



This is to certify that the
dissertation entitled

**POWER AND ACCURACY OF DETECTING LINKAGE BETWEEN
QUANTITATIVE TRAIT LOCI AND GENETIC MARKERS**

presented by

Zhiwu Zhang

has been accepted towards fulfillment
of the requirements for

Ph.D. Animal Science

degree in

Major professor

Date December 15, 1998

PLACE IN RETURN BOX to remove this checkout from your record.
 TO AVOID FINES return on or before date due.
 MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
JAN 22 2006		
JUL 05 2007		
111809		

**POWER AND ACCURACY OF DETECTING LINKAGE BETWEEN
QUANTITATIVE TRAIT LOCI AND GENETIC MARKERS**

By

Zhiwu Zhang

A DISSERTATION

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

DOCTOR OF PHILOSOPHY

Department of Animal Science

1998

ABSTRACT

POWER AND ACCURACY OF DETECTING LINKAGE BETWEEN QUANTITATIVE TRAIT LOCI AND GENETIC MARKERS

By

Zhiwu Zhang

Power of testing additive effects of quantitative trait locus (QTL) linked to genetic markers and accuracy of the parameter estimators from a QTL and polygenic additive mixed model were evaluated through simulations. The underlying conditions included QTL location, additive genetic variance due to QTL linked to markers, additive genetic variance due to polygenic effects unlinked to markers, and residual effects.

Granddaughter population designs were simulated under a mixed model with additive polygenic and QTL effects for various combinations of factors at various levels. Items of interest included: selection schemes, marker intervals, number of daughters, magnitude of QTL effects and heritability of traits. Sons in the design were selected by 4 alternative schemes of random, disruptive, truncation and stabilized. Fifty replicates were generated for each of the populations and analyzed separately by restricted maximum likelihood.

Estimates of QTL location, variance due to QTL, and polygenic and residual effects were unbiased in unselected populations. In fact, estimates of

QTL location were unbiased in populations under all selection schemes.

However, estimates of variances due to QTL, polygenic and residual effects were biased in populations under nonrandom selection schemes. Magnitude of the biases were dependent on marker intervals, numbers of daughter, magnitudes of QTL effect and heritability levels.

Selection schemes had a significant influence on the power of testing linkage between genetic markers and QTL using restricted maximum likelihood. Disruptive selection generated higher power than random selection, whereas truncation and stabilizing selections have less power than random selection.

Power of test using restricted maximum likelihood also depends on the number of daughters of each sire, marker intervals, magnitude of QTL effect and heritability. Power of test was higher with more daughters per sire, smaller marker intervals, larger magnitude of QTL effect, and higher heritability. The magnitude of the differences due to changing a factor was larger when power was less saturated by other factors.

A Dedication to My Parents

ACKNOWLEDGMENTS

This research reflects the efforts of many. My major professor, Dr. Ivan Mao, is acknowledged for his role as a mentor and friend without whom I could not have succeeded. Dr. Dennis Banks and Dr. Rob Tempelman are recognized for their assistance in the study. From the Department of Statistics, gratitude is extended to Dr. Raoul LePage for his invaluable advice in the interpretation of the results.

I thank Dr. Fernando Grignola for providing QTL analysis program and comments for the study.

Appreciation is extended to Dr. Wejun Zhao for his role in the initiative of my program at Michigan State University.

I thank my office mates and fellow graduate students, Charlie Chang, David Norris, Morten Rye, Thomas Mark Lars Nielsen, Ole Pedersen, Mara Preiler, Renee Bell and Kadir Kizilkaya for their help and friendship.

A final note of recognition is extended to my wife, Wanling, and my son, Jingwei. My wife always understands me and always loves me. My son does not always understands me but always loves me.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
GENERAL INTRODUCTION	1
CHAPTER 1	
LITERATURE REVIEW	5
Introduction	5
Resource Populations	9
Line Crosses	9
Outbreed Lines	11
Statistical Methods	13
ANOVA.....	13
Regression	14
Maximum Likelihood.....	15
BLUP Based Methods	16
Sampling Strategy.....	20
Random Selection	20
Disruptive Selection.....	20

Truncation Selection.....	21
Stabilizing Selection	22
Marker Assisted Selection.....	23
Genetic Markers	23
Advantage of MAS	25
Limitations of MAS	26
 CHAPTER 2	
METHODOLOGY	27
Population	27
Simulation of Genotypes	30
Simulation of DYDs	30
Selection Schemes	33
Statistical Analysis	34
Experimental Design	36
 CHAPTER 3	
BIASES IN GENETIC PARAMETER ESTIMATES FROM A POLYGENE AND QTL MIXED MODEL UNDER GRANDDAUGHTER DESIGN.	37
Abstract.....	37
Introduction	38
Materials and Methods.....	40
Results and Discussions	45
Conclusions and Implications	54

CHAPTER 4

POWER OF DETECTING MARKER ASSOCIATED QTL EFFECTS IN GRANDDAUGHTER DESIGN	56
Abstract.....	56
Introduction	57
Materials and Methods	60
Results and Discussions	65
Conclusions and Implications.....	77

GENERAL CONCLUSION.....	78
--------------------------------	-----------

APPENDICES

A - Expectation of Heritability Weighted DYD	79
B - Input and Output of Grignola's Program	82
C - Flow Chart of Pedigree.....	83
D - Flow Chart of Subroutine of Crossover	84
E - Flow Chart of Subroutine of Randasign	85
F - Flow Chart of Subroutine of Phenotype	86
G - FORTRAN (90) Code of Simulation Program	87
BIBLIOGRAPHY	95

LIST OF TABLES

Table 1 - Median and SE of Bias Rates (%) in the Estimation of QTL Allelic Variance	50
Table 2 - Median and SE of Bias Rates (%) in the Estimation of Polygenic Variance.	51
Table 3 - Median and SE of Bias Rates (%) in the Estimation of Residual Variances.	52
Table 4 - Power from Populations Subjected to Alternative Selection Schemes with Different Levels of Heritability (h^2) and Magnitudes of QTL Effect (v^2)	68

LIST OF FIGURES

Figure 1 - Population Structure of Granddaughter Design	28
Figure 2 - Pedigree of 20 Sires in Simulated Granddaughter Design Population	29
Figure 3 - Genetic Marker Loci and one QTL on an Autosomal Chromosome ..	31
Figure 4 - Estimates of Polygenic, QTL and Residual Variances	48
Figure 5 - Profiles of Test Statistics	66
Figure 6 - Comparison of Powers between Random and Truncation Selection Schemes	69
Figure 7 - Marginal Effects on Power of Test with Random Selection	70
Figure 8 - Effect of Number of Daughters on Power of Test under Truncation Selection Scheme	71
Figure 9 - Effect of Heritability on Power of Test under Truncation Selection Scheme	72
Figure 10 - Effect of Marker Interval on Power of Test under Truncation Selection Scheme	73
Figure 11 - Effect of Magnitude of QTL Effect on Power of Test under Truncation Selection Scheme	74

GENERAL INTRODUCTION

For many years, Animal breeders have changed the genetic composition of farm animals through selection without knowledge of the underlying genes (Bovenhuis et al., 1997). This approach is based on the assumption that a trait is controlled by infinitive genes each having an equally small effect. Due to the environmental effects and the segregation of genes from parent to offspring, accurate estimation of the breeding value of an animal depends on the amount of records on the phenotype of the individual itself and/or relatives. In general, the requirement of a large number of records postpones the age at which the animal can be selected as a parent and therefore, restricts the annual genetic progress (Bovenhuis et al., 1997).

However, genes with large effects on economically important traits have been identified. Notable examples are the double muscling gene in cattle (Hanset and Michaux, 1985a and 1985b), the gene determining halothane sensitivity in swine (Smith and Bampton, 1977), and the estrogen receptor genes influence a litter size in swine (Rothschild et al., 1994).

The possibility of finding major genes affecting economically important traits has been greatly increased due to the development of methods to detect polymorphism at the DNA level, for examples, RFLP (Paterson et al., 1988; Visscher et al., 1990; Soller and Beckman, 1990; Georges et al., 1995). Genetic

linkage maps of polymorphic molecular markers have been developed in many domestic animal species. It was possible for the first time to begin the systematic search for individual loci affecting quantitative traits of economic importance.

Unfortunately, the majority of genetic markers, especially DNA based polymorphism, are not likely to affect animal performance. They may be located closely to the genes affecting quantitative traits (Soller and Beckman, 1990).

The detection of QTL linked to genetic markers is a statistical inference. It is structured around the formal test of the null hypothesis, which proposes that the recombination rate between genetic markers and QTL is 50%. Experiments are designed to test the null hypothesis against a alternative hypothesis. Statistical analysis may either reject null hypothesis, suggesting the existence of the linkage, or may not reject it, suggesting that genetic markers are not linked to the QTL.

A statistical interface involve two kinds of errors: Type I, when a true null hypothesis is rejected, and Type II, when a false null hypothesis is accepted. The probabilities of committing Type I and II errors are denoted by α and β , respectively. The probability of not committing a Type II error is called power of test, which is $(1 - \beta)$.

Power of test is more important when nonsignificant results are obtained, since in this case, a valid assertion of those results is only possible if the power is high. On the other hand, higher power leads to larger sample size in general, which can be costly. Powers of 0.8 to 0.9 are generally used to accept null

hypotheses (Searcy-Bernal, 1994). Therefore, it is crucial to predetermine the desired power of the linkage test in evaluating candidate markers.

Once marker linked QTL effects are detected, the use of the linkage between marker loci and QTL provide additional information for increasing the accuracy of selecting the genetic difference between individuals. Efficiencies were compared between purely phenotype selection and marker assisted selection (Kashi et al., 1990; Meuwissen and Van Arendonk, 1992; Meuwissen and Goddard, 1996; Smith and Simpson, 1986; Stam, 1986; van der Beek and Van Arendonk, 1996). Marker assisted selection was more efficient than selection without using marker information in 1) early generations; 2) lowly heritability traits; 3) large populations and 4) close linkage between marker and QTL.

This conclusion is based on the assumption that the parameters of quantitative trait loci are known without error. Simulations demonstrated that overestimation of QTL variance decreased genetic gain for marker assisted selection (MAS) over the long term. For an error of 15 centiMorgan (cM) on the location of QTL, genetic superiority of MAS was reduced by 80% in the first generation than MAS without error on QTL location (Spelman and Van Arendonk, 1997b).

The power to detect linkage and accuracy of estimates of QTL parameters depended on several factors, including statistical method, sampling strategy, sample size, marker density and magnitude of QTL effects. Powers have been examined for the statistical method of ANOVA (Weller, 1990a), maximum

likelihood (Knott and Haley, 1992b; Le-Roy and Elsen, 1995; Carbonell et al., 1993; Jensen, 1989; Lander and Botstein, 1989; Knapp and Bridges, 1990; Elsen et al., 1997) and regression (Moreno-Gonzalez, 1992). Knowledge power using restricted maximum likelihood remains unknown.

The accuracy and precision of estimates using restricted maximum likelihood (REML) were evaluated in many studies for additive polygenic models in unselected and selected populations (Henderson, 1975a; Banks et al., 1985; Gianola et al., 1986; Beaumont, 1991; Sorensen and Kennedy, 1984). Grignola et al., (1996b) evaluated accuracy and precision of estimates using REML with QTL and polygenic mixed models in unselected populations.

The objective of this study was to evaluate parameter estimates from QTL and additive polygenic mixed models and to examine power of detecting QTL using restricted maximum likelihood method in combinations of different selection schemes, marker intervals, magnitude of QTL effect, heritability, and number of daughters in a granddaughter design.

Chapter 1

LITERATURE REVIEW

Introduction

For many years, Animal breeders have changed the genetic composition of farm animals through selection without knowledge of the underlying genes. Recent developments in molecular biology have changed this situation and have allowed the genes controlling traits or genetic markers linked to the genes to be identified (Bovenhuis et al., 1997).

Classical animal breeding approaches to estimate the additive genetic value of an individual depends on phenotypic observations on the individual itself and/or relatives. This approach is based on the assumption that a trait is controlled by infinitive genes each having an equally small effect. The action of individual gene cannot be observed directly and a trait is generally described in terms of summary statistics such as the heritability.

For most of the traits of interest to animal breeders, differences in phenotypic observations are determined by both genetic and environmental differences. Further more, segregation of genes takes place each time genes are transmitted from parent to offspring. As a result of these factors, accurate

estimation of the breeding value of an animal is possible only if a large number of records on the phenotype of the individual itself and/or relatives are available. In general, the requirement of a large number of records postpones the age at which the animal can be selected as a parent and therefore, restricts the annual additive genetic progress.

However, some genes with a large effect on economically important traits have been identified. Notable examples are the dwarfing gene in poultry, the Booroola gene affecting ovulation rate in sheep, the double muscling gene in cattle (Hanset and Michaux, 1985a and 1985b), the gene determining halothane sensitivity in swine (Smith and Bampton, 1977), and estrogen receptor gene controlling litter size in swine (Rothschild et al., 1994).

Unfortunately, most genetic markers, especially DNA based polymorphism, for example, RFLP, are not likely to be the alleles that affect the performance of animals. But they may be linked to the genes affecting quantitative traits. Several studies have shown that individual loci affecting quantitative traits can be detected if linked to genetic markers (Soller and Beckman, 1990). For example, a genetic marker on chromosome 8 was found to be linked to a QTL with an additive effect of 3 ova (Rathje et al., 1997).

Once genetic marker, QTL linkage is established, information of linkage provides additional information for increasing accuracy of selection, especially for the traits that are difficult to improve when using traditional selection methods (Kashi et al., 1990; Meuwissen and Van, 1992; Meuwissen and Goddard, 1996; Smith and Simpson, 1986; Stam, 1986; van der Beek and Van Arendonk, 1996).

The earliest usage of linkage between genetic marker and QTL was demonstrated by Sax (1923). However, it was restricted by the limitation of the number of genetic markers. Genetic markers are available now due the development of methods to detect polymorphism at the DNA level (Kashi et al., 1990). The genetic linkage maps of polymorphic molecular markers have been developed in many domestic animal species. It is possible, for the first time, to begin the systematic search for individual loci affecting quantitative traits of economic importance (Paterson et al., 1988; Visscher et al., 1990; Soller and Beckman, 1990; Georges et al., 1995).

There are several activities involved in identifying and utilizing the linkage between genetic markers and QTL. These include recording animals for the character of interest, typing them for genetic markers, testing for statistical associations between genetic markers and phenotypic score and, if associations are found, applying marker assisted selection in breeding schemes (Paterson et al., 1988). Two statistical inferences are involved in these activities. One is the power of test. The other is the estimation of QTL location, QTL variance, polygenic variance and residual variance.

Identification of linkage between a genetic marker and QTL is structured around the formal test of a null hypothesis (H_0), which assumes genetic markers are not linked to QTL. Statistical inference is to test H_0 against the alternative hypothesis, which propose that genetic markers are linked to the QTL. Statistical analysis may either reject H_0 , suggesting the existence of the linkage, or may not reject it, suggesting that genetic are not linked to the QTL at some level of

confidence. This statistical inference involves two types of errors: Type I error, the rejection of a true null hypothesis, and Type II error, the acceptance of a false null hypothesis. The probabilities of not committing type II is called power of test.

Statistical power is more important when nonsignificant results are obtained. In this case a valid assertion of those results is only possible if the power is high.

In order to decrease the probability of making Type II error, power of test must reach a certain level. Levels of 0.05 and 0.01 are accepted as popular significance levels for rejecting the null hypothesis, while powers of 0.8 to 0.9 are required for acceptance of a null hypotheses (Searcy-Bernal, 1994).

Power of testing linkage between genetic markers and QTL depends on many factors (e.g., statistical procedures, sample size, population structure, magnitude of QTL effect, recombination rate, heritability of traits, and sampling strategies). Because of the expense in both genotyping and generating a experimental population, efforts have been spent in optimizing the design of experiments for a optimal power.

A quasi-theoretical numerical method can be used to predict power of detecting QTL from the shape of the multidimensional expected likelihood surface (Mackinnon and Weller, 1995). However, theoretical calculations of power are not always empirically possible. Computer simulation can be employed in the situation.

Once a genetic marker associated QTL effect is detected, it is of interest to find the location of QTL in the genome, and to determine the effects of the

QTL by estimating the phenotypic variance explained by QTL linked to markers. All of these parameters are required in predictions of the breeding values in cooperating genetic marker information.

An overview is given for the characteristics of the affecting factors in identification and utilization of linkage between marker and QTL, including different resource populations; statistical methods strategies of selective genotyping and marker assisted selection.

Resource Populations

There are two primary types of data used for mapping a quantitative trait locus: data derived from linecrosses that include backcross and F_2 populations (Soller et al., 1976) and data from outcross populations (Soller and Genizi, 1978; Weller, 1990b).

Line Crosses

Most successful QTL mapping efforts described to date have exploited F_2 or backcrosses obtained from parental populations divergent for the traits of interest (Paterson et al., 1988). The main reason is that line crosses generate disequilibrium.

Linkage disequilibrium between genetic markers and QTL creates difference on a trait across marker genotypes. One way of introducing linkage disequilibrium in a population is by crossing lines that differ with respect to their allele frequencies at marker loci and QTL. Associations between genetic markers and QTL can be studied by comparing the phenotypic performance of F_2 or back cross individuals with different marker genotype configurations.

In ideal situation, all F_1 individuals are heterozygous for the marker as well as QTL. There is complete linkage disequilibrium between the marker and QTL in the F_1 . All F_1 individuals have the same linkage phase. An empirical case of such a design is the use of inbred lines.

Linecrosses are frequently used in laboratory animals (e.g., mice) and plants. For farm animals, however, inbred lines are seldom available. In addition, rearing large numbers of F_1 and F_2 individuals is possible for some farm animals (e.g., chicken) but not for others (e.g., cattle) because of the long generation intervals and costs of the experiment.

Power to detect QTL has been described for crosses of inbred lines (Soller et al., 1976; Weller, 1986; Jansen, 1993; Lander and Botstein, 1989; Luo and Woolliams, 1993; Simpson, 1989; Knott et al., 1992b and Darvasi et al., 1993).

Outbreed Lines

Linkage disequilibrium in an outcross population between a marker and linked QTL is more likely within families (Neimann-Sorensen and Robertson, 1961).

In commercial dairy cattle populations, sires often have hundreds or even thousands of daughters produced by artificial insemination. Thus, a segregating QTL can be detected by analyzing the progeny of heterozygous sires. Daughters inheriting the different sire marker alleles should also display a difference for the quantitative trait.

There is more power to perform this analysis over multiple sires rather than a single sire. Even if a sire is heterozygous for the genetic marker, he may still be homozygous for the QTL. Different marker genotypes can not show any difference on the QTL genotypes.

Further more, if the sire is heterozygous for both marker and QTL loci, linkage phase between the marker and QTL alleles may be different from different sire. Thus, analysis should be performed within sires.

The analysis can be performed within paternal half-sib families using either the daughter design or the granddaughter design (Weller, 1990a). The basic idea of the daughter design is to trace marker alleles from the sires to his daughters and to determine whether daughters that inherited alternative sire alleles differ with respect to the quantitative trait. In the daughter design, daughters of a sire are scored for markers and evaluated for the quantitative trait.

However, in the granddaughter design, the sons of a proven sire are scored for genetic markers and granddaughters are evaluated for the quantitative trait. In this latter case, the observations on the granddaughters are used to estimate the breeding value of the sons. This breeding value has a lower residual variance compared to a single observations which increases the power of the experiment (Weller et al., 1990b; Van der Beek et al., 1995).

In granddaughter design, marker associated effects measured in granddaughter generation will be halved with respect to marker associated effects measured in the daughter generation. Nevertheless, the standard error of the contrasts are smaller, so that granddaughter designs may be able to deliver equivalent power while scoring fewer individuals for the markers, the most costly part of the program. Also, it may be easier to collect blood or semen samples from sons of sires, concentrated in AI centers, than from their daughters, scattered over many farms. Power to detect QTL has been described for outbreeding populations (Knott et al., 1992b; Bovenhuis and Weller, 1994).

Because of different linkage relationship among sires, outcross populations (e.g. the daughter design) have less statistical power than line crosses. To detect a QTL with a substitution effect of 10-30% of a phenotypic standard deviation, it is necessary to determine the genetic marker genotype of thousands of daughters (Weller, 1990a). With inbred lines, the same power can be obtained by determining the genetic marker genotypes of less than 1000 progeny (Soller et al., 1976).

Statistical Methods

Statistical methods of mapping QTL vary depending the structure of populations and interpretation of the nature of QTL effect. Effect of QTL can be considered as either fixed effect or random. This leads to a different choice of statistical approaches. Fixed effect is only one that t test (Simpson, 1989), ANOVA and regression methods can deal with, while, maximum likelihood method, as well as BLUP based methods are suitable for both (Simpson, 1989; Lander and Botstein, 1989; Cowan et al., 1990; Weller, 1990a; Haley et al., 1994; Lander and Botstein, 1989; Knott and Haley, 1992a).

ANOVA

ANOVA is performed by contrasting marker genotype effects by (Soller et al., 1976). This method of analysis yields estimates of marker allele substitution effects. However, the analysis does not provide any information about the location of the QTL (e.g., the method cannot distinguish between a slightly linked QTL with a large effect and a closely linked QTL with small effect). Another disadvantage of this type of analysis is that some of the progeny cannot be assigned to one of the two parental alleles. These animals have to be excluded from the analysis, which results in reduced power. Weller and Wyler, (1992) evaluated the power of ANOVA in daughter design and granddaughter design.

Regression

Haley and Knott (1992) and Martinez and Curnow (1992) independently introduced a regression method. Regression is performed on the probability of an individual having a QTL genotype, given the genotype for the flanking markers.

ANOVA is identical to regression with specific protonts. If the probability that an animal has inherited a particular QTL allele from its parent is based only on information from single flanking informative marker, regression is equivalent to an analysis of variance. The position and the effect of the QTL cannot be disentangled.

However, QTL location can be estimated by utilizing marker bracket. This method results in an estimate of the QTL position as well as the variance explained by genotype contrasts. The probability of an individual having a QTL genotype, given the genotype for the flanking markers depends upon the location of the QTL. By moving a putative QTL along the chromosome, the most likely position of the QTL corresponds to the position with minimum residual sum of squares (Whittaker et al., 1996).

Once QTL genotype probability by marker genotype is determined, standard statistical software packages can be used for regression part of the analysis (Spelman et al., 1996; Weller et al., 1990b; Hoeschele, 1990; Cowan et al., 1990; and Goddard, 1991). Power of using regression was investigated by (Moreno-Gonzalez, 1992 and Jansen, 1994a). The distribution of the statistics of testing marker associated QTL effect was studied (Hyne and Kearsey, 1995).

Maximum Likelihood

Weller (1986) developed maximum likelihood methods to detect marker associated QTL effects. It involves maximizing and comparing the likelihood of the data under different genetic models to ascertain the most likely genetic structure. The maximum likelihood of the data under a additive polygenic model is compared with that under the combined model containing a major gene linked to markers and additive polygenic component. A significant improvement in the likelihood obtained by incorporating a major gene in the model provides evidence for a linkage between QTL and genetic markers.

Maximization is usually with respect to five parameters: mean, the additive and dominant QTL effect, recombination rate between marker and QTL, and within QTL genotype residual variance (Weller, 1986; Bovenhuis and Weller, 1994).

Interval mapping is more accurate than a single marker in estimating QTL location. The advantage depends upon the heterozygosity of the markers and the position of the QTL within the flanking markers (Darvasi et al., 1993; van der Beek et al., 1995)

The power of test depends on family size (Knott et al., 1992b), QTL effect, recombination between marker and QTL (Le-Roy and Elsen, 1995) and heritability of quantitative trait (Carbonell et al., 1993 and Jensen, 1989).

Currently, the most popular analytical method to investigate QTL is carried out by either regression, or by the maximum likelihood approach. The two methods yield very similar results (Lander and Botstein, 1989; Haley and Knott, 1992; Martinez and Curnow, 1994), the regression approach applies a more straightforward test of significance and is programmable using standard statistical packages.

BLUP Based Methods

Method of ANOVA, regression and ML were developed mainly for line cross populations. They can not fully account for the more complex data structures in outcross populations (e.g. data on several families with relationships across families, unknown linkage phases in parents, unknown number of QTL alleles in the population, and varying amounts of data information on different QTL or in different families). Best linear unbiased prediction (BLUP) based methods were developed to overcome these drawbacks.

For data that does not contain genetic marker information, BLUP has proved to be a very flexible method. BLUP can handle data with many nongenetic effects (e.g. season), with arbitrary pedigree structure, and with nonrandom mating and selection. Currently BLUP is effectively used for the prediction of breeding values of farm animals.

The prediction of an animal's breeding value is based on phenotypes of the animal itself and relatives. When only phenotypes are considered, the

contribution of observations on relatives to an animal's breeding value depends on the additive genetic relationship, *i.e.*, the average proportion of genes shared in common by descent, and the heritability of the trait. The additive relationship between individuals is formed without knowledge of the actual gene contributed from parent to its offspring.

Recently, the concept of the additive genetic relationship has been extended to the gametic relationship where paternal and maternal gametes of an animal are considered separately. The gametic relationship has been used for constructing the relationship due to dominance effects and for the analysis of gametic imprinting effects (Schaeffer et al., 1989). For these reasons it is also likely to be useful for analyzing data containing information on genetic markers, if the assumptions of BLUP are reasonably satisfied (Grignola, 1996a).

Information on an animal's genotype at a marker locus provides information on transmission of a chromosomal region from parent to offspring. If QTL are located in the chromosomal region, this information can be used to obtain a more accurate estimate of breeding values because the inheritance of alleles at the chromosomal region can be traced more precisely than inheritance at an unmarked QTL. In this case, the additive genetic value of an animal can be partitioned into additive genetic value at the marked chromosomal region and the sum of additive genetic effects of polygenes linked to markers.

Construction of variance and covariance of QTL effects linked to markers is the key factor to apply BLUP in QTL analysis. Fernando and Grossman (1989) showed that information on a single marker can be used in an animal model by

fitting additive effect for alleles at QTL linked to a genetic marker and additive polygenic effects for alleles at the remaining quantitative trait loci. Goddard (1991) extended the model to include information from more than one marker. Wang et al. (1995) presented an algorithm for additive relationship matrix construction without the requirement on the information of origin of alleles. Meuwissen and Goddard (1996) presented a method in which the covariance matrix of effects at the marked QTL is approximated. This approximation reduces the computational requirements.

The linear mixed model equation is greatly enlarged by including QTL effect linked to markers. The size of the equation can be reduced by use of a reduced animal model (Goddard, 1991). In this case, effects are only predicted for animals that are parents. Breeding values and additive QTL effects for non-parents can be obtained by back solving.

Another way of reducing the size of equation is to link phenotypes to the total additive effects and link total additive effects to QTL. An animal model method to reduce the number of equations per animal to one was presented by Van Arendonk et al. (1994) combining information on marker linked QTL and QTL unlinked to marker into one numerical relationship matrix. A reduced animal model version of Van Arendonk et al. (1994) is also developed by Saito and Iwaisaki (1996).

Hoeschele and VanRaden (1993a and 1993b) indicated that if some of the animals to be evaluated do not have marker data and do not provide relationship ties among genotyped descendants with known marker data, the

marker linked QTL equations for such animals can be eliminated. The inverse of a covariance matrix among total additive polygenic effects and the additive effects of the QTL alleles can be obtained directly. When only a small fraction of the animals are genotyped for markers and the remaining fraction do not provide marker data, the procedure of Hoeschele and VanRaden (1993a and 1993b) has the advantage of reducing the number of equations to be solved.

All of the algorithms above are based on an unknown parameter, which is the recombination rate between genetic markers and QTL. One solution is to maximize the log likelihood at each point at the marker interval by using restricted maximum likelihood with respect to the other parameters, e.g. additive variance due to allele at QTL linked to marker, additive variance due to alleles at the remaining quantitative trait loci, and residual variance. The location of QTL was estimated at the point corresponding the maximum log likelihood over the entire marker interval (Van Arendonk et al., 1994; Grignola et al., 1996a)

Restricted maximum likelihood (REML, by Patterson and Thompson, 1971) has become the method of choice for estimating the variance components in animal breeding. The first attempt for estimating position and variance contribution of a single QTL together with additive polygenic and residual variance components by REML was undertaken by Van Arendonk et al. (1994) with a single marker. Grignola et al., (1996a) extended this method to multiple markers.

Sampling Strategy

Most of the studies of detection of QTL and marker assisted selection assumed that individual sampled are randomly chosen from population. This assumption is seldom true. In detection of QTL effects linked to markers, individual with extreme phenotypic values are selected for genotyping to increase the power of test. In most livestock species, data must be obtained from existing commercial populations. Usually, such populations have been selected for many generations toward a desired breeding goal. The impacts of selection on power and estimates of parameters of QTL effects are discussed in following section.

Random Selection

Random selection assumed that individuals are randomly chosen from generation to generation. Sampled individuals share the same gene pool with base population.

Disruptive Selection

To increase power of detecting marker associated QTL effects, animals for genotyping were selected with extreme phenotypic values. The statistical power of selective genotyping for the purpose of detecting linkage between QTL and markers was investigated by Lander and Botstein (1989); Darvasi and Soller

(1992). It was found that power to detect a QTL effect is increased by selectively genotyping individuals with extreme values for the quantitative trait (Weller and Wyler, 1992; Lin and Ritland, 1996).

Mackinnon and Georges (1992) found that selection by truncation for the trait of interest significantly reduces the genetic variance of the trait, thus reducing the power to detect linked quantitative trait loci.

In the absence of selection, the estimates of QTL location, variance due to QTL and polygenic effects, and residual variance are unbiased by using REML (Grignola et al., 1996b). In another simulation study, bias of the estimated QTL effects was less than 2% in the absence of selection (Meuwissen and Goddard, 1997). The selection bias using BLUP based methods are needed to be investigated for polygenic and marker linked QTL mixed model.

Truncation Selection

Theory indicates the effects of selection can be accommodated by an appropriate model that includes all data upon which selection decisions were based, tracing back to the unselected base generation (Henderson, 1975b; Gianola et al., 1986). Sorensen and Kennedy (1984) simulated several generations of selection and omitted data from earlier generations. They concluded that the estimate of the additive genetic variance before selection was nearly unbiased when their model acknowledged all relationships that developed in previous generations.

However, results in a similar study, Van der Werf and De Boer (1990) showed that there were small biases in estimates of additive genetic variance of the base unselected generation in some cases, even with all relationships tracing back to base generation.

Impact of selection on the estimates under additive polygenic and marker linked QTL mixed models were also carried out for ANOVA, regression and maximum likelihood. Biases were found for both the population under going selection and population with selective genotyping. The bias is a function of selection intensity and the magnitude of QTL effects. Lin and Ritland (1996) showed selective genotyping can bias estimates of the recombination frequency between linked QTL. The QTL effects linked to markers tend to be overestimated increasingly with decreasing family size and true QTL effect (Georges et al., 1995). Bias of the estimate of linkage increased when parents were not a random sample from a population in linkage equilibrium (Uimari et al., 1996).

Stabilizing Selection

Stabilizing is applied when standard production of livestock is required. The individuals with large or small observation are excluded. Genetic homogeneous is desirable.

Marker Assisted Selection

The usage of linkage between genetic marker and quantitative trait loci was shown earliest by Sax (1923). However, the low number of genetic markers available has been a limiting factor for application of marker assisted selection until the recent discovery of molecular technology.

Genetic Markers

The first molecular markers used were allozymes, protein variants detected by differences in migration on starch gels in an electric field. Since 1960s, this class of markers has been extensively applied to population genetic problems. This approach was largely replaced by the method of evaluating variation directly at DNA level in mid of 1980s. A simplest approach is to digest DNA with a variety of restriction enzymes, each of which cuts the DNA at specific sequence. When the digested DNA is run on a gel under an electric current the fragment separated out according to size. The individual bands can be isolated by using labeled DNA probes that base pair complementarily to particular region of genome. This approach is called restriction fragment length polymorphisms (RFLPs). Each RFLP probe generally scores a single marker locus. The marker alleles are codominant. Heterozygotes and homozygotes can be distinguished.

Another approach is uses short primers for DNA replication via the polymerase chain reaction (PCR) to delimit fragment sizes. The fragment flanked

by primers is amplified. The primers are random short sequence. This approach is called randomly amplified polymorphic DNAs (RAPDs). RAPDs have advantages over RFLPs in that a single probe can reveal several loci once. They also require small amount of DNA.

The number of genetic markers has increased rapidly in recent years. The advent of techniques to detect molecular variation, beginning with protein electrophoresis and genetic polymorphism at level of DNA, has increased the number of markers now. These markers include multisite restriction fragment length polymorphism (RFLP) haplotypes, variable number of tandem repeat (VNTR) sequences, and polymerase chain reaction (PCR) based polymorphism (Mullis, 1990). The huge amount of genetic variability revealed by these techniques in many agriculturally important species has allowed the construction of detailed genetic maps evenly spaced throughout the genome (Paterson et al., 1988; Visscher et al., 1990; Soller and Beckman, 1990; Georges et al., 1995). Polymorphism at DNA level makes it possible to identify genotypic differences among individuals at many genomic sites. The use of information on markers is expected to accelerate genetic progress through increasing accuracy of selection, reduction of generation interval and increased selection differentials (Kashi et al., 1990). The advantages and limitations of marker assisted selection are described in following.

Advantage of MAS

Regression and BLUP are currently methods used to integrate phenotype with marker information. Lande and Thompson (1990) proposed an index incorporating the phenotype value and molecular score of individuals. The molecular score is computed from the effects attributed to markers by multiple regression of phenotype on marker genotype. BLUP, incorporating marker information, was described by Fernando and Grossman (1989). The total additive genetic value composed of the prediction of the sum of two QTL alleles linked to marker and polygene effect at remanding loci unlinked to markers.

The efficiencies were compared between purely phenotype selection and marker assisted selection (Lande and Thompson, 1990; Gimelfarb and Lande, 1994; Ruane and Colleau, 1996; Hospital and Moreau, 1997). Zhang and Smith (1992 and 1993) conducted similar simulations of marker assisted selection. However, they compared MAS not with purely phenotypic selection but rather to selection based on the BLUP estimate of an individual's breeding value. The common conclusion was that marker assisted selection was more efficient than selection without using marker information in 1) early generations; 2) low heritability traits; 3) large populations; 4) close linkage between marker and QTL.

Marker assisted selection can still be effective in populations that have been highly selected for many generations on phenotype or predicted breeding values. Georges et al. (1995) demonstrate that loci with considerable effects on milk production are still segregating in highly selected populations.

Limitations of MAS

First of all, the limitation of MAS comes from the nature of genetic composition of traits. The genetic variance may be controlled by many genes with small effects even though there may be a few loci with large effects (Shrimpton and Robertson, 1988). Traditional methods will be effective to identify the effect of these genes. Further more, due to arising of variance within populations by high mutability of polygene, traditional methods is effective to improve whatever previous gains have been made (Hill, 1982).

Another factor limiting efficiency of MAS is the error of QTL parameter estimates. The QTL parameters in MAS studies have assumed to be known without error when genetic and economic responses to MAS were estimated. However, this situation is not generally true. The variance associated QTL effects was overestimated when the analysis had low power (Wang et al., 1995).

Deduction on selection response can result from the error of the estimates of QTL parameters. For a simulated error of 15 cM on the location of QTL, genetic superiority of MAS was reduced by 80% in the first generation. Zhang and Smith (1993) showed that poorly estimated QTL effects added noise to the system and reduced selection response in marker assisted selection.

Chapter 2

METHODOLOGY

Data is generated given specified underlying parameters for a marked QTL and polygene mixed model. The estimates of the parameters are obtained by analyzing the simulated data using REML. The design of the breeding populations simulated, the simulation of genotypes on markers and QTLs, simulation of phenotypes, selection schemes, and the statistical model for analysis are described below.

Population

The granddaughter design, together with daughter design was proposed for the purpose of identifying genetic markers associated with QTL (Weller, 1990). Granddaughter design requires the genotyping of sires and their sons at polymorphic loci and the recording of phenotypes on granddaughters. There are two generations involved in daughter design: sires and daughters. Phenotyping is on daughters. Genotyping is on both sires and daughters (see figure 1).

One of the advantages of granddaughter design over daughter design is that it may be easier to collect blood or semen samples from sons of sires,

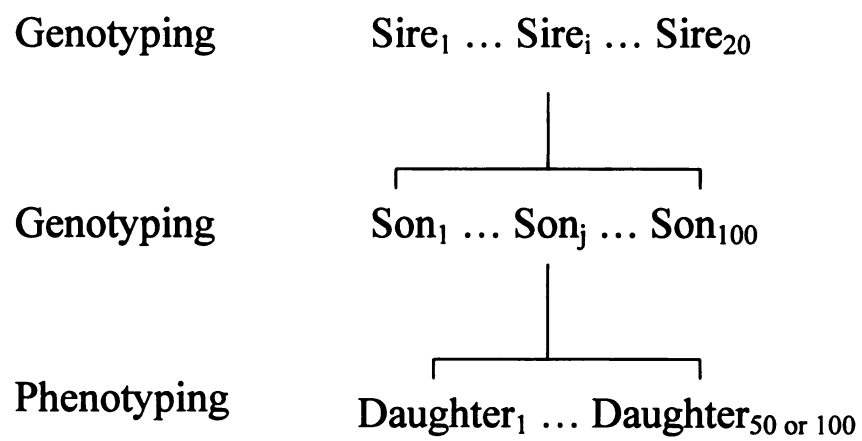


Figure 1 - Population Structure of Granddaughter Design

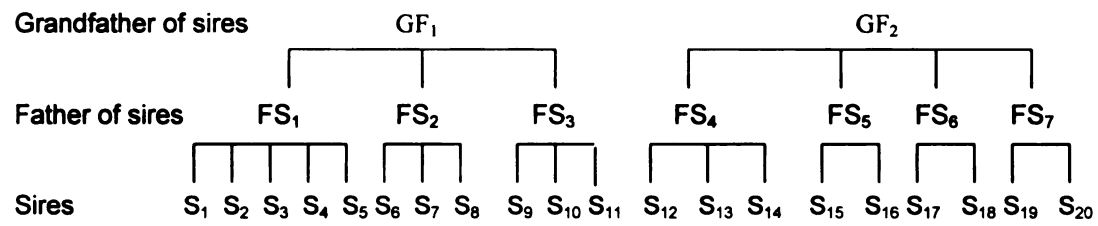


Figure 2 - Pedigree of 20 Sires in Simulated Granddaughter Design Population

concentrated in AI centers, than from their daughters, scattered over many farms. In this study, granddaughter design populations are simulated. Each population consisted of 20 sires from 9 different grandsires, just as Grignola et al (1996b). Each sire had 100 sons for a total of 2000 sons. The number of granddaughters per son is set to either 50 or 100 (see figure 2).

Simulation of Genotypes

Five equally spaced marker loci on an autosomal chromosome are simulated (see Figure 3). There are five alleles with equal frequency (20%) on each locus. A single QTL is set to the midway between the third and fourth marker loci. There are two alleles at the QTL. The gene frequencies of the two alleles are set to equal (50%). The effect of one allele is set to α , and $-\alpha$ for the other allele. There is no dominant effect. Therefore, the additive genetic variance due to QTL is $2\alpha^2$.

Simulation of DYDs

Instead of individual yield on each daughter of son (granddaughter), the daughter's yield deviation (DYD) of is used for analysis. The values of DYDs can be generated from each daughters' yield or generated for each son directly. A model with a mixture of polygenic effects and the allelic effect due to the QTL

linked to marker is defined for the phenotypic observations on granddaughters below. The trait that is recorded is called “daughter yield deviation”.

$$DYD_i = \frac{1}{N_d} \sum_{j=1}^{N_d} (v_{ij}^1 + v_{ij}^2) + 0.5u_i + \varepsilon_i$$

where DYD_i is the evaluation on son i ; N_d is the number of daughters per son; v_{ij}^1 was the additive genetic effect of one allele at QTL linked to genetic markers of granddaughter j of son i ; v_{ij}^2 was the additive genetic effect of the other allele at QTL of granddaughter j of son i ; u_i was the additive genetic effect of the polygene unlinked to genetic markers of son i ; and ε_i was the random residual effect corresponding to DYD_i . The variances of v_{ij}^1 and v_{ij}^2 were both assumed to be σ_v^2 , which was the additive genetic variance due to the QTL allele linked to genetic markers. Therefore, the total additive genetic variance was $\sigma_a^2 = 2\sigma_v^2 + \sigma_u^2$.

The u_i was generated from a normal distribution with zero mean and a (co) variance matrix of $I\sigma_u^2$ with σ_u^2 being the additive genetic variance due to polygenic effects unlinked to genetic markers. The residual effect (ε_i) was generated from normal distribution with mean of zero and variance of

$$\frac{1}{N_d} (0.75\sigma_a^2 + \sigma_e^2), \text{ where } \sigma_e^2 \text{ is the environmental variance of the}$$

granddaughters' yield. All covariances between v_{ij}^1 , v_{ij}^2 , u_i , and ε_i were assumed

to be zero. The heritability of the yield trait was defined as $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$.

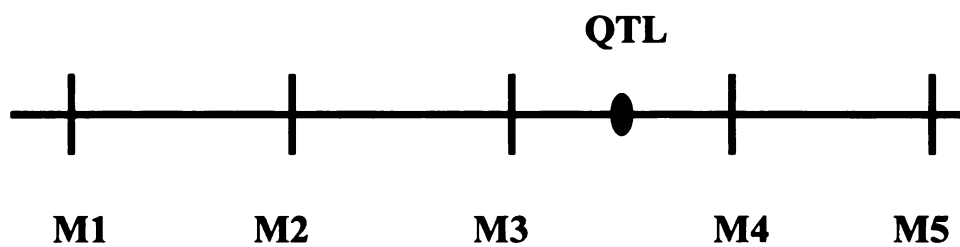


Figure 3 - Genetic Marker Loci and One QTL on an Autosomal Chromosome.

Five marker loci are equally spaced. A single QTL is at the midway between the third and fourth marker loci.

The QTL and marker alleles were generated in linkage equilibrium across base population animals and each offspring inherited a QTL-marker haplotype subject to chance recombination between the loci.

Five marker loci on an autosomal chromosome with equally spaced marker intervals were simulated. A biallelic QTL with equal frequencies (.5) was located midway between the third and fourth marker loci. This location assignment associated higher power of test than the midway location between the fourth and fifth. The two QTL allelic effects were set to α and $-\alpha$, respectively, i.e., gene action was assumed to be additive. Therefore, the variance due to QTL ($2\sigma_v^2$) was $2\alpha^2$, and thus the value of α was set to be the square root of σ_v^2 .

The magnitude of QTL effects was denoted by the ratio of QTL allelic variance to the total additive genetic variance:

$$v^2 = \frac{\sigma_v^2}{\sigma_a^2}.$$

The maximum value of v^2 is .5, in which case, σ_u^2 equals to zero.

Selection Schemes

The selection schemes refer to alternative strategies of sampling phenotyped individuals for genotyping. The idea behind selective genotyping is that scoring characters is often much less expensive than scoring genetic

markers. Hence, there may be merit in choosing a subset of phenotyped individuals for genotyping.

Most of the studies assumed that genotype individual are randomly chosen from population. However, the situation is not true in commercial populations. Usually, such populations have been selected for many generations toward a desired breeding goal. The selection could be truncated in most cases or stabilizing occasionally. The typical selection scheme for detection of QTL is disruptive, in which the uppermost and lower most fractions of scored individuals are genotyped.

The 100 sons of each sire in granddaughter design in this study are selected on their daughter yield deviations (DYD) (Van Raden and Wiggans, 1991) according to the following alternative selection schemes:

1. Random selection: Sons are randomly chosen from the population;
2. Disruptive selection: Those sons with their DYD away from the population mean by one standard deviation (SD) in either direction are chosen;
3. Truncation selection: Sons with DYD greater than the mean are chosen.
4. Stabilizing selection: Sons with DYD within the range of one positive and negative SD from the mean are chosen.

Statistical Analysis

A reduced animal model including QTL effects linked to markers and polygenic effects was used to analyze the simulated data:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{T}_u\mathbf{u} + \mathbf{Z}\mathbf{D}\mathbf{T}_v\mathbf{v} + \mathbf{e}$$

where $\mathbf{y}$ was an $N \times 1$ vector of DYD evaluations on sons with N being the number of DYDs, which was equal to the number of sons; $\boldsymbol{\beta}$ was a vector of fixed effects; $\mathbf{X}$ was the design matrix relating $\boldsymbol{\beta}$ to $\mathbf{y}$; $\mathbf{u}$ was an $N_p \times 1$ vector of the polygenic effects unlinked to the markers with N_p being the number of sires; $\mathbf{Z}$ was the incidence matrix relating elements in $\mathbf{y}$ to sons; $\mathbf{T}_u$ was a transformation matrix relating sons to sires for polygenic effect; $\mathbf{v}$ was a $2N_p \times 1$ vector of the QTL allelic effects linked to the markers; $\mathbf{D}$ was the incidence matrix relating each animal to its two QTL alleles; $\mathbf{T}_v$ was the transformation matrix relating sons to sires for QTL effect; and $\mathbf{e}$ was the vector of residual effects. The random effects were assumed to follow a normal distribution with a (co)variance structure of

$$\text{Variance} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_u \sigma_u^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_v \sigma_v^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix}$$

where $\mathbf{A}_u$ was the numerator additive polygenic relationship matrix (Henderson, 1975); σ_u^2 was the additive polygenic variance; $\mathbf{A}_v$ was the relationship matrix of QTL effects linked to markers; σ_v^2 was the QTL allelic variance; and

$$\mathbf{R} = \mathbf{I} \sigma_e^2 + \Lambda_u \sigma_u^2 + \Lambda_v \sigma_v^2$$

where $\mathbf{I}$ is an identity matrix of $N \times N$, Λ_u is the correlation matrix of Mendelian polygenic effects and Λ_v is the correlation matrix of Mendelian QTL effects. The theory of building $\mathbf{A}_v$ and Λ_v was presented by Wang et al. (1995).

An algorithm of REML by Grignola et al. (1996a) was adapted for the estimation of variance components in this study. The analysis was conducted at a number of successive positions along a chromosome. Then, the likelihood was maximized with respect to σ_u^2 , σ_v^2 , σ_e^2 and r (recombination rate) at each position. The estimated location of QTL was determined by the largest likelihood value over the grid of possible QTL locations.

Experimental Design

The objective of this study is to investigate the effects of different selection schemes, lengths of marker interval (r), magnitudes of QTL effects (v^2), heritability levels (h^2), and number of daughters (n_d) in a granddaughter design on the estimates of QTL location and variance components of QTL, polygenic and residual effects. A sub-population is defined by each combination of r (10 or 30 cM), v^2 (.125 or .25), h^2 (.05 or .4), n_d (50 or 100) and selection schemes (random, disruptive, truncation and stabilizing), for a total of 64 combinations. For each sub-population, fifty replicates are generated and are analyzed separately. Phenotypic variance ($\sigma_a^2 + \sigma_e^2$) is set to 10,000. Bias of estimation is defined as the difference between true parameter value and its estimate. The bias rate is defined as the ratio of bias to the true value of parameter. The median bias rate from 50 replicates is used as the measurement for each sub-population. Power on each combination is estimated by the proportion of significant results over the total replicates.

Chapter 3

BIASES IN GENETIC PARAMETER ESTIMATION FROM A POLYGENE AND QTL MIXED MODEL UNDER GRANDDAUGHTER DESIGN

Abstract

Statistical biases in the estimation of quantitative trait loci (QTL) location, additive genetic variance due to markers linked to QTL, additive genetic variance due to polygene unlinked to markers, and residual variance were evaluated in a simulation study. Genetic populations with a granddaughter design were simulated under a mixed model with polygenic effects and QTL effects for various combinations of selection schemes, marker intervals, numbers of daughters, magnitudes of QTL effects and heritability levels with respect to a single trait. For each of the combinations, 50 replicate populations were generated and analyzed separately by a restricted maximum likelihood algorithm.

Estimates of QTL location, variance due to segregating QTL, and polygenic and residual effects were all unbiased in unselected populations. Estimates of QTL location were unbiased not only in unselected populations, but also in populations under various selection schemes. However, estimates of variances due to QTL, polygenic and residual effects were biased significantly in

populations under nonrandom selection schemes. The magnitudes of the biases were dependent on marker intervals, numbers of daughters, magnitudes of QTL effect and heritability levels.

Introduction

The advantages of marker assisted selection (MAS) in breeding schemes have been discussed rigorously recently (Lande and Thompson, 1990; Gimelfarb and Lande, 1995; Hospital and Moreau, 1997; Spelman and Garrick, 1997). However, when genetic and economic responses to MAS were estimated in these studies, genetic parameters were assumed known without error. Simulation studies (Zhang and Smith, 1993 and Wang *et al.*, 1995) have demonstrated that the overestimation of variance due to QTL effects would decrease estimated long term genetic gain under MAS. For an error of only 15 cM in the location of QTL, estimated genetic superiority of MAS would be reduced by 80% after the first generation of selection (Spelman and Van Arendonk, 1997).

The major difference between the classical polygenic animal model and a mixture model of marker-associated QTL effects and polygenic effects is the additional (co)variance matrix of QTL effects linked to markers. The algorithm to construct this matrix for a single marker was presented by Fernando and Grossman (1989), with multiple marker extensions provided by Goddard (1991). Wang *et al.* (1995) proposed an algorithm for the case of incomplete information of the origin of marker genes, which was required in the algorithm by Fernando.

Variances of single marker associated QTL effects, additive polygenic effects, and residual effects can be estimated by restricted maximum likelihood (REML) as shown by Van Arendonk *et al.* (1994). Grignola *et al.* (1996a) expanded their method for multiple marker associations.

The accuracy and precision of the estimates using REML have been evaluated in many studies for the polygenic model in unselected and selected populations (Henderson, 1975a; Banks *et al.*, 1985; Gianola *et al.*, 1986; Beaumont, 1991; Sorensen and Kennedy, 1984). Grignola *et al.* (1996b) evaluated the accuracy and precision of the estimates using REML in a single marked QTL and polygenic mixed model in unselected populations.

In most livestock species, data for linkage analysis must be obtained from existing commercial populations. Usually, such populations had been selected for many generations toward a desired breeding goal (Vukasinovic *et al.*, 1998). One sampling strategy that has been advocated to increase power of marker associated QTL effect detection, given limited resources, is to select animals with extreme phenotypic values for genotyping. However, the frequency of favorable alleles and genetic variances are plausibly changed under selection (Mackinnon and Georges, 1992; Keightley and Bulfield, 1993), thereby potentially causing large biases in estimates of QTL effects.

Granddaughter design (GDD) proposed by Weller in 1990, together with daughter design, are two typical designs proposed for the purpose of identifying genetic markers associated with QTL. A GDD requires the genotyping of grandsires and their sons at polymorphic loci and the recording of phenotypes on

granddaughters. One practical advantage of GDD over daughter design is that it may be easier to collect blood or semen samples from sires and sons which are normally concentrated in AI centers than from their daughters which are normally scattered over many commercial farms.

The objective of this study was to investigate the effects of different selection schemes, lengths of marker interval, magnitudes of QTL effects, heritability levels, and number of daughters in a granddaughter design on the estimation of QTL location and variance components of QTL, polygenic and residual effects.

Materials and Methods

Data was generated given specified underlying parameters for a marked QTL and polygene mixed model. The estimates of the parameters were obtained by analyzing the simulated data using an algorithm of REML. The design of the breeding populations simulated, the genetic model for simulation, and the statistical model for analysis are described below.

Population

Each simulated GDD population consisted of 20 sires from 9 different grandsires as in the study by Grignola et al (1996b). Each sire had 100 sons for a total of 2000 sons. The number of granddaughters per son was set to either 50 or 100. The 100 sons of each sire were selected based on their daughter yield

deviations (DYD) (Van Raden and Wiggans, 1991) according to the following alternative selection schemes:

1. Random selection: Sons were randomly chosen from the population;
2. Disruptive selection: Those sons with their DYD away from the population mean by one standard deviation (SD) in either direction were chosen;
3. Truncation selection: Sons with DYD greater than the mean were chosen.
4. Stabilizing selection: Sons with DYD within the range of one positive and negative SD from the mean were chosen.

Genetic Model

A model with a mixture of polygenic effects and the allelic effect due to the QTL linked to marker was defined for the purpose of evaluation of the sons based on the phenotypic observations on granddaughters. The evaluation criterion was called daughter yield deviation, or DYD:

$$DYD_i = \frac{1}{N_d} \sum_{j=1}^{N_d} (v_{ij}^1 + v_{ij}^2) + 0.5u_i + \varepsilon_i$$

where DYD_i is the evaluation on son i ; N_d is the number of daughters per son; v_{ij}^1 was the additive genetic effect of one allele at QTL linked to genetic markers of granddaughter j of son i ; v_{ij}^2 was the additive genetic effect of the other allele at QTL of granddaughter j of son i ; u_i was the additive genetic effect of the polygene unlinked to genetic markers of son i ; and ε_i was the random residual effect corresponding to DYD_i . The variances of v_{ij}^1 and v_{ij}^2 were both assumed

to be σ_v^2 , which was the additive genetic variance due to the QTL allele linked to genetic markers. Therefore, the total additive genetic variance was $\sigma_a^2 = 2\sigma_v^2 + \sigma_u^2$.

The u_i was generated from a normal distribution with zero mean and a (co)variance matrix of $I\sigma_u^2$ with σ_u^2 being the additive genetic variance due to polygenic effects unlinked to genetic markers. The residual effect (ε_i) was generated from normal distribution with mean of zero and variance of

$$\frac{1}{N_d}(0.75\sigma_a^2 + \sigma_e^2), \text{ where } \sigma_e^2 \text{ is the environmental variance of the}$$

granddaughters' yield. All covariances between v_{ij}^1 , v_{ij}^2 , u_i , and ε_i were assumed

to be zero. The heritability of the yield trait was defined as $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$.

The QTL and marker alleles were generated in linkage equilibrium across base population animals and each offspring inherited a QTL-marker haplotype subject to chance recombination between the loci.

Five marker loci on an autosomal chromosome with equally spaced marker intervals were simulated. A biallelic QTL with equal frequencies (.5) was located midway between the third and fourth marker loci. This location assignment associated higher power of test than the midway location between the fourth and fifth. The two QTL allelic effects were set to α and $-\alpha$, respectively, i.e., gene action was assumed to be additive. Therefore, the variance due to QTL ($2\sigma_v^2$) was $2\alpha^2$, and thus the value of α was set to be the square root of σ_v^2 .

The magnitude of QTL effects was denoted by the ratio of QTL allelic variance to the total additive genetic variance:

$$v^2 = \frac{\sigma_v^2}{\sigma_a^2}.$$

The maximum value of v^2 is .5, in which case, σ_u^2 equals to zero.

Statistical Model

A reduced animal model including QTL effects linked to markers and polygenic effects was used to analyze the simulated data:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{T}_u\mathbf{u} + \mathbf{Z}\mathbf{D}\mathbf{T}_v\mathbf{v} + \mathbf{e}$$

where $\mathbf{y}$ was an $N \times 1$ vector of DYD evaluations on sons with N being the number of DYDs, which was equal to the number of sons; $\boldsymbol{\beta}$ was a vector of fixed effects; $\mathbf{X}$ was the design matrix relating $\boldsymbol{\beta}$ to $\mathbf{y}$; $\mathbf{u}$ was an $N_p \times 1$ vector of the polygenic effects unlinked to the markers with N_p being the number of sires; $\mathbf{Z}$ was the incidence matrix relating elements in $\mathbf{y}$ to sons; $\mathbf{T}_u$ was a transformation matrix relating sons to sires for polygenic effect; $\mathbf{v}$ was a $2N_p \times 1$ vector of the QTL allelic effects linked to the markers; $\mathbf{D}$ was the incidence matrix relating each animal to its two QTL alleles; $\mathbf{T}_v$ was the transformation matrix relating sons to sires for QTL effect; and $\mathbf{e}$ was the vector of residual effects. The random effects were assumed to follow a normal distribution with a (co)variance structure of

$$\text{Variance} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_u \sigma_u^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_v \sigma_v^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix}$$

where $\mathbf{A}_u$ was the numerator additive polygenic relationship matrix (Henderson, 1975); σ_u^2 was the additive polygenic variance; $\mathbf{A}_v$ was the relationship matrix of QTL effects linked to markers; σ_v^2 was the QTL allelic variance; and

$$\mathbf{R} = \mathbf{I} \sigma_e^2 + \Lambda_u \sigma_u^2 + \Lambda_v \sigma_v^2$$

where $\mathbf{I}$ is an identity matrix of $N \times N$, Λ_u is the correlation matrix of Mendelian polygenic effects and Λ_v is the correlation matrix of Mendelian QTL effects. The theory of building $\mathbf{A}_v$ and Λ_v was presented by Wang *et al.* (1995).

An algorithm of REML by Grignola *et al.* (1996a) was adapted for the estimation of variance components in this study. The analysis was conducted at a number of successive positions along a chromosome. Then, the likelihood was maximized with respect to σ_u^2 , σ_v^2 , σ_e^2 and r (recombination rate) at each position. The estimated location of QTL was determined by the largest likelihood value over the grid of possible QTL locations.

Design

A sub-population was defined by each combination of marker density (10 or 30 cM), v^2 (.125 or .25), h^2 (.05 or .4), number of daughters (50 or 100) and selection schemes (random, disruptive, truncation and stabilizing), for a total of 64 combinations. For each sub-population, fifty replicates were generated and

were analyzed separately. Phenotypic variance ($\sigma_a^2 + \sigma_e^2$) was set to 10,000.

Bias of estimation was defined as the difference between true parameter value and its estimate. The bias rate was defined as the ratio of bias to the true value of parameter. The median bias rate from 50 replicates was used to assess the relative degree of bias for each parameter within each sub-population.

Results and Discussions

QTL Location

Mackinnon and Weller (1995) showed that maximum likelihood estimation of recombination rate was inaccurate compared to estimation of other parameters. Accuracy of estimation can be improved by using interval mapping (Knott and Hally, 1992). Grignola *et al.* (1996b) showed that the estimates of QTL location were unbiased using REML within an interval mapping framework. Spelman and Van Arendonk (1997) showed that genetic gain by marker assisted selection with 5 cM error was significantly less than that achieved when the QTL position was estimated correctly. In the framework of our study, estimates of the location of QTL were unbiased in all sub-populations, and selection schemes did not affect the estimate of QTL location. The median biases in all sub-populations ranged from -1.3 to 1.1 cM and were not significantly different from zero ($P>0.05$).

Figure 4 - Estimates of Polygenic, QTL and Residual Variances

The estimates were from populations with various combinations of underlying heritability (h^2) levels, magnitudes of QTL effect (v^2), numbers of daughters (N_d), and lengths of marker interval (in cM) when undergone alternative selection schemes. L denotes the populations with .05 for h^2 , .125 for v^2 , and 50 for N_d ; H denotes the populations with .4 for h^2 , .25 for v^2 , and 100 for N_d .

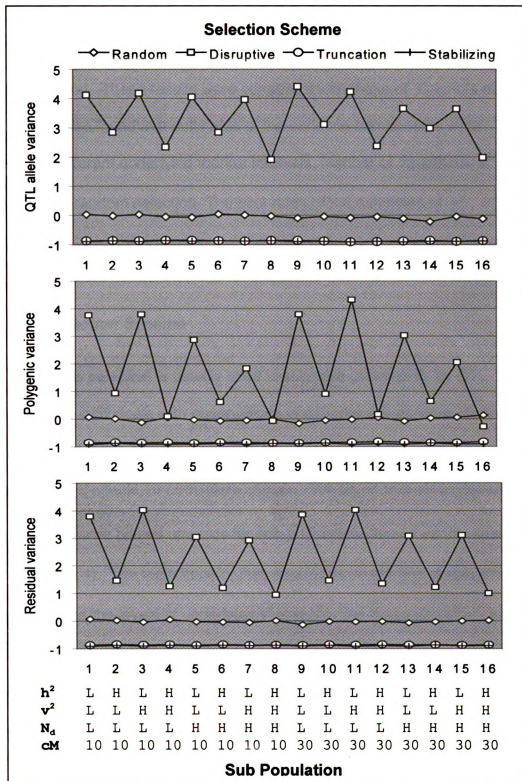


Figure 4 - Estimates of Polygenic, QTL and Residual Variances

Random Selection Scheme

Unbiased estimates of σ_v^2 , σ_u^2 and σ_e^2 were obtained under the random selection scheme. These results were consistent with those of Grignola et al. (1996b), who also found that estimates of the variance components were unbiased for different underlying marker intervals, number of daughters, v^2 and h^2 levels under random selection. The bias rates of the estimates of σ_v^2 , σ_u^2 and σ_e^2 in each sub-population were not significantly different from zero ($P>0.05$).

Disruptive Selection Scheme

Disruptive selection is usually used in selective genotyping for QTL mapping where individuals with observations of extreme values were chosen for genotyping. Under this selection scheme, those individuals with deviant polygenic or QTL effects, or both, were chosen. We found that estimates of σ_v^2 and σ_u^2 tended to be biased upwards for sub-populations based on this selection protocol (Figure 4). The magnitude of bias depended on number of daughters per son (N_d), and the levels of underlying v^2 and h^2 . The bias rates of σ_v^2 and σ_u^2 decreased significantly with more daughters and higher v^2 and h^2 values (Table 1 and 2). An increase in the number of daughters from 50 to 100 led to a significant reduction of bias in the estimates of σ_u^2 ($p<0.01$), but not for σ_v^2 ($p>0.05$).

Residual variance estimates could be reduced due to a decrease in heterozygosity of QTL. For the mating of sire (QQ) with dam (qq) for example,

Table 1 - Median* and SE of Bias Rates (%) in the Estimation of QTL Allelic Variance. The estimates were from populations subjected to disruptive selection where the populations differed in different levels of heritability (h^2) and proportions of additive genetic variance due to QTL effect (v^2).

v^2	h^2	Marker interval (cM)							
		10				30			
		No. of daughters				No. of daughters			
		50		100		50		100	
.125	.05	412 ^a	± 20	405 ^{bac}	± 21	441 ^a	± 28	365 ^{bdac}	± 29
	.4	284 ^{dec}	± 20	284 ^{dec}	± 16	311 ^{bdec}	± 18	298 ^{de}	± 19
.25	.05	418 ^a	± 23	396 ^{ba}	± 17	423 ^a	± 20	364 ^{bac}	± 20
	.4	234 ^{fe}	± 8	19 ^f	± 7	237 ^{fe}	± 13	198 ^f	± 11

*Medians with the same superscripts were not different significantly ($P>0.05$).

Table 2 - Median* and SE of Bias Rates (%) in the Estimation of Polygenic Variance. The estimates were from populations subjected to disruptive selection where the populations differed in different levels of heritability (h^2) and proportions of additive genetic variance due to QTL effect (v^2).

v^2	h^2	Marker interval (cM)							
		10				30			
		Number of daughters				Number of daughters			
		50		100		50		100	
.125	.05	376 ^a	± 23	286 ^{ba}	± 13	379 ^a	± 24	302 ^{ba}	± 16
	.4	93 ^c	± 6	61 ^{dc}	± 4	9 ^c	± 6	64 ^c	± 5
.25	.05	379 ^a	± 35	182 ^b	± 2	432 ^a	± 3	204 ^b	± 21
	.4	9 ^e	± 6	-7 ^e	± 6	16 ^{de}	± 1	-28 ^e	± 9

*Medians with the same superscripts were not different significantly ($P > 0.05$).

Table 3 - Median* and SE of Bias Rates (%) in the Estimation of Residual Variances. The estimates were from populations subjected to disruptive selection where the populations differed in different levels of heritability (h^2) and proportions of additive genetic variance due to QTL effect (v^2).

v ²	h ²	Marker interval (cM)											
		10						30					
		Number of daughters						Number of daughters					
		50			100			50		100			
.125	.05	379 ^a	±	15	304 ^b	±	9	385 ^a	±	18	308 ^b	±	9
	.4	146 ^c	±	3	119 ^d	±	3	146 ^c	±	3	122 ^d	±	3
.25	.05	402 ^a	±	14	291 ^b	±	8	402 ^a	±	13	310 ^b	±	8
	.4	125 ^d	±	3	94 ^e	±	3	135 ^{dc}	±	4	100 ^e	±	3

*Medians with the same superscripts were not different significantly ($P>0.05$).

the probability of tracing “Q” of offspring back to sire is 100% if no recombination exists between QTL and markers. However, the probability reduces to 50% if the genotypes of sire and dam are both “Qq”. As the proportion of homozygotes with high or low genotypic values was increased under disruptive selection, the certainty of knowing the identity by descent of markers in parent to an offspring was increased. Johnson (1992) and Wright (1991) showed that estimates of residual variance were greater with relationship coefficients set not equal to zero than set to zero. Estimates of σ_e^2 in this study showed an upward bias under disruptive selection. The bias rates of σ_e^2 estimates were smaller with greater number of daughters, v^2 and h^2 values (Table 3).

Truncation Selection Scheme

Individuals selected under truncation selection scheme were those with large genetic effects. Mackinnon and Georges (1992) showed that selection may lead to underestimation of QTL effects. This study found that estimates of σ_v^2 and σ_u^2 were generally biased downwards ($p < 0.01$) in populations that underwent truncation selection, but they were not influenced significantly by daughter numbers, marker interval length, and v^2 and h^2 levels.

The situation of estimation biases in σ_e^2 under truncation selection is opposite to that under disruptive selection. The increased uncertainty of tracing QTL from parents to progeny lead to a decrease in the relationship coefficients on QTL effects, and thus an underestimation of residual variance. This was

because the homozygotes with large effects became a larger portion of the population under the truncation selection. Estimates of σ_e^2 were generally biased downwards ($p < 0.01$) in populations that underwent truncation selection, but also they were not influenced significantly by daughter numbers, marker interval length, and v^2 and h^2 levels.

Stabilizing Selection Scheme

The individuals with either large or small QTL or polygene effects are excluded under stabilizing selection, in which individuals within the range of a defined SD from the mean were chosen. Polygene and QTL variances from populations that underwent stabilizing selection appeared to be underestimated in this study.

The individuals favored under stabilizing selection scheme were heterozygotes at QTL. The certainty of knowing the parent-offspring relationship with respect to a genetic marker was reduced from matings between heterozygous individuals. The result was similar to matings between homozygotes with large effects under the truncation selection scheme. This may help explain why no difference was found ($p > 0.05$) between populations underwent stabilizing selection and those undergoing truncation selection in estimates of σ_v^2 , σ_u^2 and σ_e^2 .

Conclusions and Implications

By using the REML method, QTL locations can be estimated without significant bias in populations under a wide array of sampling strategies. However, there were significant biases in estimates of variances due to QTL, polygene and residual effects from populations subjected to selection schemes except from those subject to only random selection. Variances due to QTL, polygene and residual effects were overestimated from populations subjected to disruptive selection, but were underestimated from populations subjected to truncation and stabilizing selection schemes.

The biases in variance estimates for QTL, polygene and residual effects declined with more daughters per sire, higher levels of underlying heritability and greater proportion of additive genetic variance that was due to QTL effects.

This study confirmed the results by Grignola *et al.* (1996b) that QTL location, variances of QTL, polygenic and residual effects could be unbiasedly estimated from populations that did not undergo selection. Furthermore, estimates of QTL location from populations undergone disruptive, truncation or stabilizing selection schemes were also unbiased. From populations undergone the same selection schemes, however, biases were evident in estimates of variances of QTL, polygenic and residual effects. Since biases in variance components can potentially reduce the advantage of marker assisted selection, which is the ultimate application of QTL discovery, estimation biases for

variances of QTL, polygenic and residual effects in various circumstances need to be monitored.

Chapter 4

POWER OF DETECTING MARKER ASSOCIATED QTL EFFECTS IN GRANDDAUGHTER DESIGN

Abstract

Power was examined by simulation for the test of marker linked QTL effect using restricted maximum likelihood. Granddaughter design populations were simulated from a mixed model with polygenic and QTL effects for each combinations of following characteristics: selection scheme, marker interval, number of daughters, magnitude of QTL effect and heritability of trait. Sons in the design were selected by four alternative selection schemes of random, disruptive, truncation and stabilized. Fifty replicates were generated for each population and were analyzed using restricted maximum likelihood. Results indicated that selection scheme has significant influence on the power of testing linkage between genetic markers and QTL using restricted maximum likelihood. Disruptive selection generated higher power than random selection, whereas truncation and stabilizing selection had less power than random selection. Power of test using restricted maximum likelihood also depended on number of daughters of each son, marker interval, magnitude of QTL effect and heritability. Power of test was higher with more daughters per sire, smaller marker interval,

larger magnitude of QTL effect, and higher heritability. The magnitude of the difference due to changing a factor was larger when the power was less saturated by other factors.

Introduction

Several studies have shown that individual loci affecting quantitative traits can be detected via linkage to genetic markers (Soller and Beckman, 1990; Weller and Wyler, 1992). Notable examples were the QTL gene for ovulation rate in swine linked to a marker on chromosome 8 with an additive effect of 3 ova (Rathje et al., 1997), the QTL gene of fat percentage linked to markers clustered on chromosome 4 of pig (Andersson et al., 1994) and the QTL with a significant effect on protein yield in dairy cattle linked to beta-lactoglobulin (Bovenhuis and Weller, 1994)

Identification of quantitative trait loci involves many animals to be genotyped and performance tested. Consequently, experimental designs need to be optimized to minimize the costs of data collection and genotyping for an appropriate power (van der Beek et al., 1995).

Most successful QTL mapping efforts described to date have exploited F_2 or backcrosses obtained from parental populations divergent for the traits of interest (Soller et al., 1976; Paterson et al., 1988). This method results in more power by introducing linkage disequilibrium and can be analyzed by standard software. Linkage analysis for outcross data structure is more complicated and more complex designs and analyses are needed.

ANOVA is the traditional method, which was performed by comparing marker genotype effects (Weller et al., 1990a; Mackinnon and Georges, 1992). The shortage of ANOVA is that no information on QTL location can be provided. Weller (1986) developed a maximum likelihood method to detect marker associated QTL effect. Location of QTL was estimated by maximization of likelihood with respect to recombination rate and other parameters (e.g., mean, additive QTL effect, and within QTL genotype residual variance). Haley and Knott (1992) and Martinez and Curnow (1992) independently introduced a regression method. Regression is performed on the probability of an individual having a QTL genotype, given the genotype for the flanking markers. The QTL location is estimated by the recombination rate with minimum residual sum square. Generally, regression generated similar result as maximum likelihood.

Method of ANOVA, regression and maximum likelihood method were developed mainly for linecross populations. They can not fully account for the more complex data structures in outcross populations, such as data on several families with relationships across families, unknown linkage phases in parents, unknown number of QTL alleles in the population.

For polygenic effects without marker information, BLUP had proved to be a very flexible method. It can handle data with many nongenetic effects, e.g. season, with arbitrary pedigree structure, and with nonrandom mating. Several contributions have been made, which lead to the method of using BLUP for the identification of linkage between markers and QTL.

Fernando and Grossman (1989) showed that information on a single marker can be used in BLUP by fitting the additive effect for alleles at QTL linked to genetic markers and additive polygenic effects for alleles at the remaining quantitative trait loci. Goddard (1991) extended the method to include information from more than one marker. Wang et al. (1995) presented an algorithm for building the relationship matrix with the genetic markers that the information on the origin of the markers is not required. Meuwissen and Goddard (1996) presented a method in which the covariance matrix of effects at the marked QTL was approximated. This approximation reduces the computational requirements.

Restricted maximum likelihood method based on BLUP was presented by Van Arendonk et al. (1994) for a single marker. Grignola et al., 1996a, 1996b) extended this method to multiple markers. Identification of markers associated QTL effect involves maximizing and comparing the likelihood of the data under different genetic models to ascertain the most likely genetic structure. The restricted maximum likelihood of error contrasts under a polygenic model were compared with that under the combined model containing a major gene and polygenic component. A significant improvement in the likelihood obtained by incorporating a major gene in the model provides evidence for a linkage between QTL and genetic markers.

Different statistical methods generated different power of test. The power of test also depends on other factors (e.g. QTL effect, marker interval, sample size etc.). The power of test has been examined for the statistical method of

ANOVA (Weller et al., 1990b), maximum likelihood (Knott and Haley, 1992b; Le-Roy and Elsen, 1995; Carbonell et al., 1993; Jensen, 1989; Lander and Botstein, 1989; Knapp and Bridges, 1990; Elsen et al., 1997) and regression (Moreno-Gonzalez, 1992; Jansen, 1994b; Hyne and Kearsey, 1995; Rebai et al., 1995). Knowledge of power using restricted maximum likelihood remains unknown.

The objective of this study was to examine the power of test using REML in circumstances of combination of different selection scheme, marker interval, magnitude of QTL effect, heritability and number of daughters in granddaughter designs.

Materials and Methods

Data was generated given specified underlying parameters for a marked QTL and polygene mixed model. The estimates of the parameters were obtained by analyzing the simulated data using an algorithm of REML. The design of the breeding populations simulated, the genetic model for simulation, and the statistical model for analysis are described below.

Population

The granddaughter design (GDD) was proposed for the purpose of identifying genetic markers associated with QTL (Weller, 1990). A GDD requires the genotyping of sires and their sons at polymorphic loci and the recording of phenotypes on granddaughters.

In this study, the base population originated from 9 unrelated grandsires that produced 20 sires, just as Grignola et al (1996b). Each sire had 100 sons for a total of 2000 sons. Each son produced 50 or 100 daughters (granddaughters of sires). The dams that produced the 20 sires, 2000 sons and the daughters were assumed to be unrelated. The 100 sons of each sire were selected based on their daughter yield deviations (DYD) (Van Raden and Wiggans, 1991) according to the following alternative selection schemes:

1. Random selection: Sons were randomly chosen from the population;
2. Disruptive selection: Those sons with their DYD away from the population mean by one standard deviation (SD) in either direction were chosen;
3. Truncation selection: Sons with DYD greater than the mean were chosen.
4. Stabilizing selection: Sons with DYD within the range of one positive and negative SD from the mean were chosen.

Genetic Model

A model with a mixture of polygenic effects and the allelic effect due to the QTL linked to marker was defined for the purpose of evaluation of the sons based on the phenotypic observations on granddaughters. The evaluation criterion was called daughter yield deviation, or DYD:

$$DYD_i = \frac{1}{N_d} \sum_{j=1}^{N_d} (v_{ij}^1 + v_{ij}^2) + 0.5u_i + \varepsilon_i$$

where DYD_i is the evaluation on son i ; N_d is the number of daughters per son; v_{ij}^1 was the additive genetic effect of one allele at QTL linked to genetic markers of

granddaughter j of son i ; v_{ij}^2 was the additive genetic effect of the other allele at QTL of granddaughter j of son i ; u_i was the additive genetic effect of the polygene unlinked to genetic markers of son i . The variances of v_{ij}^1 and v_{ij}^2 were both assumed to be σ_v^2 , which was the additive genetic variance due to the QTL allele linked to genetic markers. Therefore, the total additive genetic variance was $\sigma_a^2 = 2\sigma_v^2 + \sigma_u^2$. The u_i was generated from a normal distribution with zero mean and a (co)variance matrix of $I\sigma_u^2$ with σ_u^2 being the additive genetic variance due to polygenic effects unlinked to genetic markers; and ε_i was the residual effect corresponding to $DYD_{i.}$, which was inducted by Mendelian random effect, unknown additive genetic effect of dam and random environmental effect. The residual effect was generated from normal distribution with mean of zero and variance of $\frac{1}{N_d}(0.75\sigma_a^2 + \sigma_e^2)$, where σ_e^2 is the environmental variance of the granddaughters' yield. All covariances between v_{ij}^1 , v_{ij}^2 , u_i , and ε_i were assumed to be zero. The heritability of the yield trait was defined as $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$.

The QTL and marker alleles were generated in linkage equilibrium across base population animals and each offspring inherited a QTL-marker haplotype subject to chance recombination between the loci.

Five marker loci on an autosomal chromosome with equally spaced marker intervals were simulated. A biallelic QTL with equal frequencies (.5) was located midway between the third and fourth marker loci. This location

assignment associated higher power of test than the midway location between the fourth and fifth. The two QTL allelic effects were set to α and $-\alpha$, respectively, i.e., gene action was assumed to be additive. Therefore, the variance due to QTL ($2\sigma_v^2$) was $2\alpha^2$, and thus the value of α was set to be the square root of σ_v^2 .

The magnitude of QTL effects was denoted by the ratio of QTL allelic variance to the total additive genetic variance:

$$v^2 = \frac{\sigma_v^2}{\sigma_a^2}.$$

The maximum value of v^2 is .5, in which case, σ_u^2 equals to zero.

Statistical Model

A reduced animal model including QTL effects linked to markers and polygenic effects was used to analyze the simulated data:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{T}_u\mathbf{u} + \mathbf{ZD}\mathbf{T}_v\mathbf{v} + \mathbf{e}$$

where $\mathbf{y}$ was an $N \times 1$ vector of DYD evaluations on sons with N being the number of DYDs, which was equal to the number of sons; $\boldsymbol{\beta}$ was a vector of fixed effects; $\mathbf{X}$ was the design matrix relating $\boldsymbol{\beta}$ to $\mathbf{y}$; $\mathbf{u}$ was an $N_p \times 1$ vector of the polygenic effects unlinked to the markers with N_p being the number of sires; $\mathbf{Z}$ was the incidence matrix relating elements in $\mathbf{y}$ to sons; $\mathbf{T}_u$ was a transformation matrix relating sons to sires for polygenic effect; $\mathbf{v}$ was a $2N_p \times 1$ vector of the QTL allelic effects linked to the markers; $\mathbf{D}$ was the incidence matrix relating each animal to

its two QTL alleles; T_v was the transformation matrix relating sons to sires for QTL effect; and e was the vector of residual effects. The random effects were assumed to follow a normal distribution with a (co)variance structure of

$$\text{Variance} \begin{bmatrix} u \\ v \\ e \end{bmatrix} = \begin{bmatrix} A_u \sigma_u^2 & 0 & 0 \\ 0 & A_v \sigma_v^2 & 0 \\ 0 & 0 & R \end{bmatrix}$$

where A_u was the numerator additive polygenic relationship matrix (Henderson, 1975); σ_u^2 was the additive polygenic variance; A_v was the relationship matrix of QTL effects linked to markers; σ_v^2 was the QTL allelic variance; and

$$R = I \sigma_e^2 + \Lambda_u \sigma_u^2 + \Lambda_v \sigma_v^2$$

where I is an identity matrix of $N \times N$, Λ_u is the correlation matrix of Mendelian polygenic effects and Λ_v is the correlation matrix of Mendelian QTL effects. The theory of building A_v and Λ_v was presented by Wang *et al.* (1995).

An algorithm of REML by Grignola *et al.* (1996a) was adapted for the estimation of variance components in this study. The analysis was conducted at a number of successive positions along a chromosome. Then, the likelihood was maximized with respect to σ_u^2 , σ_v^2 , σ_e^2 and r (recombination rate) at each position. The estimated location of QTL was determined by the largest likelihood value over the grid of possible QTL locations.

Design

A sub-population was defined by each combination of marker density (10 or 30 cM), v^2 (.125 or .25), h^2 (.05 or .4), number of daughters (50 or 100) and selection schemes (random, disruptive, truncation and stabilizing), for a total of 64 combinations. For each sub-population, fifty replicates were generated and were analyzed separately. Phenotypic variance ($\sigma_a^2 + \sigma_e^2$) was set to 10,000. Power on each combination was estimated by the proportion of significant results over 50 replicates.

Results and Discussions

Profile of Test Statistics

Twice the log likelihood ratio was used as statistics for test of linkage between QTL and genetic markers. This statistic has a Chi-square distribution with one degree of freedom. There are 50 replicates for each sub population. Figure 5 shows the profile of test statistics in the sub population with truncation selection scheme, 30 cM marker interval, 50 daughters, $v^2=0.125$ and $h^2=0.05$. The significant threshold for test statistics is 3.841 from Chi-square distribution with one degree of freedom. There were 17 replicates with test statistics beyond the threshold. Power was estimated at 0.34.

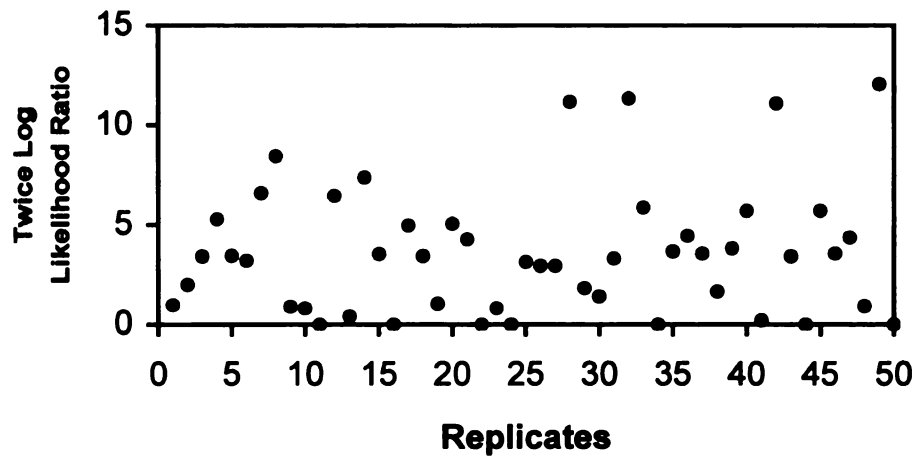


Figure 5 - Profiles of Test Statistics. The statistics were the twice log likelihood ratio from 50 replicates in sub-population with truncation selection scheme, 30 centi Morgan marker interval, 50 daughters, $v^2=0.125$ and $h^2=0.05$. Each point represents an observation of the test statistic.

random selection, whereas truncation and stabilizing selection have less power than random selection.

Power was saturated in the sub-populations under both the disruptive and random selection schemes. They reached the maximum except for those populations with larger marker interval (30 cM), fewer daughters (50), lower QTL allele variance ratio (0.125) and lower heritability (0.05) with random selection.

Disruptive selection increased the proportion of homozygote individuals with either large or with small effect on QTL. QTL variance was increased. The certainty of knowing the identity of markers in offspring to a parent increased. Therefore disruptive selection had more power of testing the QTL effect linked to markers. Significant increases in power for disruptive selection compared to random selection was founded in the situation where power was not saturated in random selection.

Mackinnon and Georges (1992) found that selection by truncation for the trait of interest significantly reduced the difference between marker genotype means and thus reduced the power to detect linked quantitative trait loci by ANOVA method. Preferred individuals under truncation selection were the homozygote with large QTL effects. The proportions of homozygote with small effects and heterozygotes were decreased. This lead to the reduction of QTL

Table 4 - Power* from Populations Subjected to Alternative Selection Schemes with Different Levels of Heritability (h^2) and Magnitudes of QTL Effect (v^2).

Marker Interval (cM)	Number of Daughters	Selection Scheme	Magnitude of QTL Effect (v^2)			
			0.125 h^2		0.25 h^2	
			0.05	0.4	0.05	0.04
10	50	Random	1.00 ^a	1.00 ^a	1.00 ^a	1.00 ^a
		Disruptive	1.00 ^a	1.00 ^a	1.00 ^a	1.00 ^a
		Truncation	0.56 ^{ij}	0.98 ^{ab}	0.94 ^{abcd}	0.98 ^{ab}
		Stabilizing	0.62 ^{hi}	0.96 ^{abc}	0.92 ^{bcde}	1.00 ^a
	100	Random	1.00 ^a	1.00 ^a	1.00 ^a	1.00 ^a
		Disruptive	1.00 ^a	1.00 ^a	1.00 ^a	1.00 ^a
		Truncation	0.76 ^{fgh}	1.00 ^a	1.00 ^a	0.98 ^{ab}
		Stabilizing	0.80 ^{efg}	0.98 ^{ab}	1.00 ^a	1.00 ^a
30	50	Random	0.84 ^{def}	1.00 ^a	1.00 ^a	1.00 ^a
		Disruptive	1.00 ^a	1.00 ^a	1.00 ^a	1.00 ^a
		Truncation	0.34 ^k	0.76 ^{fgh}	0.66 ^{ghi}	0.98 ^{ab}
		Stabilizing	0.40 ^{jk}	0.70 ^{fghi}	0.70 ^{fghi}	1.00 ^a
	100	Random	0.98 ^{ab}	0.98 ^{ab}	1.00 ^a	1.00 ^a
		Disruptive	1.00 ^a	1.00 ^a	1.00 ^a	1.00 ^a
		Truncation	0.58 ^{hij}	0.90 ^{bcde}	0.80 ^{efg}	1.00 ^a
		Stabilizing	0.56 ^{ij}	0.88 ^{bcde}	0.92 ^{bcde}	1.00 ^a

*Powers with the same superscripts were not different significantly ($P>0.05$).

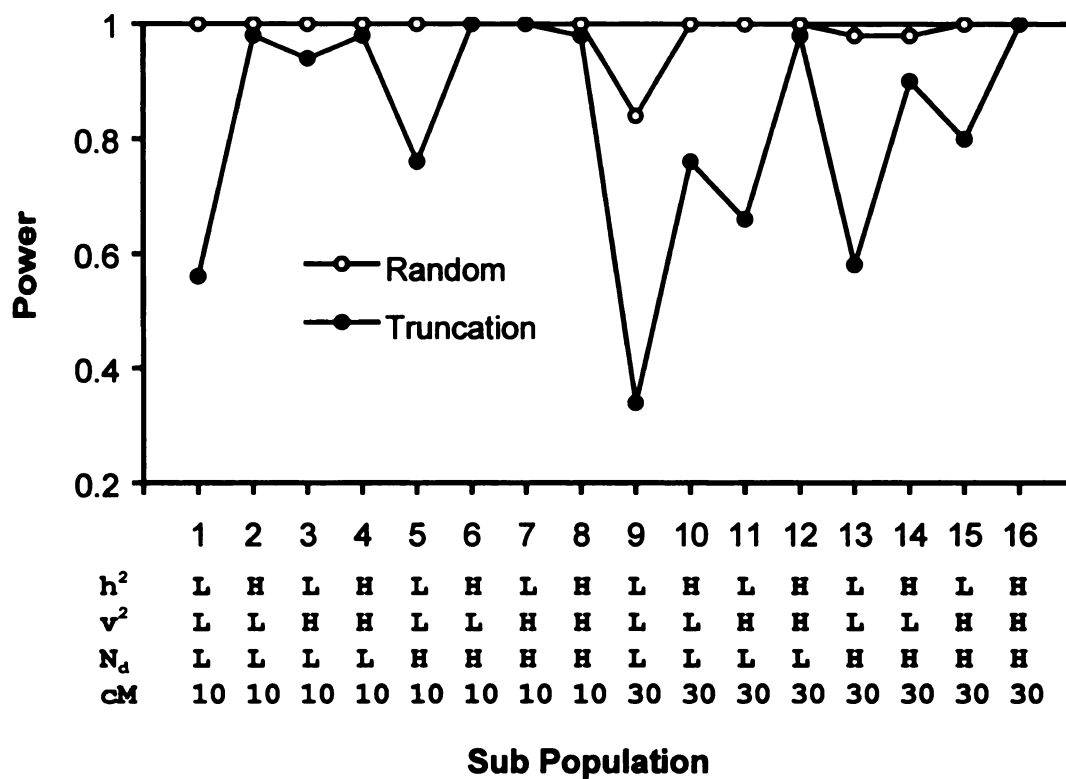


Figure 6 - Comparison of Powers between Random and Truncation Selection Schemes. The comparison are across populations with various combinations of underlying heritability (h^2) levels, magnitudes of QTL effect (v^2), numbers of daughters (N_d), and lengths of maker interval (in cM). L denotes the populations with .05 for h^2 , .125 for v^2 , and 50 for N_d ; H denotes the populations with .4 for h^2 , .25 for v^2 , and 100 for N_d .

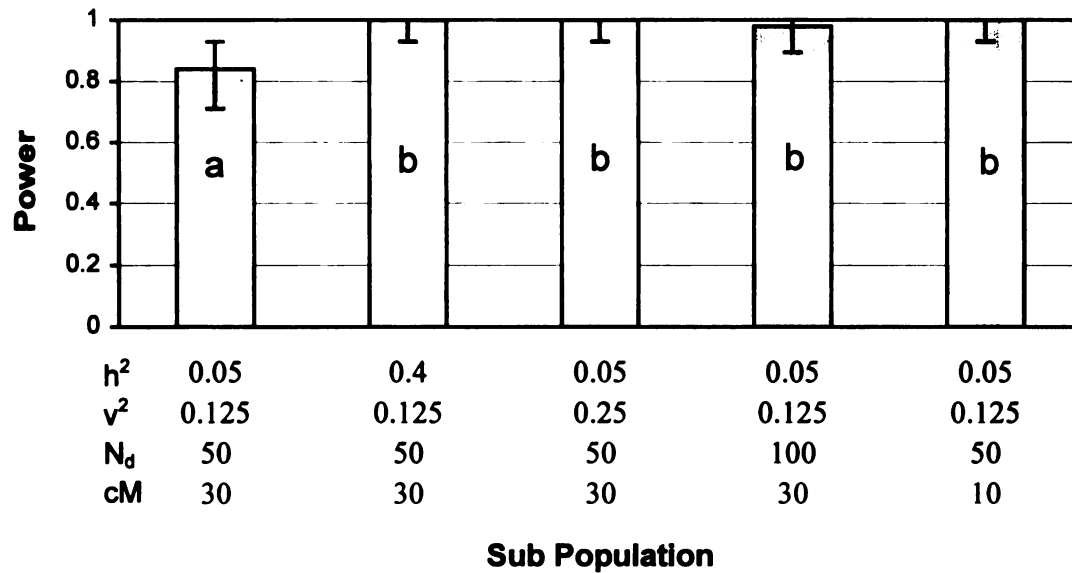


Figure 7 - Marginal Effects on Power of Test with Random Selection. The effects are from heritability (h^2), magnitude of QTL effect (v^2), number of daughters (N_d) and marker interval (cM). The bars with different symbols (a and b) are different at 0.05 level. Error bars indicate 95% confidence intervals.

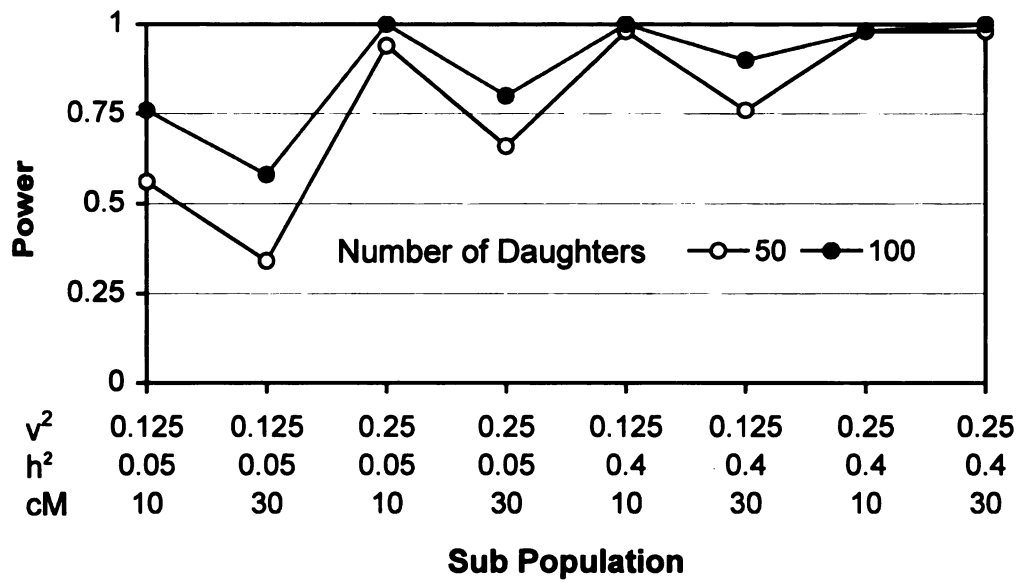


Figure 8 - Effect of Number of Daughters on Power of Test Under Truncation Selection Scheme.

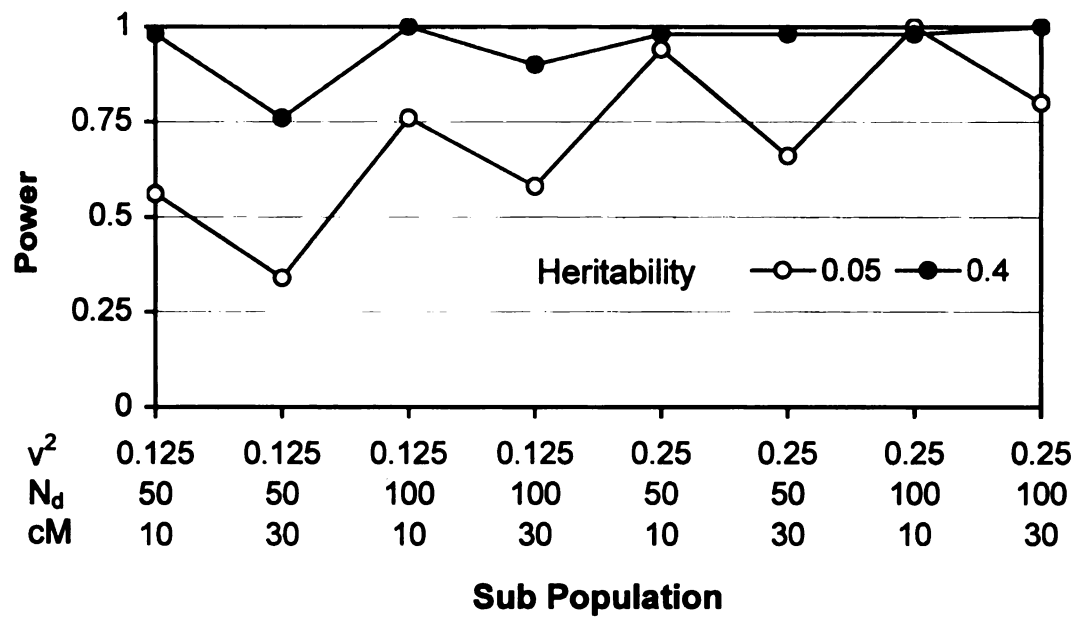


Figure 9 - Effect of Heritability on Power of Test Under Truncation Selection Scheme.

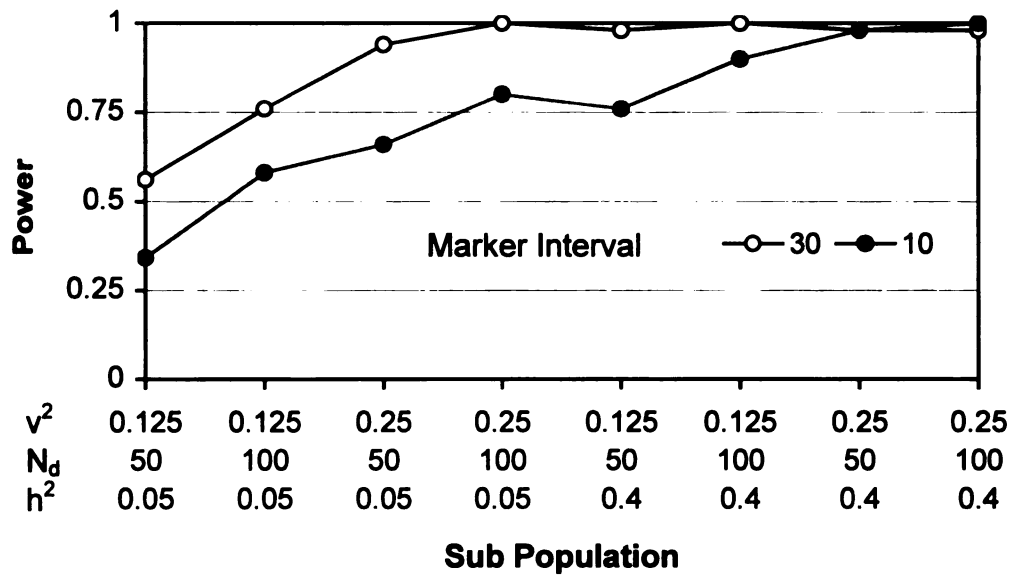


Figure 10 - Effects of Marker Interval on Power of Test under Truncation Selection Scheme.

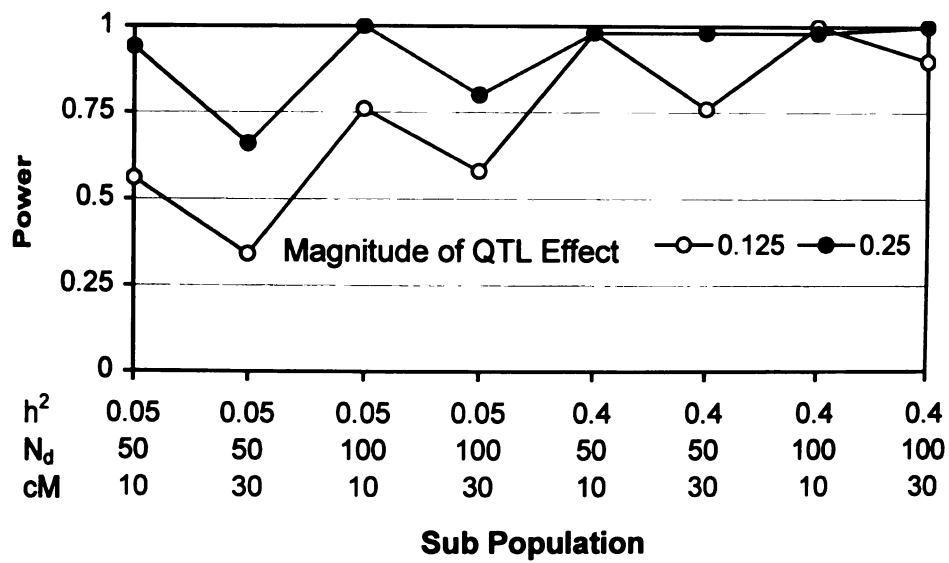


Figure 11- Effects of Magnitude of QTL Effect on Power of Test Under Truncation Selection Scheme.

Selection Schemes

Selection schemes have great impact on power of test of linkage between QTL and genetic markers. Disruptive selection generated higher power than variance. However, the certainty of knowing that a genetic marker in offspring was identical to a parent was decreased. Therefore truncation selection leads to reduction of power. Significant reductions of truncation selection to random selection were found in most sub-populations in this study (Figure 6).

The behavior of stabilizing selection was similar to truncation selection, whereas the reason for the reduction of power by stabilizing selection was different from truncation. Heterozygotes were the favorable individual under stabilizing selection. The proportions were reduced for the homozygotes with large or small genotype effects. Stabilizing selection reduced QTL variance. No difference between truncation and stabilizing selection was found in this study ($p>0.05$).

Other Factors

Power in sub-populations under disruptive selection was saturated. Exploring the impact of number of daughters, marker interval, QTL effect and heritability was based on other selection schemes. Further, the analysis focused on random and truncation selection schemes since no difference was found between truncation and stabilizing selection.

The result of this study showed the power of test depended on number of daughters, marker interval, magnitude of QTL effect and heritability. Power was

increased with more daughters, closer marker interval, larger QTL effect and higher heritability.

Under random selection, the power was 0.84 with $h^2=0.05$, $v^2=0.125$, 50 daughters and 30 cM marker interval. Significant increases of power ($p<0.05$) were found by changing the level of anyone of these factors; higher h^2 , higher v^2 , more daughters, or closer marker interval (Figure 7).

Under truncation selection, higher heritability, larger magnitude of QTL effect, more daughters on shorter marker interval generally generated higher power (Figure 8-11). Reverse situations were found in two sub populations (Figure 9 and Figure 11). However, the differences reversed were not significant ($p>0.05$, see Table 4).

The difference of power between two levels of a factor was larger when power was less saturated by other factors.

The results indicated that when v^2 was 0.125, a 10 cM marker interval increased power by 65% when compared to a 30 cM marker interval with 50 daughters and heritability of 0.05 under truncation selection. Whereas, when v^2 was 0.25, a 10 cM marker interval increased power by 42% when compared to a 30 cM marker interval with the same situation.

When marker interval was 30 cM, a $v^2=0.125$ increased power by 94% when compared to when v^2 was 0.25 and heritability was 0.05 with 50 daughters under truncation selection. Whereas, when marker interval was 10 cM, a $v^2=0.125$ increased power by 68% when compared to a v^2 of 0.25 with same situation.

Conclusions and Implications

Selection scheme has significant influence on the power of testing linkage between genetic marker and QTL using restricted maximum likelihood. Disruptive selection generated higher power than random selection, whereas truncation and stabilizing selection had less power than random selection.

Power of test using restricted maximum likelihood also depended on number of daughters per sire, marker interval, magnitude of QTL effect and heritability. Power of test was higher with more daughters per sire, smaller marker interval, larger magnitude of QTL effect, and higher heritability.

The magnitude of the difference due to changing a factor was larger when the power was less saturated by other factors.

GENERAL CONCLUSION

This study confirmed that estimates of QTL location, polygene variance, QTL allele variance and residual variance were unbiased in unselected populations. The unbiasedness was released to the number of daughters, marker densities, magnitude of QTL effect and heritability of trait.

Estimates of QTL location were unbiased even in selected populations. However, estimates of QTL variance, polygenic variance and residual variance were biased in all selection schemes except random selection.

Variances of polygenic, QTL and residual effects were overestimated with disruptive selection and underestimated with truncation and stabilizing. The biases with disruptive selection reduced with more daughters, larger magnitude of QTL effect and higher heritability of traits.

The selection schemes has significant influence on the power of testing linkage between genetic markers and QTL using residual maximum likelihood. Disruptive selection generated higher power than random selection, whereas truncation and stabilizing selection had less power than random selection.

Power of test increased with more daughters of each sire, smaller marker intervals, larger magnitude of QTL effect and higher heritability.

The magnitude of the difference due to changing a factor was larger when power was less saturated by other factors.

APPENDICES

APPENDIX A

APPENDIX A

Expectation of Heritability Weighted DYD

Let $y_1, \dots, y_n$ be the n daughters' yields of a son and μ be the population mean, then daughters' yield deviation (DYD) can be expressed as following:

$$\begin{aligned} \text{DYD} &= \frac{1}{n} [(y_1 - \mu) + \dots + (y_n - \mu)] \\ &= \frac{1}{n} (0.5s + 0.5d_1 + m_1 + e_1 + \dots + 0.5s + 0.5d_n + m_n + e_n) \\ &= 0.5s + \frac{1}{n} (0.5d_1 + \dots + 0.5d_n + m_1 + \dots + m_n + e_1 + \dots + e_n) \end{aligned}$$

where s is additive genetic effect of sire of daughter, d_i is additive genetic effect of dam of daughter i , m_i is Mendelian random effect of daughter i , e_i is environmental effects.

$$\begin{aligned} \text{Var}(\text{DYD}) &= \sigma_s^2 + \frac{1}{n} (\sigma_d^2 + 0.5\sigma_a^2 + \sigma_e^2) \\ &= 0.25\sigma_a^2 + \frac{1}{n} (0.25\sigma_a^2 + 0.5\sigma_a^2 + \sigma_e^2) \\ &= 0.25\sigma_a^2 + \frac{1}{n} (0.75\sigma_a^2 + \sigma_e^2) \end{aligned}$$

The variance of DYD was decomposed into two parts. The first part ($0.25\sigma_a^2$) is the variance of additive genetic of son. The other part is variance of residual effect. Therefore, the heritability of DYD is the proportion of $.25\sigma_a^2$ over $\text{Var}(\text{DYD})$ if DYD is analyzed with weight of one.

$$\frac{1}{n} (0.75\sigma_a^2 + \sigma_e^2) = \frac{0.75\sigma_a^2 + \sigma_e^2}{0.25n\sigma_a^2} 0.25\sigma_a^2$$

Let

$$w = \frac{0.75\sigma_a^2 + \sigma_e^2}{0.25n\sigma_a^2}$$

If DYDs are analyzed in REML with weight of $\frac{1}{w}$, the expectation of residual effect on DYD is $0.25\sigma_a^2$. Therefore, the expectation of heritability of DYD is 0.5.

Here, w can be proved as:

$$w = \frac{1 - \text{Reliability}}{\text{Reliability}}$$

where $\text{Reliability} = \frac{n}{n + K}$

$$K = \frac{4 - h^2}{h^2}$$

The derivation is following:

Under sire model, the heritability is four times of the sire variance over the total variance. The sire variance equal to 25% of additive genetic variance. The total variance is sum of sire variance and residual variance. The residual variance equal to 75% of additive genetic variance plus environmental variance. Therefore,

$$h^2 = \frac{4\sigma_s^2}{\sigma_s^2 + (0.75\sigma_a^2 + \sigma_e^2)}$$

$$k = \frac{4 - h^2}{h^2}$$

$$= \frac{0.75\sigma_a^2 + \sigma_e^2}{\sigma_s^2}$$

$$\text{Reliability} = \frac{n}{n + K}$$

$$= \frac{\sigma_s^2}{\sigma_s^2 + \frac{0.75\sigma_a^2 + \sigma_e^2}{n}}$$

$$w = \frac{1 - \text{Reliability}}{\text{Reliability}}$$

$$= \frac{0.75\sigma_a^2 + \sigma_e^2}{n\sigma_s^2}$$

$$= \frac{0.75\sigma_a^2 + \sigma_e^2}{0.25n\sigma_a^2}$$

Hence, there is an option of treating heritability of DYD as known (0.5) in REML analysis when phenotypic observation is DYD.

APPENDIX B

APPENDIX B

Input and Output of Grignola's Program

Input:

1. Pedigree of sires and sons
2. Genotypes of sires and sons
3. DYDs of sons
4. Heritability of yield

Output:

1. Estimate of QTL location
2. Estimate of heritability of DYD
3. v^2
4. Residual variance of DYD (σ_{DYDe}^2)

Estimates of polygenic, QTL and residual variance on yield can be directly calculated from the output of Grignola's program.

Let h^2 be the heritability of yield and h_{DYD}^2 be the estimate of heritability of DYD, then variance of polygenic(σ_a^2), QTL (σ_v^2) allele and residual effect (σ_e^2) on yield are calculated as following:

$$\sigma_a^2 = \frac{4h_{DYD}^2 \sigma_{DYDe}^2}{1 - h_{DYD}^2}$$

$$\sigma_v^2 = v^2 \sigma_a^2$$

$$\sigma_e^2 = \frac{(1 - h^2) \sigma_a^2}{h^2}$$

APPENDIX C

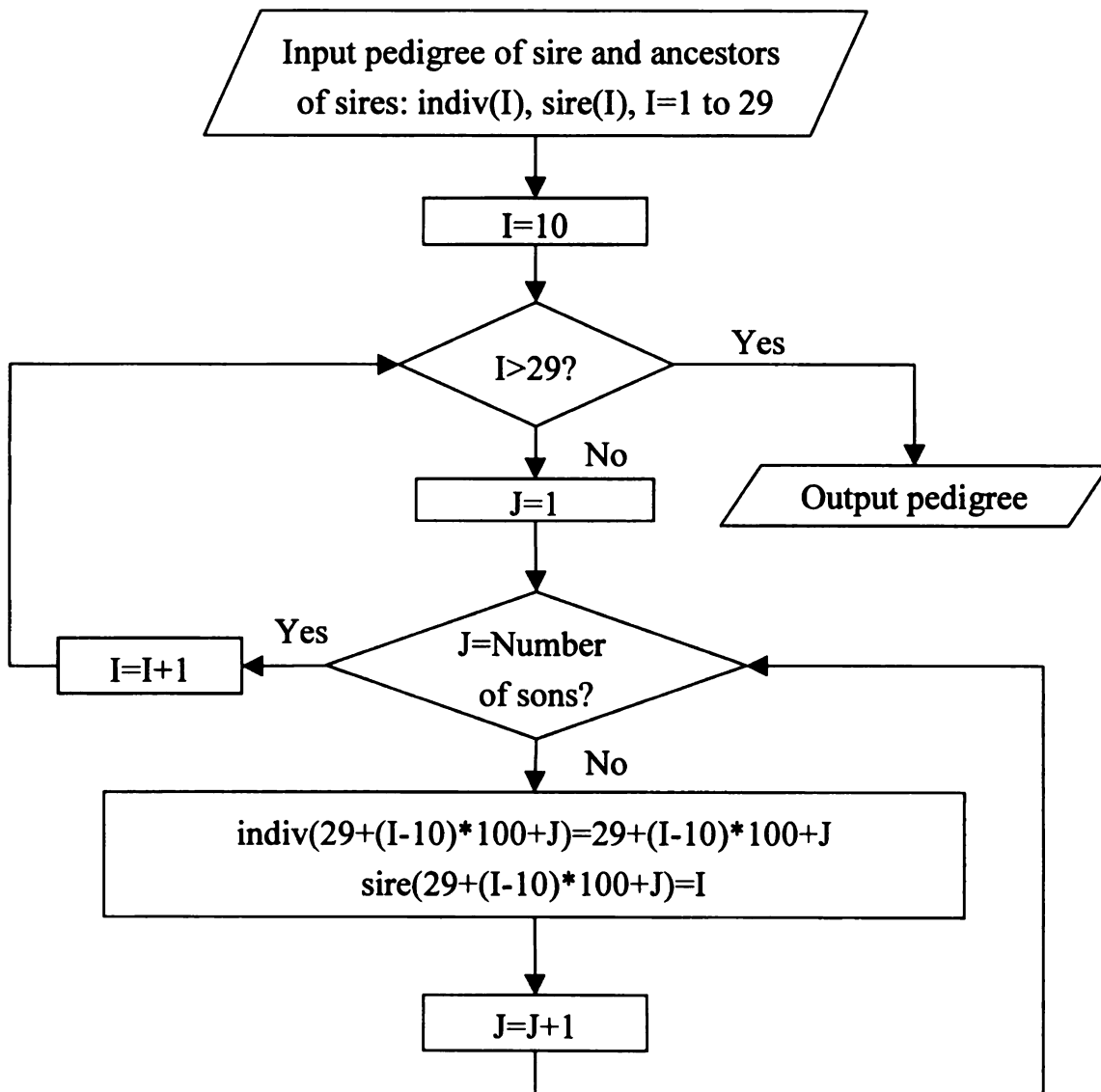
APPENDIX C

Flow Chart of Pedigree

Function: Generating pedigree of sons from pedigree of sires.

Input: Pedigree of sire and ancestors of sires

Output: Pedigree of sons

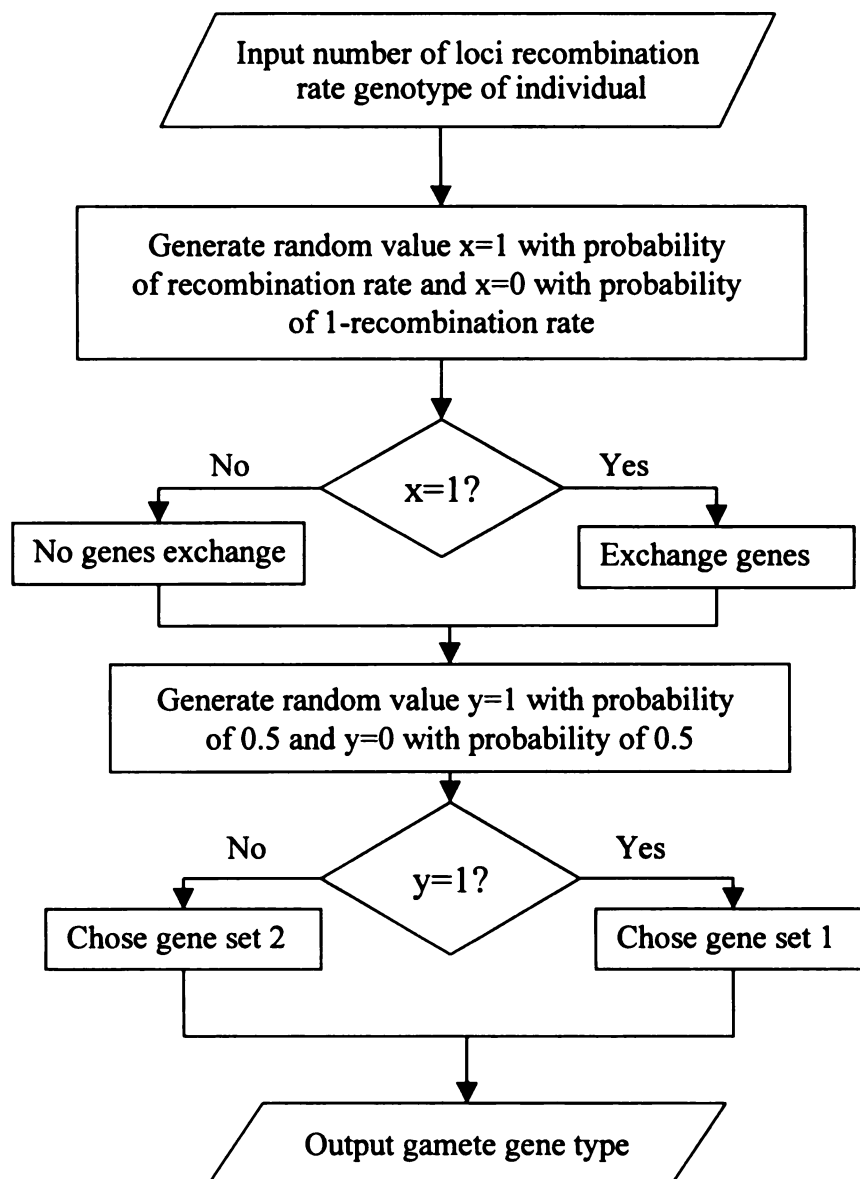


APPENDIX D

APPENDIX D

Flow Chart of Subroutine of crossover

Function: Generating a gamete from an individual.
Input: Number of gene loci, recombination rates, genotype of individual
Output: Gene type of gamete



APPENDIX E

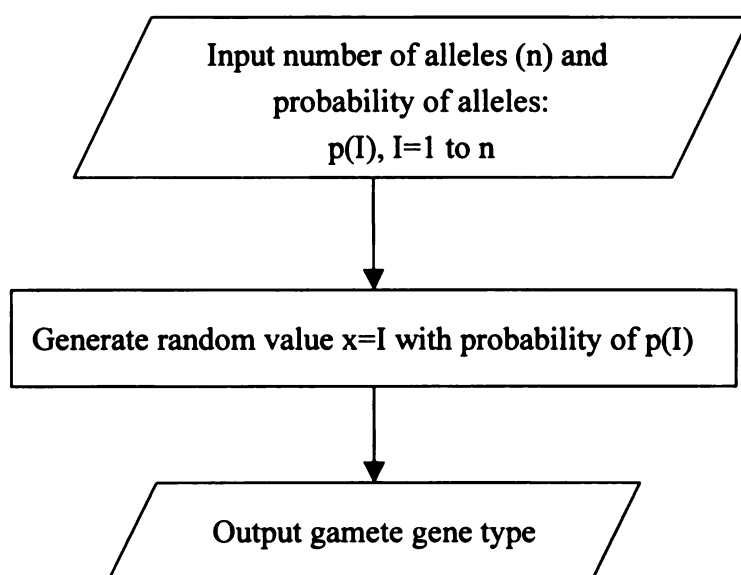
APPENDIX E

Flow Chart of subroutine of randasign

Function: Generating a gamete from a probability distribution.

Input: Allele probability

Output: Gene type of gamete

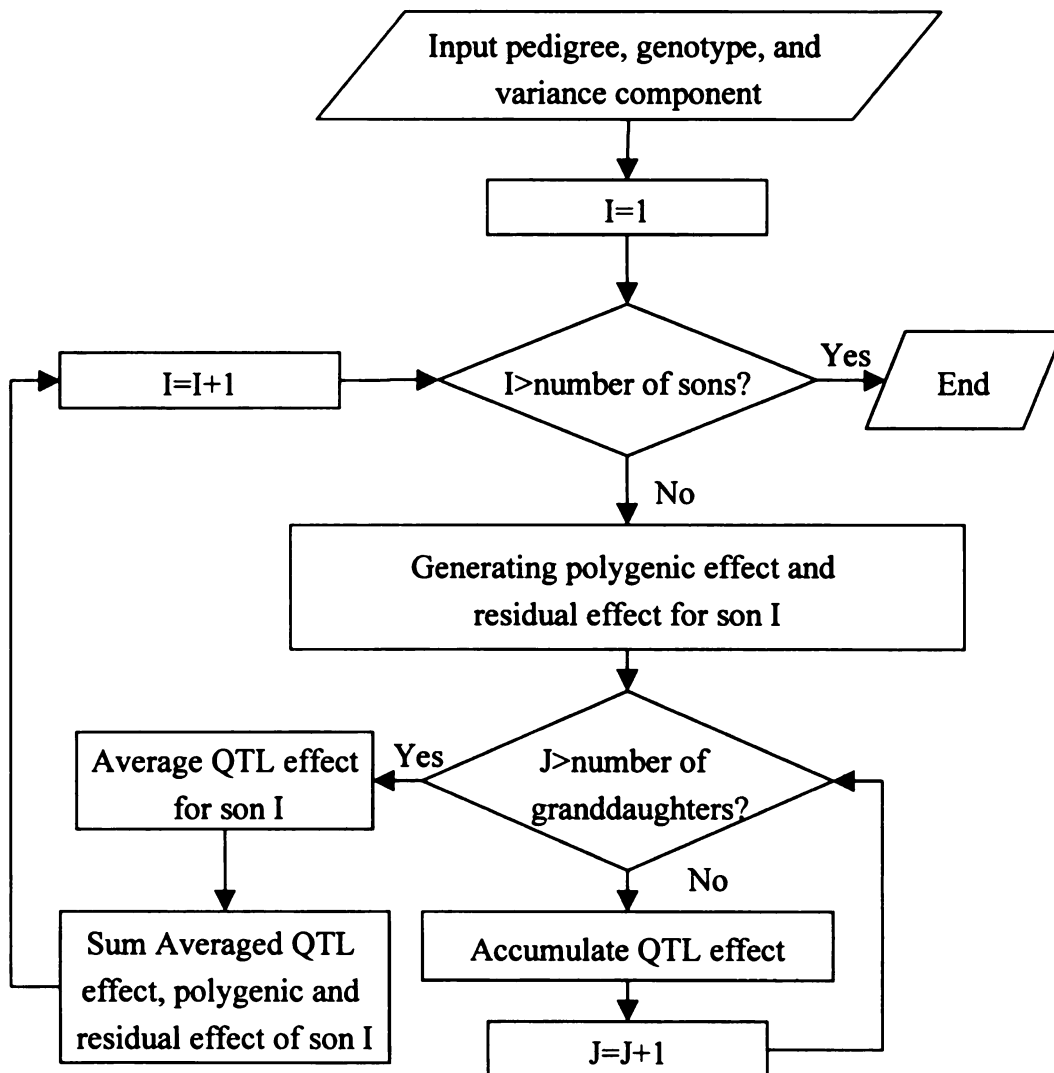


APPENDIX F

APPENDIX F

Flow Chart of subroutine of Phenotype

Function: Generating DYD
Input: Pedigree, genotype, and variance component
Output: DYD



APPENDIX G

APPENDIX G

FORTRAN (90) Code of Simulation Program

```
C*****
C
C      Simulation program
C      Written by Zhiwu Zhang
C      Last update: April 12, 1998
C
C*****
      PROGRAM simu
      USE MSIMSL

C      -----
C      Declare variables
C      -----

      common intvl,markfr,qtlfr,rintvl,rmarkfr,rqtlfr
      PARAMETER  (nmkloci=5,
*                nmkal=5,
*                nloci=nmkloci+1,
*                nqtlal=5,
*                ndaughter=1,
*                qtlposi=4,
*                ngsire=100,
*                nsire=10,
*                nson=1,
*                nparent=ngsire+ngsire*nsire, !sires+ancestors
*                nfparent=ngsire*nsire)      !sires

      INTEGER      seed,
*                idparent(nloci,2),
*                idgam(nloci),
*                indiv(nparent+nfparent*nson),
*                sire(nparent+nfparent*nson),
*                idmale(nparent+nfparent*nson,nloci),
*                idfemale(nparent+nfparent*nson,nloci),
*                irmk(nmkloci),
*                irqtl(1),
*                ntemp(nmkloci),
*                nqtlaray(nqtlal)

      REAL          recomb(nloci-1),
*                probmk(nmkal),
*                probqtl(nqtlal),
*                dyd,
*                qtlefect,
*                eqtlal(nqtlal),
*                weight,
```



```

*          nrmdev(1),
*          varu,
*          varv,
*          vare,
*          apoly(nparent+nfparent*nson),
*          epoly(nparent+nfparent*nson),
*          temp,
*          stdp,alpha,h2,dydh2,v2

C -----
c Set parameters
C -----

seed=0
stdp=100
h2=0.4
v2=0.125
data recomb/.3,.3,.15,.15,.3/
data probmk/.2,.2,.2,.2,.2/
data probqtl/.50,.50,.0,.0,.0/
sampling=1.0

vara=h2*stdp*stdp
varv=v2*vara
alpha=sqrt(varv/(probqtl(1)*probqtl(2)))
eqtlal=0
eqtlal(1)=0.5*alpha
eqtlal(2)=-.5*alpha

varu=vara-2*varv
vare=(1-h2)*stdp*stdp
dydh2=vara/(vara+(vare+0.75*vara)/ndaughter)
c weight is function of h2 and number of daughter
temp=(4-h2)/h2
temp=ndaughter/(ndaughter+temp)
weight=temp/(1-temp)
selection=sampling*sqrt(.25*vara+(vare+0.75*vara)/ndaughter)
C -----
c output parameter
C -----

open(10,file='mremlpq',status='unknown')    ! parameter file of
mremlpq
open(11,file='pedmkqtl.dat',status='unknown')    ! output of
marker and atl
open(12,file='dydzw.dat',status='unknown')    ! output of dyd
open(9,file='simulate.par',status='unknown')    ! parameter file

write(9,*) 'Number of marker loci          ',nmkloci
write(9,*) 'Number of marker alleles        ',nmkal
write(9,*) 'Number of QTL allels           ',nqtlal
write(9,*) 'Number of daughters            ',ndaughter
write(9,*) 'Number of sons per sire ',nson
write(9,*) 'Order of QTL among markers',qtlposi

```

```

write(9,*) 'Seed for random number      ',seed
write(9,*) ''
write(9,*) '                        var(v)      var(u)
var(e)'
write(9,99991) varv,varu,vare
temp=0
do i=1,qtlposi-2
temp=temp+recomb(i)
end do
temp=temp+recomb(qtlposi-1)
write(9,*) '                        v2      h2      dydh2
DQTL'
write(9,99991) v2,h2,dydh2,temp

write(9,*) ''
write(9,*) 'Recobination rate'
write(9,99991) recomb
write(9,*) 'QTL allele   Frequ'
write(9,99991) probqtl
write(9,*) 'QTL effect   '
write(9,99991) eqtlal
write(9,*) 'Marker   Frequency'
write(9,99991) probmk
write(9,*) ''
write(9,*) 'Weight                        ',weight

close (9)
99991 FORMAT (10X,7F10.2)

C      -----
c      output parameter file "mrem1pq"
C      -----
write(10,1020) .125, .50, 1000.0, .75
write(10,1020) .499, .99, 9000.0, 1.20
write(10,1020) .001, .01, 0.1, .01
write(10,1040) 0
write(10,1020) 0.2
write(10,1020) .0, .3, .6, .9, 1.2
write(10,1040) 5,5,5,5,5
do i=1,nparent
do j=1,5
write(10,1030) i,j,0.2,0.2,0.2,0.2,0.2
end do
end do
write(10,1020) 0.01, 1.19
write(10,1020) 0.01,0.01
1020 FORMAT (1X,5F10.3)
1030 FORMAT (1X,2I4,5F5.2)
1040 FORMAT (1X,5I4)

C      end of output paramaters

```

```

write(*,*) 'It is generating data, please wait.'

C -----
c   Generating pedigree
c   by Zhiwu Zhang, August 30, 1997
C -----
c   pedigree of grandsire
c   do i=1,ngsire
c   indiv(i)=i
c   sire(i)=0
c   end do

c   pedigree of sire
c   do i=1,ngsire
c   do j=1,nsire
c   indiv(ngsire+(i-1)*nsire+j)=ngsire+(i-1)*nsire+j
c   sire(ngsire+(i-1)*nsire+j)=i
c   end do
c   end do

c   pedigree of son
c   do i=1,ngsire*nsire
c   do j=1,nson
c   indiv(ngsire+ngsire*nsire+(i-1)*nson+j)=
c   *      ngsire+ngsire*nsire+(i-1)*nson+j
c   sire(ngsire+ngsire*nsire+(i-1)*nson+j)=
c   *      ngsire+i
c   end do
c   end do

C -----
c   seed 0 for random, non zero for fixed initial
C -----
CALL RNSET (SEED)

99998 FORMAT (2I6,6I3)

do i=1,nparent+nfparent*nson
C -----
C   generating marker and qtl information
c   by Zhiwu Zhang, August 20, 1997
C -----
777 if (sire(i).eq.0) then
c   unknown sire
c   male side
c       call ransign(nmkal,probm, nmkloci, irmk)
c       do j=1, qtlposi-1
c       idmale(i,j)=irmk(j)

```

```

        end do
        do j=qtlposi+1,nloci
            idmale(i,j)=irmk(j-1)
        end do
        call ransign(nqtlal,probqtl,1,irqtl)
        idmale(i,qtlposi)=irqtl(1)

c      femal side
        call ransign(nmkal,probmkn,nmkloci,irmk)
        do j=1,qtlposi-1
            idfemale(i,j)=irmk(j)
        end do
        do j=qtlposi+1,nloci
            idfemale(i,j)=irmk(j-1)
        end do
        call ransign(nqtlal,probqtl,1,irqtl)
        idfemale(i,qtlposi)=irqtl(1)
c      known sire
        else
c      male side
            do j=1,nloci
                idparent(j,1)=idmale(sire(i),j)
                idparent(j,2)=idfemal(sire(i),j)
            end do
            call crossover(nloci,recomb,idparent,idgam)
            do j=1,nloci
                idmale(i,j)=idgam(j)
            end do
c      femal side
            call ransign(nmkal,probmkn,nmkloci,irmk)
            do j=1,qtlposi-1
                idfemale(i,j)=irmk(j)
            end do
            do j=qtlposi+1,nloci
                idfemale(i,j)=irmk(j-1)
            end do
            call ransign(nqtlal,probqtl,1,irqtl)
            idfemale(i,qtlposi)=irqtl(1)
        end if

C      -----
c      polygene effect
c      by Zhiwu Zhang, September 10, 1997
C      -----
        CALL RNNOR (1, nrmdev)
        if (sire(i).eq.0) then
c      unknown sire
            apoly(i)=sqrt(varu)*nrmdev(1)
        else
            apoly(i)=sqrt(0.75*varu)*nrmdev(1)
            *          +0.5*apoly(sire(i))
        end if

```

```

C -----
C Mendilian and residual effect
C -----
      if (i .gt. nparent) then
      CALL RNNOR (1, nrmdev)
      epoly(i)=sqrt((0.75*varu+vare)/ndaughter)*nrmdev(1)

C -----
C QTL effect
C by Zhiwu Zhang, September 9, 1997
C -----

      nqtlaray=0
c      proparation for male side
            do k=1,nloci
            idparent(k,1)=idmale(i,k)
            idparent(k,2)=idfemal(i,k)
            end do

      do j=1,ndaughter
c      female side
            call ransign(nqtlal,probqtl,1,irqtl)
            nqtlaray(irqtl(1))=nqtlaray(irqtl(1))+1
c      male side
            call crossover(nloci,recomb,idparent,idgam)
            nqtlaray(idgam(qtlposi))=nqtlaray(idgam(qtlposi))+1
      end do
      qtlefect=0
      do j=1,nqtlal
            qtlefect=qtlefect+nqtlaray(j)*eqtlal(j)
      end do
      qtlefect=qtlefect/ndaughter

      dyd=qtlefect+0.5*apoly(i)+epoly(i)
c      1 Random sampling
c      keep nothing
c -----
c      2 Extrem sampling
c      if (abs(dyd).lt. selection) goto 777
c -----
c      3 over average sampling
c      if (dyd .lt. 0.0) goto 777
c -----
c      4 middle sampling
c      if (abs(dyd).gt. selection) goto 777
c -----
c      5 high selection sampling
c      if (dyd .lt. selection) goto 777

c      output DYD information
      write(12,99997) indiv(i),sire(i),1,dyd,weight

      end if

```

```

        do j=1, qtlposi-1
            ntemp(j)=idmale(i,j)*10+idfemal(i,j)
        end do

        do j=qtlposi+1, nloci
            ntemp(j-1)=idmale(i,j)*10+idfemal(i,j)
        end do

C      output pedigree and marker information
        write(11,99998) indiv(i), sire(i), 0,
        *      (ntemp(j), j=1, nmkloci)

        end do      !      do i=1, nparent+nfparent*nson

99997 FORMAT (1I6,1I6,1I2,2F20.10)
        end

C      -----
c      subroutine for cross over
c      by Zhiwu Zhang, August 28, 1997
C      -----
        subroutine crossover(nloci, recomb, idparent, idgam)
c      nloci number of marker or qtl loci
c      recomb      recombination rate
c      idparent      Identification of alleles on parent
c      idgam Identification of alleles on gamete

c      for checking
        common intvl, rintvl
        integer      intvl(5),
        *      rintvl(5)
c      end of checking

        PARAMETER      (nr=1)
        INTEGER          ir(nr), idparent(nloci,2),
        *      idgam(nloci), temp
        REAL              recomb(nloci-1), prob(2)

c      cross or not
        do i=1, nloci-1
            prob(1)=recomb(i)
            prob(2)=1-recomb(i)
            call ransign(2, prob, nr, ir)
            call ransign(2, prob, nr, ir)

```

```

      if (ir(1) .eq. 1) then
c      cross over

c      for checking
      intvl(i)=intvl(i)+1
c end of checking

      do j=i+1,nloci
      temp=idparent(j,1)
      idparent(j,1)=idparent(j,2)
      idparent(j,2)=temp
      end do
      end if
      end do

C      randomly chose one of two gamates
      prob(1)=.5
      prob(2)=.5
      call ransign(2,prob,nr,ir)
      if (ir(1) .eq. 1) then
c      option 1: chose first
      do i=1,nloci
      idgam(i)=idparent(i,1)
      end do
      else
c      option 2: non cross
      do i=1,nloci
      idgam(i)=idparent(i,2)
      end do
      end if

      end

C      -----
c      subroutine to sign alleles randomly
c      by Zhiwu Zhang, August 25, 1997
C      -----
      subroutine ransign(nmal,probs,nr,ir)
c      nmal  number of marker alleles
c      probs frequency of marker alleles
c      nr      total number to generate
c      ir      array to store random nuber generated

      INTEGER IMIN, IOPT, IWK(nmal), NOUT,ir(nr)
      REAL  WK(nmal),probs(nmal)
CALL UMACH (2, NOUT)
      IMIN      = 1
      IOPT      = 0
      nmass=nmal
c      CALL RNSET (SEED)
      CALL RNGDA (NR, IOPT, IMIN, NMASS, PROBS, IWK, WK, IR)
      END

```

BIBLIOGRAPHY

BIBLIOGRAPHY

- Andersson, L., C. S. Haley, H. Ellegren, S. A. Knott, M. Johansson, K. Andersson, L. Andersson Eklund, I. Edfors Lilja, M. Fredholm, I. Hansson, J. Hakansson, and K. Lundstrom. 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science Washington* 263:1771-1774.
- Banks, B. D., I. L. Mao, and J. P. Walter. 1985. Robustness of the restricted maximum likelihood estimator derived under normality as applied to data with skewed distributions. *J. Dairy Sci.* 68:1785-1792.
- Beaumont, C. 1991. Comparison of Henderson's Method I and restricted maximum likelihood estimation of genetic parameters of reproductive traits. *Poult. Sci.* 70:1462-1468.
- Bovenhuis, H., J. A. M. Van Arendonk, G. Davis, J. M. Elsen, C. S. Haley, W. G. Hill, P. V. Baret, D. J. S. Hetzel, and F. W. Nicholas. 1997. Detection and mapping of quantitative trait loci in farm animals. *Livestock Production Science* 52:135-144.
- Bovenhuis, H., and J. I. 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.
- Carbonell, E. A., M. J. Asins, M. Baselga, E. Balansard, and T. M. Gerig. 1993. Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid populations. *Theor. Appl. Genet.* 86:411-416.
- Cowan, C. M., M. R. Dentine, R. L. Ax, and L. A. Schuler. 1990. Structural variation around prolactin gene linked to quantitative traits in an elite Holstein sire family. *Theor. Appl. Genet.* 79:577-582.
- Darvasi, A., and M. Soller. 1992. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.* 85:353-359.

- Darvasi, A., A. Weinreb, V. Minke, J. I. Weller, and M. Soller. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134:943-951.
- Elsen, J. M., S. Knott, P. I. Roy, C. S. Haley, and P. Le Roy. 1997. Comparison between some approximate maximum-likelihood methods for quantitative trait locus detection in progeny test designs. *Theor. Appl. Genet.* 95:1-2.
- Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. *Gen. Sel. Evol.* 21:467-477.
- Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele, and X. Zhao. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139:907-920.
- Gianola, D., J. L. Foulley, and R. L. Fernando. 1986. Prediction of breeding values when variances are not known. *Genet. Select. Evol.* 18:485-497.
- Gimelfarb, A., and R. Lande. 1994. Simulation of marker assisted selection for non additive traits. *Genet. Res.* 64:127-136.
- Gimelfarb, A., and R. Lande. 1995. Marker-assisted selection and marker-QTL associations in hybrid populations. *Theor. Appl. Genet.* 91:522-528.
- Goddard, M. E. 1991. A mixed model for analyses of data on multiple genetic markers. *Theor. Appl. Genet.* 83:878-886.
- Grignola, F. E., I. Hoeschele, and B. Tier. 1996a. Mapping quantitative trait loci in outcross populations via residual maximum likelihood: I. Methodology. *Genet. Sel. Evol.* 28:479-490.
- Grignola, F. E., I. Hoeschele, Q. Zhang, and G. Thaller. 1996b. Mapping quantitative trait loci in outcross populations via residual maximum likelihood: II. A simulation study. *Genet. Sel. Evol.* 28:491-504.
- Haley, C. S., and S. A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69:315-324.
- Haley, C. S., S. A. Knott, and J. M. Elsen. 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136:1195-1207.
- Hanset, R., and C. Michaux. 1985a. On the genetic determination of muscular hypertrophy in the Belgian White and Blue cattle breed. I. Experimental data. *Genet. Select. Evol.* 17:359-368.

- Hanset, R., and C. Michaux. 1985b. On the genetic determination of muscular hypertrophy in the Belgian White and Blue cattle breed. II. Population data. *Genet. Select. Evol.* 17:369-386.
- Henderson, C. R. 1975a. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423-447.
- Henderson, C. R. 1975b. Use of relationships among sires to increase accuracy of sire evaluation. *J. Dairy Sci.* 58:1731-1738.
- Hill, W. G. 1982. Prediction of response to artificial selection from new mutations. *Genet. Res.* 40:255-278.
- Hoeschele, I. 1990. Potential gain from insertion of major genes into dairy cattle. *J. Dairy Sci.* 73:2601-2618.
- Hoeschele, I., and P. M. VanRaden. 1993a. Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. *Theor. Appl. Genet.* 85:953-960.
- Hoeschele, I., and P. M. VanRaden. 1993b. Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. *Theor. Appl. Genet.* 85:946-952.
- Hospital, F., and L. Moreau. 1997. More on the efficiency of marker assisted selection. *Theor. Appl. Genet.* 95:1181-1189.
- Hyne, V., and M. J. Kearsey. 1995. QTL analysis: further uses of 'marker regression'. *Theor. Appl. Genet.* 91:471-476.
- Jansen, R. C. 1993. Interval mapping of multiple quantitative trait loci. *Genetics* 135:205-211.
- Jansen, R. C. 1994a. Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics* 138:871-881.
- Jansen, R. C. 1994b. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136:1447-1455.
- Jensen, J. 1989. Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. *Theor. Appl. Genet.* 78:613-618.

- Johnson, Z.B., D.W. Wright, C.J. Brown, J.K. Bertrand, A.H. Brown. 1991 Effect of including relationships in the estimation of genetic parameters of beef calves. *J. of Anim. Sci.* 70: 78-88.
- Kashi, Y., E. Hallerman, and M. Soller. 1990. Marker-assisted selection of candidate bulls for progeny testing programmes. *Anim. Prod.* 51:63-74.
- Keightley, P. D., and G. Bulfield. 1993. Detection of quantitative trait loci from frequency changes of marker alleles under selection. *Genet. Res.* 62:195-203.
- Knapp, S. J., and W. C. Bridges. 1990. Using molecular markers to estimate quantitative trait locus parameters power and genetic variances for unreplicated and replicated progeny. *Genetics* 126:769-777.
- Knott, S. A., and C. S. Haley. 1992a. Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* 132:1211-1222.
- Knott, S. A., C. S. Haley, and R. Thompson. 1992b. Methods of segregation analysis for animal breeding data: A comparison of power. *Heredity* 68:299-311.
- Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743-756.
- Lander, E. S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RnP linkage maps. *Genetics* 121:185-199.
- Le-Roy, P., and J. M. Elsen. 1995. Numerical comparison between powers of maximum likelihood and analysis of variance methods for QTL detection in progeny test designs: the case of monogenic inheritance. *Theor. Appl. Genet.* 90:65-72.
- Lin, J. Z., and K. Ritland. 1996. The effects of selective genotyping on estimates of proportion of recombination between linked quantitative trait loci. *Theor. Appl. Genet.* 93:1261-1266.
- Luo, Z. W., and J. A. Woolliams. 1993. Estimation of genetic parameters using linkage between a marker gene and a locus underlying a quantitative character in F2 populations. *Heredity* 70:245-253.
- Mackinnon, M. J., and M. A. J. Georges. 1992. The effects of selection of linkage analysis for quantitative traits. *Genetics* 132:1177-1185.

- Mackinnon, M. J., and J. I. Weller. 1995. Methodology and accuracy of estimation of quantitative trait loci parameters in a half-sib design using maximum likelihood. *Genetics* 141:755-770.
- Martinez, 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.
- Martinez, O., and R. N. Curnow. 1994. Missing markers when estimating quantitative trait loci using regression mapping. *Heredity* 73:198-206.
- Meuwissen, T. H. E., and M. E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. *Genet. Sel. Evol.* 28:161-176.
- Meuwissen, T. H. E., and M. E. Goddard. 1997. Estimation of effects of quantitative trait loci in large complex pedigrees. *Genetics* 146:409-416.
- Meuwissen, T. H. E., and A. Van, J. A. M. 1992. Potential improvements in rate of genetic gain from marker-assisted selection dairy cattle breeding schemes. *J. Dairy Sci.* 75:1651-1659.
- Moreno-Gonzalez, J. 1992. Genetic models to estimate additive and non-additive effects of marker-associated QTL using multiple regression techniques. *Theor. Appl. Genet.* 85:435-444.
- Mullis, K. B. 1990. The unusual origin of the polymerase chain reaction. *Scientific American* 262:56-65.
- Neimann-Sorensen, A., and A. Robertson. 1961. The associations between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agricultura Scandinavica* 11:163-196.
- Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln, and S. D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335:721-726.
- Patterson, H. D., and R. Thompson. 1971. Recovery of inter block information when block size are unequal. *Biometrics* 58:545-554.
- Rathje, T. A., G. A. Rohrer, and R. K. Johnson. 1997. Evidence for Quantitative trait loci affecting ovulation rate in pigs. *J. Anim. Sci.* 75:1486-1494.
- Rebai, A., B. Goffinet, and B. Mangin. 1995. Comparing power of different methods for QTL detection. *Biometrics* 51:87-99.

- Rothschild, M., C. Jacobsen, D. Vaske, C. Tuggle, L. H. Wang, T. Short, G. Eckardt, S. Sasaki, A. Vincent, and D. McLaren. 1994. A major gene for litter size in pigs. 5th World congress of genetic applied to livestock production 21:225-229.
- Ruane, J., and J. J. Colleau. 1996. Marker-assisted selection for a sex-limited character in a nucleus breeding program. J. Dairy Sci. 79:1666-1678.
- Saito, S., and H. Iwaisaki. 1996. A reduced animal model with elimination of quantitative trait loci equations for marker-assisted selection. Genet. Select. Evol. 28:465-477.
- Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:522-560.
- Schaeffer, L. R., B. W. Kennedy, and J. P. Gibson. 1989. The inverse of the gametic relationship matrix. J. Dairy Sci. 72:1266-1272.
- Searcy-Bernal, R. 1994. Statistical power and aquacultural research. Aquaculture 127:371-388.
- Shrimpton, A. E., and A. Robertson. 1988. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II Distribution of third chromosome bristle effects within chromosome sections. Genetics 118:445-459.
- Simpson, S. P. 1989. Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor. Appl. Genet. 77:815-819.
- Smith, C., and P. R. Bampton. 1977. Inheritance of reaction to halothane anaesthesia in pig. Genet. Res. 29:287-292.
- Smith, C., and S. P. Simpson. 1986. The use of genetic polymorphism in livestock improvement. Anim. Prod. 20:1-10.
- Soller, M., and J. S. Beckman. 1990. Marker-based mapping of quantitative trait loci using replicated progenies. Theor. Appl. Genet. 80:205-208.
- Soller, M., T. Brody, and A. Genizi. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47:35-39.

- Soller, M., and A. Genizi. 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* 34:47-55.
- Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of response to selection using least squares and mixed model methodology. *J. Anim. Sci.* 58:1097.
- Spelman, R. J., W. Coppieters, L. Karim, J. A. M. Van Arendonk, and H. Bovenhuis. 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144:1799-1808.
- Spelman, R., and D. Garrick. 1997a. Utilisation of marker assisted selection in a commercial dairy cow population. *Livestock Production Science* 47:139-147.
- Spelman, R. J., and J. A. M. Van Arendonk. 1997b. Effect of inaccurate parameter estimates on genetic response to marker assisted selection in outbred population. *J. Dairy Sci.* 80:3399-3410.
- Stam, p. 1986. The use of marker loci in selection for quantitative characters. In: *Exploiting new technologies in animal breeding*. Oxford, UK, pp 170-182.
- Uimari, P., G. Thaller, and I. Hoeschele. 1996. The use of multiple markers in a Bayesian method for mapping quantitative trait loci. *Genetics* 143:1831-1842.
- Van Arendonk, J. A. M., B. Tier, and B. P. Kinghorn. 1994. Use of multiple genetic markers in prediction of breeding values. *Genetics* 137:319-329.
- van der Beek, S., J. A. M. Van Arendonk, A. F. Groen, S. Van der Beek, and J. A. M. Van Arendonk. 1995. Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. *Theor. Appl. Genet.* 91:6-7.
- van der Beek, S., and J. A. M. Van Arendonk. 1996. Marker assisted selection in an outbred poultry breeding nucleus. *J. Anim. Sci.* 62:171-180.
- Van der Werf, J. H. J., and I. J. M. De Boer. 1990. Estimation of additive genetic variance when base populations are selected. *J. Anim. Sci.* 68:3124-3132.
- VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. *J. Dairy Sci.* 74:2737-2746.
- Visscher, P. M., M. Mackinnon, and C. S. Haley. 1990. Efficiency of marker assisted selection. *Anim. Biotechnol.* 8:99-106.

- Vukasinovic, N., M. L. Martinez, and A. E. Freeman. 1998. Mapping quantitative trait loci under selection. 6th World Congress of Genetic Applied to Livestock Production 26:261-264.
- Wang, T., R. L. Fernando, S. van der Beek, and S. M. Grossman. 1995. Covariance between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.* 27:251-274.
- Weller, J. I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* 42:627-610.
- Weller, J. I., Y. Kashi, and M. Soller. 1990a. Estimation of sample size necessary for genetic mapping of quantitative traits in dairy cattle using genetic markers. *J. Dairy Sci.* 73:.
- Weller, J. I., Y. Kashi, and M. Soller. 1990b. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy Sci.* 73:2525-2537.
- Weller, J. I., and M. Pen. 1994. Detection and mapping quantitative trait loci in segregating populations: Theory and experimental results. In: *Proceedings Fifth World Congress on Genetics Applied to Livestock Production*, Guelph 21:213-220.
- Weller, J. I., and A. Wyler. 1992. Power of different sampling strategies to detect quantitative trait loci variance effects. *Theor. Appl. Genet.* 83:582-588.
- Whittaker, J. C., R. Thompson, and P. M. Visscher. 1996. On the mapping of QTL by regression of phenotype on marker-type. *Heredity* 77:23-32.
- Wright, D.W., Z.B. Johnson, C.J. Brown, S. Wildeus. 1991. Variance and covariance estimates for weaning weight of Senepol cattle. *J. of Anim. Sci.* 69: 3945-3951
- Xu, S. 1995. A comment on the simple regression method for interval mapping. *Genetics.* 141: 1657-1659
- Zhang, W., and C. Smith. 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theor. Appl. Genet.* 83:813-820.
- Zhang, W., and C. Smith. 1993. Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. *Theor. Appl. Genet.* 86:492-496.

MICHIGAN STATE UNIV. LIBRARIES



31293017876560