A General Framework for Analyzing the Genetic Architecture of Developmental Characteristics

Rongling Wu,*,†,[1] Chang-Xing Ma,* Min Lin* and George Casella*

**Department of Statistics, University of Florida, Gainesville, Florida 32611 and †Institute of Statistical Genetics, Zhejiang Forestry University, Lin'an, Zhejiang 311300, People's Republic of China*

Manuscript received September 24, 2003
Accepted for publication November 24, 2003

ABSTRACT

The genetic architecture of growth traits plays a central role in shaping the growth, development, and evolution of organisms. While a limited number of models have been devised to estimate genetic effects on complex phenotypes, no model has been available to examine how gene actions and interactions alter the ontogenetic development of an organism and transform the altered ontogeny into descendants. In this article, we present a novel statistical model for mapping quantitative trait loci (QTL) determining the developmental process of complex traits. Our model is constructed within the traditional maximum-likelihood framework implemented with the EM algorithm. We employ biologically meaningful growth curve equations to model time-specific expected genetic values and the AR(1) model to structure the residual variance-covariance matrix among different time points. Because of a reduced number of parameters being estimated and the incorporation of biological principles, the new model displays increased statistical power to detect QTL exerting an effect on the shape of ontogenetic growth and development. The model allows for the tests of a number of biological hypotheses regarding the role of epistasis in determining biological growth, form, and shape and for the resolution of developmental problems at the interface with evolution. Using our newly developed model, we have successfully detected significant additive × additive epistatic effects on stem height growth trajectories in a forest tree.

THE evolution of complex organisms, such as animals and plants, does not result simply from the direct transformation of adult ancestors into adult descendants, but rather involves a cascade of developmental processes that produce the new features of each generation. An increasing number of evolutionary studies have been launched to determine the genetic or developmental changes in the rate or timing of developmental processes that must take place to derive a particular phenotype from its ancestor (Rice 1997; Raff 2000; Rougvie 2001). A general view is that the evolution of developmental processes is affected by both the environment and many genes that act singly and in interaction with each other (Lynch and Walsh 1998). However, to accurately predict the direction and rate of trait evolution, a detailed genetic architecture of how genes act and interact to control various stages of development must be quantified.

The genes predisposing for a phenotypic character that displays continuous variation among individuals are referred to as quantitative trait loci (QTL). The genetic effect or variance of QTL includes two components, *additive*, due to the cumulation of breeding values, and *nonadditive*, due to allelic (dominant) or nonallelic (epistatic) interactions. Epistatic interactions between different loci can be further partitioned into different types: additive × additive, additive × dominant (or dominant × additive), and dominant × dominant. The presence of epistasis implies that the influence of a gene on the phenotype depends critically upon the context provided by other genes. In the past, the estimation of the additive and nonadditive genetic architecture of a quantitative trait was based on the phenotypes of related individuals (Lynch and Walsh 1998), although this has minimal power to detect the nonadditive genetic variances, especially epistatic variance because epistasis contributes little to the resemblance among relatives (Cheverud and Routman 1995).

The advent of DNA-based linkage maps opens a novel avenue for precisely estimating the genetic architecture of developmental traits (Vaughn *et al.* 1999). Current statistical methods proposed to detect the main and interaction effects of QTL are based on the phenotypes of a quantitative trait measured at a limited set of landmark ages. More recently, Wu *et al.* (2002, 2003a,b) and Ma *et al.* (2002) have derived a powerful functional mapping method for estimating the dynamic changes of QTL effects during a course of ontogenetic growth through the implementation of universal growth laws (West *et al.* 2001) and the structured residual (co)variance matrix among different time points (see Kirkpatrick and Heckman 1989; Kirkpatrick *et al.* 1990, 1994; Pletcher and Geyer 1999). This method has proven to be statistically more powerful and more precise be-

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

cause of a reduced number of parameters being estimated and to be biologically more meaningful due to the consideration of biological principles underlying trait development (Wu *et al.* 2002). However, this model has not incorporated the estimation process of epistatic interactions and, thus, cannot examine the role of the entire genetic architecture in developmental trajectories.

In this article, we extend the functional mapping method to map any QTL (including additive, dominant, and epistatic) that transforms allelic and/or nonallelic effects into final phenotypes during a continuous process of development represented as ontogenetic trajectories or a path through phenotype-time space (Alberch *et al.* 1979; Wolf *et al.* 2000). We derive special procedures to estimate and test the impact of epistasis on trait growth because a growing body of evidence now shows that epistasis plays a more important role in determining developmental changes than originally thought (Rice 1997, 2000; Wolf *et al.* 2000). Epistasis can trigger an effect on the evolution of development across different levels of biological organization and these include the molecular mechanisms of gene expression and genetic architecture, the evolution of sex and recombination, the genetic coadaptation and its associated outbreeding depression, adaptive evolution, and the very process of speciation (Wolf *et al.* 2000). We use a maximum-likelihood-based method, implemented with the expectation-maximization (EM) algorithm, to estimate QTL locations and genetic effects on growth differentiation. Compared with current mapping methods, our method of incorporating growth trajectories tends to be more powerful and more precise in QTL detection and effect estimation, as demonstrated in an example using forest tree data. In practice, our method is economically more feasible than previous methods because it needs a smaller size of genotyped samples to obtain adequate power for QTL detection through the use of repeated measurements for each individual. It can be anticipated that the method proposed in this article will have potential implications for understanding the origin and evolution of development and the contributions of epistatic effects to evolutionary changes in the process of development.

GROWTH EQUATIONS AND MIXTURE MODEL

Growth equations: When growth g is plotted against time t, different forms of growth curves will appear. Among these forms, a logistic growth curve (also referred to as the sigmoid curve of growth; Niklas 1994) is one of the most ubiquitous, having been derived from fundamental physiological and physical principles (West *et al.* 2001). The logistic growth curve can be mathematically described by

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

where a is the asymptotic value of g when $t \to \infty$, $a/(1 + b)$ is the value of g at $t = 0$, and c is the relative rate of growth (Bertalanffy 1957). The logistic growth curve consists of two phases, exponential and asymptotic. The overall form of the curve is determined by different combinations of parameters a, b, and c.

In evolutionary biology, a question of how a population evolves on a logistic curve is determined by how selection acts on the growth and by the local geometry of the curve itself. Some geometric properties of the growth curve have straightforward biological interpretations. For example, the slope of the logistic curve at any given time point measures the degree to which the value of growth is sensitive to a change in age:

$$\frac{dg(t)}{dt} = cg(t)\left[1 - \frac{g(t)}{a}\right]. \tag{2}$$

Such a slope represents the rate of growth at a given time. Thus, if the slope at a point is low, then that value of growth is locally buffered against age changes. The rate of growth drops off linearly as the overall size approaches some limit.

From a growth curve, we can derive the timing (t_1) of the inflection point, at which the exponential phase ends and the asymptotic phase begins (Niklas 1994). For the logistic curve, t_1 is derived as

$$t_1 = \frac{\ln b}{c}. \tag{3}$$

The inflection point is thought to play an important role in shaping ontogenetic growth and development.

The area under the logistic curve at an interval $[t_1\ t_2]$ describes the capacity of a given organism to grow over time. Such an area is the integral of the logistic curve, expressed as

$$\begin{aligned} G[t_1\ t_2] &= \int_{t_1}^{t_2} \frac{a}{1 + be^{-ct}}dt \\ &= \frac{a}{c}[\ln(b + e^{-ct_2}) - \ln(b + e^{-ct_1})]. \end{aligned} \tag{4}$$

Quantitative genetic model: We start with a simple F_2 population of size n derived from two homozygous lines. Consider two segregating QTL responsible for a quantitative trait, $\mathcal{Q}_k$ and $\mathcal{Q}_l$, with three genotypes, Q_kQ_k, Q_kq_k, q_kq_k, and Q_lQ_l, Q_lq_l, q_lq_l, respectively. Three genotypes at a QTL are denoted by j_k that takes 2, 1, or 0 depending on the number of capitalized alleles. The genotypic value of a two-QTL genotype (j_kj_l) can be expressed by a linear model,

$$\begin{aligned} g_{j_kj_l} = \mu &+ x_k\alpha_k + z_k\beta_k + x_l\alpha_l + z_l\beta_l + w_{\alpha\alpha}i_{\alpha\alpha} + w_{\alpha\beta}i_{\alpha\beta} \\ &+ w_{\beta\alpha}i_{\beta\alpha} + w_{\beta\beta}i_{\beta\beta}, \end{aligned} \tag{5}$$

where μ is the overall mean; α_k, β_k and α_l, β_l are the additive and dominant effects of the two QTL, respectively; and $i_{\alpha\alpha}$, $i_{\alpha\beta}$, $i_{\beta\alpha}$, and $i_{\beta\beta}$ are the epistatic effects between the

two QTL due to additive × additive, additive × dominant, dominant × additive, and dominant × dominant interactions, respectively. The dummy variables in Equation 5 are defined as

$$x_k = \begin{cases} 1 & \text{for genotype } Q_kQ_k \text{ at } \mathcal{Q}_k \\ 0 & \text{for genotype } Q_kq_k \text{ at } \mathcal{Q}_k \\ -1 & \text{for genotype } q_kq_k \text{ at } \mathcal{Q}_k, \end{cases}$$

$$x_l = \begin{cases} 1 & \text{for genotype } Q_lQ_l \text{ at } \mathcal{Q}_l \\ 0 & \text{for genotype } Q_lq_l \text{ at } \mathcal{Q}_l \\ -1 & \text{for genotype } q_lq_l \text{ at } \mathcal{Q}_l, \end{cases}$$

$$z_k = \begin{cases} 1 & \text{for genotype } Q_kq_k \text{ at } \mathcal{Q}_k \\ 0 & \text{otherwise}, \end{cases}$$

$$z_l = \begin{cases} 1 & \text{for genotype } Q_lq_l \text{ at } \mathcal{Q}_l \\ 0 & \text{otherwise}, \end{cases}$$

with $w_{\alpha\alpha} = x_kx_l$, $w_{\alpha\beta} = x_kz_l$, $w_{\beta\alpha} = z_kx_l$, and $w_{\beta\beta} = z_kz_l$. These two QTL can be mapped using a genetic linkage map constructed from molecular markers. There are two possibilities for the locations of the two QTL: (1) the two QTL are located on two different marker intervals or (2) the two QTL are located on the same marker interval. Consider QTL $\mathcal{Q}_k$ bracketed by two flanking markers $\mathcal{M}_u$ and $\mathcal{M}_{u+1}$. The recombination fractions are denoted by r_u, r_{k_1}, and r_{k_2}, respectively, between the two markers, between $\mathcal{M}_u$ and $\mathcal{Q}_k$, and between $\mathcal{Q}_k$ and $\mathcal{M}_{u+1}$. The conditional probability of a given QTL genotype, conditional upon the marker genotypes for F_2 progeny i, can be generally expressed as

$$\pi_{ij_k} = \text{Prob}(i = j_k | \mathcal{M}_u, \mathcal{M}_{u+1}, r_{k_1}, r_{k_2}, r_u),$$

which depends on the location of the QTL on the marker interval, characterized by r_{k_1} and r_{k_2}. Considering all possible two-marker genotypes and QTL genotypes, π_{ij_k} forms a (9×3) matrix. If a second QTL $\mathcal{Q}_l$ is located on a different marker interval $[\mathcal{N}_v, \mathcal{N}_{v+1}]$, the conditional probabilities of a two-QTL genotype given marker intervals for progeny i are the product of the corresponding probabilities of a one-QTL genotype, *i.e.*,

$$\pi_{ij_kj_l} = \text{Prob}(i = j_k | \mathcal{M}_u, \mathcal{M}_{u+1}, r_{k_1}, r_{k_2}, r_u) \times \text{Prob}(i = j_l | \mathcal{N}_v, \mathcal{N}_{v+1}, r_{l_1}, r_{l_2}, r_v), \quad (6)$$

which forms a (81×9) conditional probability matrix. If two QTL are located on the same marker interval, the conditional probability is expressed as

$$\pi_{ij_kj_{k+1}} = \text{Prob}(i = j_kj_{k+1} | \mathcal{M}_u, \mathcal{M}_{u+1}, r_{k_1}, r_{k_2}, r_{k_3}, r_u). \quad (7)$$

In Equation 7, denote r_{k_1}, r_{k_2}, and r_{k_3} to be the recombination fractions between marker $\mathcal{M}_u$ and QTL $\mathcal{Q}_k$, between QTL $\mathcal{Q}_k$ and $\mathcal{Q}_{k+1}$, and between $\mathcal{Q}_{k+1}$ and marker $\mathcal{M}_{u+1}$, respectively. For two QTL on the same interval, $\pi_{ij_kj_{k+1}}$ forms a (9×9) matrix.

Logistic mixture model for mapping epistatic QTL: Unlike traditional statistical models, in which marker information is associated with phenotypic values measured at one time point, our model intends to map QTL for an infinite-dimensional trait expressed as a function of time (Kirkpatrick and Heckman 1989; Kirkpatrick *et al.* 1990, 1994; Pletcher and Geyer 1999). Yet, the modeling of the functional relationship of a trait with time needs to be based on the measurements made at a finite number of landmark ages. It is reasonable to assume that the phenotypes of an infinite-dimensional trait measured at all time points 1, . . . , m for each putative QTL genotype group follow a multivariate normal density,

$$f_{j_kj_l}(\mathbf{y}) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{g}_{j_kj_l})^T\Sigma^{-1}(\mathbf{y} - \mathbf{g}_{j_kj_l})\right],$$

where $\mathbf{g}_{j_kj_l}$ is the vector of the expected genotypic values of the trait measured for m times for a two-QTL genotype j_kj_l at $\mathcal{Q}_k$ and $\mathcal{Q}_l$ and Σ is the residual variance-covariance matrix of the phenotypes measured at different times. Assuming that the two putative QTL jointly affect the growth process, $\mathbf{g}_{j_kj_l}$ can be modeled by a growth equation. For the logistic curve of Equation 1, we have

$$\mathbf{g}_{j_kj_l} = [g_{j_kj_l}(t)]_{m\times 1} = \left[\frac{a_{j_kj_l}}{1 + b_{j_kj_l}e^{-c_{j_kj_l}t}}\right]_{m\times 1},$$

where each group of growth parameters (a, b, c) corresponds to a different QTL genotype. To increase the model's flexibility, the residual (co)variance matrix Σ need to be structured using the first-order autoregressive [AR(1)] model (Davidian and Giltinan 1995), expressed as

$$\Sigma = \sigma^2\begin{bmatrix} 1 & \rho & \cdots & \rho^{m-1} \\ \rho & 1 & \cdots & \rho^{m-2} \\ \cdots & \cdots & \ddots & \cdots \\ \rho^{m-1} & \rho^{m-2} & \cdots & 1 \end{bmatrix}, \quad (8)$$

in which we assume variance stationarity, *i.e.*, there is the same residual variance (σ^2) for growth at different ages, and covariance stationarity, *i.e.*, the covariance of growth between different ages, decreases proportionally (in correlation ρ) with increased time interval (*cf.* Pletcher and Geyer 1999). There are two advantages when the structured matrix (8) is used. First, an explicit expression of the determinant and inverse of Σ can be derived, which facilitates parameter estimation. Second, with such an expression, the growth-model-based mapping approach can be applied for an arbitrary number of time points.

The assumption of variance stationarity can be satisfied by transforming both sides (TBS) of the growth equation (1), as proposed by Carroll and Ruppert (1984). The transformation at the left side of Equation 1 can lead to a homogeneous variance over times, whereas the transformation at the right side of Equation

1 can preserve the biological properties of growth parameters (a, b, c). Thus, with the TBS model, the favorable advantages of structuring Σ according to (8) can be preserved.

We formulate the likelihood function of the phenotypic data with m-dimensional measurements as

$$L(\mathbf{\Omega}) = \prod_{i=1}^{n}\left[\sum_{j_k=0}^{2}\sum_{j_l=0}^{2}\pi_{ij_kj_l}f_{j_kj_l}(\mathbf{y}_i)\right], \quad (9)$$

where the vector $\mathbf{\Omega} = (a_{j_kj_l}, b_{j_kj_l}, c_{j_kj_l}, r_{k_1}, r_{l_1}, \rho, \sigma^2)^T$ contains unknown parameters for the QTL effects, QTL position, and residual (co)variances. The position parameters r_{k_1} and r_{l_1} depend on whether the two QTL are tested on different intervals or the same interval.

The EM algorithm: The maximum-likelihood estimates (MLEs) of the unknown parameters under a two-QTL model can be computed by implementing the EM algorithm (Dempster *et al.* 1977; Lander and Botstein 1989). We have incorporated the growth law (1) into the mixture-based likelihood function (9) and derived the log-likelihood equations to estimate $\mathbf{\Omega}$. In the E step, calculate the expected conditional (posterior) probability of a two-QTL genotype j_kj_l given marker genotypes for progeny i,

$$\prod_{ij_kj_l} = \frac{\pi_{ij_kj_l}f_{j_kj_l}(\mathbf{y}_i)}{\sum_{j_k=0}^{2}\sum_{j_l=0}^{2}\pi_{ij_kj_l}f_{j_kj_l}(\mathbf{y}_i)}.$$

In the M step, these posterior probabilities are used to solve the unknown parameters on the basis of the log-likelihood equations. The E and M steps are iterated until the estimates converge.

In practical computations, the QTL position parameters can be viewed as nuisance parameters because two putative QTL can be searched at given positions throughout the entire linkage map. The amount of support for the QTL at particular map positions is often displayed graphically through the use of likelihood maps or profiles, which plot the likelihood-ratio test statistics as a function of map positions of the two putative QTL.

HYPOTHESIS TESTS

Different from traditional mapping approaches, our functional mapping for epistatic QTL allows for the tests of a number of biologically meaningful hypotheses. These hypothesis tests can be a *global* test for the existence of significant QTL, a *local* test for the genetic effect on growth at a particular time point, a *regional* test for the overall effect of QTL on a particular period of growth process, or an *interaction* test for the change of QTL expression across ages.

Global test: Testing whether specific QTL exist to affect the shape of growth trajectories is a first step toward the understanding of the genetic architecture of complex phenotypes. The genetic control over entire growth trajectories can be tested by formulating the following hypotheses:

$$\begin{aligned} &H_0:\ a_{22} = \ldots = a_{00},\ b_{22} = \ldots = b_{00},\ c_{22} = \ldots = c_{00} \\ &H_1:\ \text{Not all of these equalities above hold.} \end{aligned} \quad (10)$$

H_0 states that no QTL affect growth trajectories (the reduced model), whereas H_1 proposes that such QTL do exist (the full model). The test statistic for testing the hypotheses (10) is calculated as the log-likelihood ratio of the reduced to the full model,

$$LR = -2[\ln L(\tilde{\mathbf{\Omega}}) - \ln L(\hat{\mathbf{\Omega}})], \quad (11)$$

where $\tilde{\mathbf{\Omega}}$ and $\hat{\mathbf{\Omega}}$ denote the MLEs of the unknown parameters under H_0 and H_1, respectively. The LR is asymptotically χ^2 distributed with 9 d.f. An empirical approach for determining the critical threshold is based on permutation tests, as advocated by Churchill and Doerge (1994). By repeatedly shuffling the relationships between marker genotypes and phenotypes, a series of the maximum log-likelihood ratios are calculated, from the distribution of which the critical threshold is determined.

We can also test the global effects of different genetic components, additive, dominant, and epistatic, on the shapes of entire growth curves. The hypothesis for testing the additive effect of QTL $\mathcal{Q}_k$ on overall growth curves can be formulated as

$$\begin{aligned} &H_0:\ \alpha_k(t) = 0 \\ &H_1:\ \alpha_k(t) \neq 0, \end{aligned} \quad (12)$$

which is equivalent to testing the difference of the full model with no restriction and the reduced model with a restriction:

$$\sum_{j_l=0}^{2} g_{2j_l}(t) = \sum_{j_l=0}^{2} g_{0j_l}(t). \quad (13)$$

Thus, the data can be fit by one less unknown parameter under the reduced model (H_0) of (12) than under the full model (H_1). An empirical approach for determining the critical threshold for the hypothesis test of (12) is based on simulation studies. Phenotypic data following a multivariate normal density are simulated for different groups of QTL genotypes whose time-dependent expected values are restricted using Equation 13. These simulated data that include no additive effect due to QTL $\mathcal{Q}_k$ are analyzed. The threshold value is determined on the basis of the distribution of the likelihood ratios (LRs) obtained from simulation replicates.

The test for the dominant effect, $\beta_k(t)$, of QTL $\mathcal{Q}_k$ is equivalent to testing the difference of the full model with no restriction and the reduced model with a restriction:

$$2\sum_{j_l=0}^{2} g_{1j_l}(t) = \sum_{j_l=0}^{2}[g_{2j_l}(t) + g_{0j_l}(t)]. \quad (14)$$

Similarly, the additive and dominant effects of QTL $\mathcal{Q}_l$ at a time point t can be tested using the following restrictions, respectively:

$$\sum_{j_k=0}^{2} g_{j_k 2}(t) = \sum_{j_k=0}^{2} g_{j_k 0}(t), \tag{15}$$

$$2\sum_{j_k=0}^{2} g_{j_k 1}(t) = \sum_{j_k=0}^{2} [g_{j_k 2}(t) + g_{j_k 0}(t)]. \tag{16}$$

The test for the epistatic effects between the two QTL is equivalent to testing the differences of the full model with no restriction and the reduced model with a restriction,

$$g_{22}(t) + g_{00}(t) = g_{20}(t) + g_{02}(t), \tag{17}$$

for the additive × additive effect;

$$2g_{21}(t) + g_{02}(t) + g_{01}(t) = 2g_{01}(t) + g_{20}(t) + g_{00}(t), \tag{18}$$

for the additive × dominant effect;

$$2g_{12}(t) + g_{20}(t) + g_{00}(t) = 2g_{01}(t) + g_{22}(t) + g_{12}(t), \tag{19}$$

for the dominant × additive effect; and

$$4g_{11}(t) + g_{22}(t) + g_{20}(t) + g_{02}(t) + g_{00}(t) = 2g_{21}(t) + 2g_{12}(t) + 2g_{10}(t) + 2g_{01}(t), \tag{20}$$

for the dominant × dominant effect. In fact, these formulations of the restrictions for testing epistatic interactions of different kinds on growth trajectories are a simple extension of Cheverud and Routman's (1995) epistatic model for specifying the epistasis for a stationary trait. Simulation studies are performed to determine the critical thresholds for hypotheses 14–20.

Local test: The local test can test the significance of the main (additive or dominant) effect of each QTL and the interaction (epistatic) effect between the two QTL on growth measured at a time point (t^*) of interest. The tests of additive and dominant effects of individual QTL and their epistatic effects can be made on the basis of the corresponding restrictions given in Equations 13–20. For example, the hypothesis for testing the additive effect of QTL $\mathcal{Q}_k$ on growth at a given time t^* can be formulated as

$$\begin{aligned} &H_0: \ \alpha_k(t^*) = 0 \\ &H_1: \ \alpha_k(t^*) \neq 0, \end{aligned} \tag{21}$$

which is equivalent to testing the difference of the full model with no restriction and the reduced model with a restriction:

$$\sum_{j_l=0}^{2} g_{2j_l}(t^*) = \sum_{j_l=0}^{2} g_{0j_l}(t^*). \tag{22}$$

Regional test: It is likely that an important developmental event often occurs in a time interval rather than simply at a time point. The question of how QTL exert their effects on a period of growth process $[t_1, t_2]$ can be tested using a regional test approach based on the areas ($G_{j_k j_l}[t_1, t_2]$, Equation 4) covered by growth curves. The hypothesis test for the genetic effect on a period of growth process is equivalent to testing the difference between the full model with no restriction and the reduced model with a restriction. The types of restriction used are similar to Equations 13–20, depending on the additive effects, dominant effects, or epistatic effects of different kinds.

Interaction test: The effects of QTL may change with age, which suggests the occurrence of QTL × age interaction effects on the growth process. The differentiation of $g(t)$ with respect to time t represents a slope of the growth curve (growth rate, Equation 2). If the slopes at a particular time point t^* are different between the curves of different QTL genotypes, this means that significant QTL × age interaction occurs between this time point and the next. The test for QTL × age interaction can be formulated with the restriction

$$\sum_{j_l=0}^{2} \frac{dg_{2j_l}(t^*)}{dt} = \sum_{j_l=0}^{2} \frac{dg_{0j_l}(t^*)}{dt} \tag{23}$$

for the additive effect of QTL $\mathcal{Q}_k$. The tests for QTL × age interactions due to the other genetic effects can be formulated using Equations 14–20.

The effect of QTL × age interaction on growth can be examined during the entire growth trajectories. The global test of QTL × age interaction due to the additive effect of QTL $\mathcal{Q}_k$ can be formulated with the restriction

$$\int_0^m \left[\sum_{j_l=0}^{2} \frac{dg_{2j_l}(t)}{dt} - \sum_{j_l=0}^{2} \frac{dg_{0j_l}(t)}{dt} \right]^2 dt = 0. \tag{24}$$

The restriction (24) means that there is the same slope at every time point between the two logistic curves of QTL genotypes Q_kQ_k and q_kq_k, thus suggesting that the additive effect of $\mathcal{Q}_k$ does not lead to significant QTL × age interaction on entire growth trajectories. Similar global tests of QTL × age interactions due to the other genetic effects can also be made, depending on the types of restrictions as shown in Equations 14–20.

Test for the timing of development: During its ontogenetic growth, an organism would experience various developmental events. The genetic determination of the timing of development sheds light on the theoretical integration of evolution and development (Raff 2000; Rougvie 2001). Using our functional mapping model, the genotypic differences in the timing (t_l) of the inflection point of maximum growth rate can be tested. According to Equation 2, the test for such a genotypic difference due to the additive effect of QTL $\mathcal{Q}_k$ is based on the restriction,

$$\sum_{j_l=0}^{2} \frac{\ln b_{2j_l}}{r_{2j_l}} = \sum_{j_l=0}^{2} \frac{\ln b_{0j_l}}{r_{0j_l}}. \tag{25}$$

TABLE 1

MLEs of the positions of two QTL, each bracketed by a different marker interval, QTL effects described by growth parameters (*a*, *b*, *c*), and phase probabilities for the QTL located within a marker interval on linkage group D16 in an interspecific hybrid population of Populus

	Position		QTL genotype			
	$\Im_1$	$\Im_2$	$Q_1q_1Q_2q_2$	$Q_1q_1q_2q_2$	$q_1q_1Q_2q_2$	$q_1q_1q_2q_2$
Marker interval	TC/CTG-730–TC/CTG-735	AG/CTT-570–AG/CTT-595				
Phase probability[a]	0.95	0.90				
$\hat{a}$			18.80	22.43	25.49	17.56
$\hat{b}$			6.86	10.87	89.05	6.33
$\hat{c}$			0.4607	0.4730	0.7113	0.4648
t_1			4.18	5.04	6.31	3.97

[a] Linkage phase here is meant between the uppercase allele of the QTL and the dominant alleles of the flanking markers for both intervals.

The tests of the control of other genetic components over the timing of the inflection point can be similarly made.

A CASE STUDY

Materials: The power of our statistical model for mapping QTL affecting growth trajectories was demonstrated by a case study in poplar trees. A *Populus deltoides* clone (designated I-69) was used as a female parent to mate with an interspecific *P. deltoides* × *P. nigra* clone (designated I-45) as a male parent (Wu *et al.* 1992). A total of 450 1-year-old rooted hybrid seedlings from this cross were planted at a spacing of 4 × 5 m at a forest farm near Xuchou City, Jiangsu Province, China. The total stem heights and diameters were measured at the end of each of the growing seasons for each tree. Two parent-specific genetic linkage maps each composed of 19 linkage groups (roughly representing 19 haploid chromosomes in poplar) were constructed from restriction fragment length polymorphism, amplified fragment length polymorphism, and microsatellite markers for this hybrid progeny (Yin *et al.* 2002) and used for the genetic mapping of QTL affecting complex traits of economical importance in forest trees.

Methods: Yin *et al.* (2002) used Grattapaglia and Sederoff's (1994) pseudo-test backcross strategy to construct two linkage maps each corresponding to a parent. Each testcross marker for these two parent-specific maps is heterozygous in one parent and null in the other. Because the two parents, I-69 and I-45, are heterozygous, there is no consistent linkage phase among dominant alleles of different markers on the same linkage group; some are in a coupling phase whereas others are in a repulsion linkage phase (Yin *et al.* 2002). Thus, unlike QTL mapping in inbred-line crosses, we need to determine the correct linkage phase between the QTL and the markers flanking it for the pseudo-test backcross mapping population (Lin *et al.* 2003). Our functional mapping proposed above was modified to incorporate the uncertainty of the QTL-marker linkage phase into the likelihood function (appendix).

Detection of QTL with significant epistasis: The statistical model built upon a universal logistic growth law (West *et al.* 2001) is used to map epistatic QTL responsible for growth trajectories in poplars. All of the 19 linkage groups were scanned on a 2-cM scale for the existence of a pair of QTL at different genomic locations. We have successfully detected a few pairs of genomic locations at which two QTL interact to affect stem growth trajectories in poplar. Table 1 and Figure 1 illustrate an example in which a pair of QTL located at the ninth interval [TC/CTG-730, TC/CTG-735] and the thirteenth interval [AG/CTT-570, AG/CTT-595] of linkage group D16 (Yin *et al.* 2002) were found to show significant epistatic effects on stem height growth. The uppercase alleles of these two QTL were observed to be in a coupling phase with dominant alleles of their respective flanking coupling markers. The maximum (93.1) of the landscape of the log-LR test statistics across the linkage group (Figure 1), greater than the genome-wide threshold at the significance level $\alpha = 0.05$ ($LR_T = 85.6$) estimated from permutation tests, justifies the adequacy of a two-QTL model incorporating growth curves.

To show the advantages of our functional mapping model, the same data set was analyzed by traditional univariate and multivariate interval mapping approaches. Univariate interval mapping applied to the most differentiated heights at the oldest age (year 11) measured detected no QTL, whereas multivariate interval mapping for three representative ages (years 1, 6, and 11) suggested a marginal QTL at the significance level $\alpha = 0.10$ (results not shown). These results indicate that our functional mapping approach is statistically more powerful for detecting QTL from a given mapping population.

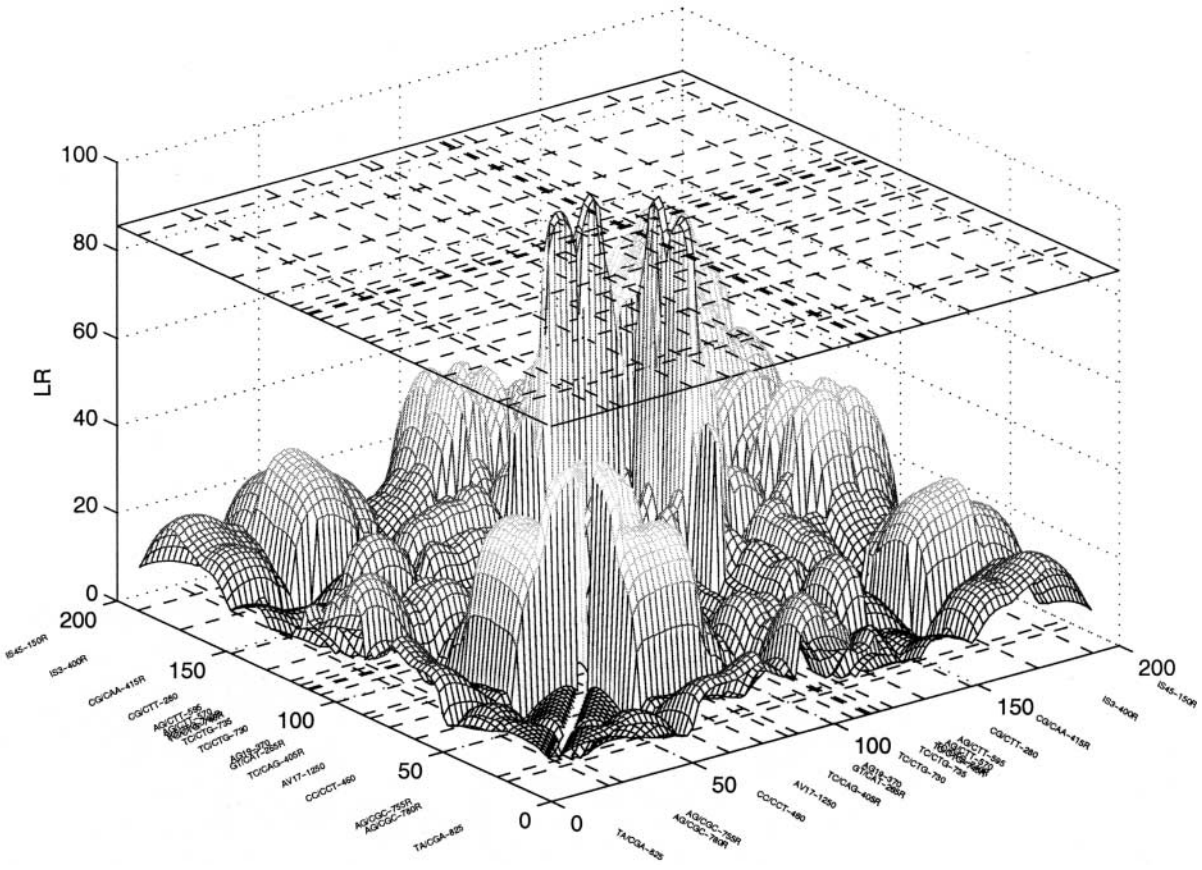

FIGURE 1.—The landscape of the log-likelihood ratio for two epistatic QTL across the same linkage group D16 containing 18 markers in poplar (YIN *et al.* 2002). Two QTL, one located on the ninth marker interval [TC/CTG-730, TC/CTG-735] and the other located on the thirteenth interval [AG/CTT-570, AG/CTT-595], were detected to epistatically determine the overall shape of stem height growth. The plot denotes the critical threshold for the existence of significant QTL. The peak of the landscape corresponding to the genomic locations of two QTL is indicated.

The pseudo-test backcross used here allows for only the significance tests of the additive effects of the two QTL and their additive × additive epistasis. The thresholds at the $\alpha = 0.05$ level for these tests were calculated by simulation studies with the restrictions 13 and 17, respectively. By comparing the maximum value of the LRs from the functional mapping approach with these thresholds, we found that the additive × additive epistasis has a significant impact on the overall differences of growth curve shapes in stem height growth, whereas these two QTL each display marginally significant effects.

To address a possible violation of the constant variance assumption in the matrix (8), we incorporate the TBS model (CARROLL and RUPPERT 1984) into our functional mapping framework. Similar results about the estimation of QTL positions and effects were obtained from the TBS-based mapping approach (data not shown). Yet, the TBS-based mapping approach provides more precise estimates of growth curve parameters, with sampling errors reduced by 20–50% compared to those from untransformed data.

The dynamic pattern of QTL effect: The growth curves of height are drawn using the estimates of logistic parameters (Table 1) for four genotypes at two interactive QTL located on linkage group D16 (Figure 2). As described in HYPOTHESIS TESTS, our functional mapping approach can be used to test various genetic hypotheses related to the developmental process on the basis of estimated growth parameters. The additive effect of the QTL located on the thirteenth marker interval is significant throughout the entire growth process measured, but its sign is altered when trees develop into age 7–8 years (Figure 2). The QTL located on the eighth marker interval has nonsignificant additive effect on growth, but it interacts significantly with the QTL on the thirteenth interval. It is not surprising that significant QTL × age interactions are detected on height growth given the change of the signs of the additive and additive × additive effects (Figure 2).

Genetic control over the inflection point: Equation 3 describes the coordinates of the inflection point where the exponential phase ends and the asymptotic phase begins (NIKLAS 1994). The difference in the coordinates between different genotypes provides important information about the genetics and evolution of growth trajectories. If different growth curves predicted by a QTL

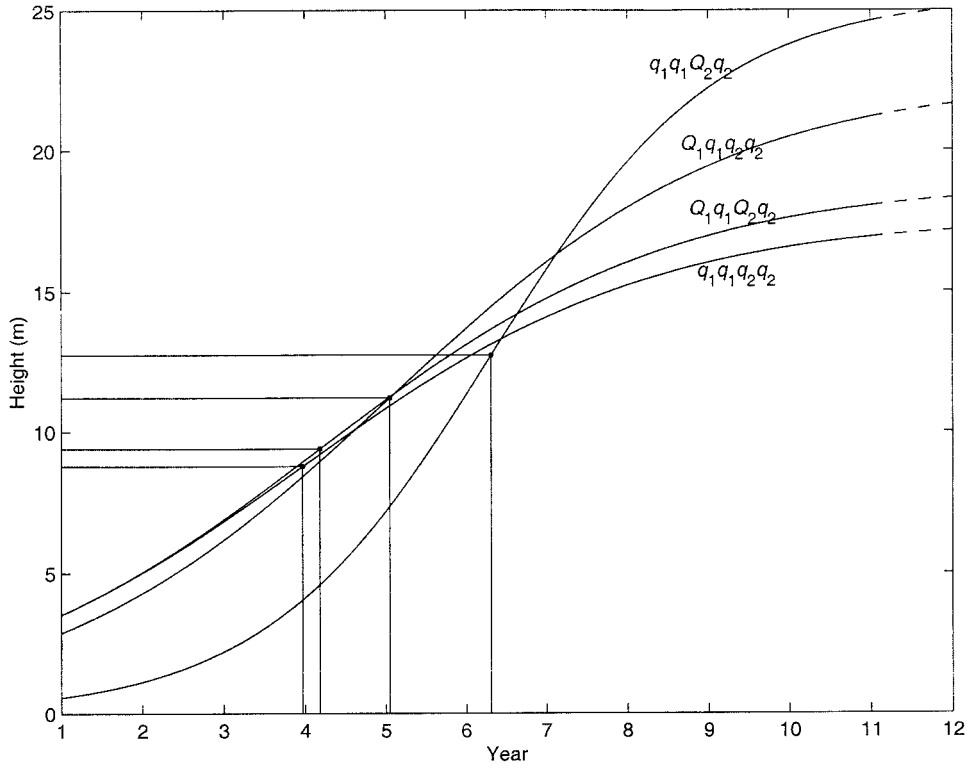

FIGURE 2.—The growth curves of four different QTL genotypes drawn using parameter sets (*a*, *b*, *c*) in Table 1 for two QTL detected on the same linkage group D16. The coordinates of the inflection point for each curve are indicated by the horizontal and vertical lines. The differentiation pattern of growth curves beyond the maximum observed age (11 years), affected by the QTL, is represented by extended broken curves.

have different ages at the inflection point, this indicates that the inflection point is under genetic determination. In our example of poplars, significant additive effect due to the QTL on the thirteenth marker interval and significant additive × additive effect on the timing of the inflection point are detected. The additive effect of the QTL on the thirteenth interval delays the occurrence of the inflection point by about 0.4 year, whereas the additive × additive effect causes the inflection point to occur 0.8 year earlier (Figure 2). Because the inflection point occurs at a time of maximum growth rate, the genetic control of growth trajectory implies that it can be genetically modified to increase a tree's capacity to effectively acquire spatial resources.

DISCUSSION

Increasing evidence has emerged for the role of complex genetic architecture in regulating the ontogenetic development of embryological phenotypes (CHEVERUD *et al.* 1983; ATCHLEY 1984; ATCHLEY and ZHU 1997; VAUGHN *et al.* 1999; CARLBORG *et al.* 2003) and, ultimately, shaping the evolutionary process of organismic form (WOLF *et al.* 2000). However, traditional genetic approaches are limited in estimating nonadditive effects. If epistasis is assumed to be absent, as in most quantitative genetic studies, COCKERHAM's (1963) model based on a mating design provides a nice estimate of the dominant variance. WU (1996) extended Cockerham's quantitative genetic model to estimate the epistatic variances on the basis of clonal replicates. If both parents and offspring are cloned, Wu's model can estimate the entire additive × additive epistatic variance, a partial additive × dominant epistatic variance, and a partial dominant × dominant epistatic variance.

In this article, we present a statistical approach for mapping any QTL that exert various genetic effects on growth trajectories based on a genetic linkage map. Our approach is unique in that it detects and estimates genetic effects due to allelic/nonallelic actions and interactions of QTL from physiological and developmental principles of growth. This uniqueness makes our approach advantageous in two aspects and leads us to construct a conceptual framework of evolutionary developmental biology (ARTHUR 2002).

First, we integrate growth equations into a statistical mapping framework to map developmental QTL that guide the trajectories of organ growth and development. Separate QTL analyses of growth at different time points are not powerful to follow the dynamics of QTL effects since relationships of growth at different times are not considered. Multivariate QTL analysis combining all time points takes into account these age-dependent relationships (KOROL *et al.* 2001), but it quickly becomes intractable when the number of time points increases. By fitting the expected genetic values at different time points by growth curves (WEST *et al.* 2001) and the residual (co)variance matrix by the AR(1) model (DAVIDIAN and GILTINAN 1995), our approach estimates a considerably reduced number of parameters,

which thus increases significantly the power to detect epistasis. Second, because biological principles are incorporated, our approach sheds better light on the integration of development and epitasis. Our approach allows for the understanding of the genetic basis for growth and development at the cutting edge of biology (Raff 2000; Arthur 2002). Growth and development are two different but related biological processes. It is likely that these two processes share some common genetic mechanisms. By mapping the timing of various developmental events, *e.g.*, the time to first flower, on the growth curves of QTL genotypes, we can test whether the same growth genes are also involved in regulating reproductive behaviors.

The estimation precision of additive, dominant, and epistatic effects on growth at particular time points, at time intervals, or during the entire growth process depends on the estimation precision of the parameters describing growth curves. Our earlier studies using a real example have demonstrated that growth parameters of different QTL genotypes can be precisely estimated (Ma *et al.* 2002; Wu *et al.* 2002, 2004). The sampling errors of the growth parameters estimated from Fisher's information index are less than one-tenth of the parameter values when a modest sample size (90) was used. Moreover, these sampling errors can be reduced if functional mapping is based on the transform-both-sides model, as advocated by Carroll and Ruppert (1984), for relaxing the variance stationarity assumption in the AR(1) model. The estimation precision of parameters in the real example is consistent with that from our simulation studies. It is inferred from these results that the estimation of the genetic architecture of a quantitative trait from our functional mapping approach should be adequately precise, although direct assessments of the sampling errors for the gene effect estimation from our approach are needed. In addition, it is possible to modify our analysis model to consider nonstationary variance-covariance structures using structured antedependent models, as proposed by Nunez-Anton (1997) and Nunez-Anton and Zimmerman (2000). Such a modification can be viewed as an alternative to the current method based on the variance-stationary assumption.

Genetic effects are expressed differentially during ontogeny (Rice 2000; Wolf *et al.* 2000), as also documented in our case study of poplar trees. Epistasis is regarded as a force to determine the direction of organismic evolution. This controversial view can now be tested and assessed, using this model and implementing the estimation process of epistasis. From an evolutionary perspective, however, our approach should be modified to consider population genetic properties of QTL alleles, QTL genotypes, and the relationship between the QTL and markers. If a mapping population is constructed using a natural population, the theory of linkage disequilibrium mapping should be integrated within the functional mapping framework. Linkage disequilibrium-based mapping provides a powerful tool for fine-scale mapping of complex traits (Lou *et al.* 2003) and, thus, the combination of this mapping strategy with our functional mapping can gain better insights into the genetic basis of development and evolution.

We thank two anonymous referees for their constructive comments on this manuscript. This work is partially supported by an Outstanding Young Investigators Award of the National Science Foundation of China (30128017), a University of Florida Research Opportunity Fund (02050259), and a University of South Florida Biodefense grant (7222061-12) to R.W. The publication of this manuscript is approved as journal series no. R-09586 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.

Arthur, W., 2002 The emerging conceptual framework of evolutionary developmental biology. Nature **415:** 757–764.

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.

Bertalanffy, V. L., 1957 Quantitative laws for metabolism and growth. Q. Rev. Biol. **32:** 217–231.

Carlborg, O., S. Kerje, K. Schutz, L. Jacobsson, P. Jensen *et al.*, 2003 A global search reveals epistatic interaction between QTL for early growth in the chicken. Genome Res. **13:** 413–421.

Carroll, R. J., and D. Ruppert, 1984 Power-transformations when fitting theoretical models to data. J. Am. Stat. Assoc. **79:** 321–328.

Cheverud, J. M., and E. J. Routman, 1995 Epistasis and its contribution to genetic variance components. Genetics **139:** 1455–1461.

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.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

Cockerham, C. C., 1963 Estimation of genetic variances, pp. 53–94 in *Statistical Genetics and Plant Breeding*, edited by W. D. Hanson and H. F. Robinson. Pub. 982, National Academy of Science-National Research Council, Washington, DC.

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.

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

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.

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.

Lin, M., X.-Y. Lou, M. Chang and R. L. Wu, 2003 A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. Genetics **165:** 901–913.

Lou, X.-Y., G. Casella, R. C. Littell, M. C. K. Yang, J. A. Johnson *et al.*, 2003 A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis. Genetics **163:** 1533–1548.

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

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

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.

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

Raff, R. A., 2000 Evo-devo: the evolution of a new discipline. Nat. Rev. Genet. **1:** 74–79.

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.

Rice, S. H., 2000 The evolution of developmental interactions: epistasis, canalization and integration, pp. 82–98 in *Epistasis and the Evolutionary Process*, edited by J. B. Wolf, E. D. Brodie III and M. J. Wade. Oxford University Press, London/New York/Oxford.

Rougvie, A. E., 2001 Control of developmental timing in animals. Nat. Rev. Genet. **2:** 690–701.

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.

West, G. B., J. H. Brown and B. J. Enquist, 2001 A general model for ontogenetic growth. Nature **413:** 628–631.

Wolf, J. B., E. D. Brodie, III and M. J. Wade, 2000 *Epistasis and the Evolutionary Process.* Oxford University Press, London/New York/Oxford.

Wu, R. L., 1996 Detecting epistatic genetic variance with a clonally replicated design: models for low- vs. high-order nonallelic interaction. Theor. Appl. Genet. **93:** 102–109.

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., C. X. Ma, M. Chang, R. C. Littell, S. S. Wu *et al.*, 2002 A logistic mixture model for detecting major genes governing growth trajectories. Genet. Res. **79:** 235–245.

Wu, R. L., C. X. Ma, M. C. K. Yang, M. Chang, U. Santra *et al.*, 2003a Quantitative trait loci for growth in Populus. Genet. Res. **81:** 51–64.

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

Wu, R. L., C. X. Ma, M. Lin, Z. H. Wang and G. Casella, 2004 Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. Biometrics (in press).

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.

Communicating editor: M. W. Feldman

APPENDIX: ESTIMATION OF QTL PARAMETERS IN A PHASE-UNKNOWN PSEUDO-TEST BACKCROSS

Suppose an interval is flanked by two markers, $\mathcal{M}_u$ and $\mathcal{M}_{u+1}$, whose dominant nonalleles are in a coupling phase. A putative QTL, $\mathcal{Q}_k$, on this interval may have different phases relative to the marker alleles. Let p_k be the probability of an uppercase QTL allele (Q_k) to be in a coupling phase with the dominant marker alleles. Thus, $1 - p_k$ is the probability of Q_k to be in a repulsion phase with the dominant marker alleles. A second QTL, Q_l, that can be located on different marker intervals N_v and N_{v+1}, or on the same marker intervals M_u and M_{u+1}, also has two different QTL-marker phase probabilities, p_l and $1 - p_l$. The combination of the two QTL leads to four possible (unknown) phase combinations with probabilities p_kp_l, $p_k(1 - p_l)$, $(1 - p_k)p_l$, and $(1 - p_k)(1 - p_l)$ if the allelic assignment of one QTL is independent of that of the second QTL. The estimation of unknown parameters should be based on a most likely phase combination (determined by p_k and p_l) inferred from the data. The likelihood functions for the four phase combinations are denoted by L_{11}, L_{12}, L_{21}, and L_{22}, respectively.

As shown in Lin *et al.* (2003), the mixture-likelihood function of these four phase combinations can be written as

$$L(\boldsymbol{\Omega}, p_k, p_l) = p_kp_lL_{11}(\boldsymbol{\Omega}) + p_k(1 - p_l)L_{12}(\boldsymbol{\Omega}) + (1 - p_k)p_lL_{21}(\boldsymbol{\Omega}) + (1 - p_k)(1 - p_l)L_{22}(\boldsymbol{\Omega}),$$

where

$$L_{11}(\boldsymbol{\Omega}) = \prod_{i=1}^{n}\sum_{j_k=0}^{1}\sum_{j_l=0}^{1}[\pi^{11}_{ij_kj_l}f_{ij_kj_l}f(\mathbf{y})],$$

$$L_{12}(\boldsymbol{\Omega}) = \prod_{i=1}^{n}\sum_{j_k=0}^{1}\sum_{j_l=0}^{1}[\pi^{12}_{ij_kj_l}f_{ij_kj_l}f(\mathbf{y})],$$

$$L_{21}(\boldsymbol{\Omega}) = \prod_{i=1}^{n}\sum_{j_k=0}^{1}\sum_{j_l=0}^{1}[\pi^{21}_{ij_kj_l}f_{ij_kj_l}f(\mathbf{y})],$$

$$L_{22}(\boldsymbol{\Omega}) = \prod_{i=1}^{n}\sum_{j_k=0}^{1}\sum_{j_l=0}^{1}[\pi^{22}_{ij_kj_l}f_{ij_kj_l}f(\mathbf{y})].$$

The probabilities ($\pi_{ij_kj_l}$) of two-QTL genotypes, conditional upon marker genotypes for progeny i, are derived for different phase combinations for the two QTL located on different marker intervals (Table A1) or for the two QTL on the same interval (Table A2). The EM algorithm has been implemented to obtain the MLEs of vector $\boldsymbol{\Omega}$ and the phase probabilities of the two QTL p_k and p_l. In Tables A1 and A2, the conditional probabilities are given for the QTL genotypes bracketed by two coupling dominant markers or repulsion dominant markers.

TABLE A1

Conditional probabilities of QTL genotypes given the marker genotypes for $\mathcal{M}_u$ and $\mathcal{M}_{u+1}$ under different marker-QTL linkage phases

	Coupling markers				Repulsion markers			
	Marker-QTL phase 1		Marker-QTL phase 2		Marker-QTL phase 1		Marker-QTL phase 2	
Marker genotype	*Qq*	*qq*	*Qq*	*qq*	*Qq*	*qq*	*Qq*	*qq*
$M_u m_u M_{u+1} m_{u+1}$	1	0	0	1	$1-\theta$	θ	θ	$1-\theta$
$M_u m_u m_{u+1} m_{u+1}$	$1-\theta$	θ	θ	$1-\theta$	1	0	0	1
$m_u m_u M_{u+1} m_{u+1}$	θ	$1-\theta$	$1-\theta$	θ	0	1	1	0
$m_u m_u m_{u+1} m_{u+1}$	0	1	1	0	θ	$1-\theta$	$1-\theta$	θ

$\theta = r_{k_1}/r_u$, where r_u is the recombination fraction between two markers $\mathcal{M}_u$ and $\mathcal{M}_{u+1}$ and r_{k_1} is the recombination fraction between the marker $\mathcal{M}_u$ and the QTL $\mathcal{Q}_k$. Double crossover is ignored.

TABLE A2

Conditional probabilities of the QTL genotypes of two QTL, $\mathcal{Q}_k$ and $\mathcal{Q}_{k+1}$, bracketed by two markers, $\mathcal{M}_u$ and $\mathcal{M}_{u+1}$, conditional on the marker genotypes under different marker-QTL linkage phases

Marker-QTL phase	Marker genotype	$Q_k q_k Q_{k+1} q_{k+1}$	$Q_k q_k q_{k+1} q_{k+1}$	$q_k q_k Q_{k+1} q_{k+1}$	$q_k q_k q_{k+1} q_{k+1}$
		Two coupling markers			
Combination 11	$M_u m_u M_{u+1} m_{u+1}$	1	0	0	0
	$M_u m_u m_{u+1} m_{u+1}$	$1-\theta_1-\theta_2$	θ_2	0	θ_1
	$m_u m_u M_{u+1} m_{u+1}$	θ_1	0	θ_2	$1-\theta_1-\theta_2$
	$m_u m_u m_{u+1} m_{u+1}$	0	0	0	1
Combination 12	$M_u m_u M_{u+1} m_{u+1}$	1	1	0	0
	$M_u m_u m_{u+1} m_{u+1}$	θ_2	$1-\theta_1-\theta_2$	θ_1	0
	$m_u m_u M_{u+1} m_{u+1}$	0	θ_1	$1-\theta_1-\theta_2$	θ_2
	$m_u m_u m_{u+1} m_{u+1}$	0	0	0	0
Combination 21	$M_u m_u M_{u+1} m_{u+1}$	0	0	1	0
	$M_u m_u m_{u+1} m_{u+1}$	θ_2	θ_1	$1-\theta_1-\theta_2$	0
	$m_u m_u M_{u+1} m_{u+1}$	0	$1-\theta_1-\theta_2$	θ_1	θ_2
	$m_u m_u m_{u+1} m_{u+1}$	0	1	0	0
Combination 22	$M_u m_u M_{u+1} m_{u+1}$	0	0	0	1
	$M_u m_u m_{u+1} m_{u+1}$	θ_1	0	θ_2	$1-\theta_1-\theta_2$
	$m_u m_u M_{u+1} m_{u+1}$	$1-\theta_1-\theta_2$	θ_2	0	θ_1
	$m_u m_u m_{u+1} m_{u+1}$	1	0	0	0
		Two repulsion markers			
Combination 11	$M_u m_u M_{u+1} m_{u+1}$	$1-\theta_1-\theta_2$	θ_2	0	θ_1
	$M_u m_u m_{u+1} m_{u+1}$	1	0	0	0
	$m_u m_u M_{u+1} m_{u+1}$	0	0	0	1
	$m_u m_u m_{u+1} m_{u+1}$	θ_1	0	θ_2	$1-\theta_1-\theta_2$
Combination 12	$M_u m_u M_{u+1} m_{u+1}$	θ_2	$1-\theta_1-\theta_2$	θ_1	0
	$M_u m_u m_{u+1} m_{u+1}$	0	1	0	0
	$m_u m_u M_{u+1} m_{u+1}$	0	0	1	0
	$m_u m_u m_{u+1} m_{u+1}$	0	θ_1	$1-\theta_1-\theta_2$	θ_2
Combination 21	$M_u m_u M_{u+1} m_{u+1}$	θ_2	θ_1	$1-\theta_1-\theta_2$	0
	$M_u m_u m_{u+1} m_{u+1}$	0	0	1	0
	$m_u m_u M_{u+1} m_{u+1}$	0	1	0	0
	$m_u m_u m_{u+1} m_{u+1}$	0	$1-\theta_1-\theta_2$	θ_1	θ_2
Combination 22	$M_u m_u M_{u+1} m_{u+1}$	θ_1	0	θ_2	$1-\theta_1-\theta_2$
	$M_u m_u m_{u+1} m_{u+1}$	0	0	0	1
	$m_u m_u M_{u+1} m_{u+1}$	1	0	0	0
	$m_u m_u m_{u+1} m_{u+1}$	$1-\theta_1-\theta_2$	θ_2	0	θ_1

$\theta 1 = r_{k_1}/r_u$, $\theta_2 = r_{k_2}/r$, where r_u, r_{k_1}, and r_{k_2} are the recombination fractions between markers $\mathcal{M}_u$ and $\mathcal{M}_{u+1}$, marker $\mathcal{M}_u$ and QTL $\mathcal{Q}_k$, and QTL $\mathcal{Q}_k$ and $\mathcal{Q}_{k+1}$, respectively. Double crossover is ignored.