not work well in Bayesian functional mapping. Also, Bayesian sampling for more parameters in other parametric models, such as the structured antedependence model (SAD), is up for investigation in future.

The proposed method can be easily extended from a BC population to DH, RIL, F2 and outbred populations. In other words, it can analyze not only the additive effects on dynamic traits in BC, DH and RIL populations but also both additive and dominant effects in F_2 and outbred populations. The computer program of our method in MATLAB is available and free upon request. Given the maximum number of QTL, besides the Bayesian shrinkage estimation, stochastic search variable selection (SSVS, Yi et al. 2003a) and Bayesian model choice (Yi et al. 2003b, 2005) can also be used for sampling QTL parameters in the fixed model space. Simulation studies demonstrate that SSVS and the shrinkage estimation perform almost equally well (Wang et al. 2005). Bayesian model choice can save the computing time, as only those positions with significant genetic effects are drawn. Therefore, once we take into account the interactions between QTL in multiple dynamic trait loci model, Bayesian model choice should be an optimal mapping strategy, compared with the Bayesian shrinkage analysis and SSVS.

Acknowledgments This work was partially supported by the National Natural Science Foundation of China (30972077).

References

Gao HJ, Yang RQ (2006) Composite interval mapping of QTL for dynamic traits. Chin Sci Bull 51(15):1857–1862

Gianola D, Perez-Enciso M, Toro MA (2003) On marker-assisted prediction of genetic value: beyond the ridge. Genetics 163(1):347–365

Giraldo J (2003) Empirical models and Hill coefficients. Trends Pharmacol Sci 24(2):63–65

Henderson CR Jr (1982) Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. Biometrics 38(3):623–640

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136(4):1447–1455

Kirkpatrick M, Heckman N (1989) A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. J Math Biol 27(4):429–450

Kirkpatrick M, Lofsvold D, Bulmer M (1990) Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124(4):979–993

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121(1):185–199

Ma CX, Casella G, Wu R (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. Genetics 161(4):1751–1762

Macgregor S, Knott SA, White I, Visscher PM (2005) Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. Genetics 171(3):1365–1376

Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271(5255):1582–1586

Schaeffer LR (2004) Application of random regression models in animal breeding. Livest Prod Sci 86(3):35–45

Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. Genetics 159(1):371–387

Sillanpää MJ, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics 148(3):1373–1388

Sillanpää MJ, Arjas E (1999) Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. Genetics 151(4):1605–1619

Tobalske BW, Hedrick TL, Dial KP, Biewener AA (2003) Comparative power curves in bird flight. Nature 421(6921):363–366

Wang H, Zhang YM, Li X, Masinde GL, Mohan S, Baylink DJ, Xu S (2005) Bayesian shrinkage estimation of quantitative trait loci parameters. Genetics 170(1):465–480

Weng QM, Wu WR, Li WM, Liu HQ, Tang DZ, Zhou YC, Zhang QF (2000) Construction of an RFLP linkage map of rice using DNA probes from two different sources. J Fujian Agric Univ 29(2):129–133

Wu R, Ma CX, Littell RC, Wu SS, Yin T, Huang M, Wang M, Casella G (2002) A logistic mixture model for characterizing genetic determinants causing differentiation in growth trajectories. Genet Res 79(3):235–245

Wu R, Ma CX, Zhao W, Casella G (2003) Functional mapping for quantitative trait loci governing growth rates: a parametric model. Physiol Genomics 14(3):241–249

Wu R, Ma CX, Lin M, Casella G (2004a) A general framework for analyzing the genetic architecture of developmental characteristics. Genetics 166(3):1541–1551

Wu RL, Ma CX, Lin M, Wang ZH, George C (2004b) Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. Biometrics 60(4):729–738

Xu S (2007) Derivation of the shrinkage estimates of quantitative trait locus effects. Genetics 177(2):1255–1258

Yang R, Xu S (2007) Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. Genetics 176(2):1169–1185

Yang R, Tian Q, Xu S (2006) Mapping quantitative trait loci for longitudinal traits in line crosses. Genetics 173(4):2339–2356

Yang R, Gao H, Wang X, Zhang J, Zeng ZB, Wu R (2007) A semiparametric approach for composite functional mapping of dynamic quantitative traits. Genetics 177(3):1859–1870

Yi N, Xu S (2000a) Bayesian mapping of quantitative trait loci for complex binary traits. Genetics 155(3):1391–1403

Yi N, Xu S (2000b) Bayesian mapping of quantitative trait loci under the identity-by-descent-based variance component model. Genetics 156(1):411–422

Yi N, George V, Allison DB (2003a) Stochastic search variable selection for identifying multiple quantitative trait loci. Genetics 164(3):1129–1138

Yi N, Xu S, Allison DB (2003b) Bayesian model choice and search strategies for mapping interacting quantitative trait loci. Genetics 165(2):867–883

Yi N, Yandell BS, Churchill GA, Allison DB, Eisen EJ, Pomp D (2005) Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. Genetics 170(3):1333–1344

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136(4):1457–1468

Zhang YM, Xu S (2005) Advanced statistical methods for detecting multiple quantitative trait loci. Recent Res Devel Genet Breed 2(1):1–23

Zhao W, Chen YQ, Casella G, Cheverud JM, Wu R (2005) A non-stationary model for functional mapping of complex traits. Bioinformatics 21(10):2469–2477

Zhou Y, Li W, Wu W, Chen Q, Mao D, Worland J (2001) Genetic dissection of heading time and its components in rice. Theor Appl Genet 102(8):1236–1242