

A unifying statistical model for QTL mapping of genotype \times sex interaction for developmental trajectories

Wei Zhao,¹ Changxing Ma,¹ James M. Cheverud,² and Rongling Wu¹

¹Department of Statistics, University of Florida, Gainesville, Florida 32611; and ²Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri 63110

Submitted 27 May 2004; accepted in final form 5 August 2004

Zhao, Wei, Changxing Ma, James M. Cheverud, and Rongling Wu. A unifying statistical model for QTL mapping of genotype \times sex interaction for developmental trajectories. *Physiol Genomics* 19: 218–227, 2004. First published August 10, 2004; doi:10.1152/physiolgenomics.00129.2004.—Most organisms display remarkable differences in morphological, anatomical, and developmental features between the two sexes. It has been recognized that these sex-dependent differences are controlled by an array of specific genetic factors, mediated through various environmental stimuli. In this paper, we present a unifying statistical model for mapping quantitative trait loci (QTL) that are responsible for sexual differences in growth trajectories during ontogenetic development. This model is derived within the maximum likelihood context, incorporated by sex-stimulated differentiation in growth form that is described by mathematical functions. A typical structural model is implemented to approximate time-dependent covariance matrices for longitudinal traits. This model allows for a number of biologically meaningful hypothesis tests regarding the effects of QTL on overall growth trajectories or particular stages of development. It is particularly powerful to test whether and how the genetic effects of QTL are expressed differently in different sexual backgrounds. Our model has been employed to map QTL affecting body mass growth trajectories in both male and female mice of an F₂ population derived from the large (LG/J) and small (SM/J) mouse strains. We detected four growth QTL on chromosomes 6, 7, 11, and 15, two of which trigger different effects on growth curves between the two sexes. All the four QTL display significant genotype-sex interaction effects on the timing of maximal growth rate in the ontogenetic growth of mice. The implications of our model for studying the genetic architecture of growth trajectories and its extensions to some more general situations are discussed.

EM algorithm; functional mapping; growth trajectories; mice; QTL \times sex interaction

FUNCTIONAL MAPPING HAS EMERGED as a powerful statistical method for detecting quantitative trait loci (QTL) that affect complex phenotypes undergoing developmental changes (21, 36, 38, 39). This method integrates mathematical models of growth curves with quantitative and molecular genetics theory within a QTL mapping framework. It does not directly estimate the expected means of different QTL genotypes at different time points, but instead fits these means using time-dependent growth curves derived from fundamental biological principles (2, 32). Thus the estimation of QTL effects on growth is equivalent to the estimation of model parameters describing the shape of growth curves.

Article published online before print. See web site for date of publication (<http://physiolgenomics.physiology.org>).

Address for reprint requests and other correspondence: R. Wu, Dept. of Statistics, 533 McCarty Hall C, Univ. of Florida, Gainesville, FL 32611 (E-mail: rwu@stat.ufl.edu).

Although it is originally conceived to map the dynamic effects of QTL on growth trajectories, functional mapping has been extended to study the genetic mechanisms for many other biological or biomedical processes including allometric scaling, thermal norm reaction, HIV-1 dynamics, tumor progression, biological clock, and drug response (31, 37). The results from real data analyses or computer simulations in each of these extensions indicate that functional mapping displays tremendous power to detect “dynamic” QTL. However, in real life, the influences of specific QTL on biological processes are far more complicated than what we have modeled thus far with gene actions of single QTL. It is impossible that the expression of QTL effects is isolated from the environment in which the organism is reared or from the genetic background of the organism. Mackay (22) documented a considerable body of evidence for genotype \times sex, genotype \times environment, and gene \times gene (epistatic) interactions at the individual QTL level for a number of morphological traits and longevity in *Drosophila*. Sex- or environment-specific QTL were found to affect growth, morphological and physiological traits, or susceptibility to different diseases in other species, such as humans (25) and mice (4, 13, 23, 30). Although no studies have been able to quantify the impacts of interacting QTL on the developmental process of a trait, these QTL must commonly exist to potentially direct organismal development toward the best utilization of resources in heterogeneous environments (22).

The motivation of this study is to develop a unifying statistical model for mapping QTL that display interactions with environments, sexes, or genetic backgrounds during ontogenetic growth. By quantifying how QTL effects on growth trajectories vary between males and females, in different environments, and in different genetic backgrounds, this model has power to unravel the genetic architecture of complex phenotypes in a comprehensive context of development, ecology, and evolution. Since the studies of these three “interactions” should be based on different genetic and statistical strategies, we have to report the results separately for each of them. In our earlier work (38), models for epistatic and QTL-environment interaction effects have been described. In this article, we present a different model for studying QTL-sex interaction for developmental trajectories. This article is organized as follows. In the first section, **METHODOLOGY**, we present basic methodologies for functional mapping of quantitative traits displaying genotype-sex interactions. In the second section, **HYPOTHESIS TESTS**, we describe a number of hypothesis tests including the molecular genetic basis of genotype-sex interactions. In the third section, **RESULTS**, we report the results from this model for a worked example in an F₂ progeny of

mice. The advantages of this model and its implications are discussed in the DISCUSSION.

METHODOLOGY

The finite mixture model. The statistical foundation of functional mapping for growth curves is based on a finite mixture model in which each curve fitted by a finite set of measurements with τ time points for any individual, arrayed by $\mathbf{y} = [\mathbf{y}(1), \dots, \mathbf{y}(\tau)]$, is assumed to have arisen from one of a known or unknown number of components, each component being modeled by a multivariate normal distribution density (20). Assuming that there are J genotypes contributing to the variation among different curves, this mixture model is expressed as

$$\mathbf{y} \sim p(\mathbf{y}|\pi, \varphi, \eta) = \pi_1 f_1(\mathbf{y}; \varphi_1, \eta) + \dots + \pi_J f_J(\mathbf{y}; \varphi_J, \eta) \quad (1)$$

where $\pi = (\pi_1, \dots, \pi_J)^T$ are the mixture proportions (i.e., genotype frequencies) which are constrained to be nonnegative and sum to unity; $\varphi = (\varphi_1, \dots, \varphi_J)^T$ are the component-specific (or genotype-specific) parameters, with φ_j being specific to component j ; and η is a parameter which is common to all components.

The likelihood function. Consider a standard F₂ design, initiated with two contrasting homozygous inbred lines. This F₂ progeny of size n contains two subpopulations, one composed of n_1 male siblings and the other composed of n_2 female siblings ($n_1 + n_2 = n$), in each of which there are three groups of genotypes at a locus. A composite genetic linkage map that integrates the two sexes is constructed with molecular markers, aimed at the identification of QTL affecting an age-dependent trait, such as body size or body weight. In practice, these longitudinal data are only observed at a finite set of times, but can be modeled by a growth curve, such as logistic or sigmoid functions.

Suppose there is a segregating QTL with alleles Q and q that affects growth curves or trajectories of both the male and female progeny from the F₂ population, but to different extents. For a particular genotype j ($j = 2$ for QQ , 1 for Qq , and 0 for qq) of this QTL, a set of parameters describing its growth curve is denoted by $G_{j(k)}$ with $k = 1$ denoting the male progeny and 2 denoting the female progeny. The comparisons of these parameters among the three different QTL genotypes across different sexes can determine whether and how this putative QTL differently affects growth trajectories for the two sexes.

The likelihood function of longitudinal data measured for the two sex F₂ populations, $\mathbf{y}_{i(k)}$, at this hypothesized QTL is constructed as

$$L(\pi, \varphi, \eta | \mathbf{y}) = \prod_{i=1}^n [\pi_2 f_2(\mathbf{y}_i; \varphi_2, \eta) + \pi_1 f_1(\mathbf{y}_i; \varphi_1, \eta) + \pi_0 f_0(\mathbf{y}_i; \varphi_0, \eta)] \quad (2)$$

$$= \prod_{k=1}^2 \prod_{i=1}^{n_k} [\pi_2 f_2(\mathbf{y}_{i(k)}; \varphi_{2(k)}, \eta_k) + \pi_1 f_1(\mathbf{y}_{i(k)}; \varphi_{1(k)}, \eta_k) + \pi_0 f_0(\mathbf{y}_{i(k)}; \varphi_{0(k)}, \eta_k)] \quad (3)$$

$$= \prod_{k=1}^2 L_k(\pi, \varphi_k, \eta_k | \mathbf{y}_k) \quad (4)$$

where the mixture proportions (π_j values) denote QTL genotype frequencies. Equation 2 assumes that all progeny are independent in their responses to time, whereas Eq. 3 assumes that the two sexes are independent, in which we can calculate the likelihood function $L_k(\pi, \varphi_k, \eta_k | \mathbf{y}_k)$ separately for each sex and combine the two likelihood functions through their multiplication. Traditional mapping methods (20) can be readily extended to accommodate the multivariate nature of time-dependent traits. The multivariate normal distribution of progeny i for each sex k measured at τ time points is expressed as

$$f(\mathbf{y}_{i(k)}; \varphi_{j(k)}, \eta_k) = \frac{1}{(2\pi)^{\tau/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{y}_{i(k)} - \mu_{j(k)}) \Sigma^{-1} (\mathbf{y}_{i(k)} - \mu_{j(k)})^T \right] \quad (5)$$

where $\mathbf{y}_{i(k)} = [\mathbf{y}_{i(k)}(1), \dots, \mathbf{y}_{i(k)}(\tau)]$ is a vector of sex-specific observation measured at τ time points, and component-specific parameters for different sexes, φ_k , are specified by $\mu_{i(k)} = [\mu_{i(k)}(1), \dots, \mu_{i(k)}(\tau)]$, a vector of expected values for genotype j at different points. At a particular time t , the relationship between the observation and expected mean can be described by a linear regression model,

$$\mathbf{y}_{i(k)}(t) = x_i \mu_{j(k)}(t) + e_{i(k)}(t)$$

where x_i is the indicator variable denoted as 1 if a QTL genotype j is considered for subject i and 0 otherwise; $e_{j(k)}(t)$ is the residual error that is iid normal with the mean of zero and the variance of $\sigma_k^2(t)$. The errors at two different time points, t_1 and t_2 , are correlated with the covariance of $\text{cov}_k(t_1, t_2)$. These (co)variances comprise a $(\tau \times \tau)$ matrix Σ_k whose elements are the common parameter η of the mixture model (Eq. 1).

The determination of the value for the indicator variable describing the genotypes of the QTL for progeny i is not obvious. According to the interval mapping theory, it is possible to do so if we use the segregation information of the known flanking markers that bracket the QTL. Suppose this QTL is bracketed by two flanking markers M_l (with alleles M_l and m_l) and M_{l+1} (with alleles M_{l+1} and m_{l+1}). Thus the QTL genotype frequencies in the F₂ population (denoted by π_j) should be expressed as the conditional probabilities of the unknown QTL genotypes given the known marker genotypes. Table 1 tabulates these conditional probabilities, generally expressed as $\pi_{j|i}$, where $j|i$ stands for QTL genotype j given a particular marker genotype for progeny i . We rewrite the likelihood function of sex-specific longitudinal data ($\mathbf{y}_{i(k)}$) and marker information genotyped for both sexes (K) as

$$L(\pi, \varphi, \eta | \mathbf{y}, \mathbf{K}) = \prod_{k=1}^2 \prod_{i=1}^{n_k} \left[\sum_{j=0}^2 \pi_{j|i} f_j(\mathbf{y}_{i(k)}; G_{j(k)}, \Theta_k) \right] \quad (6)$$

where Θ_k is a set of sex-specific parameters. Note that Eq. 6 is different from Eq. 3, as the former does not make use of marker information, whereas the latter does.

Modeling the mean vector and (co)variance matrix. The estimation of the mean vector $\mu_{j(k)}$ and the (co)variance matrix Σ_k is statistically difficult because they involve too many unknown parameters given a possible sample size. Also, such direct estimation does not take into account the biological

Table 1. Conditional probabilities of QTL genotypes upon marker genotypes for M_l and M_{l+1} in the F_2 and its endosperm tissue

Marker		Joint Marker-QTL Genotype Frequency		
Genotype	Frequency	QQ	Qq	qq
$M_l M_l M_{l+1} M_{l+1}$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r_1)^2(1-r_2)^2$	$\frac{1}{2}r_1r_2(1-r_1)(1-r_2)$	$\frac{1}{4}r_1^2r_2^2$
$M_l M_l M_{l+1} m_{l+1}$	$\frac{1}{2}(1-r)r$	$\frac{1}{2}r_2(1-r_1)^2(1-r_2)$	$\frac{1}{2}r_1(1-r_1)(1-2r_2+2r_2^2)$	$\frac{1}{2}r_1^2r_2(1-r_2)$
$M_l M_l m_{l+1} m_{l+1}$	$\frac{1}{4}r^2$	$\frac{1}{4}(1-r_1)^2r_2^2$	$\frac{1}{2}r_1r_2(1-r_1)(1-r_2)$	$\frac{1}{4}r_1^2(1-r_2)^2$
$M_l m_l M_{l+1} M_{l+1}$	$\frac{1}{2}(1-r)r$	$\frac{1}{2}r_1(1-r_1)(1-r_2)^2$	$\frac{1}{2}r_2(1-r_2)(1-2r_1+2r_1^2)$	$\frac{1}{2}r_1(1-r_1)r_2^2$
$M_l m_l M_{l+1} m_{l+1}$	$\frac{1}{2}(1-2r+2r^2)$	$r_1r_2(1-r_1)(1-r_2)$	$\frac{1}{2}(1-2r_1+2r_1^2)(1-2r_2+2r_2^2)$	$r_1r_2(1-r_1)(1-r_2)$
$M_l m_l m_{l+1} m_{l+1}$	$\frac{1}{2}(1-r)r$	$\frac{1}{2}r_1(1-r_1)r_2^2$	$\frac{1}{2}r_2(1-r_2)(1-2r_1+2r_1^2)$	$\frac{1}{2}r_1(1-r_1)(1-r_2)^2$
$m_l m_l M_{l+1} M_{l+1}$	$\frac{1}{4}r^2$	$\frac{1}{4}r_1^2(1-r_2)^2$	$\frac{1}{2}r_1r_2(1-r_1)(1-r_2)$	$\frac{1}{4}(1-r_1)^2r_2^2$
$m_l m_l M_{l+1} m_{l+1}$	$\frac{1}{2}(1-r)r$	$\frac{1}{2}r_1^2r_2(1-r_2)$	$\frac{1}{2}r_1(1-r_1)(1-2r_2+2r_2^2)$	$\frac{1}{2}r_2(1-r_1)^2(1-r_2)$
$m_l m_l m_{l+1} m_{l+1}$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}r_1^2r_2^2$	$\frac{1}{2}r_1r_2(1-r_1)(1-r_2)$	$\frac{1}{4}(1-r_1)^2(1-r_2)^2$

The conditional probability (π_{ji}) of the quantitative trait locus (QTL) genotype given a marker genotype is calculated as the ratio of the joint marker-QTL genotype frequency over the marker genotype frequency. r_1 , r_2 , and r are the recombination fractions between marker M_l and the QTL, between the QTL and marker M_{l+1} , and between the two flanking markers.

principles of growth and development. We incorporate the universal growth law, as described by a logistic equation, into the estimation process of the likelihood function (Eq. 6). Thus the mean value of QTL genotype j at time t is expressed by

$$\mu_{j(k)} = \frac{a_{j(k)}}{1 + b_{j(k)}e^{-c_{j(k)}t}}$$

where the growth parameter set $G_{j(k)} = (a_{j(k)}, b_{j(k)}, c_{j(k)})$ describes the asymptotic growth, initial growth, and relative growth rate, respectively (24). With this growth equation, we only need to estimate the growth parameters, rather than estimate genotypic means at every point, to detect genotypic differences in growth. This can significantly reduce the number of unknown parameters to be estimated, especially when the number of time points is large.

Similarly, the (co)variance matrix can be structured with an appropriate model. Statistical analysis of longitudinal data has established a number of structural models that capture most of the information contained in the matrix (12). Here, we use a first-order autoregressive [AR(1)] model to model the structure of the matrix, which is based on two assumptions: first, the variance σ^2 is constant over time; and second, the correlation decays in a proportion of ρ purely with time interval. With the AR(1) model, we only need to estimate $\Theta_k = (\rho_k, \sigma_k^2)$, instead of all elements in the matrix. The advantage of such a matrix-structuring model is to reduce the number of unknown parameters, without losing the information of the matrix. There are many other structural models that may be more advantageous over the stationary AR(1) model, but the choice of an optimal model in a particular situation should be based on statistical tests, as described in Zimmerman and Núñez-Antón (41).

Computational algorithms. As classified above, the unknown parameters that build up the likelihood function (Eq. 6)

include the curve parameters, matrix-structuring parameters, and the QTL genotype frequencies specified by QTL position measured in terms of the recombination fractions (r_1 or r_2) between the QTL and its flanking markers (see Table 1). Arrayed by $\Omega = (\Omega_k)_{k=1}^2 = (G_{j(k)}, \Theta_k, r_1)_{k=1}^2$, these unknowns can be estimated through differentiating the log-likelihood function of Eq. 2 with respect to each unknown, setting the derivative equal to zero, and solving the log-likelihood equations. This estimation process can be implemented with the EM algorithm as described below.

The log-likelihood function of growth and marker data for sex k based on Eq. 6 is given by

$$\log L_k(\Omega_k | \mathbf{y}_k, \mathbf{K}) = \sum_{i=1}^{n_k} \log \sum_{j=0}^2 [\pi_{ji} f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k)]$$

with the derivative with respect to any element Ω_l in Ω_k

$$\begin{aligned} \frac{\partial}{\partial \Omega_l} \log L_k(\Omega_k | \mathbf{y}_k, \mathbf{K}) &= \sum_{i=1}^{n_k} \sum_{j=0}^2 \frac{\pi_{ji} \frac{\partial}{\partial \Omega_l} f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k)}{\sum_{j=0}^2 [\pi_{ji} f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k)]} \\ &= \sum_{i=1}^{n_k} \sum_{j=0}^2 \frac{\pi_{ji} f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k)}{\sum_{j=0}^2 [\pi_{ji} f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k)]} \frac{\partial}{\partial \Omega_l} \log f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k) \\ &= \sum_{i=1}^{n_k} \sum_{j=0}^2 \Pi_{ji} \frac{\partial}{\partial \Omega_l} \log f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k) \end{aligned}$$

where we define

$$\Pi_{j|i} = \frac{\pi_{j|i} f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k)}{\sum_{j=0}^2 [\pi_{j|i} f_j(\mathbf{y}_{i(k)}, \mathbf{K}; G_{j(k)}, \Theta_k)]} \quad (7)$$

which could be thought of as a posterior probability that progeny i with a particular marker genotype has QTL genotype j . We then implement the EM algorithm with the expanded parameter set $\{\Omega, \Pi\}$, where $\Pi = \{P_{j|i}\}$. Conditional on Π (the E step; Eq. 7), we solve for

$$\frac{\partial}{\partial \Omega_l} \log L_k(\Omega_k | \mathbf{y}_k, \mathbf{K}) = 0 \quad (8)$$

to get the estimates of Ω (the M step; Eq. 8). The estimates are then used to update Π , and the process is repeated between Eqs. 7 and 8 until convergence. The values at convergence are the maximum likelihood estimates (MLEs) of Ω . The iterative expressions of estimating Ω from the previous step were given in Ma et al. (21) and Wu et al. (39). In Wu et al. (36), approximate estimates of the samplings errors from Fisher’s information matrices were given.

As usual, the QTL position parameter can be viewed as a known parameter, because a putative QTL can be searched at every 1 or 2 cM on a map interval bracketed by two markers throughout the entire linkage map. The amount of support for a QTL at a particular map position is often displayed graphically through the use of likelihood maps or profiles, which plot the likelihood ratio test statistic as a function of map position of the putative QTL (20).

HYPOTHESIS TESTS

Different from traditional mapping approaches, our functional mapping for longitudinal traits allows for the tests of a number of biologically meaningful hypotheses (38). 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 and “interaction” test for the change of QTL expression across times. These tests at different levels can be formulated to test the effects of QTL × sex interaction on the shape of growth.

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

$$\begin{cases} H_0 : & G_{j(k)} \equiv G_k, k = 1, 2 \\ H_1 : & \text{Not all the equalities in } H_0 \text{ hold} \end{cases} \quad (9)$$

The H_0 states that there are no QTL affecting growth trajectories and the three genotypic curves in each sex overlap (the reduced model), whereas the H_1 proposes that such QTL do exist (the full model). The test statistic for testing the hypotheses in Eq. 9 is calculated as the log-likelihood ratio (LR) of the reduced to the full model:

$$LR = -2[\log L(\tilde{\Omega} | \mathbf{y}; \mathbf{K}) - \log L(\hat{\Omega} | \mathbf{y}; \mathbf{K})] \quad (10)$$

where $\tilde{\Omega}$ and $\hat{\Omega}$ denote the MLEs of the unknown parameters under H_0 and H_1 , respectively. The LR is asymptotically χ^2 -distributed with 12 degrees of freedom. An empirical ap-

proach for determining the critical threshold is based permutation tests, as advocated by Churchill and Doerge (11). By repeatedly shuffling the relationships between marker genotypes and phenotypes, a series of the maximum LR values are calculated, from the distribution of which the critical threshold is determined.

After a significant QTL is detected, the next test is about the effect of this QTL on growth in each sex. This will use the same form as shown in Eq. 9, but focusing on a sex. It is interesting to test whether the QTL interacts with sex to affect growth trajectories. Such a null hypothesis test can be formulated as

$$G_{j(1)} \equiv G_{j(2)}, j = 2, 1, 0 \quad (11)$$

which states that any two curves between the same QTL genotype from different sexes overlap. However, when two sex-specific curves with the same QTL genotype are approximately parallel to each another, the area under the curve ($A_{j(k)}$) is an appropriate criterion for this QTL × sex interaction test, expressed as

$$\begin{aligned} A_{j(k)} &= \int_0^\tau \frac{a_{j(k)}}{1 + b_{j(k)} e^{-c_{j(k)} t}} dt \\ &= \frac{a_{j(k)}}{c_{j(k)}} [\ln(b_{j(k)} + e^{c_{j(k)} \tau}) - \ln(b + 1)]. \end{aligned}$$

In this case, the null hypothesis for testing QTL × sex interaction can be formulated as

$$A_{j(1)} - A_{j(2)} \equiv A_1 - A_2, j = 2, 1, 0 \quad (12)$$

i.e., the difference between the areas under curves for different sexes is set equal for the three QTL genotypes.

In addition to testing overall genetic effects on growth trajectories, our model allows for the tests of the additive and dominant effect as well as their interaction effects with sexes. Wu et al. (38) proposed detailed procedures for making these specific tests, all of which can be directly used or modified for this study.

Local test. The local test can test for the significance of the genetic effect of QTL and QTL × sex interaction effect on growth traits measured at a time point (t^*) of interest. For example, the hypothesis for testing the effect of QTL on growth at a given time t^* can be formulated as

$$\begin{cases} H_0 : & \mu_{j(1)}(t^*) = \mu_1(t^*), \mu_{j(2)}(t^*) = \mu_2(t^*) \\ H_1 : & \text{Not all the equalities in } H_0 \text{ hold} \end{cases} \quad (13)$$

which is equivalent to testing the difference of the full model with no restriction and the reduced model with a restriction as set in the null hypothesis.

Regional test. Sometimes we are interested in testing the difference of growth trajectories in a time interval rather than simply at a time point. The question of how a QTL exerts its effects on a period of growth trajectories $[t_1, t_2]$ can be tested using a regional test approach based on the areas,

$$\begin{aligned} A_{j(k)} &= \int_{t_1}^{t_2} \frac{a_{j(k)}}{1 + b_{j(k)} e^{-c_{j(k)} t}} dt \\ &= \frac{a_{j(k)}}{c_{j(k)}} [\ln(b_{j(k)} + e^{c_{j(k)} t_2}) - \ln(b_{j(k)} + e^{c_{j(k)} t_1})] \end{aligned}$$

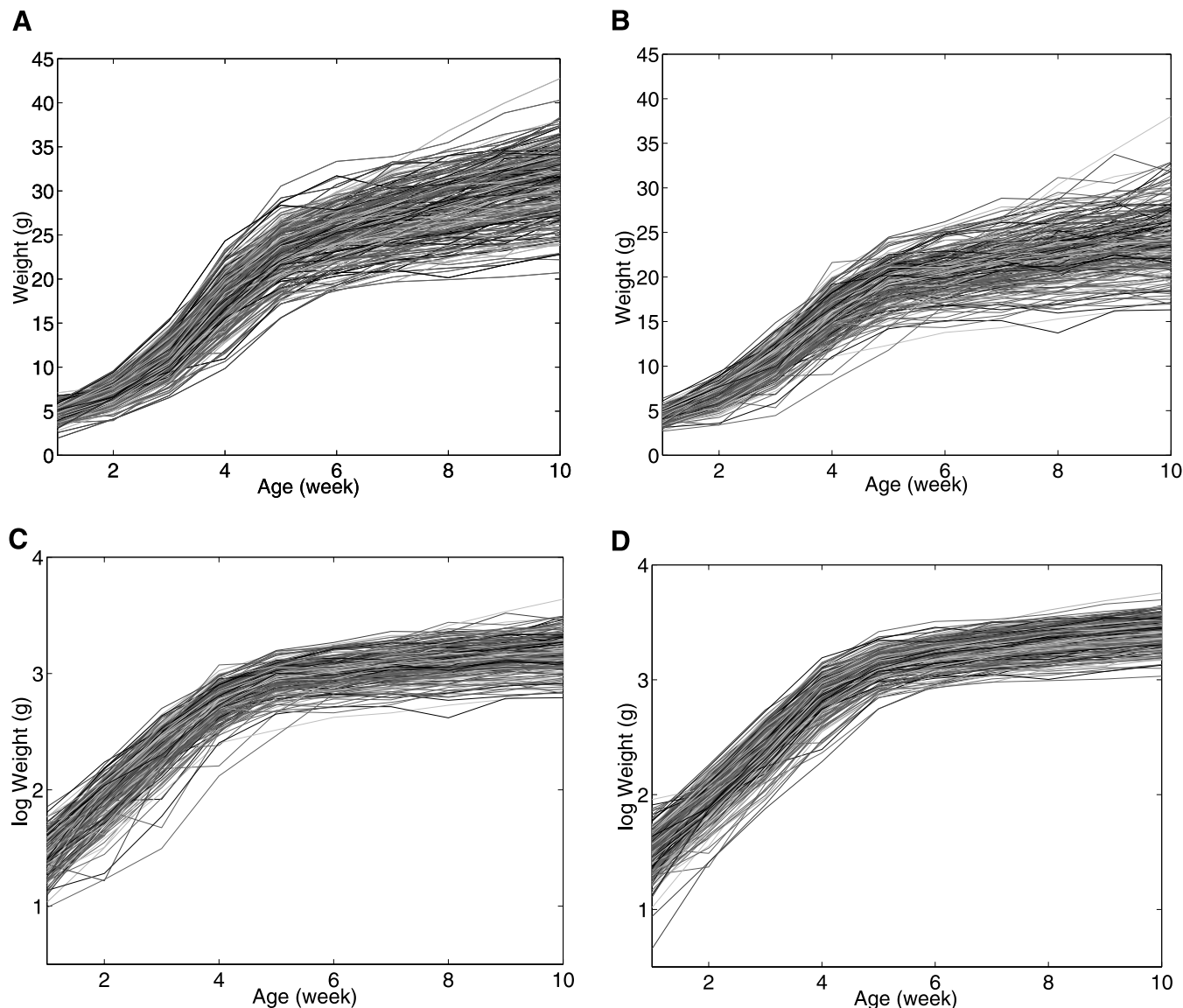


Fig. 1. Plots of body mass vs. age for 259 male (A) and 243 female (B) mice in an F₂ progeny derived from LG/J and SM/J strains (30). To display sex-specific differences, body mass was not corrected for the sex effect. The log-transformed data are plotted separately (C and D).

covered by load curves. The hypothesis test for the genetic effect on a period of growth process is equivalent to testing the difference between the full model without no restriction and the reduced model with a restriction.

Interaction test. The effects of QTL may change with time, which suggests the occurrence of QTL × time interaction effects on growth trajectories. The differentiation of growth with respect to time t represents growth rate. If the growth rates at a particular time point t^* are different between the curves of different QTL genotypes, then this means that significant QTL × time interaction occurs between this time point and next.

Test for biologically important parameters. There are a number of biological parameters that can be used to evaluate the developmental characteristics of growth. The logistic growth curve can be used to determine the coordinates of a biologically important point in the entire growth trajectory, the

“inflection point,” where the exponential phase ends and the asymptotic phase begins (24). The time at the inflection point corresponds to the time point at which a maximum growth rate occurs. The time

$$t_{I(j(k))}$$

and growth

$$\mu(t_{I(j(k))})$$

at the inflection point for QTL genotype j from sex k can be derived as

$$\begin{cases} t_{I(j(k))} = \frac{\ln b_{j(k)}}{c_{j(k)}} \\ \mu(t_{I(j(k))}) = \frac{a_{j(k)}}{2} \end{cases} \quad (14)$$

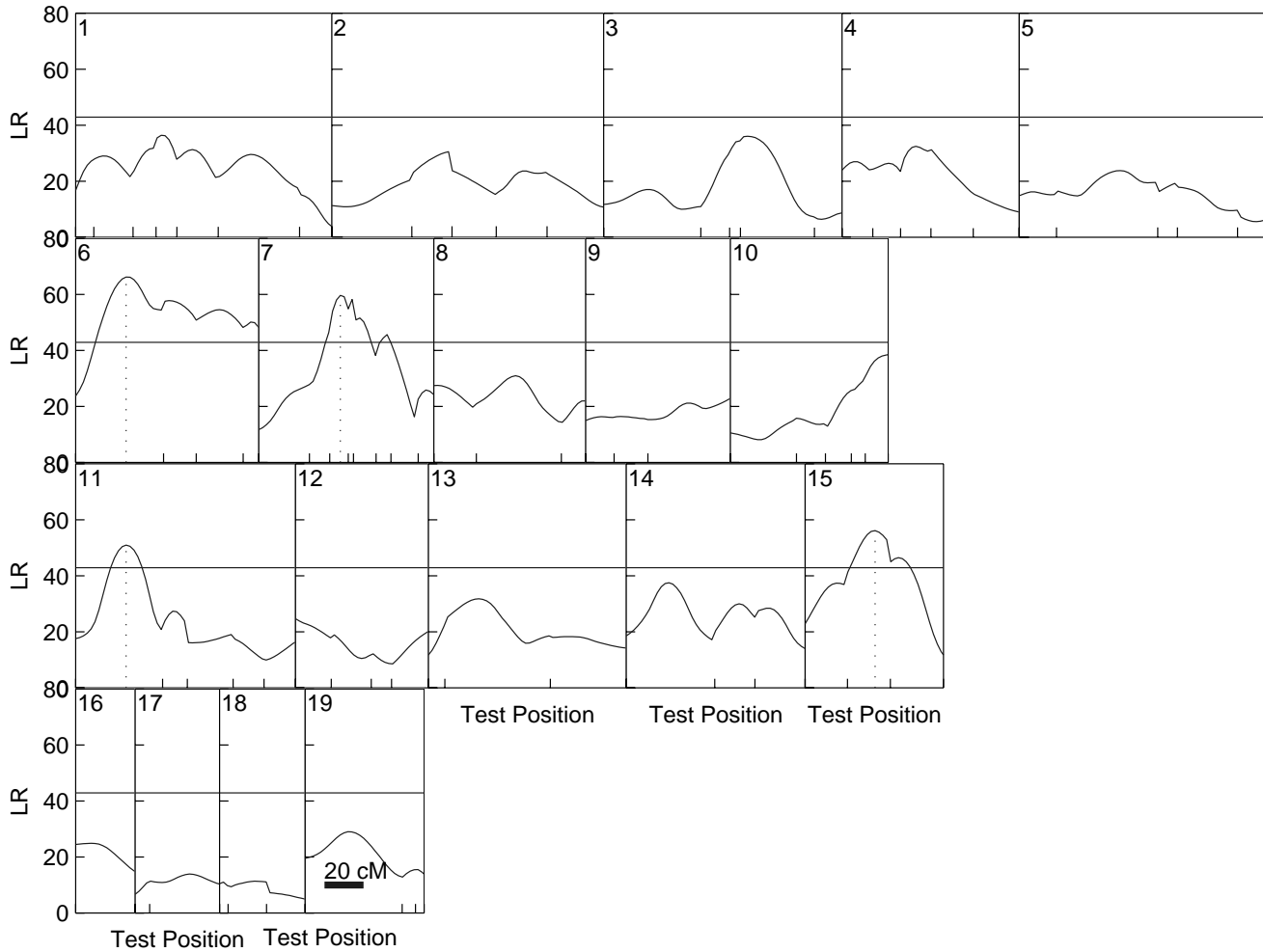


Fig. 2. The profile of the log-likelihood ratios (LR) between the full and reduced (no QTL) model for body mass growth trajectories across the entire genome using the linkage map constructed from microsatellite markers (30). The genomic positions corresponding to the peak of the curve, as indicated by vertical dot lines, are the maximum likelihood estimates (MLEs) of the quantitative trait loci (QTL) positions. The genome-wide threshold value for claiming the existence of QTL is given as the horizontal line. Tick marks on the x-axis represent the positions of markers on the linkage group, whose names are given in Vaughn et al. (30).

The difference in the coordinates between different genotypes provides important information about the genetics and evolution of growth trajectories (24). The genotypic differences in time and growth at the inflection point of maximum growth rate can be tested. The test for the genotypic difference is based on the restriction

$$\frac{\ln b_{j(1)}}{c_{j(1)}} = \frac{\ln b_1}{c_1}, \frac{\ln b_{j(2)}}{c_{j(2)}} = \frac{\ln b_2}{c_2}, j = 2,1,0 \quad \text{for } t_{I(j(k))}$$

and

$$a_{j(1)} = a_1, a_{j(2)} = a_2, j = 2,1,0 \quad \text{for } \mu(t_{I(j(k))}).$$

RESULTS

We used the joint statistical model to map sex-specific QTL that affect growth trajectories in an animal model system: the mouse. Cheverud et al. (8) constructed a linkage map with 75 microsatellite markers for 535 F₂ mice derived from two strains, the large (LG/J) and small (SM/J). The same cross

experiment was replicated by Vaughn et al. (30) to produce a new F₂ population of 259 male and 243 female mice. A molecular linkage map based on the new F₂ population was constructed from 96 polymorphic loci. Vaughn et al. (30) reported the construction of this map with a total map distance of ~1,780 cM (in Haldane units) and an average interval length of ~23 cM.

The F₂ progeny were measured for their body mass 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, and parity (30). To study the genetic architecture of sexual differences in body mass growth, the raw weights were not corrected for the effect due to sex. Figure 1 illustrates growth curves of body weights separately for the male (Fig. 1A) and female F₂ mice (Fig. 1B). On average, males display different growth trajectories than females, with the former being heavier at all time points than the latter. Substantial variation in growth curve among different animals in each sex suggests that specific QTL may be involved in shaping developmental trajectories.

Our model is employed to search for growth QTL through a genome-wide scanning approach. The assumption of constant variance for the AR(1) model may not be true in our data, as indicated by increased variance with ages in both males and females (Fig. 1). Wu et al. (39) proposed a transformation approach, called the “transform-both-sides” (TBS) model by Carroll and Ruppert (5), to reduce variance heteroscedasticity and, therefore, increase the power of the model. This TBS-based mapping model can also preserve the biological meanings of curve parameters. In this study, we incorporate the TBS-based model through log transformation into the functional mapping framework for analyzing QTL \times sex interactions. As shown in Fig. 1, *C* and *D*, the log transformation can lead to relatively constant variances in body mass growth for both males and females, although a more effective transformation approach should be estimated simultaneously with the model parameters (3, 5). Four peaks of the LR profile throughout the genome were found on chromosomes 6, 7, 11, and 15 (Fig. 2). The LR values at these peaks were detected to be

greater than the genome-wide critical threshold value at the $\alpha = 0.05$ significance level determined on the basis of 100 permutation tests. This thus suggests the existence of significant QTL for growth curves at the corresponding LR peaks on these chromosomes.

The three growth curves each determined by a genotype at each of these significant QTL are drawn separately for males and females (Fig. 3) using the MLEs of curve parameters ($\hat{G}_{j(k)}$; Table 2) from our model. The growth trajectories of the same QTL genotype are different between the two sexes, suggesting that the genetic expression of QTL is affected by sex-related background. In general, these four QTL start to exert their effects on growth when the mice age 3 or 4 wk. These ages are just the timing at which maximal growth rate occurs (inflection point; Fig. 3). After the inflection point, the QTL effects tend to increase with age.

We further tested how the QTL interact with sex to affect growth trajectories. Based on the hypothesis test given in Eq. 12, we calculate the LR values for QTL sex interaction effects

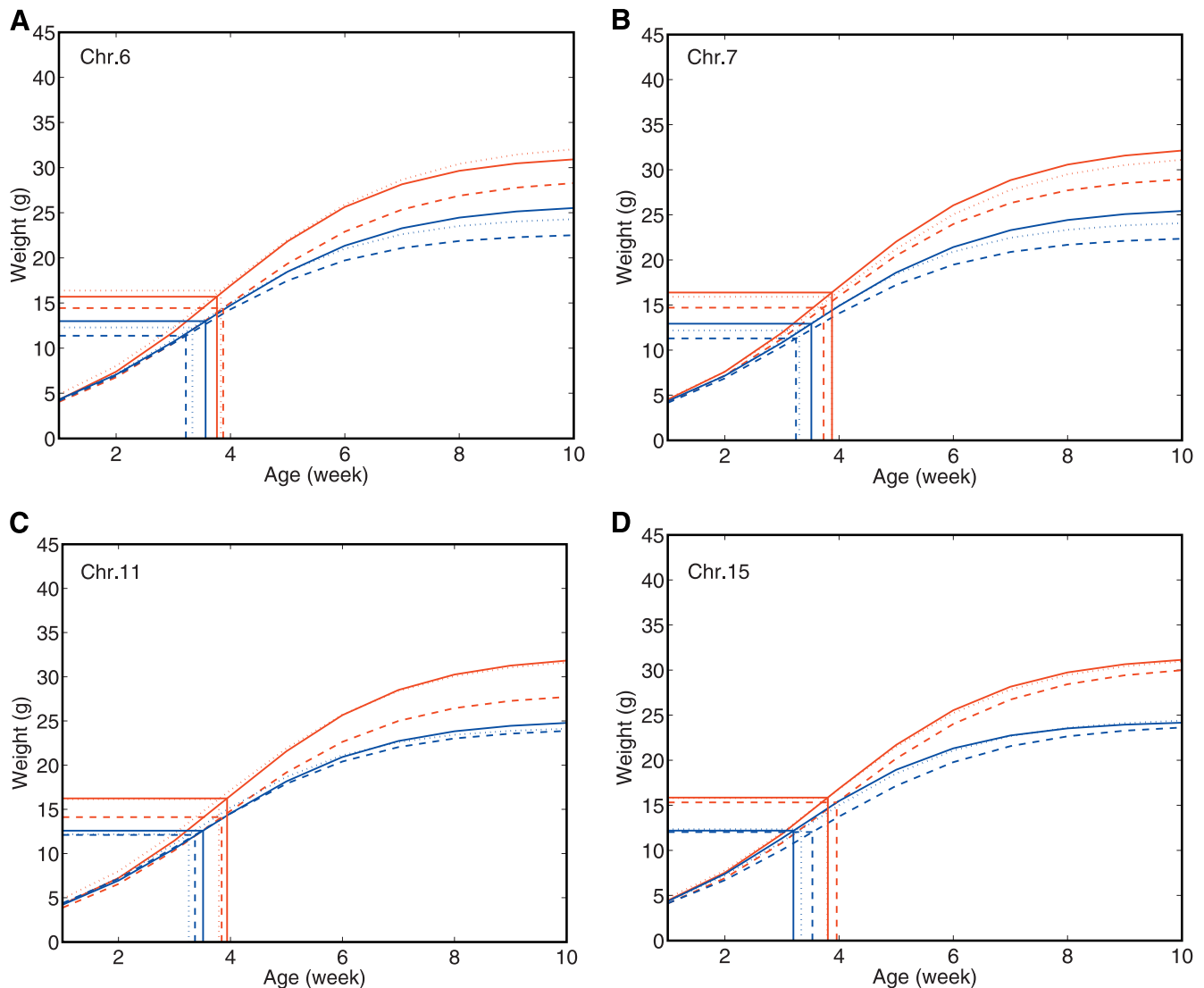


Fig. 3. Three growth curves each presenting three groups of genotypes, *QQ* (solid curves), *Qq* (dot curves), and *qq* (broken curves), in the male (red) and female (blue) mice at the QTL, detected by our joint model, on chromosomes 6, 7, 11, and 15. The times at the inflection point ($t_{l(j(k))}$) are indicated by the vertical lines each corresponding to a QTL genotype in each sex.

Table 2. The MLEs of the QTL position, QTL effects described by growth parameters, residual variance, and correlation in two different sexes of the F₂ mouse population

	Position, cM	qq			Qq			QQ			Residual	
		a _{k0}	b _{k0}	c _{k0}	a _{k1}	b _{k1}	c _{k1}	a _{k2}	b _{k2}	c _{k2}	σ _k ²	ρ _k
Chromosome 6	26											
Males		31.39	6.35	0.6687	32.75	5.71	0.6154	28.87	6.14	0.6320	0.0161	0.7723
Females		25.36	4.91	0.6376	24.34	4.73	0.6662	23.89	4.64	0.6617	0.0203	0.8224
Chromosome 7	42											
Males		32.77	6.27	0.6383	31.81	5.89	0.6168	29.42	5.95	0.6534	0.0197	0.8123
Females		24.52	4.58	0.6609	24.50	4.68	0.6547	24.01	4.72	0.6532	0.0194	0.8317
Chromosome 11	26											
Males		32.47	6.65	0.6446	32.29	5.63	0.6187	28.23	6.32	0.6485	0.0152	0.7622
Females		24.69	4.86	0.6484	24.74	4.70	0.6669	23.80	4.54	0.6587	0.0178	0.8142
Chromosome 15	36											
Males		31.70	6.15	0.6478	31.60	5.71	0.6235	30.65	6.46	0.6306	0.0188	0.8039
Females		24.64	4.64	0.6785	24.39	4.74	0.6523	24.26	4.67	0.6525	0.0200	0.8212

Position indicates the map distance in centimorgans from the first marker on a chromosome. MLE, maximum likelihood estimate; growth parameter ($G_{j(k)} = (a_{j(k)}, b_{j(k)}, c_{j(k)})$); σ_k², residual variance; ρ_k, residual correlation.

for all the four QTL (Table 3). Significant interactions were detected for QTL on chromosomes 6 and 11. The QTL on chromosome 6 is significant for both males and females (Fig. 3A), but the modes of gene action are different between the two sexes. In males, this QTL appears to be over-dominant because the heterozygote (Qq) outgrows the better homozygote (QQ). But in females, this QTL operates in a partial dominant fashion since the heterozygote is between the two homozygotes. Given these analyses, the QTL on chromosome 6 triggers significant interaction effects with sex through a so-called “allelic sensitivity” mechanism in which phenotypic changes result from differential expression of the same QTL (35). The second interacting QTL on chromosome 11 has a different mechanism. It exerts an effect on growth only in one sex (Fig. 3C). Although this QTL affects growth trajectories in a dominant fashion in males, it displays a nonsignificant effect in females. Thus this QTL may use a “regulatory mechanism” to affect differentiation in growth curves in that phenotypic changes rely upon the formation of novel genes.

Males reach the inflection point 3–5 days earlier than females. As demonstrated in Fig. 3, the QTL with significant effects on overall growth curves also affect the timing of the occurrence of maximal growth rate. Moreover, these QTL interact significantly with sexes to affect the timing of the inflection point, as shown by the LR values calculated according to the section *Test for biologically important parameters* (in HYPOTHESIS TESTS, above) (Table 3). With different growth curves, each corresponding to a QTL genotype, we can investigate possible pleiotropic effects of each of these growth QTL

Table 3. LR values and the corresponding P values for testing the genotype-sex interaction effects on overall growth curves and the timing of the inflection point for the four QTL detected on chromosomes 6, 7, 11, and 15

QTL Position	Overall Growth Curve	Timing of the Inflection Point
Chromosome 6	7.76 (0.0207)	13.60 (0.001)
Chromosome 7	3.09 (0.2136)	11.64 (0.003)
Chromosome 11	17.95 (0.0001)	16.68 (0.0002)
Chromosome 15	1.68 (0.4307)	18.79 (0.0001)

LR, log-likelihood ratio; P values are in parentheses.

on many other 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.

DISCUSSION

From a developmental perspective, the sexes are an important force to generate different optimal reproductive phenotypes (26). However, because the sexes represent different environments in which homologous traits are expressed, a conflict that cause genotype sex interactions may arise (6). It has now been recognized that the interaction between genotype and sex plays a central role in creating and maintaining extensive genetic variation for quantitative traits in heterogeneous environments (22). The statistical estimation of genotype × sex interaction at the individual QTL level, although extremely important in genetic and evolutionary studies, is difficult, due to the fact that the phenotype may not replicate across sexes. Cheverud and colleagues (9, 13, 30) used the general linear model to make a genome-wide scan for genotype × sex interactions at the QTL affecting a quantitative trait. While most quantitative traits undergo distinct developmental changes and, thus, should be described by multiple phenotypes measured at different time points (1, 7), the understanding of QTL × sex interaction effects on developmental trajectories has become a fundamental theme for evolutionary and developmental genetics.

Thanks to the recent development of functional mapping models for QTL mapping of longitudinal traits (21, 36, 38, 39), we have been able to develop a joint statistical framework for addressing this basic question of how QTL genotypes interact with sex in a coordinated manner to regulate ontogenetic trajectories in organisms. Our model is both statistically and biologically advantageous over existing QTL mapping models. First, it permits the mean vector and (co)variance matrix to be structured, thus largely reducing the parameters being estimated. The modeling of the mean vector composed of the expected means at different time points is based on biological

phenomena of interest that have been established from fundamental biochemical or physiological principles. For example, change in the size of any biological entity as a function of time follows a particular growth law that can be described by sigmoid or logistic curves (32). The incorporation of such a growth law does not only reduce the number of unknown parameters, but also, more importantly, can be well supported by biological principles. The results from such analyses are expected to be more biologically relevant.

Some specific patterns occur for the structure of a (co)variance matrix for longitudinal traits (12, 41). A considerable body of statistical literature has documented the parametric approaches for structuring the (co)variance matrices (14–16, 27, 28). Statistical theories have well been established to select an optimal approach for fitting a particular (co)variance matrix (41). Although the simple AR(1) model through TBS transformation (5) works well for the age-dependent (co)variance matrix in our example, it can be readily extended to implement more sophisticated structural models within our functional mapping framework. In particular, nonparametric or semiparametric approaches will be likely to substantially contribute to the structuring of (co)variance matrices (40).

Second, well beyond existing models, our model allows for tests of a number of developmentally or ecologically meaningful questions at the interplay among different biological disciplines. Our earlier model incorporating epistatic effects due to different QTL can shed light on the genetic control mechanisms over developmental aspect of growth (38). The model proposed in this article preserves all favorable features of the model of Wu et al. (38). The current model allows for integration between genetics, ecology, and development. More specifically, it can test when a sex-specific QTL starts to trigger an effect on growth, how long this effect takes during the time course of growth, and in which gene action mode it can alter developmental trajectories. These questions, once incorporated into an evolutionary genetic model, will help to address some long-standing debates regarding organismal development and evolution (17, 29).

Our model was used to reanalyze published data on body mass growth in an F₂ progeny population derived from large and small strains. We have successfully detected four QTL that affect growth trajectories. Statistical tests suggest that two of them on chromosomes 6 and 11 are sex-specific in trait control over overall growth curves. These two QTL use different mechanisms to alter growth patterns between the two sexes. In general, our results from the joint functional mapping model support the previous findings using single-trait mapping models (30). For example, two QTL on chromosomes 6 and 7 were detected in both studies. More QTL detected in Vaughn et al. (30) may be due to their doubled sample size and less stringent thresholds they used to reduce the type II error rate. Compared with single-trait approaches, our functional mapping that capitalizes on biological principles and correlated information has proved to be more powerful for detecting significant QTL and to provide biologically more relevant results (21).

It should be pointed out that the power of our model is affected by the ways in which to model the mean vector and the structure of the residual (co)variance matrix. Preliminary tests are needed for the goodness-of-fit of observed growth data to a particular growth equation and (co)variance structuring model. The model presented in this article can be further

extended to incorporate other development- or reproduction-related traits. Growth cannot be viewed as a trait isolated from morphological architecture, developmental timing, or reproductive fitness. Cheverud et al. (9) investigated the genetic control of adiposity. The influences of QTL on the size and shape of mandibular molars were analyzed in some detail (10, 18, 19, 34). It is possible to integrate the genetic control of other traits into our functional mapping framework, which will allow for the identification of pleiotropic QTL on different traits. In the model reported here, we assume that single QTL controls the differentiation in growth trajectories between the two sexes. Given the ubiquity of epistatic effects on development, however, it is essential to model QTL-QTL interactions in this sex-typical mapping model. Also, such QTL-QTL interactions are not necessarily derived from the same genome rather than from different genomes (male and female), as pointed out in a recent study by Wolf (33). The QTL-QTL interaction from different genomes, called genome-genome or individual-individual interaction, should be considered in our model to gain greater insights into the genetic and developmental aspects of ontogenetic growth.

ACKNOWLEDGMENTS

The publication of this manuscript is approved as journal series R-09205 by the Florida Agricultural Experiment Station.

GRANTS

This work is partially supported by National Institutes of Health Grant DK-52514 to J. M. Cheverud and by an Outstanding Young Investigators Award (no. 30128017) of the National Natural Science Foundation of China and the University of Florida Research Opportunity Fund (no. 02050259) to R. Wu.

REFERENCES

1. Atchley WR and Zhu J. Developmental quantitative genetics, conditional epigenetic variability and growth in mice. *Genetics* 147: 765–776, 1997.
2. Bertalanffy VL. Quantitative laws in metabolism and growth. *Q Rev Biol* 32: 217–231, 1957.
3. Box GEP and Cox DR. An analysis of transformations. *J Royal Stat Soc Ser B* 26: 211–252, 1964.
4. Butterfield RJ, Roper RJ, Rhein DM, Melvold RW, Haynes L, Ma RZ, Doerge RW, and Teuscher C. Sex specific QTL govern susceptibility to Theiler's murine encephalomyelitis virus-induced demyelination (TMEVD). *Genetics* 163: 1041–1046, 2003.
5. Carroll RJ and Ruppert D. Power-transformations when fitting theoretical models to data. *J Am Stat Assoc* 79: 321–328, 1984.
6. Chapman T, Arnqvist G, Bangham J, and Rowe L. Sexual conflict. *Trends Ecol Evol* 18: 41–47, 2003.
7. Cheverud JM. Quantitative genetics and developmental constraints on evolution by selection. *J Theor Biol* 110: 155–171, 1984.
8. Cheverud JM, Routman EJ, Duarte FAM, Swinderen BV, Cothran K, and Perel C. Quantitative trait loci for murine growth. *Genetics* 142: 1305–1319, 1996.
9. Cheverud JM, Vaughn T, Pletscher LS, Peripato A, Adams E, Erickson C, and King-Ellison K. Genetic architecture of adiposity in the cross of Large (LG/J) and Small (SM/J) inbred mice. *Mamm Genome* 12: 3–12, 2001.
10. Cheverud JM, Ehrich TH, Vaughn TT, Koreishi SF, Linsey RB, and Pletscher LS. Pleiotropic effects on mandibular morphology. II. Differential epistasis and genetic variation in morphological integration. *J Exp Zool Part B Mol Dev Evol* In press.
11. Churchill GA and Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971, 1994.
12. Diggle PJ, Heagerty P, Liang KY, and Zeger SL. *Analysis of Longitudinal Data*. Oxford, UK: Oxford University Press, 2002.
13. Ehrich TH, Vaughn TT, Koreishi S, Linsey RB, Pletscher LS, and Cheverud JM. Pleiotropic effects on mandibular morphology. I. Devel-

- opmental morphological integration and differential dominance. *J Exp Zool Zool Part B Mol Dev Evol* 296: 58–79, 2003.
14. **Kirkpatrick M and Heckman N.** A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J Math Biol* 27: 429–450, 1989.
 15. **Kirkpatrick M, Hill WG, and Thompson R.** Estimating the covariance structure of traits during growth and aging, illustrated with lactation in dairy cattle. *Genet Res* 64: 57–69, 1994.
 16. **Kirkpatrick M, Lofsvold D, and Bulmer M.** Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124: 979–993, 1990.
 17. **Klingenberg CP.** Heterochrony and allometry: the analysis of evolutionary change in ontogeny. *Biol Rev Camb Philos Soc* 73: 79–123, 1998.
 18. **Klingenberg CP, Leamy LJ, Routman EJ, and Cheverud JM.** Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics* 157: 785–802, 2001.
 19. **Klingenberg CP, Leamy LJ, and Cheverud JM.** Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. *Genetics* 166: 1909–1921, 2004.
 20. **Lander ES and Botstein D.** Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199, 1989.
 21. **Ma CX, Casella G, and Wu RL.** Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161: 1751–1762, 2002.
 22. **Mackay TFC.** Quantitative trait loci in *Drosophila*. *Nat Rev Genet* 2: 11–20, 2001.
 23. **Mogil JS, Richards SP, O'Toole LA, Helms ML, Mitchell SR, Kest B, and Belknap JK.** Identification of a sex-specific quantitative trait locus mediating nonopioid stress-induced analgesia in female mice. *J Neurosci* 17: 7995–8002, 1997.
 24. **Niklas KL.** *Plant Allometry: The Scaling of Form and Process*. Chicago, IL: University of Chicago, 1994.
 25. **North KE, Martin LJ, Dyer T, Comuzzie AG, and Williams JT.** HDL cholesterol in females in the Framingham Heart Study is linked to a region of chromosome 2q. *BMC Genet* 4, Suppl 1: S98, 2003.
 26. **Nuzhdin SV, Pasyukova EG, Dilda C, and Mackay TFC.** Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 94: 9734–9739, 1997.
 27. **Pletcher SD and Geyer CJ.** The genetic analysis of age-dependent traits: modeling the character process. *Genetics* 153: 825–835, 1999.
 28. **Pletcher SD and Jaffrezic F.** Generalized character process models: estimating the genetic basis of traits that cannot be observed and that change with age or environmental conditions. *Biometrics* 58: 157–162, 2002.
 29. **Rice SH.** The analysis of ontogenetic trajectories: when a change in size or shape is not heterochrony. *Proc Natl Acad Sci USA* 94: 907–912, 1997.
 30. **Vaughn TT, Pletscher LS, Peripato A, King-Ellison K, Adams E, Erikson C, and Cheverud JM.** Mapping quantitative trait loci for murine growth: a closer look at genetic architecture. *Genet Res* 74: 313–322, 1999.
 31. **Wang ZH and Wu RL.** A statistical model for high-resolution mapping of quantitative trait loci determining human HIV-1 dynamics. *Stat Med* 23: 3033–3051, 2004.
 32. **West GB, Brown JH, and Enquist BJ.** A general model for ontogenetic growth. *Nature* 413: 628–631, 2001.
 33. **Wolf JB.** Genetic architecture and evolutionary constraint when the environment contains genes. *Proc Natl Acad Sci USA* 100: 4655–4660, 2003.
 34. **Workman MC, Leamy LJ, Routman EJ, and Cheverud JM.** Analysis of quantitative trait locus effects on the size and shape of mandibular molars in mice. *Genetics* 160: 1573–1586, 2002.
 35. **Wu RL.** The detection of plasticity genes in heterogeneous environments. *Evolution* 52: 967–977, 1998.
 36. **Wu RL, Ma CX, Chang M, Littell RC, Wu SS, Yin TM, Huang MR, Wang MX, and Casella G.** A logistic mixture model for characterizing genetic determinants causing differentiation in growth trajectories. *Genet Res* 19: 235–245, 2002.
 37. **Wu RL, Ma CX, Littell RC, and Casella G.** A statistical model for the genetic origin of allometric scaling laws in biology. *J Theor Biol* 217, 275–287, 2002.
 38. **Wu RL, Ma CX, Lin M, and Casella G.** A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics* 166: 1541–1551, 2004.
 39. **Wu RL, Ma CX, Lin M, Wang ZH, and Casella G.** Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. *Biometrics* 60: 729–738, 2004.
 40. **Wu WB and Pourahmadi M.** Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* 90: 831–844, 2003.
 41. **Zimmerman DL and Núñez-Antón V.** Parametric modeling of growth curve data: an overview (with discussions). *Test* 10: 1–73, 2001.