Theor Appl Genet (2006) 113:288–300
DOI 10.1007/s00122-006-0294-2

ORIGINAL PAPER

QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato

M. Malosetti · R. G. F. Visser · C. Celis-Gamboa · F. A. van Eeuwijk

Received: 5 January 2006 / Accepted: 19 April 2006 / Published online: 20 May 2006
© Springer-Verlag 2006

Abstract The improvement of quantitative traits in plant breeding will in general benefit from a better understanding of the genetic basis underlying their development. In this paper, a QTL mapping strategy is presented for modelling the development of phenotypic traits over time. Traditionally, crop growth models are used to study development. We propose an integration of crop growth models and QTL models within the framework of non-linear mixed models. We illustrate our approach with a QTL model for leaf senescence in a diploid potato cross. Assuming a logistic progression of senescence in time, two curve parameters are modelled, slope and inflection point, as a function of QTLs. The final QTL model for our example data contained four QTLs, of which two affected the position of the inflection point, one the senescence progression-rate, and a final one both inflection point and rate.

Introduction

The availability of efficient molecular marker systems has facilitated breeders to identify quantitative trait loci (QTLs) underlying the expression of economically important traits in animals and crops. Typically, quantitative traits are observed at a fixed time point in development (e.g. at harvest in plants) and QTLs are detected by linking that phenotypic information with molecular data through adequate statistical models. Such QTL models describe the trait state at the measurement time as a function of molecular information (markers) reflecting the polymorphisms at the DNA level. The understanding of the genetic basis of a quantitative trait may profit from modelling not only the final status of the trait, but also the pattern of evolution of the trait during development. In this paper we present an example of how classical QTL models can be extended to describe final trait values as well as development by integrating QTL and physiological growth models.

Studying trait development requires first having assessments of the trait at several time points during the life cycle. A simple approach to the problem of modelling multiple observations in time on the same trait within the standard QTL analysis framework would be to consider the observations at individual time points as independent traits. The analysis then consists in identifying QTLs at individual time points, followed by a qualitative comparison of the detected QTLs across the time points (Bradshaw and Stettler 1995; Price and Tomos 1997; Verhaegen et al. 1997). An inconvenience of this approach is that instead of considering development as a continuous process in time, development is segmented into discrete observational time points. The absence of a formal integration of QTLs at individual time points within the same developmental model makes it difficult to arrive at biological sensible conclusions. For example, do QTLs for different time points that occur closely together on

Communicated by J. L. Jannink

M. Malosetti (✉) · R. G. F. Visser · C. Celis-Gamboa · F. A. van Eeuwijk
C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC), Laboratory of Plant Breeding, Wageningen University, P.O. Box 386, 6700 AJ Wageningen, The Netherlands
e-mail: marcos.malosetti@wur.nl

a chromosome represent a case of pleiotropy (one QTL expressing itself at more than one time point) or close linkage (two QTLs appearing closely together)? In addition to the biological limitations, the approach suffers from statistical limitations, as QTLs are detected and their effects estimated without considering the existing correlations between observations on successive time points. Wu et al. (1999) proposed to address the problem of the non-independence of trait values over time by using a multivariate regression procedure in which the observations at different time points constitute a set of responses, while the marker information generates a set of predictors. Although this approach alleviates the problem of ignoring the correlations between consecutive observations, still it is not biologically attractive as it describes the developmental process as a discrete collection of time points. Besides, it requires all the individuals to be measured at the same time points, which is often impossible.

From a biological point of view, there is a need for a unified modelling framework to investigate the genetics of trait development. In addition, such a modelling framework should be flexible enough to be adapted to the usually non-linear trait responses over time. Physiology-inspired growth models are attractive candidates to play a central role in that. Examples of commonly used models in biology are the linear and exponential growth models (that assume non-limiting growth), and the family of S-shaped curves where growth converges to a maximum (Schnute 1981). A desirable property of such growth models is that the state of the trait is described at any moment in development rather than at a discrete collection of time points at which the trait was actually observed. More importantly, the development process is described by a reduced number of curve parameters that can be interpreted in biological terms (e.g. the relative rate of growth, moment of maximum growth rate, etc). The variability in growth (development) trajectories between individual genotypes is reflected by genotype-specific curve parameters. A link between physiological models and QTL models can be established by modelling the genetic basis of the growth curve parameters in terms of QTLs. In a physiological QTL model, the phenotypic response is predicted from the curve parameters in combination with environmental input and/or time, where the curve parameters are linear functions of underlying QTLs, unspecified polygenes, and environmental and developmental noise. Examples of this approach have been presented recently for leaf growth in rice (*Oryza sativa* L.) (Wu et al. 2002b), stem diameter growth in trees (*Populus spp.*) (Ma et al. 2002), leaf elongation in maize (*Zea mays* L.) (Reymond et al. 2003), and flowering time in barley (*Hordeum vulgare* L.) (Yin et al. 2005).

A relatively straightforward approach to combine growth modelling with QTL analysis consists of the following two-step procedure. In the first step, observations at successive time points are used to estimate individual-specific parameters of a given growth curve model, where after in the second step, conventional QTL analysis is applied to the curve parameter estimates of the first step, interpreting these estimates as standard phenotypic traits. This strategy has been proven to produce satisfactory results (Reymond et al. 2003; Yin et al. 2005) and has as a strong point in favour, its relatively simple implementation. However, this two-step approach will be far from optimal when curve parameters are estimated imprecisely. The main drawback of the two-step procedure is that in the QTL analysis the uncertainty in parameter estimates is not taken into account and neither is the correlation between the parameters, with as a consequence possible loss of power for QTL detection and incorrect estimates and standard errors (Verbeke and Molenberghs 2000).

An immediate approach to the modelling of growth trajectories was formulated by Ma et al. (2002). They combine logistic growth curves and QTL mapping within a mixture model approach, modelling growth curves parameters as a function of molecular marker information. The approach is implemented within an expectation-maximization (EM) algorithm and proved to be powerful and to produce accurate estimates of QTL effects and positions (Wu et al. 2002a, 2003a). The methodology was further generalized to allow changing growth rates during development (Wu et al. 2003b).

As we consider mixture model approaches to be somewhat inflexible with respect to the inclusion of additional design and treatment features, we have developed an alternative to the modelling of growth curves that is based on a non-linear extension of classical mixed models. We start below with the description of some well known deterministic growth curves and proceed from there to the definition of flexible genetic models for growth curves in the context of non-linear mixed models, where curve parameters are governed by QTLs. We illustrate our methodology by an example consisting of multiple observations on leaf senescence in a diploid potato cross (*Solanum phureja* L. × *Solanum tuberosum* L.) (Celis-Gamboa 2002). We conclude on a multiple QTL model using composite interval mapping (Jansen and Stam 1994; Zeng 1994), thereby extending a single QTL non-linear mixed

model approach as used by Rodriguez-Zas et al. (2002) for mapping QTLs affecting lactation curves in diary cattle.

Materials and methods

Data

The potato phenotypic and molecular data used in this research was produced by Celis-Gamboa (2002) who conducted an extensive evaluation of a diploid offspring originating from an interspecific cross between *S. phureja* and *S. tuberosum*. We give a brief description of the cross and the data set that we used in this research, but for more detailed information (e.g. with respect to the genetic background of maternal and paternal genotypes of the cross) the reader is referred to the original publications (Celis-Gamboa 2002, 2003). Many traits were evaluated at various time points spaced at intervals of 1–2 weeks, but we concentrate on plant senescence as observed for 205 genotypes. Senescence was defined by a score on a scale from 1 to 7 that expressed the overall condition of the plant (1 = all leaves green and 7 = all leaves brown/yellow). Although measured at an ordinal scale, it turned out that for all practical and statistical purposes, senescence could be treated as a continuous variable. We used the evaluations taken at 64, 75, 89, 96, 110, 125, 140, 155, 170, and 185 days after planting. The population was genotyped on the basis of amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995). As AFLP is a dominant marker system, for which band presence and absence is scored, and as potato is an outbreeding crop species, a cross between a maternal and a paternal plant produces three types of polymorphic bands: (1) bands that are heterozygously present in the maternal genotype (*S. phureja*) and homozygously absent in the paternal genotype (*S. tuberosum*), (2) bands that are heterozygously present in both parents, and (3) bands that are homozygously absent in the maternal genotype and heterozygously present in the paternal parent. While markers in groups 1 and 3 segregate in a 1:1 ratio (presence:absence), those in group 2 segregate in a 3:1 ratio. Maps were constructed by Celis-Gamboa (2002) following the pseudo-testcross approach (Grattapaglia and Sederoff 1994), leading to one map for the *S. phureja* genome and another map for the *S. tuberosum* genome. In this paper, we used the *S. tuberosum* map, which contained 178 markers (type 3 markers) distributed over 12 linkage groups, spanning a total of 784 cM.

The phenotypic model

We will describe a general QTL methodology for developmental traits, but to keep notation and treatment simple we will restrict ourselves to the logistic curve, that in a preliminary study fitted the senescence curves well (Celis-Gamboa 2002). The logistic curve will serve as an example for the wider class of non-linear developmental curves in which other examples are the exponential, Gompertz and Richards curve. Assume that a generic, non-stochastic model for the progression of the senescence score for a potato genotype in our example data set is given by the following logistic curve:

$$f(t) = \gamma + \frac{\delta}{1 + \exp^{-\beta(t-\alpha)}} \tag{1}$$

where each of the four parameters has its own biological interpretation; α is the point in time when half of the process (cycle) has been completed and to which we will refer as mid-senescence, β represents the senescence progression rate at mid-senescence, to which we will refer below as rate, γ is the lower asymptote (minimum score value), and δ is the difference between the lower and the upper asymptote, difference between minimum and maximum score value (Fig. 1).

The generic model of Eq. 1 can be converted into a stochastic, genotype specific model by making the curve parameters dependent on the genotype and introducing an error term. When we write for the logistic function in Eq. 1, $f(t; \alpha_i, \beta_i, \gamma_i, \delta_i)$, and when we assume that the lower and upper asymptote are the same for all genotypes then the developmental process for genotype i becomes

$$\underline{y}_i(t) = f(t; \alpha_i, \beta_i, \gamma, \delta) + \underline{e}_i(t) \tag{2}$$

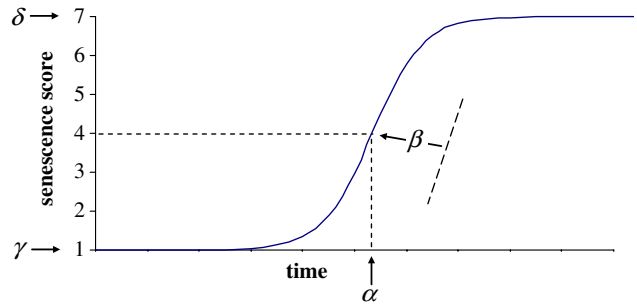

Fig. 1 Schematic representation of senescence progression and the relationship between the parameters and the curve characteristics

where the senescence score for genotype i at time point t, $\underline{y}_i(t)$, is a function of the fixed parameters α_i, β_i, γ, and δ. In the senescence example the parameters γ and δ are constrained to be equal for all genotypes since the minimum and maximum senescence scores were the same for all the genotypes (minimum score is 1 so $\gamma = 1$, and maximum score is 7 so $\delta = 6$). Therefore, these parameters will be omitted from model descriptions below. Finally, the error term $\underline{e}_i(t)$is assumed to be independently normally distributed with mean 0 and variance σ^2. For the type of data we are studying, some form of autocorrelation between subsequent measurements on the same plant would have been plausible. However, the observations in time for senescence on the potato genotypes came from different plants, as each time a small part of the experimental plots was harvested. As expected, for our senescence data we found no indication for autocorrelation in a model with genotype specific rates and mid-senescence times. Therefore, we refrain from modelling the autocorrelation for the residual error in this paper, although our description can be extended to allow for it. For the sake of clarity, here and hereafter, we will underline the random variables in the model formula.

Model (2) is an example of a fixed, non-linear model, where the only random term is the residual error. In such a fixed model, the estimation of the genotype specific curve parameters is based on the observations for the individual genotypes, no information from other genotypes is used. In contrast, within a mixed model formulation, parameter estimates would be based on a combination of observations for the specific genotype and observations on other genotypes. Furthermore, a mixed model would provide a more realistic model as the inaccuracy and imprecision of the parameter estimates as well as the correlation between the parameters is explicitly taken into account. A mixed model specification for senescence can then be:

$$\underline{y}_i(t) = f(t; \alpha_0 + \underline{a}_i^*, \beta_0 + \underline{b}_i^*) + \underline{e}_i(t), \quad (3)$$

where the main difference with model (2) is that the mid-senescence and rate parameters in model (3) are now modelled as a combination of fixed and random effects. Model (3) is a so-called subject-specific model (Davidian and Giltinan 2003), where the fixed parameters α_0 and β_0, for mid-senescence and rate respectively, are common to all genotypes, while the random deviations $\underline{a}_i^*$ and $\underline{b}_i^*$, for mid-senescence and rate respectively, are specific to the genotypes. For the random parameters $\underline{a}_i^*$and $\underline{b}_i^*$ we assume a multi-normal distribution with zero mean and variance-covariance matrix $\Sigma^* = \begin{bmatrix} \sigma^2_{a*} & \\ \sigma_{a*\,b*} & \sigma^2_{b*} \end{bmatrix}$. The residual term $\underline{e}_i(t)$is assumed to be independent and normally distributed with mean 0 and variance σ^2.

Model (3) is an example of a non-linear mixed model. Such models are very suitable for the simultaneous modelling of growth curves for a collection of organisms. A good overview of theory and applications of non-linear mixed models is given by Davidian and Giltinan (2003).

The QTL model: including molecular information in the model

In model (3), no genetic information other than the identification of the genotypes themselves is considered. The genetic variance for the mid-senescence and rate parameter, σ^2_{a*} and σ^2_{b*}, is caused by the variation in all the genes affecting the curve parameters. The inclusion of molecular marker information in the model would allow to evidence specific chromosomal regions (QTLs) as contributing to the genetic variation of those parameters. A straightforward extension of model (3) that accounts for variation due to QTLs consists in introducing an extra fixed term for the contribution of putative QTLs to the curve parameters. Hence, a single QTL mixed model, with the QTL affecting both mid-senescence time and rate is:

$$\underline{y}_i(t) = f(t; \alpha_0 + X_{im}\alpha_m + \underline{a}_i; \beta_0 + X_{im}\beta_m + \underline{b}_i) + \underline{e}_i(t), \quad (4)$$

where α_m and β_m represent the fixed effects of a putative QTL locus at position m on the mid-senescence and rate parameters respectively, and $\underline{a}_i$ and $\underline{b}_i$ are phenotypic individual-specific random residuals for both the curve parameters with mean zero and variance-covariance matrix $\Sigma = \begin{bmatrix} \sigma^2_a & \\ \sigma_{a\,b} & \sigma^2_b \end{bmatrix}$. The residual term $\underline{e}_i(t)$ is assumed to be independently normally distributed with mean 0 and variance σ^2. In the case of marker regression, i.e., QTL detection tests are performed only at genomic positions coinciding with marker loci (Lynch and Walsh 1998), the indicator variable X_{im} is a simple function of the observed genotype for marker locus m. For our potato example, X_{im} can be taken equal to 1 whenever an AFLP band is present, and the individual is actually heterozygous at that locus, while X_{im} is 0 for band absence, and the individual is homozygous at that locus. The estimated QTL effects for α_m and β_m can then be interpreted as the difference

in mid-senescence and rate, respectively, between individuals being heterozygous and homozygous.

Using the more powerful approach of interval mapping (Lander and Botstein 1989), requires the term X_{im} to give a probabilistic statement about the QTL genotypes at any particular chromosome position and not exclusively at marker positions. For most standard populations obtained from crosses between inbred lines, the information from flanking markers can be used to estimate at any chromosome position the conditional probabilities of the possible QTL genotypes (Lynch and Walsh 1998). For our potato data, X_{im} represents the probability of the heterozygous state of the QTL at position m, and the complement $(1-X_{im})$ is the probability of the homozygous state for the same QTL.

The QTL analysis: scanning the genome for QTLs

After the definition of a model framework, the next step is to identify within the set of all possible models the best one for the data under study. The vast amount of possible models makes an exhaustive search infeasible, so we need to define a strategy to guide us throughout the model space in search of the best model. There is no unique best strategy for doing this, and model-search strategies within QTL mapping are therefore still a matter of discussion. We propose a procedure in which we: (1) do a genome scan based on simple interval mapping (SIM) assuming QTLs will affect both rate and mid-senescence, (2) identify a set of potential cofactors (covariables to be used in further genome scans) based on the results of the SIM run, (3) choose a final set of cofactors for inclusion in composite interval mapping (CIM) by backward selection from the cofactor set of the previous step, (4) do CIM, and (5) select a final multi-QTL model by testing whether the selected QTLs from CIM indeed affect both rate and mid-senescence or only one of the two.

For the SIM step, we fit model (4) along the chromosomes at particular step length by maximum likelihood. For the potato data we used an interval of 4 cM (or shorter when consecutive markers mapped less than 4 cM apart). At each chromosome position, we test for the global effect of the QTL on the curve trajectory by a log-likelihood ratio test comparing a full model with a QTL affecting both rate and mid-senescence against a reduced model without QTL effects. The log-likelihood ratio is defined as:

$$\mathrm{LR} = -2\ln\left[\frac{\text{Likelihood reduced model}}{\text{Likelihood full model}}\right], \tag{5}$$

where ln stands for the natural logarithm. Statistical significance for the test can be assessed by comparing LR with a chi-square distribution with two degrees of freedom. A correction for multiple testing is required. Based on simulations (results not shown), we suggest a Bonferroni correction with a genome wide test level of $\alpha_g = \alpha/n$ with α the test level for an individual test and n = [length of genome in cM/10]. For example, with $\alpha = 0.05$ and a genome of 1,000 cM, $\alpha_g = 0.05/[1{,}000/10] = 0.0005$. The LR is plotted along the chromosomes to produce a profile where QTLs and their most probable positions are indicated by peak values exceeding the defined significance critical LR value. The peaks observed in the LR plots from the SIM analysis form a set of potential cofactors. Before doing a CIM analysis a definite set of cofactors is selected by a backward selection procedure starting from a model including all putative cofactors and then testing for the effect of removing each cofactor from the set by a LR test. The cofactor whose removal produces the lowest non-significant LR is removed after which the procedure is repeated until all cofactor produce a significant LR test after being removed from the set.

When applying CIM, background genetic variation caused by other QTLs is controlled by including cofactors in the model (Jansen and Stam 1994; Zeng 1994). A CIM scan will fit the following model along the chromosome:

$$\begin{aligned}\underline{y}_i(t) = & f(t;\, \alpha_0 + \sum_{c\varepsilon C} X_{ic}\alpha_c + X_{im}\alpha_m + \underline{a}_i;\, \beta_0 \\ & + \sum_{c\varepsilon C} X_{ic}\beta_c + X_{im}\beta_m + \underline{b}_i) + \underline{e}_i(t)\end{aligned} \tag{6}$$

where $\sum_{c\varepsilon C} X_{ic}\alpha_c$ and $\sum_{c\varepsilon C} X_{ic}\beta_c$ represent QTLs affecting mid-senescence and rate at positions other than the position under evaluation, while $X_{im}\alpha_m$ and $X_{im}\beta_m$ stand for the QTL under test. During a CIM genome scan a LR test is performed comparing model (6), at position m, with the following model (7):

$$\underline{y}_i(t) = f(t;\, \alpha_0 + \sum_{c\varepsilon C} X_{ic}\alpha_c + \underline{a}_i;\, \beta_0 + \sum_{c\varepsilon C} X_{ic}\beta_c + \underline{b}_i) + \underline{e}_i(t). \tag{7}$$

For our potato example, we chose to do a LR test at every marker position and at every 4 cM, starting from the closest marker. For evaluations close to a cofactor, we defined a window of 10 cM at either side of the cofactor within which the particular cofactor was temporarily removed from the model.

LR test profiles are produced for the overall test of the QTL affecting the curve trajectory, and QTLs are

identified at profile peak values provided that the peak exceeds the threshold LR value. The critical value for CIM is taken equal to that for SIM. The QTL model selected after CIM looks like:

$$\underline{y}_i(t) = f(t; \alpha_0 + \sum_{q \varepsilon Q} X_{iq}\alpha_q + \underline{a}_i; \beta_0 + \sum_{q \varepsilon Q} X_{iq}\beta_q + \underline{b}_i) + \underline{e}_i(t), \quad (8)$$

with Q the set of selected QTLs.

A final refinement consists in searching for more parsimonious QTL models by testing whether QTLs affects both mid-senescence and rate or only one of the two parameters. We compare the reduced models (9a) and (9b) with the full model (8)

$$\underline{y}_i(t) = f(t; \alpha_0 + \sum_{q \varepsilon Q*} X_{iq}\alpha_q + \underline{a}_i; \beta_0 + \sum_{q \varepsilon Q} X_{iq}\beta_q + \underline{b}_i) + \underline{e}_i(t), \quad (9a)$$

$$\underline{y}_i(t) = f(t; \alpha_0 + \sum_{q \varepsilon Q} X_{iq}\alpha_q + \underline{a}_i; \beta_0 + \sum_{q \varepsilon Q*} X_{iq}\beta_q + \underline{b}_i) + \underline{e}_i(t), \quad (9b)$$

where Q^* is equal to Q, the set of selected QTLs in CIM, minus the QTL for which the significance of mid-senescence or rate parameter is tested. The QTL effect on a particular parameter is unimportant whenever a non-significant LR follows from removal of that particular parameter from the model. Assuming the presence of a QTL for either or both of mid-senescence and rate at the test position, the LR statistic follows a Chi-square distribution with one degree of freedom.

Fitting the models

Non-linear mixed models like the ones discussed above can be fitted with the SAS macro NLINMIX that is available from anonymous ftp (http://www.support.sas.com/ctx/samples/index.jsp?sid=539) (Littell et al. 1996). We fitted our models using maximum likelihood in combination with the estimation method of Lindstrom and Bates (1990), which was specified in NLINMIX by the option settings 'method=ml' and 'expand=eblup'. To stabilize the variance of the rate parameter we worked with its log transform. Detailed descriptions and comparisons of estimation and inference procedures can be found in Davidian and Giltinan (2003). Before fitting the various models, we calculated genetic predictors, X_{im}, at a grid of positions along the genome following an algorithm described by Jiang and Zeng (1997). These predictors were then included in the models as explanatory variables.

Results

For our senescence data in potato, the LR profile for SIM showed peaks above the critical threshold on chromosomes V and VI (Fig. 2). Chromosome V was important in determining the senescence curve as it contained three significant peaks, at approximately 34, 59, and 86 cM (we will identify those positions as putative cofactors c51, c52, and c53, respectively). These peaks are a first indication that several putative QTLs could be located on chromosome V, albeit such conclusions should be taken with caution at this stage of the analysis. The reason is that this result was derived from a one-QTL model and we cannot rule out a 'ghost' QTL arising as a consequence of close proximity to neighbouring QTLs. Two peaks were observed close to the beginning of chromosome VI, at 0 and 17 cM, and we identify them as putative cofactors c61, and c62, respectively.

The backward selection procedure selected only some of the cofactors in the initial set to be included in the next round of mapping (Table 1). This was expected as the initial set of cofactors included points rather close to each other on the chromosome, so it was likely that some of them had turned up because of other QTLs nearby on the chromosome. The selected positions for cofactors were the points on chromosome V at 34 and 86 cM (c51 and c53, respectively), and the point at 0 cM on chromosome VI (c61).

In comparison with SIM, CIM showed a simpler picture on chromosome V, and allowed to detect an extra QTL on chromosome XII (Fig. 2). The lower level of the LR profile for CIM in comparison to SIM is due to the upward bias from neighbouring QTLs for SIM. On chromosome V, the two CIM peaks were at 34 and 90 cM. These two QTLs will hereafter be referred to as Q_{5A} and Q_{5B}, respectively. On the initial part of chromosome VI two peaks were observed, at 0 and 7 cM. Given the short distance between both peaks we are inclined to explain this by possible inaccuracies in marker positions on the map rather than by the presence of two QTLs in this region. Since the position at 0 cM gave the highest peak value, we identify the QTL at that position and we refer to it as Q_6. Finally, on chromosome XII one QTL was evidenced by a peak around 36 cM (Q_{12}). A second peak around 11 cM was just above the significance threshold. The final model from the CIM scan was thus a QTL model with five QTLs. However, the second QTL on chromosome 12, although significant in the CIM scan, disappears when comparing the final 5-QTL model with a 4-QTL model without this QTL by means of a LR test.

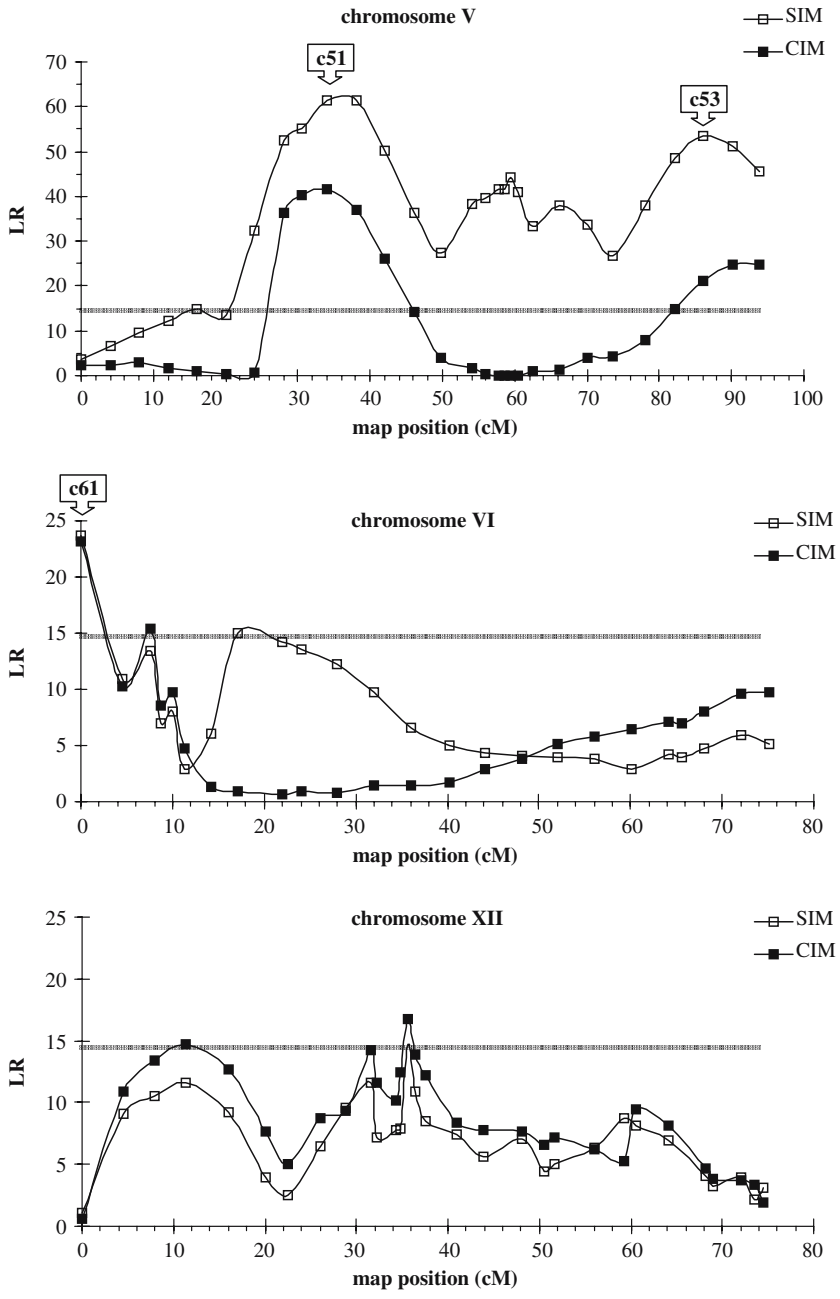

Fig. 2 LR profiles of simple interval mapping scan (*open squares*) and composite interval mapping (*filled squares*) scan for QTLs affecting the parameters describing plant senescence in potato. The profiles correspond to the LR statistic comparing a model with a QTL affecting the senescence curve and one without such a QTL. The *horizontal line* corresponds to the critical LR threshold corresponding to the cumulative upper limit of a chi-square distribution with two degrees of freedom and a genome-wise test level of 0.05 (14.6). The positions of the three cofactors c51, c53, and c61 are indicated by the *boxes with downward arrow heads*

In summary, after the CIM step, we have detected four QTLs affecting the leaf senescence trajectory curve. The aim of the next step is to answer the question how each of these QTLs affects the development curve. Three alternatives are possible: (a) the QTL only affects rate, (b) the QTL only affects mid-senescence, or (c) the QTL affects both rate and mid-senescence. The results of the LR used to compare the different situations for each of the four QTLs are presented in Table 2. For Q_{5A} the LR was not significant when removing the rate parameter indicating that the effect of this QTL on the rate is unimportant. Therefore, Q_{5A} affects the senescence curve by altering the mid-senescence parameter. The other QTL on chromosome V, Q_{5B}, affected both mid-senescence and rate since the drop of any of the corresponding terms from the model produced a significant LR test. In summary, on chromosome V there are two QTLs affecting senescence although in a different way, while the effect of Q_{5A} is related only to the moment at which half of the process is completed, Q_{5B} has an effect on both the moment at and the rate with which the senescence progresses once the process is launched. A different situation was observed for Q_6 that only

Table 1 Results of a backward selection procedure to exclude redundant cofactors from the QTL model. At each step the LR is calculated between the full model (the best from previous step) and a reduced model in which one cofactor (indicated with a minus sign) is removed. The best model in each step is indicated in bold. In the last step none of the models was superior to the full model and therefore none of them is in bold

	Cofactors in the model	–2LL[a]	LR[b]
Step 0	**c51 + c52 + c53 + c61 + c62**	**752.2**	
Step 1	–c51	775.7	23.4**
	–c52	**752.4**	**0.2**
	–c53	769.6	17.3**
	–c61	758.6	6.3*
	–c62	753.5	1.3
Step 2	–c51	789.4	37.0**
	–c53	774.8	22.4**
	–c61	759.1	6.7**
	–c62	**753.5**	**1.1**
Step 3	–c51	791.1	37.7**
	–c53	775.0	21.5**
	–c61	776.7	23.2**

*$P < 0.05$; **$P < 0.01$

[a] *LL* log likelihood

[b] LR=–2LL(reduced)–[–2LL(full)]

Table 2 LR between the full model (QTL affecting both rate and mid-senescence parameters) and a reduced model in which either the rate or the mid-senescence parameter are excluded from the model. The used critical LR value corresponds to the 0.99-upper limit of a chi-square distribution with one degree of freedom (6.63)

	LR after dropping QTL effect for	
	Rate	Mid-senescence
Q_{5A}	5.9	46.2**
Q_{5B}	20.2**	11.3**
Q_6	20.6**	4.2
Q_{12}	0.6	15.0**

**$P < 0.01$

affected the rate parameter (the drop of the mid-senescence parameter from the model produced a non-significant LR). Finally, Q_{12} affected only the mid-senescence parameter, a similar type of action as Q_{5A}. So, our final multi-QTL model included two QTLs affecting only mid-senescence (Q_{5A} and Q_{12}), one QTL affecting only the rate (Q_6), and one QTL affecting both parameters (Q_{5B}). From the estimated QTL effects we calculated predictions for the senescence values at individual time points and then correlated those predictions with observed senescence values for each genotype individually. The average squared correlation coefficient between predicted and observed senescence values across genotypes was 0.74, proving a satisfactory goodness of fit of the QTL model to the data.

In Table 3 we present the point estimates (and the corresponding 95 confidence intervals) of the model parameters associated with each QTL as provided by SAS. The parameters represent the difference between the heterozygous QTL genotype Q^tQ^p and the homozygous QTL genotype Q^pQ^p, with superscript *t* for *Solanum tuberosum* and *p* for *Solanum phureja*. For example, the heterozygous $Q^t_{5A}Q^p_{5A}$ reaches 50% development of senescence approximately 26 days earlier (effect of –25.8) than the homozygous $Q^p_{5A}Q^p_{5A}$. Conversely, for the other two QTLs affecting the mid-senescence parameter (Q_{5B} and Q_{12}), the estimates had a positive sign, meaning that the heterozygous genotypes $Q^t_{5B}Q^p_{5B}$ and $Q^t_{12}Q^p_{12}$ will attain the 50% of senescence approximately 15 and 12 days later than the homozygous genotypes $Q^p_{5B}Q^p_{5B}$ and $Q^p_{12}Q^p_{12}$ respectively. Before describing the QTL estimates associated with rate, it needs to be mentioned that the values presented in Table 3 are back-transformations from a logarithmic scale, so rather than being additive they represent multiplicative effects. Therefore, the heterozygous QTL genotype Q^tQ^p will increase the rate with respect to the homozygous Q^pQ^p genotype when the parameter value is larger than 1 and will reduce the rate when the value is smaller than 1. The rate parameter is reduced by a factor of 0.58 when at Q_{5B} the genotype contains one copy of the *S. tuberosum* allele ($Q^t_{5B}Q^p_{5B}$) in comparison to the genotype consisting of only *S. phureja* alleles ($Q^p_{5B}Q^p_{5B}$). In biological terms the lower the rate the slower the progression of senescence. Finally, for Q_6 the presence of the *S. tuberosum* allele determines a slower progression of the senescence process (a reduction in the rate parameter by a factor of 0.72) in comparison to the progression observed in the homozygous $Q^p_6Q^p_6$genotype.

An overall visualization of the QTL effects in time can be made by comparing the predicted progression of senescence for hypothetical genotypes differing in allele composition for the detected QTLs. For example, in Fig. 3a the comparison is made between two hypothetical genotypes differing only in the constitution of Q_{5A} while being homozygous Q^pQ^p for the other three QTLs. The plot shows that the heterozygous genotype $Q^t_{5A}Q^p_{5A}$ senesce earlier than the homozygous genotype $Q^p_{5A}Q^p_{5A}$ although the rate of progression is the same as both curves are parallel. The plot reflects the fact that this particular QTL does not affect the speed with which the senescence progresses, but only the timing of senescence. Figure 3b illustrates another situation in which the genotypes $Q^t_{5B}Q^p_{5B}$ and $Q^p_{5B}Q^p_{5B}$ differ in both the moment at which they reach 50% senescence (later in the heterozygous than in the homozygous genotype) and in the rate of senescence (faster in the homozygous than in the heterozygous

Table 3 Parameter estimates and the associated 95% confidence intervals from the fit of a model with four QTLs. The QTLs are located on chromosome V at 34 cM (Q_{5A}), and at 90 cM (Q_{5B}), on chromosome VI at 0 cM (Q_6), and on chromosome XII at 36 cM (Q_{12})

	Rate[a]			Mid-senescence		
	Estimate	Lower	Upper	Estimate	Lower	Upper
Intercept	0.43	0.35	0.52	105.9	97.5	114.3
Q_{5A}	NE	NE	NE	–25.8	–33.4	–18.3
Q_{5B}	0.58	0.46	0.72	14.6	7.2	21.9
Q_6	0.72	0.60	1.15	NE	NE	NE
Q_{12}	NE	NE	NE	11.7	5.8	17.7

NE no effect

[a]Back-transformed from a $\log_{10}$ scale, therefore the effects are multiplicative rather than additive

genotype). The plot in Fig. 3c shows the situation for Q_6 in which the inflection point is attained at approximately the same time but the rate is faster for the homozygous genotype ($Q_6^pQ_6^p$) than for the heterozygous genotype ($Q_6^tQ_6^p$). Finally, in Fig. 3d we present the comparison between genotypes with a different constitution at Q_{12}. While the rate of progression of senescence is similar for both the homozygous ($Q_{12}^pQ_{12}^p$) and the heterozygous ($Q_{12}^tQ_{12}^p$) genotypes (parallel trajectory curves), the process occurs earlier in the former than in the latter.

Discussion

The study of the genetics underlying the variation of economically important traits in plants and animals has been largely driven by models that describe traits at a fixed moment in time. The obvious reason for this is that, in general, traits are observed at the time of greatest importance, whether this is biological or economical. For example, plants typically are observed at fixed stages in development as anthesis, maturity, and harvest time. However, any phenotype is an integrated function over time with at least one of the arguments depending on gene (QTL) action under specific (and generally changing) environmental conditions. The inclusion of a time dimension in the QTL model would allow addressing questions related to the effect of QTLs throughout the life cycle. To that purpose, we presented a modelling framework that integrates growth models and QTL modelling. There are several reasons that make this approach attractive: (1) it brings together genetics and physiology, thereby gaining interpretability of results from both perspectives, (2) it provides breeders and geneticists with better opportunities to detect QTLs as the genetic variation in question is increased by considering the entire developmental process and not just the final trait state, (3) it provides breeders with insights into the most effective way to genetically shape trait development to attain breeding objectives, (4) it provides geneticists with clues about possible QTL functions that can be used to follow up on the search for candidate genes, (5) it often confers higher power for finding QTLs than analyses at individual time points.

To illustrate some of the above points, we compared our non-linear mixed model analysis of senescence with a series of individual time point analyses. For the analyses per time point we took a procedure typically followed by breeders, namely a CIM analysis by a special purpose QTL package, in our case MapQTL (van Ooijen 2004). LR profiles for CIM analyses by MapQTL of individual time points are shown in Fig. 4 together with the LR profile of the non-linear mixed model analysis. The LR profiles for both types of analysis are comparable as they involve two parameters in both cases, although not the same ones, of course. In Fig. 4, genome wide test levels of 0.05 are included for both the individual time point analyses, where a Bonferroni correction for multiple traits has been applied, and the non linear mixed model analysis. The set of time-point analyses revealed less QTLs than the non-linear mixed model analysis. Only the QTLs on chromosome V were detected by some of the time point analyses, while neither the QTL on chromosome VI nor the one on chromosome XII was detected at any of the time points. In addition, the biological information of this latter QTL analysis is rather limited and difficult to interpret in comparison to our non linear mixed model approach. For example, the QTL on chromosome V had an effect on days 89, 96, 110 (the largest effect), and 125, but not on days 75 and 155. Questions on pleiotropy versus close linkage also arise easily from individual time points analyses, without there being a simple way to resolve them.

By modelling leaf senescence in potato, we showed that we were able to not only identify four different

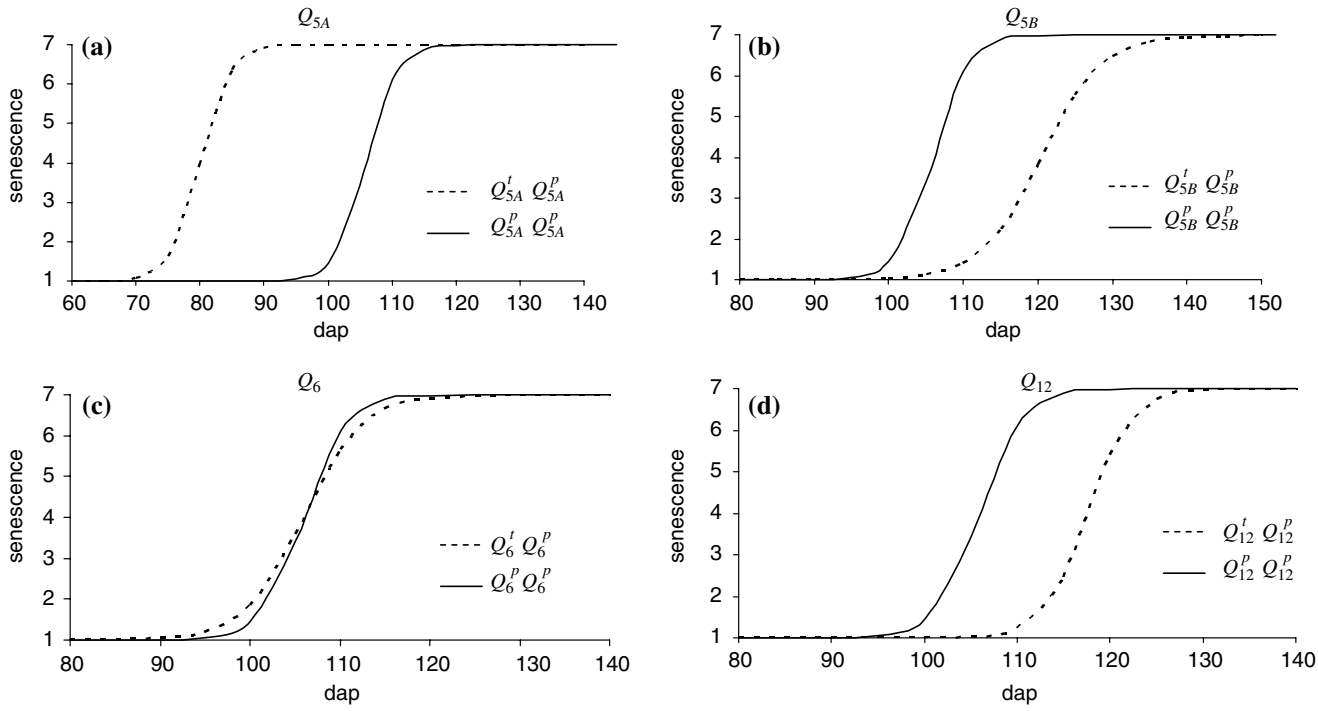

Fig. 3 Predicted senescence development of genotypes differing in genotype for each of the four detected QTLs; **a** Q_{5A}, **b** Q_{5B}, **c** Q_6, and **d** Q_{12}. The two alternative genotypes at the QTLs are heterozygous with one allele from *S. tuberosum* and the other from *S. phureja* (Q^tQ^p), and homozygous carrying both alleles from S. phureja (Q^pQ^p)

QTLs affecting the process of leaf senescence, but also to distinguish between QTLs that delay senescence and QTLs that affect the rate with which senescence advances. The distinction between such types of QTL effects reflects two types of senescence-response associated with stay-green genotypes: a delayed onset of senescence (also called type A response), and a slow progression of senescence (or type B response) (Thomas and Howarth 2000). The underlying mechanisms for these two types of senescence responses are probably different. For example, in *Sorghum bicolor L.* some stay-green genotypes show type A response while others show type B response (Borrell et al. 2000). The fact that we found QTLs determining delayed senescence and QTLs affecting the rate of progression that largely did not co-locate is in accordance with possibly different underlying mechanisms affecting the rate and the timing of senescence.

Selection of QTL alleles affecting senescence curve parameters should be more effective than the selection of alleles for QTLs detected by analysis of senescence responses at individual time points. From a plant breeder's point of view, the independence of QTLs shaping different aspects of the senescence curve creates good prospects for independent selection of the underlying parameters. From the point of view of a geneticist, interested in identifying and ultimately cloning genes, these results provide insights that can be used for finding relations between the detected QTLs and earlier described so-called senescence-associated genes (SAGs). For example, in the model species *Arabidopsis thaliana* different regulatory genes have been described for the onset of senescence and for the rate of senescence, respectively (Gepstein et al. 2003).

The statistical methodology used in this paper, a one-step approach modelling parameters for growth curves and underlying QTLs simultaneously, is technically more complex than a two-step approach in which first curve parameters are estimated for each genotype individually and next these curve parameter estimates are introduced as traits in a traditional QTL analysis. Although the philosophy behind both approaches is similar, the one-step procedure, in our case based on a mixed model formulation, has an advantage over the two-step approach from the inferential point of view. The statistical argument for a mixed model is that it offers extra flexibility for modeling the data more realistically by accounting for the different sources of variation and the resulting correlation structures present in the data. This does not mean that the two-step strategy cannot lead to satisfactory results. The differences between the performances of both approaches will depend on the complexity of the data under study.

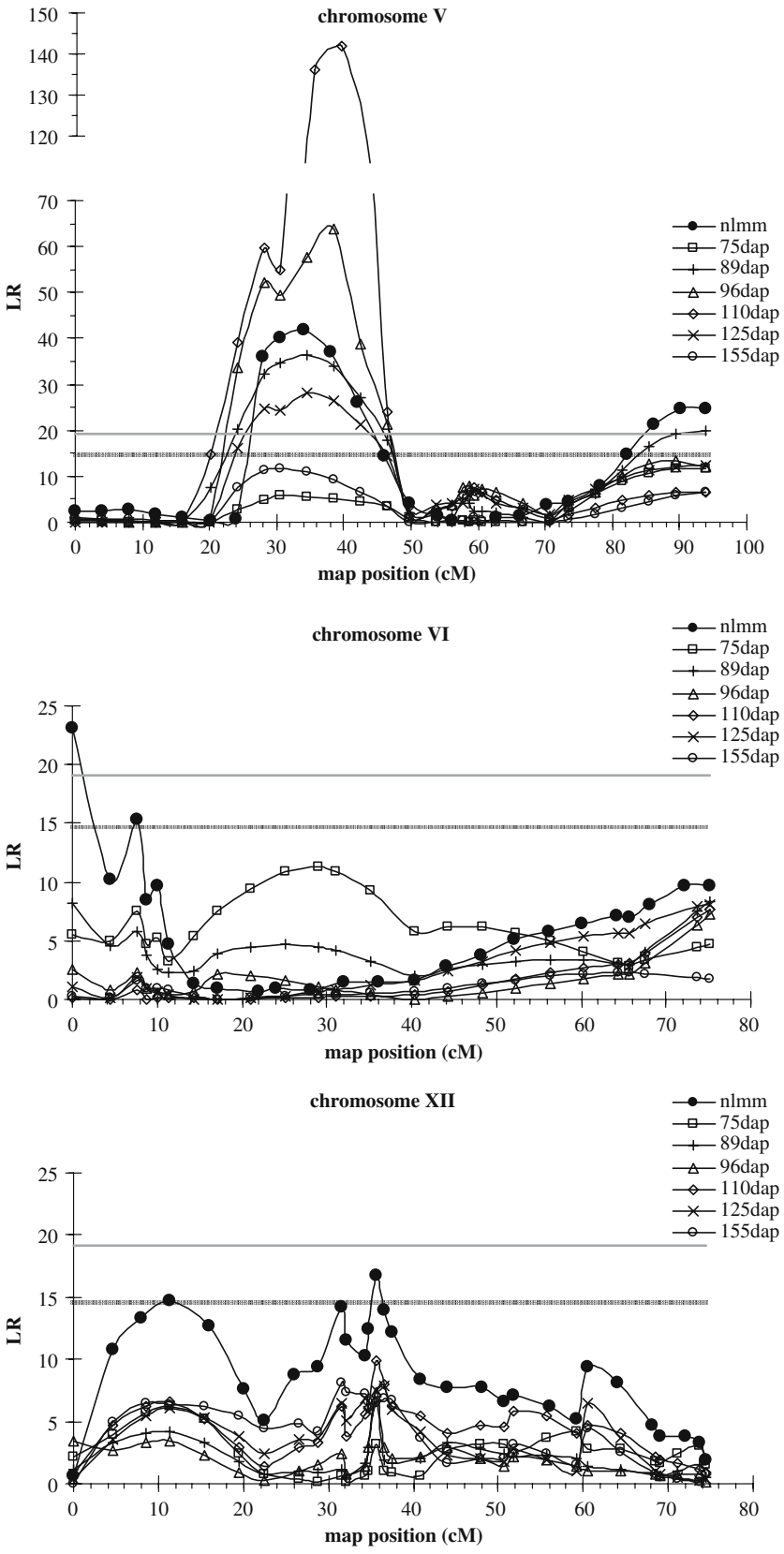

Fig. 4 LR profiles of a composite interval mapping scan based on a non-linear mixed model for senescence and standard composite interval mappings for senescence as measured at a number of time points during development. Thresholds for the LR statistic at a genome-wise level of 0.05 are given as *horizontal lines* with the *upper grey line* for the individual time point analyses and the *lower black line* for the non-linear mixed model analysis (*dap* days after planting, *nlmm* non-linear mixed model)

The advantages of the one-step non-linear mixed model approach come at the price of a higher demand on statistical modeling skills and higher computing loads for inference. It is encouraging, however, to observe that for both the theory and practice of non-linear mixed models the literature grows quickly (Davidian and Giltinan 2003), so that for practitioners more examples come available as well as more user friendly software. The estimation procedure for non-linear mixed models that we used for our potato example, is widely used in practice and it is acknowledged to work well (Davidian and Giltinan 2003). The procedure has been shown to be robust against non-normality of the random effects and possible model misspecifications (Hartford and Davidian 2000).

As a competitor to the use of non-linear mixed models for the QTL modeling of growth trajectories, mixture model approaches using EM estimation procedures have been proposed (Ma et al. 2002; Wu et al. 2002a, 2003a). The difference between the mixture model approach and our mixed model approach is that we approximate the mixture density for the phenotype in relation to the possible QTL genotypes for a particular place at the genome (Jansen 1992, 1993; Zeng 1993, 1994) with a normal density following the regression approach of Haley and Knott (1992) and Martínez and Curnow (1992). Although, the mixture model is more exact, in practice the differences between mixture and regression/mixed model approaches are small (Haley and Knott 1992; Kao 2000). An attractive advantage of regression/mixed models is the higher flexibility for including multiple environments, genotype by environment interaction and QTL by environment interaction (Malosetti et al. 2004), and experimental design details (Smith et al. 2001; Verbyla et al. 2003), because only standard statistical software with regression/mixed model facilities is required in which it is comparably easy to update model definitions.

The mixed model approach for identifying the genetic factors governing development can also be applied for investigating the relation between phenotype and genotype in its dependence on environmental factors. The approach is compatible with the philosophy of gene-to-phenotype models like the ones presented by Reymond et al. (2003) and Yin et al. (2005), a promising field for the understanding of complex genotype by environment interaction. The inherent complexity of quantitative trait performance under different environmental circumstances requires more elaborate models without losing biological relevance. It is precisely in this context that the fruitful combination of statistical and eco-physiological models is put forward as a promising tool (Tardieu 2003; van Eeuwijk et al. 2005).

A challenge that we took up in this paper is to make physiologists, geneticists, and breeders acquainted with the possibilities for including developmental trajectories in QTL analysis by means of non-linear mixed models. We feel that the advantages for biological interpretation of our approach outweigh the extra efforts in statistics.

Acknowledgments We are grateful to Cajo ter Braak and H-P Piepho for their discussion and comments on the manuscript. We also thank the Editor and two anonymous reviewers whose comments helped to improve the final version of the manuscript.

References

Borrell AK, Hammer GL, Douglas ACL (2000) Does maintaining green leaf area in Sorghum improve yield under drought? I leaf growth and senescence. Crop Sci 40:1026–1037

Bradshaw HD, Stettler 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

Celis-Gamboa C (2002) The life cycle of the potato (*Solanum tuberosum* L): from crop physiology to genetics. Ph.D. Thesis. Wageningen University, Wageningen

Celis-Gamboa C, Struik PC, Jacobsen E, Visser RGF (2003) Temporal dynamics of tuber formation and related processes in a crossing population of potato (*Solanum tuberosum*). Ann Appl Biol 143:175–186

Davidian M, Giltinan DM (2003) Nonlinear models for repeated measurement data: an overview and update. J Agric Biol Environ Stat 8:387–419

van Eeuwijk FA, Malosetti M, Yin X, Struik PA, Stam P (2005) Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. Aust J Agric Res 56:883–894

Gepstein S, Sabehi G, Carp M-J, Hajouj T, Nesher MFO, Yariv I, Dor C, Bassani M (2003) Large-scale identification of leaf senescence-associated genes. Plant J 36:629–642

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121–1137

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hartford A, Davidian M (2000) Consequences of misspecifying assumptions in nonlinear mixed effects models. Comput Stat Data Anal 34:139–164

Jansen RC (1992) A general mixture model for mapping quantitative trait loci by using molecular markers. Theor Appl Genet 85:252–260

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205–211

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455

Jiang CJ, Zeng ZB (1997) Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica 101:47–58

Kao CH (2000) On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. Genetics 156:855–865

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

Lindstrom MJ, Bates DM (1990) Nonlinear mixed effects models for repeated measures data. Biometrics 46:673–687

Littell RC, Milliken GA, Stroup WW, Wolfinger RR (1996) SAS system for mixed models. SAS Institute Inc, Cary

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland

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

Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137:139–145

Martínez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–488

van Ooijen JW (2004) MapQTL® 5, software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V., Wageningen, The Netherlands

Price AH, Tomos AD (1997) Genetic dissection of root growth in rice (*Oriza sativa* L) II: mapping quantitative trait loci using molecular markers. Theor Appl Genet 95:143–152

Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol 131:664–675

Rodriguez-Zas SL, Southney BR, Heyen DW, Lewin HA (2002) Detection of quantitative trait loci influencing dairy traits using a model for longitudinal data. J Dairy Sci 85:2681–2691

Schnute J (1981) A versatile growth model with statistically stable parameters. Can J Fish Aquat Sci 38:1128–1140

Smith AB, Cullis B, Appels R, Campbell AW, Cornish GB, Martin D, Allen HM (2001) The statistical analysis of quality traits in plant improvement programs with application to the mapping of milling yield in wheat. Aust J Agric Res 2:1207–1219

Tardieu F (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8:9–14

Thomas H, Howarth CJ (2000) Five ways to stay green. J Exp Bot 51:329–337

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

Verbyla A, Eckermann PJ, Thompson R, Cullis B (2003) The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. Aust J Agric Res 54:1395–1408

Verhaegen D, Plomion C, Gion J-M, Pointel M, Costa P, Kremer A (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

Vos P, Hogers R, Bleeker M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

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

Wu R, Ma C-X, Chang M, Littell RC, Wu SS, Yin T, Wang M, Casella G (2002a) A logistic mixture model for characterizing genetics determinants causing differentiation in growth trajectories. Genet Res 79:235–245

Wu W, Zhou Y, Li W, Mao D, Chen Q (2002b) Mapping of quantitative trait loci based on growth models. Theor Appl Genet 105:1043–1049

Wu R, Ma C-X, Yang MC, Chang M, Littell RC, Santra U, Wu SS, Yin T, Huang M, Wang M, Casella G (2003a) Quantitative trait loci for growth trajectories in *Populus*. Genet Res 81:51–64

Wu R, Ma C-X, Zhao W, Casella G (2003b) Functional mapping for quantitative trait loci governing growth rates: a parametric model. Physiol Genomics 14:241–249

Yin X, Struik PC, van Eeuwijk FA, Stam P, Tang J (2005) QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. J Exp Bot 56:967–976

Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90: 10972–10976

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