Genet. Res., Camb. (2012), **94**, *pp.* 85–95. © Cambridge University Press 2012
doi:10.1017/S0016672312000249

Bayesian B-spline mapping for dynamic quantitative traits

JUN XING[1], JIAHAN LI[2], RUNQING YANG[3,4]*, XIAOJING ZHOU[5] AND SHIZHONG XU[6]

[1] *Department of Gastroenterology, Tumor Hospital of Harbin Medical University, Harbin 150086, People's Republic of China*
[2] *Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, USA*
[3] *Animal Science and Veterinary Medicine, Heilongjiang Bayi Agricultural University, Daqing, 163319, People's Republic of China*
[4] *School of Agriculture and Biology, Shanghai Jiaotong University, Shanghai 200240, People's Republic of China*
[5] *Department of Mathematics, Heilongjiang Bayi Agricultural University, Daqing, 163319, People's Republic of China*
[6] *Department of Botany and Plant Science, University of California, Riverside, CA 92521, USA*

(*Received 14 September 2010; revised 15 February 2012; accepted 20 February 2012*)

Summary

Owing to their ability and flexibility to describe individual gene expression at different time points, random regression (RR) analyses have become a popular procedure for the genetic analysis of dynamic traits whose phenotypes are collected over time. Specifically, when modelling the dynamic patterns of gene expressions in the RR framework, B-splines have been proved successful as an alternative to orthogonal polynomials. In the so-called Bayesian B-spline quantitative trait locus (QTL) mapping, B-splines are used to characterize the patterns of QTL effects and individual-specific time-dependent environmental errors over time, and the Bayesian shrinkage estimation method is employed to estimate model parameters. Extensive simulations demonstrate that (1) in terms of statistical power, Bayesian B-spline mapping outperforms the interval mapping based on the maximum likelihood; (2) for the simulated dataset with complicated growth curve simulated by B-splines, Legendre polynomial-based Bayesian mapping is not capable of identifying the designed QTLs accurately, even when higher-order Legendre polynomials are considered and (3) for the simulated dataset using Legendre polynomials, the Bayesian B-spline mapping can find the same QTLs as those identified by Legendre polynomial analysis. All simulation results support the necessity and flexibility of B-spline in Bayesian mapping of dynamic traits. The proposed method is also applied to a real dataset, where QTLs controlling the growth trajectory of stem diameters in Populus are located.

1. Introduction

Dynamic traits are those that change over time in the developmental process of life or other quantitative factors (e.g. environmental condition). These traits are often observed in the fields of biology and medicine, such as growth and developmental traits, milk production, egg production and drug response. Any development in plant and animal experiences both systematic and individual-specific processes, and quantitative trait loci (QTLs) are genes across the whole genome that control the systematic component of this developmental process. Mapping QTL of dynamic traits can be conducted in various ways: the simplest approach consists of performing single-trait analysis at each time point, this is the least restrictive approach in the sense that no parametric restrictions are imposed on the curve shape that is formulated by observed data points. However, the single-trait approach may not be efficient because the inference of QTL effects at each time point does not benefit from 'borrowing' information from other time points. A natural alternative is to conduct a multiple-trait analysis in which measurements at each time point are regarded as different traits. Although the multiple-trait method takes into account the correlations among measurements, it can only identify the QTL for the measured time points. If too many measuring points exist, the solution to multivariate analysis will be infeasible. In contrast, random regression (RR)

* Corresponding author: School of Agriculture and Biology, Shanghai Jiaotong University, Shanghai 200240, People's Republic of China. Tel: (8621) 34206146. Fax: (8621) 34206146. E-mail: runqingyang@sjtu.edu.cn

analysis (e.g. Henderson, 1982; Jamrozik *et al*., 1997; Schaeffer, 2004) fits the dynamic pattern of the genetic effect for each QTL, which can not only detect the QTL-controlled dynamic trajectories but also infer QTL at arbitrary time points (Macgregor *et al*., 2005; Yang *et al*., 2006, 2007; Yang & Xu, 2007). Along with these ideas, Wu and his colleagues (Ma *et al*., 2002; Wu *et al*., 2002, 2003, 2004) have developed a functional-mapping strategy that uses some structured models to fit residuals (Jin *et al*., 2010). But in RR analysis, fitting time-dependent residuals by polynomials can take advantages of the many well-developed inference methods available in linear model literature.

In order to map QTL for traits with dynamic patterns, both parametric and non-parametric models are used to describe the change of genotypic effects over time. Emphasizing on the interpretability of mapping results, earlier functional mapping (Ma *et al*., 2002; Wu *et al*., 2002, 2003, 2004) fits QTL phenotypic effects by biologically meaningful mathematical functions. However, its applicability is limited due to non-linearity and the pool of candidate mathematical functions (Yang *et al*., 2006, 2007, 2009; Yang & Xu, 2007). In contrast, Legendre polynomials have been widely used for mapping dynamic traits. In addition to the flexibility of fitting biological curves with arbitrary shapes, the Legendre polynomial is a linear model, and thus theories and algorithms that developed in linear models could be applied directly to estimate QTL parameters. However, although higher-order polynomials are capable of modelling changes in means and variances along a continuous scale, such polynomials often overemphasize on observations at the extremes and may result in Runge's phenomenon. That is, the goodness of fit to curve decreases with the order of polynomials, due to oscillations at two extremes of the curve (de Boor, 2001). Alternatively, the splines that construct curves from pieces of lower-degree polynomials smoothed at selected pointed (knots) are more commonly used in non-parametric data analysis. As a particular type of spline curve, B-splines yield the same fit as splines based on truncated power functions, but have better numerical properties (Ruppert *et al*., 2003). The applications of the B-splines to mapping QTL for dynamic traits have been firstly discussed by Yang *et al*. (2006) and introduced by Yang *et al*. (2009), respectively.

Frameworks for mapping dynamic trait loci have been developed from the interval-mapping procedure under maximum likelihood to the Bayesian-mapping method. Compared to interval mapping that detects one QTL at a time based on a single QTL model, Bayesian mapping based on a multiple QTL model can simultaneously identify multiple QTLs across the entire genome, which greatly enhances the statistical power of QTL detection. In this paper, using B-spline to model the dynamics of population mean, QTL effects and individual-specific time-dependent environmental errors, we establish a multiple QTL model for mapping dynamic trait loci and estimate QTL parameters using the Bayesian shrinkage method. Through computer simulations, we compare the performance of Bayesian-mapping and interval-mapping methods, as well as the flexibilities of B-splines and Legendre polynomials in the QTL mapping of dynamic quantitative traits.

2. Methods

(i) *Genetic model of dynamic traits*

We now use a backcross design as an example to describe the genetic model for dynamic traits. Based on Mendel's law of inheritance, there are two possible genotypes in a backcross population at any given locus, denoted by Qq and qq, respectively, with equal frequencies. Let $y_i(t)$ be the phenotypic value of individual i measured at time t, which can be described by the following linear model:

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

for $i=1, 2, \ldots, n$, where n is the number of individuals, q is the maximum number of QTLs evaluated in the genome, $\mu(t)$ is the population mean at time t, x_{ij} is the genotype indicator variable (defined as 1 for one genotype and -1 for the alternative genotype) for the ith individual at the jth locus, $\alpha_j(t)$ ($j=1, 2, \ldots, q$) is the genetic value of the jth QTL at time t, $\beta_i(t)$ is an individual-specific time-dependent environmental error with an i.i.d. $N[0, \sigma^2_\beta(t)]$ distribution and ε_i is an individual-specific time-independent environmental error with an i.i.d. $N(0, \sigma^2_\varepsilon)$ distribution. This is a mixed effects model with $\mu(t)$ and $\alpha_j(t)$ being the fixed effects and $\beta_i(t)$ being the random effect. The purpose of the QTL mapping is to estimate $\alpha_j(t)$, the time-dependent functional genetic effect of locus j, for $j=1, 2, \ldots, q$.

All the model parameters, except σ^2_ε, are functions of time. The functional relationships between parameters and time may be described by B-splines. Define $\psi(t)=[\psi_{0,\,p}(t) \quad \psi_{1,\,p}(t) \quad \ldots \quad \psi_{r,\,p}(t)]$ as a covariable of the B-splines with k knots and p-order polynomial, where $r=k-p-2$ (see Appendix A). Also define $\boldsymbol{\mu}=[\mu_0 \quad \mu_1 \quad \ldots \quad \mu_r]^{\mathrm{T}}$ as a vector of population means, which is time independent. The time-dependent population mean $\mu(t)$ may be described as a linear combination of $\boldsymbol{\mu}$ weighted by the basis of the B-splines, i.e. $\mu(t)=\boldsymbol{\psi}(t)\boldsymbol{\mu}$. Similarly, we can describe other parameters using the same B-splines, e.g. $\alpha_j(t)=\boldsymbol{\psi}(t)\boldsymbol{\alpha}_{\mathbf{j}}$, where $\boldsymbol{\alpha}_j=[\alpha_{j0} \quad \alpha_{j1} \ldots \alpha_{jr}]^{\mathrm{T}}$ for $j=1, 2, \ldots, q$ and $\beta_i(t)=\boldsymbol{\psi}(t)\boldsymbol{\beta}_{\boldsymbol{i}}$, where $\boldsymbol{\beta}_i=[\beta_{i0} \quad \beta_{i1} \ldots \beta_{ir}]^{\mathrm{T}}$ for $i=1, 2, \ldots, n$. Since $\boldsymbol{\beta}_{\mathbf{i}}$ is treated as an RR

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

effect, we assume that $\boldsymbol{\beta_i}$ is i.i.d. $N(0, \boldsymbol{\Sigma_\beta})$ with $\boldsymbol{\Sigma_\beta}$ being the covariance matrix of RR effect for individual-specific time-dependent environmental errors.

With the B-spline reparameterization, model (1) is now rewritten as a linear function of the time-independent parameters:

$$y_i(t) = \boldsymbol{\psi}(t)\boldsymbol{\mu} + \sum_{j=1}^{q} x_{ij}\boldsymbol{\psi}(t)\boldsymbol{\alpha}_j + \boldsymbol{\psi}(t)\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad (2)$$

The phenotypic values for each individual are collected at $m+1$ fixed time points, t_0, t_1, …, t_m, denoted by a vector $\mathbf{y_i} = [y_i(t_0)\ y_i(t_1)\ \ldots\ y_i(t_m)]^T$. Define

$$\boldsymbol{\psi} = [\,\boldsymbol{\psi}^{\mathrm{T}}(t_0) \quad \boldsymbol{\psi}^{\mathrm{T}}(t_1) \cdots \boldsymbol{\psi}^{\mathrm{T}}(t_m)\,]$$

as an $(r+1)\times(m+1)$ matrix. In matrix notation, the linear model for $\boldsymbol{y_i}$ is

$$\boldsymbol{y}_i = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\mu} + \sum_{j=1}^{p} x_{ij}\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\alpha}_j + \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad (3)$$

where $\boldsymbol{\varepsilon}_i = [\,\varepsilon_{it_0} \cdots \varepsilon_{it_m}\,]^{\mathrm{T}}$ is an $(m+1)\times 1$ vector for the residual effects assumed $\varepsilon_i \sim N(\mathbf{0}, \mathbf{I}\boldsymbol{\sigma_\varepsilon}^2)$. The expectation of model (3) is

$$\mathrm{E}(\mathbf{y}_i|\boldsymbol{\mu}, \boldsymbol{\alpha}_j) = \mathbf{U}_i = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\mu} + \sum_{j=1}^{p} x_{ij}\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\alpha}_j$$

and the covariance matrix is

$$\mathrm{Var}(\mathbf{y}_i|\boldsymbol{\mu}, \boldsymbol{\alpha}_j) = \mathbf{V} = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\Sigma_\beta}\boldsymbol{\psi} + \mathbf{I}\sigma_\varepsilon^2.$$

(ii) *Bayesian B-spline mapping*

In Bayesian mapping analysis for dynamic traits, the observed data are phenotypes $\boldsymbol{y} = \{\boldsymbol{y}_i\}$ for $i = 1, 2, \ldots, n$ and marker information $\boldsymbol{M} = \{\boldsymbol{M}_j\}$ for $i = 1, 2, \ldots, n$. Parameters $\boldsymbol{\theta}$ include population mean $\boldsymbol{\mu}$, QTL regression effects $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_j\}$ for $j = 1, 2, \ldots, q$, RR effect for individual-specific time-dependent environmental errors $\boldsymbol{\beta} = \{\boldsymbol{\beta}_i\}$ for $i = 1, 2, \ldots, n$, QTL positions $\boldsymbol{\lambda} = \{\lambda_j\}$ for $j = 1, 2, \ldots, q$, the QTL genotype indicator variable $\boldsymbol{x} = \{x_{ij}\}$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, q$, prior covariance matrice of QTL regression effects $\boldsymbol{A} = \{\boldsymbol{A_j}\}$ for $j = 1, 2, \ldots, q$, covariance matrix of RR effects for individual-specific time-dependent environmental error $\boldsymbol{\Sigma_\beta}$ and residual variance σ_ε^2.

(a) *Likelihoods*

Given unknown parameters, the observed data are conditionally independent (Sen & Churchill, 2001; Wang *et al.*, 2005) so that

$$p(\mathbf{y}, \mathbf{M}|\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{M}, \boldsymbol{\theta})p(\mathbf{M}|\boldsymbol{\theta}),$$

where we have the likelihood based on the model (3)

$$p(\mathbf{y}|\mathbf{M}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(\mathbf{y}_i|\mathbf{M}, \boldsymbol{\theta}) \propto |\mathbf{V}|^{-n(m+1)2} \exp\left[\sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{U}_i)^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{U}_i)\right] \quad (4)$$

and

$$p(\mathbf{M}|\boldsymbol{\theta}) = p(\mathbf{M}|\mathbf{x}, \boldsymbol{\lambda}) = \prod_{i=1}^{q} \frac{p(\mathbf{M}_i, \mathbf{x}_i|\boldsymbol{\lambda})}{p(\mathbf{x}_i|\boldsymbol{\lambda})},$$

which is derived from a Markov model under the assumption of no segregation interference (Wang *et al.*, 2005).

(b) *Prior specification*

In the Bayesian shrinkage analysis for QTL mapping, the number of QTLs, q, is treated as a constant (see Wang *et al.* 2005 and Yang *et al.* 2007 for justification). The prior distribution for $\boldsymbol{\mu}$ is uniform. The prior distribution for each of the genetic effects is multivariate normal, i.e. $p(\boldsymbol{\alpha_j}|\boldsymbol{A_j}) = N(\mathbf{0}, \boldsymbol{A_j})$ for all $j = 1, 2, \ldots, q$, where $\boldsymbol{A_j}$ has its own prior, $p(\boldsymbol{A_j}) = \mathrm{IW}(b_0, \boldsymbol{\Gamma_A})$, an inverse Wishart distribution with b_0 and $\boldsymbol{\Gamma_A}$ being hyperparameters. The prior distribution for $\boldsymbol{\Sigma_\beta}$ is also inverse Wishart, i.e. $\boldsymbol{\Sigma_\beta} \sim \mathrm{IW}(d_0, \boldsymbol{\Gamma_\beta})$, where d_0 and $\boldsymbol{\Gamma_\beta}$ are hyperparameters. The prior for σ_ε^2 is inverted chi-square distribution $\mathrm{IC}(v_e, (v_e S_e)^{-1})$ with v_e and S_e being hyperparameters, and $p(\lambda_j) = 1/l_j$ for $j = 1, 2, \ldots, q$, where l_j is the distance between the two neighbouring QTLs (Sillanpää & Arjas, 1998, 1999; Wang *et al.*, 2005). The joint prior distribution of all the parameters is

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\mu})p(\boldsymbol{\Sigma_\beta})p(\sigma_\varepsilon^2)\prod_{j=1}^{q} p(\boldsymbol{\alpha}_j|\mathbf{A}_j)p(\mathbf{A}_j). \quad (5)$$

Combining the conditional density of the data with the prior distribution of parameters, we obtain the joint distribution of the data and parameters, which is proportional to the joint posterior distribution of the parameters,

$$p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{M}) \propto p(\mathbf{y}, \mathbf{M}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (6)$$

(c) *Markov chain Monte Carlo (MCMC) sampling for QTL parameters*

This joint posterior distribution is the target distribution from which parameters are sampled. Due to analytically intractable joint posterior distribution, the MCMC methods such as Gibbs sampler (Gelman *et al.*, 1995) and Metropolis–Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970) are used to

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

sample each parameter conditional on all other parameters. Except for QTL genotypes without the closed form of conditional posterior distribution, other unknown parameters can be drawn by Gibbs samplers from their conditional posterior distributions (see Appendix B for details of derivation).

Considering that the genotype of QTL closely depends on the QTL position, we adopt Metropolis–Hastings algorithm to sample jointly the QTL position and relative genotype for one locus at a time. Each locus is drawn from a variable interval whose boundaries are the positions of adjoining QTLs (Wang *et al*., 2005; Zhang & Xu, 2005).

Genotypes of missing markers were generated randomly in each iteration on the basis of the probability inferred jointly from the nearest non-missing flanking markers and the phenotypes (Wang *et al*., 2005). The probability of missing marker genotype estimated by the flanking markers is treated as the prior probability. After incorporating the marker (QTL) effects through the phenotype, the probability becomes the posterior probability, which is used to generate the missing marker genotype.

In summary, the MCMC algorithm is described in the following steps:

(1) Initialize all variables with values sampled from their prior distributions.
(2) Update the population means with a sample from a multivariate normal with mean (B.1) and covariance matrix (B.2) in Appendix B.
(3) Update the genetic effects for each QTL with sample from a multivariate normal with mean (B.3) and covariance matrix (B.4) in Appendix B.
(4) Update the covariance matrix for each QTL with a sample from the inverse Wishart (B.5) of Appendix B.
(5) Update the RR effects for individual-specific time-dependent environmental errors with a sample from a multivariate normal with mean (B.6) and covariance matrix (B.7) in Appendix B.
(6) Update the covariance matrix of RR effects for individual-specific time-dependent environmental errors with a sample from the inverse Wishart (B.8) of Appendix B.
(7) Update the residual variance with a sample from a scaled inverse chi-square (B.9) of Appendix B.
(8) Update the QTL position for each marker interval.
(9) Update the genotypes for each QTL.
(10) Impute the genotypes of missing markers.
(11) Repeat steps (2)–(10) until the Markov chain reaches a desirable length.

(d) *Post-Bayesian analysis*

In conventional Bayesian mapping, the marginal posterior distribution of the QTL position can be shown by plotting the number of hits by the QTL in a short interval against the genome location (Sillanpää & Arjas, 1998, 1999; Yi & Xu, 2000*a*, *b*; Wang *et al*., 2005). The curve is called QTL intensity profile. If an interval contains a QTL, we expect that the QTL intensity profile within the interval shows a peak. Otherwise, the intensity profile appears flat (uniform). The intensity profile only provides us a signal of 'peak' at possible QTL, but it is unable to answer whether the effects of the QTL with higher intensity are statistically significant or not. To address this problem, we used a Wald test to determine statistical significance from a frequentist perspective (see Yang *et al*. 2007 for theoretical justification). Let us denote the QTL intensity profile by $f(\lambda)$, which is a function of the genome location, and the test statistic of the overall QTL effect by $T^2(\lambda)$, which is

$$T^2(\lambda) = \begin{cases} \boldsymbol{\alpha}^{\mathrm{T}}(\lambda)\mathbf{S}_{\boldsymbol{\alpha}}^{-1}(\lambda)\boldsymbol{\alpha}(\lambda), & f(\lambda) > f_0, \\ 0, & f(\lambda) \leqslant f_0, \end{cases} \tag{7}$$

where $T^2(\lambda)$ is the Wald test statistic, $\boldsymbol{\alpha}(\lambda)$ is the vector for posterior means of the QTL regression effects, $\boldsymbol{S}_{\boldsymbol{\alpha}}(\lambda)$ is the estimated posterior sample covariance matrix for the QTL regression effects and the f_0 is the flat intensity in the interval without the peak. Under the null hypothesis, i.e. there is no QTL at position λ, $T^2(\lambda)$ will have an asymptotic chi-square distribution with r degrees of freedom. Therefore, a critical value of chi-square distribution may be used to declare statistical significance at position λ. Generally, there are higher intensity and larger genetic effect at the position where QTL exists. Using the statistic $T^2(\lambda)$ profile to indicate the locations of the QTL, most of the intervals will have $T^2(\lambda)$ of zeros due to the lower intensities and thus only the intervals with possible significant QTL effects and higher intensities will show clear peaks.

3. Simulation

To evaluate the efficiency of Bayesian B-spline mapping for dynamic traits and the flexibility of B-spline in this framework, we conduct three simulations: (1) comparing the statistical power of QTL detection between the proposed method and the interval-mapping method (Yang *et al*., 2006); (2) fitting the simulated data generated by B-spline using the Legendre polynomial-based approach and (3) fitting the simulated data generated by the Legendre polynomials using B-spline based Bayesian mapping approach for dynamic traits.

For the first two scenarios, we simulate a backcross population including 150 segregating individuals. The

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

Table 1. *QTL regression effects for B-splines used in simulation*

	True parameter					
QTL no.	Position	α_1	α_2	α_3	α_4	α_5
1	23	2·28	0·65	1·52	0·20	2·50
2	56	2·30	−0·39	1·31	0·85	4·20
3	148	3·55	0·36	−1·02	−1·77	2·40
4	153	2·05	2·00	1·24	−0·68	2·80
5	267	−1·05	−0·96	1·54	−0·84	2·52
6	332	3·44	−1·24	−1·23	2·00	3·10
7	338	2·12	1·68	1·25	2·12	−2·00
8	476	−2·06	−0·31	3·93	−2·83	−3·13
9	522	2·75	2·25	2·50	1·50	1·25
10	574	1·60	1·60	1·80	2·30	3·00

61 co-dominant markers are evenly spaced on the chromosome fragment of 600 cM long. We put ten QTLs governing the trajectory of a dynamic trait along the genome. Assume that changes in phenotype of the trait and QTL additive effects over time follow B-splines with four knots and polynomial segments of order 2. The array of measurement time points for each individual was designated as 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100, from which four knot points were chosen at 10, 40, 70 and 100. The regression coefficients for overall mean were set to

$$\boldsymbol{\mu} = (\,8{\cdot}13 \quad 25{\cdot}14 \quad 12{\cdot}27 \quad 0{\cdot}83 \quad 24{\cdot}71\,)^{\mathrm{T}}.$$

The covariance matrix of RR effects for individual-specific time-dependent environmental errors was

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \begin{bmatrix} 1{\cdot}04 & 0{\cdot}171 & -0{\cdot}315 & 0{\cdot}100 & 0{\cdot}280 \\ 0{\cdot}171 & 1{\cdot}586 & -0{\cdot}035 & 0{\cdot}041 & 0{\cdot}410 \\ -0{\cdot}315 & -0{\cdot}035 & 1{\cdot}736 & 0{\cdot}287 & 0{\cdot}050 \\ 0{\cdot}100 & 0{\cdot}041 & 0{\cdot}287 & 0{\cdot}772 & 0{\cdot}320 \\ 0{\cdot}280 & 0{\cdot}410 & 0{\cdot}050 & 0{\cdot}320 & 1{\cdot}250 \end{bmatrix},$$

and the variance for random experimental error was $\sigma_\varepsilon^2 = 2{\cdot}0$. The locations of the ten simulated QTLs and their regression genetic effects are shown in Table 1. The cumulative proportion of phenotypic variance from measurement time point 10–100 contributed by an individual QTL ranged from 0·026 to 0·157, as calculated in Yang *et al.* (2006), the total genetic variance contributed by all ten QTLs was 0·903. The trajectories for each QTL are shown in Fig. 1. They are categorized into three groups according to the shape of curves.

Based on the simulated parameters above, a vector of the phenotypic values for individual was randomly generated by $\mathbf{y}_i = \boldsymbol{\psi}^{\mathrm{T}}\left(\boldsymbol{\mu} + \sum_{j=1}^{10}\boldsymbol{\alpha}_j + \boldsymbol{\beta}_i\right) + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\beta}_\mathbf{i}$ and $\boldsymbol{\varepsilon}_\mathbf{i}$ were the vectors of random numbers sampled from $N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ and $N(\mathbf{0}, \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2)$, respectively. The simulated data were analysed using both the Bayesian B-spline-mapping method and the interval-mapping method based on maximum likelihood (Yang *et al.*, 2006).

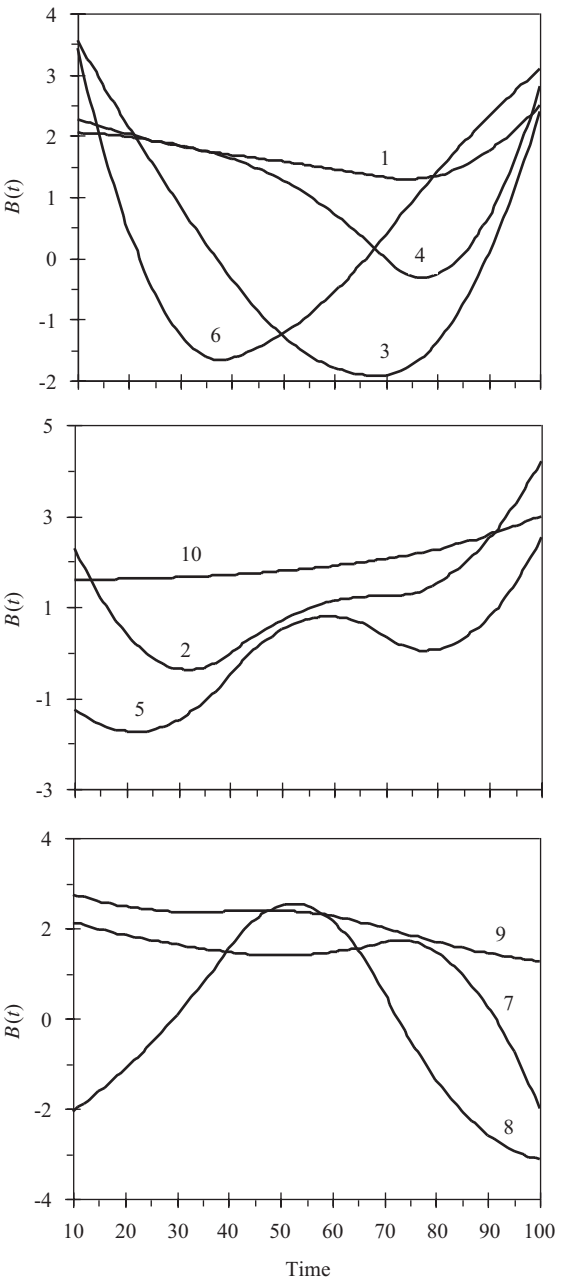

Fig. 1. Changes in genetic effects for ten simulated QTLs with time.

In the MCMC-implemented Bayesian analysis with the moving interval approach, we include only 20 QTLs in the working model. By fitting the phenotypic B-spline of each individual, the initial value of the overall mean ($\boldsymbol{\mu}$) is determined as population mean for regression coefficients, the covariance matrix of RR effects for individual-specific time-dependent environmental error ($\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$) is initialized with population covariance matrix for regression coefficients, and residual variance (σ_ε^2) starts with population mean for residual error variances of phenotypic B-splines. The

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

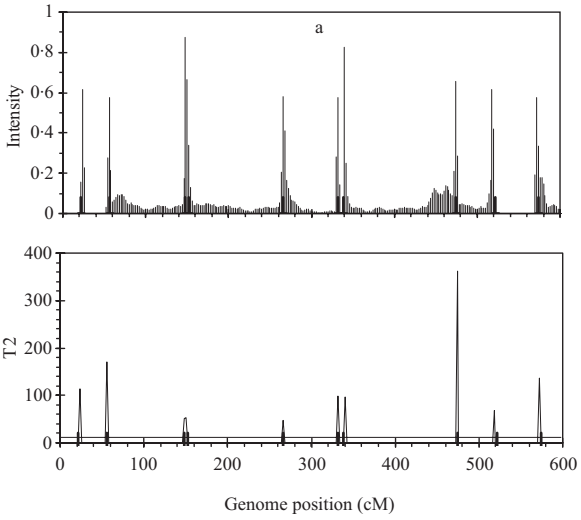

Fig. 2. The QTL intensity profile (above) and T^2 test statistic profile (below) over the entire genome obtained with Bayesian B-spline mapping. The true positions of the simulated QTL are represented by black needles. In T^2 test statistic plot, the horizontal line indicates the empirical critical value of 11·05 when the type I error rate was 5%.

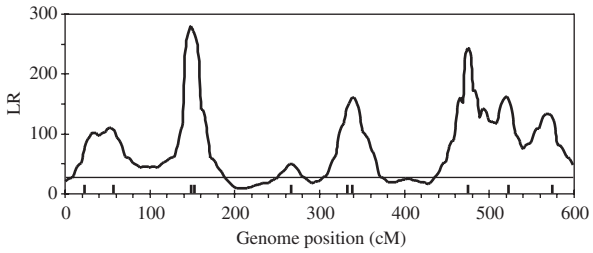

Fig. 3. The likelihood ratio statistic profiles estimated with interval-mapping analysis for the entire genome. The horizontal line indicates the empirical critical value of 25·01 when the type I error rate was 5%, which is estimated from 500 permutation tests.

regression coefficients for genetic effects of all QTLs ($\boldsymbol{\alpha}$) are initialized as zero. Both the hyperparameters b_0 and d_0 are taken to be 5 and the prior covariance $\boldsymbol{\Gamma}_A$ and $\boldsymbol{\Gamma}_\beta$ were assigned to be the identity matrix $\boldsymbol{I}$ and $0{\cdot}5\boldsymbol{I}$, respectively. The initial value of λ_j takes the middle point of the interval where the jth QTL resides. The initial value of genotype indicator x_{ij} is sampled from the probability of x_{ij} conditional on the flanking markers. The Gibbs sampler was run for 30 000 cycles after discarding the first 6000 cycles as the burn-in period. The chain was thinned to reduce serial correlation by saving one observation in every 30 cycles and thus the posterior sample contained 1000 samples for post-MCMC analysis. The simulation experiment was replicated five times. Herein, we only report the result of one replicate because there is very little variation in the mapping result among the replicates.

Figure 2 shows the profiles for QTL intensity and T^2 statistic estimated with Bayesian analysis along with the true locations of simulated QTL, where the critical value of T^2 is 11·05, that is, the critical value of chi-square distribution with 5 degree of freedom at the significance level of 5%. It can be seen that either the QTL intensity or test statistic profile in Bayesian analysis was able to detect nine simulated QTLs clearly. In contrast, interval-mapping analysis only detected seven simulated QTLs (see Fig. 3). Although the two analyses have no power to separate closely linked QTL within the same interval of markers, such as the QTL3 and QTL4, the Bayesian analysis easily detected the linked closely QTL that located in different marker intervals, such as QTL6 and QTL7.

The estimated QTL locations and effects obtained from Bayesian analysis are summarized in Table 2 along with the true parameters. Clearly, Bayesian analysis is capable of accurately estimating the QTL locations and effects. The posterior means (standard deviations) for population mean and covariance matrix of RR effects for individual-specific time-dependent environmental errors are estimated as

$$\hat{\boldsymbol{\mu}} = [\,8{\cdot}08(0{\cdot}17) \quad 24{\cdot}87(0{\cdot}20) \quad 12{\cdot}56(0{\cdot}18) \quad 0{\cdot}93(0{\cdot}19) \quad 24{\cdot}69(0{\cdot}15)\,]^{\mathrm{T}}$$

and

$$\hat{\Sigma}_{\boldsymbol{\beta}} = \begin{bmatrix} 0{\cdot}789(0{\cdot}368) & -0{\cdot}245(0{\cdot}282) & -0{\cdot}135(0{\cdot}265) & -0{\cdot}224(0{\cdot}251) & 0{\cdot}096(0{\cdot}220) \\ -0{\cdot}245(0{\cdot}282) & 0{\cdot}769(0{\cdot}391) & 0{\cdot}317(0{\cdot}248) & 0{\cdot}250(0{\cdot}240) & 0{\cdot}272(0{\cdot}215) \\ -0{\cdot}135(0{\cdot}265) & 0{\cdot}317(0{\cdot}248) & 1{\cdot}065(0{\cdot}423) & 0{\cdot}404(0{\cdot}220) & 0{\cdot}472(0{\cdot}235) \\ -0{\cdot}224(0{\cdot}251) & 0{\cdot}250(0{\cdot}240) & 0{\cdot}404(0{\cdot}220) & 0{\cdot}616(0{\cdot}291) & 0{\cdot}179(0{\cdot}206) \\ 0{\cdot}096(0{\cdot}220) & 0{\cdot}272(0{\cdot}215) & 0{\cdot}472(0{\cdot}235) & 0{\cdot}179(0{\cdot}206) & 0{\cdot}657(0{\cdot}257) \end{bmatrix},$$

respectively. The posterior mean (standard deviation) of residual variance is $\hat{\boldsymbol{\sigma}}^2_\varepsilon = 1{\cdot}985$ (0·098). Although there are relatively large estimated errors for the covariance matrix of RR effects for individual-specific time-dependent environmental errors, nine detected QTLs collectively contribute 0·891 of the total accumulative phenotypic variance calculated by $\hat{\mathbf{V}} = \boldsymbol{\psi}^{\mathrm{T}}\hat{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\psi} + \mathbf{I}\hat{\sigma}^2_\varepsilon$, which is very close to the true value used in the simulation. In contrast, the interval-mapping analysis overestimated the effects of detected QTLs (see Table 3) and is unable to estimate the total accumulative proportion of phenotypic variance

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

Table 2. *The estimated posterior means* (*posterior standard deviations*) *for QTL positions and regression effects for B-spines obtained with Bayesian B-spline mapping*

	Estimates					
QTL no.	Position	α_1	α_2	A_3	α_4	α_5
1	24	2·19 (0·34)	0·40 (0·43)	1·53 (0·40)	0·68 (0·43)	2·45 (0·32)
2	56	2·74 (0·43)	−1·07 (0·50)	1·50 (0·44)	1·31 (0·48)	3·97 (0·39)
3	–	–	–	–	–	–
4	150	2·00 (0·63)	2·46 (0·72)	1·17 (0·62)	−1·36 (0·74)	2·93 (0·57)
5	266	−0·77 (0·37)	−0·33 (0·46)	0·97 (0·43)	−0·83 (0·48)	2·51 (0·39)
6	332	3·55 (0·67)	−2·01 (0·59)	−0·05 (0·48)	2·14 (0·58)	3·33 (0·62)
7	340	1·92 (0·61)	1·76 (0·69)	0·71 (0·45)	2·10 (0·66)	−2·07 (0·60)
8	474	−2·15 (0·36)	−0·75 (0·43)	4·25 (0·38)	−3·25 (0·41)	−2·73 (0·32)
9	518	2·39 (0·51)	2·40 (0·63)	1·92 (0·49)	1·23 (0·52)	0·91 (0·52)
10	572	1·17 (0·46)	1·34 (0·55)	2·59 (0·51)	2·90 (0·51)	3·08 (0·42)

– indicates that the position or regression effect cannot be estimated by the Bayesian B-spline mapping.

contributed by the detected QTL, since in this framework there are no unique and accurate estimates of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and σ_{ε}^2.

For the simulated dataset, we replace the B-spline with Legendre polynomials of orders 4, 5 and 6, respectively, to identify ten simulated QTLs. Legendre polynomials of order 4 and the B-spline mentioned above have the same number of regression coefficients. Meanwhile, we expect that polynomials of orders 5 and 6 yield higher goodness of fit to the simulated dataset than that of order 4. Fitting results indicate that although QTL intensity profiles for the three polynomials (not shown) all arise clear signal at each simulated locus, significant loci identified with these three polynomials are less than those with Bayesian B-spline mapping. As expected, polynomials of higher order may find more QTLs than those of low orders (Figure 4).

Next, we generate the simulated data with Legendre polynomials and take a look at the detecting result with Bayesian B-spline mapping. For the same design of experiment implemented above, we describe changes in population mean, QTL genetic effects and individual-specific time-dependent environmental errors with Legendre polynomial of order 3. The regression effects for ten simulated QTLs are listed in Table 4. The population mean is $\boldsymbol{\mu} = [45 \quad 44 \quad -1 \quad -7]^T$, the covariance matrix of RR effects for individual-specific time-dependent environmental errors is

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \begin{bmatrix} 1{\cdot}042 & 0{\cdot}171 & -0{\cdot}035 & 0{\cdot}100 \\ 0{\cdot}171 & 0{\cdot}086 & 0{\cdot}041 & 0{\cdot}032 \\ -0{\cdot}035 & 0{\cdot}041 & 0{\cdot}087 & 0{\cdot}052 \\ 0{\cdot}100 & 0{\cdot}032 & 0{\cdot}052 & 0{\cdot}076 \end{bmatrix},$$

and the residual variance is $\sigma_{\varepsilon}^2 = 2{\cdot}0$. Note that measuring time points are assigned as 5, 8, 13, 21, 26, 32 and 39.

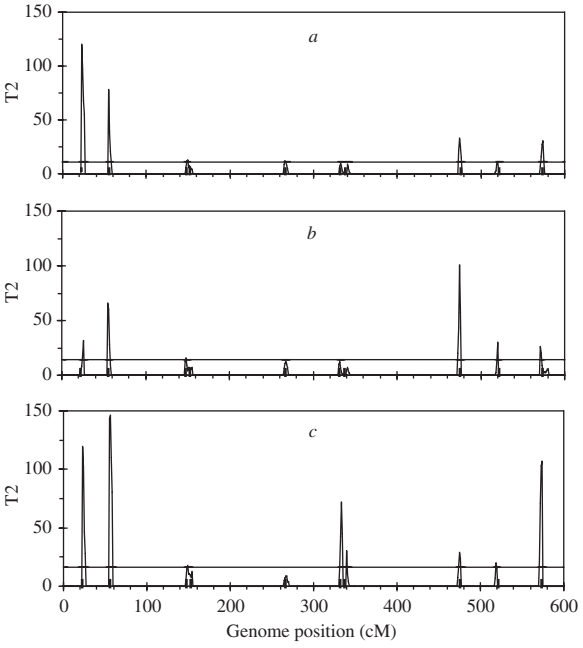

Fig. 4. The T^2 statistic profiles for the entire genome obtained with Bayesian mapping based on Legendre polynomials of 4 (*a*), 5 (*b*) and 6 (*c*) order. The true positions of the simulated QTL are represented by black needles. The horizontal lines indicate the empirical critical values when the type I error rate was 5%, which are 11·07 for (*a*), 12·59 for (*b*) and 14·07 for (*c*), respectively.

Yang & Xu (2007) have analysed the simulated data with Bayesian shrinkage estimation based on Legendre polynomials and identified nine of these simulated QTLs. For the same simulated dataset, Bayesian B-spline mapping also has the ability to obtain the same mapping result as Legendre polynomial analysis for the simulated data (results not shown). Herein, the knot points are chosen as 5, 20 and 39 in the used B-spline.

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

Table 3. *The estimated posterior means* (*standard deviations*) *of QTL regression effects for B-spline obtained with interval mapping*

	Estimates					
QTL No.	Position	α_1	α_2	A_3	α_4	α_5
1	–	–	–	–	–	–
2	57	3·98(0·82)	0·26(0·36)	3·66(0·67)	0·74(0·56)	6·10(0·82)
3	149	5·52(0·66)	2·08(0·39)	1·28(0·62)	−2·82(0·49)	5·58(0·79)
4	–	–	–	–	–	–
5	267	1·10(0·82)	−0·46(0·34)	2·32(0·61)	−0·16(0·51)	3·46(0·86)
6	–	–	–	–	–	–
7	339	4·58(0·82)	0·10(0·42)	1·02(0·64)	3·14(0·53)	0·86(0·86)
8	475	0·90(0·69)	1·20 (0·46)	6·02(0·71)	−1·78(0·48)	−0·24(0·76)
9	521	2·42(0·65)	3·08(0·39)	5·32(0·74)	0·90(0·57)	1·58(0·78)
10	569	2·84(0·76)	2·42(0·39)	3·88(0·67)	3·00(0·42)	3·56(0·82)

– indicates that the position or regression effect cannot be estimated by the interval mapping.

Table 4. *QTL regression effects for Legendre polynomials used in the simulation experiment*

QTL	α_0	α_1	α_2	α_3
1	0·00	1·65	2·52	1·20
2	2·34	2·08	1·37	1·18
3	2·55	1·36	−2·02	−1·27
4	1·05	−2·57	1·24	−1·10
5	1·12	1·68	1·25	0·00
6	2·94	0·00	−1·68	1·72
7	1·82	−0·80	−1·20	−0·80
8	1·28	−0·50	−1·17	1·32
9	2·00	−1·25	0·00	−1·18
10	1·75	1·30	−1·45	1·17

4. Example

The data published by Ma *et al*. (2002) are used to demonstrate the application of the proposed method to real data from dynamic traits. The analysed trait is the growth of the stem diameter of *Populus* trees measured annually in 11 years. The mapping population consists of 78 progeny of pseudo-backcross derived from the triple hybridization of *Populus*. A genetic linkage map has been constructed using 90 markers distributed on 23 linkage groups. Since the growth of the stem diameter follows the S-shaped curve, Ma *et al*. (2002) fit change of QTL genotypic value by Logistic curve within the interval-mapping framework. Subsequently, Yang *et al*. (2006, 2009) replaced Logistic curve with Legendre polynomials and B-splines, respectively, and obtained good results.

In the Bayesian mapping procedure used here, the maximum number of QTLs is set to 100. Initial values of all unknown parameters and all the hyperparameters are determined according to the method used in the simulation. The algorithm used in this analysis is the same as that used in the analysis of simulated data. The B-spline with three knots and polynomial segments of order 2 is used to model the dynamic of population mean and genetic effect for each QTL in model (1). When the knots are chosen as 1, 6 and 11 years, the proposed method detects seven QTLs on linkage group D3-1, D4, D9, D10, D15 and D17 that control the growth trajectory of stem diameter of *Populus*. Parameter estimates of detectable QTLs are listed in Table 5. As a comparison, we also choose polynomials of orders 3, 4 and 5 and find only the polynomials of order 4 identifies the same QTLs as what we have from the B-spline approach (results not shown), demonstrating the capability of B-spline to replace polynomial in Bayesian mapping for dynamic quantitative traits. Moreover, in the B-spline, if internal knot points are taken as 5 and 7 years, respectively, mapping results are almost the same for either the Bayesian-mapping method or the interval-mapping method. This implies that the efficiency of the proposed method strongly depends on the choice of internal knots. For the real dataset, interval mapping based on model (1) with a single QTL only locates two QTLs on linkage group D10 and D17. However, functional mapping developed by Ma *et al*. (2002) only find QTLs on linkage group D10. Yang *et al*. (2009) also detect seven QTLs with B-spline-based interval mapping, but only two QTLs on linkage group D4 and D10 are consistent with our findings in this paper. This difference may be due to different mapping methods and the covariance structure specification for residuals.

5. Discussion

We successfully extended the Bayesian shrinkage method to mapping multiple QTLs for dynamic traits, and demonstrated by the simulated data analysis that

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

Table 5. *Parameter estimates of QTLs obtained from Bayesian B-spline mapping for stem diameters in* Populus

QTL no.	Linkage group	Marker interval	Position	Regression genetic effect (standard error)			
1	D3-1	CT/CAG-350R ~ CT/CAG-505	89·5	−0·075(0·032)	0·201(0·205)	−0·106(0·032)	0·023(0·103)
2	D4	AQ2-1220 ~ TT/CGA-395	243·5	0·053(0·033)	−0·233(0·124)	0·157(0·108)	0·116(0·144)
3	D9	Q9-1700R ~ CG/CCT-655	89·3	−0·014(0·039)	−0·334(0·289)	0·088(0·035)	0·228(0·162)
4	D10	TC/CAA-540 ~ TC/CTG-750	60·4	0·089(0·034)	−0·221(0·242)	0·061(0·035)	0·217(0·128)
5	D15	TT/CGT_640 ~ AA/CGT_690	79·7	0·068(0·041)	−0·375(0·167)	0·255(0·041)	0·190(0·159)
6	D15	AV14_1500R ~ TC/CAC_420	133·6	−0·088(0·040)	0·247(0·181)	−0·070(0·037)	−0·012(0·142)
7	D17	CA/CCG-820R ~ TC/CAG-350R	64·6	−0·018(0·018)	0·360(0·242)	−0·138(0·018)	−0·073(0·154)

Bayesian mapping is able to effectively and accurately detect QTL governing dynamic traits. The Bayesian-mapping analysis of dynamic traits was based on RR model, in which B-spline was chosen to simultaneously describe the dynamics of population means, genetic effects of multiple QTL and other environmental factors over time. This overcomes the disadvantages of Logistic curve used in traditional functional mapping, due to its non-additivity. Our model can be considered as a general framework for mapping QTL, for instance, which could be reduced to repeatability model for multiple traits when the order of B-spline was set to 0, and could be reduced to a single-trait model when all the covariates of measuring time were 1.

As illustrated in simulations and real data analysis, B-spline can capture complicated patterns but it is highly sensitive to the choice of knots. Therefore, the choice of knots should be the key elements of model specification in Bayesian B-spline mapping. Theoretically, too many knots lead to the over-fitting of the data, while too few knots lead to under-fitting. Some authors have proposed automatic schemes for optimizing the number and the positions of the knots (Friedman & Silverman, 1989; Kooperberg & Stone, 1991, 1992). In particular, the choice of knots can be easily dealt with by using penalized or Bayesian estimation methods (e.g. Whaba, 1990; Rupert *et al.*, 2003). Due to many different B-splines nested for each QTL and individual-specific environmental effects, however, Bayesian shrinkage analysis with the choice of a large number of knots will become computationally expensive and thus difficult to map dynamic trait loci. If only those positions with significant genetic effects are drawn with Bayesian model choice (Yi *et al.*, 2005; Min & Czado, 2011), the computational time can be greatly reduced. Here, we use the same B-spline to fit the changes in population means, genetic effects of multiple QTLs and other environmental factors over time, according to the averages fit of phenotypic values at different measurement points. Apparently, this choice for B-spline may not be optimal.

Since regression coefficients of B-spline determine the shape of dynamic trajectory, Bayesian B-spline mapping proposed here infers the QTL controlling the dynamic trajectory of dynamic quantitative traits by the Wald test statistic derived from posterior samples for QTL regression effects. By substituting each group of posterior sampling values for QTL regression effects into B-spline, we can obtain posterior sample for QTL effects at any dynamic point or within changing process of interest and by which infer the QTLs controlling any dynamic point or changing process of dynamic traits. According to the relationship between QTL regression effects and the parameters in biological meaningful model such as logistic curve, as derived in Yang & Xu (2007), we can also infer those QTLs controlling the biological meaningful characters. Therefore, it is easy to answer the criticism on the polynomial approach when applied to any curve is the interpretability of the polynomial regression coefficients in mapping QTL for dynamic traits.

6. Acknowledgment

The research was partially supported by the National Natural Science Foundation of China (30972077 and 31110103065) to R. Y.

Appendix A. The B-spline

Given k knots with $t_0 \leqslant t_1 \leqslant \ldots \leqslant t_{k-1}$, B-spline of degree p is a parametric curve composed of a linear combination of basis B-splines $\psi_{i,p}(t)$:

$$B(t) = \sum_{i=0}^{k-p-2} b_i \psi_{i,p}(t),$$

where b_i are called control points or de Boor points.

The $k-p-1$ basis B-splines of degree p can be defined, for $p = 0, 1, \ldots, k-2$, using the Cox–de Boor recursion formula. Basis functions of degree $p=0$ have values of unity for all points in a given interval, and zero otherwise. For the jth interval given by knots t_j and t_{j+1} with $t_j < t_{j+1}$,

$$\psi_{j,0}(t) = \begin{cases} 1, & \text{if } t_j < t < t_{j+1} \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } j = 0, 1, \ldots, k-2.$$

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

Higher-degree basis functions $\psi_{i,\,p}(t)$ for $p>0$, are then determined recursively from the values of the lower degree basis functions and the width of the adjoining intervals between knots. The general relationship is

$$\psi_{j,p}(t)=\frac{t-t_j}{t_{j+p}-t_j}\psi_{j,p-1}(t)+\frac{t_{j+p+1}-t}{t_{j+p+1}-t_{j+1}}\psi_{j+1,p-1}(t),$$
$$\text{for } j=0,\ 1,\ \ldots,\ k-p-2.$$

Obviously, there are a limited number of non-zero basis functions of lower order for each p.

Appendix B. Conditional posterior densities used for Gibbs samplers

The conditional posterior distribution for each unknown parameter can be derived from the joint posterior distribution density by fixing other parameters.

The conditional posterior distribution of μ is multivariate normal with mean

$$\mathrm{E}(\boldsymbol{\mu}|\ldots)=(n\boldsymbol{\psi}\mathbf{V}^{-1}\boldsymbol{\psi}^{\mathrm{T}})^{-1}\boldsymbol{\psi}\mathbf{V}^{-1}\times\sum_{i=1}^{n}\left(\mathbf{y}_i-\sum_{j=1}^{q}x_{ij}\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\alpha}_j\right)\tag{B1}$$

and covariance matrix

$$\mathrm{Var}(\boldsymbol{\mu}|\ldots)=(n\boldsymbol{\psi}\mathbf{V}^{-1}\boldsymbol{\psi}^{\mathrm{T}})^{-1},\tag{B2}$$

where the $\boldsymbol{\mu}|\ldots$ notation is a short expression for conditional variable $\boldsymbol{\mu}$ given all the data and other variables. Note that all variables appearing on the right-hand side of eqn (B1) are actual values sampled in a previous iteration.

The conditional posterior distribution of the jth QTL effects is multivariate normal with mean

$$\mathrm{E}(\boldsymbol{\alpha}_j|\ldots)=\sum_{i=1}^{n}\mathbf{A}_j\boldsymbol{\psi}(\mathbf{V}+x_{ij}^2\boldsymbol{\psi}^{\mathrm{T}}\mathbf{A}_j\boldsymbol{\psi})^{-1}x_{ij}\times\left(\mathbf{y}_i-\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\mu}-\sum_{j'\neq j}^{q}x_{ij'}\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\alpha}_{j'}\right),\tag{B3}$$

and covariance matrix

$$\mathrm{Var}(\boldsymbol{\alpha}_j|\ldots)=\mathbf{A}_j-\sum_{i=1}^{n}x_{ij}^2\mathbf{A}_j\boldsymbol{\psi}(\mathbf{V}+x_{ij}^2\boldsymbol{\psi}^{\mathrm{T}}\mathbf{A}_j\boldsymbol{\psi})^{-1}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{A}_j.\tag{B4}$$

The conditional posterior distribution of $\boldsymbol{A_j}$ is inverse Wishart with degrees of freedom b_0+1 and covariance matrix $(\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^{T}+\boldsymbol{\Gamma}_{\mathbf{A}}{}^{-1})^{-1}$, i.e.

$$\mathbf{A}_j\sim\mathrm{IW}\left[b_0+1,\ (\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^{\mathrm{T}}+\boldsymbol{\Gamma}_{\mathbf{A}}^{-1})^{-1}\right].\tag{B5}$$

The conditional posterior distribution of the individual-specific $\boldsymbol{\beta_i}$ is a multivariate normal distribution with mean

$$\mathrm{E}(\boldsymbol{\beta}_i|\ldots)=\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\psi}\mathbf{V}^{-1}(\mathbf{y}_i-\mathbf{U}_i)\tag{B6}$$

and covariance matrix

$$\mathrm{Var}(\boldsymbol{\beta}_i|\ldots)=\boldsymbol{\Sigma}_{\boldsymbol{\beta}}-\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\psi}\mathbf{V}^{-1}\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}.\tag{B7}$$

The conditional posterior distribution of $\boldsymbol{\Sigma_\beta}$ is inverse Wishart, i.e.

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\sim\mathrm{IW}\left[d_0+n,\left(\sum_{i=1}^{n}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^{\mathrm{T}}+\boldsymbol{\Gamma}_{\boldsymbol{\beta}}^{-1}\right)^{-1}\right]\tag{B8}$$

The conditional posterior distribution for residual variance σ_ε^2 is a scaled inverse chi-square with parameters $n(m+1)+v_e$ and $h_eS_e+\left(\sum_{i=1}^{n}\boldsymbol{\varepsilon}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i\right)^{-1}$, where $h_e=v_e+n(m+1)$ and

$$\boldsymbol{\varepsilon}_i=\mathbf{y}_i-\mathbf{U}_i-\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\beta}_i.\tag{B9}$$

References

de Boor, C. (2001). *A Practical Guide to Splines*, Vol. 27, 2nd edn. New York: Springer-Verlag.

Friedman, J. H. & Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling. *Technometrics* **31**, 3–21.

Gelman, A., Carlin, J. B., Stern H. S. & Rubin, D. B. (1995). *Bayesian Data Analysis*. New York: Chapman & Hall.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

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

Jamrozik, J., Schaeffer, L. R. & Dekkers, J. C. M. (1997). Genetic evaluation of dairy cattle using test day yields and random regression model. *Journal of Dairy Science* **80**, 1217–1226.

Jin, T., Li, J., Guo, Y., Zhou, X., Yang, R. & Wu, R. (2010). An optimal strategy for functional mapping of dynamic trait loci. *Genetical Research* **3**, 1–8.

Kooperberg, C. & Stone, C. J. (1991). A study of logspline density estimation. *Computational Statistics and Data Analysis* **12**, 327–348.

Kooperberg, C. & Stone, C. J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics* **1**, 301–328.

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

Macgregor, S., Knott, S. A., White, I. & Visscher, P. M. (2005). Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. *Genetics* **171**, 1365–1376.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

Min, A. & Czado, C. (2011). Bayesian model selection for D-vine pair-copula constructions. *Canadian Journal of Statistics* **2**, 239–258.

https://doi.org/10.1017/S0016672312000249 Published online by Cambridge University Press

Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). *Semi-parametric Regression*. New York: Cambridge University Press.

Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Science* **86**, 35–45.

Sen, Ś. & Churchill, G. A. (2001). A statistical framework for quantitative trait mapping. *Genetics* **159**, 371–387.

Sillanpää, M. J. & Arjas, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**, 1373–1388.

Sillanpää, M. J. & Arjas, E. (1999). Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* **151**, 1605–1619.

Wang, H., Zhang, Y. M., Li, X., Masinde, G. L., Mohan, S., Baylink, D. J. & Xu, S. (2005). Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics* **170**, 465–480.

Wahba, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics 59*. Philadelphia, PA: SIAM.

Wu, R., Ma, C. X., Littell, R. C., Wu, S. S., Yin, T., Huang, M., Wang, M. & Casella, G. (2002). A logistic mixture model for characterizing genetic determinants causing differentiation in growth trajectories. *Genetical Research* **79**, 235–245.

Wu, R. L., Ma, C. X., Lin, M., Wang, Z. H. & George, C. (2004). Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. *Biometrics* **60**, 729–738.

Wu, R., Ma, C. X., Zhao, W. & Casella, G. (2003). Functional mapping for quantitative trait loci governing growth rates: a parametric model. *Physiological Genomics* **14**, 241–249.

Yang, J., Wu, R. & Casella, G. (2009). Nonparametric functional mapping of quantitative trait loci. *Biometrics* **65**, 30–39.

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

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

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

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

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

Yi, N., Yandell, B. S., Churchill, G. A., Allison, D. B., Eisen, E. J. & Pomp, D. (2005). Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* **170**, 1333–1344.

Zhang, Y. M. & Xu, S. (2005). Advanced statistical methods for detecting multiple quantitative trait loci. *Recent Research Developments in Genetics and Breeding* **2**, 1–23.