Regression Analysis for Mapping Dynamic Trait QTL in Outbred Population

Huijiang Gao[1], Runqing Yang[2], Wenzhong Zhao[1], Hua Sun[1], Shizhong Xu[3], Yuchun Pan[2]

1. School of Animal Science, Northeast Agriculture University, Harbin, Heilongjiang 150030, China
2. School of Agriculture and Biology, Shanghai Jiaotong University, Shanghai 201101, China
3. University of California, Riverside, CA 92521, USA

Abstract: By using Legendre polynomials to model dynamic changes of each genetic effect, a mathematic model was constructed for mapping dynamic trait loci. Other than the number of estimated effects increasing several times, there were no differences between the mapping principles for dynamic trait and the general quantitative trait loci underlying the same genetic design. Therefore, given the prerequisite that, residual errors at each observed point have the same distribution and are independent from each other, it is feasible to use regression analysis (REG) as an alternative of maximum likelihood (ML) for simplifying the estimated model parameters and detects QTL for dynamic traits. The new strategy for mapping dynamic trait loci is developed in the context of an outbred population with three-order Legendre polynomial as a sub-model. The properties of regression analysis are demonstrated and compared with maximum likelihood via replicated Monte Carlo simulations. The factors considered include individual number, test-day frequency, marker frequency and accumulative heretability, the level combinations of which were based on orthogonal table. It showed that REG performs equally well with ML, but deficient in accuracy and precision of residual variance estimating. The detection power of both methods depends greatly on size of accumulative heretability. [Nature and Science. 2004;2(4):68-78].

Key Words: dynamic trait; QTL mapping; Legendre polynomial; regression analysis; simulation

1. Introduction

The growth development of tissues and organs and change of performance with time of life or certain quantitative factors belong to the category of dynamic trait, which is of major interest to developmental genetics in plant and animal breeding studies.

Existing methods for genetic mapping of a dynamic trait can be synoptically divided into two classes: dynamic point-based method and dynamic model-based method. The dynamic point-based method treated each dynamic point as a different trait and genetically map dynamic trait loci by using separate analysis (Cheverud *et al.* 1996; Nuzhdin *et al.* 1997; Verhaegen *et al.* 1997; Emebiri *et al.* 1998; Wu *et al.* 1999), mutitrait analysis (Jiang & Zeng 1995; Korol *et al.* 1995; Ronin *et al.* 1995; Eaves *et al.* 1996; Knott and Haley 2000) or conditional analysis for these phenotypes at each dynamic points (Yan et al, 1998a, b; Wu et al, 2002). However, this analysis approach can just provide partial estimates of genetic control over dynamic traits, because it was not possible to capture the whole information about infinite dynamic points in the dynamic process. Also, it is required to observe in equilibrium phenotypes at dynamic points from each individual except separate analysis. A dynamic model-based method proposed recently attempted to detect QTL affecting the changing process of dynamic trait via analyzing the phenotypic function (curve) on time in life and other quantitative factors. In papers just published, Rodriguez-Zas *et al*. (2002) developed a longitudinal-linkage analysis approach for mapping QTL affecting the shape and scale of lactation curves for production and health traits in dairy cattle in outbred population. By combining growth models with QTL mapping, W. R. Wu et al (2002) proposed an approach of QTL mapping based on growth models. Both two approaches are archived by two steps: the first is to fit the dynamic curve of each individual or line with a theoretical or empirical mathematical model and then map QTLs based on the estimated model parameters. As a rule, two-step estimation

performs lower statistical power because the estimated errors of model parameters are not considered in the second step. Based on functional mapping strategy (Ma *et al* 2002; Wu et al 2002; 2003), the models for dynamic trait mapping QTL were established by using Logistic curves to describe the effects of QTL genotypes on the growth process of a forest tree. Although effects of QTL genotypes can be estimated by one step, it is difficult to estimate the additive and interaction effects of QTL and fail to extent to arbitrary population for estimating simultaneously multiple QTL genetic effects. Enlightened by the random regression model (Henderson 1982; Schaeffer & Dekkers, 1994; Schaeffer, 2002), said the Legendre polinomial may be nested into each QTL effect in general QTL mapping model to describe the change of QTL effects on dynamic traits and so the new mathematic model for mapping dynamic trait loci can be constructed for mapping one or multiple QTLs in various population (Yang et al, 2003).

Maximum likelihood, implemented with the EM algorithm, was used at the beginning of interval mapping advent. It can fully capture the information of data distribution, but could not be implemented in complex data structure because of the difficulty in computations. Haley and Knott (1992) introduced regression analysis method to interval mapping by substituting the probabilities of each QTL genotype under the markers for the indicator variable of each genotype and proved that it can save much time in computation and get similar results as ML. Other than the number of estimated effects increased several times, there were no differences between the mapping principles for dynamic trait and the general quantitative trait under the same genetic design. The objective in this study is to apply regression analysis to simplify and estimate model parameters and detect QTL for dynamic traits and also compare with maximum likelihood, simulation experiments in outbred populations.

2. Mapping principle

2.1 Genetic design

The population exampled to demonstrate the application of QTL mapping for dynamic trait loci in animals in this study is the full-sibs of outbred cross, which is most general in animal populations. Under Mendelian inheritance laws, there are four possible genotypes denoted by $Q_1^sQ_1^d$, $Q_1^sQ_2^d$, $Q_2^sQ_1^d$ and $Q_2^sQ_2^d$. If Legendre polynomial is used to model the genetic change with time or other factors of dynamic trait or the effect of QTL, the statistical model describe the phenotypic value of individual i at j test-day may be

$$y_{ij} = \mathrm{T}'_{ij}\mu + X_{i1}\mathrm{T}'_{ij}\mathrm{g}_1 + X_{i2}\mathrm{T}'_{ij}\mathrm{g}_2 + X_{i3}\mathrm{T}'_{ij}\mathrm{g}_3 + X_{i4}\mathrm{T}'_{ij}\mathrm{g}_4 + e_{ij}$$
$$= \mathrm{T}'_{ij}\mu + \sum_{k=1}^{4} X_{ik}\mathrm{T}'_{ij}\mathrm{g}_k + e_{ij}$$

$i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, m$; $k = 1, 2, 3, 4$.

Where, $\boldsymbol{\mu}$ is the regression coefficient vector of the Legendre polynomial for population means; $\mathbf{T}'_{ij} = \begin{bmatrix} 1 & P_1(\tau_{ij}) & \cdots & P_s(\tau_{ij}) \end{bmatrix}'$ with τ and s representing standardized time t and the order of Legendre polynomial selected, respectively; X_{ik} is the indicator variable of QTL genotype, valued 1 when the genotype of individual i is the k or 0 when the other. g_k denotes vector of fixed regression effects corresponding to genotype k.

2.2 Statistical method

Maximum likelihood. The phenotypes of the traits at all time points for each genotype group follow a multivariate normal distribution:

$$f_k(y_{ij}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_{ij} - \mathrm{T}'_{ij}\mu - \mathrm{T}'_{ij}\mathrm{g}_k)^2}{2\sigma^2} \right]$$

Actually, the genotype of individual i is unobservable, and it must be inferred by markers as conditional probability. Then the likelihood function of the outbred cross with m-dimensional measurements can be represented by a multivariate mixture distribution:

$$L(\mu, \mathrm{g}, \sigma^2) = \prod_{i=1}^{n}\prod_{j=1}^{m}\left[\sum_{k=1}^{4} p_{ik} f_k(\mathrm{y}_{ij}) \right]$$

The unknown parameters about QTL can be estimated using maximum likelihood implementing the EM algorithm. The steps of EM algorithm can be

recapitulated as

(1) Set the initial values of each parameter $\Omega^{(0)} = [\mu^{(0)}\ g^{(0)}\ \sigma^{2(0)}]$;

(2) Calculate the posterior probabilities of each QTL genotype $p_{ik} = \frac{p_{ik} f_k(\mathrm{y}_{ij})}{\sum_{k=1}^{4} p_{ik} f_k(\mathrm{y}_{ij})}$;

(3) E-step: calculate the expectations under the posteriors;

(4) M-step: estimate the MLE of each parameter.

Let $\mathrm{g}' = [\mathrm{g}_1 \quad \mathrm{g}_2 \quad \mathrm{g}_3 \quad \mathrm{g}_4]$,

$$\mathrm{A} = \begin{bmatrix} \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p'_{i1}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p'_{i2}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p'_{i3}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p'_{i4}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} \\ & \sum_{i=1}^{n}p'_{i1}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij_{ij}} & 0 & 0 & 0 \\ & & \sum_{i=1}^{n}p'_{i2}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & 0 & 0 \\ & & & \sum_{i=1}^{n}p'_{i3}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & 0 \\ \textit{sym.} & & & & \sum_{i=1}^{n}p'_{i4}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} \end{bmatrix}$$

$$\mathrm{C}' = \begin{bmatrix} \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} & \sum_{i=1}^{n}p'_{i1}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} & \sum_{i=1}^{n}p'_{i2}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} & \sum_{i=1}^{n}p'_{i3}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} & \sum_{i=1}^{n}p'_{i4}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} \end{bmatrix}$$

then $[\mu^{(1)} \vdots\ \mathrm{g}^{(1)}] = \mathrm{A}^{-1}\cdot\mathrm{C}$ $\quad \hat{\sigma}^{2(1)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{4}p'_{ik}\sum_{j=1}^{m}(y_{ij} - \mathrm{T}_{ij}\mu^{(1)} - \sum \mathrm{T}_{ij}\mathrm{g}_k^{(1)})^2$.

With $\mathbf{\Omega}^{(0)}$ substituted by $\mathbf{\Omega}^{(1)}$, iterations are then made between step (2), (3) and (4) and terminated when a predetermined criterion is satisfied. The final estimates are to be the MLEs.

In practical computations, the QTL position parameter θ can be viewed as fixed because a putative QTL can be searched at every 1 or 2 cM on a map interval bracketed by two markers throughout the entire linkage map. The amount of support for a QTL at a particular map position is often displayed graphically through likelihood profiles. The position of the largest and most significant is considered the most possible that the putative QTL lying.

Regression analysis. Different from ML, the indicator variables for each genotype in the statistical model are first replaced by their conditional expectations given marker information, i.e.,

$$y_{ij} = \mathrm{T}'_{ij}\mu + \sum_{k=1}^{4}\hat{X}_{ik}\mathrm{T}'_{ij}\mathrm{g}_k + e_{ij}\ .$$

Where, $\hat{X}_{ik} = E(X_{iki} \mid \mathrm{MN})$, MN represented the flanking markers. The estimation and statistical test of QTL effects can be accomplished using conventional multiple regression analysis by treating $\hat{X}_{ik}$ as observed values.

The conditional probabilities of each QTL genotype, $\Pr(\mathrm{Q} \mid \mathrm{MN})$, were calculated by multipoints estimation method (Rao and Xu, 1998; Xu et. 1998).

Let $\Pr(Q \mid \mathrm{MN}) = p_{ik}$, then $\hat{X}_{ik} = E(X_{ik} \mid \mathrm{MN}) = [p_{i1} \quad p_{i2} \quad p_{i3} \quad p_{i4}]$, and

$$\mathrm{A}=\begin{bmatrix} \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i1}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i2}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i3}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i4}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} \\ & \sum_{i=1}^{n}p_{i1}^{2}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i1}p_{i2}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i1}p_{i3}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i1}p_{i4}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} \\ & & \sum_{i=1}^{n}p_{i2}^{2}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i2}p_{i3}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i1}p_{i4}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} \\ & & & \sum_{i=1}^{n}p_{i3}^{2}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} & \sum_{i=1}^{n}p_{i3}p_{i4}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} \\ sym. & & & & \sum_{i=1}^{n}p_{i4}^{2}\sum_{j=1}^{m}\mathrm{T}'_{ij}\mathrm{T}_{ij} \end{bmatrix}$$

$$\mathrm{C}'=\begin{bmatrix} \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} & \sum_{i=1}^{n}p_{i1}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} & \sum_{i=1}^{n}p_{i2}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} & \sum_{i=1}^{n}p_{i3}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} & \sum_{i=1}^{n}p_{i4}\sum_{j=1}^{m}\mathrm{T}'_{ij}y_{ij} \end{bmatrix}$$

The estimates of each parameter can be calculated by

$$[\mu\vdots\ \ \mathrm{g}]=\mathrm{A}^{-1}\cdot\mathrm{C}$$

$$\hat{\sigma}^{2}=\frac{1}{nm-nk}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\mathrm{T}_{ij}\mu-\sum_{k=1}^{4}p_{ik}\mathrm{T}_{ij}\mathrm{g}_{k})^{2}$$

Ulteriorly, the means to detect the possible position of the putative QTL in linkage map is the same as the ML's.

2.3 Estimation for genetic effects

In an outbred family, the genotypic value of a progeny can be partitioned according to the following genetic model,

$$g_{kij}=a_{ki}^{s}+a_{kj}^{d}+d_{kij}\quad(i=1,2;\ j=1,2).$$

Where a_{k1}^{s} and a_{k2}^{s} are the effects of the two alleles in the sire, a_{k1}^{d} and a_{k2}^{d} are the effects of the two alleles in the dam, and d_{kij} are the dominance deviations. Because these effects are not estimable, some constraints are required. These constraints are

$$\sum_{i=1}^{2}a_{ki}=\sum_{j=1}^{2}a_{kj}=\sum_{i=1}^{2}d_{kij}=\sum_{j=1}^{2}d_{kij}=0.$$

Under these constraints there are three independent estimable effects, which are conveniently defined as:

（1） $a_{k}^{s}=a_{k1}^{s}-a_{k2}^{s}$, the average effect of allelic substitution for the sire;

（2） $a_{k}^{d}=a_{k1}^{d}-a_{k2}^{d}$, the average effect of allelic substitution for the dam;

（3） $d_{k}=d_{k11}-d_{k21}-d_{k12}+d_{k22}$, the interaction between alleles of the sire and the dam (dominance effect).

The three genetic effects, $\gamma_{k}=[\mathrm{a}_{k}^{s}\quad\mathrm{a}_{k}^{d}\quad\mathrm{d}_{k}]$, are estimated by $\gamma_{k}=\mathrm{H}^{T}\mathrm{g}_{k}$.

Where

$$\mathrm{H}=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 1 \\ \frac{1}{2} & -\frac{1}{2} & -1 \\ -\frac{1}{2} & \frac{1}{2} & -1 \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}.$$

2.4 Hypothesis tests

The hypothesis about the existence of a QTL affecting an overall process of dynamics can be formulated as

H_0: $\gamma_{k}=0$ H_A: $\gamma_{k}\neq 0$ (k=1,2,3)

That H_0 is rejected means there has a QTL existent in the map.

The test statistics for testing the hypothesis is calculated as the log-likelihood ratio (LR) of the full over reduced model:

$$\mathrm{LR}=-2\log\left[\frac{L(\tilde{\Omega})}{L(\hat{\Omega})}\right].$$

Where $\tilde{\Omega}$ and $\hat{\Omega}$ denote the ML estimates of the unknown parameters under H_0 and H_1, respectively.

3. Simulation studies

The properties of the method, especially unbiaseness, standard errors of the parameter estimation and statistical power, were investigated numerically via Monte Carlo simulations with a full-sib population. On a single chromosome segment of length 100 cM consider and simulate 6, 11 and 21 evenly spaced codominant markers covered,

respectively, and a single QTL was supposed locating at position 27 cM. The test-day of the data observed a range from 1 to 150. Alleles inherited at each position in individuals were generated by the linkage phase inherited from estimation.

Assume the effects of the QTL of dynamic trait were fixed in statistical model and described its different contribution to phenotypic values at different time points by three-orders Legendre polynomial. The first four terms of Legendre polynomial were

$$P_0(\tau)=1, \qquad P_1(\tau)=\tau,$$

$$P_2(\tau)=\frac{1}{2}(3\tau^2-1), \qquad P_3(\tau)=\frac{1}{2}(5\tau^3-3\tau)$$

Let the mean values of regression effect in population be 0, then the phenotypic value of individual *i* at time *j* can be calculated as follow according to QTL genotypes of each individual and test project of this population, given the additive effect (a_0^s for sire and a_0^d for dam) and dominance effects (d_0) of the parents and the residual variance σ_0^2.

$$y_{ij}=\frac{1}{2}(\mathrm{T}'_{ij}\mathrm{a}_0^s+\mathrm{T}'_{ij}\mathrm{a}_0^d)+\frac{1}{4}\mathrm{T}'_{ij}\mathrm{d}_0+\xi\cdot\sigma_0$$

Where, ξ is a random number of standardized normal distribution.

Given effect, a_0 ($\mathrm{a}_0^s=\mathrm{a}_0^d=a_0$) and d_0, the variance and covariance of each regression effect were calculated as $G_{a^s}=G_{a^d}=\frac{1}{4}\mathrm{a}'_0\mathrm{a}_0$, $G_d=\frac{1}{16}\mathrm{d}'_0\mathrm{d}_0$.

Combined with residual variance σ^2, the heredity in each point of the dynamic process can be calculated as

$$h_i^2=\frac{g(i)}{g(i)+\sigma_0^2}$$

Where, $g(t)=T_i(G_a+G_d)T_i'$ with

$$\mathrm{T}'_i=\begin{bmatrix}1 & P_1(t_i) & \cdots & P_s(t_i)\end{bmatrix}'.$$

The phenotypic value, additive effect value, dominance effect value at each dynamic points were calculated under each given genetic regresson effect, $\mathrm{a}'_0=\begin{bmatrix}1.58 & 0.65 & -1.02 & 1.65\end{bmatrix}$ and $\mathrm{d}'_0=\begin{bmatrix}1.77 & 0.36 & -0.71 & 1.27\end{bmatrix}$, and graphically represented in Figure 1a and 1b.

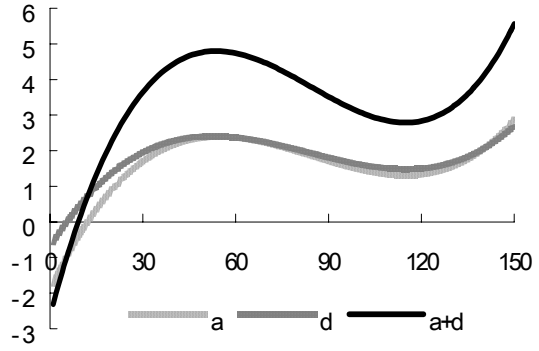

Figure 1a. Dynamic curve of total genetic value, additive and dominance effect value

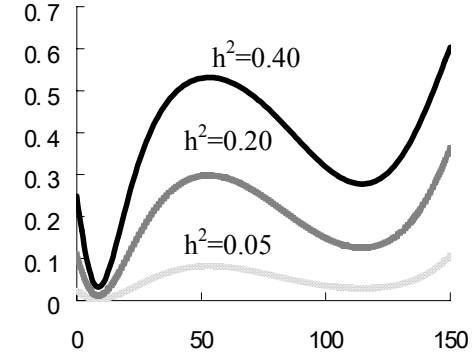

Figure 1b. Heritability change at each dynamic point under three accumulative accumulative heretability levels

The factors considered included individual number, test-day frequency, marker number and heredity level and each taken three levels as 0.05, 0.20 and 0.40 were taken in accumulative heretability; 100, 300 and 500 in individual; 6, 11 and 21 in marker; 5, 10 and 15 in test-day frequency.

Since the combinations of factors considered were too much, the experiments of simulation were arranged according to orthogonal table (Table 1). The simulation was repeated 100 times for each situation. The mean estimate and the standard error of an estimated parameter, calculated from 100 replicated simulations, were showed in Table 2-4.

The statistical power was determined by counting the proportion of the number of runs (over 100 replicates) with test statistic values greater than an empirical threshold. The empirical threshold value was obtained by choosing the 95-th and 99-th percentiles of the highest test statistic over 1000 additional runs under the null model (no QTL is segregating). Table 2 shows that ML method provides slightly higher threshold values than regression analysis, but these critical values hardly depend on each factor.

Table 1. Experimental factor and its combinations

Groups	Factors			
	Heretabilities	Marker frequencies	Individual numbers	Test-day frequencies
1	0.05	6	100	5
2	0.05	11	300	10
3	0.05	21	500	15
4	0.2	6	300	15
5	0.2	11	500	5
6	0.2	21	100	10
7	0.4	6	500	10
8	0.4	11	100	15
9	0.4	21	300	5

Table 2. Empirical critical values of the test statistics in different combinations（95%, 99%）

Groups	Methods	95%	99%
1	ML	21.72	24.24
	REG	21.03	23.50
2	ML	19.68	21.53
	REG	19.54	21.38
3	ML	18.87	20.64
	REG	18.86	20.64
4	ML	18.34	19.98
	REG	18.34	19.98
5	ML	19.53	20.00
	REG	18.41	19.88
6	ML	19.98	21.96
	REG	19.85	21.83
7	ML	17.95	19.71
	REG	17.86	19.58
8	ML	19.57	21.52
	REG	19.59	21.57
9	ML	19.44	21.33
	REG	19.44	21.33

Table 3 showed the estimates (and standard deviation) of QTL position, residual variance and the powers of regression analysis for detection compared with ML in different combinations. Generally speaking, the two methods show virtually no differences, which is consistent with the observation of Xu (1998) and Haley and Knott's (1992) comparison of ML and REG in mapping QTL for general quantitative trait. The methods behave as expected: a high level factor tends to produce an unbiased estimate of the QTL position and small estimation errors. Both methods perform well in detection power, which may reach 98% when individual number over 300 even only the 5% variation of the trait was controlled by the QTL. The power of QTL detection in regression analysis was above 65% when the individual number was just 100. If only analysis single records of a trait, the number of individual needed that ensure detection power be of 95% would be 500 in the same situation. It sufficiently shows the complement of test-day frequency and individual number.

Both methods perform well in estimating the residual variance with little estimation error and high precision under every combination. Although ML behaves better than REG in the same situation, the estimates of the two methods are both tending to the true values as the sample size, which defined as the product of individual number and test-day frequency increasing.

Table 3. Estimates of QTL Position, Residual Variance and the Powers for Detection Calculated from 1000 Repeated Simulations

Groups	Methods	cM	σ^2	Statistical power	
				α =0.05	α =0.01
1	ML	28.735(0.541)	33.106(0.0806)	79	69
	REG	29.188(0.537)	36.949(0.0775)	77	71
2	ML	27.249(0.0784)	34.895(0.0294)	100	100
	REG	27.277(0.0801)	35.047(0.0308)	99	98
3	ML	26.988(0.0314)	35.018(0.0181)	100	100
	REG	26.988(0.0311)	35.017(0.0191)	100	100
	True	27.00	7.379		
4	ML	27.179(0.0633)	7.350(0.00597)	99	99
	REG	27.168(0.0787)	7.283(0.00719)	99	99
5	ML	27.017(0.0383)	7.329(0.00676)	100	100
	REG	27.071(0.0412)	7.336(0.00857)	100	100
6	ML	26.913(0.0576)	7.262(0.0104)	100	100
	REG	26.889(0.0602)	7.298(0.0144)	100	100
	True	27.00	2.767		
7	ML	26.880(0.0415)	2.758(0.00224)	100	100
	REG	26.786(0.0625)	2.717(0.00394)	100	100
8	ML	27.018(0.0601)	2.728(0.00376)	100	100
	REG	27.035(0.0654)	2.635(0.00818)	100	100
9	ML	27.001(0.0242)	2.760(0.00199)	100	100
	REG	27.007(0.0275)	2.7037(0.00628)	100	100

Table 4. Mean Estimates and Standard Deviations (in parentheses) of Effects of QTLCalculated from 1000 Repeated Simulations

Groups	Methods	a_0^s	a_1^s	a_2^s	a_3^s	a_0^d	a_1^d	a_2^d	a_3^d	d_0	d_1	d_2	d_3
	True	1.58	0.65	-1.02	1.65	1.58	0.65	-1.02	1.65	1.77	0.36	-0.71	1.27
1	ML	1.55	0.64	-0.98	1.65	1.61	065	-1.02	1.62	1.68	0.35	-0.65	1.13
		(0.024)	(0.036)	(0.045)	(0.057)	(0.025)	(0.037)	(0.045)	(0.057)	(0.058)	(0.092)	(0.102)	(0.138)
	REG	1.54	0.64	-0.98	1.68	1.58	0.66	-1.00	1.64	1.70	0.39	-0.63	1.07
		(0.023)	(0.037)	(0.045)	(0.057)	(0.023)	(0.037)	(0.043)	(0.057)	(0.057)	(0.086)	(0.104)	(0.137)
2	ML	1.58	0.63	-1.01	1.66	1.58	0.64	-1.01	1.65	1.78	0.32	-0.74	1.34
		(0.008)	(0.013)	(0.015)	(0.018)	(0.008)	(0.013)	(0.016)	(0.018)	(0.016)	(0.027)	(0.033)	(0.040)
	REG	1.58	0.63	-1.01	1.66	1.58	0.64	-1.01	1.65	1.78	0.33	-0.75	1.34
		(0.008)	(0.013)	(0.015)	(0.018)	(0.008)	(0.013)	(0.016)	(0.018)	(0.016)	(0.027)	(0.033)	(0.040)
3	ML	1.59	0.66	-1.02	1.66	1.58	0.65	-1.05	1.66	1.76	0.37	-0.71	1.30
		(0.005)	(0.008)	(0.010)	(0.012)	(0.005)	(0.008)	(0.010)	(0.012)	(0.009)	(0.016)	(0.021)	(0.023)
	REG	1.59	0.65	-0.99	1.65	1.59	0.67	-1.06	1.63	1.77	0.38	-0.65	1.30
		(0.004)	(0.008)	(0.011)	(0.011)	(0.005)	(0.084)	(0.095)	(0.011)	(0.010)	(0.015)	(0.021)	(0.023)
4	ML	1.58	0.64	-1.02	1.63	1.58	0.65	-1.02	1.65	1.76	0.35	-0.70	1.27
		(0.004)	(0.005)	(0.007)	(0.007)	(0.004)	(0.005)	(0.006)	(0.008)	(0.007)	(0.011)	(0.015)	(0.017)
	REG	1.58	0.65	-1.02	1.63	1.59	0.65	-1.02	1.65	1.76	0.35	-0.70	1.28
		(0.004)	(0.005)	(0.007)	(0.008)	(0.005)	(0.006)	(0.007)	(0.008)	(0.010)	(0.012)	(0.016)	(0.019)
5	ML	1.58	0.65	-1.01	1.66	1.58	0.65	-1.03	1.66	1.78	0.38	-0.71	1.30
		(0.004)	(0.006)	(0.008)	(0.010)	(0.004)	(0.006)	(0.008)	(0.009)	(0.008)	(0.013)	(0.016)	(0.020)
	REG	1.58	0.65	-1.02	1.66	1.58	0.65	-1.03	1.66	1.78	0.38	-0.71	1.29
		(0.004)	(0.006)	(0.008)	(0.010)	(0.004)	(0.006)	(0.008)	(0.009)	(0.008)	(0.014)	(0.016)	(0.020)
6	ML	1.57	0.65	-1.03	1.66	1.58	0.64	-1.02	1.64	1.78	0.35	-0.73	1.29
		(0.006)	(0.010)	(0.012)	(0.141)	(0.006)	(0.010)	(0.012)	(0.014)	(0.012)	(0.020)	(0.025)	(0.029)
	REG	1.57	0.65	-1.03	1.66	1.58	0.64	-1.02	1.64	1.78	0.35	-0.73	1.29
		(0.006)	(0.010)	(0.012)	(0.014)	(0.006)	(0.010)	(0.012)	(0.014)	(0.013)	(0.020)	(0.025)	(0.029)
7	ML	1.58	0.65	-1.02	1.64	1.58	0.65	-1.03	1.65	1.77	0.36	-0.71	1.28
		(0.002)	(0.003)	(0.004)	(0.004)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.007)	(0.008)	(0.009)
	REG	1.58	0.65	-1.02	1.65	1.58	0.66	-1.02	1.66	1.79	0.36	-0.71	1.28
		(0.003)	(0.003)	(0.004)	(0.006)	(0.004)	(0.004)	(0.005)	(0.006)	(0.010)	(0.008)	(0.010)	(0.014)
8	ML	1.57	0.65	-1.01	1.63	1.58	0.66	-1.02	1.63	1.77	0.36	-0.72	1.27
		(0.003)	(0.005)	(0.007)	(0.008)	(0.003)	(0.005)	(0.007)	(0.008)	(0.006)	(0.011)	(0.014)	(0.015)
	REG	1.58	0.64	-1.01	1.63	1.58	0.66	-1.03	1.64	1.78	0.36	-0.73	1.29
		(0.004)	(0.005)	(0.007)	(0.008)	(0.004)	(0.005)	(0.007)	(0.008)	(0.007)	(0.011)	(0.014)	(0.016)
9	ML	1.58	0.65	-1.02	1.65	1.58	0.65	-1.03	1.65	1.77	0.37	-0.71	1.29
		(0.002)	(0.003)	(0.004)	(0.004)	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)	(0.007)	(0.008)
	REG	1.57	0.64	-1.02	1.66	1.58	0.65	-1.02	1.64	1.77	0.37	-0.70	1.25
		(0.003)	(0.005)	(0.006)	(0.007)	(0.003)	(0.005)	(0.006)	(0.008)	(0.006)	(0.010)	(0.012)	(0.015)

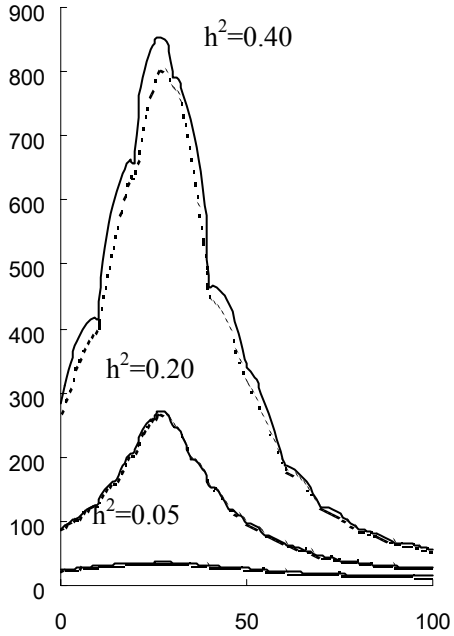

Figure 2a. Comparison of the likelihood ratio test statistic profiles of REG (dotted line) with ML (solid line) under three accumulative levels heritability with 100 individuals

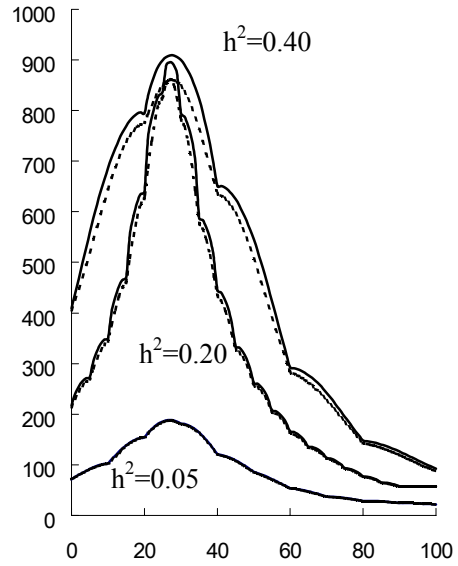

Figure 2b. Comparison of the likelihood ratio test statistic profiles of REG (dotted line) with ML (solid line) at three levels of QTL heritability with with 300 individuals

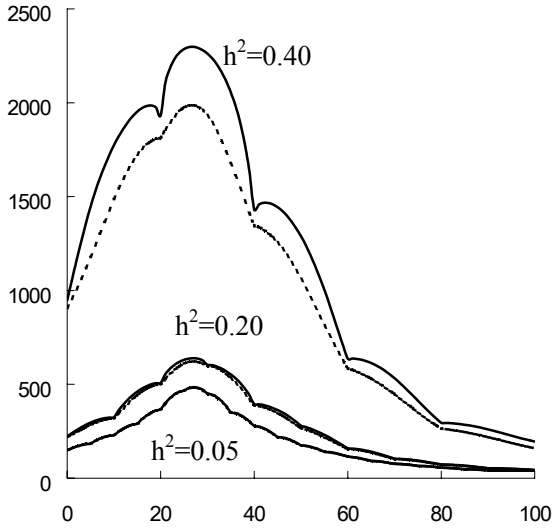

Figure 2c. Comparison of the likelihood ratio test statistic profiles of REG (dotted line) with ML (solid line) at three levels of QTL heritability with 500 individuals

To compare the two methods in QTL detection, the likelihood ratio test statistic profiles were plotted against the chromosome position (Figure 2a-c). It is easily seen that the profiles of the two methods almost overlapped under the heritbility of low level. The test statistic of ML method seems apparently greater than that of REG method. Both methods could hardly detect the existence of QTL under low heritability, but both showed high ability in detection for the trait with high heritability.

4. Discussion and Conclusion

QTL mapping is a typical problem of the regression with uncertain independent variables. It is well known that ML is the optimal method for this problem because the distributions of the unknown independent variables are fully taken into account (Lander and Botstein 1989; Jansen 1994; Zeng 1994). Under normal residual distribution, the REG method can provide a first-order approximation to the ML in the sense that the unobserved independent variables replaced by their conditional expectations (Haley and Knott 1992; Xu 1998). Other than time number of estimated effects increased, there were no differences between the mapping principles for dynamic trait and the general quantitative trait loci by using Legendre polynomial to describe QTL genetic effects

underlying same genetic design. If residual variance of each observation point is supposed to be equal and of normal distribution, the unknown indicator variable for QTL genotypes can be replaced by its expectation and REG method could be used to estimate the parameter of QTL for mapping dynamic traits. The estimation for the residual variance tends to be slightly expanded whereas the unknown QTL genotypes simply replaced by their conditional probabilities given maker information.

The process of simulation analysis indicated that REG method is of absolute advantage in computational algorithm and speed and of certain detection power compared with the ML.

The results of simulation showed that REG performs equally well with ML, but deficient in accuracy and precision of residual variance estimating. The detection power of both methods depends greatly on heredity level, but the power of detection would be sufficient with individuals above 300, makers over 6 and test-day frequency at 5% or more even if in lower heritability.

Correspondence to:

Runqing Yang

School of Agriculture and Biology

Shanghai Jiaotong University

Shanghai 201101, China

E-mail: runqingyang@sjtu.edu.cn

References:

[1] Cheverud JM, Rutledge JJ, Atchley WR. Quantitative genetics of development- genetic correlations among age-specific trait values and the evolution of ontogeny. Evolution 1983;37: 895-905.

[2] CHANGJIAN JIANG, ZHAOBANG ZENG, 1995. Multiple Trait Analysis of Genetic Mapping for Quantitative Trait Loci. Genetics, 140: 1111-27

[3] CHANGXING MA, GEORGE C, RONGLING WU, 2002. Functional mapping of quantitative trait loci underlying the character process: A theoretical framework. Genetics, 61: 1751-62

[4] DEMPSTER A P, LAIRD N M, RUBIN D B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc, 39: 1-38

[5] DEMPSTER, A P, N M LAIRD AND D B RUBIN, 1977 Maximum likelihood from incomplete data via EM algorithm. J. R. Stat. Soc. Ser. B **39:** 1–38

[6] EAVES L J, NEALE M C, MAES H., 1996. Multivariate multipoint linkage analysis of quantitative trait loci. Behav Genet, 26: 519-25

[7] EMEBIRI, L. C., M. E. DEVEY, A. C. MATHESON AND M. U. SLEE, 1998. Age-related changes in the expression of QTLs for growth in radiata pine seedlings. Theor. Appl. Genet. **97:** 1053-61.

[8] HALEY C S, KNOTT S A, 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity, 69: 315-24

[9] HENDERSON C R JR., 1982. Analysis of covariance in the mixed model: Higher level, no homogenous, and random regressions. Biometrics, 38: 623-40

[10] JANSEN R C., 1994. Controlling the Type I and Type Ⅱ Errors in Mapping Quantitative Trait Loci. Genetics, 138: 871-81

[11] KNOTT S A, HALEY C S., 2000. Multitrait least squares for quantitative trait loci detection. Genetics. 156: 899-911

[12] KOROL A B, RONIN Y I, KIRZHNER V M., 1995. Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics, 140: 1137-1147 [9] Wu, W. R., W. M. Li, D. Z. Tang, H. R. Lu and A. J. Worland, 1999. Time-related mapping of quantitative trait loci underlying tiller number in rice. Genetics. **151:** 297–303.

[13] LANDER E S， BOTSTEIN D, 1989. Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. Genetics, 121: 185-99

[14] Nuzhdin SV, Pasyukova EG,. Sex-spesific quantitative trait loci affecting longevity in *Drosophila melanogaster*. Proc Natl Acad Sci USA 1997;94:9734-9.

[15] RODRIGUEZ-ZAS S L., B. R. SOUTHEY, D. W. HEYEN, AND H. A. LEWIN, 2002. Detection of quantitative trait loci influencing dairy traits using a model for longitudinal data. J Dairy Sci, 85: 2681-91

[16] RONIN Y L, KIRZHNER V M, KOROL A B., 1995. Linkage between loci of quantitative traits and marker loci: multitrait analysis with a single marker. Theor. Appl. Genet, 90: 776-86

[17] RONGLING WU, CHANG-XING MA, MYRONCHANG, RAMON C. LITTELL, SAMUEL S.WU, TONGMING YIN, MINREN HUANG, MINGXIU WANG and GEORGECASELLA 2002. A logistic mixture model for characterizing genetic determinants causing diåerentiation in growth trajectories. Genet. Res. Camb. 79: 235-45.

[18] RONGLING WU, CHANG-XING MA, MARK C. K. YANG, MYRON CHANG,RAMON C. LITTELL, UPASANA SANTRA, SAMUEL S. WU, TONGMING YIN,MINREN HUANG, MINXIU WANG AND GEORGE CASELLA, 2003. Quantitative trait loci for growth trajectories in *Populu*. Genet. Res. Camb., 81:51-64

[19] RAO, S. Q. AND S. XU, 1998. Mapping quantitative trait loci for categorical traits in four-way crosses. Heredity 81:214-24.

[20] SCHAEFFER L R, DEKKERS J C M., 1994. Random regressions in animal models for test-day production in dairy cattle. Proc. 5thWCGALP, Canada, 18: 443

[21] SCHAEFFER L R., 2002. Application of random regression model in animal breeding. J Dairy Sci., 83: Suppl 1

[22] VERHAEGEN, D., C. PLOMION, J. M. GION, M. POITEL, P. COSTA *ET AL.*, 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.

[23] XU, S. AND D. D. G. GESSLER, 1998. Multipoint genetic mapping of quantitative trait loci using a variable number of sibs per family. Genetical Research, 71:73-83.

[24] WU R L, MA C X, ZHU J & GEORGE C., 2002. mapping epigenetic quantitative trait loci (QTL) altering a developmental trajectory. Genome, 45:28-33

[25] WU W R, ZHOU Y, LI W, MAO D, CHEN Q., 2002. Mapping of quantitative trait loci based on growth models. Theoretical and Applied Genetics, 105: 1043--9

[26] XU S., 1998. Further investigation on the regression method of mapping quantitative trait loci. Heredity, 80: 364-73

[27] YAN J Q, ZHU J, HE C X, 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–-74

[28] YAN J Q, ZHU J, HE C X, BENMOUSSA M, WU P. 1998b Molecular diss.ection of developmental behavior of plant height in rice(*Oryza sativa* L.). Genetics, 150:1257–-65

[29] ZENG Z B., 1994. Precision mapping of quantitative trait loci. Genetics, 136: 1457-68