Functional mapping for quantitative trait loci governing growth rates: a parametric model

Rongling Wu,1,2 Chang-Xing Ma,1 Wei Zhao,1 and George Casella1
1Department of Statistics, University of Florida, Gainesville, Florida 32611; and 2Laboratory of Statistical Genetics, Zhejiang Forestry College, Lin'an, Zhejiang 311300 PR, China

Submitted 3 February 2003; accepted in final form 25 June 2003

GROWTH IS A PHYSIOLOGICAL process by which the organism assimilates or transforms essential, nonliving nutrients into living protoplasm. Every growth process includes two different components, hypertrophy (increase in cell size) and hyperplasia (increase in cell number) (30). Knowledge about growth is always important in understanding the origin of life and, ultimately, altering growth trajectories beneficial to humans. Various physiological mechanisms underlying a growth process, such as hormonal, nutritional, or environmental, have been proposed, and now have become a fundamental topic covered in general physiology textbooks. Growth as a dynamic, complex phenotype is also under genetic control, as is well documented in a considerable body of literature on quantitative genetic research (1, 15–17, 27, 37). The precise details of the individual’s growth are influenced by the particular variants of the genes, or alleles, that comprise its own genetic constitution. However, the relationships between the details of observable growth form of the individual and its underlying genetic constitution are never simple and obvious (36).

The recent molecular marker technologies have provided a powerful means for unraveling the genetic architecture of a quantitative trait including growth (22, 23). Using genetic linkage maps constructed from polymorphic markers, one can identify individual loci or quantitative trait loci (QTL) responsible for growth differentiation and then be map these on the genome (4, 6, 31, 39). These empirical studies of growth measured at discrete time points, however, have two limits. First, they did not treat growth as a continuous process (as it should be), and therefore, the QTL identified cannot reveal an entire picture of genetic control over ontogenetic growth. Second, some of these studies attempted to map QTL affecting body gain of mouse in a particular time interval, or “growth rate” (20), but growth rates may vary during different stages of growth and also growth rates at different ages are not completely independent of each other (1, 2). It is crucial to map QTL that underlie age-specific growth rates by considering their dependence during the entire process of growth.

To overcome the first limit, we proposed a parametric model for incorporating growth curve equations into a mapping approach, which can detect the dynamic QTL affecting an entire growth process (21, 38). Using this so-called “functional mapping” model, one can determine when a responsible QTL starts or ceases its effect on growth during an individual’s ontogeny. Also, based on an example of forest trees, our model has proven more powerful to detect significant QTL than traditional interval mapping originally proposed by Lander and Botstein (19). In this article, we intend to overcome the second limit by extending our functional mapping model to map QTL affecting the overall growth rate in organisms. This extended model is based on the mechanistic argument for the relationship between growth rates and ages that is described by a parametric function. The QTL parameters determining differences in overall growth rate are estimated using a maximum likelihood-based method, implemented with the EM algorithm. Using this new model, we successfully detect a major QTL affecting age-specific growth rates of stem diameters in a forest tree. The implications and extensions of this model for un-
understanding the genetic basis of development are discussed.

PARAMETRIC MODELING OF GROWTH RATES

According to one of the pioneering mathematical biologists, Ludwig von Bertalanffy (3), growth in weight, length, or size will occur whenever the anabolic or metabolic rate exceeds the rate of catabolism. Thus the ratio of these two processes indicates the occurrence and change of growth. When the ratio approaches and then drops below unity, growth will decrease and eventually cease. Bertalanffy also noted that the metabolic rate of an animal (in fact, any organism, as proven by subsequent researchers; Refs. 34–36) scales as the kth power of its weight, but the catabolic rate is proportional to the weight itself. Therefore, the growth rate, i.e., the difference between the metabolic and catabolic rates becomes

\[
\frac{dG}{dt} = \eta G^k - \kappa G
\]

where \( G \) represents the growth at time \( t \), and \( \eta \) and \( \kappa \) are the constants of metabolism and catabolism, respectively. For small values of \( k \), integration of Eq. 1 leads to the growth equation

\[
G = \left(\frac{\eta}{\kappa} - \frac{\eta}{\kappa} G_0^{1-k} \right) e^{-\left(1-k\right)\frac{\ln \left(b^k\right)}{k}}
\]

where \( G_0 \) is the growth at \( t = 0 \). This growth function is sigmoidal (S-shaped), approaching asymptotically the value \( \left(\frac{\eta}{\kappa}\right)^{\frac{1}{k(1-k)}} \) as \( t \to \infty \). Such an S-shaped pattern of growth includes an exponential growth stage, an asymptotic growth stage, and the point of inflection at which these two stages are connected (25). At the point of inflection, an organism displays maximum growth per unit time. After the substitutions

\[
\begin{align*}
  a^{1-k} & = \frac{\eta}{\kappa} \\
  b & = \frac{\eta}{\kappa} G_0^{-k} \\
  r & = (1 - k)\kappa
\end{align*}
\]

where \( a \) is the limiting growth, \( b \) is a parameter related to the growth at \( t = 0 \), and \( r \) is the “rate constant” that determines the spread of the curve along the time axis, Richards (28) rewrites Eq. 2 as

\[
G = a\left(1 + be^{-rt}\right)^{\frac{1}{1-k}}
\]

Equation 3, referred to as the Richards growth model, can be reduced to three well-known growth functions

\[
G = \begin{cases} 
  a(1 - be^{-rt}) & \text{Monomolecular curve when } k = 0 \\
  ae^{-be^{-rt}} & \text{Gompertz curve when } k = 1 \\
  a & \text{Autocatalytic or logistic curve when } k = 2 \\
  \frac{1}{1 + be^{-rt}} & 
\end{cases}
\]

The monomolecular curve actually describes the asymptotic phase of the logistic growth curve. The logistic curve is symmetric about the inflection point at which \( t = (\ln b)/r \) and \( G = a/2 \). Yet, the Gompertz curve is not symmetric, with the inflection point occurring at \( t = (\ln b)/r \) and \( g = a/2 \). These three curves, with different flexibilities, may best fit growth data collected from a different species, organ, or in a different environment.

Growth rate curves. The change in \( G \) per change in time is the absolute rate of growth as defined in Eq. 1. With substitutions (28), the general forms of absolute growth rate \( (H) \) and relative growth rate \( (G) \) can be now written, respectively, as

\[
H = \frac{dG}{dt} = rG - \frac{G}{1-k} \left[ \left(\frac{\eta}{G}\right)^{1-k} - 1 \right] \tag{5}
\]

and

\[
R = \frac{1}{G} \frac{dG}{dt} = \frac{r}{1-k} \left[ \left(\frac{\eta}{G}\right)^{1-k} - 1 \right] \tag{6}
\]

The special forms of absolute growth rate and relative growth rate are expressed as

\[
\begin{align*}
  H &= r(a - G) & \text{Monomolecular curve when } k = 0 \\
  H &= rG \ln \left(\frac{a}{G}\right) & \text{Gompertz curve when } k = 1 \\
  H &= rG \left(1 - \frac{G}{a}\right) & \text{Autocatalytic or logistic curve when } k = 2
\end{align*}
\]

(7)

and

\[
\begin{align*}
  R &= r\left(\frac{a}{G} - 1\right) & \text{Monomolecular curve when } k = 0 \\
  R &= r \ln \left(\frac{a}{G}\right) & \text{Gompertz curve when } k = 1 \\
  R &= r \left(1 - \frac{G}{a}\right) & \text{Autocatalytic or logistic curve when } k = 2
\end{align*}
\]

(8)

The absolute and relative growth rates described in Eqs. 7 and 8 are a function of time \( t \), and thus the QTL underlying age-specific rates can be mapped within the framework constructed in Refs. 21 and 38.

THE STATISTICAL METHOD

Consider a backcross in which there are two groups of genotypes at a locus. A marker-based genetic linkage map is constructed using this backcross, aimed at the identification of QTL affecting a quantitative trait. Consider a trait, such as body size or body weight, which can be modeled by one of the three growth curves in Eq. 4 based on an infinite number of measurements during growth trajectories. Assume that a pleiotropic biallelic QTL affecting age-specific growth rates is segregating in the backcross population. This
QTL is bracketed by two flanking markers $M_i$ and $M_{i+1}$, each with possible genotypes $M_im_i$ and $m_im_m$ as well as $M_{i+1}m_{i+1}$ and $m_{i+1}m_{i+1}$. For a particular genotype $j$ ($j = 1$ for $Qq$ or 0 for $qq$) at this QTL, the parameters describing its growth rate 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 rate curves.

Assume that all $n$ progeny in the backcross are measured at each of $(m + 1)$ time points. The change of the trait phenotype ($y_i$) of progeny $i$ over two successive time points $[t_i, (t_i + 1)]$, i.e., the absolute growth rate (AGR), is calculated as $\Delta y_i(t) = y_i(t + 1) - y_i(t)$. The corresponding relative growth rate (RGR) is calculated as $z_i(t) = \Delta y_i(t) / y_i(t)$. Below, the description of our model will be based on AGR unless it should be specified. The value of $\Delta y_i(t)$ due to the QTL located on an interval flanked by markers $M_i$ and $M_{i+1}$ can be expressed by a linear statistical model

$$\Delta y_i(t) = \sum_{j=0}^{1} x_{ij}H_j(t) + e_i(t)$$

(9)

where $x_{ij}$ 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, $H_j(t)$ is the growth rate of the $j$th QTL genotype 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 $N(0, \sigma^2(t))$. The probability with which $x_{ij}$ takes 1 or 0 depends on the two-locus genotype of the flanking markers $M_i$ and $M_{i+1}$ and the position of the QTL on the marker interval. The probability of a QTL genotype ($Qq$ or $qq$) conditional upon the four genotypes of the flanking markers ($M_iM_iM_{i+1}m_{i+1}$, $M_im_iM_{i+1}m_{i+1}$, $m_im_im_{i+1}m_{i+1}$, and $m_im_mM_{i+1}m_{i+1}$) for progeny $i$ in a backcross population was expressed as

$$\pi_{ij} = \text{Prob}(x_{ij} = j, l + 1, 0)$$

where $\theta$ is the ratio of the recombination fractions between marker $M_i$ and the QTL over between the two markers.

The growth rates for individual $i$ at all time points $1, 2, \ldots, m + 1$ for each QTL genotype group, $\Delta y_{ij} = [\Delta y_{ij}(1), \Delta y_{ij}(2), \ldots, \Delta y_{ij}(m)]$, follow a multivariate normal density

$$f(\Delta y_{ij}) = \frac{1}{(2\pi)^{m/2}v^1/2} \exp\left\{-\frac{1}{2}(\Delta y_{ij} - H_j)^T\Sigma^{-1}(\Delta y_{ij} - H_j)\right\}$$

(10)

where $H_j$ is the vector of the expected genotypic values of AGR measured at $m$ time intervals, and $\Sigma$ is the residual variance-covariance matrix of $\Delta y_{ij}$. Indeed, $H_j$ can be modeled by the growth rate curve of Eq. 5 as

$$H_j = [H_j(t)]_{1 \times m} = \left[\begin{array}{c}r_jG_j(t) \left(\frac{a_j}{1 - k_j} - 1\right) \end{array}\right]_{1 \times m}$$

and $\Sigma$ can be assumed identical between the two QTL genotypes and modeled using the AR(1) model [Davidian and Giltinan (8); Verbeke and Molenberghs (32)] as

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho^{m-1} \\ \rho & 1 & \cdots & \rho^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \cdots & 1 \end{bmatrix}$$

The matrix $\Sigma$ of Eq. 10 assumes variance stationarity, i.e., there is the same residual variance ($\sigma^2$) for the growth rate at each time, and covariance stationarity, i.e., the covariance of growth rate between different times decreases proportionally (in $\rho$) with increased time interval (27). These two assumptions can be relaxed using an approach proposed in Ref. 26.

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

$$L(\Omega) = \prod_{i=1}^{N} \sum_{j=1}^{2} \pi_{ij}f(\Delta y_{ij})$$

(11)

where the vector $\Omega = (a_j, b_j, r_j, k_j, \theta, \rho, \sigma^2)^T$ contains 11 unknown parameters to be estimated for the QTL effect, QTL position, and residual (co)variances.

The maximum likelihood estimates (MLEs) of the unknown parameters under the pleiotropic model can be computed by implementing the EM algorithm (9, 19, 21, 38). In the ($\tau + 1$)th iteration, the E-step (i.e., expectation) calculates the posterior probabilities of progeny $i$ having a QTL genotype $j$ as

$$\Pi_{ij}^{(\tau+1)} = \frac{\pi_{ij}f(\Delta y_{ij})^{(\tau+1)}}{\sum_{j=0}^{1} \pi_{ij}f(\Delta y_{ij})^{(\tau+1)}}$$

(12)

In the M-step (i.e., maximization), parameters are estimated from the score equations obtained by differentiating the log-likelihood of Eq. 11 with respect to each unknown and letting the derivatives equal zero. The E and M steps are iterated until a convergence criterion is satisfied. The converged values are the MLEs. The standard errors of the MLEs are estimated using the inverse of the Fisher information matrix. The derivation of this matrix needs the second derivatives of the log-likelihood of Eq. 11 with respect to the unknown parameters.

**HYPOTHESIS TESTS**

A number of biologically meaningful hypotheses can be tested based on the growth rate-based genetic model. The hypothesis about the existence of a QTL affecting an overall growth rate curve can be formulated as

$$H_0: a_1 = a_0, b_1 = b_0, r_1 = r_0, k_1 = k_0$$

$$H_1: \text{at least one of the equalities above does not hold}$$

where $H_0$ corresponds to the reduced model, in which the data can be fit by a single growth rate curve, and $H_1$ corresponds to the full model, in which there exist two different rate curves to fit the data. The test statistic for testing the hypotheses of Eq. 13 is calculated as the log-likelihood ratio (LR) of the full model over the reduced model.
Age-specific weights can be determined on the basis of proportionality to the actual rate, i.e., \( w(t) = \frac{dG}{dt} \) itself (28). With this we have the weighted growth rate mean as

\[
\int_{-\infty}^{\infty} \left( \frac{dG}{dt} \right)^2 dt \int_{-\infty}^{\infty} \frac{dG}{dt} dt = \frac{1}{a} \int_{0}^{\infty} \frac{dG}{dt} dG = \frac{ar}{2k + 2} \quad (17)
\]

Similarly, a weighted mean relative growth rate may be determined exactly as was the corresponding absolute rate, i.e.,

\[
1_a \int_{0}^{\infty} \left( \frac{1}{G} \cdot \frac{dG}{dt} \right) dG = \frac{1}{a} \int_{0}^{\infty} \frac{r}{1-k} \left[ \frac{a}{G} \right]^{1-k} - 1 \right] dG = \frac{r}{k}
\]

To test for the genetic difference of the mean absolute or relative growth rate conditioned by the QTL, we impose the restrictions

\[
\frac{a_r r_f}{2k_1 + 2} = \frac{a_r r_0}{2k_0 + 2}
\]

and

\[
\frac{r_1}{k_1} = \frac{r_0}{k_0}
\]

respectively. The corresponding tests are performed by comparing the full model and the reduced model with these restrictions.

RESULTS

A case study. An example of a forest tree is used to demonstrate the power of our statistical model for mapping QTL affecting growth rates. The study material, as described in Ref. 21, was derived from the interspecific hybridization of Populus (poplar), P. deltoides and P. euramericana. A total of 450 hybrids were planted at a spacing of 4 × 5 m in the complete randomized design in a field trial near Xuchou City, Jiangsu Province, China. The total stem heights and diameters measured at the end of each of the first 11 growing seasons are used for our QTL analysis.

Genetic linkage maps based on the pseudo-test backcross design were constructed using 90 genotypes randomly selected from the 450 hybrids with random amplified polymorphic DNAs (RAPDs), amplified fraction length polymorphisms (AFLPs), and inter-simple sequence repeats (ISSR) (40). These parent-specific maps comprise the 19 largest linkage groups, which roughly represent 19 pairs of chromosomes. In this article, we use one of the linkage groups from the P. deltoides parent map to map QTL affecting growth rates based on the model described above.

As tested in Ma et al. (21), diameter growth of each of the 90 mapped genotypes followed a different S-shaped growth curve, all belonging to the logistic curve type (\( k = 2 \)). This implies that it is appropriate to build our model upon this nearly universal growth law (35). We calculated the difference of diameter between two successive ages for each tree, which is used as the
estimate of age-specific AGR. These growth rates, plotted against ages, are illustrated in Fig. 1, in which all trees can be statistically fit using a AGR curve. But the differences of AGR curve among different genotypes imply the existence of possible QTL underlying these differences. Our model used to detect such QTL was derived on the basis of the logistic AGR curve of Eq. 7. The EM algorithm was derived to obtain the MLEs of unknown parameters describing the logistic curves ($a_j$, $b_j$, $r_j$) and residual variance and covariance. The log-likelihood equations used to obtain these MLEs in the M step are provided in the APPENDIX.

Using our AGR-based model, we successfully detected a QTL affecting diameter growth rate curves located on linkage group 9, which is composed of 10 markers (Fig. 2). The critical value for claiming the existence of this AGR QTL is determined on the basis of the permutation tests proposed by Doerge and Churchill (11). We used 1,000 permutation tests to obtain the empirical estimate of the critical value as 18.2 and 24.1 at the significance levels $P = 0.05$ and 0.01, respectively. The profile of the log-likelihood ratios (LR) of the full vs. reduced model (Eq. 13) across the interval between markers CG/CCT-620 and CA/CAG-660.

The AGR-based model preserves the advantage of functional mapping proposed by Wu et al. (38) and Ma et al. (21) in detecting the dynamic change of QTL expression over time because the parameters determining the curve shape of a different QTL genotype are estimated. The growth curves of diameter are drawn using the estimates of the logistic parameters for two genotypes at the QTL detected on linkage group 9 (Fig. 3). This QTL is detected to be inactive until trees grew to about 6 yr in the field (Fig. 3A). And its effect on diameter growth increased with ages. At 11 yr old, genotype $Qq$ has diameter growth 4.1 cm more than its counterpart $qq$. This difference appears to increase after age 11 yr, as predicted from the logistic curves estimated.

This QTL not only affects overall diameter growth curves in poplars, but also governs the difference in age-specific AGR (Fig. 3B). In fact, the influence of this QTL on AGR starts at 3 yr old, about 3 yr earlier than its influence on growth curves. After 3 yr, this QTL triggers a consistently large effect on AGR until the last measurement. Using Eq. 17, we further tested for and eventually detected the significant effect of this QTL on the overall growth rate during the entire process of growth. Some physiological and bioenergetic constraints (35) can be used to explain the difference in QTL control mechanisms over growth curves and growth rate curves.

Simulation. The advantage of permutation tests as an approach to determining the critical values (11) is that they automatically provide an empirical estimate of the power to detect a QTL and of the probability of the type I error for QTL detection. But permutation tests do not provide information about the biases and precision of parameter estimates. Here, we perform a simulation study to examine the sampling variances of the MLEs of the logistic parameters fitted. The scale used in this simulation mimics the example assuming 90 progeny, 11 measurement points, and 10 markers as spaced in linkage group D9 (Fig. 2). Two groups of QTL genotypes (and therefore two groups of growth param-

Fig. 1. Plots of stem diameter growth rates vs. ages for each of the 90 genotypes used to construct linkage maps in poplar hybrids [Yin et al. (40)]. The absolute growth rates of these genotypes can be well fit by a particular logistic growth rate curve. The x- and y-axes of the plots denote age (in years) and stem diameter growth rate (in cm), respectively.
eters, $a_j$, $b_j$, $r_j$) are hypothesized. The hypothesized growth parameters are those estimated from the example described above. We carried out 500 simulation replicates to estimate the growth parameters and the residual variance and correlation. The sampling errors of these estimates among the replicates are around or below 10% of their means, thus suggesting that our AGR-based model provides reasonably precise estimates of growth parameters and other parameters.

**DISCUSSION**

In this article, we present a parametric model for mapping QTL affecting age-specific growth rates based on the mechanistic relationship between two different biological processes, metabolism and catabolism (3, 21). This new model can be viewed as being complementary to functional mapping proposed by Wu et al. (38) and Ma et al. (21) aimed at mapping QTL for overall growth trajectories, because the growth curve and growth rate curve are regarded as two related but different developmental features.

In an earlier study, we discussed the advantages of incorporating growth laws into a mapping theory in increasing QTL mapping power and making mapping results more relevant to an understanding of the genetics of development and evolution (21, 38). Our new
model based on growth rate curves preserves these advantages and, on the other hand, possesses two new significant merits. First, by identifying QTL governing growth rates, this new model has direct implications for understanding the genetic control of growth alterations in various environments. Growth rates are frequently associated with an individual’s age, physiological status, hormonal contents, and adaptability to changing environments. The QTL that specifically controls the growth rates at different ages can be better used to understand the genetic architecture of organismal adaptation.

Second, to facilitate the mathematical treatments of the age-specific residual variance-covariance matrix in our earlier model, the simple AR(1) model (Eq. 10) is used, which can provide a general formula for parameter estimation in the M step of the EM algorithm (see the Appendix). Such a structured matrix also has potential in increasing the precision of parameter estimation because of a reduced number of unknown parameters being estimated (7). A possible limitation of the AR(1) model, however, is the assumption of variance stationarity over ages. This assumption is, in fact, violated for any growth data due to the fact that growth differentiation tends to be larger at older ages than younger ages. Unlike growth curves, growth rate curves have less heterogeneities among different ages (data not shown), a favorable condition for deploying the more tractable AR(1) model in our growth rate-based mapping model. Despite this, however, it is still worthwhile structuring the covariance matrix using various statistical methods, e.g., the random coefficient model of Laird and Ware (18) and the nonparametric model of Diggle and Verbyla (10), and assessing the fit of these methods to the data.

Although the outputs of our approach are represented by the parameters underlying differences in the overall shapes of growth rate curves, they can be described by regular genetic parameters, i.e., additive or dominant QTL effect and the percentage of the total phenotypic variance explained by a QTL. With these genetic parameters, our approach can increase the power of discriminating various important hypotheses that concern the genetic architecture of developmental features (1, 29, 31). The knowledge about the developmental change of QTL expression helped to formulate an effective early selection strategy based on marker information.

The approach proposed in this study can be extended to other situations, such as an F_2 design and a full-sib family, to deal with linked QTL of epistasis, to combine it with selective genotyping, to adopt composite or multiple interval mapping, or to implement other biologically meaningful growth curves, such as the double logistic model for human body growth trajectories (5). The choice of an appropriate growth model should be based on its optimal fit to observational data. In this study, we assume that there are specific QTL for growth rates during the whole period of growth. But it is possible that there exist QTL affecting the growth acceleration process. Nath and Moore (24) and Gergorczyk (14) derived the “growth acceleration curve” based on the third derivative of the Richards function. The growth acceleration curve is given by

\[ I = \frac{r^2G}{(1-k)^2} \left[ 1 - \left( \frac{a}{G} \right)^{1-k} \right] \left[ \frac{r}{1-k} \left( \frac{a}{G} \right)^{1-k} \right]^2 - rk \left[ 1 - \left( \frac{a}{G} \right)^{1-k} \right] \left( \frac{a}{G} \right)^{1-m} \]

From the second derivative of the Richards function, we can determine the inflection point of the growth curve at which absolute growth rate (AGR) is maximal during growth process. In our model, it is straightforward to formulate a hypothesis test about the effect of an AGR QTL on the inflection point. Using the above third derivative, Nath and Moore (24) and Gergorczyk (14) provided the equations for calculating the two points (P_1 and P_2) of inflection of the second derivative growth rate curve. At point P_1, the first inflection of the growth rate curve, maximum acceleration of growth occurs, whereas at point P_2, the second inflection of the growth rate curve, maximum deceleration of growth takes place. With these two points, one can separate the entire growth period into three phases: the exponential phase (from 0 to P_1), the linear phase (from P_1 to P_2), and the aging phase (from P_2 to \( \infty \)). These three phases can be incorporated into our mapping strategy to carefully investigate how a detected AGR QTL acts during these three different phases. With such information, it is possible to understand the genetic basis of development and design an efficient breeding plan to alter growth trajectories via marker-assisted selection.

Our functional mapping model for mapping growth rates assumes that the relationship between growth rate and age can be described by mathematical functions. However, in practice, this parametric approach may be limited by two facts. First, only one or two stages of growth are measured, so it is not feasible to use a prescribed function to model the growth data. Second, to adequately model the growth process from birth to adulthood in an organism, one need use a sophisticated function composed of too many parameters to be precisely estimated. Also, even the best model may miss some facets of the entire growth process, leading to biased results. When the number of time points measured is inadequate to be fitted by a function or when the entire growth process cannot be adequately described by a function, nonparametric methods for functional fitting (also called “smoothing methods”) that do not need to prescribe a parametric growth model should be devised. In statistics, a variety of methods for nonparametric function fitting, such as splines and kernel estimators, have been proposed (12, 13) and can be integrated into our functional mapping framework.

The publication of this manuscript was approved as Journal Series no. R-09203 by the Florida Agricultural Experiment Station.
DISCLOSURES

This work is supported by National Natural Science Foundation of China Outstanding Young Investigator Award 30128017 (to R. Wu), University of Florida Research Opportunity Fund 02050259, and University of South Florida Biodefense Grant 7222061-12 (to R. Wu).

APPENDIX

We derived the EM algorithm for estimating the QTL location, the growth rate parameters and residual (co)variance. Below, we give the log-likelihood functions used to estimate these parameters in $\Omega$ in the M step.

Denote

$$
\alpha_j(t) = \frac{ab_je^{-\gamma t}(1 + b_je^{-\gamma r})}{(1 + b_je^{-\gamma r})^2},
$$

$$
\beta_j(t) = \frac{ab_je^{-\gamma r}(1 + b_je^{-\gamma t})}{(1 + b_je^{-\gamma t})^2},
$$

$$
\gamma_j(t) = \frac{e^{-\gamma r}[1 - tr_j + b_1(1 + tr_j)e^{-\gamma r}]}{(1 + b_je^{-\gamma r})^2},
$$

where $j = 0, 1, t = 1, \ldots, m$.

We have

$$
a'_j = \sum_{i=1}^{n} \sum_{t=1}^{m-1} (\alpha_i(t) - \alpha_j(t+1))(\frac{e^{-\gamma t}(1 + b_je^{-\gamma r})}{(1 + b_je^{-\gamma r})^2}) - \frac{(1 - \rho^2)\alpha_i(t)e^{-\gamma r}}{(1 + b_je^{-\gamma r})^2}
\sum_{i=1}^{n} \Pi_{ij}
$$

$$
b'_j = \sum_{i=1}^{n} \sum_{t=1}^{m-1} \left(\frac{\beta_i(t) - \beta_j(t+1)}{(1 + b_je^{-\gamma r})^2} - \frac{(1 - \rho^2)\beta_i(t)e^{-\gamma r}}{(1 + b_je^{-\gamma r})^2}\right)
\sum_{i=1}^{n} \Pi_{ij}
$$

$$
r'_j = \sum_{i=1}^{n} \sum_{t=1}^{m-1} \left(\frac{\gamma_i(t) - \gamma_j(t+1)}{(1 + b_je^{-\gamma r})^2} - \frac{(1 - \rho^2)\gamma_i(t)e^{-\gamma r}}{(1 + b_je^{-\gamma r})^2}\right)
\sum_{i=1}^{n} \Pi_{ij}
$$

and

$$
(\sigma^2)' = \sum_{j=0}^{n} \sum_{i=1}^{n} \left(\frac{(y_i(t) - \mu_i(t)) - p(y_i(t + 1) - \mu_i(t + 1))^2}{(1 - \rho^2)(\mu_i(m) - \mu_i(m))^2}\right)
\sum_{i=1}^{n} \Pi_{ij}
$$

where

$$
A = \sum_{j=0}^{n} \sum_{i=1}^{n} \sum_{t=1}^{m-1} (y_i(t) - \mu_i(t))(y_i(t+1) - \mu_i(t+1))
$$

$$
B = \sum_{j=0}^{n} \sum_{i=1}^{n} \sum_{t=2}^{m} (y_i(t) - \mu_i(t))^2 - (y_i(m) - \mu_i(m))^2
$$

$$
C = \sum_{j=0}^{n} \sum_{i=1}^{n} \sum_{t=1}^{m-1} ((y_i(t) - \mu_i(t)) - p(y_i(t + 1) - \mu_i(t + 1))^2 + (1 - \rho^2)(\mu_i(m) - \mu_i(m))^2)/\sigma^2
$$

Parameters $[a', b', r', a_0, b_0, r'_0, p', (\sigma^2)']$ denote the new estimator of $\Omega$ in the next iterative step. The MLE of the QTL position ($\theta$) is obtained using a grip approach, by assuming and moving a particular QTL position throughout the chromosome and searching for a maximum LR at an optimal position.

REFERENCES


Physiol Genomics • VOL 14 • www.physiolgenomics.org