

Mapping quantitative trait Loci for ordinal traits using the generalized linear model in half-sib designs

Zong-Jun YIN^{a,b*}, Qin ZHANG^{a**}

^a Key Laboratory of Animal Genetics and Breeding of the Ministry of Agriculture, State Key Laboratory for Agrobiotechnology, College of Animal Science and Technology, China Agricultural University, Beijing 100094, P.R. China

^b College of Animal Science and Technology, Anhui Agricultural University, Hefei 230036, P.R. China

(Received 16 December 2004 – Accepted 12 December 2005)

Abstract – Most QTL mapping methods share a common assumption: that the phenotype follows a normal distribution. Many phenotypes of interest, however, do not satisfy this assumption. In this paper, a methodology of QTL mapping for ordinal traits based on the framework of the generalized linear model (GLM) is presented. The location and effect of the putative QTL were estimated using the maximum likelihood method. The efficiency and the power of the proposed method were compared with that of the method based on the linear model (LM) in various conditions (QTL effect, heritability, phenotypic incidence, and number of categories of phenotypes) via simulation. A daughter design with multiple families and a total of 500 individuals was applied. The results showed that the GLM approach had certain advantages over the LM approach in power of QTL detection and QTL position estimation for ordinal traits. The estimates of the QTL position were 0.11~1.59 cM (0.78 on average) less biased with smaller standard errors. The power of QTL detection was 1.6 ~ 10.9% (5.1% on average) higher. In addition, the power and the accuracy of QTL mapping depended on the effect of the putative quantitative trait loci and the value of heritability. With the increase of the QTL effect from 0.05 to 0.3, the biases of the QTL position estimates reduced 0.4 to 3.6 cM and the power increased 27 to 56% under different heritabilities. With the increase of heritability from 0.1 to 0.4, the biases reduced 0.24 to 3.1 cM and the power increased 5% to 35% under a different QTL effect.

ordinal traits / threshold models / generalized linear model / QTL mapping / maximum likelihood

Résumé – Cartographie de loci à effets quantitatifs (QTL) pour des caractères ordinaux en utilisant le modèle linéaire généralisé dans des familles de demi-germains. La plupart des méthodes de cartographie des QTL ont en commun un même principe : le phénotype suit une distribution normale. Or, de nombreux phénotypes d'intérêt ne satisfont pas à cette condition. Dans cet article, la méthodologie de cartographie des QTL pour des caractères ordinaux basée sur le modèle

* Corresponding author: yinzongjun@yahoo.com.cn

** qzhang@cau.edu.cn

linéaire généralisé (GLM) est présentée. La position et l'effet du QTL supposé sont estimés à l'aide de la méthode du maximum de vraisemblance. L'efficacité et la puissance des tests proposés sont comparées par simulation à ceux fondés sur le modèle linéaire (LM) dans des conditions variables (effet des QTL, héritabilité, incidence phénotypique et nombre de catégories phénotypiques). Une population expérimentale selon un schéma familles de filles de mêmes pères et un total de 500 individus a été utilisée. Les résultats montrent que GLM a certains avantages sur LM pour la puissance de détection des QTL et l'estimation de leur position en ce qui concerne les caractères ordinaux. L'estimation de la position du QTL était 0,11~1,59 cM (0,78 en moyenne) moins biaisée, avec des écart-types moindres et la puissance de détection était 1,6~10,9 % (5,1 % en moyenne) donc plus grande. La puissance et la précision de la cartographie des QTL dépendent de la variance des QTL supposés et de la valeur de l'héritabilité. Lorsque la proportion de variance due au QTL augmente de 0,05 à 0,3, les biais d'estimation des positions de QTL sont diminués de 0,4 à 3,6 cM en fonction des situations testées et la puissance de détection est augmentée de 27 à 56 % quelle que soit l'héritabilité. Lorsque l'héritabilité augmente de 0,1 à 0,4, les biais sont réduits de 0,24 à 3,1 cM suivant les simulations et la puissance est augmentée de 5 à 35 % quelle que soit la variance des QTL.

caractères ordinaux / modèle à seuils / modèle linéaire généralisé / cartographie des QTL / maximum de vraisemblance

1. INTRODUCTION

Many economically important traits in livestock vary in a discontinuous manner but are not inherited in a simple Mendelian fashion. For example, disease resistance traits in animals may be recorded in binary or ordered categories. These traits, normally called categorical or threshold traits [5], are presumably controlled by many genes with expressions modified by environmental effects like quantitative traits. In human genetics, such traits are termed as complex discrete traits [20]. Owing to the recent advances in marker technology, the molecular dissection of heritable quantitative traits into their individual Mendelian components or quantitative trait loci (QTL) is now conceivable. Novel statistical techniques to detect and map individual genes affecting quantitative traits have been developed and subsequently refined. Most methods of QTL mapping are based on the interval mapping approach, using either least squares regression or maximum likelihood [6, 8, 13, 15, 26, 30]. The QTL concept is also used for traits with discrete distribution for which a continuous underlying process can be assumed [25]. However, most QTL mapping methods share a common assumption that the phenotypes

follow a normal distribution and hence are not appropriate for categorical traits. Thus, new methods are needed for QTL mapping for categorical traits. However, mapping QTL for categorical traits is more challenging than for continuously distributed traits due to the nonlinear relationship between the observed phenotype and unobservable genetic effects, especially when the mapping population contains multiple outbred families [3, 14, 26]. The threshold model or generalized linear model (GLM) would be the key techniques for QTL mapping for categorical traits.

Methods for QTL mapping for binary traits, which are special categorical traits with just two categories, have been developed in line-crossing experiments. Using the threshold model, Hackett and Weller [9] and Xu and Atchley [27] developed methods to map QTL controlling the underlying liability of binary traits. These methods were based on a logit probability model and illustrated using a back-cross population as an example and can easily be extended for ordinal traits, since the threshold model is also valid for this type of traits. Rao and Xu [19] adapted the logit probability model used for the binary trait to the analysis of ordinal traits in four-way crosses. In the GLM framework,

Spyrides-Cunha et al. [21] developed a proportional odds model for testing the presence of QTL at single-markers. Here, we present the methodology of QTL mapping for binary and ordinal traits in the GLM framework for half-sib design. The location and effect of the putative QTL were estimated using the maximum likelihood method. The efficiency and power of the proposed GLM approach were compared with those of the simple linear model (LM) where discrete data are treated as if they were continuous.

2. METHODS

Under the assumption that ordinal traits are controlled by polygenes, the discrete phenotype (Y) is assumed to be controlled by a latent variable, referred to as liability (z), which is considered to be continuous and normally distributed and can be described by the usual linear model

$$z = x'\beta + G + g + e$$

where z is the liability, $x'\beta$ represents all fixed effects, G is the QTL genotypic value, g is the polygenic effect with a distribution of $N(0, \sigma_g^2)$, and e is the residual with a distribution of $N(0, \sigma_e^2)$. The discrete phenotypes are determined by a set of underlying thresholds ($v_1 < v_2 < \dots < v_{c-1}$), which define the observed categories, Y , on an ordinal scale 1, 2, ..., c , i.e.

$$v_{k-1} < Z < v_k \Leftrightarrow Y = k, \quad k = 1, 2, \dots, c, \\ v_0 = -\infty, v_c = +\infty.$$

For binary traits, the observed two categories are determined by one fixed threshold (v). When the value of the liability is above the threshold value, an individual shows one status of phenotype on the observed scale, otherwise, it shows the other status of the phenotype.

The maximum likelihood method for the half-sib design developed by Weller

et al. [24] is a convenient method for QTL mapping for continuously distributed traits. Empirical studies in animal breeding suggest that the linear model is robust for departures from normality in the line cross population [13]. It may be interesting to show the comparison between the GLM and the traditional linear method (LM) in a half-sib population. For half-sib design, assuming a single QTL with alleles Q and q coming from heterozygous sires and a random dam population, the distribution of the liability of the offspring is a mixture of three distributions. The probability density function of the liability of individual j can be written as

$$f(z_j) = \sum_{i=1}^3 fre_{j(i)} \cdot f_{(i)}(z_j)$$

where $fre_{j(i)} = \Pr(G = i|M)$ represents the probability of individual j having QTL genotype i ($i = QQ, Qq, qq$) conditional on marker genotype information, which can be calculated from the known recombination rates and marker genotypes [16], $f_{(i)}(z_j)$ represents the normal distribution density of z_j given QTL genotype i , with the same variance and different mean ($a, 0, -a$) for different QTL genotype, where a is the additive effect of the putative QTL. However, because the liability (z) was not observable, we have to construct the likelihood function from the discrete phenotypic value.

Consider N progeny in a half-sib design and let Y_j and G_j be the phenotypic and genotypic values of progeny j , respectively. Y_j can be transformed into a multivariate vector, $W_j = [w_{j1} \ w_{j2} \ \dots \ w_{jc}]^T$, where

$$w_{jk} = \begin{cases} 1 & \text{if } Y_j = k \\ 0 & \text{if } Y_j \neq k \end{cases} \quad k = 1, 2, \dots, c.$$

Conditional on the genotypic value of progeny j , the transformed multivariate phenotype W_j follows a multinomial distribution with $\Pr(w_{jk} = 1|G_j) = P_{jk}$.

Therefore, the likelihood of observing W_j is $f_j(G_j) = \prod_{k=1}^c P_{jk}^{w_{jk}}$. Since G_j were not observable, all possible values of G_j must be considered and the expected likelihood takes the average $f_j(G_j)$ weighted by the QTL genotype probabilities. So,

$$\begin{aligned} f_j &= \sum_{i=1}^3 pr(G_j = Q_i) f_j(G_j = Q_i) \\ &= \sum_{i=1}^3 pr(G_j = Q_i) \prod_{k=1}^c P_{jk}^{w_{jk}} \end{aligned}$$

where Q_i represents the genotypic value of QTL genotype i , with $Q_1 = a$, $Q_2 = 0$, and $Q_3 = -a$. We now examine the link between the genotypic value and the phenotype. Under the assumption of a normal distribution for e , P_{jk} is defined by

$$\begin{aligned} P_{jk} &= \Phi(v_k - x'\beta - g_j - G_j) \\ &\quad - \Phi(v_{k-1} - x'\beta - g_j - G_j) \end{aligned}$$

where $\Phi(\xi)$ stands for the cumulative standard normal distribution function with argument ξ . Analysis involving $\Phi(\xi)$ is referred to as probit analysis. A logistic model is employed to approximate $\Phi(\xi)$ for estimation purpose [1]. The logistic model is expressed by $\psi(x) = \frac{\exp(x)}{1+\exp(x)}$. The approximate relationship between a probit model and a logistic model is $\Phi(x) \approx \psi(cx)$, where $c = \pi/\sqrt{3}$ [17]. Therefore,

$$\begin{aligned} P_{jk} &\approx \frac{\exp\{c[v_k - (x'\beta + g_j + G_j)]\}}{1 + \exp\{c[v_k - (x'\beta + g_j + G_j)]\}} \\ &\quad - \frac{\exp\{c[v_{k-1} - (x'\beta + g_j + G_j)]\}}{1 + \exp\{c[v_{k-1} - (x'\beta + g_j + G_j)]\}}. \end{aligned}$$

In QTL analysis, the probability $pr(G_j = Q_i)$ can be inferred through the conditional genotype probability $fre_{j(i)} = \Pr(G_j = i|M)$. So, the likelihood function for

progeny j is now rewritten as

$$\begin{aligned} f_j &= \sum_{i=1}^3 pr(G_j = Q_i) f_j(G_j = Q_i) \\ &= \sum_{i=1}^3 pr(G = i|M) \prod_{k=1}^c P_{jk}^{w_{jk}} \end{aligned}$$

and the overall likelihood function for N progeny is $L = \prod_{j=1}^N f_j$. The parameters σ_g^2 , σ_e^2 and a can be estimated via maximum likelihood estimation (MLE). The test of H_0 vs. H_1 can be carried out by using the likelihood ratio (LR) statistic. The distribution of LR can be approximated to a chi-square distribution with two degrees of freedom [22, 27] or empirically derived via simulation under the null hypothesis [19].

3. SIMULATION STUDY

A simulation study was carried out to investigate the efficiency and behavior of the method proposed above.

3.1. Experimental design

A half-sib design was used for linkage analysis of putative QTL and genetic markers. There were 10 sires, each mating with 50 dams and producing 50 half-sib progeny. So, the total number of progeny was 500. All sires and dams were unrelated with each other and matings between sires and dams were random. The traits were determined by many genes of small effect each and by a relatively major QTL with two alleles, Q and q , with frequencies of p and $1-p$, respectively. The putative QTL was situated between two flanking biallelic markers, with distances d_1 and d_2 to the left and right marker, respectively. The distance between the two markers was D and was assumed known. The Haldane mapping function was assumed when transforming the distances to

recombination rates. The linkage phase between QTL and markers was assumed to be known or estimated with a high degree of accuracy.

3.2. Data simulation

The flanking markers were 45.8 cM apart with a QTL located between them with distances 17.8 and 28 cM to the left and right marker, respectively. Paternal marker-QTL haplotypes and maternal marker haplotypes received by progeny were sampled according to the population frequencies of 0.5 for all alleles and the linkage between QTL and marker loci.

Two kinds of ordinal traits with two and four ordered categories, respectively, were simulated. For both kinds of traits, two sets of incidences were considered, (0.20, 0.60, 0.15, 0.05) and (0.20, 0.40, 0.40, 0.20) for the four-category trait, and (0.80, 0.20) and (0.50, 0.50) for the two-category trait. The QTL effect was expressed as the ratio of QTL variance (σ_G^2) to the total genetic variance ($\sigma_{GT}^2 = \sigma_G^2 + \sigma_g^2$), i.e. $\Delta_{QTL} = \sigma_G^2 / \sigma_{GT}^2$. The genotypic values of the three QTL genotypes were defined as a , 0, and $-a$, respectively. For the sires and dams, the QTL genotypic values were determined according to Falconer's model [5] and the polygenic effects were generated via sampling from $N(0, \sigma_g^2)$ for given heritability (h^2), Δ_{QTL} , and σ_e^2 . For the offspring, the QTL genotypic values were generated according to the haplotypes produced by their parents. The polygenic effects were generated as $u = 0.5g_s + 0.5g_d + m$, where g_s and g_d represent the polygenic value of the sire and dam, respectively, and m represents the Mendelian sampling effect, which follows a normal distribution with mean 0 and variance $\sigma_m^2 = 0.25\sigma_g^2$. The liability values were generated as the sum of the polygenic effect, QTL genotypic value, and random residual effect which was generated from $N(0, \sigma_e^2)$. The liability values

were transformed to discrete phenotypes in accordance with the fixed thresholds, which were determined according to the given incidences.

In the simulation, three levels of QTL effect ($\Delta_{QTL} = 0.05, 0.10, \text{ and } 0.30$) and two levels of heritability of the liability ($h^2 = 0.1 \text{ and } 0.4$) were considered to examine their influence on the accuracy and power of QTL mapping. For each parameter combination 1 000 replicates were simulated. For each replicate, the presence of a QTL was tested and the parameters (position and QTL genotypic value) were estimated using the method presented in Section 2, and for comparison, the linear model method described by Weller et al. [24] as well.

3.3. Hypothesis testing and power of the test

The critical values for testing the presence of a QTL within the marker interval were obtained by generating an empirical distribution of the test statistic (LR) by simulating 2 000 replicates under the null hypothesis of no QTL being present. The power of the test was calculated as the percentage of the replicates for which the null hypothesis was rejected among the 1 000 replicates simulated under the alternative hypothesis (QTL is present).

4. RESULTS

4.1. Estimation of QTL position

The means and standard errors of the estimated QTL position using the two methods are given in Table I. The results showed that in all cases the means of the estimated QTL position obtained by the GLM approach were closer to the true value with smaller standard errors than that obtained by the linear model approach. But the differences were generally small. Compared

Table I. Mean estimates of QTL position and their standard errors (in parentheses) over 1 000 replicates (the true position is 17.8 cM from the left marker).

Type of trait	Incidences	Heritability	Contribution of QTL variance	Mean estimates (SE)	
				GLM	LM
Binary	0.20, 0.80	0.10	0.05	22.31(7.42)	23.90(7.83)
			0.10	20.57(6.87)	21.25(7.26)
			0.30	16.34(6.36)	16.12(6.75)
		0.40	0.05	19.23(6.48)	20.02(6.86)
			0.10	16.47(4.65)	19.73(5.27)
			0.30	18.82(4.46)	19.36(4.74)
	0.50, 0.50	0.10	0.05	22.16(7.41)	22.63(7.67)
			0.10	21.21(5.88)	21.32(6.54)
			0.30	18.46(5.35)	19.38(6.13)
		0.40	0.05	19.44(5.34)	21.38(6.10)
			0.10	18.46(4.15)	19.15(5.28)
			0.30	18.22(4.21)	18.61(4.73)
Four categories	0.20, 0.60, 0.15, 0.05	0.10	0.05	20.77(6.34)	22.31(7.12)
			0.10	19.28(4.56)	20.02(6.23)
			0.30	18.55(5.27)	19.28(5.61)
		0.40	0.05	17.19(3.52)	19.76(5.72)
			0.10	17.12(3.15)	19.23(4.05)
			0.30	18.19(2.76)	18.34(3.43)
	0.20, 0.30, 0.30, 0.20	0.10	0.05	21.36(7.25)	21.72(6.89)
			0.10	19.43(5.45)	21.28(5.72)
			0.30	18.47(4.57)	19.36(5.55)
		0.40	0.05	18.67(4.78)	19.23(5.75)
			0.10	18.55(3.53)	18.96(3.79)
			0.30	17.46(3.21)	18.58(3.64)

with the results for the four-category trait, the estimates of QTL position for the binary trait had larger deviations from the true position and the standard errors of the estimates were also larger for binary categories. In general, the accuracy of QTL position estimates increased with the increase of the heritability and QTL effect.

4.2. Estimation of QTL genotypic value

The means and standard errors of the estimated QTL genotypic value (a) using the

two methods are given in Table II. The differences of the means of the estimates from both methods were generally small (from 0.0009 to 0.0319) although the estimates from the LM approach were slightly accurate in most cases, but the standard errors from the GLM approach were smaller than those from the LM approach in all cases. The differences were getting smaller when the incidences of different categories were getting balanced. In general, for both methods, the QTL effects were significantly overestimated when the QTL effects were

Table II. Mean estimates of QTL genotypic value and their standard errors (in parentheses) over 1 000 replicates.

Type of trait	Incidences	Heritability	Contribution of QTL variance	True genotypic value	Mean estimates(SE)	
					GLM	LM
Binary	0.20, 0.80	0.10	0.05	0.1054	0.1233(0.0876)	0.1349(0.1035)
			0.10	0.1491	0.1146(0.0923)	0.1275(0.1127)
			0.30	0.2582	0.1926(0.1245)	0.2147(0.1334)
		0.40	0.05	0.2582	0.2025(0.1343)	0.2163(0.1526)
			0.10	0.3651	0.3125(0.1212)	0.3236(0.1324)
			0.30	0.6325	0.5643(0.1426)	0.5762(0.1558)
	0.50, 0.50	0.10	0.05	0.1054	0.1321(0.0756)	0.1367(0.0847)
			0.10	0.1491	0.1187(0.0734)	0.1223(0.1032)
			0.30	0.2582	0.2178(0.0741)	0.2169(0.0962)
		0.40	0.05	0.2582	0.2326(0.1121)	0.2247(0.1068)
			0.10	0.3651	0.3319(0.1207)	0.3364(0.1326)
			0.30	0.6325	0.5856(0.1133)	0.5813(0.1365)
Four categories	0.20, 0.60, 0.15, 0.05	0.10	0.05	0.1054	0.2245(0.1077)	0.1926(0.0923)
			0.10	0.1491	0.2336(0.0829)	0.2447(0.1130)
			0.30	0.2582	0.3108(0.0932)	0.2879(0.1025)
		0.40	0.05	0.2582	0.3163(0.1346)	0.2904(0.0780)
			0.10	0.3651	0.4126(0.1055)	0.3975(0.1246)
			0.30	0.6325	0.5882(0.1221)	0.5746(0.1416)
	0.20, 0.30, 0.30, 0.20	0.10	0.05	0.1054	0.2155(0.1142)	0.2038(0.1038)
			0.10	0.1491	0.2167(0.1081)	0.2132(0.1231)
			0.30	0.2582	0.2935(0.0827)	0.2828(0.0964)
		0.40	0.05	0.2582	0.2963(0.1133)	0.2831(0.1049)
			0.10	0.3651	0.3997(0.1174)	0.3825(0.1323)
			0.30	0.6325	0.6022(0.1336)	0.5811(0.1358)

small and underestimated when the QTL effects were large. The biases for the four-category trait were generally larger than those for the binary trait. Very interestingly, for the binary trait, the bias increased with the increase of the QTL effect and heritability, while for the four-category trait, it decreased in the same situations. For all different QTL effects and heritabilities, the biases were smaller when the incidences of different categories were balanced.

4.3. Power of QTL detection

The empirical statistical powers of the QTL detection at 0.05 significance levels are given in Table III for the two methods. The GLM method, in general, had a higher power than the LM method, especially in the cases of binary trait, low heritability, and small QTL effect, where the power of the GLM method were about 10% higher than that of the LM method.

Table III. Power of QTL detection using the GLM and LM methods based on empirical thresholds obtained from 2 000 simulations at 5% significance level.

Type of trait	Incidences	Heritability	Contribution of QTL variance	Power (%)	
				GLM	LM
Binary	0.20, 0.80	0.10	0.05	28.4	17.5
			0.10	35.9	32.7
			0.30	55.3	50.5
		0.40	0.05	35.2	29.6
			0.10	50.3	48.7
			0.30	90.8	82.4
	0.50, 0.50	0.10	0.05	28.7	18.1
			0.10	37.3	32.6
			0.30	56.2	51.1
		0.40	0.05	35.4	29.5
			0.10	52.0	49.6
			0.30	91.8	83.2
Four categories	0.20, 0.60, 0.15, 0.05	0.10	0.05	36.5	30.2
			0.10	49.7	46.6
			0.30	83.8	81.3
		0.40	0.05	43.5	39.4
			0.10	59.4	55.0
			0.30	92.2	87.5
	0.20, 0.30, 0.30, 0.20	0.10	0.05	35.5	31.3
			0.10	51.9	46.6
			0.30	85.2	81.8
		0.40	0.05	40.8	36.4
			0.10	61.5	57.1
			0.30	92.8	89.3

When plotting the averages of the test statistics over 1000 replicates against the map positions, it showed that all curves were peaking near the true QTL position (17.8 cM) (figure not shown). In general, the power increased with the increase of the QTL effect and the heritability. For the same QTL effect and heritability, the power for the four-category trait was higher than that for the binary trait. However, the trait incidence showed very little effect on the power of QTL detection in all cases.

5. DISCUSSION

In this paper, we developed a novel technique in the GLM framework based on the principles of the threshold model [19, 27] under a generalized linear model (GLM) to map QTL for a binary and ordinal trait in the half-sib design. The developed method was tested and compared with the usual linear model method through a set of simulations. The results showed that the threshold model has certain advantages in QTL

location estimation (Tab. I) and power of QTL detection over the linear model method (Tab. III). The estimates of QTL position were 0.11~1.59 cM (0.78 on average) less biased with smaller standard errors and the power of QTL detection was 1.6~10.9% (5.1% on average) higher. The greatest advantage was observed for the binary trait with a low heritability and small QTL effect.

Statistical methods of QTL mapping in line crosses and outbred populations have been developed for binary characters [4, 9, 11, 27]. In a GLM framework, Visscher et al. [23] suggested a logistic regression approach for binary traits using either a probit or a logit link function, and found that both link functions gave similar results. When compared to the classical linear regression (RIM) applied on binary data, the logistic regression gave similar power of QTL detection. Yi and Xu [29] developed a fixed-model approach to map QTL for complex binary traits for multiple full-sib families in outbred populations, and showed that their method was efficient when there were a small number of large families. As the number of families increases, the fixed model approach becomes inefficient because of the large number of parameters to be estimated. Yi and Xu [28] therefore proposed a random model approach in which the effect of each allele was treated as a random variable so that a single variance rather than individual allelic effects were to be estimated and tested. Kadarmideen et al. [12] set up a (likelihood-based) generalized interval mapping method to map QTL for complex binary traits in multi-family half-sib designs based on the threshold theory and implemented using a Newton-Raphson algorithm for fitting the likelihood. They showed that the least-squares-based method and their generalized interval mapping method had similar power of QTL detection. Methods for binary traits can easily be extended for ordi-

nal traits, since the threshold model is also valid for this type of traits. Rao and Xu [19] adapted the logit probability model used for binary traits to the analysis of ordinal traits in four-way crosses. In a GLM framework, Spyrides-Cunha et al. [21] developed a proportional odds model for testing the presence of QTL at single-markers.

Among all the studies on mapping QTL for discrete traits, only a few studies compared the proposed method to standard linear model interval mapping in terms of power of QTL detection [12] in the outbred population. Mixed threshold models would be more beneficial than mixed linear models for the estimation of random effects for categorical traits because genetic variances are linked to the incidence of the trait for linear models [7]. Such mixed model situations arise in QTL mapping for binary or other non-normal traits when implemented via complex pedigree analysis or random models involving estimation of QTL, polygenic and residual variance components [28]. In fact, Yi and Xu [28] reported that generalized interval mapping has more power than RIM to detect QTL for binary traits, which strengthens the need for threshold models in mapping QTL for categorical traits. The advantages of the threshold model over the linear model could increase as complexity increases. One such example is from Quaas et al. [18], who showed that interactions found for calving ease data on the observed binary scale were much smaller than on the liability scale. This again supports the use of threshold models for the analysis of categorical data.

Although the linear model method (LM) is statistically inappropriate for analysis of discrete data, it is simpler to use and less computer intensive than the threshold-model method (GLM). Given this property of LM, we investigated the equivalence of the GLM and LM methods for QTL mapping for discrete data. The results showed that the two methods were

in general similar in terms of estimation of QTL location and QTL effect. The observed similarity is as expected based on reports from several studies which indicated similarity between linear and threshold models for QTL interval mapping with binary or ordinal traits [9, 11, 12, 19, 27]. In this study, however, significant differences were observed between LM and GLM in power for traits with lower heritability and smaller QTL effects. In addition, the advantage of the GLM method was also related to the number of categories of the trait phenotype. For the binary trait, the GLM method showed larger advantage than for the four-category trait.

In this study we only considered the half-sib population, but for the general population, the GLM method can also be applied. The main differences exist in inferring the linkage phase and the conditional probability of the putative QTL genotype. In the general population, it is complex to infer the conditional probability of the QTL genotype using the identity-by-descent (IBD) of the QTL and all marker loci [2, 14]. Here, we only demonstrated the GLM method of QTL mapping in a half-sib design using a single QTL model. In reality, multiple QTL may be present. Zeng [31] and Heather [10] explored a solution for such a problem, which is referred to as composite interval mapping. For the situation of existing multiple QTL, we can add other QTL effects in the linear model of the liability to fit this situation.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers for their constructive comments. This study was funded by the National Natural Science Foundation of China (Grant No. 30430500), the National Key Basic Research Program of China, the Natural Science Fund of Anhui Province (Grant No. 050410204), and the Education Committee of Anhui Province, China (Grant No. 2004kj151 and 2002jq126).

REFERENCES

- [1] Bonney G.E., Regressive logistic models for familial disease and other binary traits, *Biometrics* 42 (1986) 611–625.
- [2] Charlier C., Farnir F., Berzi P., Vanmanshoven P., Brouwers B., Vromans H., Georges M., Identify-by-descent mapping of recessive traits in livestock: application to map the bovine syndactyly locus to chromosome 15, *Genome Res.* 6 (1996) 580–589.
- [3] Chen W.M., Broman K.W., Liang K.Y., Quantitative trait linkage analysis by generalized estimating equations: Unification of variance components and Haseman-Elston Regression, *Genet. Epidemiol.* 26 (2004) 265–272.
- [4] Ding X.D., Zhang Q., Xu R.H., Wang Y.C., Pedigree transmission disequilibrium test for QTL mapping of threshold trait, *Chinese Sci. Bull.* 49 (2004) 1347–1353.
- [5] Falconer D.S., Mackay T.F.C., *Introduction to Quantitative Genetics*, 4th ed., Longman, England, 1996.
- [6] Feingold E., Regression-based quantitative trait locus mapping in the 21st century, *Am. J. Hum. Genet.* 71 (2002) 217–222.
- [7] Gianola D., Theory and analysis of threshold characters, *J. Anim. Sci.* 56 (1982) 1079–1096.
- [8] Grignola F.E., Hoeschele I., Tier B., Mapping quantitative trait loci in out-cross population via Residual Maximum Likelihood. I. Methodology, *Genet. Sel. Evol.* 28 (1996) 479–490.
- [9] Hackett C.A., Weller J.I., Genetic mapping of quantitative trait loci for traits with ordinal distributions, *Biometrics* 51 (1995) 1252–1263.
- [10] Heather J., Cordell J.A., Todd N.J., Statistical modeling of interlocus interactions in a complex disease, *Genetics* 158 (2001) 357–367.
- [11] Kadarmideen H.N., Dekkers J.C.M., Generalized marker regression and interval QTL mapping methods for binary traits in half-sib family designs, *J. Anim. Breed. Genet.* 118 (2001) 297–309.
- [12] Kadarmideen H.N., Janss L.L.G., Dekkers J.C.M., Power of quantitative trait locus mapping for polygenic binary traits using generalized and regression interval mapping in multi-family half-sib designs, *Genet. Res.* 76 (2000) 305–317.

- [13] Knott S.A., Haley C.S., Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses, *Genet. Res.* 60 (1992) 139–151.
- [14] Kruglyak L., Lander E.S., Complete multi-point sib-pair analysis of qualitative and quantitative traits, *Am. J. Hum. Genet.* 57 (1995) 439–454.
- [15] Liu Y, Jansen G.B., Lin C.Y., Quantitative trait loci mapping for dairy cattle production traits using a maximum likelihood method, *J. Dairy Sci.* 87 (2004) 491–500.
- [16] Lander E.S., Botstein D., Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics* 121 (1989) 185–199.
- [17] McCullagh P., Nelder J.A., *Generalized linear models*, 2nd ed., Chapman-Hall, London, 1989.
- [18] Quaas R.L., Zhao Y., Pollack E.J., Describing interactions in dystocia scores with a threshold model, *J. Anim. Sci.* 66 (1988) 396–399.
- [19] Rao S.Q., Xu S.Z., Mapping quantitative trait loci for ordered categorical traits in four-way crosses, *Heredity* 81 (1998) 214–224.
- [20] Risch N., Merikangas K., The future of genetic studies of complex human diseases, *Science* 273 (1996) 1516–1517.
- [21] Spyrides-Cunha M.H., Demetrio C.G.B., Camargo L.E.A., Proportional odds model applied to mapping of disease resistance genes in plants, *Genet. Mol. Biol.* 23 (2000) 223–227.
- [22] Thaller G., Dempfle L., Hoeschele I., Maximum likelihood analysis of rare binary traits under different modes of inheritance, *Genetics* 143 (1996) 1819–1829.
- [23] Visscher P.M., Haley C.S., Knott S.A., Mapping QTLs for binary traits in backcross and F2 populations, *Genet. Res.* 68 (1996) 55–63.
- [24] Weller J.I., Kashi Y., Soller M., Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle, *J. Dairy Sci.* 73 (1990) 2525–2537.
- [25] Wright S., An analysis of variability in number of digits in an inbred strain of guinea pigs, *Genetics* 19 (1934) 506–536.
- [26] Xu S., Further investigation of the regression method for mapping quantitative trait loci, *Heredity* 80 (1998) 364–373.
- [27] Xu S., Atchley W.R., Mapping quantitative trait loci for complex binary diseases using line crosses, *Genetics* 143 (1996) 1417–1424.
- [28] Yi N.J., Xu S.Z., Mapping quantitative trait loci for complex binary traits in outbred population, *Heredity* 82 (1999) 668–676.
- [29] Yi N., Xu S., A random model approach to mapping quantitative trait loci for complex binary traits in outbred populations, *Genetics* 153 (1999) 1029–1040.
- [30] Zhang Q., Boichard D., Hoeschele I., et al., Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree, *Genetics* 149 (1998) 1959–1973.
- [31] Zeng Z.B., Precision mapping of quantitative trait loci, *Genetics* 136 (1994) 1457–1468.

