A Mechanistic Model for Genetic Machinery of Ontogenetic Growth

Rongling Wu,* Zuoheng Wang,* Wei Zhao* and James M. Cheverud†
*Department of Statistics, University of Florida, Gainesville, Florida 32611 and †Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri 63110

Manuscript received August 5, 2004
Accepted for publication September 10, 2004

ABSTRACT

Two different genetic mechanisms can be proposed to explain variation in growth trajectories. The allelic sensitivity hypothesis states that growth trajectory is controlled by the time-dependent expression of alleles at the deterministic quantitative trait loci (dQTL) formed during embryogenesis. The gene regulation hypothesis states that the differentiation in growth process is due to the opportunistic quantitative trait loci (oQTL) through their mediation with new developmental signals. These two hypotheses of genetic control have been elucidated in the literature. Here, we propose a new statistical model for discerning these two mechanisms in the context of growth trajectories by integrating growth laws within a QTL-mapping framework. This model is developed within the maximum-likelihood context, implemented with a grid approach for estimating the genomic positions of the deterministic and opportunistic QTL and the simplex algorithm for estimating the growth curve parameters of the genotypes at these QTL and the parameters modeling the residual (co)variance matrix. Our model allows for extensive hypothesis tests for the genetic control of growth processes and developmental events by these two types of QTL. The application of this new model to an F2 progeny in mice leads to the detection of deterministic and opportunistic QTL on chromosome 1 for mouse body mass growth. The estimates of QTL positions and effects from our model are broadly in agreement with those by traditional interval-mapping approaches. The implications of this model for biological and biomedical research are discussed.

GROWTH, defined as the change of size with age, is characteristic of every living entity on this planet. Growth analysis by various mathematical, biophysical, or physiological approaches has fascinated the students of biology for many decades (Brody 1945; von Bertalanffy 1957; West et al. 2001). Mechanistic modeling of growth processes and patterns has been instrumental in predicting meat production in agriculture (Scanes 2003) and stimulated the integration of developmental events into evolutionary studies (Rice 1997). In medicine, mathematical models constructed in terms of the difference between anabolism and catabolism have been used to predict dynamic changes of tumors over time (Norton 1988; Hart et al. 1998) and these models display important clinical consequences by deciphering the natural history of the disease.

Growth, from an evolutionary standpoint, represents the capability of an organism to sense differences in the internal or external environment, to redirect the developmental trajectory to better suit current environmental conditions, and to thereby increase fitness (Vinicius and Lahr 2003). A general view derived from fundamental physical and physiological principles is that growth is more expeditious during the early exponential phase since growth rate is self-accelerating or autocatalytic and continues to proceed albeit at a lower rate later when growth rate becomes self-inhibiting (West et al. 2001). However, there is a ubiquitous phenomenon that some organisms produce distinctly different growth through several phases during development, while others display more subtle developmental differences. These differences in growth form or trajectory among organisms are thought to provide fuel for adaptation and evolution through alterations of the timing and rate of developmental events (Rice 1997).

The genetic control of growth has been studied using various statistical models. Meyer (1998) proposed random-regression models to understand the genetic control of growth trajectories for animal breeding, whereas Kirkpatrick et al. (1990) derived a series of genetic models for infinite-dimensional traits in the evolutionary context. These models are particularly effective to study the genetic architecture of growth by modeling the (co)variance matrix for growth traits measured at different time points, although none of them can unravel the genetic mechanisms underlying growth differentiation at the molecular level.

The differentiation of growth trajectories can be attributed to a complex interaction between deterministic and opportunistic factors. The deterministic factors predispose an organism toward a specific growth shape (prototype), whereas the opportunistic factors modify it in response to the unique environment the organism
experiences. The deterministic factors represent a suite of inherited loci formed during embryogenesis, but the opportunistic factors that can be genetic or environmental occur postembryonically through continuous growth. Genetic opportunistic factors are an array of new genes that are activated by the regulatory system of the organism in response to changing environments. Environmental opportunistic factors include various predictable environmental changes and unpredictable stochastic errors.

One of the most important tasks in growth analysis is to discern the genetic mechanisms underlying growth from environmental influences. The effects of deterministic factors or deterministic quantitative trait loci (dQTL) on growth are triggered by allelic sensitivity, which is represented as the differential expression of alleles at these QTL across development. As opposed to dQTL, opportunistic (or indeterministic) quantitative trait loci (oQTL) regulate growth trajectories by responding to specific environmental cues to turn on or adjust the expression of structural genes that directly influence the growth. Thus, deterministic effects due to dQTL occur at the level of the whole growth curve, basically affecting the whole growth process, while opportunistic effects due to oQTL cause deviations from the growth model by adjusting growth rate and acceleration curves in response to developmental signals. Although they are not mutually exclusive, the deterministic and opportunistic mechanisms contribute differently to differentiation in growth trajectories. Wu (1998), for the first time, used QTL-mapping approaches to test the relative importance of these two mechanisms in shaping phenotypic plasticity. The “allelic sensitivity” hypothesis proposes that the QTL affecting the sensitivity of a trait to variation in an environment variable should map to the same regions as those that explain the variation in the across-environment mean trait value among genotypes. The “gene regulation” hypothesis predicts that the QTL regions that explain the variation among genotypes in the sensitivity of a trait to an environmental variable will be distinct from QTL that contribute to the variation in the trait among the genotypes within a given environment. Currently, QTL-mapping approaches have been used to study the phenotypic plasticity of life span in Drosophila (Leips and MACKAY 2000) and glucosinolate accumulation under methyl jasmonate treatment in Arabidopsis (Kliebenstein et al. 2003).

Here, we develop a novel statistical model for understanding the genetic architecture of growth trajectories by integrating the two underlying genetic mechanisms, allelic sensitivity and gene regulation, and the physiological principles behind growth and development in the QTL-mapping context. Wu and coworkers have proposed a host of statistical models for functional mapping of QTL responsible for the shape of growth trajectories using the growth data measured at many discrete time points (Wu et al. 2002, 2003a,b; Ma et al. 2002). These models have been successfully used to identify QTL affecting stem size growth in a forest tree and extended to map QTL for cancer growth, drug response, and human immunodeficiency virus (HIV)-1 dynamics. Beyond these previous models, the new model proposed in this article can provide greater insights into the physiological control mechanisms of growth and thereby provide a more rapid way to alter growth patterns toward a predetermined direction. We used body weight data in an F2 hybrid progeny of mice (Cheverud et al. 1996) to investigate the usefulness and power of this new model.

A GENERAL FRAMEWORK FOR FUNCTIONAL MAPPING

Mixture model: An arbitrary individual randomly chosen from a mapping population must carry one and only one of possible QTL genotypes at putative QTL with a particular probability. The central theme of genetic mapping of QTL using DNA-based polymorphic markers, such as single-nucleotide polymorphisms (SNPs) and microsatellites, is to estimate the probability that each individual carries a QTL genotype and, therefore, calculate the genetic effects of the QTL on the phenotype. Statistically, this problem can be described by a mixture model, in which the likelihood of each individual is the sum of the density functions \( f \) of observed phenotypes \( y \) for different QTL genotypes weighted by the corresponding genotype frequencies. Assuming that there are \( L \) QTL genotypes, this mixture model is expressed as

\[
\gamma \sim p(y|\pi, \phi, \eta) = \pi_1 f(y; \phi_1, \eta) + \ldots + \pi_L f(y; \phi_L, \eta),
\]

where \( \pi = (\pi_1, \ldots, \pi_L)^T \) are the mixture proportions that are constrained to be nonnegative and sum to unity; \( \phi = (\phi_1, \ldots, \phi_L)^T \) are the component-specific parameters, with \( \phi_l \) being specific to component \( l \); and \( \eta \) is a parameter that is common to all components. \( T \) stands for the vector transpose.

The estimation of QTL genotype frequencies (specified by \( \pi \)), QTL effects (specified by \( \phi \)), and the residual variance within each QTL genotype (specified by \( \eta \)) involves three steps. First, derive the unknown QTL genotype frequencies in the given mapping population conditional upon known marker genotype according to Mendelian (for linkage analysis) or population genetic principles (for linkage disequilibrium analysis). Second, model the genetic effects and residual variance (including residual covariances for a multivariate analysis) specified in the distribution on the basis of quantitative genetic theory. Third, develop an efficient computational algorithm to solve all the unknown parameters in the mixture described by Equation 1.
The occurrence of CIN, due to gene mutation, results in chromosomal instability (CIN) in tumor initiation. Tumor (sigmoidal) growth curves are found to result proposed to estimate these population parameters. Brody's universal growth law that can be described by mathematical functions. A considerable body of evidence suggests that the kinetics of ontogenetic growth of biological entities obey a universal growth law that can be described by mathematical functions (Brody 1945; Niklas 1994; Guiot et al. 2003). West et al. (2001) derived a conceptual model from basic cellular mechanisms and physical principles to explain why such a growth law is followed by any tissue, organ, and organism. In cancer research, universal tumor (sigmoidal) growth curves are found to result from chromosomal instability (CIN) in tumor initiation (Rasnake and Duesberg 1999; Fabarius et al. 2002). The occurrence of CIN, due to gene mutation, results in whole chromosomes or large fractions of chromosomes being gained or lost during cell division. This results in an imbalance in the number of chromosomes per cell (aneuploidy) and an enhanced rate of loss of heterozygosity (Stock and Bialy 2003).

Instead of estimating the genetic effects on the phenotypes measured at different time points, our function-mapping model estimates the parameters that describe growth curves (called the curve parameters). Also, different from traditional multivariate mapping approaches that estimate an entire unstructured (co)variance matrix (Jiang and Zeng 1995), functional mapping estimates only the parameters that characterize the structure of the (co)variance matrix (called the matrix parameters). The estimation of the curve parameters depends on practical biological problems; for example, cancer growth can be fit by a sigmoidal curve whereas HIV-1 dynamics are fit by a biexponential function. The best way in which the (co)variance matrix is structured for a data set should be chosen on the basis of modeling tests. For instance, stationary autoregressive models may be enough for some growth data sets, but for many other data sets nonstationary antedependence (AD) models may be more appropriate (Núñez-Antón and Zimmerman 2000; Zimmerman and Núñez-Antón 2001).

### Step 3—computational algorithms

A number of computational algorithms can be used to solve the mixture model of Equation 1 containing the unknown parameters $\Omega = (\pi, \phi, \eta)$. These algorithms include least-squares regression analysis, a maximum-likelihood method implemented with the EM algorithm, and a Bayesian approach implemented with a Markov chain Monte Carlo algorithm. For our functional mapping, the simplex algorithm, widely used in operations research (Nelder and Mead 1965), is found to provide a fast and precise maximum-likelihood estimate (MLE) of the curve and matrix parameters (Zhao et al. 2004). Additional advantages of the simplex algorithm are that it is derivative free and easy to implement.

For linkage analysis, the genotypic frequencies of QTL genotypes are functions of the recombination fraction between the QTL and markers. The estimation of these recombination fractions can be obtained using a grid approach by viewing the QTL position as a known parameter between two flanking markers. The profile for the support of the likelihood for the existence of a QTL is drawn across a linkage group and the peak of the profile corresponds to the position of the QTL over the genome. For LD analysis, the QTL-marker genotypic frequencies are determined by a group of population genetic parameters, i.e., the QTL and marker allele frequencies and QTL-marker linkage disequilibrium. Recently, a closed-form solution for the EM algorithm was proposed to estimate these population parameters through the estimation of haplotype frequencies (Lou et al. 2003).

A MECHANISTIC MODEL FOR MAPPING GROWTH

### Statistical models for growth, growth rate, and growth acceleration

The size of a biological entity is measured at a finite set of $\tau$ time points, so we have a finite set of data on each individual, which can be considered as a multivariate trait vector, $y = [y(1), \ldots, y(\tau)]$. As discussed in the Introduction, growth is controlled by two types of genetic machineries, one being the dQTL and the other being the oQTL. Cumulative growth for individual $i$ measured at time $t$ can be partitioned into two different components in terms of the underlying dQTL and residual effects. Statistically, this can be expressed as

$$y_i(t) = g_i(t) + e_i(t),$$

where $g_i(t)$ is the deterministic genetic component that affects growth trajectories, and $e_i(t)$ is the residual component including the aggregation of other genetic components and random environmental errors.
Growth rate for an individual during a time interval is the difference between the growth measured at two different time points, expressed as

$$
\Delta y_i(t, t + 1) = y_i(t + 1) - y_i(t).
$$

Incorporating the component partitioning of growth as described by Equation 2, we have

$$
\Delta y_i(t, t + 1) = [g_i(t + 1) - g_i(t)] + [e_i(t + 1) - e_i(t)]
+ \Delta \delta_i(t, t + 1).
$$

(3)

The first term of Equation 3 describes the allelic sensitivity of the dQTL from time t to t + 1. We assume that the residual increment [e_i(t + 1) - e_i(t)] from time t to t + 1 includes the opportunistic genetic component stimulated by new environmental signals [\Delta \delta_i(t, t + 1)] and random errors [\Delta \xi_i(t, t + 1)]. Thus, the second term of Equation 3 describes the gene regulation control of the oQTL that are newly activated by environmental signals from time t to t + 1. As compared to the growth model (Equation 2) that can identify only the deterministic genetic control, the growth rate model (Equation 3) can identify both deterministic and opportunistic genetic control. Whereas the determinant gene affects growth through its allelic sensitivity, as indicated by \(g_i(t + 1) - g_i(t)\), the indeterminate gene may control growth through regulatory interactions.

Growth acceleration \(\Delta y_i(t, t + 1)\) is defined as the difference between growth rates at different time intervals, i.e.,

$$
\Delta y_i(t, t + 1) = \Delta y_i(t + 1) - \Delta y_i(t - 1), t.
$$

On the basis of Equation 3, we have

$$
\Delta y_i(t, t + 1) = [g_i(t + 1) - 2g_i(t) + g_i(t - 1)] + \Delta \delta_i(t, t + 1)
+ \Delta \xi_i(t, t + 1),
$$

(4)

where the residual acceleration \([\Delta \delta_i(t, t + 1)]\) contains opportunistic genetic control \([\Delta \delta_i(t, t + 1)]\) and random error components \([\Delta \xi_i(t, t + 1)]\).

**Growth equations:** A number of mathematical functions have been proposed to deal with time-dependent growth. Our analysis is based on a general Richards (1959) function for growth, which is expressed as

$$
W(t) = a(1 + be^{-c})^{1/(1-b)}, \quad \text{with } k > 1, b > 1,
$$

(5)

where \(W\) is the value of an entity’s size attribute at time \(t\), \(a\) is the asymptotic value of size, \(b\) is a parameter to position the curve on the time axis, \(c\) is the growth rate constant, and \(k\) is the shape parameter of the curve. Differentiating Equation 5, we obtain

$$
dW \frac{dt}{dt} = \frac{c}{k - 1} \left[ 1 - \left( \frac{W^{k - 1}}{a} \right) \right],
$$

(6)

defined as the absolute growth rate (AGR) of size. The division of AGR by \(W\), i.e., \((1/W)(dW/dt)\), is known as the relative growth rate (RGR). Differentiating the RGR we obtain

$$
\frac{d^2W}{dt^2} = \frac{c}{k - 1} \left[ 1 - k \left( \frac{W^{k - 1}}{a} \right) \right],
$$

(7)

which is the absolute growth acceleration of \(W\).

Figure 1 illustrates the shapes of growth curve, growth rate curve, and growth acceleration curves based on Equations 5, 6, and 7, respectively. Genetic differentiation in growth curve is due to interactions between the dQTL and oQTL, with the dQTL contributing to the mean growth trajectory, as specified by Equation 5, and the oQTL contributing to the deviation of the trajectory by adjusting growth rate (Equation 6) and acceleration curves (Equation 7) in response to developmental signals.

**Genetic settings:** As analyzed using Equations 5, 6, and 7, growth, growth rate, and growth acceleration provide different information about the genetic control of ontogenetic growth. Let us consider three QTL, of which one is the dQTL with alleles \(A\) and \(a\) and the other two are the oQTL that regulate growth rate and acceleration, respectively. The oQTL associated with growth rate has alleles \(B\) and \(b\), and the oQTL associated with growth acceleration has alleles \(C\) and \(c\). These QTL are segregating in an \(F_2\) population. Each of the three QTL may be bracketed by the same pair of flanking markers, \(A - N\), or a different pair of flanking markers, \(M - N\), bracketing the dQTL, \(M - N\), bracketing the oQTL, or \(M - N\), bracketing the acceleration oQTL. The QTL genotypes at a QTL are unknown, but can be predicted from observed genotypes of flanking markers on the basis of their linkage relationship with the QTL. The conditional probability of a QTL genotype, conditional upon known marker genotypes, is expressed in terms of the recombination fractions among the markers and QTL. Such conditional probabilities form the vector for the proportions of components (\(\pi\)) contained in the mixture model.

We use a set of curve parameters \(\Omega_{\delta} = (a_\delta, b_\delta, c_\delta, k_\delta)\) to determine the growth curve of genotype \(u\) at the dQTL (\(u = 0\) for \(aa\), 1 for \(Aa\), and 2 for \(AA\)), \(\Omega_{\delta} = (a_\delta, b_\delta, c_\delta, k_\delta)\) to determine the growth rate curve of genotype \(v\) at the rate oQTL (\(v = 0\) for \(bb\), 1 for \(Bb\), and 2 for \(BB\)), and \(\Omega_{\delta} = (a_\delta, b_\delta, c_\delta, k_\delta)\) to determine the growth acceleration curve of genotype \(w\) at the acceleration QTL (\(w = 0\) for \(cc\), 1 for \(Cc\), and 2 for \(CC\)). The genotypic values of the dQTL for growth at different time points are fit by Equation 5, whereas the genotypic values of the rate oQTL for growth rate and of the acceleration oQTL for growth acceleration at different time intervals are fit by Equations 6 and 7, respectively. As seen below, \(\Omega_{\delta}\) and \(\Omega_{\delta}\) are jointly estimated on
the basis of Equation 3, and $\mathbf{\Omega}_{c_d}$ and $\mathbf{\Omega}_{c_w}$ are jointly estimated on the basis of Equation 4.

In our new functional mapping for growth rate or growth acceleration, we propose to simultaneously estimate the joint QTL genotypes at the dQTL and oQTL on the basis of the observed joint marker genotypes at two pairs of flanking markers. The conditional probabilities of joint QTL genotypes given joint marker genotypes can be derived separately for two different cases. First, if the two QTL are linked with different marker intervals, these conditional joint QTL genotype frequencies for individual $i$ carrying a particular marker genotype ($\pi_{ui}$ or $\pi_{wi}$) are calculated as the product of the conditional QTL genotype frequencies ($\pi_{dij}$, $\pi_{dij}$, $\pi_{dij}$) as predicted by respective marker genotypes, i.e.,

$$\pi_{ui} = \pi_{dij} \pi_{dij}$$

or

$$\pi_{wi} = \pi_{dij} \pi_{dij}.$$  

Note that the conditional probability given a marker genotype is equivalent to the conditional probability given a particular individual $i$ because we know the marker genotype of each individual. Second, if two QTL are linked with the same marker interval, the conditional joint QTL genotype frequencies at these two QTL given the corresponding marker pairs should be derived on the basis of classic Mendelian genetic principles (Wu and Casella 2005).

**A mechanism-based mixture model:** The traditional genetic mapping model of a longitudinal trait assumes a simple genetic control mode over growth. For example, the growth model based on Equation 2 can detect only dQTL that govern growth through the life of an organism. Here, we propose to use the growth rate (Equation 3) or acceleration model (Equation 4) to characterize more precise genetic control mechanisms underlying growth. Such mechanisms include both dQTL and oQTL control. Assume that the growth of an entity is measured at $\tau$ discrete time points for an $F_2$ population of size $n$. Growth rates are calculated as the differences of growth between two adjacent time points. For these growth rates to reflect the scale of growth of an individual, we need to include the measurement at one of the two extreme points. Without loss of generality, the growth measured at the first time point is included in growth rate data for individual $i$, arrayed by $\Delta y = [y_1(1), \Delta y(1), 2, \ldots, \Delta y(\tau - 1, \tau)]$. The likelihood function of all the growth rate data ($\Delta y$) can be formulated by a mixture model,

$$L(\mathbf{\Omega} | \Delta y) = \prod_{i=1}^{n} \left[ \sum_{l=1}^{g} \pi_{li} f_l(\Delta y) + \ldots + \pi_{li} f_l(\Delta y) \right],$$

where the unknown vector $\mathbf{\Omega}$ includes the QTL position in terms of the conditional QTL genotype probabilities given marker genotypes, i.e., mixture proportions $\pi = (\pi_1, \ldots, \pi_L)^T$, the QTL genotypic means and residual (co)variances contained in the multivariate normal distribution for QTL genotype $l$, $f(\Delta y)$, with the $\tau$-dimensional mean vector $\mathbf{m}$ and the $\tau \times \tau$ (co)variance matrix $\Sigma$. We have

$$\mathbf{m} = \begin{bmatrix} g(1), g(2) - g(1) + \Delta g(1, 2), \ldots, g(\tau) - g(\tau - 1) + \Delta g(\tau - 1, \tau) \end{bmatrix},$$

which contains the allelic sensitivity, $g(t + 1) - g(t)$,
due to the deterministic QTL $\Delta$ (fit by Equation 5) and the regulatory interaction, $\Delta \delta(t, t+1)$, due to the oQTL (fit by Equation 6). Thus, to model the mean vector in Equation 9, we need to estimate only the curve parameters for the growth ($\Omega_{ay}$) and growth rate curves ($\Omega_{by}$).

The estimation of all elements in $\Sigma$ is not efficient. An approach for structuring the matrix for repeated measurements should be used. One of the simplest approaches is the first-order autoregressive [AR(1)] model in which we need to estimate only two parameters, residual variance ($\sigma^2$) and residual correlation per unit time interval ($\rho$). The main advantages of the AR(1) model include easy mathematical manipulation and increased power. However, this structuring approach needs two strong assumptions, i.e., constant residual variance over time and residual correlation decaying in a proportion ($\rho$) with time interval. Wu et al. (2004b) implemented a transform-both-sides (TBS) model to make the first assumption more realistic. Many other parametric approaches have been proposed to structure the matrix and some of them may need the estimation of more detailed genetic architecture of complex phenotypes. The choice of the most parsimonious approach is discussed in Núñez-Antón and Zimmerman (2000).

Similarly, the opportunistic QTL for growth acceleration ($\delta y$) can be mapped by formulating a likelihood function expressed as

$$L(\Omega|\delta y) = \prod_{i=1}^{n} \left[ \pi_{i0} f_i(\delta y) + \ldots + \pi_{i1} f_i(\delta y) \right].$$

The mean vector in this case is given by

$$\mathbf{m} = [g(3) - 2g(2) + g(1) + \delta \delta(2, 3), \ldots, g(\tau) - 2g(\tau - 1) + g(\tau - 2) + \delta \delta(\tau - 1, \tau)],$$

and the (co)variance matrix, $\Sigma$, can be structured using an appropriate approach.

**Computational algorithms:** The likelihood functions described by Equations 8 and 10 contain the unknown vector $\Omega$ composed of the curve ($\Omega_{ay}$, $\Omega_{by}$, $\Omega_{ay}$) and matrix parameters [$\rho$, $\sigma^2$ if the AR(1) model is used] and unknown QTL positions. In QTL mapping, QTL positions can be viewed as a fixed parameter (Wu and Casella 2005). When the QTL are assumed to be at particular positions, conditional joint QTL genotype frequencies ($\pi$) given marker genotypes are known. With known QTL positions, we need to estimate only curve and matrix parameters. By moving QTL positions every 4 cM along a genome, a landscape for the likelihood-ratio test statistics determined by two genomic positions each for a different QTL can be drawn.

The estimation of the curve and matrix parameters can be based on the EM algorithm. But this may be extremely difficult because the nonlinear equations contained in our model are very complicated to differentiate. We have implemented the simplex algorithm (Nelder and Mead 1965) to solve the likelihood equations (Zhao et al. 2004).

**Hypothesis Tests**

Our model allows for a number of hypothesis tests to examine the genetic control of growth processes. These hypothesis tests can be about the control of a QTL over entire growth trajectories, growth at specific time points or time intervals, the timing at which growth proceeds to reach a critical size, or the timing of developmental features of interest (see Ma et al. 2002; Wu et al. 2003a, 2004a). Also, whether the deterministic and opportunistic QTL affect growth performances independently or in an integrative way can be tested. All these tests are helpful to address biological questions related to the control mechanisms of ontogenetic growth.

**The existence of the growth QTL:** Testing whether specific QTL exist to affect the presence of QTL for growth is a first step toward the understanding of the detailed genetic architecture of complex phenotypes. The genetic control over entire growth trajectories can be tested by formulating the following hypotheses,

$H_0$: $\Omega_{ay} = \Omega_{ay}; \Omega_{by} = \Omega_{by}; \Omega_{ay} = \Omega_{ay}$

$H_1$: at least one of the equalities above does not hold,

when growth rate data are analyzed, or

$H_0$: $\Omega_{ay} = \Omega_{ay}; \Omega_{by} = \Omega_{by}; \Omega_{ay} = \Omega_{ay}$

$H_1$: at least one of the equalities above does not hold,

when growth acceleration data are analyzed. The $H_0$ states that there are no QTL affecting growth trajectories (the reduced model), whereas the $H_1$ proposes that such QTL do exist (the full model). The test statistic for testing the hypotheses is calculated as the log-likelihood ratio of the reduced to the full model,

$$LR = -2 \left[ \ln L(\hat{\Omega}_{ay}, \hat{\Omega}_{by}, \hat{\Omega}_{ay}; \hat{\rho}, \hat{\sigma}^2) - \ln L(\tilde{\Omega}_{ay}, \tilde{\Omega}_{by}, \tilde{\Omega}_{ay}; \hat{\rho}, \hat{\sigma}^2) \right],$$

where the tilde and the carat denote the MLEs of the unknown parameters under $H_0$ and $H_1$, respectively. The LR is asymptotically chi-square distributed with 16 d.f. when large samples of individuals and a large number of time points within each individual are used. 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.

**The existence of dQTL:** Our model can test the existence of dQTL for growth. This test relies on linkage...
mapping in which particular QTL can map to testable genomic positions.

The null hypothesis for this test is given by

\[ \Omega_{d0} = \Omega_{d1} = \Omega_{d2} \]

The \( H_2 \) is the same as Equation 12 for the growth rate model and Equation 13 for the growth acceleration model.

**The existence of oQTL:** The null hypothesis for testing the existence of oQTL can be formulated by

\[ \Omega_{a0} = \Omega_{a1} = \Omega_{a2} \]

for the growth rate model and

\[ \Omega_{o0} = \Omega_{o1} = \Omega_{o2} \]

for the growth acceleration model.

In practice, it is interesting to test whether the dQTL and oQTL are identical. This test can be formulated by assuming that there is the same genomic position for the two QTL and that the curve parameters affected by dQTL and oQTL are identical among the corresponding QTL genotypes. Under the growth rate and acceleration model, these two assumptions are expressed, respectively, as

\[ \text{position}(d) = \text{position}(\beta); \quad \Omega_{a0} = \Omega_{a0}; \quad \Omega_{a1} = \Omega_{a1}; \quad \Omega_{a2} = \Omega_{a2}, \]

and

\[ \text{position}(a) = \text{position}(\delta); \quad \Omega_{o0} = \Omega_{o0}; \quad \Omega_{o1} = \Omega_{o1}; \quad \Omega_{o2} = \Omega_{o2}. \]

Wu et al. (2004a) proposed a functional mapping model for testing epistatic interaction effects on growth trajectories. This model can be extended in this study to test if the dQTL and oQTL interact to affect growth curves, growth rate curves, or growth acceleration curves.

**RESULTS**

Our newly developed model was used to map age-dependent QTL for body growth in a model animal system—mouse. Cheverud et al. (1996) constructed a linkage map based on 75 microsatellite markers in 535 \( F_2 \) progeny derived from two strains, Large (LG/J) and Small (SM/J). The \( F_2 \) hybrids were weighed at 10 weekly intervals starting at age 7 days. The raw weights were corrected for the effects of each covariate due to dam, litter size at birth, parity, and sex. Figure 2 describes growth curves, growth rate curves, and growth acceleration curves for body weights of the \( F_2 \) mice. Growth, growth rate, and growth acceleration data of all mice, except for one outlier, were well fit by Equations 5–7, respectively (data not shown). This outlier was excluded from any further analysis.

In this example, we employed the AR(1) model to model the (co)variance matrix for mouse body mass measured at 11 different time points. The residual (co)variance matrix (\( \Sigma \)) for the growth rate data is derived on the basis of the AR(1) model for the growth data. The covariance between the growth at time point 1 and the growth rate between two adjacent time points \((t, t + 1)\) is derived as

\[
\text{cov}[y(1), \Delta y(t, t + 1)] = \text{cov}[y(1), y(t + 1)] - \text{cov}[y(1), y(0)]
\]

\[
= (\rho^t - \rho^{t+1}) \sigma^2. \tag{14}
\]

Assuming two different time points \( t_k \geq t_0 \), the covariance between these two growth rates is derived as

![Figure 2.—Plots of body weight growth vs. ages for each of the 534 mice used to construct linkage maps in an F2 progeny population. The growth of these genotypes can be well fitted by a particular logistic curve.](image)
The log-likelihood ratio (LR) test statistics for claiming is at 20–22 cM from D1Mit20. Separate analyses suggested mouse chromosome 1 responsible for the growth in an additive fashion (the heterozygote $Aa$) and the inverse largely facilitates the extension of our model to any number of time points measured and can also significantly increase the computational efficiency.

Our model identified two significant QTL located on mouse chromosome 1 responsible for the growth process of body weight. Figure 3 gives the landscape of the log-likelihood ratio (LR) test statistics for claiming the existence of a dQTL ($\Delta$) and an oQTL ($\mathcal{O}$) on chromosome 1. The peak of the landscape that corresponds to the MLEs of the locations of these two QTL has the LR value of 391, well beyond the critical threshold (202) obtained from 100 permutation tests. As indicated in Figure 3, dQTL and oQTL are located at 30.3 and 93.6 cM, respectively, from one end of the chromosome. The hypothesis test about whether the dQTL and oQTL are the same QTL is rejected, suggesting that mass growth in mice is genetically controlled with two different mechanisms, such as allelic sensitivity and gene regulation.

The three growth curves for the dQTL, each corresponding to a QTL genotype, $AA$, $Aa$, or $aa$, in the $F_2$ population, are drawn in Figure 4A using the MLEs of the curve parameters (Table 1). As expected, the LG/J and SM/J strains contribute the positive allele $A$ or the negative allele $a$ to increased body mass growth, respectively. Statistical tests suggest that the three curves are significantly different and therefore the QTL detected triggers a significant impact on entire growth trajectories in mouse body mass. It is found that this QTL starts to display increased effects after age 4–6 weeks. This period seems to be a transition phase for gene action, before which the QTL exerts a small effect on body growth in an additive fashion (the heterozygote $Aa$ and homozygote $AA$ overlap), but after which the overdominant effect becomes important (the heterozygote $Aa$ is higher than the homozygote $AA$; Figure 4A). This dQTL was also detected by separate analyses of the growth measured at individual time points (Cheverud et al. 1996). The QTL location estimated by separate analyses is at 20–22 cM from D1Mit20. Separate analyses sug-
Mapping QTL for Growth

Figure 4.—Growth curves (A) and growth rate curves (B) of the three QTL genotypes at the deterministic QTL detected on mouse chromosome 1. The three QTL genotypes are denoted by AA, Aa, and aa, respectively.

Suggested that this QTL triggers an increased effect on body growth after age 6 weeks, exhibits an overdominant effect, and has the positive allele contributed by the LG/J strain. All these findings are broadly in agreement with those in our model. But our model derived from physiological and genetic mechanisms is biologically more relevant and statistically more precise and powerful than simple separate analyses. We further analyzed the difference in growth rate among the three genotypes at the dQTL detected on mouse chromosome 1 (Figure 4B). It appears that this QTL affects growth rate after age 4 weeks.

The detected oQTL modifies body growth in response to the developmental change of mice. As compared to the dQTL, this QTL displays smaller effects on growth trajectories (Table 1). Like the dQTL, the positive allele $B$ for increased body size growth is contributed by the LG/J strain. Because of small estimates of parameter $b$ (related to the initial value of growth), the three curves of the QTL genotypes ($BB$, $Bb$, and $bb$) at the oQTL tend to be straight lines (Figure 5). Like dQTL, the oQTL affects mouse body mass growth in an overdominant fashion. It seems that our detected oQTL has also been detected by separate analyses of growth by CHEV-
TABLE 1
Parameter estimates of a deterministic and opportunistic QTL on mouse chromosome 1 affecting growth curves in an F_2 progeny of mice

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Deterministic QTL</th>
<th>Opportunistic QTL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA Aa aa BB Bb bb</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>33.27 38.18 28.65</td>
<td>15.54 18.82 9.07</td>
</tr>
<tr>
<td>b</td>
<td>5.94   7.26  9.31</td>
<td>1.01 \times 10^{-4} 1.08 \times 10^{-4} 1.36 \times 10^{-4}</td>
</tr>
<tr>
<td>r</td>
<td>0.58   0.59  0.72</td>
<td>0.55 0.35 0.53</td>
</tr>
<tr>
<td>k</td>
<td>1.80   1.79  1.85</td>
<td>1.16 1.00 1.76</td>
</tr>
<tr>
<td>Position</td>
<td>30.3</td>
<td>93.6</td>
</tr>
<tr>
<td>\sigma^2</td>
<td></td>
<td>3.8701</td>
</tr>
<tr>
<td>\rho</td>
<td></td>
<td>0.7881</td>
</tr>
<tr>
<td>LR</td>
<td></td>
<td>391</td>
</tr>
</tbody>
</table>

The positions of the QTL are indicated by the distances from the first marker on chromosome 1.

erud et al. (1996). These authors identified a QTL at the last marker interval D1Mit14 and D1Mit17, but it triggered smaller effects on growth than the QTL detected between markers D1Mit20 and D1Mit7.

DISCUSSION

Growth is a complex trait that is controlled by actions and interactions of polygenes and sensitive to environmental changes. The genetic mapping of growth provides a powerful tool for the identification of genes affecting complex traits. To increase mapping power and precision of parameter estimation, we need to base our mapping approach on the physiological mechanisms for growth and the genetic mechanisms causing growth differentiation. In our earlier publications, a number of statistical models were proposed to integrate the physiological mechanisms of growth, described by mathematical functions (West et al. 2001), within a QTL-mapping framework (Ma et al. 2002; Wu et al. 2002, 2003a,b, 2004a,b). In this article, we present a comprehensive model for mapping QTL affecting ontogenetic growth by considering different genetic control mechanisms for growth curve, growth rate curve, and growth acceleration curve.

Two distinct mechanisms underlie ontogenetic growth: allelic sensitivity, that is, the differential allelic expression of the same dQTL across ages, and regulatory gene control, in which new oQTL are activated as the organism grows. Our model proposed in this article can identify these two mechanisms underlying ontogenetic growth and development. It allows for the test of a number of biologically important variables in organismic development, for example, the timing of maximal growth rate and the duration of linear growth phase (Ma et al. 2002). The knowledge about the genetic control of the timing...

FIGURE 5.—Growth rate curves of the three QTL genotypes at the opportunistic QTL detected on mouse chromosome 1. The three QTL genotypes are denoted by BB, Bb, and bb, respectively.
of developmental features can enhance our understanding of the genetic mechanisms underlying growth trajectories.

Our model provides information about the dynamic changes of allelic expression over time. Its application to a case study of mouse growth basically supports the previous detection of QTL by traditional mapping approaches (Cheverud et al. 1996). However, our model has greater power to detect significant QTL and provides more precise estimates of QTL parameters than traditional interval mapping does for single-time growth because our model simultaneously analyzes growth data measured at all possible time points. Furthermore, with the different growth curves, each corresponding to a QTL genotype, our model can be extended to investigate possible pleiotropic effects of the growth QTL on many different developmental events, such as the timing of sexual maturity and reproductive fitness, or biomedically important traits, such as metabolic rate and fatness. We can therefore integrate growth and development, which are historically regarded as two different biological problems, into a comprehensive framework under which their common or unique underlying genetic machineries are identified.

The maximum-likelihood-based statistical theory has proven to be effective for deriving our QTL-mapping methodology incorporated by growth laws (Ma et al. 2002; Wu et al. 2004a,b). However, in this study, attempting to embed more sophisticated developmental principles (and therefore more unknown parameters) in the mapping model, we encounter a considerable computational load although it is not a prohibitive factor. Also, it is likely that sample size and the number of time points used in practice are not sufficiently large, which affects the statistical inference of our model based on traditional maximum-likelihood theory. These considerations make it worthwhile to derive our functional mapping model in the Bayesian context (Robert and Casella 1999).

Our model will be particularly useful in biomedical research. For example, in cancer clinics, the characterization of the timing at which the exponential growth phase begins and the linear growth phase ends in tumor growth enables us to determine the time from initiation to clinical symptoms and from first symptoms to serious clinical problems. Furthermore, it will determine how long after treatment the tumor will take to recur if the treatment is unsuccessful. It may also determine the outcome of a treatment that takes several weeks or months to complete, such as in many radiotherapy or chemotherapy schedules. Knowledge of a tumor’s growth rate and growth potential can therefore be important for both prognosis and treatment planning (Izumi et al. 2003).

With the availability of the complete genome sequence of humans, it seems feasible to map and clone the genes that are responsible for the observed variations in growth parameters. Finding the genes and functions that underlie variation in growth characteristics is going to be an exciting challenge for future research. Such work holds great promise for development of gene therapies to prevent and control cancers.

We thank Bruce Walsh and an anonymous reviewer for their thoughtful comments that have improved the presentation of this manuscript. This work is supported by National Institutes of Health grant DK52514 to J.M.C. and an Outstanding Young Investigator Award of the National Natural Science Foundation of China (30128017), a University of Florida Research Opportunity Fund (02050250), and a University of South Florida Bioendowsee grant (1222061-12) to R.W. The publication of this manuscript was approved as Journal Series no. R-10572 by the Florida Agricultural Experiment Station.

LITERATURE CITED


Communicating editor: M. Feldman