Theor Appl Genet (2002) 105:1043–1049
DOI 10.1007/s00122-002-1052-8

W. Wu · Y. Zhou · W. Li · D. Mao · Q. Chen

Mapping of quantitative trait loci based on growth models

Received: 26 June 2001 / Accepted: 4 January 2002 / Published online: 14 September 2002
© Springer-Verlag 2002

Abstract An approach called growth model-based mapping (GMM) of quantitative trait loci (QTLs) is proposed in this paper. The principle of the approach is to fit the growth curve of each individual or line with a theoretical or empirical growth model at first and then map QTLs based on the estimated growth parameters with the method of multiple-trait composite interval mapping. In comparison with previously proposed approaches of QTL mapping based on growth data, GMM has several advantages: (1) it can greatly reduce the amount of phenotypic data for QTL analysis and thus alleviate the burden of computation, particularly when permutation tests or simulation are performed to estimate significance thresholds; (2) it can efficiently analyze unbalanced phenotype data because both balanced and unbalanced data can be used for fitting growth models; and (3) it may potentially help us to better understand the genetic basis of quantitative trait development because the parameters in a theoretical growth model may often have clear biological meanings. A practical example of rice leaf-age development is presented to demonstrate the utility of GMM.

Keywords Quantitative trait loci · Growth model-based mapping · Development · Leaf age · Rice

Introduction

Molecular markers enable us to map quantitative trait loci and, at the same time, estimate their individual effects. In general, QTL mapping is carried out based on the trait performance at a certain time or stage (often the end) of ontogeny. This strategy of QTL mapping can be called time-fixed mapping (TFM; Wu et al. 1999). An alternative strategy is to map QTLs based on the trait performance at a series of times or stages of ontogeny. This strategy can be called time-related mapping (TRM; Wu et al. 1999). Two approaches have been proposed for analyzing the time-related phenotype data for QTL mapping. An intuitive approach used in most TRM studies is to map QTLs by analyzing the trait performance at each observation time or time interval separately (Bradshaw and Settler 1995; Sondur et al. 1995; Cheverud et al. 1996; Plomion et al. 1996; Price and Tomos 1997; Verhaegen et al. 1997; Yan et al. 1998a, b). By comparing QTLs detected at different times or time intervals, the expression dynamics of individual QTLs could be inferred. This kind of approach can be called separate time-related mapping (STRM). Considering that a correlation may exist between phenotypes observed at adjacent times, Yan et al. (1998a, b) developed a so-called conditional mapping (CM) method for STRM analysis. The authors suggest that the CM method could provide an estimate of the net effect of a QTL expressed at each time interval.

Alternatively, Wu et al. (1999) proposed a joint time-related mapping (JTRM) approach, which treats the developmental process of a quantitative trait as a whole for QTL mapping by jointly analyzing all phenotype data measured at different times or time intervals with the method of multiple-trait composite interval mapping (MCIM; Jiang and Zeng 1995). With a practical example of tiller number development in rice the authors demonstrated that JTRM and STRM could detect the same set of QTLs, but JTRM has the merits of getting a comprehensive estimate of each QTL's position and plotting a complete curve of the expression dynamics for each QTL. In addition, JTRM may also potentially increase the statistical power of QTL mapping because it makes use of the information on correlations between different times or time intervals in the analysis.

Although the current TRM approaches (STRM and JTRM) can reveal the differential activities of QTLs during ontogeny, it does not take the pattern of trait-value growth into account. The pattern of trait-value growth

Communicated by Q. Zhang

W. Wu (✉) · Y. Zhou · W. Li · D. Mao · Q. Chen
College of Crop Science,
Fujian Agriculture and Forestry University,
Fuzhou, Fujian 350002, People's Republic of China
e-mail: zjweiren@pub5.fz.fj.cn
Tel.: +86-591-3789176, Fax: +86-591-3789176

has long been an interesting subject in physiological studies because it may reflect the biological mechanisms of quantitative trait development. The pattern of trait-value growth can usually be quantitatively described with a theoretical or an empirical growth model (Grizzle and Allen 1969; Koops 1986; France et al. 1996) such as the Gompertz model (Winsor 1932), the Richards model (Richards 1959) and the Weibull model (Yang et al. 1978). The usefulness of a growth model is that it allows precise description of a complicated growth process with only a few parameters, which, in many cases, may have definite biological meanings, depending on the trait and the organism concerned. Practical examples can be seen from studies on human-height growth (Preece and Baines 1978; Jolicoeur et al. 1988; Kanefuji and Shohoji 1990). Therefore, growth models are helpful for understanding the biological mechanisms conferring quantitative trait development. We thus have reason to expect that combining growth models with QTL mapping would enable a better understanding of the developmental genetic basis of quantitative traits.

In this paper, we propose an approach of QTL mapping based on growth models, called growth model-based mapping (GMM), and compare it with JTRM via a practical example.

The principle of growth model-based QTL mapping

At the phenotypic level, the development of a quantitative trait appears as a process of trait value growth, which can be plotted as a curve in a two-dimensional space defined by time as one axis and trait value as the other. In general, every trait has a characteristic pattern (or shape) of the growth curve, but the details of the growth curve may vary greatly among genotypes (see Fig. 2a). In other words, the growth curve of a quantitative trait is genetically controlled. Hence, by associating growth curves (phenotypes) of a trait with molecular marker-types (genotypes), it is possible to detect QTLs controlling the development of the trait.

A straightforward and intuitive approach for mapping QTLs underlying the development of a quantitative trait, by testing the association between growth curves and marker-types, is to integrate marker-type data and all phenotype data measured at sequential times or time intervals into a multivariate linear regression analysis with the method of MCIM (taking the DH population as an example), i.e.,

$$y_{jk}=b_{0k}+b^*_k x^*_j+\sum_{l}^{m} b_{lk}x_{jl}+\varepsilon_{jk} \quad (j=1,\ldots,n;\ k=1,\ldots,t), \tag{1}$$

where y_{jk} is the phenotypic value of DH line j for trait (or time or time-interval) k; b_{0k} is the mean of the model for trait k; b^*_k is the additive effect of the putative QTL on trait k; x^*_j is an indicator variable of the putative QTL's genotypes, taking values of 1 for genotype Q_1Q_1 and –1 for Q_2Q_2, with probabilities depending on the genotypes of flanking markers; b_{lk} is the partial regression coefficient of y_{jk} on marker l; x_{jl} is an indicator variable of genotypes of marker (cofactor) l, taking values of 1 for genotype M_1M_1 and –1 for M_2M_2; ε_{jk} is the random error for trait k in DH line j; and n, m and t are the numbers of DH lines, markers selected as cofactors and traits (or times or time intervals) to be analyzed, respectively. This approach is actually the principle of JTRM (Wu et al. 1999). However, the approach does not take the mathematical pattern of the growth curve into account.

In principle, the growth curve of a trait can be described by a theoretical or an empirical growth model, i.e.,

$$y = f(t;\theta)+\varepsilon, \tag{2}$$

where t is time; y is the trait value at time t; $\theta = (\theta_1, \theta_2, \ldots, \theta_m)'$, a vector of parameters, which defines a point in an m-dimensional space (see Fig. 2b); and ε is the random error following a normal distribution with mean = 0. Consider an ideal situation, where the growth curve is perfectly fitted by the growth model and there is no random error ($\varepsilon \equiv 0$). In this case, the growth curve is completely determined by the growth-parameter vector θ. Namely, there is a definite corresponding relationship between the growth curve and the growth-parameter vector. In the terms of analytical mathematics (Lang 1978), function (2) is a mapping from the set of growth-parameter vectors (Θ) to the set of growth curves (C), i.e., f: $\Theta \rightarrow$ C. This means that the variation observed in C (or among growth curves; Fig. 2a) should be a reflection of the variation in Θ (or among growth-parameter vectors; Fig. 2b), and so the association between growth curves and marker-types should be a reflection of the association between growth-parameter vectors and marker-types. Thus, we find an alternative approach for mapping QTLs underlying the development of a quantitative trait based on testing the association between growth curves and marker-types; namely, to integrate marker-type data and estimated growth parameters (derived phenotype data) into a multivariate linear regression analysis with the method of MCIM. This is the method of GMM. A model similar to (1) can be used. The only difference is that the symbol y_{jk} in the model of GMM stands for the kth growth parameter (instead of the trait value measured at the kth observation time or time-interval) of the jth DH line.

An example

To verify the usefulness of GMM, we compared it with JTRM using data of the leaf age growth in rice. The leaf age (or the age indicated by the leaves on the main culm) of a plant at a given time is defined by Zhou et al. (2001) as:

$$\text{Leaf age} = \text{Number of developed leaves} + \frac{\text{Length of the developing leaf}}{\text{Final length of the developing leaf}}.$$

The experiment was carried out with a DH population consisting of 111 lines derived from a cross between an *indica* rice variety Gui-630 and a *japonica* rice variety Taiwanjing. The leaf age of each plant was investigated

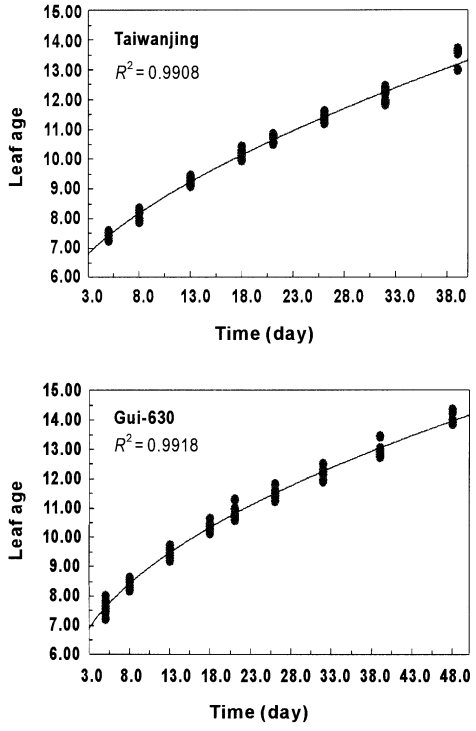

Fig. 1 Growth curves of the leaf age of parental lines Taiwanjing and Gui-630 fitted by GM-I. R^2: coefficient of determination

every 5–7 days beginning on the 30th day after sowing untill the end of leaf growth. For details of the experiment, see Zhou et al. (2001).

A simple empirical growth model was found to fit the data of the leaf-age growth of each DH line well within the time span of observation (Fig. 1; all the coefficients of determination obtained by the model were greater than 95%, with an average of 98.33 ± 1.15%): i.e.,

$$y = c + at^b + \varepsilon, \tag{3}$$

where y is leaf age, t is time (the origin of t was set at the 25th day after sowing), a, b and c are parameters, and ε is the error. Both the fitted growth curves and their corresponding estimated growth-parameter vectors showed great variation (Fig. 2), implying that the leaf-age growth is genetically controlled in rice. In addition, all the three estimated growth parameters showed an approximately symmetrical frequency distribution (Fig. 3). This indicates that the growth-parameter vector θ = (a, b, c) had the statistical features of multiple quantitative traits suitable for MCIM analysis.

In order to examine whether using different growth models could affect the results of GMM, we identified another empirical growth model, which can also fit the leaf-age growth data very well within the time span of observation (all the coefficients of determination obtained by the model were greater than 95%, with an average of 98.35 ± 1.15%) but is a little more complicated than (3): i.e.,

$$y = \frac{a + bt}{1 + ct + dt^2} + \varepsilon, \tag{4}$$

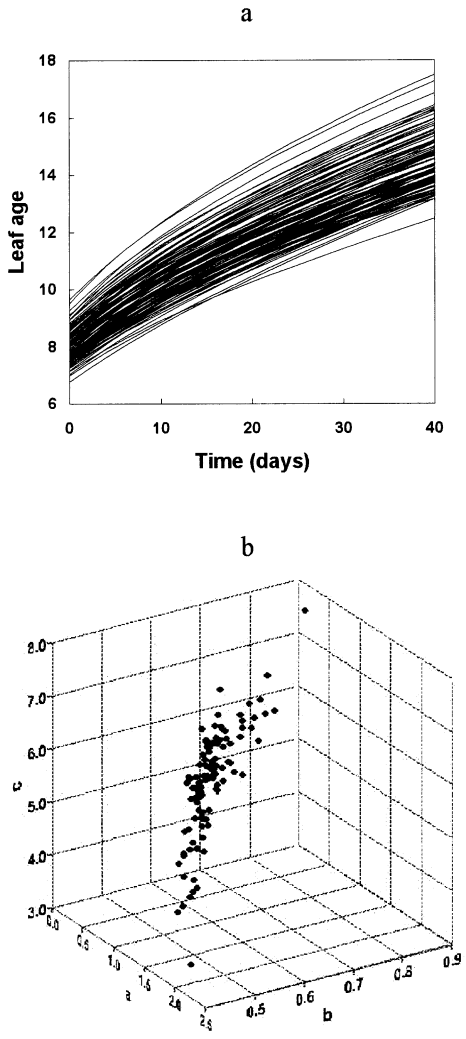

Fig. 2 Growth curves of the leaf age of 111 DH lines fitted by GM-I (**a**) and their corresponding points in the growth-parameter space (**b**)

where a, b, c and d are parameters; and all other symbols have the same meanings as those in (3). Similar to (3), all the estimated growth parameters in (4) also showed an approximately symmetrical frequency distribution (Fig. 3). Therefore, the growth-parameter vector θ = (a, b, c, d) also possessed the statistical features of multiple quantitative traits suitable for MCIM analysis.

For the convenience of address, we shall call (3) and (4) growth model I (GM-I) and growth model II (GM-II) and, correspondingly, abbreviate the QTL mapping analyses based on the two growth models as GMM-I and GMM-II, respectively.

Both GMM (including GMM-I and GMM-II) and JTRM were performed using the method of MCIM based on the least-squares estimation (Wu et al. 1999). As our main purpose was to check whether growth curves (measurable trait values) and growth parameters (derived trait values) supply equivalent genetic information for QTL mapping, we did not use the practically observed, but expected, data of leaf ages at the 30th, 34th, 38th, …,

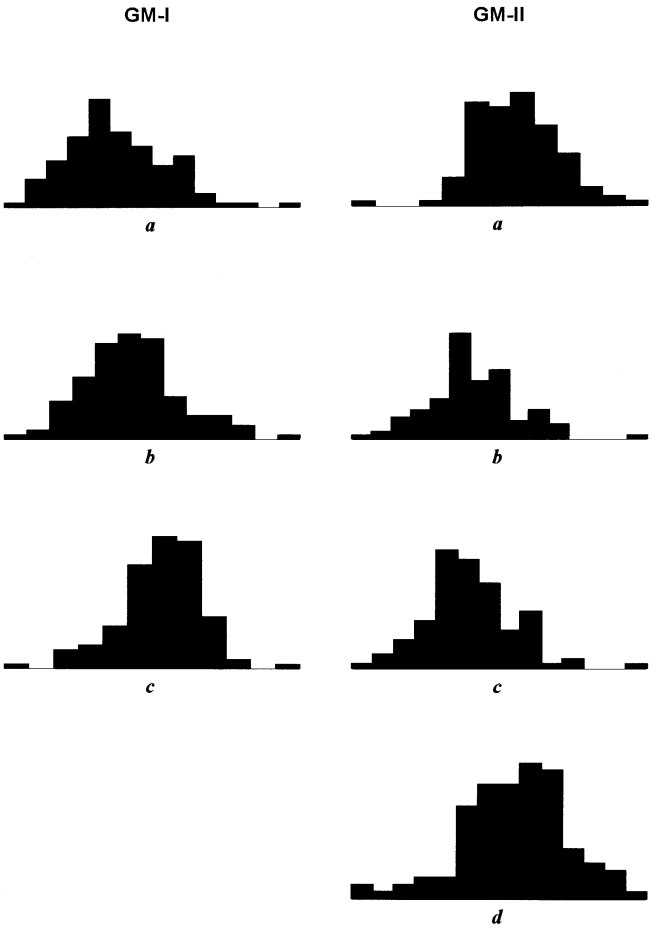

Fig. 3 Frequency distributions of growth parameters of GM-I and GM-II

70th days after sowing, estimated by the fitted GM-I for the JTRM (see also Discussion for another reason). Besides, considering that using different cofactors in the regression model of MCIM may significantly affect the results of QTL mapping, we did not screen for cofactors but used a full model containing all markers (Zeng 1994). However, because the total number of markers exceeds the sample size, we performed MCIM on each chromosome separately. A 20 cM-wide window was set on each side of the marker interval being tested throughout the MCIM analysis.

Permutation tests (Churchill and Doerge 1994) with 5,000 replicates were conducted to estimate LOD significance thresholds. Since the MCIM analyses were performed on each chromosome separately, the permutation tests were also conducted on each chromosome separately. In order to have a genome-wise significance level of 0.05, a nominal significance level of $1 - (1 - 0.05)^{1/12} = 0.00427$ for each chromosome was used (note: rice genome n = 12).

Under such a significance level, a total of four QTLs were detected on chromosomes 2, 8, 10 and 12 (Fig. 4), designated as *qLA2*, *qLA8*, *qLA10* and *qLA12*. Among the three methods, GMM-I detected all four QTLs, while GMM-II and JTRM only detected three (*qLA2*, *qLA10* and *qLA12*) and two (*qLA10* and *qLA12*) of them, respectively.

Following Wu et al. (1999), after QTL mapping the curves of accumulated expression, the expression rate of the additive effect [$a(t)$ and $a'(t)$] and the proportion of phenotypic variation explained (R^2) by each QTL could be estimated using a multiple linear regression model containing all the detected QTLs (Fig. 5). Similar to the case of MCIM, we again did not use the practically observed but the expected phenotype data of leaf age and, in addition, the leaf age growth rate estimated by the fitted GM-I for the multiple regression analyses.

It is evident from Fig. 5 that the expression dynamics of the QTLs were quite different. The expression rate of the additive effect of *qLA10* dropped dramatically within the time span, while that of *qLA12* almost remained constant. For *qLA2* and *qLA8*, the expression rates decreased at the beginning until reaching zero and then increased again, but the acting direction reversed. On average, over the time span, *qLA10* and *qLA12* had the largest effects and contributions to the explained phenotypic variation, followed by *qLA2*, while *qLA8* exhibited a very small effect and contribution. This is consistent with their LOD scores (Fig. 4).

Comparing GMM-I and GMM-II, we see that they generated quite similar LOD profiles and detected the same set of QTLs except for *qLA8*. As we have seen that *qLA8* showed a very small effect and its LOD score was just slightly over the significance threshold in GMM-I, it could probably be a false positive. Hence, neglecting *qLA8*, we see that GMM-I and GMM-II are basically equivalent. In other words, using different growth models may not greatly affect the result of GMM, as long as the growth models can fit the growth data equally well.

Comparing GMM and JTRM, we see that their LOD profiles were also very similar in shape, particularly in the regions around the detected QTLs, and they also detected the same set of QTLs except for *qLA2* (ignoring *qLA8*). This indicates that GMM and JTRM are approximately equivalent, but GMM seems to be a little more powerful than JTRM (at least in the present example).

Discussion

We have demonstrated the utility of GMM and its approximate equivalence to JTRM. However, perhaps one may still be concerned with the rationality of employing a linear multivariate regression model to analyze growth parameters because they have a nonlinear relationship with growth curves – the directly measurable trait values. However, this is not a problem. In fact, although the theory of modern quantitative genetics is basically based on linear models and proves very successful in explaining and predicting the genetic laws of quantitative traits, no one concerns about whether practical traits meet the linearity assumption or not, because no one ensures if there is really such a simple linear rela-

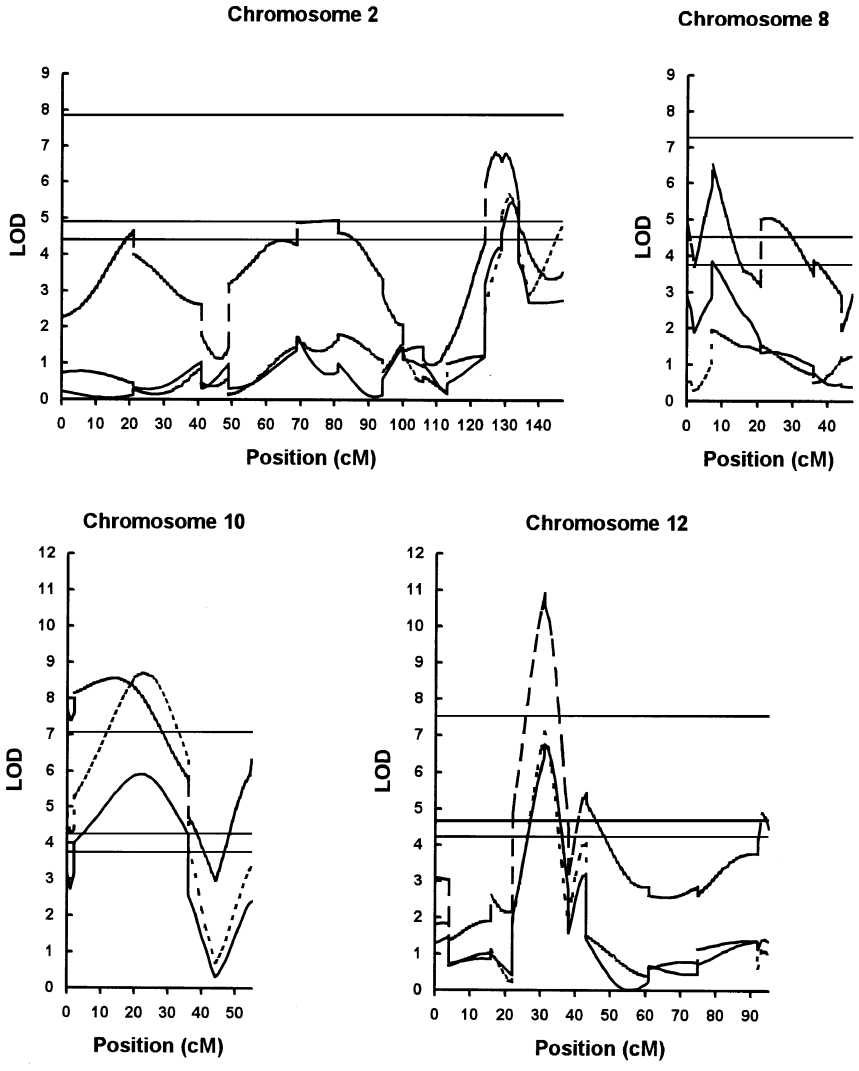

Fig. 4 LOD profiles yielded by JTRM (*long dashed lines*), GMM-I (*solid lines*) and GMM-II (*short dashed lines*). *Horizontal lines* indicate the thresholds at a genome-wise significance level of 0.05 for JTRM (top), GMM-I (bottom) and the GMM-II (middle)

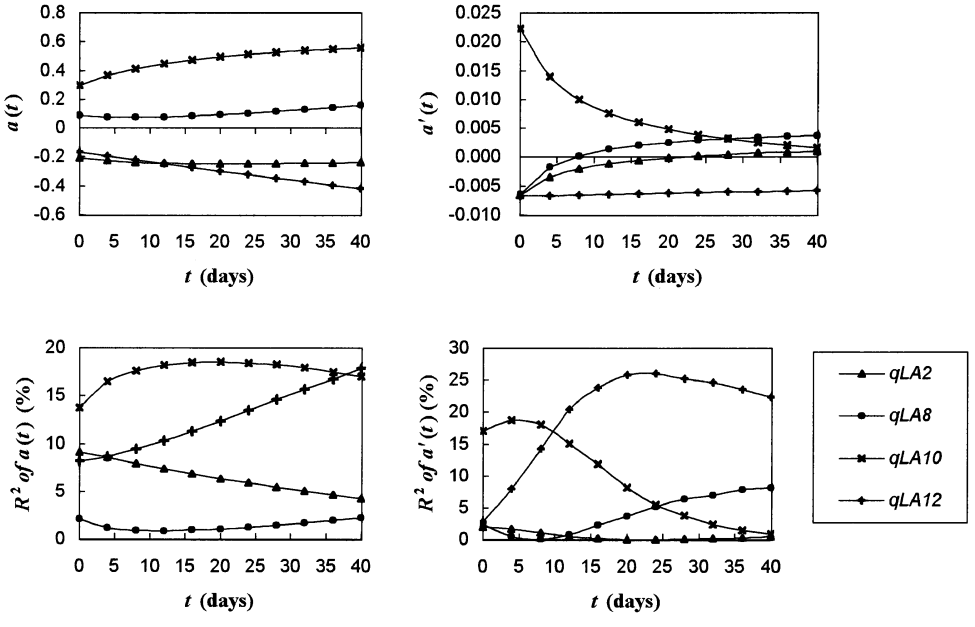

Fig. 5 Expression dynamics of the detected QTLs

tionship in nature. In addition, it is also impossible to expect all traits to meet the linearity assumption if some do. For instance, if we assume that sub-characters (e.g. grain length, width and thickness) follow linear genetic models, then their attributed super-character (e.g. grain weight ∝ length × width × thickness) will not, and vice versa. In short, whether linear or nonlinear, the real relationship between trait phenotypes and gene effects may not much matter for quantitative genetic analysis. Thus, nonlinear transformation of phenotype data (or scale transformation) is often used in practical studies as long as it is helpful for analyzing and understanding the genetic problems (Mather and Jinks 1982).

It should be noted that the growth model is a special nonlinear transformation because it does not transform one value to another value, but a curve to a vector. Here, we have actually taken the developmental process of a quantitative trait as a complex trait. But this complex trait is special because its phenotype does not appear as a single value but a curve or a vector. For ease of description, we would call it a *process trait*. Interestingly, it seems that a process trait must be treated as a whole in the GMM analysis. In the present example, we also conducted QTL mapping with individual growth parameters, but no QTL was detected. This implies that each growth parameter alone is not appropriate to be taken as a trait.

In comparison with JTRM, GMM has several advantages. First, GMM can greatly reduce the amount of phenotype data involved in the MCIM analysis. In the present example, the dimension of the trait value matrix used in JTRM was 111 (lines) × 11 (traits, i.e. observation times), while those used in GMM-I and GMM-II reduced to 111 (lines) × 3 (traits, i.e. growth parameters) and 111 × 4, respectively. The reduction of phenotype data for MCIM can greatly save computation time. This is particularly meaningful when permutation tests or simulation is performed to estimate significance thresholds.

Second, GMM can efficiently analyze unbalanced phenotype data while JTRM cannot. For instance, suppose that we used a DH population for time-related QTL mapping studies and planned to conduct a phenotypic investigation for a trait at a series of times (e.g. the 5th, 10th, 15th, 25th and 30th days after sowing). However, due to some unexpected causes, we could not fulfill the planned phenotype investigation on the 10th day and therefore had to investigate the residual lines on the next (11th) day. Thus, we obtained a set of unbalanced phenotype data. The phenotype data observed on the 10th and 11th days cannot be used together for JTRM, unless it is thought that the 1-day delayed investigation would not bring about errors. For GMM, however, there is no such a problem because both balanced and unbalanced data can be used for fitting growth models. So, GMM allows a more flexible way of phenotype investigation, which would be more convenient for practical studies. In the present example of rice, the leaf age data were actually unbalanced. This was another reason that we did not use the observed phenotype data for the JTRM.

Third, GMM may potentially help us better understand the genetic basis of quantitative trait development because the growth parameters may often have clear biological meanings. For example, in the human-height growth model proposed by Preece and Baines (1978), three out of five parameters have clear physiological meanings, i.e. adult height, approximate age (E) at which the pubertal growth spurt occurs, and height reached by the child at age E. It can be expected that the QTL effects on these parameters with clear biological meanings would possibly provide useful information about the genetic and the physiological mechanisms of quantitative trait development.

If we have no theoretical growth model for a trait, we have to look for an empirical model. Certainly, the most important criterion for identifying the best model is the goodness of fit. But sometimes, we may find several empirical growth models suitable for a trait. In this case, we may prefer to choose the simplest one with fewest parameters, since we have seen in the above example that using different growth models may not greatly affect the result of GMM, as long as the growth models can fit the growth data equally well.

In principle, the method of GMM is potentially applicable to studying any quantitative trait. GMM can be used to analyze the whole developmental process, as well as only a part (as in the present example) of a trait. Sometimes, a very complicated developmental process needs be divided into several phases and described by a multiphasic growth model (Koops 1986; Koops et al. 1987). In principle, such multiphasic growth models are also suitable for GMM analysis.

The principle of GMM can also be utilized to study QTLs underlying biological response to certain environmental stresses, such as cold, drought, submergence and irradiation. The response changes with the intensity (dose) of stress as a curve. In general, the dose-response curve can be described by a theoretical or empirical function containing several parameters. Hence, QTL mapping can be performed based on these parameters. Moreover, GMM can even be applied to a more complicated situation, where the development of a quantitative trait is exposed to a gradient of stress (environments) so that the trait-value growth process will appear as a curved surface in a three-dimensional space defined by time, stress and trait value, as long as an appropriate growth model can be identified.

Acknowledgements This work was supported in part by grants from the National Natural Science Foundation of China (project 39670411), the High Technology Research and Development Program of China (project 101-02-01-03) and the Rockefeler Foundation.

References

Bradshaw HD, Settler RF (1995) Molecular genetics of growth and development in populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. Genetics 139:963–973

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

Churchill GA, Doerge RW (1994) Empirircal threshold values for quantitative trait mapping. Genetics 138:963–971

France J, Dijkstra J, Thornley JHM, Dhanoa MS (1996) A simple but flexible growth function. Growth Dev Aging 60:71–83

Grizzle JE, Allen M (1969) Analysis of growth and dose response curve. Biometrics 25:357–381

Jiang C, Zeng Z-B (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127

Jolicoeur P, Pontier J, Pernin M-O, Sempe M (1988) A lifetime asymptotic growth curve for human height. Biometrics 44:995–1003

Kanefuji K, Shohoji T (1990) On a growth model of human height. Growth Dev Aging 54:155–165

Koops WJ (1986) Multiphasic growth curve analysis. Growth 50:169–177

Koops WJ, Grossman M, Michalska E (1987) Multiphasic growth curve analysis in mice. Growth 51:372–382

Lang S (1978) Analysis I (6th printing). Addison-Wesley Publishing Company, Reading, Massachusetts

Mather K, Jinks JL (1982) Biometrical genetics. University Press, Cambridge

Plomion C, Durel C-E, O'Malley DM (1996) Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. Theor Appl Genet 93:849–858

Preece MA, Baines MJ (1978) A new family of mathematical models describing the human growth curve. Ann Human Biol 5:1–24

Price AH, Tomos AD (1997) Genetic dissection of root growth in rice (*Oryza sativa* L.). II. Mapping quantitative trait loci using molecular markers. Theor Appl Genet 96:143–152

Richards FJ (1959) A flexible growth function for empirical use. J Exp Bot 10:290–300

Sondur SN, Manshardt RM, Stiles JI (1995) Genetics of growth rate and flowering time in papaya (*Carica papaya* L.). J Quant Trait Loci (http://probe.nalusda.goppv:8000/otherdocs/jqtl/jqtl1995-04/jq18r2.html/)

Verhaegen D, Plomion C, Gion J-M, Poitel M, Costa P (1997) Quantitative trait dissection analysis in Eucalyptus using RAPD markers. 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. Theor Appl Genet 95:597–608

Winsor SB (1932) The Gompertz curve as a growth curve. Proc Natl Acad Sci USA 18:1–8

Wu W-R, Li W-M, Tang D-Z, Lu H-R, Worland AJ (1999) Time-related mapping of quantitative trait loci underlying tiller number in rice. Genetics 151:297–303

Yan JQ, Zhu J, He CX, Benmoussa M, Wu P (1998a) Quantitative trait loci analysis for the developmental behavior of tiller number in rice (*Oryza sativa* L.). Theor Appl Genet 97:267–274

Yan JQ, Zhu J, He CX, Benmoussa M, Wu P (1998b) Molecular dissection of developmental behavior of plant height in rice (*Oryza sativa* L.). Genetics 150:1257–1265

Yang RC, Kozak A, Smith JHG (1978) The potential of Weibull-type functions as a flexible growth curves. Can J For Res 8:424–431

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zhou Y, Li W, Wu W, Chen Q, Mao D, Worland AJ (2001) Genetic dissection of heading time and its components in rice. Theor Appl Genet 102:1236–1242