Heredity (2008) 101, 321–328

www.nature.com/hdy

ORIGINAL ARTICLE

A nonlinear mixed-effect mixture model for functional mapping of dynamic traits

W Hou[1], H Li[1], B Zhang[2], M Huang[2] and R Wu[1,2,3]
[1]*Department of Statistics, University of Florida, Gainesville, FL, USA;* [2]*The Key Laboratory of Forest Genetics and Gene Engineering, Nanjing Forestry University, Nanjing, Jiangsu, People's Republic of China and* [3]*UF Genetics Institute, University of Florida, Gainesville, FL, USA*

Functional mapping has emerged as a next-generation statistical tool for mapping quantitative trait loci (QTL) that affect complex dynamic traits. In this article, we incorporated the idea of nonlinear mixed-effect (NLME) models into the mixture-based framework of functional mapping, aimed to generalize the spectrum of applications for functional mapping. NLME-based functional mapping, implemented with the linearization algorithm based on the first-order Taylor expansion, can provide reasonable estimates of QTL genotypic-specific curve parameters (fixed effect) and the between-individual variation of these parameters (random effect). Results from simulation studies suggest that the NLME-based model is more general than traditional functional mapping. The new model can be useful for the identification of the ontogenetic patterns of QTL genetic effects during time course.
Heredity (2008) **101,** 321–328; doi:10.1038/hdy.2008.53; published online 9 July 2008

Keywords: mixture model; nonlinear mixed-effect model; growth trajectories; QTL mapping; poplar

Introduction

Dynamic traits change their phenotypes with time or other independent variables. A profound understanding of the genetic control of a dynamic trait should include the timing of the underlying genes to turn on and off in a time course, the duration of genetic main and interaction effects, the pleiotropic effects of the genes on various developmental events, and sensitivity of the genes in response to environmental signals. Functional mapping, emerging as a next-generation statistical method for genetic mapping, has proven to be powerful for addressing the above-mentioned issues by mapping ontogenetic quantitative trait loci (QTL) for complex dynamic traits (Ma *et al*., 2002; Wu *et al*., 2003a, b, 2004a, b, c; reviewed in Wu and Lin, 2006; Yang *et al*., 2006; Yang and Xu, 2007). The fundamental idea of functional mapping is to jointly model the mean covariance structure within a mixture model framework for dynamic traits longitudinally measured at different time points by using parametric or nonparametric approaches. If there exist biologically meaningful mathematical equations for longitudinal curves, such as growth equation (West *et al*., 2001), biexponential curve for HIV dynamics (Ho *et al*., 1995), Fourier series approximation for cell cycle (Spellman *et al*., 1998) and power equation for allometric scaling (West *et al*., 1997), parametric approaches can be implemented to estimate the mathematical parameters that define the shapes of curves for a QTL genotype expressed as a mixture component, instead of directly estimating the QTL genotypic means at all different time points.

As a type of time series data, longitudinal traits exhibit a strong autocorrelation between successive time points. Structuring such a time-dependent covariance matrix by a stationary or nonstationary approach can increase the model's stability, robustness and statistical power to detect QTL. The approaches for modeling the covariance structure in functional mapping have been based on autoregressive (AR; Ma *et al*., 2002) or antedependence models (Zhao *et al*., 2005). In all such modeling work, repeated measurements are assumed to be independent among different subjects and, thus, only within-subject covariance structures have been considered. In a general setting of longitudinal data analysis, three components of random variability in the modeling process should be distinguished, that is, the random effects that stem from heterogeneity between individual profiles, serial correlation between observations within sampling unit and measurement error (Davidian and Giltinan, 1995, 2003; Diggle *et al*., 2002). Thus, although the approximation of a covariance structure merely based on serial correlations in current functional mapping is thought to be parsimonious, it may have serious limitations that would prevent a wide implication of functional mapping. These limitations are shown in the following aspects.

First, the mathematical parameters for individual longitudinal curves with the same QTL genotype may not be independent among subjects. The ignorance of among-subject dependence for the curve parameters would overestimate the genetic effect of QTL on longitudinal trajectories. To draw a valid statistical inference for longitudinal data, random effects that capture heterogeneity among subjects should be considered, in

Correspondence: Professor R Wu, Department of Statistics, University of Florida, Gainesville, FL 32611, USA.
E-mail: rwu@stat.ufl.edu
Received 10 March 2008; revised 25 April 2008; accepted 30 April 2008; published online 9 July 2008

a conjunction with direct modeling of the within-subject correlation (Chi and Reinsel, 1989; Schabenberger, 1995). Second, the curve parameters may be affected by an array of biological or demographic covariates, such as age, sex, race and body weight. For those biological covariates, it is possible that they are under the control of genetic systems that are the same as, or different from, those for the longitudinal traits under consideration. Testing the difference of genetic control for different traits or processes presents an interesting and challenging genetic issue (Lynch and Walsh, 1998).

In this article, nonlinear mixed-effect (NLME) models, or hierarchical nonlinear models, will be incorporated into the context of functional mapping based on a mixture model, aimed to circumvent the above-mentioned limitations of current functional mapping strategies. Since their first emergence in the early 1980s (Beal and Sheiner, 1982; reviewed in Davidian and Giltinan, 1995), NLME models have quickly become a popular statistical method for studying longitudinal data (Lindstrom and Bates, 1990). More recently, a number of extensions and modifications to better suit new challenges have been developed (Davidian and Giltinan, 1995; Vonesh *et al*., 2002; Wu, 2002, 2004a, b). The major advantage of NLME models lies in their capacity and flexibility to model various structures of covariance matrices. Also, they display a unique ability to accommodate a general intraindividual covariance structure for unbalanced data where measurements are sparse for some subjects and different subjects receive different measurement patterns.

The application of NLME models is a promising approach for improving parameter estimation and valid inferences of longitudinal data including pharmacokinetics and HIV dynamics. Mixed-effect models may also be appealing to genetic studies by increasing the flexibility of QTL mapping for response curves (Rodriguez-Zas *et al*., 2002; Malosetti *et al*., 2006). However, because the statistical properties behind this technique have not been explored, its application lacks sensible justifications. Also, although these published works can model interindividual variation in curve parameters, they have a limited flexibility to model the common genetic basis shared by biologically meaningful curve parameters and other biological variables. The purpose of this study is to develop NLME models for estimating the ontogenetic pattern of the genetic control of complex dynamic traits and examine the statistical behavior of this technique through extensive simulation studies. We will integrate NLME models and mixture models within the framework of functional mapping to increase the vision of this mapping method.

Functional mapping

Nonlinear mixed-effects model

The purpose for the development of functional mapping is to map the temporal effects of QTL on longitudinal traits. Consider a mapping population of n individuals, in which a total of J QTL genotypes at different loci are segregating to affect time-dependent phenotypes of a trait. All the individuals are genotyped for multiple polymorphic markers that construct a genetic linkage map and phenotyped for a longitudinal trait measured at a finite set of time points.

Let $\mathbf{t}_i = \{t_{i\tau}\}_{\tau=1}^{T_i}$ be the vector of times for individual i measured at T_i time points and $\mathbf{y}_i = \{y_i(t_{i\tau})\}_{\tau=1}^{T_i}$ be the vector for longitudinal phenotypic measurements of individual i. The time points may be unbalanced among individuals and unequally spaced during measurements. The phenotypic value of the trait for individual i affected by the putative QTL can be described by a two-stage NLME model, expressed as:

Stage 1 (individual-level model): The response value of individual i across different time points is described by

$$\mathbf{y}_i = \sum_{j=1}^{J} \xi_{ij} g(\beta_{ij}; \mathbf{t}_i) + \varepsilon_i \quad (1)$$

where ξ_{ij} is the indicator variable defined as 1 if individual i carries QTL genotype j and 0 otherwise, g is a nonlinear function of β_{ij} and $\mathbf{t}_i$, β_{ij} is a $(q \times 1)$ vector of individual-specific unknown curve parameters and ε_i is a $(1 \times T_i)$ error term, usually assumed to have a normal distribution with mean vector 0 and within-individual covariance matrix Σ_i. Note that Σ_i is a $(T_i \times T_i)$ serial covariance matrix, which can be structured by a set of parameters (Diggle *et al*., 2002).

Stage 2 (population level): The parameters that define the curve shape of individual i with QTL genotype j can be expressed as

$$\beta_{ij} = \mathbf{A}_i \beta_j + \mathbf{B}_i \mathbf{b}_{ij} \quad (2)$$

where β_j is a $(p \times 1)$ vector of the unknown population parameters for QTL genotype j, $\mathbf{b}_{ij}$ is a $(k \times 1)$ vector of the random effects, assumed to be normally distributed with mean vector 0 and $(k \times k)$ between-individual covariance matrices $\mathbf{D}_j$, and $\mathbf{A}_i$ and $\mathbf{B}_i$ are design matrices of size $q \times p$ and $q \times k$ for β_j and $\mathbf{b}_{ij}$, respectively. This stage captures the interindividual systematic and random variation. This model in a general form can handle any kinds of nonlinear function and the design matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ can vary for different groups, covariates or even for different individuals.

Mixture model-based likelihood

The statistical foundation for QTL mapping with molecular markers is a finite mixture model. According to the mixture model, the trait value of an individual is assumed to have arisen from one (and only one) of J QTL genotype groups or mixture components, each component with a relative proportion and being modeled by a normal distribution density.

The likelihood of unknown parameters given the longitudinal measurements ($\mathbf{y}$) and marker information ($\mathbf{M}$) for the mapping population is formulated, in terms of a mixture model, as

$$\begin{aligned} L(\omega, \beta, \mathbf{b}_i, \theta | \mathbf{y}, \mathbf{M}) &= \prod_{i=1}^{n} \sum_{j=1}^{J} [\omega_{j|i} f_j(\mathbf{y}_i | \beta_j, \mathbf{b}_{ij}, \theta)] \\ &= \int \left(\sum_{j=1}^{J} \omega_{j|i} f_j(\mathbf{y}_i | \beta_j, \mathbf{b}_{ij}, \theta) \right) f(\mathbf{b}_{ij} | \mathbf{D}_j) \, d(\mathbf{b}_{ij}) \quad (3) \\ &= \sum_{j=1}^{J} \omega_{j|i} \int f_j(\mathbf{y}_i | \beta_j, \mathbf{b}_{ij}, \theta)) f(\mathbf{b}_{ij} | \mathbf{D}_j) d\mathbf{b}_{ij}, \end{aligned}$$

where $\omega = \{\omega_{j|i}\}_{j=1}^{J}$ are the QTL genotype frequencies which are constrained to be nonnegative and sum to

unity, $\beta = \{\beta_j\}_{j=1}^{J}$ and $\mathbf{b}_i = \{\mathbf{b}_{ij}\}_{j=1}^{J}$ are the component (or QTL genotype)-specific parameters, with β_j and $\mathbf{b}_{ij}$ being specific to QTL genotype j, θ is the common parameters to all QTL genotypes, which is the set of unknown parameters that construct $\mathbf{D}_j$ and Σ_i.

The mixture proportion or QTL genotype frequency $\omega_{j|i}$ depends on the type of mapping population, such as the backcross, recombinant inbred lines, F_2 or natural population. The frequencies of QTL genotypes can be inferred by observed marker genotypes because markers and QTL are assumed to be cosegregating in the mapping population. Assume that a putative QTL is located between two flanking markers that bracket the QTL. Thus, the mixture proportions, $\omega_{j|i}$, can be expressed as the conditional probabilities of QTL genotypes given the flanking marker genotype of individual i. The conditional probability can be derived in terms of recombination fractions between the QTL and each of the two markers and between the two markers.

Computational algorithm

There are three types of parameters that define the likelihood (3), which are the mixing proportions of QTL genotypes conditional on marker genotypes (ω), QTL genotype-specific curve parameters (β, $\mathbf{b}_i$) and the covariance-structuring parameters (θ). The mixing proportions (ω) are expressed in terms of the recombination fractions between the markers and QTL and, therefore, the genomic location of the QTL (converted by the map function). In practice, ω, that is, the QTL location, can be treated as a constant because a putative QTL can be searched at every 1 or 2 cM on an interval of two flanking makers throughout the entire linkage group. The log-likelihood ratio (LR) test statistics are plotted against the linkage map distance. The linkage map position corresponding to a peak of the log-LR plot will be determined as the maximum-likelihood estimate (MLE) of the QTL location. Thus, on each scanning location of a QTL, the mixture likelihood will only depend on β_j, $\mathbf{b}_{ij}$ and θ. This grid approach is computationally simple, but cannot provide the estimate of the confidence interval of the QTL location estimate. Chen (2005) derived an algorithm for simultaneously estimating the standard errors and confidence intervals of the estimates of QTL effects and locations within the mixture model framework.

From the likelihood (3), the estimates of β_j, $\mathbf{b}_{ij}$ and θ will need to jointly maximize the posterior distribution function $f_j(\mathbf{y}_i|\beta_j, \mathbf{b}_{ij}, \theta) f(\mathbf{b}_{ij}|\mathbf{D})$ weighted by $\omega_{j|i}$. But an inference based on the maximization of this distribution is difficult because its expectation is not linear for these unknowns. A few statistical approaches have been developed to obtain the MLEs of β_j, $\mathbf{b}_{ij}$ and θ, and they include numerical evaluation of the integral (Davidian and Giltinan, 1995, 2003), Monte Carlo expectation maximization (EM) algorithm (Wu, 2002, 2004a) and approximations to the nonlinear likelihood function (Tierney and Kadane, 1986; Lindstrom and Bates, 1990; Wolfinger, 1993). Here, we will use a linearization approximation method by using the first-order Taylor expansion to approximate the nonlinear expectation function (Beal and Sheiner, 1982; Lindstrom and Bates, 1990).

For individual i, the mixture-based NLME models (1) and (2) are rewritten into a single equation, expressed in matrix notation as

$$\mathbf{y}_i = \sum_{j=1}^{J} \xi_{ij} g(\beta_j, \mathbf{b}_{ij}; \mathbf{t}_i) + \varepsilon_i. \tag{4}$$

By taking the first-order Taylor expansion of $g(\beta_j, \mathbf{b}_{ij}; \mathbf{t}_i)$, Equation (4) is linearized to become a linear mixed-effect (LME) model expressed as

$$\tilde{y}_i = \sum_{j=1}^{J} \xi_{ij} (\mathbf{W}_i \beta_j + \mathbf{Z}_i \mathbf{b}_{ij}) + \varepsilon_i \tag{5}$$

where

$$\tilde{y}_i = \mathbf{y}_i - \sum_{j=1}^{J} \xi_{ij} g(\hat{\beta}_j, \hat{\mathbf{b}}_{ij}; \mathbf{t}_i) + \sum_{j=1}^{J} \xi_{ij} (\mathbf{W}_i \hat{\beta}_j + \mathbf{Z}_i \hat{\mathbf{b}}_{ij}),$$

with the $\mathbf{W}_i$ and $\mathbf{Z}_i$ composed of time-dependent elements:

$$W_i(\mathbf{t}_i) = \frac{\partial g(\beta_j, \mathbf{b}_{ij}; \mathbf{t}_i)}{\partial \beta_j^{\mathrm{T}}},$$

and

$$Z_i(\mathbf{t}_i) = \frac{\partial g(\beta_j, \mathbf{b}_{ij}; \mathbf{t}_i)}{\partial \mathbf{b}_{ij}^{\mathrm{T}}}.$$

According to Laird and Ware (1982), the estimates of $\mathbf{b}_{ij}$ and β_j under the LME model are approximated by

$$\hat{\mathbf{b}}_{ij} = \hat{\mathbf{D}}_j \hat{\mathbf{Z}}_i^{\mathrm{T}} (\Sigma_i + \hat{\mathbf{Z}}_i \hat{\mathbf{D}}_j \hat{\mathbf{Z}}_i^{\mathrm{T}})^{-1} (\tilde{y}_i - \hat{\mathbf{W}}_i \hat{\beta}_j) \tag{6}$$

$$\begin{aligned} \hat{\beta}_j = &\left(\sum_{i=1}^{n} \hat{\mathbf{W}}_i^{\mathrm{T}} (\Sigma_i + \hat{\mathbf{Z}}_i \hat{\mathbf{D}}_j \hat{\mathbf{Z}}_i^{\mathrm{T}})^{-1} \hat{\mathbf{W}}_i \right)^{-1} \\ &\times \sum_{i=1}^{n} \hat{\mathbf{W}}_i^{\mathrm{T}} (\Sigma_i + \hat{\mathbf{Z}}_i \hat{\mathbf{D}}_j \hat{\mathbf{Z}}_i^{\mathrm{T}})^{-1} \tilde{\mathbf{y}}_i. \end{aligned} \tag{7}$$

Also, QTL genotype-specific curve parameters β_j can be estimated, along with covariance matrix parameters θ, by maximizing the approximate-likelihood function expressed as

$$L(\beta, \theta|\tilde{\mathbf{y}}_i) = \prod_{i=1}^{n} [\omega_{j|i} f_j(\tilde{\mathbf{y}}_i)],$$

where

$$\begin{aligned} f_j(\tilde{\mathbf{y}}_i) = &\frac{1}{(2\pi)^{T/2} |\Sigma_i + \hat{\mathbf{Z}}_i \hat{\mathbf{D}}_j \hat{\mathbf{Z}}_i^{\mathrm{T}}|^{1/2}} \\ &\times \exp\left[-\frac{1}{2} (\tilde{\mathbf{y}}_i - \hat{\mathbf{W}}_{ij} \beta_j)^{\mathrm{T}} \right. \\ &\left. \times (\Sigma_i + \hat{\mathbf{Z}}_i \hat{\mathbf{D}}_j \hat{\mathbf{Z}}_i^{\mathrm{T}})^{-1} (\tilde{\mathbf{y}}_i - \hat{\mathbf{W}}_{ij} \beta_j) \right]. \end{aligned}$$

The simplex algorithm implemented with the MatLab function fminsearch can be used to obtain the MLEs of β_j and θ (Lagarius *et al*., 1998).

Hypotheses

A significant advantage of functional mapping is that it can perform a number of biologically meaningful hypotheses based on the mathematical model of longitudinal curves. Most importantly, the existence of a QTL

that exerts an effect on an overall growth curve should first be tested and this can be formulated as

$$\left\{\begin{array}{l} H_0\colon \beta_j \equiv \beta(j=1,\ldots,J) \\ H_1\colon \text{at least one of the equalities above does not hold} \end{array}\right\} \tag{8}$$

at least one of the equalities above does not hold (1), where H_0 corresponds to the reduced model, in which the data can be fit by a single mathematical curve, and H_1 corresponds to the full model, in which there exist different longitudinal curves to fit the data. The log-likelihood values L_0 and L_1 under the H_0 and H_1 are calculated. The test is performed with a log-LR statistic

$$\mathrm{LR} = -2\ln(L_0/L_1). \tag{9}$$

To determine the significance of the LR test, we use the critical threshold generated by permutation tests (Churchill and Doerge, 1994). By repeatedly shuffling the relationships between marker genotypes and phenotypes, a series of the maximum log-LRs are calculated, from the distribution of which the critical threshold is obtained. The LR statistic is plotted against test locations for all the linkage groups. A location of a high peak of LR that is beyond the threshold is considered corresponding to the position of QTL.

In addition, the hypothesis test for the time at which the detected QTL turns on or off its effect on longitudinal trajectories can be performed, by comparing the difference of the expected means between different genotypes at various time points. Within the functional mapping framework, the effect of the QTL on a period of time course and its interaction with age can also be tested (Wu *et al*., 2004a).

A worked example

Mapping population

Here we reanalyzed a published data set for QTL mapping of growth trajectories (Ma *et al*., 2002) to demonstrate the utilization of NLME-incorporated functional mapping. The plant materials used were derived from the interspecific hybridization (F_1) between Eastern Cottonwood (*Populus deltoides*) and Canadian poplar (*P. euroamericana*). Different from inbred lines that need an advanced-generation design for mapping, outcrossing species like trees can make use of a controlled cross of F_1, in which genes are segregating in different patterns because of heterozygous parents. Grattapaglia and Sederoff (1994) proposed a pseudotest backcross design to perform QTL mapping in such an F_1 cross for outcrossing species. This design capitalizes on the so-called testcross markers that are segregating in one parent but null in the second parent. Thus, two different linkage maps can be constructed for an outbred cross, each derived from a different heterozygous parent.

The hybrid poplars for QTL mapping were planted at a spacing of 4×5 m at a forest farm near Xuzhou City, Jiangsu Province, China. Total stem heights and diameters measured at the end of each of 11 growing seasons are used in this example. A subset (90) of hybrid trees randomly selected from the original population were used to construct two parent-specific genetic linkage maps with random amplified polymorphic DNAs, amplified fraction length polymorphisms and inter-simple sequence repeats (Yin *et al*., 2002). Using NLME-based functional mapping, we attempt to locate QTL affecting stem diameter growth trajectories on the linkage map derived from the *P. deltoides* parent. Individuals with missing joint genotypes for a given pair of markers were excluded from our analysis.

Model formulation

The growth of the stem diameter can be well fit by a logistic equation expressed as

$$g(t) = \frac{a}{1+b\mathrm{e}^{-rt}} \tag{10}$$

where a is the asymptotic or limiting value of g when $t \to \infty$, $a/(1+b)$ is the initial value of g when $t=0$ and r is the relative rate of growth (von Bertalanffy, 1957). Given this growth equation, we express the growth of individual i by

$$y_i(t) = \xi\left(\frac{a_1}{1+b_1\mathrm{e}^{-r_1 t}}\right) + (1-\xi)\left(\frac{a_0}{1+b_0\mathrm{e}^{-r_0 t}}\right) + \varepsilon_i(t)$$

where indicator ξ equals 1 or 0 for QTL genotype Qq and qq, respectively. Growth parameters (a, b, r) are QTL genotype-specific, subscripted by the genotype notation. For simplicity, we only model interindividual variation for parameters a and b by a simple linear regression

$$\left\{\begin{array}{l} a_{1i} = a_1 + b_{ai} \\ b_{1i} = b_1 + b_{bi} \end{array}\right. \quad \text{for QTL genotype Qq}$$

$$\left\{\begin{array}{l} a_{0i} = a_0 + b_{ai} \\ b_{0i} = b_0 + b_{bi} \end{array}\right. \quad \text{for QTL genotype qq}$$

where random effects $\mathbf{b}_i = (b_{ai}, b_{bi})$ are assumed to be genotype-invariant, normally distributed with mean vector zero and diagonal covariance matrix

$$\mathbf{D} = \begin{bmatrix} v_a^2 & 0 \\ 0 & v_b^2 \end{bmatrix}.$$

In this analysis, $\varepsilon_i(t)$ is assumed to display a normal distribution with mean vector zero and the first-order AR (AR(1)) covariance matrix specified by two parameters ρ and σ^2 (Ma *et al*., 2002). The AR(1) model assumes that the variance (σ^2) is time-invariant and correlation decays in a proportion ρ with time lag. These two assumptions can be relaxed by introducing more complicated non-stationary models (Zhao *et al*., 2005).

QTL scanning and estimation

The NLME-based mapping model is used to genome-wide scan for all possible QTL, their existence and chromosomal distribution. We detect two QTL on linkage groups 9 and 10 that affect diameter growth trajectories in poplar trees. Figure 1 illustrates a plot of the LRs between the full (there is a QTL) and reduced model (there is no QTL) across all the linkage groups. These two detected QTL are located at 111.1 cM from the first left marker on linkage group D9 and 12 cM from the first left marker on linkage group D10 because the LR peaks (34.77 and 33.33) at these positions far exceed the genome-wide critical threshold (31.64). Permutation tests were performed to determine the empirical threshold for declaring the genome-wide existence of QTL throughout all the linkage groups.

The MLEs of the curve parameters for each of two QTL genotypes, Qq and qq, and the parameters that model the

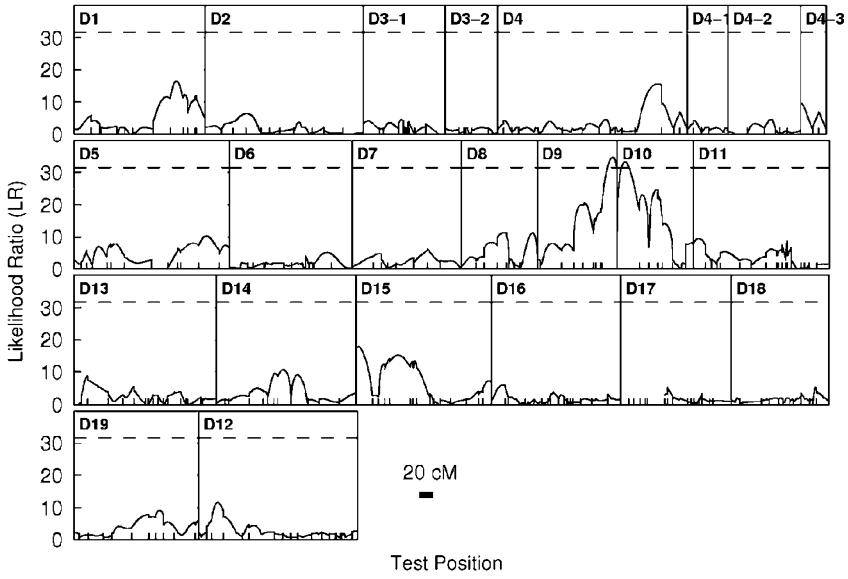

Figure 1 The profile of log-likelihood ratio (LR) between the full and reduced model estimated from the nonlinear mixed-effect (NLME) model for analyzing stem diameter growth trajectories in an interspecific poplar hybrid progeny across all the linkage groups. The dashed line is the 1% cutoff point from permutation tests. The linkage map used is one constructed with heterozygous markers from the *Populus deltoides* parent (Yin *et al*., 2002).

structure of the variance matrix are tabulated in Table 1, along with the approximate standard errors of these estimates estimated from Fisher's information matrix. All the parameters can be estimated with reasonable precision. The MLEs of the curve parameters in Table 1 were used to draw growth curves at each QTL for diameter growth (Figure 2). The pattern of the differentiation in growth curves between two QTL genotypes at each QTL suggests that these two detected QTL do not trigger an effect on growth at an early stage of tree development, but are activated at age 5–6 years and keep operational afterwards. The timing of QTL to be switched on seems to be concordant with the emerging age of intertree competition for resources availability. These results broadly support those obtained from traditional functional mapping (TFM, Ma *et al*., 2002).

Table 1 MLEs of QTL genotype-specific parameters that define stem diameter growth trajectories in poplar trees from the NLME model

Parameters	*MLE*		
	Genotype Qq		*Genotype qq*
D9			
a_j	28.28 (0.7044)		23.85 (0.6272)
b_j	12.24 (0.8472)		11.52 (0.8115)
r_j	0.56 (0.0124)		0.63 (0.0156)
σ^2		4.27 (0.4899)	
ρ		0.91 (0.0117)	
v_a^2		10.50 (1.3069)	
v_b^2		7.28 (1.0007)	
LR		34.77	
D10			
a_j	27.85 (0.6820)		23.66 (0.6923)
b_j	12.44 (0.7770)		11.85 (0.8888)
r_j	0.56 (0.0117)		0.64 (0.0168)
σ^2		4.12 (0.4322)	
ρ		0.91 (0.0110)	
v_a^2		11.07 (1.4863)	
v_b^2		7.83 (1.3992)	
LR		33.33	

Abbreviations: LR, likelihood ratio; MLE, maximum-likelihood estimate.
The standard errors of the MLEs are given in the parentheses.

Monte Carlo simulation

In order to examine the statistical properties of the NLME model for QTL mapping, two different Monte Carlo simulation strategies were performed. The simulation studies mimic the example of poplar trees with two sample sizes (80 and 200). For the first strategy, data are simulated according to the NLME model, whereas, for the second strategy, data are simulated according to TFM by Ma *et al*. (2002). In both cases, only serial correlations are modeled with the AR(1) process. The simulated data sets under different strategies are analyzed, respectively, by the NLME and TFM models. Such reciprocal designs are thought to be helpful for the methodological comparison of QTL mapping.

As expected, if the data are simulated by the NLME model, the NLME model displays better estimation accuracy and precision of parameters than does the TFM model (Table 2). The NLME model can precisely estimate the QTL location, but the TFM fails to do so. Also, compared to the TFM model, the NLME model is more advantageous for convergence under the same convergence criterion. For the data simulated under the TFM model, the two analytical models, NLME and TFM, perform similarly in the precision of parameter estimation and power (Table 3). The estimates of heritability by the two models are consistent with the true value. Tables 2 and 3 give the results for a sample size of 80. Increased sample sizes tend to blur the difference between the two models (results not shown). In general, it can be suggested that the NLME model covers the TFM model and, thus, can be used in a broader range of data types than the TFM model.

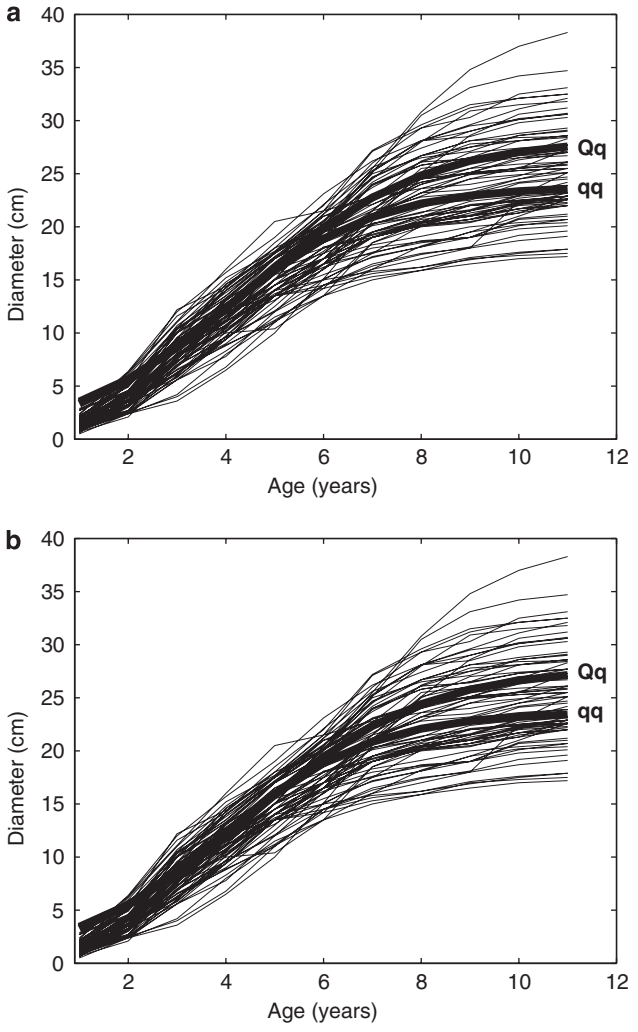

Figure 2 Two growth curves (foreground) each presenting one of the two groups of genotypes at the major quantitative trait loci (QTL) detected on linkage groups D9 and D10. The differentiation of growth curves shows the effect of the QTL on growth trajectories.

The two models are similar in computational efficiency. On a desktop (CPU 2.4 gHz and memory 512 mb), both models use about 20 min per simulation round for the data simulated with the NLME model. Yet, the TFM model uses more time when it has a convergence problem. For the data simulated with the TRM model, the NLME model still uses about 20 min, but the TRM model is faster (using about 15 min).

Discussion

In all the organisms, the development of morphological, anatomical and physiological traits takes place in characteristic ontogenetic periods. Effective modeling of the genetic control of particular physiological alterations emerging in the course of the developmental process (from their early onset until their late consequences) requires the use of adequate statistical models. Some basic statistical models for the genetic study of developmental dynamics have been proposed, in an attempt to identify the ontogenetic genetic factors or QTL that control the structure and function of a developmental system (Wu *et al*., 1999, 2003a, b, 2004a, b, c; Ma *et al*., 2002; Zhao *et al*., 2005; Wu and Lin, 2006; Yang *et al*., 2006; Yang and Xu, 2007). These so called functional mapping models have been expanded into various genetic fields related to biomedical sciences, such as cancer growth (Liu *et al*., 2005), HIV dynamics (Wang and Wu, 2004; Wang *et al*., 2006) and drug response (Lin and Wu, 2005).

The central idea of functional mapping is to model the mean vector and covariance matrix structure by parametric or nonparametric approaches. Previous functional mapping approaches have modeled the structure of the covariance matrix by considering autocorrelation components, but ignoring other sources that also affect the covariance structure, such as random effects and

Table 2 MLEs of parameters for data set simulated by NLME model with a sample size of 80 obtained from NLME-incorporated and traditional functional mapping

Parameters	*True values*	*MLE*	
		NLME	*TFM*
a_1	28.28	28.06 (0.6863)	27.94 (0.8505)
b_1	12.24	12.78 (1.2345)	12.57 (1.0337)
r_1	0.56	0.56 (0.0156)	0.57 (0.0190)
a_0	23.85	23.53 (0.6424)	24.09 (1.3549)
b_0	11.52	12.09 (1.2365)	11.51 (0.8971)
r_0	0.63	0.63 (0.0185)	0.62 (0.0199)
ρ	0.91	0.90 (0.0225)	0.96 (0.0094)
σ^2	4.27	4.13 (0.8863)	12.61 (1.7198)
v_a^2	10.25	10.03 (2.2443)	—
v_b^2	7.28	10.49 (5.7530)	—
v_c^2	0.001	0.0006 (0.0007)	—
QTL location	111.10	111.65 (4.5246)	102.41 (18.2223)
Convergence rate	—	100%	60%

Abbreviations: MLE, maximum-likelihood estimate; NLME, nonlinear mixed-effect; QTL, quantitative trait loci; TFM, traditional functional mapping.
The standard errors of the MLEs are given in the parentheses.

Table 3 MLEs of parameters for data set simulated by the TFM model with a sample size of 80 obtained from NLME-incorporated and traditional functional mapping

Parameters	*True values*	*MLE*	
		NLME	*TFM*
a_1	28.28	28.32 (0.3799)	28.32 (0.3775)
b_1	12.24	12.31 (0.9705)	12.29 (0.9670)
r_1	0.56	0.56 (0.0148)	0.56 (0.0149)
a_0	23.85	23.82 (0.3376)	23.82 (0.3392)
b_0	11.52	11.72 (0.8351)	11.70 (0.8579)
r_0	0.63	0.63 (0.0169)	0.63 (0.0174)
ρ	0.91	0.90 (0.0122)	0.91 (0.0108)
σ^2	4.27	3.99 (0.4732)	4.17 (0.4462)
v_a^2	0.0	0.23 (0.3609)	—
v_b^2	0.0	0.58 (1.1707)	—
v_r^2	0.0	0.0001 (0.0002)	—
QTL location	111.10	110.67 (3.5292)	110.57 (3.3062)
Convergence rate	—	100%	100%

Abbreviations: MLE, maximum-likelihood estimate; NLME, nonlinear mixed-effect; QTL, quantitative trait loci; TFM, traditional functional mapping.
The standard errors of the MLEs are given in the parentheses.

measurement errors (Diggle *et al*., 2002). The study presented in this article is aimed to generalize functional mapping to model the effects of random effects on the parameter estimation of functional mapping and its relevant hypothesis tests, thus broadening the visibility of functional mapping. The incorporation of random effects with functional mapping based on NLME models (Beal and Sheiner, 1982; Lindstrom and Bates, 1990; Davidian and Giltinan, 1995, 2003; Vonesh *et al*., 2002; Wu, 2002, 2004a, b) is robust; in that it can provide sufficient power to detect ontogenetic QTL for longitudinal data measured at uneven spaces and irregularly for different subjects.

The NLME-incorporated functional mapping model has been used to analyze a published growth data set in poplar trees. As compared to previous simpler functional mapping (TFM) (Ma *et al*., 2002), the new model generates agreeable results for the detection of QTL, their chromosomal locations and ontogenetic effects during a time course. However, simulation studies based on reciprocal designs, that is, the data simulated and, then, analyzed by NLME and TFM models, respectively, suggest that whereas QTL contained in the TFM-simulated data can be detected by both models, QTL in the NLME-simulated data can only well be detected by the NLME model. All this implies that the NLME model is more general and can be used more widely in practice than the TFM model.

Perhaps, the most significant advantage of NLME-based functional mapping is its flexibility to extend the idea of functional mapping to a broad spectrum of biological and biomedical areas (see also Malosetti *et al*., 2006). NLME models include two-stage hierarchical characterization of intra- and intersubject variation. In the first stage, any form of parametric models can be incorporated that are defined by biologically meaningful mathematical parameters; for example, growth rate parameter in the growth equation (West *et al*., 2001) is related to the developmental status of an organism in a time period. These mathematical parameters may be correlated with other physiological variables or expressed differently under different environmental conditions or genetic backgrounds. The genetic control of these biological phenomena can be integrated into the second stage of the NLME model at which specific underlying QTL can be modeled, estimated and tested.

Statistics inference of longitudinal measurements based on the NLME model has received considerable attention in recent years because of its flexibility to incorporate the correlation within repeated measurements, between-individual variation and covariates (Vonesh *et al*., 2002; Wu, 2002, 2004a, b; Davidian and Giltinan, 2003). The NLME model has been recently extended to take into account censoring and covariate measured with errors (Wu, 2002), missing covariates (Wu, 2004a) and nonignorable dropouts (Wu, 2004b). In addition, to clearly describe the NLME model, we constructed our model framework in the context of interval mapping. More recently, Xu and group have developed a series of shrinkage models that allow a genome-wide search for all possible QTL (Xu, 2003, 2007; Wang *et al*., 2005). These multiple QTL models taking into account epistatic interactions between different QTL can be incorporated into the NLME model. All these statistical and genetic extensions can be incorporated into functional mapping, which will provide a powerful means for characterizing the developmental machinery of the genetic control of complex traits at the interplay between trait formation and progression and the environment in which the organism is grown. The computer code for the statistical method proposed in this article can be available from the corresponding author.

Acknowledgements

We thank Associate Editor, Dr Shizhong Xu, and the three anonymous referees for their constructive comments on the manuscript. The preparation of this manuscript is partially supported by grants from NSF (0540745) and the National Natural Science Foundation of China (09-95671 and 30230300).

References

Beal SL, Sheiner LB (1982). Estimating population kinetics. *Crit Rev Biomed Eng* **8**: 195–222.

Chen Z (2005). The full EM algorithm for the MLEs of QTL effects and positions and their estimated variances in multiple interval mapping. *Biometrics* **61**: 474–480.

Chi EM, Reinsel GC (1989). Models for longitudinal data with random effects and AR (1) errors. *J Am Stat Assoc* **84**: 452–459.

Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.

Davidian M, Giltinan D (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall: New York.

Davidian M, Giltinan DM (2003). Nonlinear models for repeated measurements: an overview and update. *J Agric Biol Environ Stat* **8**: 387–419.

Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002). *Analysis of Longitudinal Data*. Oxford University Press: Oxford, UK.

Grattapaglia D, Sederoff RR (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**: 1121–1137.

Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV infection. *Nature* **373**: 123–126.

Lagarius JC, Reeds JA, Wright MH, Wright PE (1998). Convergence properties of the Neler–Mead simplex method in low dimensions. *SIAM J Optim* **9**: 112–147.

Laird NM, Ware JH (1982). Random effects models for longitudinal data. *Biometrics* **38**: 963–974.

Lin M, Wu RL (2005). Theoretical basis for the identification of allelic variants that encode drug efficacy and toxicity. *Genetics* **170**: 919–928.

Lindstrom MJ, Bates DM (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**: 673–687.

Liu T, Zhao W, Tian LL, Wu RL (2005). An algorithm for molecular dissection of tumor progression. *J Math Biol* **50**: 336–354.

Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer: Sunderland, MA, USA.

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

Malosetti M, Visser RGF, Celis-Gamboa C, van Eeuwijk FA (2006). QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. *Theor Appl Genet* **113**: 288–300.

Rodriguez-Zas SL, Southney BR, Heyen DW, Lewin HA (2002). Detection of quantitative trait loci influencing dairy traits using a model for longitudinal data. *J Dairy Sci* **85**: 2681–2691.

Schabenberger O (1995). The use of ordinal response methodology in forestry. *Forest Sci* **41**: 321–336.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB *et al.* (1998). Comprehensive identification of cell-cycle regulated genes in *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **95**: 14863–14868.

Tierney L, Kadane JB (1986). Accurate approximations for posterior moments and marginal densities. *J Am Stat Assoc* **81**: 82–86.

von Bertalanffy L (1957). Quantitative laws in metabolism and growth. *Q Rev Biol* **32**: 217–231.

Vonesh EF, Wang H, Nie L, Majumdar D (2002). Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed-effects models. *J Am Stat Assoc* **97**: 271–283.

Wang ZH, Hou W, Wu RL (2006). A statistical model to analyze quantitative trait locus interactions for HIV dynamics from the virus and human genomes. *Stat Med* **25**: 495–511.

Wang ZH, Wu RL (2004). A statistical model for high-resolution mapping of quantitative trait loci determining human HIV-1 dynamics. *Stat Med* **23**: 3033–3051.

Wang H, Zhang Y-M, Li X, Masinde GL, Mohan S, Baylink DJ *et al.* (2005). Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics* **170**: 465–480.

West GB, Brown JH, Enquist BJ (1997). A general model for the origin of allometric scaling laws in biology. *Science* **276**: 122–126.

West GB, Brown JH, Enquist BJ (2001). A general model for ontogenetic growth. *Nature* **413**: 628–631.

Wolfinger RD (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**: 791–795.

Wu L (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *J Am Stat Assoc* **97**: 955–964.

Wu L (2004a). Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. *J Am Stat Assoc* **32**: 700–709.

Wu L (2004b). Nonlinear mixed-effects models with nonignorably missing covariates. *Can J Stat* **32**: 27–37.

Wu RL, Lin M (2006). Functional mapping? How to map and study the genetic architecture of dynamic complex traits. *Nat Rev Genet* **7**: 229–237.

Wu RL, Ma C-X, Lin M, Casella G (2004a). A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics* **166**: 1541–1551.

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

Wu RL, Ma CX, Lou XY, Casella G (2003a). Molecular dissection of allometry, ontogeny, and plasticity: a genomic view of developmental biology. *Bioscience* **53**: 1041–1047.

Wu RL, Ma C-X, Zhao W, Casella G (2003b). Functional mapping of quantitative trait loci underlying growth rates: a parametric model. *Physiol Genomics* **14**: 241–249.

Wu RL, Wang ZH, Zhao W, Cheverud JM (2004c). A mechanistic model for genetic machinery of ontogenetic growth. *Genetics* **168**: 2383–2394.

Wu W-R, Li W-M, Tang D-Z, Lu H-R, Worland AJ (1999). Time-related mapping of quantitative trait loci underlying tiller number in rice. *Genetics* **151**: 297–303.

Xu S (2003). Estimating polygenic effects using markers of the entire genome. *Genetics* **163**: 789–801.

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

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

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

Yin TM, Zhang XY, Huang MR, Wang MX, Zhuge Q, Zhu LH *et al.* (2002). Molecular linkage maps of the *Populus* genome. *Genome* **45**: 541–555.

Zhao W, Chen YQ, Casella G, Cheverud JM, Wu RL (2005). A nonstationary model for functional mapping of complex traits. *Bioinformatics* **21**: 2469–2477.