Physiol Genomics 25: 458–469, 2006.
First published February 7, 2006; doi:10.1152/physiolgenomics.00181.2005.

Functional mapping for genetic control of programmed cell death

Yuehua Cui,[1,2] Jun Zhu,[3] and Rongling Wu[1]

[1]*Department of Statistics, University of Florida, Gainesville, Florida;* [2]*Department of Statistics and Probability, Michigan State University, East Lansing, Michigan; and* [3]*College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, Zhejiang, China*

Submitted 25 July 2005; accepted in final form 25 January 2006

Cui, Yuehua, and Rongling Wu. Functional mapping for genetic control of programmed cell death. *Physiol Genomics* 25: 458–469, 2006. First published February 7, 2006; doi:10.1152/physiolgenomics.00181.2005.—"Naturally occurring" or "programmed" cell death (PCD) in which the cell uses specialized cellular machinery to kill itself is a ubiquitous phenomenon that occurs early in organ development. Such a cell suicide mechanism that enables metazoans to control cell number and eliminate cells threatening the organism's survival has been thought to be under genetic control. In this report, we develop a novel statistical model for mapping specific genes or quantitative trait loci (QTL) that are responsible for the PCD process based on polymorphic molecular markers. This model incorporates the biological mechanisms of PCD that undergoes two different developmental stages, exponential growth and polynomial death. We derived a parametric approach to model the exponential growth and a nonparametric approach based on the Legendre function to model the polynomial death. A series of stationary and nonstationary models has been used to approximate the structure of the covariance matrix among cell numbers at a multitude of different times. The statistical behavior of our model is investigated through simulation studies and validated by a real example in rice.

quantitative trait loci; semiparametric model; mean-covariance structure model; EM-Simplex algorithm; order selection

THE QUESTION OF HOW AN ORGANISM develops into a fully functioning adult from a mass of undifferentiated cells has always attracted top researchers in diverse areas of developmental biology (19). Fundamentally, to produce a functioning adult form, a living organism should coordinate various complementary and sometimes antagonistic processes, which include cell proliferation and programmed cell death (PCD), or apoptosis, during its development (8). The molecular and genetic characterization of these processes has been useful to identify the specific signaling pathways that underlie a delicate balance between cell proliferation and PCD (32) and, in the ultimate, enhance our understanding of the roots of disease such as cancer (17, 43). In particular, PCD has been thought to be a universal phenomenon that occurs in predictable patterns in response to environmental or developmental clues, whose study has become one of the most fascinating areas in all of biology over the last decade (16, 19).

Studies of the genetic control of development have used simple model systems, the nematode *Caenorhabditis elegans* and the fruitfly *Drosophila*, from which views have been established that PCD involves specific genes and proteins and that these genes and proteins interact within the cells that die (13, 35, 31, 18). For the adult hermaphrodite *C. elegans* forms, there are 1,090 somatic cells, of which 131 die by apoptosis. Each apoptosis process is characterized by four stages (31): *1*) decision about whether a cell should die or assume another fate, *2*) death, *3*) engulfment of the dead cell by phagocytes, and *4*) degradation of the engulfed corpse. Each of these stages is regulated by a number of genes. Mutations affecting the final three stages affect all somatic cells, whereas genes affecting the death verdict affect very few cells.

The molecular genetic pathway that defines PCD in *C. elegans* and *Drosophila* provides a basis for understanding apoptosis in more complex organisms, including higher plants and humans, because genes responsible for PCD are evolutionarily conserved (18). Given the unique complexity of genetic pathways of these species, however, a rigorous, detailed, and analytic approach should be developed on its merit that allows for the genome-wide identification of genes for apoptosis in any complex organism. Genetic mapping based on molecular markers (23, 44, 26), by superimposing real biological phenotypes on genome sequence and structural polymorphisms, can provide an unbiased view of the network of gene actions and interactions of quantitative trait loci (QTL) that build a complex phenotype like PCD.

Unlike traditional QTL mapping for a complex trait, the mapping of PCD must incorporate the dynamic feature of this developmental process. Although this presents one of the most difficult tasks in genetic studies, some of the key issues have been overcome by Wu and colleagues (27, 37–41), who proposed a so-called "functional mapping" to map and identify specific QTL that underlie the developmental changes of complex traits. The rationale of functional mapping is to express the genotypic values of QTL at a series of time points in terms of a continuous growth function with respect to time *t*. Under this formulation, the parameters describing longitudinal trajectories, rather than time-dependent genotypic values as carried out in traditional mapping strategies, are estimated within a maximum likelihood framework. Also, unlike traditional strategies, functional mapping estimates the parameters that model the structure of the covariance matrix among a multitude of different time points, which, therefore, largely reduces the number of parameters being estimated for variances and covariances, especially when the dimension of data is high.

In this article, we develop a novel statistical model for the genome-wide scan of QTL that guide PCD toward an active process of cell death. This model incorporates two sequentially distinct stages of the developmental process into the mapping framework constructed within the context of Gaussian mixture models. The first stage, growth, has proven to obey some universal growth law that can be modeled mathematically by curve parameters (5). Although no proper mathematical equation can describe the second stage, death, which is subject to a fast exponential decay of cells followed by a slowly decreasing function (4), a nonparametric approach based on the Legendre

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, Univ. of Florida, Gainesville, FL 32611 (e-mail: rwu@stat.ufl.edu).

function is derived to model the PCD process. The combination of parametric modeling of the growth process and nonparametric modeling of the death process lays a foundation for semiparametric functional mapping of PCD. We implement a nonstationary mean-dependent covariance model to characterize the structure of the covariance matrix among cell numbers measured at a multitude of different times. The statistical behavior of our model is investigated through simulation studies. The utility of the model in a real example of rice suggests that our model can be useful in practice.

DIFFERENT PHASES OF PCD

In general, the whole process of PCD can be described by five reasonably distinct phases (Fig. 1; Ref. 15): lag, exponential, declining growth rate, a stationary phase, and death. Each of the phases is defined below.

Lag Phase

The lag phase is the initial growth phase, during which cell number remains relatively constant before rapid growth. During this phase the organism prepares to grow, and unseen biochemical changes, cell division, and differentiation of tissues occur during this time.

Exponential Phase

During the exponential phase the tissues are growing and dividing rapidly to take advantage of abundant nutrients. Growth rate, as a measure of the increase in biomass over time, is determined from the exponential phase. Growth rate is one important way of expressing the relative success of an organism in adapting to the biotic or abiotic environment imposed upon it. The duration of the exponential phase depends on the growth rate and the abundance of nutrients to support tissue growth. If the growth phase is plotted (time on *x*-axis and biomass on logarithmic *y*-axis), the exponential phase will be straightened out.

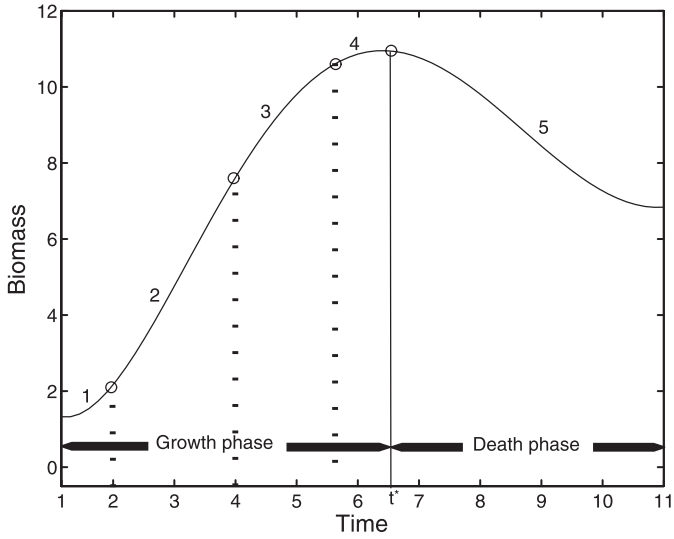


Fig. 1. A typical example of programmed cell death (PCD) that includes 5 different stages: *1*) lag, *2*) exponential, *3*) declining growth, *4*), stationary, and *5*) death.

Declining Growth

Declining growth normally occurs when either a specific requirement for cell division is limiting or something else is inhibiting reproduction. During this phase growth slows or the death rate increases. As a result, the initiation of new tissues and the senescence and death of old ones start to come into equilibrium. This phase typically occurs as nutrients become limiting for growth.

Stationary Phase

Tissues enter the stationary phase when net growth is zero, and within a matter of time cells may undergo dramatic biochemical changes. The nature of the changes depends on the growth-limiting factor. The shutdown of many biochemical pathways as the stationary phase proceeds means that the longer the cells are held in this condition the longer the lag phase will be when cells are returned to good growth conditions.

Death Phase

When cell metabolism can no longer be maintained, the death rate of a tissue is generally very rapid (4). The steepness of the decline is often more marked than that represented in the accompanying growth figure.

The duration and extent of each phase will depend on the organism and the environmental conditions. For example, if tissues from the stationary phase are supplied with fresh nutrients, the lag phase will be longer than for the case of tissues from the declining phase. For growing tissues from the exponential phase, organisms supplied with fresh nutrients will likely skip the lag phase. If the growth nutrient is rich, organisms will remain in the exponential growth phase for a longer period and produce a greater biomass. Furthermore, their rate of growth in the exponential phase may also be greater.

Growth curves must be drawn from a series of growth measurements at different times during the growth curve. Mathematical equations have been derived to model the growth from the lag to stationary phases (34), although there is no specific mathematical equation for the death phase.

STATISTICAL MODEL

The Mixture Model-Based Likelihood

Consider a standard backcross design, initiated with two contrasting homozygous inbred lines, in which there are two genotypes at each locus. Assume that a genetic linkage map covering the entire genome has been constructed with polymorphic markers, aimed to identify QTL responsible for PCD. There are a certain number of QTL forming J genotypes that affect PCD. The statistical foundation for functional mapping of these QTL is based on a finite mixture model. According to this mixture model, each PCD curve, y_i, for a backcross progeny (i) longitudinally measured at T time points is assumed to have arisen from one (and only one) of these J QTL genotypes (called components in statistics), with each being modeled by a multivariate normal distribution, i.e.,

$$y_i \alpha p(y_i|\omega,\varphi,\eta) = \omega_1 f(y_i;\varphi_1,\eta) + \ldots + \omega_J f(y_i,\varphi_J,\eta), \qquad (1)$$

where p is the mixture of multiple multivariate normal distri-

butions, each denoted by f, and $\omega = (\omega_1,\ldots,\omega_J)'$ are the mixture proportions (i.e., QTL genotype frequencies) of J QTL genotypes, which are constrained to be nonnegative and

$$\sum_{j=1}^{J} \omega_j = 1,$$

$\varphi = (\varphi_1,\ldots,\varphi_J)'$ are the QTL genotype-specific parameters, and η are parameters that are common to all QTL genotypes.

Now we use the genetic linkage map constructed by molecular markers to detect and map the underlying QTL for PCD over the entire genome. Consider such a segregating QTL with alleles Q and q in the backcross population. This QTL cannot be directly observed but rather is to be inferred on the basis of known marker information. In QTL interval mapping, we will use two flanking markers to infer the genotype of a QTL that is hypothesized to be located between the two markers (23). The recombination fraction is a linkage parameter that can describe the genetic distance between two given loci. Let θ, θ_1, and θ_2 be the recombination fractions between the two markers M_1 and M_2, between the left marker M_1 and the QTL, and between the QTL and the right marker M_2, respectively. On the basis of segregation and transmission of genes from the parent to progeny, one can derive the conditional probabilities of an unknown QTL genotype, conditional on the known marker genotypes, in terms of these recombination fractions. The unknown parameters that specify the position of QTL within a marker interval are arrayed in Ω_s, where s denotes the QTL position.

According to functional mapping (27), the mixture-based likelihood function of the longitudinal PCD trait (y) and marker information (M) collected in the backcross population at this hypothesized QTL with two genotypes (denoted by $j = 1, 0$) is constructed as

$$L(\Omega_s,\Omega_u,\Omega_v|y,M) = \prod_{i=1}^{n}[\omega_{1/i} f(y_i|\Omega_{u1},\Omega_v) + \omega_{0/i} f(y_i|\Omega_{u0},\Omega_v)], \quad (2)$$

where $\omega_{1|i}$ and $\omega_{0|i}$ contained in Ω_s are the mixture proportions corresponding to the frequencies of different QTL genotypes for a progeny i, expressed as the conditional probabilities of QTL genotypes given marker genotypes for this progeny; $\Omega_u = (\Omega_{u1}, \Omega_{u0})$ contains the parameters that model time-dependent means for genotype j; and Ω_v contains the parameters that model the structure of the residual covariance matrix that is assumed to be common to all mixtures. Different from parameters φ and η in the original mixture model (1), which are unstructured, Ω_{u_j} and Ω_v are the mathematical parameters that model the mean-covariance structure.

The multivariate normal distribution of each mixture for progeny i measured for T time points is expressed as

$$f(\boldsymbol{y}_i,M|\Omega_{u_j},\Omega_v) = \frac{1}{(2\pi)^{T/2}|\sum|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{u}_j)\sum^{-1}(\boldsymbol{y}_i - \boldsymbol{u}_j)^T\right], \quad (3)$$

where $\boldsymbol{y}_i = [y_i(1),\ldots,y_i(T)]$ is a vector of observation for progeny i and $\boldsymbol{u}_j = [u_j(1),\ldots,u_j(T)]$ is a mean vector for all the progeny with genotype j. At a particular time point (say t), the relationship between the observation and mean can be described by a linear regression model,

$$y_i(t) = \xi_i u_1(t) + (1 - \xi_i)u_0(t) + e_i(t),$$

where ξ_i is an indicator variable of progeny i for QTL genotype defined as 1 for $j = 1$ and 0 for $j = 0$ and $e_i(t)$ is the residual error which is *iid* normal with zero mean and variance $\sigma^2(t)$. The errors for progeny i at two different time points, t_1 and t_2, are correlated with covariance $\text{cov}(y_i(t_1),y_i(t_2))$. The variances and covariances comprise the covariance matrix Σ, whose elements are the common parameters specified by Ω_v. With these general settings, the statistical challenge becomes how to model the mean process and how to structure the covariance matrix.

Semiparametric Modeling of Mean Vector

In a broad sense, the entire PCD process for a particular individual i can be divided into two phases, growth and death (Fig. 1). Let t_i^* be the transition time point that marks the end of the growth phase and the beginning of the death phase. The mean vector in the multivariate normal distribution of PCD (*Eq. 3*) for individual i that carries QTL genotype j can now be specified by two mean subvectors expressed as

$$u_{j|i} = (\boldsymbol{u}_{Gj|i},\boldsymbol{u}_{Dj|i}) \quad (4)$$

where $\mathbf{u}_{Gj|i}$ and $\mathbf{u}_{Dj|i}$ correspond to the growth and death vectors before and after t_i^*, respectively.

Parametric model of growth phase. The process of growth (before t_i^*) follows universal growth laws and can be described by biologically meaningful mathematical functions. As a nearly universal biological law for living systems, the sigmoidal (or logistic) growth function can be fitted to capture age-specific change during the growth phase (34). The logistic function is mathematically described for individual i by

$$g_i(t) = \frac{a_i}{1 + b_i e^{-c_i t}}, \text{ with } t \approx [1,t_i^*] \quad (5)$$

where a_i is the asymptotic or limiting value of g_i when $t \to \infty$, $a_i/(1 + b_i)$ is the initial value of g_i when $t = 0$, and c_i is the relative rate of growth (5). Thus with these three parameters, one can uniquely determine the shape of PCD in the growth phase for individual i that carries QTL genotype j and have the time-dependent mean vector for *Eq. 4* specified by

$$\boldsymbol{u}_{Gj|i} = [\boldsymbol{u}_{Gj|i}(1),\ldots,\boldsymbol{u}_{Gj|i}(t_i^*)] = \left[\frac{a_j}{1 + b_j e^{-c_j}},\ldots,\frac{a_j}{1 + b_j e^{-c_j t_i^*}}\right] \quad (6)$$

If different genotypes at a putative QTL have different combinations of the parameters (a_j, b_j, c_j), this implies that this QTL plays a role in governing the difference of growth trajectories.

Nonparametric model of death phase. Because no particular mathematical function can be used to describe the death phase (after t_i^*), the nonparametric approach based on the orthogonal Legendre polynomial (LEP) is used. The flexibility of LEP will greatly increase the robustness of functional mapping.

With appropriate order r, the time-dependent genotypic values for different QTL genotypes in the death phase can be

fitted by the orthogonal LEP. A family of such polynomials with normalized time t' is denoted by

$$P(t') = [P_0(t'), P_1(t'), \ldots, P_r(t')]^T$$

and a vector of genotypic-related, time-independent values with order r is denoted by

$$\boldsymbol{v}_j = [v_{0j}, v_{1j}, \ldots, v_{rj}]^T$$

where $\mathbf{v}_j$ is called the base genotypic vector for QTL genotype j and the parameters within the vector are called the base genotypic means. The normalized time t' is obtained by adjusting the original measurement time t to match the orthogonal function range $[-1, 1]$, by

$$t' = -1 + \frac{2(t - t_{\min})}{t_{\max} - t_{\min}}$$

where $t_{\min}$ and $t_{\max}$ are the first and last time point, respectively.

With these specifications, the time-dependent genotypic values, $\boldsymbol{u}_{Dj|i(t)}$ in the death phase can be described as a linear combination of $\boldsymbol{v}_j$ weighted by series of the polynomials, i.e.,

$$\boldsymbol{u}_{Dj}(t) = \boldsymbol{v}_j^T P(t') \quad (7)$$

Thus for individual i whose QTL genotype is j, we use the following expression to model genotype means in the death phase

$$\boldsymbol{u}_{Dj|i} = [\boldsymbol{u}_{Dj|i}(t_i^*), \ldots, \boldsymbol{u}_{Dj|i}(T)] \quad (8)$$

This approach has great flexibility in modeling longitudinal data that cannot be fitted by a parametric model. By choosing an appropriate order, the nonparametric model can better capture the intrinsic pattern of developmental PCD. The number of parameters can be reduced if the order of the polynomial should be less than the number of time points.

Modeling Covariance Structure

To model the covariance structure for longitudinal data, we need to make the following assumptions: *1*) the error $e_i(t)$ in *Eq. 1* is normally distributed with mean zero and variance $\sigma^2(t)$, and *2*) the error $e_i(t)$ is independent and identically distributed among different individuals. A number of statistical models have been used to model the covariance structure (12). In earlier functional mapping, the first-order autoregressive [AR(1)] model was used (27), which is expressed as

$$\sigma^2(1) = \ldots = \sigma^2(T) = \sigma^2 \quad (9)$$

for the variance, and

$$\sigma(t_1, t_2) = \sigma^2 \rho^{|t_2 - t_1|} \quad (10)$$

for the covariance between any two time intervals t_1 and t_2, where $0 < \rho < 1$ is the proportion parameter with which the correlation decays with time lag. The parameters that model the structure of the (co)variance matrix are arrayed in Ω_v.

To remove the heteroscedastic problem of the residual variance, which violates a basic assumption of the simple AR(1) model, two approaches can be used. The first approach is to model the residual variance by a parametric function of time, as originally proposed by Pletcher and Geyer (29). However, this approach must implement additional parameters for characterizing the age-dependent change of the variance. The second approach is to embed Carroll and Ruppert's (7) transform-both-sides (TBS) model into the growth-incorporated finite mixture model (40), which does not need any more parameters. Both empirical analyses with real examples and computer simulations suggest that the TBS-based model can increase the precision of parameter estimation and computational efficiency. Furthermore, the TBS model preserves original biological means of the curve parameters, although statistical analyses are based on transformed data.

The TBS-based model displays the potential to relax the assumption of variance stationarity, but the covariance stationarity issue remains unsolved. Zimmerman and Núñez-Antón (47) proposed a so-called structured antedependence (SAD) model to model the age-specific change of correlation in the analysis of longitudinal traits. The SAD model has been employed in several studies and displays many favorable properties (48). Zhao et al. (46) incorporated the first-order SAD [SAD(1)] model into modeling of the covariance matrix.

In this article, we use a different modeling approach that is as simple as the AR(1) and as flexible as the SAD(1). This approach has two steps. In *step 1*, the intraindividual correlation structure is modeled. In many cases, a systematic pattern of correlation is evident, which may be characterized accurately by a relatively simple model. The intraindividual correlation among the time-dependent elements of e_i for individual i is assumed to follow a pattern, expressed as

$$\mathrm{corr}(e_i) = \boldsymbol{R}_i(\varphi)$$

where the correlation matrix $\mathbf{R}_i(\varphi)$ is a function of a vector of correlation parameters φ. The correlation structure can be described by the AR(1) model in which $\varphi = \rho$ (*Eq. 10*).

In *step 2*, time-dependent variances are specified according to Horwitz's rule in analytical chemistry. This rule proposes that there exists an empirical relationship between concentration and variance (1). Thus we can similarly model time-dependent variances for individual i by considering its genotypic means at various time points, expressed as

$$\sigma_i^2(t) = \sigma^2 u_{j|i}^2(t)$$

Therefore, the corresponding covariance matrix can be modeled by

$$\Sigma_{j|i} = \mathrm{cov}(y) = \sigma^2 u_{j|i}^{\frac{1}{2}} \mathbf{R}_i(\varphi) u_{j|i}^{\frac{1}{2}} \quad (11)$$

where $u_{j|i} = \mathrm{diag}[u_{j|i}^2(1), \ldots, u_{j|i}^2(T)]$. Because the covariance structure is modeled as a function of genotypic mean, it is called the mean-covariance (M-C) model. The M-C model has great flexibility for modeling the covariance matrix of the PCD process. The unknown parameters to be estimated in the M-C model are arrayed in $\Omega_v = (\sigma^2, \varphi)$ with accepted means.

Computation Algorithms

The EM algorithm (10) has served as a standard approach for obtaining the maximum likelihood estimates (MLEs) of the parameters in traditional QTL mapping (23). This algorithm has also been used for functional mapping of longitudinal traits to obtain the MLEs of $(\Omega_s, \Omega_u, \Omega_v)$ for the likelihood (*Eq. 2*) (27, 40). The EM algorithm is implemented with two steps: *1*) the E step, in which the posterior probability of a QTL

genotype given the marker genotype of progeny i is calculated using

$$\Pi_{j|i} = \frac{\omega_{1|i} f(y_i|\Omega_{u_j},\Omega_v)}{\omega_{1|i} f(y_i|\Omega_{u_1}\Omega_v) + \omega_{0|i} f(y_i|\Omega_{u_0},\Omega_v)}$$

and *2*) the M step, in which the calculated $\Pi_{j|i}$ values are used to solve the log-likelihood equations, aimed to estimate (Ω_s, Ω_{u_j}, Ω_v) defining the QTL position, the QTL genotype-specific curve parameters, and the parameters that model the covariance matrix, respectively.

An iterative procedure between the E and M steps will be processed until convergence.

The estimates at convergence are regarded as the MLEs of the unknown parameters. In practice, the QTL position parameter Ω 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.

In functional mapping for PCD, $\Omega_u = (\Omega_G,\Omega_D)$ and Ω_v are contained in complex nonlinear equations, and therefore it is difficult to derive a closed form for their MLEs. The Nelder-Mead simplex algorithm as a direct search method for nonlinear unconstrained optimization, originally proposed by Nelder and Mead (28), can be used to estimate these parameters (45). This algorithm attempts to minimize a scalar-valued nonlinear function using only function values, without any derivative information (explicit or implicit). The algorithm uses linear adjustment of the parameters until some convergence criterion is met.

However, because of the complex nonlinear function being minimized by simplex algorithm, it cannot always guarantee the correct convergence of covariance parameters during the minimization process. This consequently results in negative infinity of the log-likelihood function, and convergence will never be reached. Because of these concerns, we used the simplex algorithm to estimate the mean parameters, namely, the logistic curve and Legendre polynomial parameters, and the EM algorithm to estimate the parameters that model the structure of the covariance matrix (see APPENDIX).

Under the joint modeling framework, two mean functions, growth and death, must be connected. Two constraints are imposed to make the PCD curve continuous at the transition time point t_i* for each individual. The first constraint is to make the growth mean equal to the death mean at time t_i*. The second constraint is that the two functions have the same score at time t_i*. These two constraints are expressed as

$$\begin{cases} u_{Gj|i}(t_i^*) = u_{Dj|i}(t_i^*) \\ \dfrac{\partial}{\partial t_i^*} u_{Gj|i}(t_i^*) = \dfrac{\partial}{\partial t_i^*} u_{Dj|i}(t_i^*) \end{cases}$$

With these constraints, we obtain the expressions of one growth parameter and one death parameter for any QTL genotype j. For example, if the Legendre polynomial order is 3, we can solve the equations to obtain the estimates of a_j and v_{1j} as follows,

$$\begin{cases} a_j = \dfrac{2[v_{0j} - 0.5(3t'^2_i + 1)v_{2j} - 5t'^3_i v_{3j}](1 + b_j e^{-c_j t_i^*})^2}{1 + b_j e^{-c_j t_i^*} - b_j c_j (\Delta t) t_i' e^{-c_j t_i^*}} \\ v_{1j} = \dfrac{a_j b_j c_j (\Delta t) e^{-c_j t_i^*}}{2(11 + b_j e^{-c_j t_i^*})} - 3t_i' v_{2j} - 0.5(15t'^2_i - 3)v_{3j} \end{cases}$$

where t_i' is the adjusted time for t_i and $\Delta t = t_{max} - t_{min}$.

It is possible that the algorithm described above may generate local maxima for the likelihood surface. An empirical approach for reducing the possibility of local maxima is to use multiple sets of initial values of the parameters. The initial values are determined in the light of parameter estimates from the data by assuming that no QTL is involved. We will obtain the global maxima when no further increase of the likelihood is found in a space of parameters.

Legendre Order Selection

To determine which order of the LEP best fits the data, we must select the optimal order. One of the popular model selection criteria is the Akaike information criterion (AIC) (1). The AIC value at a particular order r is calculated by

$$\text{AIC} = -2\ln L(\hat{\Omega}_G,\hat{\Omega}_D,\hat{\Omega}_v|r) + 2 \text{ dimension } (\Omega_G,\Omega_D,\Omega_v|r), \quad (12)$$

where $\hat{\Omega}_G = \{\hat{a}_j,\hat{b}_j,\hat{c}_j\}\}_{j=0}^{1}$ and $\hat{\Omega}_D = \{\hat{u}_{0j},\hat{b}_{1j}, \ldots, \hat{u}_{rj}\}\}_{j=0}^{1}$ are the MLEs of parameters for the growth curve function and the Legendre polynomial of order r, Ω_v contains the MLEs of the covariance parameters, and dimension($\Omega_G,\Omega_D,\Omega_v|r$) represents the number of free parameters under order r. The optimal order is one that displays the minimum AIC value.

Another model selection criterion to determine the optimal order of the Legendre function is the Bayesian information criterion (BIC) (30), which is calculated by

$$\text{BIC} = -2\text{lnL}(\hat{\Omega}_G,\hat{\Omega}_D,\hat{\Omega}_v|\text{r}) + \text{dimension } (\Omega_G,\Omega_D,\Omega_v|\text{r})\ln(\text{n}), \quad (13)$$

where all the parameters are defined similarly as above except that n is the total number of observations at a particular time point. Because the BIC adjusts the effect of sample size, the model selected by the BIC will be more parsimonious. Other criteria, such as those proposed for high-dimension parametric models (25), can also be used.

Calculating Curve Heritability

It is easy to calculate the heritability level (H^2) when traits are measured at a single time point, but for longitudinal traits heritability calculation is difficult. We propose two ways to do it: *1*) Calculate H^2 at a single time point t where the traits show the highest variation. For a backcross design, the genetic variation is given by $\sigma_G^2 = \frac{1}{4}[u_1(t) - u_0(t)]$, where $u_j(t)$ (j = 1,0) is the genetic mean for genotype j at time t, and the heritability $H^2(t) = \sigma_G^2(t)/\sigma_\epsilon^2(t)$ is calculated, where $\sigma_\epsilon^2(t)$ is the residual variance at time t.

2) Calculate H^2 with the area under curve (AUC). Functional mapping maps the dynamic gene effect over time. The genetic variation explained by the entire measurement period is more informative than that by individual time point. We propose to calculate the genetic variation by $\sigma_G^2 = \frac{1}{4}(AUC_1 - AUC_0)^2$,

where AUC_1 and AUC_0 are the AUC for two different genotypes. The AUC is calculated by

$$AUC_j = \int_{t_1}^{t^*} \frac{a_j}{1 + b_j e^{c_j}} dt + \int_{t_1}^{t_T'} \mathbf{P}_r(t') v_j dt'$$

$$= \frac{a_j}{c_j} [\ln(b_j + e^{c_j t^*}) - \ln(b_j + e^{c_j t_1})] + \int_{t'^*}^{t_T'} \mathbf{P}_r(t') v_j dt'$$

where t^* is the transition time point, a_j, b_j, and r_j are the growth parameters for genotype j, $\mathbf{P}_r(t')$ is a vector of LEP with order r, u_j is the base genotypic mean parameters, and t' is the adjusted time point.

Hypothesis Testing

One of the major advantages of functional mapping is that it allows for a number of hypothesis tests to examine the genetic control mechanisms of growth throughout development and in response to varying environmental or developmental clues. Wu et al. (39) have formulated some of these hypothesis tests, which include the global test of genetic effects on the entire developmental process, the regional test of genetic control over a particular developmental period of interest, and the point test for the timing of developmental events. From a genetic perspective, we can also test how different genetic action modes play a role in regulating the developmental process. With such a complete set of tests, we are able to address biological questions related to the genetic control mechanisms of PCD traits.

Testing whether specific QTL exist to affect the PCD process is a first step toward the understanding of the detailed genetic architecture of complex phenotypes. The genetic control over the entire developmental process of PCD can be tested by formulating the following hypotheses:

$$\begin{cases} H_0\text{: } u_{G1} = u_{G0}, u_{D1} = u_{D0} \\ H_{1a}\text{: at least one of the above equalities does not hold} \end{cases} \quad (14)$$

H_0 states that there is no QTL affecting the dynamic PCD process (the reduced model), whereas H_{1a} proposes that such a QTL does 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 as given below:

$$LR = -2[\log L(\tilde{\Omega}|y,M) - \log L(\hat{\Omega}|y,M)], \quad (15)$$

where $\tilde{\Omega}$ and $\hat{\Omega}$ denote the MLEs of the unknown parameters under H_0 and H_{1a}, respectively. The critical threshold value for declaring the presence of QTL can be empirically calculated based on the permutation tests (9).

Other hypotheses can be made to test whether the detected QTL only controls the growth phase with the following alternative hypothesis:

$$H_{1b}: u_{G1} \neq u_{G0} \text{ and } u_{D1} = u_{D0} \quad (16)$$

or whether the detected QTL only controls the death phase with the following alternative:

$$H_{1b}: u_{G1} = u_{G0} \text{ and } u_{D1} \neq u_{D0} \quad (17)$$

The critical thresholds for the above two hypotheses can be determined with simulation studies. Only when both H_{1b} and H_{1c} are rejected, is the detected QTL thought to pleiotropically affect the growth and death phases.

The proposed model can be used to test the influence of QTL on growth in different stages of development, lag, exponential, declining growth, stationary phase, and death. These tests can be based on the AUC during a time course of interest. Simulation studies are used to determine the critical thresholds for each test.

RESULTS

A Worked Example

We use the proposed model here to analyze a real data set of rice. Two inbred lines, semidwarf IR64 and tall Azucena, were crossed to generate an F_1 progeny population. By doubling haploid chromosomes of the gametes derived the heterozygous F_1, a doubled haploid (DH) population of 123 lines was founded (20). Such a DH population is equivalent to a backcross population because its marker segregation follows 1:1. With 123 DH lines, Huang et al. (20) genotyped 135 RFLP and 40 isozyme and RAPD markers to construct a genetic linkage map, based on the Kosambi function, of length 2,005 cM with an average distance of 11.5 cM, representing a good coverage of 12 rice chromosomes.

The 123 DH lines and their parents, IR64 and Azucena, were planted in a randomized complete design with two blocks. Each block was divided into different plots, each containing eight plants per line. Starting from 10 days of transplanting, tiller numbers were measured every 10 days for five central plants in each plot until all lines had headed. We used the means of the two blocks for QTL analysis.

Figure 2 illustrates the dynamics of tiller numbers for each DH line measured at 9 time points. Tiller growth is thought to be an excellent example of PCD in plants (16) because it experiences several developmental stages when rice grows. At an early stage, tiller numbers increase dramatically, corresponding to the vegetative phase in rice. During the reproductive phase, the increase of tiller numbers declines with the initiation of the panicle, the emergence of the flag leaf (the last

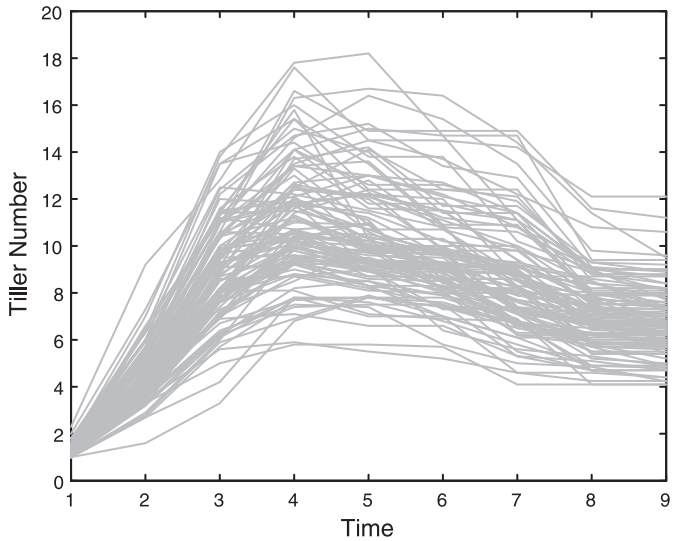


Fig. 2. Dynamic changes of the number of tillers for 123 doubled haploid lines of rice as an example of PCD in plants.

leaf), and booting, heading, and flowering of the spikelets. Tillers that do not bear panicles are called ineffective tillers and will be killed, leading to the death phase. The number of ineffective tillers is a closely examined trait in plant breeding because they are undesirable for commercial varieties. Ineffective tillers result in many unwanted problems in rice such as the overconsumption of nutrition and competition of space. The genetic control system plays an important role in reducing overproduced tillers and balancing the rice metabolism system for optimal use efficiency of nutrients.

Our semiparametric model was used to map specific QTL that determine the dynamic changes of tiller number during ontogeny. Although the growth phase of tiller number can be well modeled by a logistic equation defined by parameters *a*, *b*,and *c* (*Eq. 5*), no proper equation can be used to model the death phase. For this reason, a nonparametric approach based on the Legendre polynomial function is adopted in the framework of QTL mapping. However, this encounters the issue of order determination. To detect the best order of the Legendre polynomial function for this rice data set, we calculated the AIC and BIC for various orders (Table 1). Both criteria provide the consistent result that the death phase of tiller number can be best explained by a Legendre polynomial of order 3.

By genomewide scanning for QTL at every 2 cM within each marker interval across 12 rice chromosomes, our model identified three major QTL that trigger their effects on the overall PCD process of tiller number. As shown by the genome-wide log-likelihood ratio (LR) profile in Fig. 3, these three QTL are located between markers RG146 and RG345 and between markers RZ730 and RZ801 on chromosome 1 and on marker RZ792 on chromosome 9. Of these three detected QTL, the first is significant genome-wide, whereas the other two are significant chromosome-wide, all at the 5% significance level based on the critical thresholds determined from the permutation tests.

To know more about the behavior of the detected QTL, we tabulated the MLEs of curve parameters that specify the growth phase and genotypic basis effects that specify the death phase, along with the approximate standard errors of the estimates (Table 2). All the parameters that specify the growth and death phases for different QTL genotypes ($j = 1$ for *QQ* or 0 for *qq* in the DH population) are estimated with reasonably high precision as shown by the standard errors, although the estimation precision tends to be better for the growth parameters than for the death parameters (Table 2). The parameters that model the structure of the covariance matrix based on the M-C model can also be well estimated, suggesting good behavior of our model.

Table 1. *Model selection for Legendre polynomial orders based on AIC and BIC values under M-C covariance-structuring model*

	Order				
Selection Criterion	0	1	2	3	4
AIC	2437.8	1096	759.74	**559.87**	670.65
BIC	2453.9	1109.4	775.77	**578.58**	692.03

AIC, Akaike information criterion; BIC, Bayesian information criterion; M-C, mean-covariance. The minimum values for AIC and BIC are shown in boldface.

Using the MLEs of parameters for the growth and death phases, we draw the developmental trajectories of tiller number for the two different QTL genotypes (Fig. 4). Each QTL shows a unique developmental pattern over time. For example, the dynamic process of genetic effects for the QTL located between markers RZ730 and RZ801 on chromosome 1 is different from those for the other two QTL. Statistical tests based on *Eqs. 16* and *17* show that the QTL detected between markers RG146 and RG345 on chromosome 1 and on marker RZ792 on chromosome 9 merely control the growth phase, whereas the second QTL on chromosome 1 controls the entire developmental process ($P < 0.05$).

Simulation

We performed a series of simulation studies to examine the statistical properties of the model. Six equidistant markers are simulated from a backcross population and are ordered as M_1–M_6 on a linkage group with a length of 100 cM. The Haldane map function was used to convert the map distance into the recombination fraction. Different heritability levels ($H^2 = 0.1$ vs. 0.4) and different sample sizes ($n = 100$ vs. 200) were considered in the simulation study to examine the model's performances under different scenarios. The putative QTL is located between markers M_3 and M_4, at 48 cM from the first marker. Data are simulated by assuming that the QTL controls the entire developmental process. The simulated data have nine continuous time points. The means at different time points used to model the covariance matrix based on the M-C model are the average of the two genotypic means.

Table 3 lists the results from the simulation; the true parameters are given in the first column. In general, our model can provide reasonable estimates of the QTL positions and the growth and death parameters determined by the QTL, with estimation precision depending on heritability level and sample size. In all cases of different sample sizes and heritabilities, the maximum values of the LR landscapes from 100 simulation replicates are beyond the critical thresholds at the $\alpha = 0.001$ level determined from 1,000 permutation tests for the simulated data, suggesting that our model has 100% power to detect QTL in these conditions. The precision of parameter estimation is evaluated in terms of the square root of the mean squared errors (SMSE) of the MLEs. The QTL positions and effects can be better estimated when the PCD trait has a higher than lower heritability or when the sample size is larger rather than smaller (Table 3). However, the increase of H^2 from 0.1 to 0.4 leads to more significant improvement for the estimation precision than the increase of *n* from 100 to 200. For example, the SMSE of the growth parameter c_0 for QTL genotype *qq* reduces by more than onefold when H^2 is increased from 0.1 to 0.4 for a given sample size, whereas such reduction is much smaller when *n* is increased from 100 to 200 for a given heritability. This suggests that in practice it is more important to manage experiments to reduce residual errors (and therefore increase H^2) than to simply increase sample size.

DISCUSSION

The growth of any tissue, whether normal or malignant, is determined by the quantitative relationship between the rate at which cells proliferate and the rate at which cells die. Depending on how the rate of cell proliferation is compromised or

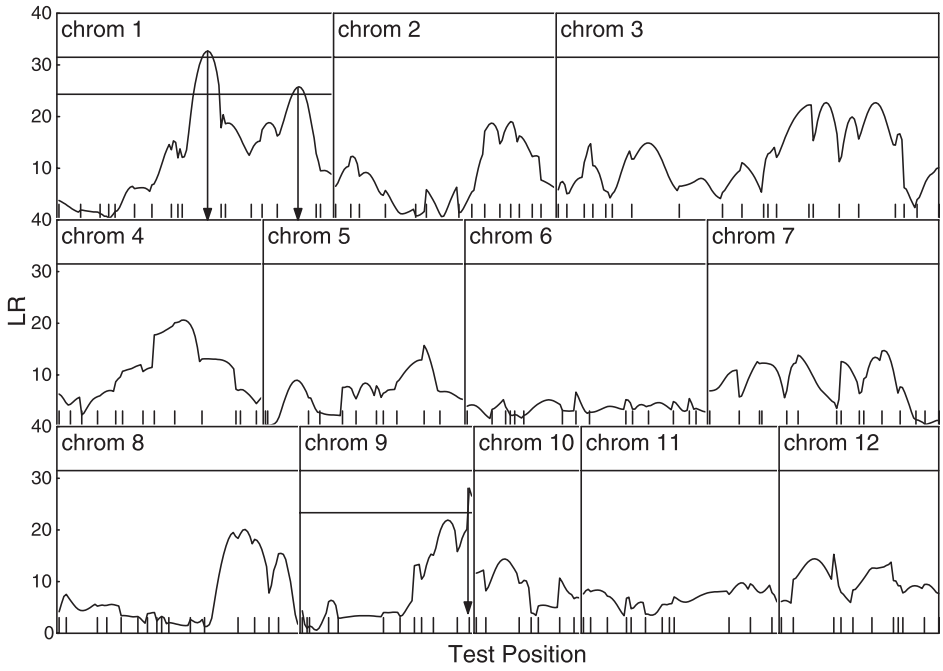


Fig. 3. The profile of the log-likelihood ratios (LR) between the full [there is a quantitative trait locus (QTL)] and reduced (there is no QTL) models for tiller number trajectories across the 12 rice chromosomes. The genomic positions corresponding to the peak of the curve are the maximum likelihood estimates of the QTL localization (indicated by arrows). The threshold value for claiming the existence of QTL is given as the horizontal dotted line for the genome-wide level and the dashed line for the chromosome-wide level. The positions of markers on the linkage groups (20) are indicated at ticks.

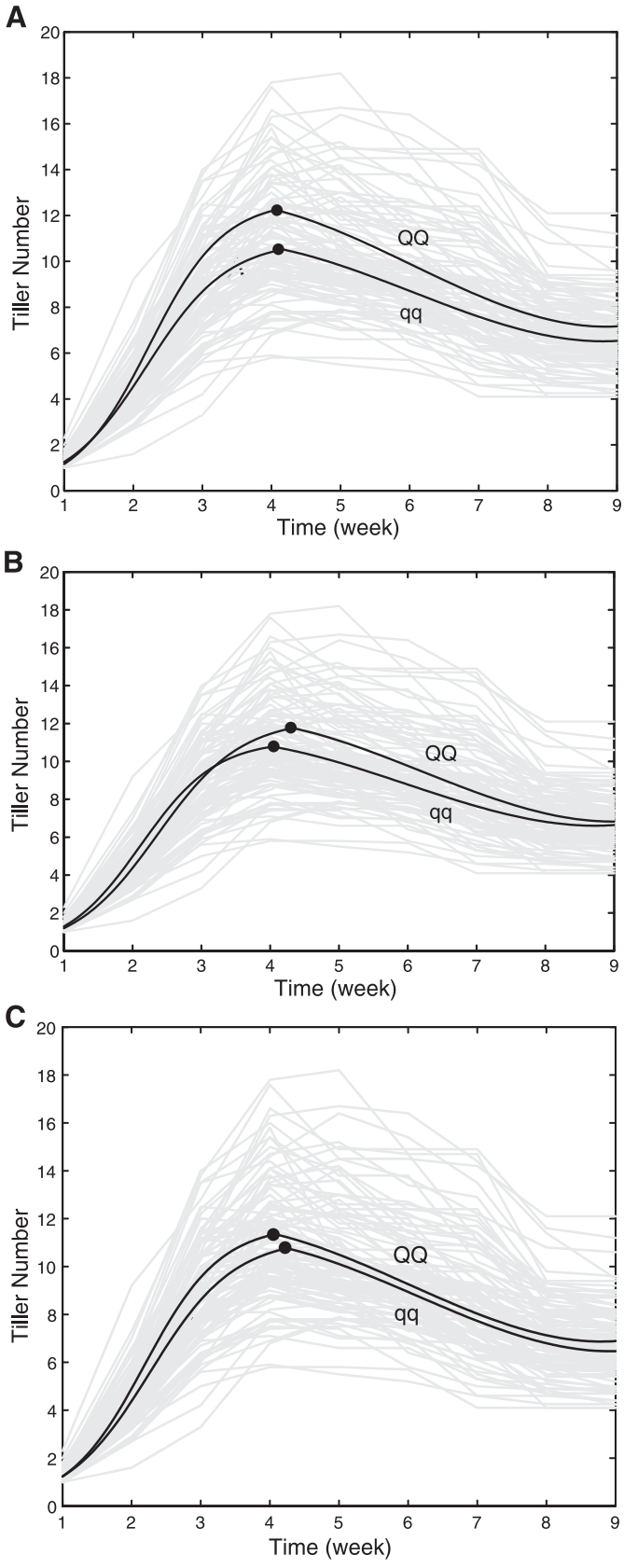
coordinated with the rate of cell death, all tissues will undertake two distinct processes, growth or death, throughout their development (21). Unlike the detailed framework for cell proliferation, understanding of the initiation of cell death and the cellular mechanics of this process is still in its infancy. Cell death can occur as an active and orderly process of development, a process described by the term programmed cell death (PCD) (19).

PCD, also referred to as apoptosis, appears to be a universal feature of animal development, and abnormalities in PCD have been associated with a broad variety of human diseases, including certain cancers and neurodegenerative disorders (17). In plants, PCD is also ubiquitous for essential development and survival, including xylogenesis, reproduction, senescence, and pathogenesis (16). To make PCD control and execution efficient, particular genetic mechanisms should be involved in regulating and modulating this process in response to various developmental and environmental stimuli. The use of organisms with simple structure like the nematode *C. elegans* and *Drosophila* has led to the identification of numerous genes responsible for PCD (13, 18, 19. 43). The 2002 Nobel Prize in Physiology or Medi-

Table 2. *MLEs of growth and death parameters for QTL genotypes QQ and qq at significant QTL detected on different chromosomes for tiller numbers in a doubled haploid rice population*

Parameters	Chromosome 1	Chromosome 2*	Chromosome 9*
QTL position, cM	120	198	119.1
Marker interval	RG146–RG345	RZ730–RZ801	RZ792
Growth parameters			
a_0	10.5973	11.0824	10.7188
b_0	7.4938 (0.1986)	7.5167 (0.2757)	7.8547 (0.3538)
c_0	1.7257 (0.0230)	1.8486 (0.0276)	1.6747 (0.0365)
a_2	12.3823	11.5461	11.5739
b_2	9.6068 (0.3759)	8.9779 (0.3390)	8.3427 (0.3948)
c_2	1.8834 (0.0321)	1.661 (0.0296)	1.8311 (0.0376)
Death parameters			
u_{00}	8.5001 (0.2190)	8.6888 (0.3173)	8.5837 (0.3191)
u_{10}	−2.4290	−2.7785	−2.4647
u_{20}	0.1008 (0.0529)	0.13257 (0.0713)	0.0989 (0.0668)
u_{30}	0.5118 (0.0415)	0.5685 (0.0545)	0.5312 (0.0528)
u_{02}	9.6225 (0.3701)	9.2706 (0.3111)	9.1128 (0.3525)
u_{12}	−3.2068	−2.6222	−2.8732
u_{22}	0.1472 (0.0829)	0.0964 (0.0669)	0.1030 (0.0706)
u_{32}	0.6575 (0.0670)	0.5765 (0.0531)	0.5784 (0.0569)
Covariance parameters			
σ^2	0.0493 (0.0051)	0.0508 (0.0038)	0.0535 (0.0051)
ρ	0.8747 (0.0120)	0.8738 (0.0055)	0.8786 (0.0112)

Values are maximum likelihood estimates (MLEs; with approximate SE in parentheses) for quantitative trait locus (QTL) genotypes *QQ* (subscript 2) and *qq* (subscript 0) on chromosomes 1, 2, and 9. *Chromosome-wide significant QTL. See text for definitions.



cine was awarded to three scientists because of their discoveries of the genetic regulation of PCD (19).

Although the use of simple model systems can provide a wealth of information about the genetics of PCD for more complicated organisms like animals and higher plants, some key questions cannot be addressed well without a direct use of these organisms. Recent development of high-throughput molecular technologies has made it possible to generate a massive amount of genetic and genomic data for any organism almost without limit. This thus presents a pressing need for the development of vigorous, detailed, and analytical approaches to unravel the genetic control and regulation mechanisms that underlie the PCD process. In this article, we have for the first time developed a statistical model that can make a systematic scan of QTL for PCD across the entire genome with a well-covered genetic linkage map. This model has been validated by a real example for the PCD process of tiller number in rice. Three QTL were detected to affect tiller number trajectories during a growing season in the field. The locations for two of the QTL detected on chromosome 1 are consistent with those estimated from basic interval mapping of single traits (42), but the third QTL detected on chromosome 9 was previously undetected by interval mapping. It seems that our model has not been able to detect many other QTL detected by Yan et al. (42), which may be due to the difference in the threshold criteria used to claim the existence of a QTL between the two approaches.

The rationale for this PCD mapping model is similar in spirit to that established in earlier functional mapping of growth curves (27, 39–41, 46). Both types of models for functional mapping integrate the mathematical aspects of biological principles into the framework for QTL mapping constructed on the basis of Gaussian mixture models. Their significant advantages compared with traditional genetic mapping models (23) lie in increased model flexibility, stability, and statistical power for QTL detection (reviewed in Ref. 36). This is because functional mapping only needs to estimate a few number of mathematical parameters that define the curve, rather than estimating many genotypic means at all time points. Functional mapping implemented with PCD can provide a quantitative platform for testing the interplay between gene actions and PCD processes. The statistical power of functional mapping is further increased by taking advantage of structuring the covariance matrix with a much lower number of parameters.

It can also be seen that the PCD mapping model proposed here is different from earlier published functional mapping approaches purely based on parametric modeling. The present model splits the PCD process into two sequentially distinct phases—growth and death. As in a parametric model, universal growth laws (34) are used to model the growth phase, whereas nonparametric modeling is performed for the death phase. This combination of parametric and nonparametric models, referred to as semiparamet-

Fig. 4. Two curves for the dynamic changes of tiller numbers, each presenting two groups of genotypes, *QQ* and *qq*, at each of the three QTL, detected between markers RG146 and RG345 (*A*) and between markers RZ730 and RZ801 (*B*) on chromosome 1 and on marker RZ792 on chromosome 9 (*C*). Tiller number trajectories for all the individuals studied are indicated in shaded background.

Table 3. *MLEs of QTL position and model parameters derived from 100 simulation replicates*

	$H^2 = 0.1$		$H^2 = 0.4$	
True Parameters	$n = 100$	$n = 200$	$n = 100$	$n = 200$
QTL position				
$s = 48$ *cM*	46.222 (4.0302)	45.98 (3.1937)	46.02 (3.268)	46.06 (2.6907)
Growth parameters				
$a_2 = 15.033$				
$b_2 = 8.324$	8.3705 (0.3109)	8.3317 (0.2207)	8.3404 (0.1218)	8.3211 (0.0911)
$c_2 = 1.814$	1.8085 (0.0354)	1.812 (0.0239)	1.8124 (0.0125)	1.813 (0.0097)
$a_0 = 10.926$				
$b_0 = 7.602$	7.6648 (0.4536)	7.6559 (0.3050)	7.5957 (0.1809)	7.6274 (0.1348)
$c_0 = 1.522$	1.5442 (0.0588)	1.5255 (0.0294)	1.5219 (0.0211)	1.5224 (0.0126)
Death parameters				
$u_{02} = 9.817$	9.8818 (0.4282)	9.8442 (0.2480)	9.8363 (0.1515)	9.8257 (0.1001)
$u_{12} = -6.453$				
$u_{22} = -0.366$	−0.3623 (0.0952)	−0.3692 (0.0609)	−0.3661 (0.0346)	−0.3673 (0.0257)
$u_{32} = 0.958$	0.9695 (0.0841)	0.9554 (0.0532)	0.9589 (0.0326)	0.9565 (0.0231)
$u_{00} = 7.893$	8.1185 (0.4238)	7.9525 (0.2909)	7.8879 (0.1295)	7.9137 (0.1204)
$u_{10} = -3.681$				
$u_{20} = -0.208$	−0.2127 (0.0932)	−0.2164 (0.0584)	−0.2056 (0.0343)	−0.2109 (0.024)
$u_{30} = 0.625$	0.6461 (0.0812)	0.6277 (0.0605)	0.6247 (0.032)	0.6260 (0.0244)
Covariance parameters				
$\sigma^2 = 0.1194$	0.1096 (0.0161)	0.1167 (0.0099)		
$\sigma^2 = 0.0199$			0.0198 (0.002)	0.0197 (0.0015)
$\rho = 0.85$	0.8415 (0.0187)	0.8475 (0.0114)	0.8484 (0.0161)	0.8478 (0.0113)

Values are MLEs (square roots of mean square errors in parentheses). The location (s) of the putative QTL is described by the map distances (in cM) from the first marker of the linkage group (100 cM long). H^2, heritability level.

ric modeling, is aimed at overcoming the problem of a death phase that cannot be described mathematically. Although it is feasible for a nonparametric approach to model any form of curve including the entire growth-death curve as in Fig. 1, this approach does not make full use of biological information contained in the growth phase and, therefore, is likely to lose the advantages of parametric functional mapping in the biological interpretation of results (see Ref. 39).

The nonparametric part of our semiparametric model is based on Legendre polynomial approaches. As shown by Kirkpatrick and Heckman (22), Legendre polynomials have several favorable properties for curve fitting which include *1*) the functions are orthogonal; *2*) they are flexible to fit sparse data; *3*) higher orders are estimable for high levels of curve complexity; and *4*) computation is fast because of good convergence. Other nonparametric regression methods using kernel estimates have been considered for the mean structure of growth curve data by Altman (2), Boularan et al. (6), Wang and Ruppert (33), and Ferreira et al. (14).

Relative to nonparametric modeling of the mean structure, nonparametric covariance modeling has received little attention. Leonard and Hsu (24) derived a Bayesian approach for nonparametric estimates of the covariance structure. Diggle and Verbyla (11) used kernel-weighted local linear regression smoothing of sample variogram ordinates and of squared residuals to provide a nonparametric estimator for the covariance structure without as assuming stationarity. It is appealing to incorporate these nonparametric or semiparametric approaches into our functional mapping framework to increase the model's flexibility.

APPENDIX

The MLEs of the parameters $\Omega = (\Omega_u, \Omega_v)$ are derived as follows, with the symbol * denoting the estimates of parameters from the previous step. The values of $(\Omega_u^* \Omega_v^*)$ estimated from the following equations will be used to provide new estimators of (Ω_u, Ω_v) in the next step.

The first derivative of the log-likelihood function in *Eq. 2* with respect to specific parameter Ω_λ is given by

$$\frac{\partial}{\partial \Omega_\lambda} \log L(\Omega|y) = \sum_{i=1}^{n} \sum_{j=0}^{1} \frac{\omega_{j|i} \dfrac{\partial}{\partial \Omega_\lambda} f_j(y_i|\Omega)}{\sum_{j'=0}^{1} \omega_{j'|i} f_{j'}(y_i|\Omega)}$$

$$= \sum_{i=1}^{n} \sum_{j=0}^{1} \frac{\omega_{j|i} f_j(y_i|\Omega)}{\sum_{j'=0}^{1} \omega_{j'|i} f_{j'}(y_i|\Omega)} \frac{\partial}{\partial \Omega_\lambda} \log f_j(y_i|\Omega)$$

$$= \sum_{i=1}^{n} \sum_{j=0}^{1} \Pi_{j|i} \frac{\partial}{\partial \Omega_\lambda} \log f_j(y_i|\Omega)$$

where we define

$$\Pi_{j|i} = \frac{\omega_{j|i} f_j(y_i|\Omega)}{\sum_{j'=0}^{1} \omega_{j'|i} f_{j'}(y_i|\Omega)} \qquad (A1)$$

The MLEs of the parameters contained in (Ω_m, Ω_v) are obtained by solving

$$\frac{\partial}{\partial \Omega_\psi} \log L(\Omega|y) = 0 \qquad (A2)$$

However, the MLEs cannot be obtained directly because of the mixture distribution problem. The EM-Simplex algorithm was applied to estimate the parameters. For the M-C covariance model with AR(1) correlation structure, we have $\Sigma = \sigma^2 u^{1/2} \mathbf{R}(\rho) u^{1/2}$, where $u = \text{diag}\{\bar{u}^2(1), \ldots, \bar{u}^2(T)\}$, $\bar{u}(t)$ refers to the average mean for the two genotypes at time t, and $\mathbf{R}(\rho)$ is the AR(1) correlation matrix. This model has the following properties:

$$\log |\Sigma| = T \log (\sigma^2) + (T-1) \log (1-\rho^2) + \sum_{t=1}^{T} \log (u_j(t))$$

and

$$\frac{\partial}{\partial \rho} (y_i - u_j)\, \Sigma^{-1} (y_i - u_j)^T = \frac{1}{\sigma^2(\rho^2-1)^4} \left\{ -2 \sum_{t=1}^{T-1} \frac{[y_i(t) - u_j(t)][y_i(t+1) - u_j(t+1)]}{u_j(t)u_j(t+1)} + \frac{2[y_i(1) - u_j(1)]\rho}{u_j^2(1)} + \frac{2[y_i(T) - u_j(T)]\rho}{u_j^2(T)} \right\}$$

Therefore, by solving *Eq. A2*, we have

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \sum_{j=0}^{1} \Pi_{j|i} (y_i - \hat{u}_j)^T \hat{u}^{-1/2} \mathrm{R}(\hat{\rho}) \hat{u}^{-1/2} (y_i - \hat{u}_j)}{nT}$$

and

$$\hat{\rho} = \frac{n\sigma^2(T-1)\hat{\rho}^{*3} + \mathrm{C}(\hat{\rho}^{*2} + 1)}{n(T-1)\hat{\sigma}^2 - \mathrm{D}}$$

where

$$\Gamma(\hat{\rho}) = \begin{bmatrix} -\frac{1}{\hat{\rho}^2-1} & \frac{\hat{\rho}}{\hat{\rho}^2-1} & 0 & \cdots & 0 \\ \frac{\hat{\rho}}{\hat{\rho}^2-1} & -\frac{\hat{\rho}+1}{\hat{\rho}^2-1} & \frac{\hat{\rho}}{\hat{\rho}^2-1} & \cdots & 0 \\ 0 & \frac{\hat{\rho}}{\hat{\rho}^2-1} & -\frac{\hat{\rho}+1}{\hat{\rho}^2-1} & \frac{\hat{\rho}}{\hat{\rho}^2-1} & \vdots \\ \vdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \frac{\hat{\rho}}{\hat{\rho}^2-1} & -\frac{\hat{\rho}+1}{\hat{\rho}^2-1} & \frac{\hat{\rho}}{\hat{\rho}^2-1} \\ 0 & \cdots & \cdots & \frac{\hat{\rho}}{\hat{\rho}^2-1} & -\frac{1}{\hat{\rho}^2-1} \end{bmatrix}$$

$$\mathrm{C} = \sum_{i=1}^{n} \sum_{j=0}^{1} \Pi_{j|i} \sum_{t=1}^{T-1} \frac{[y_i(t) - \hat{u}_j(t)][y_i(t+1) - \hat{u}_j(t+1)]}{\bar{u}_j(t)\bar{u}_j(t+1)}$$

$$\mathrm{D} = \sum_{i=1}^{n} \sum_{j=0}^{1} \Pi_{j|i} \left\{ \frac{[y_i(1) - \hat{u}_j(1)]^2}{\bar{u}^2{}_j(1)} + \frac{[y_i(T) - \hat{u}_j(T)]^2}{\bar{u}^2{}_j(T)} \right\}$$

Estimation Procedures

1) E step: Given initial values for (Ω_u, Ω_v), calculate the posterior probability matrix $\Pi = \Pi_{j|i}$ in *Eq. A1*.

2) M step: With the mean parameters contained in Ω_u from the previous step, calculate the mean vector **u** and update the covariance parameters $\hat{\sigma}^2$ and $\hat{\rho}$. These updated covariance parameters are used in the simplex step to maximize the mean parameters contained in Ω_u.

3) The above procedures are iteratively repeated until a certain convergence criterion is met. The converging values are the MLEs of the parameters.

ACKNOWLEDGMENTS

We thank Dr. Jun Zhu for providing the rice data set that made it possible to write this manuscript.

GRANTS

This work was partially supported by National Science Foundation Grant 0540745 to R. Wu.

REFERENCES

1. **Akaike H.** A new look at the statistical model identification. *IEEE Trans Automatic Control* AC-19: 716–723, 1974.
2. **Altman NS.** Kernel smoothing of data with correlated errors. *J Am Stat Assoc* 90: 508–515, 1990.
3. **Atkinson AC.** Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *J R Stat Soc C* 52: 261–278, 2003.
4. **Balaban NQ, Merrin J, Chait R, Kowalik L, and Leibler S.** Bacterial persistence as a phenotypic switch. *Science* 305: 1622–1625, 2004.
5. **von Bertalanffy L.** Quantitative laws in metabolism and growth. *Q Rev Biol* 32: 217–231, 1957.
6. **Boularan J, Ferre L, and Vieu P.** Growth curves: a two-stage nonparametric approach. *J Stat Plan Infer* 38: 327–350, 1994.
7. **Carroll RJ and Ruppert D.** Power-transformations when fitting theoretical models to data. *J Am Stat Assoc* 79: 321–328, 1984.
8. **Cashio P, Lee TV, and Bergmann A.** Genetic control of programmed cell death in *Drosophila melanogaster*. *Semin Cell Dev Biol* 16: 225–235, 2005.
9. **Churchill GA and Doerge RW.** Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971, 1994.
10. **Dempster AP, Laird NM, and Rubin DB.** Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 39: 1–38, 1977.
11. **Diggle PJ and Verbyla AP.** Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* 54: 401–415, 1998.
12. **Diggle PJ, Heagerty P, Liang KY, and Zeger SL.** *Analysis of Longitudinal Data*. Oxford, UK: Oxford Univ. Press, 2002.
13. **Ellis HM and Horvitz HR.** Genetic control of programmed cell death in the nematode *C. elegans*. *Cell* 44: 817–829, 1986.
14. **Ferreira E, Nunez-Anton V, and Rondriguez-Poo J.** Kernel regression estimates of growth curves using nonstationary correlated errors. *Stat Prob Lett* 34: 413–423, 1997.
15. **Fogg GE and Thake B.** *Algal Cultures and Phytoplankton Ecology*. Madison, WI: Univ. of Wisconsin Press, 1987.
16. **Greenberg JT.** Programmed cell death: a way of life for plants. *Proc Natl Acad Sci USA* 93: 12094–12097, 1996.
17. **Hanahan D and Weinberg RA.** The hallmarks of cancer. *Cell* 100: 57–70, 2000.
18. **Horvitz HR.** Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. *Cancer Res* 59: 1701s-1706s, 1999.
19. **Horvitz HR.** Worms, life, and death (Nobel Lecture). *Chembiochem* 4: 697–711, 2003.
20. **Huang N, Parco A, Mew T, Magpantay G, McCough SR, Guiderdoni E, Xu J, Subudhi PK, Angeles ER, and Khush GS.** RFLP mapping of isozymes, RAPD and QTL for grain shape, brown planthopper resistance in a doubled haploid rice population. *Mol Breed* 3: 105–113, 1997.
21. **Jacobson MD, Wei M, and Raff MC.** Programmed cell death in animal development. *Cell* 88: 347–354, 1997.
22. **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.
23. **Lander ES and Botstein D.** Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199, 1989.
24. **Leonard T and Hsu J.** Bayesian inference for a covariance matrix. *Ann Stat* 20: 1669–1696, 1992.
25. **Linhart H and Zucchini W.** *Model Selection*. New York: Wiley, 1986.
26. **Lynch M and Walsh B.** *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer, 1998.
27. **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.
28. **Nelder JA and Mead R.** A simplex method for function minimization. *Computer J* 7: 308–313, 1965.
29. **Pletcher SD and Geyer CJ.** The genetic analysis of age-dependent traits: modeling the character process. *Genetics* 153: 825–835, 1999.
30. **Schwarz G.** Estimating the dimension of a model. *Ann Stat* 6: 461–464, 1978.
31. **Steller H.** Mechanisms and genes of cellular suicide. *Science* 267:1445–1449, 1995.

32. **Vaudry D, Falluel-Morel A, Leuillet S, Vaudry H, and Gonzalez BJ.** Regulators of cerebellar granule cell development act through specific signaling pathways. *Science* 300: 1532–1534, 2003.
33. **Wang N and Ruppert D.** Nonparametric estimation of the transformation in the transform-both-sides regression model. *J Am Stat Assoc* 90: 522–534, 1995.
34. **West GB, Brown JH, and Enquist BJ.** A general model for ontogenetic growth. *Nature* 413: 628–631, 2001.
35. **White K, Grether ME, Abrams JM, Young L, Farrell K, and Steller H.** Genetic control of programmed cell death in *Drosophila*. *Science* 264: 677–683, 1994.
36. **Wu RL and Lin M.** Functional mapping—How to map and study the genetic architecture of complex dynamic traits. *Nat Rev Genet* 7: 229–237, 2006.
37. **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.
38. **Wu RL, Ma CX, Zhao W, and Casella G.** Functional mapping of quantitative trait loci underlying growth rates: a parametric model. *Physiol Genomics* 14: 241–249, 2003.
39. **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.
40. **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.
41. **Wu RL, Wang ZH, Zhao W, and Cheverud JM.** A mechanistic model for genetic machinery of ontogenetic growth. *Genetics* 168: 2383–2394, 2004.
42. **Yan JQ, Zhu J, He CX, Benmoussa M, and Wu P.** Quantitative trait loci analysis for the developmental behavior of tiller number in rice. *Theor Appl Genet* 97: 267–274, 1998.
43. **Yuan J and Horvitz HR.** A first insight into the molecular mechanisms of apoptosis. *Cell* S116: S53–S56, 2004.
44. **Zeng ZB.** Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468, 1994.
45. **Zhao W, Wu RL, Ma CX, and Casella G.** A fast algorithm for functional mapping of complex traits. *Genetics* 167: 2133–2137, 2004.
46. **Zhao W, Chen YQ, Casella G, Cheverud JM, and Wu RL.** A non-stationary model for functional mapping of complex traits. *Bioinformatics* 21: 2469–2477, 2005.
47. **Zimmerman DL and Núñez-Antón V.** Structured antedependence models for longitudinal data. In: *Modeling Longitudinal and Spatially Correlated Data. Methods, Applications, and Future Directions*, edited by Gregoire TG, Brillinger DR, Diggle PJ, Russek-Cohen E, Warren WG, and Wolfinger R. New York: Springer, 1997, p.63–76.
48. **Zimmerman DL and Núñez-Antón V.** Parametric modeling of growth curve data: an overview. *Test* 10: 1–73, 2001.