THE EFFECTS OF POPULATION SIZE AND DOMINANCE OF QUATITATIVE TRAIT LOCI (QTL) ON THE DETECTION OF LINKAGE BETWEEN MARKERS AND QTL FOR LIVESTOCK

G. J. Jeon

Department of Breeding and Genetics, National Livestock Research Institute, R.D.A., Suwon City, Gyeonggido, 440-351, Korea

Summary

A simulation study on detection of linkage between genetic markers and QTL in backcross design was conducted. The effects of various sample sizes and the degree of QTL dominance on detection of linkage were examined by using a simple regression analysis. The results indicated that as sample size increased, the standard error of the estimated slope became smaller. When the dominance effect of QTL was complete, the estimated slope tended to be negative but was statistically not significant at all with type I error of greater than 50%. With complete linkage between genetic Marker and QTL, the estimated intercept value was smallest but the estimated slope was largest as expected. In most cases with various degree of dominance and sample sizes, when the actual recombination rate became larger, greater values were obtained for the slope except in the case of complete dominance of QTL.

(**Key Words** : Linkage Between QTL and Marker, Sample Size, Dominance Effect)

Introduction

The theoretical methodologies for detection of quautitative trait loci (QTL) linked to the genetic markers have been numerously published (Weller, 1986; Lander and Botstein, 1989; Martinez and Curnow, 1992; Knapp et al., 1990; Simpson, 1989) and for the genetic improvement of livestock, the use of markers was thoroughly reviewed by Smith and Simpson (1986). The difficulties for detection of QTL were due to many factors especially for finding the significant polymorphic DNA markers from which the experimental population was established. The most of economic traits in livestock are quantitative, or continuous, which are controlled by a very large number of genes, more likely infinite. In some cases, QTLs with major effect were reported in poultry, pigs, and sheep (Hanset, 1982). In dairy cattle, use of milk protein genes such as α-Casein, β-Casein, and κ-Casein have been used as genetic markers (Gelderman et al., 1985, Bovenhuis and Weller, 1994) and their effects on milk production traits were found significant. However, it is difficult to distinguish whether the significance was due to the QTL linked to the milk protein genes or the direct effect of the milk protein genes. For beef cattle, meat quality is of major interest and its size of heritabilities are relatively moderate to high from 0.15 to 0.60. Therefore, the application of molecular technology using DNA markers into improving meat traits are expected substantial (Sellier, 1994). To detect the linkage between QTL and genetic markers, several experimental designs were suggested. The power of detection of QTL linked to the genetic markers was higher for the design of two parental inbred lines than for a daughter design. The objectives of this study were to examine: 1) the effect of degree of QTL dominance, 2) the sample size of progeny, 3) the actual distance between QTL and markers, on the detection of the linkage between QTL and Markers using a linear regression analysis.

Materials and Methods

For the detection of QTL linked to genetic markers, two inbred parental lines homozygous for both Marker and QTL, MM/QQ (Parent 1) and mm/qq (Parent 2), were assumed. Their progeny, F_1(Mm/Qq), were backcrossed to the parental line with dominant marker genotype (MM,

[1]Address reprint requests to Dr. G. J. Jeon, Department of Breeding and Genetics, National Livestock research Institute, R.D.A., Suwon City, Gyeonggido, 440-351, Korea.

Received March 31, 1995
Accepted August 22, 1995

Parent 1). And then, the progeny data with identified marker genotypes were used for the analysis. To detect the linkage between QTL and marker, data were simulated and were analyzed by a simple linear regression. In real population, only marker genotype is possibly identified, not the QTL genotype. Therefore, it was assumed that two inbred lines were formed by high and low lines for the quantitative traits of economic interest. From mating of F_1 (Mm/Qq) backcrossed to Parent 1 (MM/QQ), the expected genotypes and their frequencies with recombination rate of r are:

Genotype :	MQ/MQ	MQ/mQ	MQ/Mq	Mm/Qq
Fequency :	(1−r)/2	(1/2)r	(1/2)r	(1−r)/2

From the above genotypes, only marker genotypes are distingushable. The detection of QTL linked to the markers can be evaluated by the usual t-test for the comparison of two population means. The hypothesis for the detection of markers linked to the QTL is given as:

$H_0 : (\mu_{MM} - \mu_{Mm}) = 0$
$H_a : (\mu_{MM} - \mu_{Mm}) \neq 0$

Then, the test statistic, t_c, is a simple t-test:

$$t_c = (\hat{\mu}_{MM} - \hat{\mu}_{Mm}) / [S^2_{pooled}(1/n_{MM} + 1/n_{Mm})]^5, \cdots\cdots[1]$$

where μ_{MM} = Mean trait value of marker genotype, MM;
μ_{Mm} = Mean trait value of marker genotype, Mm;
n_{MM}, n_{Mm} = Number of observations for MM and Mm, respectively; and
S^2_{pooled} = Pooled marker variance
$= [(n_{MM} - 1)s^2_{MM} + (n_{Mm} - 1)s^2_{Mm}] / (n_{MM} + n_{Mm} - 2)$, for s^2_{MM} and s^2_{Mm} being phenotypic variances for MM and Mm, respectively.

which is compared with the critical t-table value with degrees of freedom of $(n_{MM} + n_{Mm} - 2)$. With the same principle as for the t-test [1], the exact test result can be achieved by a simple regression analysis as:

$$y = b_0 + b_1 x + e \cdots\cdots[2]$$

where y = trait values for marker genotype(phenotypic records);
x = Coded marker genotypic value (1 = MM, 0 = Mm);
e = Random residual
b_0 = Intercept; and
b_1 = Slope.

Expected value for b_1 was derived as:

$b_0 = E(y) - b_1E(x)$,
where $E(y) = 1/2\ (\mu_{MM} + \mu_{Mm})$;
$E(x) = (1+0)/2 = 0.5$;
b_1 = $Cov(x,y)/V(x)$, for $V(x) = p(1-p)$ and $p = 0.5$.
= $Cov(x,y)/0.25$, for $Cov(x,y) = (1-2r)d^2$ (Weir, 1994)
= $1/4(1-2r)d^2$, for $d = (\mu_{QQ} - \mu_{Qq})$

The results from both [1] and [2] give exactly the same value of type I error with equal power of test. The power of the test is practically difficult due to unknown true values of parameters. The power of test is a function of the type II error of β and is denoted as $(1-\beta)$. One of the use for the power of the test is to detect the difference between a parameter and a specific value. For statistical test, both of α and β are important but are antagnistic to each other. If α increases, β decreases, vice versa. To increase the power of the test, the three ways are possible (Gill, 1978) such that: 1) increase the sample size, 2) reduce the experimental error, and 3) increase the type I error.

TABLE 1. THE PARAMETERS USED FOR THE SIMULATION IN THE STUDY

Mating Scheme : Backcross design

Distribution of marker effect :
1) $MM = (1-r)N(\mu_{QQ}, \sigma^2) + (r)N(\mu_{Qq}, \sigma^2)$
2) $Mm = (r)N(\mu_{QQ}, \sigma^2) + (1-r)N(\mu_{Qq}, \sigma^2)$

$\mu_{MM} = (1-r)\mu_{QQ} + (r)\mu_{Qq}$
$\mu_{Mm} = (r)\mu_{QQ} + (1-r)\mu_{Qq}$

Additive genetic value of QTL : E(QQ) = 300, E(qq) = 0

Variance of QTL : $V(QQ) = V(Qq) = V(qq) = 45^2$

Dominance of QTL :
1) No dominance effect (E(Qq) = 1/2[E (QQ) + E (qq)] = 150
2) Various degree of dominance : D expressed in terms of Qq/QQ
- No dominance = 0%
- Partial dominance = 10%, 20%, 30%, 50%
- Complete dominance = 100%

Sample Size of progeny : 50, 100, 1,000

Recombination rate(r) : 0.05, 0.10, 0.25, 0.45, 0.49, 0.5 (6 levels)

Simulation of Data:

The base populations assumed were two inbred parental lines which were homozygous for both Markers and QTL (MM/QQ and mm/qq). The marker and QTL were assumed diallelic systems, M, m for marker alleles and Q and q for QTL alleles, respectively. The populations were designed as the backcross experiment in which the F1(Mm/Qq) was backcrossed to the parental

line of MM/QQ. The data were generated using the RANNOR procedure of SAS(1988) as:

$y_{MM} \sim (1-r)N(\mu_{QQ}, \sigma^2) + (r)N(\mu_{Qq}, \sigma^2)$

$y_{Mm} \sim (r)N(\mu_{QQ}, \sigma^2) + (1-r)N(\mu_{Qq}, \sigma^2)$

For each population, five replications were generated.

Results and Discussion

The linkage between QTL with various degree of dominance and markers for various recombination rates was examined for different sample sizes. As sample size increased, the standard error of the estimated slope (b_1) tended to decrease, as expected. However, the dominant QTL allele was complete to the recessive QTL allele, eg., D=1, then the standard error was increased (table 2 to 7). Within a given sample size, the degree of dominance of QTL did not significantly affect the size of standard error. However, as the degree of dominance increased, the estimated intercept value was increased while the slope was decreased. For all cases except with the complete dominance effect of QTL, the test result was extremely highly significant ($p < 0.0001$). When the complete dominance of QTL existed, the estimated slopes were not significant at all regardless of the degree of linkage, which indicated that when complete dominance of QTL exists, the detection of linkage is not possible even with the existence of tight linkage. The estimated slope for sample sizes over 100 showed the negative but did not significantly differ from zero. From the result, it can be concluded that only considering the type I error, the necessary sample size over 100 progeny record was quite robust to detect the linkage between marker and QTL. And also, as the actual recombination rate between marker and QTL became larger, the estimated slope also became larger. As r value became smaller, the probability of permitting the type I error was greater. The estimated slope values were decreased as the actual r became larger but the standard error of the slope estimate was not significantly changed. The simple t-test from comparing two populations was exactly the same result as the one from the regression analysis.

TABLE 2. DETECTION OF LINKAGE BETWEEN MARKERS AND QTL WITH VARIOUS SAMPLE SIZES AND THEIR SIGNIFICANCE OF TYPE I ERROR (α) FOR r = 0.05.

D	N	b_0	b_1	Pr > t[1]
0.0	50	9.50±6.09[3]	268.12±8.61	0.0001
0.1	50	44.83±5.20	242.26±7.36	0.0001
0.2	50	75.03±6.10	211.33±8.63	0.0001
0.3	50	92.68±6.70	204.51±9.49	0.0001
0.5	50	157.63±5.77	141.30±8.16	0.0001
1.0[2]	50	302.56±6.18	1.84±8.74	0.8339
0.0	100	16.43±4.13	264.09±5.85	0.0001
0.1	100	51.82±4.41	232.06±6.24	0.0001
0.2	100	71.59±4.34	219.31±6.13	0.0001
0.3	100	104.13±3.99	184.77±5.64	0.0001
0.5	100	156.81±4.58	135.84±6.47	0.0001
1.0[2]	100	302.68±4.48	−2.77±6.33	0.6626
0.0	1,000	15.30±1.33	267.65±1.88	0.0001
0.1	1,000	40.72±1.34	246.20±1.89	0.0001
0.2	1,000	71.14±1.36	217.43±1.92	0.0001
0.3	1,000	101.84±1.30	188.28±1.84	0.0001
0.5	1,000	157.70±1.33	133.54±1.88	0.0001
1.0[2]	1,000	300.77±1.36	−1.29±1.92	0.5013

[1] Type I error from Test for Ho : $b_1 = 0$.
[2] Complete dominance of QTL($\mu_{QQ} = \mu_{Qq}$).
[3] Standard Error.
b_0 = Intercept, b_1 = slope.
D = Degree of dominance for QTL (Qq/QQ).
N = Progeny size (equal for each marker genotype: $N_{MM} = N_{Mm}$).

TABLE 3. DETECTION OF LINKAGE BETWEEN MARKER AND QTL WITH VARIOUS SAMPLE SIZES AND THEIR SIGNIFICANCE OF TYPE I ERROR (α) FOR r = 0.10

D	N	b_0	b_1	Pr > t[1]
0.0	50	23.48±5.56[3]	246.85±7.86	0.0001
0.1	50	53.09±5.74	215.62±8.12	0.0001
0.2	50	83.12±6.67	198.89±9.57	0.0001
0.3	50	103.11±5.79	176.54±8.19	0.0001
0.5	50	163.95±5.39	128.29±7.63	0.0001
1.0[2]	50	291.05±6.16	12.23±8.71	0.1632
0.0	100	31.90±4.01	242.98±5.67	0.0001
0.1	100	59.01±3.68	217.50±5.21	0.0001
0.2	100	74.90±4.11	200.55±5.81	0.0001
0.3	100	113.81±4.55	172.05±6.44	0.0001
0.5	100	165.32±3.83	114.83±5.41	0.0001
1.0[2]	100	301.16±4.38	13.04±6.19	0.2301
0.0	1,000	28.40±1.28	239.38±1.81	0.0001
0.1	1,000	97.37±1.31	215.60±1.85	0.0001
0.2	1,000	83.38±1.33	192.38±1.88	0.0001
0.3	1,000	112.03±1.30	168.34±1.84	0.0001
0.5	1,000	165.70±1.28	120.92±1.81	0.0001
1.0[2]	1,000	299.52±1.31	0.69±1.85	0.7080

TABLE 4. DETECTION OF LINKAGE BETWEEN MARKERS AND QTL WITH VARIOUS SAMPLE SIZES AND THEIR SIGNIFICANCE OF TYPE I ERROR (α) FOR r = 0.25

D	N	b_0	b_1	Pr > t[1]
0.0	50	63.62±4.93[3]	155.48±6.97	0.0001
0.1	50	101.73±4.83	136.46±6.83	0.0001
0.2	50	123.17±5.46	122.30±7.73	0.0001
0.3	50	140.01±4.36	100.84±6.16	0.0001
0.5	50	195.08±5.04	72.49±7.13	0.0001
1.0[2]	50	304.00±4.99	−10.08±7.05	0.1562
0.0	100	71.20±3.47	148.25±4.91	0.0001
0.1	100	103.26±3.36	129.96±4.75	0.0001
0.2	100	124.31±3.12	119.68±4.41	0.0001
0.3	100	146.93±3.56	104.19±5.03	0.0001
0.5	100	187.21±3.60	74.21±5.09	0.0001
1.0[2]	100	303.02±3.66	−1.83±5.18	0.7246
0.0	1,000	74.14±1.10	150.16±1.56	0.0001
0.1	1,000	99.01±1.12	133.10±1.59	0.0001
0.2	1,000	119.05±1.12	121.60±1.58	0.0001
0.3	1,000	142.27±1.10	104.95±1.57	0.0001
0.5	1,000	185.90±1.11	74.06±1.57	0.0001
1.0[2]	1,000	301.29±1.13	−0.29±1.60	0.8588

TABLE 5. DETECTION OF LINKAGE BETWEEN MARKERS AND QTL WITH VARIOUS SAMPLE SIZES AND THEIR SIGNIFICANCE OF TYPE I ERROR (α) FOR r = 0.45

D	N	b_0	b_1	Pr > t[1]
0.0	50	131.56±4.46[3]	27.90±6.30	0.0001
0.1	50	149.43±4.59	28.88±6.49	0.0001
0.2	50	171.14±4.07	19.21±5.75	0.0012
0.3	50	185.17±4.71	16.83±6.66	0.0131
0.5	50	220.70±4.90	3.71±6.93	0.5937
1.0[2]	50	301.45±4.70	−4.21±6.65	0.5279
0.0	100	136.38±3.21	27.31±4.55	0.0001
0.1	100	148.80±3.07	26.48±4.33	0.0001
0.2	100	167.41±3.11	26.52±4.39	0.0001
0.3	100	189.23±3.19	16.72±4.52	0.0003
0.5	100	210.80±3.19	22.58±4.51	0.0001
1.0[2]	100	300.17±3.05	−6.23±4.31	0.1503
0.0	1,000	135.77±1.03	29.87±1.45	0.0001
0.1	1,000	151.92±1.02	26.50±1.45	0.0001
0.2	1,000	168.65±1.01	22.15±1.42	0.0001
0.3	1,000	185.52±1.04	21.18±1.47	0.0001
0.5	1,000	217.70±1.02	13.81±1.44	0.0001
1.0[2]	1,000	300.05±1.04	−0.15±1.47	0.9197

TABLE 6. DETECTION OF LINKAGE BETWEEN MARKERS AND QTL WITH VARIOUS SAMPLE SIZES AND THEIR SIGNIFICANCE OF TYPE I ERROR (α) FOR r = 0.49

D	N	b_0	b_1	Pr > t[1]
0.0	50	143.50±4.52[3]	11.51±6.39	0.0746
0.1	50	170.06±4.38	−1.92±6.20	0.7579
0.2	50	173.50±5.17	1.01±7.31	0.8900
0.3	50	194.36±4.32	−0.88±6.12	0.8854
0.5	50	227.68±4.47	4.70±6.32	0.4586
1.0[2]	50	301.97±4.46	0.20±6.31	0.9752
0.0	100	149.78±3.12	9.09±4.41	0.0406
0.1	100	161.93±3.26	8.07±4.61	0.8130
0.2	100	177.20±2.97	4.85±4.41	0.8488
0.3	100	192.68±3.17	4.08±4.48	0.3648
0.5	100	225.12±3.06	−0.03±4.33	0.5192
1.0[2]	100	299.73±3.22	−0.41±4.55	0.7683
0.0	1,000	146.61±1.00	6.18±1.41	0.0001
0.1	1,000	162.20±1.97	5.27±1.37	0.0001
0.2	1,000	176.96±1.98	4.84±1.39	0.0005
0.3	1,000	191.38±1.02	5.50±1.44	0.0001
0.5	1,000	223.33±0.99	3.56±1.40	0.0110
1.0[2]	1,000	300.21±1.01	−0.67±1.43	0.6392

TABLE 7. DETECTION OF LINKAGE BETWEEN MARKERS AND QTL WITH VARIOUS SAMPLE SIZES AND THEIR SIGNIFICANCE OF TYPE I ERROR(α) FOR r = 0.50

D	N	b_0	b_1	Pr > t[1]
0.0	50	154.54±4.08[3]	0.98±5.77	0.8657
0.1	50	165.47±4.42	2.76±6.26	0.6606
0.2	50	171.40±3.98	1.48±5.63	0.7930
0.3	50	193.14±4.51	2.96±6.38	0.6436
0.5	50	222.46±4.82	−3.51±6.81	0.6079
1.0[2]	50	296.17±4.29	2.18±6.07	0.7206
0.0	100	146.03±3.11	2.76±4.39	0.5306
0.1	100	167.15±2.95	−0.65±4.17	0.8758
0.2	100	181.69±3.01	−2.21±4.26	0.6040
0.3	100	192.94±3.26	10.40±4.61	0.6252
0.5	100	226.95±3.21	−1.32±4.54	0.7726
1.0[2]	100	297.58±3.13	4.75±4.43	0.2851
0.0	1,000	150.35±1.01	−1.97±1.43	0.1687
0.1	1,000	166.38±1.00	−1.27±1.42	0.3698
0.2	1,000	180.98±1.01	−0.37±1.42	0.7964
0.3	1,000	195.06±0.98	−2.24±1.38	0.1043
0.5	1,000	224.67±1.01	−0.64±1.43	0.6548
1.0[2]	1,000	300.25±1.00	0.23±1.41	0.8729

Conclusion

To find significant genetic markers which can account for the genetic variability is a quite difficult task. Once two inbred parental lines were established, the detection of QTL linked to the specified marker genotype is relatively simple. The method for the detection of linkage between QTL and genetic marker (Non-functional gene) is either a simple t-test between two marker genotype, MM and Mm, or equivalently the test for slope being non-zero from simple regression analysis. However, if the QTL linked to the marker has an effect of complete dominance, then the detection of linkage between QTL and the markers was impossible even with the tight linkage. And also, as the actual linkage map distance, r, increases, the size of progeny sample must be increased especially when the actual r approaches toward 0.5.

Literature Cited

Bovenhuis, H. and J. L. Weller. 1994. Mapping and Analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. Genetics 137:267-280.

Gelderman, H., U. Pieper and B. Roth. 1985. Effects of marker chromosome sections on milk performance in cattle. Theor. Appl. Genet. 70:138-146.

Gill, J. L. 1978. Design and Analysis of Experiments in the animal and medical sciences. Volume 1. The Iowa State Univ. Press. Ames, Iowa.

Hanset, R. 1982. Major genes in animal production, examples and perspectives: Cattle and pigs. In proceedings of the 2nd World Congress of Genetics Applied to Livestock Production, Madrid, Oct, 4-8, 1982, 8:439-452. Madrid: Editorial Garsi.

Knapp, S. J., W. C. Bridges and D. Birkes. 1990. Mapping quantitative loci using molecular marker linkage maps. Theor. Appl. Genet. 79:583-592.

Lander E. S. and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121:185-199.

Maritinez, O. and R. N. Curnow. 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. 85:480-488.

SAS. 1988. SAS Language Guide for Personal Computers, Release 6.03 ed. Cary, NC:SAS Inst. Inc.

Sellier, P. 1994. The future role of molecular genetics in the control of meat production and meat quality. Meat Sci. 36:29-44.

Simpson, S. P. 1989. Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms usmg inbred lines. Theor. Appl. Genet. 77:815-819.

Smith, C. and S. P. Simpson. 1986. The use of genetic polymorphism in livestock improvement. J. Anim. Breedg. Genet. 103:205-217.

Weir, B. S. 1994. Analysis of molecular genetic information. North Carolina State Univ. Press.

Weller, J. I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics. 42:627-640.