

Functional Mapping of Quantitative Trait Loci Underlying the Character Process: A Theoretical Framework

Chang-Xing Ma,^{*,†} George Casella* and Rongling Wu^{*,1}

^{*}Department of Statistics, University of Florida, Gainesville, Florida 32611 and [†]Department of Statistics, Nankai University, Tianjin 300071, China

Manuscript received September 10, 2001

Accepted for publication May 6, 2002

ABSTRACT

Unlike a character measured at a finite set of landmark points, function-valued traits are those that change as a function of some independent and continuous variable. These traits, also called infinite-dimensional characters, can be described as the character process and include a number of biologically, economically, or biomedically important features, such as growth trajectories, allometric scalings, and norms of reaction. Here we present a new statistical infrastructure for mapping quantitative trait loci (QTL) underlying the character process. This strategy, termed *functional mapping*, integrates mathematical relationships of different traits or variables within the genetic mapping framework. Logistic mapping proposed in this article can be viewed as an example of functional mapping. Logistic mapping is based on a universal biological law that for each and every living organism growth over time follows an exponential growth curve (*e.g.*, logistic or S-shaped). A maximum-likelihood approach based on a logistic-mixture model, implemented with the EM algorithm, is developed to provide the estimates of QTL positions, QTL effects, and other model parameters responsible for growth trajectories. Logistic mapping displays a tremendous potential to increase the power of QTL detection, the precision of parameter estimation, and the resolution of QTL localization due to the small number of parameters to be estimated, the pleiotropic effect of a QTL on growth, and/or residual correlations of growth at different ages. More importantly, logistic mapping allows for testing numerous biologically important hypotheses concerning the genetic basis of quantitative variation, thus gaining an insight into the critical role of development in shaping plant and animal evolution and domestication. The power of logistic mapping is demonstrated by an example of a forest tree, in which one QTL affecting stem growth processes is detected on a linkage group using our method, whereas it cannot be detected using current methods. The advantages of functional mapping are also discussed.

THE theoretical principle for analyzing quantitative trait loci (QTL) dates back to SAX (1923), who first associated pattern and pigment markers with seed size in beans. However, statistical methodologies for mapping QTL on a high-density linkage map of molecular markers had not been well established until LANDER and BOTSTEIN's (1989) pioneering work. These authors employed an expectation-maximization (EM)-implemented maximum-likelihood approach, proposed by DEMPSTER *et al.* (1977), to map QTL on a particular chromosomal interval bracketed by two flanking markers. This so-called interval mapping method was later improved by including markers from other intervals as covariates to control the overall genetic background (JANSEN and STAM 1994; ZENG 1994). The improved method, called composite interval mapping by ZENG (1994), displays increased power in QTL detection because of reduced residual variance. KAO *et al.* (1999) proposed using multiple marker intervals

simultaneously to map multiple QTL of epistatic interactions throughout a linkage map. Currently, an upsurge of QTL mapping methodologies has been developed to consider various situations regarding different marker types (dominant or codominant), different marker spaces (sparse or dense), different experimental designs (F_2 /backcross or full-sib family), or different mapping populations (autogamous or allogamous). The statistical methods used for QTL mapping in the literature include regression analyses (HALEY and KNOTT 1992; XU 1995), maximum likelihood (LANDER and BOTSTEIN 1989; ZENG 1994; KAO *et al.* 1999), and the Bayesian approach (SATAGOPAN *et al.* 1996; SILLANPAA and ARJAS 1999; XU and YI 2000). Many of these mapping methods have been instrumental in the identification of QTL responsible for variation in various complex traits important to agriculture, forestry, biomedicine, or biological research (TANKSLEY 1993; WU *et al.* 2000; MACKAY 2001; MAURICIO 2001).

It should be noted, however, that many quantitative traits, such as body size and body shape, are inherently too complex to be described by a single value, because their phenotypes change with age, metabolic rate, or en-

¹Corresponding author: Department of Statistics, 533 McCarty Hall C, University of Florida, Gainesville, FL 32611.
E-mail: rwu@stat.ufl.edu

environmental stimulus. These traits, which can be expressed as a function (or stochastic process) of some independent and continuous variable, were thought of as *infinite-dimensional characters* by KIRKPATRICK and HECKMAN (1989) or *function-valued traits* by PLETCHER and GEYER (1999). The genetic determination of the character process has long intrigued students in different disciplines of biology, genetics, and breeding (*e.g.*, CHEVERUD *et al.* 1983; ATCHLEY 1984; WU *et al.* 1992; ATCHLEY and ZHU 1997; RICE 1997). A simple approach for mapping infinite-dimensional characters is to associate markers with phenotypes separately for different ages, traits, or environments and compare the differences of QTL expression across ages, traits, or environments (CHEVERUD *et al.* 1996; NUZHIDIN *et al.* 1997; VERHAEGEN *et al.* 1997; EMEBIRI *et al.* 1998; WU *et al.* 1999). However, these separate analyses cannot provide effective estimates of genetic control over infinite-dimensional characters, because they fail to capture the information about the covariances of different traits or the same trait measured at different ages or environments. Although multitrait mapping approaches can take into account simultaneously different traits or the same trait measured at different ages or environments (JIANG and ZENG 1995; KOROL *et al.* 1995; RONIN *et al.* 1995; EAVES *et al.* 1996; KNOTT and HALEY 2000), their applications are actually limited to bivariate, or at most trivariate, analyses. As the number of traits increases these multitrait analysis approaches will have a reduced ability to produce precise estimates of genetic parameters in quantitative genetic studies (SHAW 1987).

To circumvent the difficulty in manipulating a large number of correlated traits, new attempts were made by MANGIN *et al.* (1998) and KOROL *et al.* (2001), who transformed the initial trait space into a space of a lower dimension on the basis of principal component analysis or interval-specific calculation of eigenvalues and eigenvectors of the residual covariance matrix. These new attempts have, to some extent, made the genetic mapping of a large number of traits more tractable, but they still treat infinite-dimensional characters as discrete traits or eigenvalues and do not place the physiological mechanisms predisposing for the phenotypic variation of infinite-dimensional characters in a mapping framework. In real life, an infinite-dimensional character often changes its phenotype through particular physiological regulations or developmental signals in the same way that an organism tends to maximize its metabolic capacity and internal efficiency as a consequence of natural selection. Therefore, the incorporation of the underpinning physiological or developmental mechanisms of trait variation into a QTL-mapping strategy may likely produce more accurate results in terms of biological reality.

The objective of this study is to propose a general theoretical framework for embedding biological mechanisms and processes in the statistical analysis of QTL

mapping. A maximum-likelihood-based method, implemented with the EM algorithm, is used to estimate QTL locations and effects on various biological processes. The newly developed method is applied in an example to map the growth of a forest tree. Compared with current mapping methods, our method incorporating growth trajectories tends to be more powerful and more precise in QTL detection and also has greater potential to increase mapping power, precision, and resolution by reducing residual variance and the number of unknown parameters to be estimated. In practice, our method is economically more feasible than previous methods because it needs a smaller sample size to obtain adequate power for QTL detection as a result of the use of multiple measurements for each individual. It can be anticipated that the method proposed in this article will have great implications for the design of an efficient early selection program and the interface of genetics, development, and evolution.

MODELING THE CHARACTER PROCESS

Many biological processes in real life are expected to arise as curves, such as growth curves, allometric scalings, hormone profiles, and norms of reaction. A growth curve or *trajectory* represents an individual as a function that relates the age of an individual to some measure of its size. Since there are an infinite number of ages, growth trajectories can be thought of as function-valued traits (PLETCHER and GEYER 1999; JAFFREZIC and PLETCHER 2000) or infinite-dimensional characters (KIRKPATRICK and HECKMAN 1989). Other examples of function-valued traits include the continuous change of a morphological or physiological variable with body size (*allometric scaling*; NIKLAS 1994; WEST *et al.* 1997, 1999) and responsive phenotypes of a given genotype to a changing environment (*reaction norm*, VIA *et al.* 1995). The common property of these function-valued traits is that they can be described as a function (or stochastic process) of some independent and continuous variable consisting of an infinite number of points, such as age, temperature, light intensity, or biological size.

The model for QTL mapping we developed relies on concepts from functional analysis and stochastic processes. Throughout, we use growth trajectories as a concrete example to illustrate the ideas, but allometric scalings, hormone profiles, and reaction norms can be treated in the same framework with appropriate modifications.

Growth trajectory: It is well known that there are biological laws underlying growth trajectories (GOULD 1977; ALBERCH *et al.* 1979). A growth law can be visualized as the “force field” propelling a point through a phenotype space, tracing out the ontogenetic path. If the size of an organism is denoted by y , its *ontogenetic trajectory* $y(t)$ can be generated through the differential dy/dt , which models the growth rate. Many differential

functions have been established to describe the growth trajectory. Basically, they are sorted into three categories: (1) exponential, (2) saturating, and (3) sigmoidal (VON BERTALANFFY 1957; NIKLAS 1994). Each of these growth models has a common feature that the development of ontogenetic trajectory is regulated by a set of “control parameters” such as onset age of growth, offset signal for growth, growth rate during the period of growth, and initial size at the commencement of the growth period. Also, each of these growth models exhibits an initial phase of exponential growth due simply to the geometrically multiplying population of newly differentiated cells. This initial growth phase has the property that small perturbations in growth rate or onset age are amplified enormously during ontogeny. Thus, it is easy to find examples of how a small “mutation” in a growth parameter causes a series of developmental alterations that produce a phenotype qualitatively different from the normal one.

In this article, we further limit our analysis to sigmoidal, or logistic, function (PEARL 1925). The logistic curve is regarded as among the most important ones to capture the age-specific change in growth (NIKLAS 1994; WEST *et al.* 2001). The logistic growth curve as a biological law can be mathematically described by

$$g(t) = \frac{a}{1 + be^{-rt}}, \quad (1)$$

where a is the asymptotic or limit value of g when $t \rightarrow \infty$, $a/(1 + b)$ is the initial value of g when $t = 0$, and r is the relative rate of growth (VON BERTALANFFY 1957). The logistic growth curve consists of two phases, an exponential and an asymptotic. The overall form of the curve is determined by different combinations of parameters a , b , and r . If different genotypes at a putative QTL have different combinations of these parameters, this implies that this QTL plays a role in governing the difference of growth trajectories.

The logistic growth curve described in Equation 2 can be used to determine the coordinates of a biologically important point in the entire growth trajectory—the *inflection point*—where the exponential phase ends and the asymptotic phase begins (NIKLAS 1994). The time at the inflection point corresponds to the time point at which a maximum growth rate occurs. The time (t_1) and growth [$g(t_1)$] at the inflection point for a QTL genotype can be derived as

$$t_1 = \frac{\log b}{r}$$

$$g(t_1) = \frac{a}{2}. \quad (2)$$

The difference in the coordinates between different genotypes provides important information about the genetics and evolution of growth trajectories (NIKLAS 1994). Moreover, the time at the inflection point, to-

gether with the initial growth and asymptotic growth, determines exclusively the difference of two growth curves. Any two curves will not be distinguishable if they have the same values for these three variables.

STATISTICAL MODELS

Genetic design: The purpose of this article is to introduce a novel idea to QTL mapping. Hence, we suppose a simplest backcross design derived from two contrasted homozygous inbred lines. Other more complex designs, such as an F_2 or full-sib family, can also be used. In a backcross population, there are two groups of genotypes at a locus, in which a marker-based genetic linkage map is constructed, aimed at the identification of QTL affecting an age-dependent trait, such as body size or body weight. In practice, the data are observed only at a finite set of times, $1, \dots, m$, rather than a continuum, so we have only a finite set of data on each individual i , which can be considered as a multivariate trait vector, $y_i(1), \dots, y_i(m)$. This finite set of data can be modeled by a growth curve. Assume that a pleiotropic QTL of allele Q_1 and Q_2 affecting growth curves or trajectories is segregating in the backcross population. This QTL is bracketed by two flanking markers η and $\eta + 1$, each with two genotypes $M_\eta m_\eta, m_\eta m_\eta$, and $M_{\eta+1} m_{\eta+1}, m_{\eta+1} m_{\eta+1}$, respectively. For a particular genotype j ($j = 1$ for $Q_1 Q_2$ or 2 for $Q_2 Q_2$) of this QTL, the parameters describing its logistic curve are denoted by a_j, b_j , and r_j . The comparisons of these parameters between two different genotypes can determine whether and how this putative QTL affects growth trajectories.

Suppose that there are a total of N progeny in the backcross measured at each of m times. The trait phenotype of progeny i measured at time t due to the QTL located on an interval flanked by markers η and $\eta + 1$ can be expressed by a linear statistical model (KIRKPATRICK and HECKMAN 1989; LANDER and BOTSTEIN 1989; PLETCHER and GEYER 1999),

$$y_i(t) = \xi_{i1} g_1(t) + (1 - \xi_{i1}) g_2(t) + e_i(t), \quad (3)$$

where ξ_{ij} is an indicator variable for the possible genotypes of the QTL for progeny i and defined as 1 if a particular QTL genotype is indicated and 0 otherwise; $g_j(t)$ is the genotypic value of the QTL for the trait at time t ; and $e_i(t)$ is the residual effect of progeny i , including the aggregate effect of polygenes and error effect, and distributed as $N(0, \sigma_e^2(t))$. The probability with which ξ_{ij} takes 1 or 0 depends on the two-locus genotype of the flanking markers η and $\eta + 1$ and the position of the QTL on the marker interval. The probability of a QTL genotype ($Q_1 Q_2$ or $Q_2 Q_2$) conditional upon the four genotypes of the flanking markers ($M_\eta m_\eta M_{\eta+1} m_{\eta+1}, M_\eta m_\eta m_{\eta+1} m_{\eta+1}, m_\eta m_\eta M_{\eta+1} m_{\eta+1}$, and $m_\eta m_\eta m_{\eta+1} m_{\eta+1}$) for progeny i in the backcross population was expressed as

$$p_{ij} = \text{Prob}(\xi_{ij} = j | \eta, \eta + 1, \theta), \quad j = 1 \text{ or } 2,$$

where θ is the ratio of the recombination fractions between marker η and the QTL to the recombination fraction between the two markers.

Statistical methods: The phenotypes of the trait at all time points $1, \dots, m$ for each QTL genotype group follow a multivariate normal density,

$$f_j(\mathbf{y}) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp[-(\mathbf{y} - \mathbf{g}_j)^T \Sigma^{-1} (\mathbf{y} - \mathbf{g}_j) / 2],$$

where \mathbf{g}_j is the vector of the expected genotypic values of the trait for QTL genotype j measured for t times and Σ is the residual variance-covariance matrix of the phenotypes measured at different ages. Indeed, \mathbf{g}_j can be modeled by the logistic curve of Equation 2 as

$$\mathbf{g}_j = [g_j(t)]_{1 \times m} = \left[\frac{a_j}{1 + b_j e^{-r_j t}} \right]_{1 \times m}, \quad (4)$$

and Σ can be assumed identical among different genotypes and modeled using AR(1) repeated measurement errors (DAVIDIAN and GILTINAN 1995; VERBEKE and MOLENBERGHS 2000) as

$$\Sigma = \sigma_c^2 \begin{bmatrix} 1 & \rho & \dots & \rho^{m-1} \\ \rho & 1 & \dots & \rho^{m-2} \\ \dots & \dots & \dots & \dots \\ \rho^{m-1} & \rho^{m-2} & \dots & 1 \end{bmatrix}. \quad (5)$$

For simplicity, the matrix Σ of Equation 5 assumes variance stationarity, *i.e.*, there is the same residual variance (σ^2) for the trait at each time, and covariance stationarity; *i.e.*, the covariance between different measurements decreases proportionally (in ρ) with increased time interval (see also PLETCHER and GEYER 1999). These two assumptions, although providing a reasonable approximation in some situations, can be readily relaxed. In the DISCUSSION, we propose a few different approaches to the relaxation of these two assumptions.

The likelihood of the backcross progeny with m -dimensional measurements can be represented by a multivariate mixture model

$$L(\Omega) = \prod_{i=1}^N \left[\sum_{j=1}^2 p_{ij} f_j(\mathbf{y}_i) \right], \quad (6)$$

where the vector $\Omega = (a_j, b_j, r_j, \theta, \rho, \sigma^2)^T$ contains unknown parameters to be estimated for the QTL effect, QTL position, and residual (co)variances. The maximum-likelihood estimates (MLEs) of the unknown parameters for a pleiotropic QTL can be computed by implementing the EM algorithm (DEMPSTER *et al.* 1977; LANDER and BOTSTEIN 1989; ZENG 1994). The log-likelihood is given by

$$\log L(\Omega) = \sum_{i=1}^N \log \left[\sum_{j=1}^2 p_{ij} f_j(\mathbf{y}_i) \right], \quad (7)$$

with derivatives

$$\begin{aligned} \frac{\partial}{\partial \Omega_\lambda} \log L(\Omega) &= \sum_{i=1}^N \sum_{j=1}^2 \frac{p_{ij} \frac{\partial}{\partial \Omega_\lambda} f_j(\mathbf{y}_i)}{\sum_{j=1}^2 p_{ij} f_j(\mathbf{y}_i)} \\ &= \sum_{i=1}^N \sum_{j=1}^2 \frac{p_{ij} f_j(\mathbf{y}_i)}{\sum_{j=1}^2 p_{ij} f_j(\mathbf{y}_i)} \frac{\partial}{\partial \Omega_\lambda} \log f_j(\mathbf{y}_i) \\ &= \sum_{i=1}^N \sum_{j=1}^2 P_{ij} \frac{\partial}{\partial \Omega_\lambda} \log f_j(\mathbf{y}_i), \end{aligned}$$

where we define

$$P_{ij} = \frac{p_{ij} f_j(\mathbf{y}_i)}{\sum_{j=1}^2 p_{ij} f_j(\mathbf{y}_i)}, \quad (8)$$

which could be thought of as a posterior probability that progeny i have QTL genotype j . We then implement the EM algorithm with the expanded parameter set $\{\Omega, \mathbf{P}\}$, where $\mathbf{P} = \{P_{ij}, j = 1, \dots, k; i = 1, \dots, N\}$. Conditional on \mathbf{P} , we solve for

$$\frac{\partial}{\partial \Omega_\lambda} \log L(\Omega) = 0 \quad (9)$$

to get our estimates of Ω (the M step; Equation 9). The estimates are then used to update \mathbf{P} (the E step; Equation 8), and the process is repeated until convergence. The values at convergence are the MLEs of Ω . The iterative expressions of estimating Ω from the previous step are given in APPENDIX A. The standard errors of the MLEs are estimated using the inverse of the Fisher information matrix.

In practical computations, the QTL position parameter θ can be viewed as a fixed parameter because a putative QTL can be searched at every 1 or 2 cM on a map interval bracketed by two markers throughout the entire linkage map. The amount of support for a QTL at a particular map position is often displayed graphically through the use of likelihood maps or profiles, which plot the likelihood-ratio test statistic as a function of map position of the putative QTL.

HYPOTHESIS TESTS

After the MLEs of the parameters of interest are obtained, a number of biologically meaningful hypotheses can be tested on the basis of the logistic-based genetic model. First, the hypothesis about the existence of a QTL affecting an overall growth curve can be formulated as

$$H_0: a_1 = a_2, b_1 = b_2, r_1 = r_2$$

$$H_1: \text{at least one of the equalities above does not hold,} \quad (10)$$

where H_0 corresponds to the reduced model, in which the data can be fit by a single logistic curve, and H_1 corresponds to the full model, in which there exist two different logistic curves to fit the data.

Second, the hypothesis test can be performed on the

time at which the detected QTL starts to exert or ceases an effect on growth trajectories, by comparing the difference of the expected means between different genotypes at various time points. At a given time t^* , the hypothesis is

$$\begin{aligned} H_0: & g_1(t^*) = g_2(t^*) \\ H_1: & g_1(t^*) \neq g_2(t^*). \end{aligned} \tag{11}$$

If H_0 is rejected, this means that the QTL has a significant effect on variation in growth at time t^* . Testing the hypotheses (11) is equivalent to testing the difference of the model with no restriction and the model with the restriction:

$$\frac{a_1}{1 + b_1 e^{-r_1 t^*}} = \frac{a_2}{1 + b_2 e^{-r_2 t^*}}.$$

Because t^* is given, one of the six logistic parameters can be expressed as a function of the other five and, thus, there is one fewer parameter to be estimated for the model with the above restriction (the reduced model) than the model with no restriction (the full model). By scanning time points from 1 to m , one can find the time point at which the QTL starts or ceases to exert an effect on growth.

Third, the genotypic differences in time (t_1) and growth [$g(t_1)$] at the inflection point of maximum growth rate (Equation 2) can be tested. The test for the genotypic difference is based on the restriction

$$\frac{\log b_1}{r_1} = \frac{\log b_2}{r_2} \tag{12}$$

for t_1 , and

$$\frac{a_1}{2} = \frac{a_2}{2} \tag{13}$$

for $g(t_1)$.

Fourth, when there is no double ‘‘crossover’’ between the growth curves of the two QTL genotypes, the effect of QTL \times age interaction on the overall growth curve can be tested by comparing the genotypic differences at time $t = 0$ and $t = \infty$, which is expressed by the restriction

$$\frac{a_1}{1 + b_1} - \frac{a_2}{1 + b_2} = a_1 - a_2. \tag{14}$$

Similarly, the effect of QTL \times age interaction on the growth at any two different time points t_1 and t_2 can be tested with the restriction

$$\frac{a_1}{1 + b_1 e^{-r_1 t_1}} - \frac{a_2}{1 + b_2 e^{-r_2 t_1}} = \frac{a_1}{1 + b_1 e^{-r_1 t_2}} - \frac{a_2}{1 + b_2 e^{-r_2 t_2}}. \tag{15}$$

Testing QTL \times age interactions on the basis of Equations 14 and 15 can be helpful to our understanding of

the way in which QTL trigger an effect on growth and development.

The test statistics for testing the hypotheses (10–15) are calculated as the log-likelihood ratio (LR) of the full over reduced model:

$$LR = -2 \log \frac{L(\hat{\Omega})}{L(\tilde{\Omega})},$$

where $\tilde{\Omega}$ and $\hat{\Omega}$ denote the ML estimates of the unknown parameters under H_0 and H_1 , respectively. But the determination of the distribution of the LR is a difficult statistical issue. For a two-normal mixture model, like ours in this study, LOISEL *et al.* (1994) proved that the limiting distribution of LR under H_0 for (10) is a 50:50 mixture of χ^2_1 and χ^2_2 if σ is unknown. The test statistics for the other hypotheses (12–15) can be viewed as being asymptotically χ^2 distributed with 1 d.f.

EXAMPLE

The Populus map: We use an example of a forest tree to demonstrate the power of our statistical model for mapping QTL affecting growth trajectories. The study material used was derived from the triple hybridization of *Populus* (poplar). A *Populus deltoides* clone (designated I-69) was used as a female parent to mate with an interspecific *P. deltoides* \times *P. nigra* clone (designated I-45) as a male parent (WU *et al.* 1992). The hybrids between *P. deltoides* and *P. nigra* are called Euramerica poplar (*P. euramericana*). Both *P. deltoides* I-69 and *P. euramericana* I-45 were selected at the Research Institute for Poplars in Italy in the 1950s and were introduced to China in 1972. In the spring of 1988, a total of 450 1-year-old rooted three-way hybrid seedlings were planted at a spacing of 4 \times 5 m at a forest farm near Xuchou City, Jiangsu Province, China. The total stem heights and diameters measured at the end of each of 11 growing seasons are used in this example.

A genetic linkage map has been constructed using 90 genotypes randomly selected from the 450 hybrids with random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), and intersimple sequence repeats (ISSRs; YIN *et al.* 2002). This map comprises the 19 largest linkage groups for each parental map, which represent roughly 19 pairs of chromosomes. We chose linkage group 10 from the *P. deltoides* parent map to detect QTL affecting diameter growth using our newly developed method.

Logistic curves: By plotting total growth against year, it is observed that each of the 90 mapped genotypes follows the S-shaped (logistic) growth curve. Figure 1 illustrates S-shaped growth curves for individual stem diameters over 11 years. A least-squares approach was used to fit diameter growth with the logistic curve (Equation 1) for each genotype. On the basis of statistical tests, all genotypes can be well fit by a logistic curve

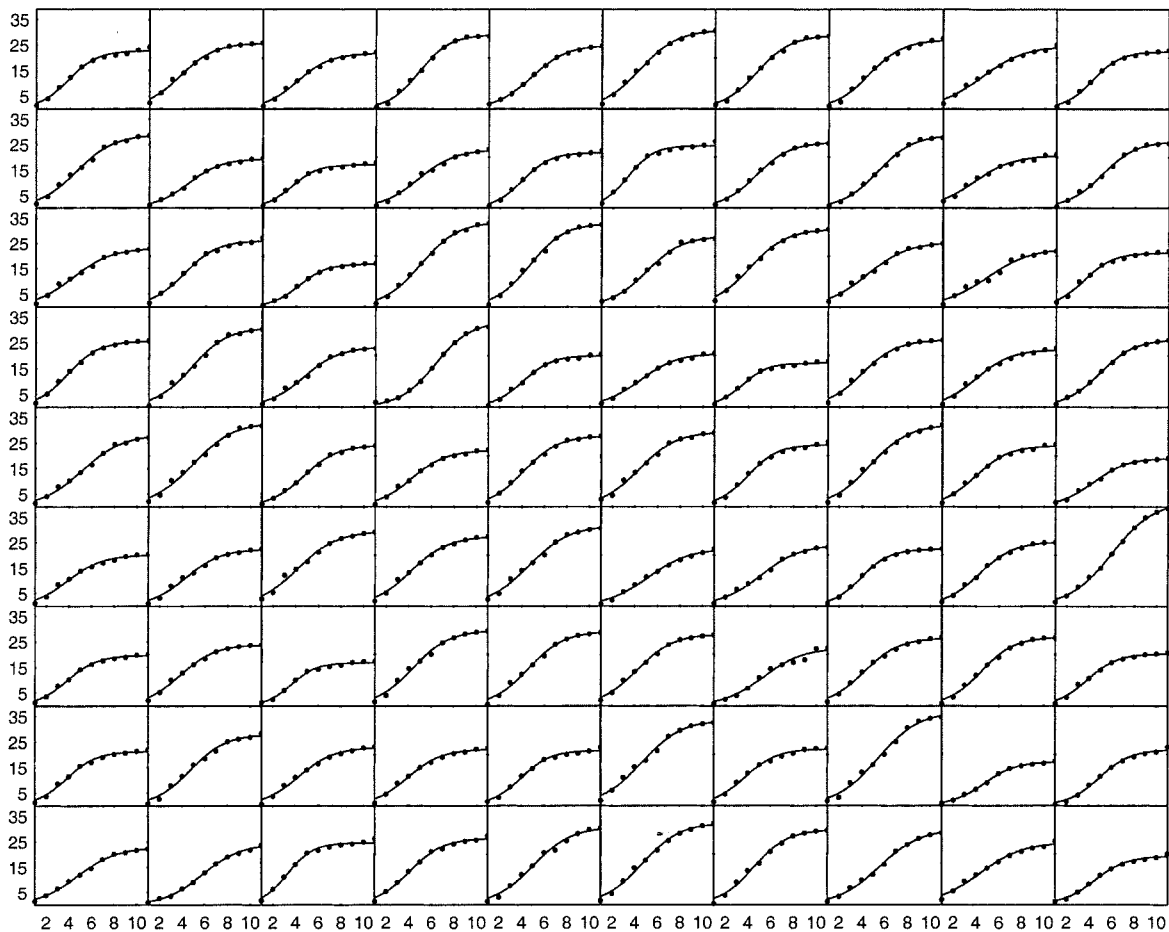


FIGURE 1.—Plots of stem diameter growth *vs.* ages for each of the 90 genotypes used to construct linkage maps in poplar hybrids (YIN *et al.* 2002). The growth of these genotypes can be well fit by a particular logistic curve. The *x*- and *y*-axes of the plots denote age (in years) and stem diameter (in centimeters).

($r^2 > 0.95$). Also, different curve shapes of these genotypes imply possible genetic control over growth trajectories. The statistical model built upon the logistic growth curve model is used to map QTL responsible for growth trajectories in diameters.

QTL detection: Using our logistic mapping model, one QTL is detected on linkage group 10 for the growth trajectory of stem diameter in the interspecific hybrids of poplar (Figure 2). The critical value for claiming the existence of QTL can be determined on the basis of the Bonferroni argument for the sparse-map case (LANDER and BOTSTEIN 1989) or by permutation tests proposed by DOERGE and CHURCHILL (1996). In this example, the chromosome-wide empirical estimate of the critical value is obtained from 1000 permutation tests. It is found that the critical values for declaring the existence of a QTL on the linkage group under consideration are 34.69 and 45.56 at the significance levels $P = 0.05$ and 0.01 , respectively. The profile of the log-likelihood ratios of the full *vs.* reduced model across the length of linkage group 10 has a clear peak at ~ 13 cM from marker CA/CCC-640R. The LR value at this peak is 51.0, well beyond

the empirical critical threshold at the significance level $P = 0.01$.

To compare the power of our method with previous methods, the same material is subjected to interval mapping (LANDER and BOTSTEIN 1989) and composite interval mapping (ZENG 1994) on the basis of the most differentiated phenotypes measured at year 11 (Figure 1). Neither of these two mapping methods can declare the existence of a significant QTL given their lower LR values. Figure 2 illustrates the result from interval mapping for diameters. No LR value from interval mapping is larger than the threshold (7.68) obtained from permutation tests at the significance level $P = 0.05$.

Similar conclusions about the difference of QTL detection between our method and current methods are obtained for many other linkage groups (results not shown). These suggest that our method incorporating logistic growth curves has greater power to detect a significant QTL than the current methods.

The dynamic pattern of QTL expression: Our method has an additional advantage; *i.e.*, it can detect the dynamic change of QTL expression over time. The growth

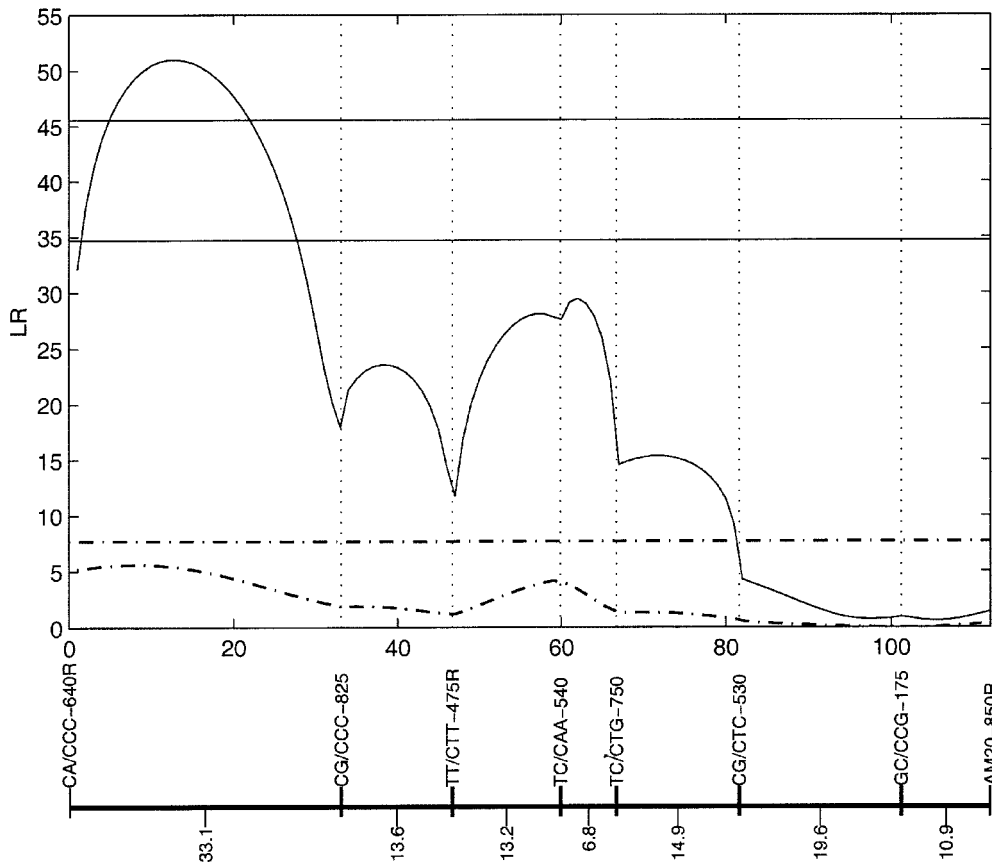


FIGURE 2.—The profile of the log-likelihood ratios between the full and reduced (no QTL) model for diameter growth trajectories across linkage group 10 in the *Populus deltoides* parent map. The genomic positions corresponding to the peak of the curve are the MLEs of the QTL localization. The solid and broken curves/lines indicate the results from our method and interval mapping, respectively. The threshold values for both methods are given as horizontal lines. The vertical dotted lines indicate the positions of markers on the linkage group (YIN *et al.* 2002), whose names are given beneath.

curves of diameter are drawn using the estimates of logistic parameters for two genotypes at the QTL detected on linkage group 10 (Figure 3). On the basis of the hypothesis test (11), this QTL is detected to be inactive until trees grew to ~ 6 years in the field. And its effect on diameter growth increased with age. At 11 years old, genotype Q_1Q_2 exhibited diameter growth 4.5 cm more than its alternative Q_2Q_2 . This difference appears to increase after age 11 years, as predicted from the logistic curves estimated (Figure 3). Apparently, this QTL interacts significantly with age to affect stem diameter growth.

If two growth curves predicted by a QTL have different ages and/or growth at the inflection point, this indicates that the inflection point is under genetic determination. It is found that the QTL detected on linkage group 10 exerts strong control over the inflection point (Figure 3). The genetic control of the inflection point suggests that the growth trajectory can be genetically modified to increase a tree's capacity to effectively acquire spatial resources.

DISCUSSION

Beyond the traditional models and tools used for quantitative genetic studies, current genome technologies permit us to dissect quantitative traits into individual locus components (QTL). Through this dissection

the genetic basis of quantitative traits can be better unraveled (MACKAY 2001) and, ultimately, genetic improvement for these complex traits can be made more efficient (TANKSLEY 1993; WU *et al.* 2000). Analyses and interpretations of genomic data, however, are strongly dependent upon the study material, data structure, genetic model, and statistical method used. As a result, considerable attention has been paid to the development of powerful experimental designs and analytical methodologies that can increase the power, precision, and resolution of QTL mapping. Currently, there have been many strategies proposed to increase QTL mapping. These include: (1) selecting two highly differentiated inbred lines to make a segregating generation, such as the F_2 or backcross; (2) increasing the sample size of a mapping population by genotyping more progeny; (3) saturating the map density using informative markers, especially in genomic regions carrying QTL; (4) using composite interval mapping and multiple interval mapping (JANSEN and STAM 1994; ZENG 1994; KAO *et al.* 1999); and (5) developing more powerful computational technologies, such as the Bayesian approach implemented with the Markov chain Monte Carlo (MCMC) algorithm (SATAGOPAN *et al.* 1996; XU and YI 2000). In this article, we propose that a simultaneous analysis of repeated measurements for a quantitative trait based on biological mechanisms can be used as an alternative strategy to enhance mapping power and precision.

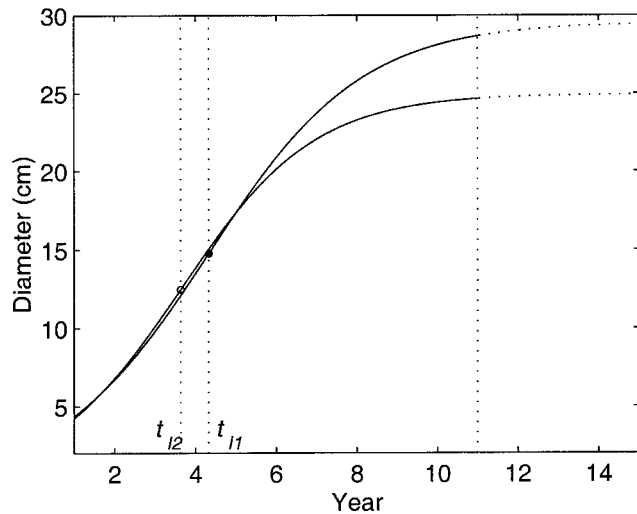


FIGURE 3.—Two growth curves each presenting two groups of genotypes at the QTL detected on linkage group 10 in the *Populus deltoides* parent map. The times at the inflection point (t_{11} and t_{12}) are indicated for the two QTL genotypes Q_1Q_2 and Q_2Q_2 , respectively. The differentiation pattern of growth curves beyond the maximum observed age (11), affected by the QTL, is represented by extended broken curves.

It is well demonstrated that increased sample sizes and marker densities can almost always improve precision in QTL mapping, but they could be economically expensive in practice. Our mapping approach for repeated measurements based on growth curves can extract maximum information about QTL effects and positions contained in an arbitrary segregating family and, thus, confers an advantage for QTL detection in the situation where a limited size of genotyped samples or a limited level of marker density is used. In an example with a small sample size ($N = 90$) using forest tree data, our logistic mixture model offers improved power to detect a number of QTL underlying stem growth, in contrast to traditional approaches based on a single trait, which do not detect any QTL. Such differences are not surprising because a single-trait analysis approach typically cannot detect the QTL of small effect (BEAVIS *et al.* 1994).

The increased detection power of our approach results from the simultaneous use of multiple measurements that are correlated due to either the effect of pleiotropic QTL or residual covariances or both. This, in principle, is similar to the result from multitrait mapping, as shown in JIANG and ZENG (1995), RONIN *et al.* (1995), EAVES *et al.* (1996), MANGIN *et al.* (1998), KNOTT and HALEY (2000), and KOROL *et al.* (1995, 2001). However, beyond these multitrait mapping approaches, our growth-based approach treats phenotypic values as a function of age, thus having the ability to analyze a quantitative trait measured at an unlimited number of time points by modeling the full, continuous growth trajectory. Moreover, instead of estimating a large number of parameters, as needed in the traditional

approaches, our approach estimates a highly reduced number of model parameters, which can make an initially high-dimensional mapping model more tractable and the estimates of QTL parameters more precise.

Composite interval mapping can improve mapping precision to some extent when multiple QTL are located on the same linkage group, but their use frequently depends upon many other factors, *e.g.*, marker spacing, the choice of markers as cofactors, and genotyped sample size (BROMAN 2001). Multiple interval mapping, proposed by Zeng and co-workers, can simultaneously model multiple marker intervals so that multiple QTL and their epistatic interactions can be estimated (KAO *et al.* 1999). Yet, a serious difficulty may be encountered when multiple interval mapping is extended to simultaneously map multiple quantitative traits or repeated measurements at different ages, because a high number of QTL effects should be modeled in these cases. Our logistic-mixture model, when built upon composite interval mapping or multiple interval mapping, can make these two approaches more tractable by reducing the number of model parameters to be estimated. In fact, evidence for more than one QTL observed on some linkage groups from our approach in the poplar example (results not shown) prompts us to build the logistic mixture model upon composite interval mapping or multiple interval mapping and provide better resolution of multiple linked QTL for growth processes.

We have used the method of maximum likelihood to estimate the unknown parameters with their MLEs. The MLEs are attractive in terms of their properties of invariance, consistency, and asymptotic efficiency. Our approach, built upon the traditional maximum-likelihood method, is readily accessible to the general genetics community. Using prior information on parameters, however, we can incorporate the logistic-mixture model in the Bayesian paradigm (SATAGOPAN *et al.* 1996; XU and YI 2000). By specifying the prior density of parameters, MCMC can be used to evaluate the posterior density and provide posterior distributions of QTL effects and positions (ROBERT and CASELLA 1999).

Although the results of our approach are quantified by differences in the parameters controlling the overall shapes of different logistic curves, they can also be interpreted as regular genetic parameters, *i.e.*, the additive or dominant effect of a QTL on growth at an arbitrary time point and the percentage of the total phenotypic variance explained by this QTL. According to classical quantitative genetics theory, the expected genetic values for QTL genotypes Q_1Q_2 and Q_2Q_2 at time t can be expressed, respectively, as

$$g_1(t) = \mu(t) + \frac{1}{2}\alpha(t)$$

$$g_2(t) = \mu(t) - \frac{1}{2}\alpha(t),$$

where $\alpha(t)$ is the additive genetic effect of the QTL detected on growth at time t , which can be solved from the above equations. The additive genetic variance of growth at time t contributed by this QTL is expressed as

$$\sigma_g^2(t) = \frac{1}{4}\alpha^2(t).$$

Thus, the percentage of the total phenotypic variance accounted for by this QTL is

$$R(t) = \frac{\sigma_g^2(t)}{\sigma_g^2(t) + \sigma_e^2(t)}.$$

These parameters described above can also be used to investigate the contribution of a QTL to growth at a point. However, it is important in practice to know how much a QTL contributes to the differentiation of overall growth curves or the differentiation of growth at a time interval. This can be formulated by calculating the integral of the difference of two logistic curves on a particular time interval. In APPENDIX B, the formula for calculating the integral of a logistic curve is given. With the genetic contributions of a QTL to growth, our approach can increase the power of discriminating various important hypotheses that concern the genetic architecture of developmental features (VAUGHN *et al.* 1999). Using the logistic mixture model, the pattern of gene expression for each different QTL can be explicitly described, thus leading to insights on fundamentally important biological questions; *e.g.*, when and how does a QTL affect the phenotype of a quantitative trait during the entire growth trajectory? How does a QTL interact with age to affect the growth and development of an organism? These questions will also have implications for applied breeding programs. If a tree breeder intends to select superior genotypes with high fiber yield at harvesting ages (say 15 years) on the basis of their early performance, a marker-assisted selection strategy incorporating a QTL like one detected on linkage group 10 from our method (Figure 3) can be expected to increase the efficiency of early selection because such a QTL predisposes for productive final fiber yield at age 15 years. With no information about the developmental change of QTL expression, however, this breeder is unable to identify and, therefore, to make use of this QTL in his early selection.

Our method can be extended to incorporate a general biological process of an organism into a QTL mapping framework. Such a process can be allometric scalings (WEST *et al.* 1997, 1999), growth models (GOULD 1977; ALBERCH *et al.* 1979), or continuous responses to the environment in which an organism is reared (VIA *et al.* 1995). However, for clarity of description, we based our analysis on growth curves only. For growth models, we further limited our analysis to sigmoidal or logistic curves. Logistic growth curves are now regarded as one

of the ubiquitous phenomena in biology, holding for every cell, organ, tissue, organism, or population, in a range from microbes (10^{-13} g) to blue whales (10^8 g), no matter what species it is derived from (WEST *et al.* 2001). The pattern of the logistic growth curve can be different among species, populations, and genotypes (HOF *et al.* 1999; ROBERT *et al.* 1999). However, it is also worthwhile to incorporate other biologically meaningful models (reviewed in NIKLAS 1994) into our analysis, as long as they fit well a dataset for particular species, environments, or developmental stages.

To incorporate a general biological process, we should first have a descriptive mathematical function that is expressed as

$$y = \begin{cases} f(x) & \text{for allometric laws} \\ g(t) & \text{for growth models,} \\ h(z) & \text{for reaction norms} \end{cases} \quad (16)$$

where y is the biological trait of interest, x is the body size, t is the age, and z is an environmental variable like temperature, nutrition, or light intensity. The forms of mathematical functions, $f(x)$, $g(t)$, and $h(z)$, which can be linear or nonlinear, are generally different, depending on specific questions of interest. Generally, the establishment of appropriate mathematical functions is based on the goodness of fit to observational data (NIKLAS 1994). Alternatively, these mathematical functions are derived from an optimality perspective. For example, WEST *et al.* (1997, 1999) proposed a fractal-like network system for the absorption and internal distribution of metabolites to explain quarter-power scaling laws pervasive in the living world. In addition, WEST *et al.* (2001) explained why the growth of an organism follows a sigmoid curve based on fundamental principles for the allocation of metabolic energy between maintenance of existing tissue and the production of new biomass.

The method proposed in this study can be extended to other situations, such as partially informative markers or dominant markers, to deal with linked QTL of epistasis or to combine it with selective genotyping. In this study, it is assumed that residual variances and covariances among different ages are stationary. This assumption simplifies the mathematical manipulation of the residual variance-covariance matrix (inversion, factorization, etc.), but may deviate from reality. The extension of our analysis to nonstationary variance-covariance structures is possible, as proposed by NUNEZ-ANTON (1997) and NUNEZ-ANTON and ZIMMERMAN (2000) in their structured antedependent models. Also, Kirkpatrick and co-workers proposed Legendre polynomials to model the dynamic changes of genetic or residual variance and covariance with age (KIRKPATRICK and HECKMAN 1989; KIRKPATRICK *et al.* 1990, 1994). These parametric models for covariance function were improved by PLETCHER and GEYER (1999) to assure the

positive definite property of the functions. These different models for covariance function with some modifications can be incorporated into our mapping strategy.

Functional mapping: Since LANDER and BOTSTEIN's (1989) interval mapping, there has been a wealth of literature reporting on the development of statistical methods for QTL mapping. The transition from a usual single- or two-trait analysis to treatment of multiple measurements from different traits significantly improves all aspects of utilization of the mapping information contained in the data. In traditional mapping strategies, the combination of statistics and molecular genetics makes it possible to identify QTL that contribute to complex traits. However, in this study we attempt to combine powerful statistics and molecular genetics with developmental mechanisms underlying biological features, relationships, and processes to shed light on the genetic basis of complex, or quantitative, traits. This new strategy, which is called *functional mapping* due to the implementation of different mathematical functions of biological means, offers four significant advantages over previous strategies when applied to QTL mapping: (1) Results from functional mapping are closer to biological reality because the underlying biological mechanisms are considered; (2) smaller sample sizes may be used to achieve adequate power and precision for QTL detection because multiple measurements on the same individuals increase precision for mapping; (3) a large number of variables can be analyzed simultaneously by treating growth or a process as a smooth curve, and also the estimates of a small number of parameters can increase the precision of parameter estimation and the flexibility of the model; and (4) functional mapping allows for the testing of different biological hypotheses and this has a direct impact on applied breeding and the developmental studies of genetics and evolution.

We thank Dr. Alan Agresti, Dr. Myron Chang, Dr. Ramon Littell and Dr. Sam Wu for their helpful discussions on this study and three anonymous referees for their constructive comments on the earlier version of this manuscript. This work is partially supported by grants from the National Science Foundation to G.C. (DMS9971586) and an Outstanding Young Investigator Award of the National Natural Science Foundation of China to R.W. (30128017). The publication of this manuscript is approved as journal series R-08640 by the Florida Agricultural Experiment Station.

LITERATURE CITED

- ALBERCH, P., S. J. GOULD, G. F. OSTER and D. B. WAKE, 1979 Size and shape in ontogeny and phylogeny. *Paleobiology* **5**: 296–317.
- ATCHLEY, W. R., 1984 Ontogeny, timing of development, and genetic variance-covariance structure. *Am. Nat.* **123**: 519–540.
- ATCHLEY, W. R., and J. ZHU, 1997 Developmental quantitative genetics, conditional epigenetic variability and growth in mice. *Genetics* **147**: 765–776.
- BEAVIS, W. D., O. S. SMITH, D. GRANT and R. FINCHER, 1994 Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. *Crop Sci.* **34**: 882–896.
- BROMAN, K. W., 2001 Review of statistical methods for QTL mapping in experimental crosses. *Lab Anim.* **30**: 44–52.
- CHEVERUD, J. M., J. J. RUTLEDGE and W. R. ATCHLEY, 1983 Quantitative genetics of development—genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution* **37**: 895–905.
- CHEVERUD, J. M., E. J. ROUTMAN, F. A. M. DUARTE, B. VAN SWINDEREN, K. COTHRAN *et al.*, 1996 Quantitative trait loci for murine growth. *Genetics* **142**: 1305–1319.
- DAVIDIAN, M., and D. M. GILTINAN, 1995 *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, London.
- DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN, 1977 Maximum likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. Ser. B* **39**: 1–38.
- DOERGE, R. W., and G. A. CHURCHILL, 1996 Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**: 285–294.
- EAVES, L. J., M. C. NEALE and H. MAES, 1996 Multivariate multipoint linkage analysis of quantitative trait loci. *Behav. Genet.* **26**: 519–525.
- EMEBIRI, L. C., M. E. DEVEY, A. C. MATHESON and M. U. SLEE, 1998 Age-related changes in the expression of QTLs for growth in radiata pine seedlings. *Theor. Appl. Genet.* **97**: 1053–1061.
- GOULD, S. J., 1977 *Ontogeny and Phylogeny*. Harvard University Press, Cambridge, MA.
- HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.
- HOF, L., L. C. P. KEIZER, I. A. M. ELBERSE and O. A. DOLSTRA, 1999 A model describing the flowering of single plants, and the heritability of flowering traits of *Dimorphothea pluvialis*. *Euphytica* **110**: 35–44.
- JAFFREZIC, F., and S. D. PLETCHER, 2000 Statistical models for estimating the genetic basis of repeated measures and other function-valued traits. *Genetics* **156**: 913–922.
- JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.
- JIANG, C., and Z.-B. ZENG, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**: 1111–1127.
- KAO, C. H., Z.-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- KIRKPATRICK, M., and N. HECKMAN, 1989 A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J. Math. Biol.* **27**: 429–450.
- KIRKPATRICK, M., D. LOFSVOLD and M. BULMER, 1990 Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* **124**: 979–993.
- KIRKPATRICK, M., W. G. HILL and R. THOMPSON, 1994 Estimating the covariance structure of traits during growth and aging, illustrated with lactation in dairy cattle. *Genet. Res.* **64**: 57–69.
- KNOTT, S. A., and C. S. HALEY, 2000 Multitrait least squares for quantitative trait loci detection. *Genetics* **156**: 899–911.
- KOROL, A. B., Y. I. RONIN and V. M. KIRZHNER, 1995 Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics* **140**: 1137–1147.
- KOROL, A. B., Y. I. RONIN, A. M. ITSKOVICH, J. PENG and E. NEVO, 2001 Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. *Genetics* **157**: 1789–1803.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- LOISEL, P., B. GOFFINET, H. MONOD and G. M. DE OCA, 1994 Detecting a major gene in an F₂ population. *Biometrics* **50**: 512–516.
- MACKAY, T. F. C., 2001 Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* **2**: 11–20.
- MANGIN, B., P. THOQUET and N. GRIMSLEY, 1998 Pleiotropic QTL analysis. *Biometrics* **54**: 88–99.
- MAURICIO, R., 2001 Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nat. Rev. Genet.* **2**: 370–381.
- NIKLAS, K. L., 1994 *Plant Allometry: The Scaling of Form and Process*. University of Chicago, Chicago.
- NUNEZ-ANTON, V., 1997 Longitudinal data analysis: non-stationary error structures and antedependent models. *Appl. Stoch. Models Data Anal.* **13**: 279–287.
- NUNEZ-ANTON, V., and D. L. ZIMMERMAN, 2000 Modeling nonstationary longitudinal data. *Biometrics* **56**: 699–705.

NUZHIDIN, S. V., E. G. PASYUKOVA, C. L. DILDA, Z-B. ZENG and T. F. C. MACKAY, 1997 Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **94**: 9734–9739.

PEARL, R., 1925 *The Biology of Population Growth*. Knopf, New York.

PLETCHER, S. D., and C. J. GEYER, 1999 The genetic analysis of age-dependent traits: modeling the character process. *Genetics* **153**: 825–835.

RICE, S. H., 1997 The analysis of ontogenetic trajectories: when a change in size or shape is not heterochrony. *Proc. Natl. Acad. Sci. USA* **94**: 907–912.

ROBERT, C. P., and G. CASELLA, 1999 *Monte Carlo Statistical Methods*. Springer, New York.

ROBERT, N., S. HUET, C. HENNEQUET and A. BOUVIER, 1999 Methodology for choosing a model for wheat kernel growth. *Agronomie* **19**: 405–417.

RONIN, Y. L., V. M. KIRZHNER and A. B. KOROL, 1995 Linkage between loci of quantitative traits and marker loci: multitrait analysis with a single marker. *Theor. Appl. Genet.* **90**: 776–786.

SATAGOPAN, J. M., Y. S. YANDELL, M. A. NEWTON and T. C. OSBORN, 1996 A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805–816.

SAX, K., 1923 The association of size difference with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**: 552–560.

SHAW, R. G., 1987 Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution* **45**: 143–151.

SILLANPAA, M. J., and E. ARJAS, 1999 Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* **151**: 1605–1619.

TANKSLEY, S. D., 1993 Mapping genes. *Annu. Rev. Genet.* **27**: 205–233.

VAUGHN, T. T., L. S. PLETSCHER, A. PERIPATO, K. KING-ELLISON, E. ADAMS *et al.*, 1999 Mapping quantitative trait loci for murine growth: a closer look at genetic architecture. *Genet. Res.* **74**: 313–322.

VERBEKE, G., and G. MOLENBERGHS, 2000 *Linear Mixed Models for Longitudinal Data*. Springer, New York.

VERHAEGEN, D., C. PLOMION, J. M. GION, M. POITEL, P. COSTA *et al.*, 1997 Quantitative trait dissection analysis in Eucalyptus using RAPD markers. I. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theor. Appl. Genet.* **95**: 597–608.

VIA, S., R. GOMULKIEWICZ, G. D. JONG, S. M. SCHEINER, C. D. SCHLICHTING *et al.*, 1995 Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol. Evol.* **10**: 212–217.

VON BERTALANFFY, L., 1957 Quantitative laws in metabolism and growth. *Q. Rev. Biol.* **32**: 217–231.

XU, S., 1995 A comment on the simple regression method for interval mapping. *Genetics* **141**: 1657–1659.

XU, S. Z., and N. J. YI, 2000 Mixed model analysis of quantitative trait loci. *Proc. Natl. Acad. Sci. USA* **97**: 14542–14547.

WEST, G. B., J. H. BROWN and B. J. ENQUIST, 1997 A general model for the origin of allometric scaling laws in biology. *Science* **276**: 122–126.

WEST, G. B., J. H. BROWN and B. J. ENQUIST, 1999 The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**: 1677–1679.

WEST, G. B., J. H. BROWN and B. J. ENQUIST, 2001 A general model for ontogenetic growth. *Nature* **413**: 628–631.

WU, R. L., M. X. WANG and M. R. HUANG, 1992 Quantitative genetics of yield breeding for Populus short rotation culture. I. Dynamics of genetic control and selection models of yield traits. *Can. J. For. Res.* **22**: 175–182.

WU, R. L., Z-B. ZENG, S. E. MCKEAND and D. M. O'MALLEY, 2000 The case for molecular mapping in forest tree breeding. *Plant Breed. Rev.* **19**: 41–67.

WU, W. R., W. M. LI, D. Z. TANG, H. R. LU and A. J. WORLAND, 1999 Time-related mapping of quantitative trait loci underlying tiller number in rice. *Genetics* **151**: 297–303.

YIN, T. M., X. Y. ZHANG, M. R. HUANG, M. X. WANG, Q. ZHUGE *et al.*, 2002 Molecular linkage maps of the Populus genome. *Genome* **45**: 541–555.

ZENG, Z-B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

APPENDIX A

In what follows, we derive the log-likelihood functions used to estimate the parameters in $\Omega = (a_j, b_j, r_j, \rho, \sigma^2)$. The symbol ' denotes the estimates of parameters from the previous step.

$$a'_j = \left(\sum_{i=1}^N P_{ij} \left[\sum_{t=1}^{m-1} (y_i(t) - \rho y_i(t+1)) \left(\frac{1}{1 + b_j e^{-r'_j t}} - \frac{\rho}{1 + b_j e^{-r'_j(t+1)}} \right) + \frac{(1 - \rho^2) y_i(m)}{1 + b_j e^{-r'_j m}} \right] \right) / \left(\sum_{i=1}^N P_{ij} \left[\sum_{t=1}^{m-1} \left(\frac{1}{1 + b_j e^{-r'_j t}} - \frac{\rho}{1 + b_j e^{-r'_j(t+1)}} \right)^2 + \frac{1 - \rho^2}{(1 + b_j e^{-r'_j m})^2} \right] \right),$$

$$b'_j = \left(b_j \sum_{i=1}^N P_{ij} \left[\sum_{t=1}^{m-1} \left(\frac{a_j}{1 + b_j e^{-r'_j t}} - \frac{a_j \rho}{1 + b_j e^{-r'_j(t+1)}} \right) \left(\frac{e^{-r'_j t}}{(1 + b_j e^{-r'_j t})^2} - \frac{\rho e^{-r'_j(t+1)}}{(1 + b_j e^{-r'_j(t+1)})^2} \right) + \frac{(1 - \rho^2) a_j e^{-r'_j m}}{(1 + b_j e^{-r'_j m})^2} \right] \right) / \left(\sum_{i=1}^N P_{ij} \left[\sum_{t=1}^{m-1} (y_i(t) - \rho y_i(t+1)) \times \left(\frac{e^{-r'_j t}}{(1 + b_j e^{-r'_j t})^2} - \frac{\rho e^{-r'_j(t+1)}}{(1 + b_j e^{-r'_j(t+1)})^2} \right) + \frac{(1 - \rho^2) y_i(m) e^{-r'_j m}}{(1 + b_j e^{-r'_j m})^2} \right] \right),$$

$$r'_j = \left(r_j \sum_{i=1}^N P_{ij} \left[\sum_{t=1}^{m-1} (y_i(t) - \rho y_i(t+1)) \left(\frac{t e^{-r'_j t}}{(1 + b_j e^{-r'_j t})^2} - \frac{(t+1) \rho e^{-r'_j(t+1)}}{(1 + b_j e^{-r'_j(t+1)})^2} \right) + \frac{(1 - \rho^2) m y_i(m) e^{-r'_j m}}{(1 + b_j e^{-r'_j m})^2} \right] \right) / \left(\sum_{i=1}^N P_{ij} \left[\sum_{t=1}^{m-1} \left(\frac{a_j}{1 + b_j e^{-r'_j t}} - \frac{a_j \rho}{1 + b_j e^{-r'_j(t+1)}} \right) \times \left(\frac{t e^{-r'_j t}}{(1 + b_j e^{-r'_j t})^2} - \frac{(t+1) \rho e^{-r'_j(t+1)}}{(1 + b_j e^{-r'_j(t+1)})^2} \right) + \frac{(1 - \rho^2) m a_j e^{-r'_j m}}{(1 + b_j e^{-r'_j m})^2} \right] \right),$$

and

$$\sigma'^2 = \frac{1}{mN(1 - \rho^2)} \sum_{j=1}^2 \sum_{i=1}^N P_{ij} \left\{ \sum_{t=1}^{m-1} [(y_i(t) - g_j(t)) - \rho(y_i(t+1) - g_j(t+1))]^2 + (1 - \rho^2)(y_i(m) - g_j(m))^2 \right\},$$

$$\rho' = \left[1 - \frac{Nm - N - B}{A} \rho^3 - \frac{B + C - Nm + N}{A} \rho \right]^{1/2},$$

where

$$A = \sum_{j=1}^2 \sum_{i=1}^N P_{ij} \sum_{t=1}^{m-1} (y_i(t) - g_j(t))(y_i(t+1) - g_j(t+1)),$$

$$B = \sum_{j=1}^2 \sum_{i=1}^N P_{ij} \left[\sum_{t=2}^m (y_i(t) - g_j(t))^2 - (y_i(m) - g_j(m))^2 \right],$$

$$C = \frac{1}{\sigma^2} \sum_{j=1}^2 \sum_{i=1}^N P_{ij} \left\{ \sum_{t=1}^{m-1} [(y_i(t) - g_j(t)) - \rho(y_i(t+1) - g_j(t+1))]^2 + (1 - \rho^2)(y_i(m) - g_j(m))^2 \right\}.$$

The values of $(a'_1, b'_1, r'_1, a'_0, b'_0, r'_0, \rho', \sigma'^2)$ estimated from

the above equations will be used to provide new estimators of Ω in the next step.

APPENDIX B

Below, we describe a mathematical procedure for calculating the integral of a logistic curve,

$$g(t) = \frac{a}{1 + be^{-rt}},$$

on the interval $[t_1, t_2]$. The integral of the curve on this interval is expressed as

$$\begin{aligned} G[t_1, t_2] &= \int_{t_1}^{t_2} \frac{a}{1 + be^{-rt}} dt \\ &= \int_{t_1}^{t_2} \frac{ae^{rt}}{e^{rt} + b} dt. \end{aligned}$$

Letting $y = b + e^{rt}$, we have

$$\frac{dy}{dt} = re^{rt}$$

and, thus,

$$\frac{dy}{r} = e^{rt} dt.$$

Also, when $t = t_1$ or t_2 , we have the limits of y as $b + e^{rt_1}$ or $b + e^{rt_2}$, respectively. Therefore, we have

$$\begin{aligned} G[t_1, t_2] &= \int_{t_1}^{t_2} \frac{a}{ry} dy \\ &= \frac{a}{r} \ln y \Big|_{b+e^{rt_1}}^{b+e^{rt_2}} \\ &= \frac{a}{r} [\ln(b + e^{rt_2}) - \ln(b + e^{rt_1})]. \end{aligned}$$