Theor Appl Genet (2004) 108:699–705
DOI 10.1007/s00122-003-1484-9

ORIGINAL PAPER

C.-X. Ma · M. Lin · R. C. Littell · T. Yin · R. Wu

A likelihood approach for mapping growth trajectories using dominant markers in a phase-unknown full-sib family

Received: 16 May 2003 / Accepted: 9 September 2003 / Published online: 28 October 2003

Abstract Dominant markers have been commonly used in mapping quantitative trait loci (QTLs) in outcrossing species, in which not much prior genome information is available. But the dominant nature of these markers may lead to reduced QTL mapping precision and power. A new statistical method is proposed to incorporate growth laws into a QTL mapping framework, under which the use of the efficiency of dominant markers can be increased. This new method can be used to identify specific QTLs affecting differentiation in growth trajectories, and further estimate the timing of a QTL to turn on, or turn off, affecting growth during the entire ontogeny of a species. Using this method based on dominant markers we have successfully mapped a QTL for stem height growth trajectories to a linkage group in a forest tree. The implications of this method for the understanding of the genetic architecture of growth using dominant markers are discussed.

Communicated by F. Salamini

C.-X. Ma · M. Lin · R. C. Littell · R. Wu (✉)
Department of Statistics,
University of Florida,
533 McCarty Hall C, Gainesville, FL 32611, USA
e-mail: rwu@stat.ufl.edu
Tel.: +352-392-3806
Fax: +352-392-8555

T. Yin
The Key National Laboratory of Forest Genetics
and Gene Engineering,
Nanjing Forestry University,
Jiangsu 210037 Nanjing, P. R. China

R. Wu
Laboraty of Statistical Genetics,
Zhejiang Forestry College,
Lin'An, Zhejiang 311300, P. R. China

Introduction

The identification of quantitative trait loci (QTLs), responsible for quantitatively inherited phenotypes based on genetic linkage maps, is of typical importance to formulate an efficient breeding plan. Numerous marker systems have been developed to construct a genetic map and they can be broadly classified into two types, dominant and codominant. Codominant markers include restriction fragment length polymorphisms, microsatellites and single nucleotide polymorphisms, whose genotyping process critically relies upon the prior information of the genomes. These markers are mostly used for those well-studied organisms with heavy investments, like Arabidopsis, mice or humans. For many other organisms, dominant markers based on polymerase chain reactions, such as random amplified polymorphism DNAs (RAPDs) and amplified fragment length polymorphisms (AFLPs), may be the simplest and most economic choice because these markers require little prior genomic information. In fact, the abundance and accessibility of dominant markers in virtually any genome has made them widely applicable in current plant and animal breeding programs.

While dominant markers display great implications for genetic analysis, their dominant nature that the homozygote of the dominant allele and the heterozygote cannot be distinguished from each other, would cause a serious loss of the information for mapping (Ritter et al. 1990). Although for a backcross design or a doubled haploid design dominant markers are as informative as codominant markers, this is actually not the case for commonly used F_2 or full-sib families. Both theoretical modelling and empirical marker analysis indicate that linkage analysis using dominant markers has low mapping accuracy, power and resolution (Maliepaard et al. 1998; Wu et al. 2002a). Several strategies have been proposed to increase the use of the efficiency of dominant markers; for example, a hidden Markov model of Lander and Green (1987), a Monte Carlo EM algorithm of Jansen (1996) and the iteratively re-weighted least-square meth-

od of Xie and Xu (1999). The basic idea of these strategies is to make use of other more informative markers to recover missing information of less informative dominant markers. Its application is based on the assumption that both dominant and codominant markers are located on the same linkage group and will not be feasible for the situation, in which a cluster of dominant markers are located on the same particular genomic region (Young et al. 1998).

Recently, Ma et al. (2002) have devised a new statistical model for increasing QTL mapping efficiency by incorporating biological principles into a mapping framework. This model implemented with the EM algorithm, can provide the identification of a QTL responsible for a particular biological process. In the case study from a forest tree, this model displays more power to detect a QTL affecting the growth trait than traditional interval mapping or composite interval mapping. In this article, we extend this idea to map complex growth traits based on dominant markers that represent a less expensive, more common marker type for under-representative outcrossing species, such as forest trees.

Since outcrossing species are usually characterized by a high level of heterozygosity, it is difficult or impossible to generate inbred lines used for linkage mapping. For these species, full-sib families derived from heterozygous parents are utilized as basic mapping materials, in which parental linkage phases between different loci are unknown a priori. Ritter et al. (1990), Maliepaard et al. (1998) and Wu et al. (2002a) proposed various approaches for determining the linkage and parental linkage phases for any type of molecular markers. In many cases, especially when markers are not fully informative, parental linkage phases cannot be precisely determined, which thus affects the estimation of the linkage. Different from these treatments, here we devise a general model for estimating the probability of parental linkage phases, which allows for a simultaneous estimation of the linkage. We used an example from poplar trees to demonstrate the power of the extended model in QTL mapping using dominant markers in outcrossing species.

Theory and method

Dominant markers have reduced power due to missing information when they are used for interval mapping of QTLs in an F_2 or full-sib family. For an F_2, initiated with two inbred lines, alleles of two flanking markers and a putative QTL bracketed by the two markers have a known coupling phase; whereas, for a full-sib family derived from two heterozygous lines, the linkage phase of the markers and the QTL is not known a priori. The derivation of our mapping model starts with a simpler phase-known F_2 family, followed by a more complicated full-sib family.

Phase-known F_2 family

Suppose there is an F_2 population, initiated with two homozygous lines, in which there are three groups of genotypes at a codominant locus. Consider an age-specific trait, such as plant height or stem dry weight, which is measured at a finite number of time points during growth trajectories. Assume that this age-specific growth trait is affected by a segregating pleiotropic QTL, bracketed by two flanking markers *M* and *N*, with three possible genotypes *MM*, *Mm* and *mm* and *NN*, *Nn* and *nn*, respectively. The three genotypes at the QTL are expressed by *QQ*, *Qq* and *qq*.

With known linkage phases, we derive 27 conditional probabilities of the three QTL genotypes given the nine marker genotypes at the two flanking markers in the F_2 population. These probabilities are a function of the recombination fractions between the marker *M* and the QTL (r_1), between the QTL and the marker *N* (r_2) and between the two markers (r). But for dominant markers each with a dominant allele and recessive allele (denoted by *o*) that, therefore, form two distinguishable phenotypes, some of these conditional probabilities will be collapsed, leading to 4×3=12 probabilities. The conditional probability of a QTL genotype *j* (*j*=1 for *QQ*, 2 for *Qq* and 3 for *qq*), conditional upon the four phenotypes of the flanking markers (*M_N_*, *M_oo*, *ooN_* and *oooo*) for the *i*th F_2 progeny was expressed as

$$\pi_{ij} = \text{Prob}(x_i = j | M, N, r_1, r_2),$$

where _ denotes the unknown allele types at the markers, and x_i is the indicator variable characterizing the QTL genotype of progeny *i*. These conditional probabilities will be used to construct a mixture of the statistical model for QTL mapping.

Phase-unknown full-sib family

For any two heterogeneous parents, P_1 and Q_2, used to generate a full-sib family, we do not know the linkage phases between alleles at markers and QTLs. Wu et al. (2002a) proposed a Bayesian approach to characterize a most likely linkage phase for linkage analysis of the markers. After the linkage phase between the markers is determined, the next step is to detect a more likely marker-QTL phase. For two coupling flanking markers (*MN*/*oo*), two possible QTL-marker phases are expressed as Φ_1=*MQN*/*oqo* and Φ_2=*MqN*/*oQo*, where the slashes are used to separate two homologous chromosomes. The probabilities, with which these two linkage phases Φ_1 and Φ_2 occur, are denoted by p_1 and $1-p_1$ for parent P_1 and p_2 and $1-p_2$ for parent P_2. Two possible linkage phases for each parent generate four different phase combinations between the two parents, $\Phi_1^P \times \Phi_1^Q$, $\Phi_1^P \times \Phi_2^Q$, $\Phi_2^P \times \Phi_1^Q$, and $\Phi_2^P \times \Phi_2^Q$, with the corresponding probabilities denoted as $p_1\ p_2$, $p_1(1-p_2)$, $(1-p_1)p_2$ and $(1-p_1)(1-p_2)$. Under different parental phase combinations, conditional probabilities of a given QTL genotype, conditional upon a given two-marker genotype, will be different, although combinations $\Phi_1 \times \Phi_2$ and $\Phi_2 \times \Phi_1$ have the same probabilities (Table 1).

Considering all possible linkage phase combinations, the mixture of the conditional probability of a QTL genotype *j*, given a marker genotype for individual *i*, can be expressed as

$$\pi_{ij} = p_1 p_2 \pi_{ij}^{11} + (p_1 + p_2 - 2p_1 p_2)\pi_{ij}^{12|21} + (1-p_1)(1-p_2)\pi_{ij}^{22},$$

where the superscript of π denote different parental phase combinations. A similar expression can also be derived for the mixture of the conditional probability of two QTL-marker phases (*MQo*/*oqN* and *Mqo*/*oQN*) when two flanking markers are in the repulsion phase (*Mo*/*oN*). Table 1 lists the mixture of the conditional probabilities of the three QTL genotypes given by the genotypes of two coupling dominant markers under different parental phase combinations.

Statistical model

The phenotypes of a quantitative trait considered in this study are a group of measurements made at *m* hallmark time points. For many traits, such as plant height and

Table 1 Joint probabilities of QTL genotypes and marker genotypes at two coupling dominant markers, bracketing the QTL in a full-sib family derived from heterozygous parents with unknown linkage phases. The conditional probability of a QTL genotype given marker-genotypes can be derived according to the Bayes theorem

Marker		QTL genotypes		
Genotype	Frequency	*QQ*	*Qq*	*qq*
Phase combination $\hat{I}_{i_1}^{P}\times\hat{I}_{i_1}^{Q}$				
*M_N_*	1/4 $(3-2r+r^2)$	1/4 $(1-r_2^2)\ (1-r_1^2)$	1/2 $(r_1^2-r_1+1)\ (r_2^2-r_2+1)$	1/4 $r_1\ r_2\ (2-r_2)\ (2-r_1)$
M_oo	1/4 $(2-r)r$	1/4 $r_2^2\ (1-r_1^2)$	1/2 $(r_1^2-r_1+1)\ (1-r_2)r_2$	1/4 $r_1\ (1-r_2)^2\ (2-r_1)$
*00N_*	1/4 $(2-r)r$	1/4 $r_1^2\ (1-r_2^2)$	1/2 $(r_2^2-r_2+1)\ (1-r_1)r_1$	1/4 $r_2\ (1-r_1)^2\ (2-r_2)$
oooo	1/4 $(1-r)^2$	1/4 $r_1^2\ r_2^2$	1/2 $(1-r_1)\ r_1(1-r_2)r_2$	1/4 $(1-r_1)^2\ (1-r_2)^2$
Phase combination $\hat{I}_{i_1}^{P}\times\hat{I}_{i_2}^{Q}/\hat{I}_{i_2}^{P}\hat{I}_{i_1}^{Q}$				
*M_N_*	1/4 $(3-2r+r^2)$	1/4 $(1-r_1+r_1^2)\ (1-r_2+r_2^2)$	1/4 $[1+2\ r_1\ r_2\ (2-r_2+r_1\ r_2-r_1)-r_1^2-r_2^2]$	1/4 $(r_2^2+1-r_2)\ (r_1^2+1-r_1)$
M_oo	1/4 $(2-r)r$	1/4 $r_2\ (1-r_1+r_1^2)\ (1-r_2)$	1/4 $[r_2^2+2\ r_1-r_1^2+2\ r_1\ r_2\ (r_2+r_1-r_1\ r_2-2)]$	1/4 $r_2\ (r_1^2+1-r_1)\ (1-r_2)$
*00N_*	1/4 $(2-r)r$	1/4 $r_1\ (1-r_1)\ (1-r_2+r_2^2)$	1/4 $[r_1^2+2\ r_2-r_2^2+2\ r_1\ r_2\ (r_2+r_1-r_1\ r_2-2)]$	1/4 $r_1\ (r_2^2+1-r_2)\ (1-r_1)$
oooo	1/4 $(1-r)^2$	1/4 $r_1\ (1-r_1)\ r_2\ (1-r_2)$	1/4 $[1-2\ r_2+\ r_2^2-2\ r_1+r_1^2+2\ r_1\ r_2\ (2-r2-r_1+r_1\ r_2)]$	1/4 $r_1\ (1-r_1)\ r_2\ (1-r_2)$
Phase combination $\hat{I}_{i_2}^{P}\times\hat{I}_{i_2}^{Q}$				
*M_N_*	1/4 $(3-2r+r^2)$	1/4 $r_1\ \ r_2\ (2-r_2)\ (2-r_1)$	1/2 $(r_1^2-r_1+1)\ (r_2^2-r_2+1)$	1/4 $(1-r_2^2)\ (1-r_1^2)$
M_oo	1/4 $(2-r)r$	1/4 $r_1\ (1-r_2)^2\ (2-r_1)$	1/2 $(r_1^2-r_1+1)\ (1-r_2)\ r_2$	1/4 $r_2^2\ (1-r_1^2)$
*00N_*	1/4 $(2-r)r$	1/4 $r_2\ (1-r_1)^2\ (2-r_2)$	1/2 $(r_2^2-r_2+1)\ (1-r_1)\ r_1$	1/4 $r_1^2\ (1-r_2^2)$
oooo	1/4 $(1-r)^2$	1/4 $(1-r_1)^2\ (1-r_2)^2$	1/2 $(1-r_1)\ r_1\ (1-r_2)\ r_2$	1/4 $r_1^2\ r_2^2$

plant diameter, we can use a particular growth curve equation including a logistic or sigmoidal equation to model the phenotypes at different time points (Bertalanffy 1957; West et al. 2001). Because of genetic and environmental influences, different individuals may display different shapes of the growth curves. The aim of this study is the identification of the QTL responsible for the differences of the growth curve shape. If no such a QTL exists, three QTL genotypes will have an identical growth curve. Otherwise, there are three different growth curves each corresponding to a QTL genotype.

The trait phenotype of progeny i measured at time t due to the QTL located on an interval flanked by markers M and N can be expressed by a linear statistical model,

$$y_i(t) = \xi_{i1}g_1(t) + \xi_{i2}g_2(t) + \xi_{i3}g_3(t) + e_i(t), \quad (1)$$

where ξ_{ij} is an indicator variable describing the genotypes of the QTL for progeny i and defined to be 1 if a particular QTL genotype is indicated, and 0 otherwise; $g_j(t)$ is the genotypic value of the QTL for the trait at time t, which can be fit using the logistic curve expressed as,

$$g_j(t) = \frac{a_j}{1 + b_j e^{c_j t}} \quad (2)$$

with a being the asymptotic or limiting value of g when $t \to \infty$, $a/(1+b)$ is the initial value of g when $t=0$ and c is the relative rate of growth (Bertalanffy 1957), and $e_i(t)$ is the residual effect of progeny i, including the aggregate effect of polygenes and error effect, and distributed as $N[0,\sigma_e^2(t)]$. The m-point residual errors follow a multivariate normal distribution, $MN(0, \Sigma)$.

To increase the model's flexibility, Ma et al. (2002) structured the residual (co)variance matrix Σ using the AR(1) model (Davidian and Giltinan 1995), expressed as

$$\sum = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho^{m-1} \\ \rho & 1 & \cdots & \rho^{m-2} \\ \cdots & \cdots & \ddots & \cdots \\ \rho^{m-1} & \rho^{m-2} & \cdots & 1 \end{bmatrix} \quad (3)$$

in which we assume the variance stationarity, i.e., there is the same residual variance (σ^2) for growth at different ages, and covariance stationarity, and that the covariance of growth between different ages decreases proportionally (in ρ) with the increased time interval (Pletcher and Geyer 1999). There are two advantages when the structured matrix (3) is used. First, general expressions of the determinant and the inverse of Σ can be derived, which facilitates the parameter estimation., with these expressions, the growth model-based mapping approach can be applied for an arbitrary number of time points.

To ensure that the errors in model (1) are homoscedastic and normal, a transformation approach, as originally coined by Box and Cox (1964), can be used for the growth phenotype y_i . Preserving the favorable advantages of the Box-Cox transformation, Carroll and Ruppert (1984) proposed a so-called transform-both-side (TBS) approach, i.e. transforming at both sides of the model (3), aimed to obtain biologically meaningful estimates of growth parameters involved in model (1). Statistically, the TBS model is expressed as

$$\log[y_i(t)] = \xi_{i1}\log[g_1(t)] + \xi_{i2}\log[g_2(t)] + \xi_{i3}\log[g_3(t)] + \varepsilon_i(t),$$

or equivalently, defining $z_i(t)=\log[y_i(t)]$,

$$z_i(t) = \xi_{i1}h_1(t) + \xi_{i2}h_2(t) + \xi_{i3}h_1(t) + \varepsilon_i(t),$$

where

$$h_j(t) = \log\left[\frac{a_j}{1 + b_j e^{c_j t}}\right], \quad (4)$$

and $\varepsilon_i(t)$ is the residual error after the log-transformation, following a multivariate normal distribution $MN(0, \Sigma^2_\varepsilon)$.

According to a general QTL mapping theory, the likelihood of a mapping population of size n with the log-transformed phenotypes (z) measured at m different time-points can be formulated as

$$L(z|\Omega)=\prod_{i=1}^{n}\left[\sum_{j=1}^{3}\pi_{ij}f_j(z_i)\right], \quad (5)$$

where, as defined above, π_{ij} is the conditional probability of QTL genotype j given a marker genotype for progeny i, and $f_j(z_i)$ is a multivariate normal density of the phenotypes with m dimensions for a QTL genotype j,

$$f_j(z_i)=\frac{1}{(2\pi)^{m/2}|\sum_\varepsilon|^{1/2}}\exp\left[-\left(z_i-h_j\right)^T\sum_\varepsilon^{-1}\left(z_i-h_j\right)/2\right],$$

where h_j is the vector of the log-transformed (see Eq. 4) expected genotypic values of the trait for the QTL genotype j at t time-points, and Σ_ε is the residual variance-covariance matrix of the phenotypes among different time points.

When a logistic growth equation is implemented in the likelihood function (5), the unknown parameters being estimated are contained in the vector $\Omega=(a_j, b_j, c_j, r_1 \text{ or } r_2, \rho, \sigma_e^2)^T$. Parameters ($a$, b, c) determine the shape of a growth curve and are characterized by the QTL effect. For a phase-unknown full-sib family, the unknown vector includes two additional parameters, the phase probabilities p_1 and p_2. The maximum likelihood estimates (MLEs) of the unknown parameters for a pleiotropic QTL can be computed by implementing the EM algorithm (Dempster et al. 1977; Lander and Botstein 1989). Ma et al. (2002) derived a general framework for implementing the EM algorithm to solve the growth model-based likelihood function (5).

One of the important issues for the precise mapping of outcrossing populations is the characterization of the correct linkage phases. In this study, we have derived a closed-form solution of p_1 and p_2. With the estimates of p_1 and p_2, we can determine the probability of a parent to have a particular linkage phase between the markers and the QTL.

Hypothesis tests

We can make two kinds of hypothesis tests, global and local. The global test is to test the existence of a significant QTL responsible for differences in the growth curve, which can be formulated as

$$\begin{cases} H_0 : a_2 = a_1 = a_0, b_2 = b_1 = b_0, c_2 = c_1 = c_0 \\ H_1 : \text{not all equalities above hold.} \end{cases} \quad (6)$$

In H_0, the data are fit by a single logistic curve, whereas in H_1 three different logistic curves are used to fit the data. The log-likelihood ratio test statistic for these two hypotheses are calculated. The critical threshold can be determined on the basis of permutation tests, as advocated by Doerge and Churchill (1996).

The local test is to test the difference among the three QTL genotypic values at a particular time in the entire growth trajectories. The local test can identify the timing of the detected QTL to turn on or off to affect the growth trajectories (see Ma et al. 2002 for a detailed discussion).

Results

An example of a forest tree is used to demonstrate the power of our statistical model for mapping QTLs affecting growth trajectories using dominant markers. The study material, as described in Yin et al. (2002), was derived from the interspecific hybridization of *Populus* (poplar), *Populus deltoides* and *Populus euramericana*. A total of 450 hybrids were planted at a spacing of 4×5 m in the complete randomized design in a field trial near Xuchou City, Jiangsu Province, China. The total stem height and diameters measured at the end of each of the first 11 growing seasons are used for our QTL analysis.

In constructing genetic linkage maps for the interspecific hybrid between *P. deltoides* and *P. euramericana* using 90 genotypes randomly selected from the 450 hybrids (Yin et al. 2002), we found a number of RAPD and AFLP dominant markers that segregate in a 3:1 ratio. These dominant markers were found to cluster on particular linkage groups, as also seen in other organisms (Young et al. 1998). We choose one of the linkage groups comprising all of the dominant markers to map the QTL affecting growth trajectories of stem diameters in the mapping population.

As tested in Fig. 1, the height growth of a subset of the 90 mapped genotypes follows a different S-shaped logistic growth curve. This implies that it is appropriate to build our model upon this nearly universal growth law (West et al. 2001). Because our mapping population was derived from two outcrossing parents, we should characterize the most likely linkage phase among the dominant markers and QTLs, to estimate the QTL effects and positions. We thus used a full-sib family model to estimate the growth parameters of different QTL genotypes and the phase probabilities for each parent.

Using the growth law-incorporated method without considering the TBS model, we successfully detected a QTL affecting the height growth curves located on a linkage group comprising six dominant markers (Fig. 2). The critical value for claiming the existence of this QTL is determined on the basis of the permutation tests proposed by Doerge and Churchill (1996). We used 1,000 permutation tests to obtain the empirical estimate of the chromosome-wide critical value as 116, at the significance level α=0.05. The profile of the log-likelihood ratios (LR) of the full vs. reduced model (Eq. 6) across the length of the linkage group, has a clear peak (119) on a small interval (3.6 cM long) between two coupling dominant markers AACGT470 and L4-320. This suggests that we have adequate evidence for the existence of this

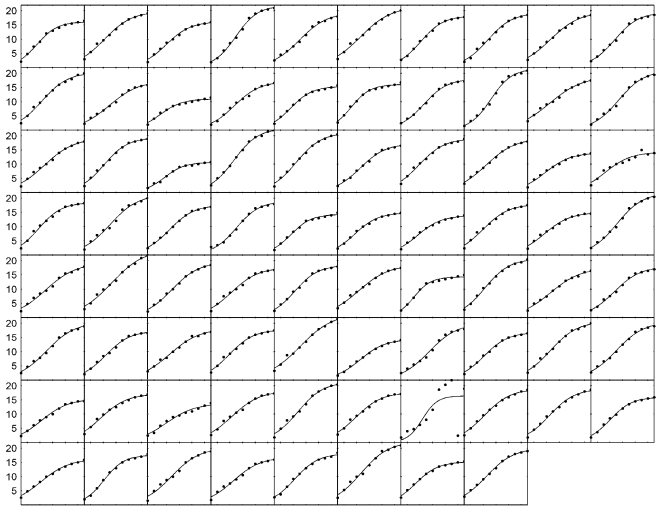

Fig. 1 Plots of stem-height growth against ages for a subset of the mapping population used to construct linkage maps in poplar hybrids (Yin et al. 2002). The growth of these genotypes can well fit a particular logistic curve. The x- and y-axes of the plots denote age (in year) and stem height (in m)

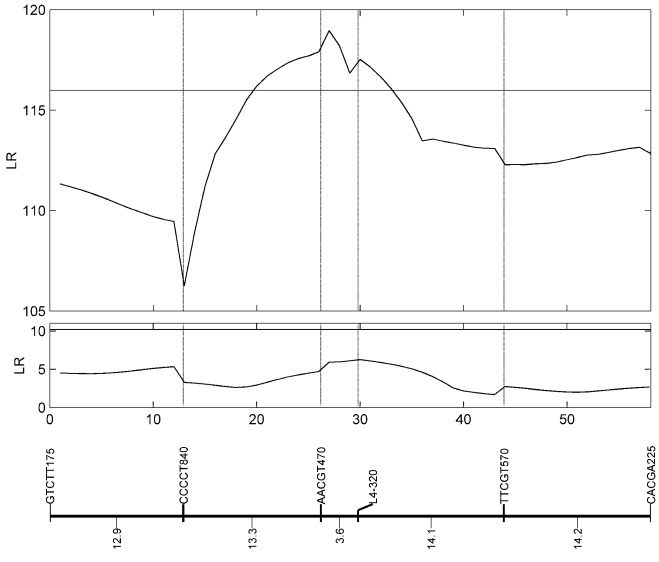

Fig. 2 The profile of the log-likelihood ratios (LR) between the full and reduced (no QTL) model for height-growth trajectories, across a linkage group comprising of dominant RAPD and AFLP markers in the *P. deltoides* × *P. euramericana* map. Upper: the genomic position corresponding to the peak of the profile is the MLE of the QTL position from our model. The threshold value for our method is given as the *horizonal line*. Lower: the profile of the LR values and the empirical threshold (*horizonal line*) are obtained for the most differentiated growth at age 11 years from traditional interval mapping. The *vertical dot lines* indicate the positions of markers on the linkage group, whose names are given beneath

QTL on this interval. The dominant alleles of the two flanking markers were found to be in a repulsion phase with the favorable allele of the QTL in parent P_1 ($\hat{p}_1 = 0$) and to be most likely in a coupling phase with the favorable allele of the QTL in parent P_2 ($\hat{p}_2 = 0.75$).

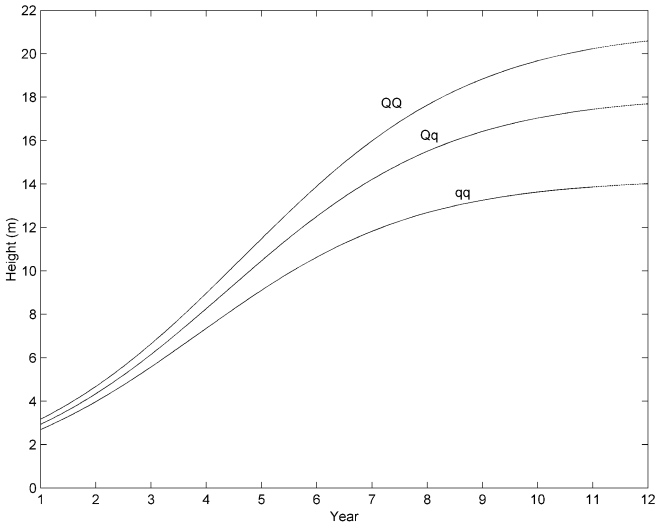

Fig. 3 Three height-growth curves each corresponding to a QTL genotype for a QTL detected on a linkage group, comprising dominant RAPD and AFLP markers in the *P. deltoides* × *P. euramericana* map. Each curve is described by group parameters, which are a_2=21.22, b_2=9.15 and c_2=0.4759 for genotype *QQ*, a_1=18.12, b_1=8.43 and c_1=0.4889 for genotype *Qq* and a_0=14.24, b_0=7.10 and c_0=0.5068 for genotype *qq*. The differentiation pattern of growth curves beyond the maximum observed-age (11 years), affected by the QTL, is represented by extended *broken curves*

To compare the power of our method with previous methods, the same material was analyzed by the more traditional interval mapping approach (Lander and Botstein 1989), based on the most differentiated phenotypes measured at the age of 11 years. No QTL could be detected to affect growth traits from this approach (Fig. 2). The peak of the LR profile from interval mapping is 5.0, largely below the threshold 10 obtained from 1,000 permutation tests, at the significance level α=0.05.

Our method can detect the dynamic change of QTL expression over time because the parameters determining the curve shape of a different QTL genotype are estimated. The growth curves of height are drawn using the estimates of logistic parameters for the three genotypes at the QTL detected on the dominant marker linkage group (Fig. 3). The three curves tend to be parallel, with genotype *QQ* on the top, genotype *qq* on the bottom and genotype *Qq* intermediate. This pattern suggests that this QTL mostly displays an additive effect on the entire period of height growth. Based on the local test, this QTL is found to be inactive until trees grew to about 6 years in the field (Fig. 3). And its effect on height growth increased with ages. At 11 years old, genotype *QQ* is about 3 m taller than genotype *Qq*, whereas genotype *Qq* is about 4 m taller than genotype *qq*. This difference appears to increase after the age of 11 years, as predicted from the logistic curves estimated.

We also perform QTL analysis for the same linkage group using the TBS-based model (Carroll and Ruppert 1984). The same QTL detected from the untransformed

data, as shown in Fig. 2, can also be detected, but the precision of parameter estimation is much higher when the data are untransformed (Result not shown). The maximum value of the log-likelihood ratio test statistics over the linkage group from the transformed data is about 30% larger than the critical threshold at the significance level α=0.05, whereas the counterpart value is 2.6% [(119–116)/116] for the untransformed data. This pronounced difference suggests that the TBS-based mapping model better fits the AR(1) model (Eq. 3) and can increase the power to detect a QTL affecting growth trajectories using dominant markers, by making the residual variance more constant over time.

Discussion

Current genome technologies have been available to dissect quantitative traits into individual locus components (QTLs), through which the genetic basis of quantitative traits can be better unravelled (reviewed by Mackay 2001). Because PCR-based markers are rich and do not rely on prior genome information, QTL mapping using dominant markers has now become a common practice in outcrossing species. Due to their partial information content, however, the use of these markers leads to reduced mapping precision and power (Ritter et al. 1990; Maliepaard et al. 1998; Wu et al. 2002a).

In this article, we develop a new statistical method for mapping a QTL affecting a continuously inherited trait. This method has two major advantages. First, it can increase the use efficiency of dominant markers by incorporating growth laws into a conventional mapping framework established by Lander and Botstein (1989). Previous methods were developed to recover the missing information of dominant markers from other more informative codominant markers on the same linkage group (Lander and Green 1987; Jansen 1996; Gessler and Xu 1999; Xie and Xu 1999). However, for a linkage group clustered with dominant markers (e.g. Young et al. 1998), these methods are limited due to the inadequacy of codominant markers. Our simultaneous mapping for repeated measurements based on a logistic curve, similar to multi-trait QTL analysis (Korol et al. 2001), can extract maximum information about QTL effects and positions contained in a data set and, thus, display increased power for QTL detection. In the poplar example used here, we detected a QTL for growth using our method, but could not do so using interval mapping. Our method proposed in this article, therefore, presents an excellent alternative to increasing the applicability of less expensive dominant markers in practice. It should be noted that the logistic growth model (Eq. 2) used in this study is only one of the growth laws describing the relationships between size and age. For a particular data set, tests for different growth equations should be made, from which an optimal equation is chosen.

Second, our method can provide more informative results regarding the genetic basis of complex traits such as growth (see Atchley 1984; Atchley and Zhu 1997; Pletcher and Geyer 1999). Results from a single time point-analysis can only detect a QTL affecting the phenotype at a particular time (Cheverud et al. 1996; Vaughn et al. 1999). Yet, our method can detect a QTL governing the entire process of growth, whose use is not limited by an increasing number of time points measured. Further, we can estimate the timing of a QTL to start or cease its effect on growth. As a by-product of this study, we obtain a closed-form solution for estimating the probabilities of parental QTL-marker linkage phases for outcrossing species. This will make our growth model-based approach broadly useful in practical outcrossing mapping programs.

As in Ma et al. (2002), we use the AR(1) model (Eq. 3) to fit the variance-covariance matrix. But this model may not always be true. We proposed, using a TBS-based model, to ascertain that the residual variance is constant over age and, therefore, preserve the favorable feature of the AR(1) to be easily manipulated. If the variance-stationarity assumption cannot be justified even after data transformation, we can use Nunez-Anton's (1997) and Nunez-Anton and Zimmerman's (2000) structured antedependent models to treat nonstationary variance-covariance structures. Or, a functional approach that describes the dynamic changes of genetic variance and covariance with ages can be used; for example, Kirkpatrick and coworkers used the Legendre polynomial (Kirkpatrick and Heckman 1989; Kirkpatrick et al. 1990, 1994) and Pletcher and Geyer's (1999) the character process model. Lund et al. (2002) proposed a random regression model to incorporate the Legendre polynomials into QTL mapping for repeated measurements. When a functional approach is implemented, the favorable properties of structuring Σ using the AR(1) model will be lost. Also, in this case, computationally more extensive numerical approaches are needed to provide the estimates of the time-dependent variances and covariances (Verbeke and Molenberghs 2000).

In addition, our analysis is based on a one-QTL model in a phase-unknown full-sib family. A general model should be developed to consider a multiple epistatically interacting QTL and multiple families derived from related or unrelated parents. If different families are chosen from a population, our functional mapping framework should be constructed on linkage-disequilibrium mapping that makes use of historical recombinants in a population (Lou et al. 2003). A joint linkage and linkage-disequilibrium analysis (Wu et al. 2002b) can also be framed to increase the efficacy of functional mapping by utilizing the recombinants that are created in both genotyped families, and at a historical time.

Acknowledgements We thank two anonymous referees for their constructive comments on this manuscript. This work is supported by an Outstanding Young Investigator Award (30128017) of the National Natural Science Foundation of China, a University of Florida Research Opportunity Fund (02050259), and a University of South Florida Biodefense grant (7222061-12) to R. W. The

publication of this manuscript was approved as Journal Series No. R-09204 by the Florida Agricultural Experiment Station.

References

Atchley WR (1984) Ontogeny, timing of development, and genetic variance-covariance structure. Am Nat 123:519–540

Atchley WR, Zhu J (1997) Developmental quantitative genetics, conditional epigenetic variability and growth in mice. Genetics 147:765–776

Bertalanffy von L (1957) Quantitative laws for metabolism and growth. Quart Rev Biol 32:217–231

Box GEP, Cox DR (1964) An analysis of transformations. J Roy Stat Soc, pp 211-252

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

Cheverud JM, Routman EJ, Duarte FAM, van Swinderen B, Cothran K, Perel C (1996) Quantitative trait loci for murine growth. Genetics 142:1305–1319

Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. Chapman and Hall, London

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via EM algorithm. J Roy Stat Sco Ser B 39:1–38

Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics 142:285–294

Gessler DDG, Xu S (1999) Multipoint genetic mapping of quantitative trait loci with dominant markers in outbred populations. Genetica 105:281–291

Jansen RC (1996) A general Monte Carlo method for mapping multiple quantitative trait loci. Genetics 142:305–311

Kirkpatrick M, Heckman N (1989) A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. J Math Biol 27:429–450

Kirkpatrick M, Lofsvold D, Bulmer M (1990) Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979–993

Kirkpatrick M, Hill WG, Thompson R (1994) Estimating the covariance structure of traits during growth and aging, illustrated with lactation in dairy cattle. Genet Res 64:57–69

Korol AB, Ronin YI, Itskovich AM, Peng J, Nevo E (2001) Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. Genetics 157:1789–1803

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in human. Proc Natl Acad Sci USA 84:2363–2367

Lou X-Y, Casella G, Littell RC, Yang MCK, Johnson JA, Wu RL (2003) A haplotype-based algorithm for multilocus linkage-disequilibrium mapping of quantitative trait loci with epistasis. Genetics 163:1533–1548

Lund MS, Sorensen P, Madsen P (2002) Linkage analysis in longitudinal data using random regression. 7th World Congress on Genetics Applied to Livestock Production, August 19–23 2002, Montpellier, France

Ma CX, Casella G, Wu RL (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. Genetics 161:1751–1762

Mackay TFC (2001) Quantitative trait loci in *Drosophila*. Nature Rev Genet 2:11–20

Maliepaard C, Alston FH, van Arkel G, et al: (1998) Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73

Nunez-Anton V (1997) Longitudinal data analysis: non-stationary error structures and antedependent models. Appl Stoch Models Data Anal 13:279–287

Nunez-Anton V, Zimmerman DL (2000) Modeling nonstationary longitudinal data. Biometrics 56:699–705

Pletcher SD, Geyer CJ (1999) The genetic analysis of age-dependent traits: modeling the character process. Genetics 153:825–835

Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125:645–654

Vaughn TT, Pletscher LS, Peripato A, King-Ellison K, Adams E, Erikson C, Cheverud JM (1999) Mapping quantitative trait loci for murine growth: a closer look at genetic architecture. Genet Res 74:313–322

Verbeke G, Molenberghs G (2000) Linear mixed models for longitudinal data. Springer, Berlin Heidelberg New York

West GB, Brown JH, Enquist BJ (2001) A general model for ontogenetic growth. Nature 413:628–631

Wu RL, Ma C-X, Painter I, Zeng Z-B (2002a) Simultaneous maximum-likelihood estimation of linkages and linkage phases over a heterogeneous genome. Theor Pop Biol 61:349–363

Wu RL, Ma C-X, Casella G (2002b) Joint linkage and linkage-disequilibrium mapping of quantitative trait loci in natural populations. Genetics 160:779–792

Xie C, Xu S (1999) Mapping quantitative trait loci with dominant markers in four-way crosses. Theor Appl Genet 98:1014–1021

Yin TM, Zhang XY, Huang MR, Wang MX, Zhuge Q, Tu SM, Zhu L-H, Wu RL (2002) The molecular linkage maps of the Populus genome. Genome 45:541–555

Young WP, Wheeler PA, Coryell VH, Keim P, Thorgaard GH (1998) A detailed linkage map of rainbow trout produced using doubled haploids. Genetics 148:839–850