Mapping Quantitative Trait Loci Using Multiple Linked Markers via Residual Maximum Likelihood

by

Fernando E. Grignola

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Genetics

APPROVED:

Ina Hoeschele, Chair

Ronald Pearson

David Notter

Eric Hallerman

William Vinson

August 19, 1996

Blacksburg, Virginia

Keywords: quantitative trait, gene mapping, residual maximum likelihood

Mapping Quantitative Trait Loci Using Multiple Linked Markers via Residual Maximum Likelihood

by

Fernando E. Grignola

Ina Hoeschele, Chairman

GENETICS

(ABSTRACT)

Mapping quantitative trait loci in outbred populations is important since development of inbred lines in livestock species is usually not feasible. Traditional genetic mapping methods, such as Least Squares and Maximum Likelihood, cannot fully accommodate complex pedigree structures, and more sophisticated methods such as Bayesian analysis are very demanding computationally. In this thesis, an alternative approach based on a Residual Maximum Likelihood method for estimation of position and variance of one or two linked QTLs and of additive polygenic and residual variances is presented. The method is based on a mixed linear model including polygenic and random QTL allelic effects. The variance-covariance matrix of QTL allelic effects and its inverse is computed conditional on incomplete information from multiple linked markers. The method is implemented using interval mapping and a derivative-free algorithm, where the required coefficient matrix of the Mixed Model Equations is derived from a Reduced Animal Model. Simulation studies based on a granddaughter design with 2000 sons, 20 sires and 9 ancestors were performed to evaluate parameter estimation and power of QTL detection. Daughter Yield Deviations of sons were simulated under three QTL models, a biallelic, a multiallelic (10 alleles), and a normal-effects model. A linkage group of five or nine markers located on the same chromosome was assumed, and genotypes were available on sons, sires and ancestors. Likelihood ratio statistics were used to test for the presence of one or two linked QTLs. Parameters were estimated quite accurately for all three QTL models, showing that the method is robust to the number of alleles at the QTL. The effect of considering or ignoring relationships in the analyses did not have a major impact on parameter estimates but reduced the power of QTL detection. In general,

power tended to decrease as the number of sons per sire, QTL contribution to additive genetic variance, or distance between QTLs was reduced. The method allowed for detection of a single QTL explaining 25% of the additive genetic variance, and for detection of two QTLs when jointly they accounted for 50% or 12.5% of the additive genetic variance. Although the REML analysis is an approximate method incorporating an expected covariance matrix of the QTL effects conditional on marker information, it is a computationally less expensive alternative to Bayesian analysis for accounting for the distribution of marker-QTL genotypes given marker and phenotypic information. For the designs studied, parameters were estimated accurately and QTLs mapped with satisfactory power.

ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Ina Hoeschele for giving me the opportunity to work on this project, as well as for her scholarly and financial support through my graduate program.

Also I would like to thank Drs. Eric Hallerman, Ronald Pearson, William Vinson and Dave Notter for serving on my graduate committee. Each of these gentlemen has greatly enriched my personal and educational experience at Virginia Tech.

My appreciation is also extended to secretaries of the Departments of Dairy Science at Virginia Tech and Animal and Range Sciences at Montana State University that helped me to get through the numerous "hoops" of graduate school and have been very good friends to have along the trail.

To my fellow graduate students at Virginia Tech and Montana State, a special thanks for their support, friendship and for helping me keep my sanity when things were tough.

I am also indebted to my family in Uruguay for their continuous support, love and encouragement.

Finally, I would like to dedicate this thesis to my wife, María. Her love, patience, unconditional support and friendship made this thesis much easier to accomplish.

TABLE OF CONTENTS

INTRODUCTION

Genetic progress in any population under artificial selection relies on the availability of genetic variation and the identification of superior animals. In dairy cattle, genetic improvement has been substantial (1-2% per year, Wiggans et al., 1988). Best Linear Unbiased Prediction (BLUP) of breeding values with an animal model (Henderson, 1984) is the current method of choice for genetic evaluations in most countries. It is based on the assumption of the additive infinitesimal gene model, meaning that the genetic variation is controlled by a very large number of segregating additive loci each of which has a small effect. Under this model and assuming no inbreeding, the proportion of the additive genetic variance in offspring that can be explained by variance among parental additive genetic values is only a quarter (a half for two parents). Therefore, half of the additive genetic variance or the Mendelian segregation variance, remains unexplained. Identification of allelic variants of genes that affect quantitative traits and can be traced from parent to offspring may be used to more accurately assess genetic potential in livestock species.

With the advent of new molecular genetics techniques, new methods for the detection of highly polymorphic marker systems have been developed, and better saturated genetic maps currently are becoming available for livestock species, especially for swine and cattle (see Archibald et al., 1994 for a review). A genetic map shows the most likely order of and distances between markers, estimated from the amount of recombination that

occurs between these loci in experimental crosses or reference families. With a sufficient number of markers, linkage or synteny groups of markers can be formed and assigned to chromosomes by means of physical maps. On a physical map, some of the markers, known to be located on individual chromosomes, provide landmarks for location and orientation of entire linkage groups. Markers will be in the same order on both the genetic and physical maps.

Given the availability of marker maps, researchers have begun to perform linkage analyses for detecting associations between marker loci and loci influencing quantitative traits (QTL). Several statistical methods for linkage analysis in humans, domestic animals and plants currently are being applied to map QTLs. Maximum likelihood, least squares or (multiple) linear regression, and Bayesian approaches are the most commonly used. Given the flexibility and widespread use of the Residual Maximum Likelihood (REML) methodology, this thesis concentrates on developing a derivative-free REML approach to map QTLs using multiple linked markers. Before defining more explicitly the objectives of this dissertation, a general overview of the utilization of genetic markers, mapping resources, and statistical methods for QTL mapping is presented.

Genetic Markers in Animal Breeding

A genetic marker is a polymorphic gene, or a DNA segment whose allelic variants are inherited in Mendelian fashion. Structural differences at the genetic level can be detected by the use of techniques such as restriction fragment length polymorphism

(RFLP) (Botstein et al., 1980), variable number of tandem repeats (VNTR) (Jeffreys et al., 1985; Nakamura et al., 1987), polymerase chain reaction (PCR) (Saiki et al., 1988) and random amplified polymorphic DNA (RAPD) (Williams et al., 1990).

Genetic markers can be used to identify QTLs controlling economically important traits in farm species. This information is very useful for several reasons. First, knowing the number of QTLs and the distribution of genetic effects at individual loci influencing the traits will make it possible to employ more realistic models describing the sources of phenotypic variation in traits and to increase response to selection. Second, it will provide fundamental knowledge about gene actions and interactions (Haley et al., 1991).

The use of markers in animal breeding may permit acceleration of the rate of genetic improvement via marker-assisted selection (MAS). Once loci with important effects have been identified, MAS could be used in arriving at optimum breeding decisions. Animals could be selected partially on the basis of marker genotypes in addition to phenotypic data on themselves or their relatives, or by pre-selection of young animals before collecting their own or progeny phenotypic data e.g. young bulls before entering progeny testing (Mackinnon and Georges, 1995; Kashi et al., 1990). Moreover, marker information can be used for parentage identification, evaluation of inbreeding, assessment of genetic distance, heterosis prediction, and possibly for introgression of genes of interest into recipient populations (Hillel et al., 1992). Good examples of candidate genes for introgression are the estrogen receptor (ESR) gene, with allelic variants identified between and within several U.S., European and Chinese breeds of pigs (Rothchild et al., 1994), and

the FecB gene in sheep (Piper et al., 1985). Both major genes increase litter size in pigs and sheep, respectively.

Even though the number and the distribution of genes controlling most economically important traits is unknown, the use of genetic markers can allow us to trace part of the Mendelian sampling variance associated with direct effects of large QTLs, with effects of marker loci linked to single QTLs, or with clusters of QTLs with large substitution effects (Dentine, 1990). Some important marker-QTL associations that have been found in dairy cattle to date include the Chromo-Probe, a marker linked to a QTL affecting milk, fat and protein yield in Holstein (Cowan et al., 1990); the Weaver gene in Brown Swiss causing a lethal deficit when occurring in homozygous form but also increasing milk production when in heterozygous state (Hoeschele and Meinert, 1990), and five loci affecting milk production in selected populations (Georges et al., 1995).

Estimates of increases in rate of genetic improvement via MAS have been reported in the range of 1% to 40% (Smith et al., 1991). However, most studies reported estimates on the order of 0-10% (Kashi et al., 1990; Mackinnon and Georges, 1995; Ruane and Colleau, 1995). Even though these figures look promising, several studies have suggested the conflict of applying intense MAS selection on the short-term and long-term genetic responses. Using the Fernando and Grossman model (1989), Ruane and Coulleau (1995) compared the relative efficiency of MAS over conventional BLUP. A single marked QTL and a large number of polygenes were simulated. They found that MAS increased QTL responses relative to BLUP, but with lower polygenic responses. Overall, the effect on total genetic response was slightly increased by MAS. In contrast to

the previous studies, Gibson (1994) and Pong-Wong and Wooliams (1994), for a model combining QTL and polygenic variances, reported that genotypic selection at the QTL might lead to a rapid genetic response in early stages of selection with a negative effect in the later generations due to a lower polygenic gain.

Mapping Resources in Cattle

Dairy Bull DNA Repository (DBDR)

With the purpose of intensifying the mapping of QTL associated with milk production, milk composition, and health traits, in 1992 the DBDR was established (Da et al, 1994). The DBDR currently is located at the University of Illinois at Urbana-Champaign. In 1990, semen samples from 35 grandsires and 4,705 of their sons were requested from several AI organizations. As of August 1995, nine North American AI organizations have contributed a total of 65,743 semen samples from 3,366 bulls to the DBDR. Semen samples from 17 of the 35 grandsires requested and from 1,910 of the 4,705 sons have been received. The total number of daughters of these sons having milk production records exceeds half a million.

The Illinois Reference/Resource Families (IRRF)

The IRRF, created in 1991, were established with the aim of constructing a male-specific linkage map and to map QTLs for growth traits in beef cattle. Nine paternal half-sib families (459 offspring of 9 sires and 393 dams) have been genotyped. The breeds

included are Aberdeen Angus, Gelbvieh, Simmental. and South Devon. The genome coverage is 1,975 centimorgans (cM), and the map consists of 269 loci (249 microsatellites and 20 structural genes). The average distance between contiguous loci is 9.73 cM, with 72.1% of the map intervals not exceeding 15 cM, and 4.9% of the intervals not exceeding 25 cM (Ma et al., 1996).

The Cattle Gene Mapping Project (BovMaP)

The construction of physical and genetic maps of the bovine genome is being carried out as an international collaboration. In Europe 32 laboratories are working on the EU-funded BovMaP project, which is part of a larger international collaboration involving laboratories in Australia, Africa, Israel, Japan, and the USA. A genetic map is being constructed through the use of an international bovine reference panel (IBRP) of full sib families which is distributed to participating laboratories through nodes in Australia (Commonwealth Scientific and Industrial Research Organization, CSIRO, Brisbane), Europe (Institut National de la Recherche Agronomique, INRA Jouy-en-Josas) and the USA (Texas A&M University). Data are collated at the CSIRO, and summary databases and information on cattle as well as on pig genomes are under development at the Roslin Institute (pigs) and INRA at Jouy-en-Josas (cattle) through the EU GEMINI project.

Using an international reference panel of 295 individuals in full sibling pedigrees, Barendese et al. (1994) developed a genetic linkage map for the bovine genome. Of the 202 DNA polymorphisms analyzed, 144 are microsatellites and 114 have been assigned to the physical map. There are 171 loci linked to at least one other locus, and the average

genetic distance is 15 cM. Regarding the genome coverage, of the 29 pairs of autosomes, 24 have loci physically assigned to them, and new linkage groups are being added. The 202 genotyped loci represent an estimated 90% of the expected 2,800 cM length of the male bovine genome.

Currently, the linkage map constructed using the IBRP consists of more than 400 markers (Lewin and Boling, 1994)

American Breeders Service (ABS) Global, Inc.

In 1994, ABS opened a DNA Research and Testing Laboratory. The new laboratory acquired gene mapping technology and a marker genotype database from a U.S. Holstein dairy cattle granddaughter design (GDD) with the purchase of Genmark. The original Genmark GDD consisted of 14 sires with 1518 progeny-tested sons (33 to 208 sons per sire with an average of 108) and more than 150,000 daughters (Georges et al., 1995). A linkage map was constructed based on 159 markers with 138 markers assigned to 27 linkage groups. These 27 linkage groups were in turn assigned to 24 of the previously defined 29 autosomal synteny groups. The mean distance between loci equaled 14.8 cM, and the genome coverage was estimated to be about 66%.

Mapping Resources in Sheep

SheepBase

SheepBase is an informational database that includes mapped loci in sheep, and is available on-line through the Internet (http://tetra.gig.usda.gov:8400/sheepgbase/manager.html). Even though the sheep genome project is not as advanced as the cattle or pig projects, great efforts are underway to achieve the goal of a 20cM map. Currently, the ovine genome map consists of 235 loci mapped to 24 autosomal chromosomes and both sex chromosomes, with no assignments to chromosomes 16 and 26 (Cockett and Jones, 1994).

Sheep reference families have been established in France, New Zealand and the U.S. In France, two full-sib Booroola Merino families, with 18 and 28 offspring respectively, have been produced. In New Zealand, reference familes include Texel, Coopworth, Booroola, Merino/Romney and Perendale/Coopworth animals. In the U.S., the Meat Animal Research Center (MARC), Clay Center, Nebraska, has produced four large half-sib families with 247 offspring. Current research is being focused on increasing the map density, characterization of the callipyge gene (improved carcass characteristics), search for closer markers for the spider lamb syndrome (semi-lethal congenital disorder) gene and for the Booroola FecB gene (increased litter size) (Cockett and Jones, 1994).

Mapping Resources in Swine

Three linkage maps are currently available through three major independent programs: the USDA-ARS linkage map with about 380, the Nordic map with 128, and the PiGMaP map with 240 genes and markers. Combining all sources, the pig gene map consists of about 700 genes and markers (Misfeldt and Jorgensen, 1994). PIGBASE is a computer database that includes information on papers published about gene mapping in the pig (http://www.ri.bbsrc.ac.uk/pigmap/pigbase/pigbase.html). It also includes genetic and physical maps of the pig genome.

The strategy to develop genetic (linkage) maps is to exploit divergent genetic breeds like the Chinese Meishan, the European Large White (Yorkshire), and the European Wild Boar. Reference pedigrees are established, and take the form of three-generational families, in which grandparents from genetically divergent breeds are crossed to produce the parental (F1) generation. Then, intercrosses of F1s produce the F2 generation used for linkage mapping (Archibald et al., 1994). Examples of mapped QTLs are regions of chromosome 4 that have been found to influence growth rate, fatness and intestinal length, which have been identified in a Wild Boar / Large White cross (Andersson et al., 1994). Fatness appears to be under the control of QTLs that map to the proximal end of chromosome 4. The QTLs for intestinal length and growth rate (from birth to 70 kg) are located distal to the "fatness" QTLs.

Methods for Detection and Mapping of QTL

Basic Approach and Development of Interval Mapping

Traditional approaches dealing with the detection and mapping of QTL were restricted to the use of inbred lines. Early experimental designs (e.g., Sax, 1923; Soller and Brody, 1976) for QTL detection involved the comparison of phenotypic means between offspring of a heterozygous parent for a marker locus inheriting one or the other alternative marker allele. The difference between these two means provides an estimate of the gene substitution effect (α) (Falconer, 1989) at the marker. To test whether the effect was significantly different from zero, a simple one-way ANOVA or linear regression analysis was carried out. In 1989, Lander and Botstein pointed out the main drawbacks of this method. If the marker is not the QTL, i.e. if there is incomplete linkage, the substitution effect may be underestimated, more progeny may be required to detect a given QTL effect, the most likely position of the QTL cannot be inferred, and with single markers analyzed one at a time, power of detection against type I error decreases as the number of tests increases. Therefore, Lander and Botstein (1989) proposed the use of flanking markers in what they termed "interval mapping". This method is based on the assumption that there is only one segregating QTL on a chromosome of interest: however, the method would not be optimal in the presence of linked QTL outside the interval. Evidence for a QTL is obtained by computing the LOD (logarithm of the odds) score at each position within the interval. The LOD score is a function of the size and position of the QTL, and is defined as the $\log_{10}$ of the maximum likelihood value given the presence

of a QTL over the maximum likelihood value given its absence in a particular interval. If the LOD score exceeds a pre-specified threshold, a QTL is said to be present at that particular position. By multiplying the LOD score by $2*\log_e(10)$ (≈ 4.605), it is possible to express the LOD score as the equivalent and more familiar $2\log_e$ likelihood ratio test statistic.

Classification of Gene Mapping Approaches

One criterion for grouping existing mapping methods is the type of QTL parameters they estimate, QTL substitution effects (on means) or QTL variance contributions.

Methods for estimating QTL Substitution Effects

In this group we can include ANOVA and simple linear regression (Dentine and Cowan, 1990; Haley and Knott, 1992; Weller et al., 1990), multiple linear regression (e.g., Martinez and Curnow, 1992; Moreno-Gonzalez, 1992; Zeng, 1993), Maximum Likelihood (ML) interval mapping postulating a biallelic QTL (e.g., Knott and Haley, 1992; Lander and Botstein, 1989; Weller, 1986), or a combination of ML interval mapping and regression on genotypes at other markers, termed composite interval mapping (Zeng, 1993, 1994). These methods have several drawbacks: (i) treating QTL gene effects as fixed can lead to an over-estimation of the true gene substitution effect (Georges et al., 1995; Hoeschele and VanRaden, 1993a,b; Smith and Simpson, 1986), (ii) they were developed mainly for line crosses (originated from inbred lines or from outbred populations) and, hence, are not suitable for application to outbred populations with

complex pedigree structures, varying amount of information at the QTL, and incomplete marker information (unknown marker linkage phases in the parents, uninformative offspring). Advantages of the Least Squares design are that (i) it is computationally inexpensive, allowing determination of empirical test statistic distributions via simulation or data permutation (Churchill and Doerge, 1994), and (ii) it easily can be modified to include multiple traits, multiple markers or QTLs and gene interactions. However, ML is very demanding computationally.

Methods for estimating QTL variances

If QTL gene effects are treated as random rather than fixed, estimates would be shrunken toward a prior mean more strongly for those QTLs mapped in smaller families and with smaller true gene substitution effect. With methods treating QTL gene effects as random, is possible to estimate not only the QTL effect and position, but also the QTL variance contribution to the total genetic variance. Three methods that treat QTL gene effects as random are Best Linear Unbiased Prediction (BLUP) of QTL effects (Fernando and Grossman, 1989; Goddard, 1992; Hoeschele, 1993; van Arendonk et al., 1994), Bayesian linkage analysis (Hoeschele et al., 1996; Hoeschele and VanRaden, 1993a,b), and ML based on a mixed linear model (Xu and Atchley, 1995). In addition to allowing estimation of QTL variance, these methods can accommodate more complex pedigree structures (e.g., relationships among sires or among sons through dams) both in fitting QTLs and polygenic effects. A major drawback is that computation can be very expensive, as in the case of ML and Bayesian linkage analysis.

BLUP has proven to be very advantageous in the estimation of breeding values, as non-genetic effects, pedigree information and selection can be accounted for. Fernando and Grossman (1989) showed how BLUP can be used to incorporate information on a single marker locus to predict additive effects (v) at the marked QTL and residual polygenic values (u). Goddard (1992) and van Arendonk et al. (1994) extended this method to deal with multiple linked markers and QTL. The model of Goddard (1992) can accommodate multiple linked marker loci with a single QTL in each marker interval, while the model of van Arendonk et al. (1994) considers multiple unlinked marker loci, each of which is linked to a single QTL. The original model of Fernando and Grossman (1989) assumed that every individual in the population has been genotyped at the marker. Later developments allowed for the inclusion of individuals without marker information (Hoeschele, 1993; Wang et al., 1995) and for considerable reductions in the number of equations by using a reduced animal model (RAM) (Cantet and Smith, 1991; Goddard, 1992), by including QTL equations only for genotyped animals and ancestors providing relationships among genotyped animals (Hoeschele, 1993), or by estimating the sum of the effects at several unlinked marked QTL (van Arendonk et al., 1994).

The BLUP method requires the computation of the inverse of the conditional (co)variance matrix of the QTL allelic effects, given the marker data. In outbred populations, marker data are not fully informative (e.g., parental origin of the marker alleles is unknown, marker linkage phases of parents are unknown). Wang et al. (1995) modified the method of Fernando and Grossman (1989) such that for computing the (co)variance matrix given single marker information, assigning paternal and maternal

origin to marker alleles was no longer required, and incomplete marker information was accommodated (Wang et al., 1995).

In the methods of Cantet and Smith (1991), Fernando and Grossman (1989), Goddard (1992), van Arendonk et al. (1994) and Wang et al. (1995), QTL effects are linked directly to records, and the model parameters are σ_u^2, σ_v^2, recombination rate between a QTL and a marker, and the error variance, σ_e^2. Hoeschele (1993) proposed an equivalent animal model linking records to breeding values (the sum of polygenic and QTL effects) and linking QTL effects to breeding values though the relationship matrix.

One method for estimating dispersion parameters is Residual Maximum Likelihood (REML), as suggested by Fernando and Grossman (1989) and used by van Arendonk et al. (1994). However, van Arendonk et al. (1994) found that using a single marker approach, the contribution of a QTL to additive genetic variance and its recombination rate are not separately estimable from a half-sib design with REML.

Applying Gene-Mapping Research to Dairy Cattle

It may be assumed that in dairy cattle, MAS should be applied within the current genetic evaluation systems (e.g. progeny testing, animal model evaluation). Any methodology leading to its use will have to consider MAS as a tool integrated to traditional selection strategies so that MAS can be cost-effective and acceptable to the industry. Currently, it appears that there are two major ways of applying MAS in dairy cattle populations:

1. *Via an animal model incorporating marker information in addition to production and pedigree data.* In this way it will be possible to predict breeding values by combining phenotypic and marker information (Hoeschele, 1993). The animal models proposed originally by Fernando and Grossman (1989) and Goddard (1989) are computationally infeasible for populations of the size of the US dairy cattle herd. However, by using a reduced animal model (RAM) (Cantet and Smith, 1991; Goddard, 1992) and by including QTL gene effects only for genotyped animals and their tie ancestors (Hoeschele, 1993), a large reduction in the number of equations can be achieved. Therefore, a methodological basis has been established for incorporating marker information in genetic evaluation systems.

Some important considerations before an animal model including marker data can be applied routinely are (i) dispersion parameters, including QTLs variances and positions, and polygenic variances will have to be estimated accurately, and (ii) investigation of robustness to the use of incorrect estimates of map positions of markers and QTL, additive genetic variance explained by the QTL, and number of alleles at the QTL have to be conducted. A comparison between the accuracies of genetic evaluations obtained with and without marker data is also necessary, including the case of falsely detected QTLs.

2. *Use the marker information for pre-selection of young bulls before entering progeny testing* (Kashi et al., 1990; Mackinnon and Georges, 1995). Around 75% of the genetic progress in dairy cattle populations originates from the selection of young bulls for testing (Dentine, 1990). Any increase in the accuracy of selection at this stage could have a great

impact on the rate of genetic improvement. Mackinnon and Georges (1995) proposed an alternative scheme to that of Kashi et al. (1990), where the flow of marker information is different. While Kashi et al. (1990) proposed pre-selecting young bulls based on marker allelic effects inherited from elite grand sires to grandsons or a "top-down" approach, Mackinnon and Georges (1995) proposed a "bottom-up" approach. In the latter, they suggest that once the best young bulls are selected (the ones becoming the next generation of bull sires) based on outstanding progeny-test evaluations, their daughters should be genotyped for markers linked to known QTLs. Combining genotypic and phenotypic information, it is then possible to identify young bulls heterozygous for the QTLs, if the difference in mean milk production between daughters receiving alternative marker alleles is greater than a pre-specified value. Then, the new generation of young bull progeny, which are half-sibs to those daughters, will be selected prior to progeny testing if they carry favorable marker alleles, using the number of favorable alleles or an index.

It is expected that in this way, each new generation of young bulls entering progeny testing will be genetically superior to a corresponding group without pre-selection of young bulls. This pre-selection scheme can make maximum use of reproductive technologies such as multiple ovulation and embryo transfer (MOET). With MOET, it will be possible to screen embryos, and increase or maintain the selection intensity for bull dams. The main difference between the methods of Mackinnon and Georges (1995) and Kashi et al. (1990) is that marker information becomes available at an earlier stage in the former method, allowing exploitation of the Mendelian sampling variance in full-sibs at a stage where there is not yet information in the conventional

progeny-testing scheme. A substantial advantage of this strategy is that the progeny testing scheme remains unaltered. Increases in genetic gains of 9%, 15% and 24% compared to conventional progeny-testing scheme are possible if selection is based on 1, 2, and 5 marker loci, respectively. These figures correspond to an increase in the mean value of the pre-selected bulls of .08, .13 and .21 phenotypic standard deviations, compared to a not pre-selected group. In addition, the authors also state that it is an economically feasible option.

Aim of this dissertation

The main objective of this dissertation is to develop a Residual Maximum Likelihood approach based on a mixed linear model, including polygenic effects and random QTL effects, to map QTL with multiple linked markers using a derivative-free algorithm.

In Chapter 1, we address the problem of simultaneous estimation of the QTL location, the proportion of the total additive variation associated with the QTL, and the residual polygenic variance using flanking markers rather than single markers. A granddaughter design, GDD, ignoring relationships among sires, is simulated under three genetic models differing in the number of alleles at the QTL (2, 10, and 2 times the number of sires). Assumptions are that sires are heterozygous at both marker loci, that linkage phase is known, and that the marker haplotype for each son inherited from the sire is perfectly known. The Derivative-Free Restricted Maximum Likelihood (REML) package (Meyer, 1989) was modified to accomplish this objective.

In Chapters 2 and 3, we apply the theory of Wang et al. (1995) to compute the (co)variance matrix among QTL effects allowing for relationships among sires, unknown linkage phases in the sires, homozygosity of different markers for different sires, and sons with unknown parental origin of marker alleles. Location of a single QTL and dispersion parameters are estimated under the same three QTL models mentioned above. Robustness of the analysis to the number of QTL alleles in the population also is investigated, as well as the effect of ignoring relationships among sires. Power of detecting QTLs with different variance contributions is quantified using a likelihood ratio test. For this purpose we replaced the DFREML package with our own FORTRAN 77 program, Multiple QTLs Residual Maximum Likelihood (MQREML).

In Chapter 4, the method is extended to accommodate two QTLs on the same chromosome. Different tests for the presence of a single vs. two QTLs on a chromosome are considered. Even though the method is applied to two QTLs, the extension is general for fitting multiple QTL. The MQREML program is modified to include either 1 or 2 QTL in the analysis.

Based on: (i) the two alternative strategies mentioned before for the implementation of MAS, (ii) the current availability of mapping populations, (iii) the dense cattle genome map, and (iv) the genetic evaluation system and progeny testing schemes being used in dairy cattle populations, I consider this research an important contribution to make MAS feasible in the near future.

Further research has been undertaken which is not included in this thesis. We have participated in an International QTL Workshop at the annual meeting of the International

Society for Animal Genetics in 1996 by analyzing simulated and real data with Least Squares, REML and Bayesian methods. This work is described in Uimari et al. (1996).

References

Andersson, L., Haley, C.S., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K., Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Håkansson, J. and Lundström, K. 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science 263, 1771-1774.

Archibald, A. L., Burt, D. W. and J. L. Williams. 1994. Gene mapping in farm animals and birds: An overview. Proceedings of the 5th world congress of genetics applied to livestock production, Guelph, Canada, 21:33-36.

Barendese, W. S. M. Armitage, L. M. Kossarek, A. Shalom, B. W. Kirkpatrik, A. M. Ryan, D. Clayton, L. Li, H. L. Neibergs, N. Zang, W. M. Grosse, J. Weiss, P. Creighton, F. McCarthy, M. Ron, A. J. Teale, R. Fries, R. A. McGraw, S. S. Moore, M. Georges, M. Soller, J. E. Womak and D. J. S. Hetzel. 1994. A genetic linkage map of the bovine genome. Genetics 6:227-235.

Botstein, D., White, R. L., Skolnik, M. and R. W. Davies. 1980. Construction of a genetic linkage map using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314.

Cantet, R. J. C. and C. Smith. 1991. Reduced animal model for marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 23:221-233.

Churchill, G. and R. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

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.

Cockett, N. E. and B. M. Jones. 1994. Annual Report of Cooperative Regional Projects. NRSP-8 Sheep Genome Committee Summary. Agricultural Genome Information Server.

Da. Y., M Ron, A. Yanai, M. Band, R. E. Everts, D. W. Heyen, J. I. Weller, G. R. Wiggans and H. A. Lewin. 1994. The dairy bull repository: A resource for mapping quantitative trait loci. Proceedings of the 5th world congress of genetics applied to livestock production, Guelph, Canada, 21:229-232.

Dentine, M. R. 1990. Using molecular biology to improve the accuracy of selection. 4th World Congress on Genetics Applied to Livestock Production, Edinburg, Scotland, 14:35-44.

Dentine, M. and C. M. Cowah. 1990. An analytical model for the estimation of chromosome substitution effects in the offspring of individuals heterozygous at a segregating marker locus. Theor. Appl. Genet. 79:775-780.

Falconer, D. S. 1989. Introduction to quantitative genetics. 3rd edition. Longman Scientific & Technical. England.

Fernando, R. L. and M. Grossman. 1989. Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargeant, L. S., Sorensen, A., Steele, M. R., Zhao, X., Womak, J. and I.Hoeschele. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920.

Gibson, J. P. 1994. Short-term gain at the expense of long-term response with selection of identified loci. Proceedings of the 5th world congress of genetics applied to livestock production, Guelph, Canada, 21:201-204

Goddard, M. 1992. A mixed model for analyses of data on multiple genetic markers. Theor. Appl. Genet. 83:878-886

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., Knott, S. A. and R. Thompson. 1991. Mapping quantitative trait loci in farm animals. 42nd Ann. Meeting of the EAAP, Comm: Animal Genetics, Session II, Berlin, PP.1-8.

Henderson, C. R. 1984. Applications of linear models in animal breeding. Canadian Cataloguing in Publication Data, Guelph, Canada. 423 pp.

Hillel, J., E. A. Dunnington and P. B. Siegel. 1992. DNA markers in poultry breeding and genetic analyses. Poultry Science Rev. 4: 169-186.

Hoeschele, I. 1993. Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. J Dairy Sci 76:1693-1713.

Hoeschele, I. and T. R. Meinert. 1990. Association of genetic defects with yield and type traits: The Weaver locus has a major effect on yield. J. Dairy Sci. 73:2503.

Hoeschele, I., P. Uimari, F. E. Grignola, Q. Zhang and K. M. Gage. 1996. Statistical mapping of polygene loci in livestock. In: Proceedings of the International Biometric Society (in press).

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.

Jeffreys, A. J., Wilson, V. and S. L. Hensen. 1991. Hypervariable "minisatellite" regions in human DNA. Nature 314:67.

Kashi, Y., E. Hallerman, E. and M. Soller. 1990. Marker-assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. 51:63-74.

Knott, S. A. and C. S. Haley. 1992. Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics 132:1211-1222

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

Lewin, H. A. and J. Boiling. 1994. Annual Report of Cooperative Regional Projects. NRSP-8 Cattle Genome Committee Summary. Agricultural Genome Information Server.

Lundén, A. 1991. Marker genes and production traits in domestic animals. Doctoral thesis, Swedish University of Agricultural Sciences. Department of Animal Breeding and Genetics. PO Box 7023-s-750 07 UPPSALA. Sweden.

Ma, R.Z., Beever, J.E., Da, Y., Green, C.A., Russ, I., Park, C., Heyen, D.W., Everts, R.E., Fisher, S.R., Overton, K.M., Teale, A.J., Kemp, S.J., Hines, H.C., Guerin, G. and H.A. Lewin. 1996. A male linkage map of the cattle (Bos taurus) genome (abstract). Journal of Heredity (in press)

Mackinnon, M. J. and M. A. J. Georges. 1996. A bottom-up approach to marker assisted selection. Genetics (submitted).

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

Meyer K (1989) Restricted Maximum Likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genet Sel Evol 21:317

Misfeldt, M. and N. Jorgensen. 1994. Annual Report of Cooperative Regional Projects. NRSP-8 Swine Genome Committee Summary. Agricultural Genome Information Server.

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.

Nakamura, Y., Leppert, M., O'Connell, P. O., Wolf, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616.

Piper, L. R., Bindon, B. M. and G. H. Davis. 1985. Genetics of reproduction in sheep (ed. R. B. Land and D. W. Robinson), Butterworths, London, pp. 115-125.

Rothschild, M. F., Jacobson, C., Vaske, D. A., Tuggle, C. K., Short, T. H., Sasaki, S., Eckardt, G. R. and D. G. McLaren. 1994. A major gene for litter size in pigs. Proceedings of the 5th world congress of genetics applied to livestock production, Guelph, Canada, 21:225-228.

Ruane, J. and J. J. Coulleau. 1995. Marker assisted selection for genetic improvement of animal populations when a single QTL is marked. Genet. Res., Camb. 66:71-83.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and H. A. Erlich. 1988. Primer directed amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:522.

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

Smith, C., J. A. M. Van Arendonk and R. Thompson. 1991. Possible uses of genetic markers in selection programs in cattle. 42nd Annual Meeting of the EAAP, Berlin.

Soller, M. and T. Brody. 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.

Uimari, P, Zhang, Q, Grignola, F. E., Hoeschele, I. and G. Thaller. 1996. Analysis of QTL Workshop I granddaughter design data using least-squares, residual maximum likelihood and Bayesian methods. Annual Meeting of the International Society for Animal Genetics, Tours, France

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.

Wang, T., R. L. Fernando, S. van der Beek and M. Grossman. 1995. Covariance between relatives for a marked quantitative trait locus. Genet Sel Evol 27:251.

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

Weller, J. I., Y. Kashi and M. Soller. 1990. Power of daughter and granddaughter designs for genetic mapping of quantitative traits in dairy cattle using genetic markers. J. Dairy Sci. 73:2525-2537.

Williams, J.G. K., Kubelik, A. R., Livak, K. J., Fafalski, J. A. and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Wiggans, G. R., I. Misztal and L. D. VanVleck. 1988. Implementation of an animal model for genetic evaluation of dairy cattle in the United States. J. Dairy Sci. 71 (suppl. 2):54.

Woolliams, J. A. and R. Pong-Wong. 1995. Short- versus long-term responses in breeding schemes. 46th Annual Meeting of the EAAP, Prague.

Xu, S. and W. R. Atchley. 1995. A random model approach to interval mapping of Quantitative Trait Loci. Genetics 141:1189-1197.

Zeng, Z-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Nat. Acad. Sci. 90:10972-10976.

Zeng Z-B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457-1468

Chapter 1

Empirical Best Linear Unbiased Prediction to Map QTL

F. E. Grignola and I. Hoeschele

Department of Dairy Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0315, USA

K. Meyer

Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia

Published in: **Proceedings of the 5th World Congress on Genetics Applied to Livestock Production. August 1994. Guelph, Ontario, Canada. Vol. 21:245-248**

SUMMARY

A derivative-free Residual Maximum Likelihood analysis was developed to map a QTL within a marker bracket and to simultaneously predict QTL effects and estimate QTL, additive genetic, and residual variances. The analysis was applied to a simulated granddaughter design typical for dairy cattle.

INTRODUCTION

Traditional methods for the mapping of Quantitative Trait Loci (QTL) include linear regression or ANOVA (e.g., Weller et al., 1990; Dentine and Cowan, 1990) and Maximum Likelihood (ML) (e.g., Lander and Botstein, 1989; Knott and Haley, 1992). These methods treat QTL substitution effects (α) as fixed. Then, the α of identified QTL are over-estimated, as predicted by Smith and Simpson (1986) and Hoeschele and VanRaden (1993a,b), and as verified via simulation by Georges et al. (1994). In the simulation, over-estimation increased with decreasing family size and decreasing true α. If QTL gene effects were treated as random rather than fixed, estimates would be shrunken toward a prior mean more strongly for those QTL mapped in smaller families and with smaller true α.

Two methods that treat QTL gene effects as random are Best Linear Unbiased Prediction (BLUP) of QTL effects (Fernando and Grossman, 1989; Goddard, 1992; Hoeschele, 1993) and Bayesian linkage analysis (Hoeschele and VanRaden, 1993a,b; Hoeschele, 1994). The former method can be viewed as a modification of linear

regression assuming gene effects to be normally rather than uniformly distributed a priori. The latter method can be viewed as a modification of ML assuming an exponential rather than uniform prior distribution of α at biallelic QTL and incorporating a prior probability of linkage depending on heritability.

Van Arendonk et al. (1993) found that the contribution of a QTL to additive genetic variance and its recombination rate (r) with a single marker were not separately estimable from a half-sib design in a BLUP/Residual ML (REML) analysis. Goddard (1992) presented BLUP for multiple linked markers, but assumed location and variances of QTL to be known. Here we show that QTL can be mapped and their variances estimated simultaneously from half-sib designs, when marker intervals rather than singletons are analyzed.

MATERIALS AND METHODS

The half-sib design considered here was a granddaughter design (GDD) consisting of several sires with many progeny-tested sons. In a GDD, marker genotypes are available on sires and sons, and phenotypes of sons are pre-adjusted daughter averages, termed daughter yield deviations (dyd) (VanRaden and Wiggans, 1991). The linear mixed model

$$dyd_i = \mathbf{x}_i'\beta + s_i + e_i \qquad [1]$$

was an animal model (AM) except that s_i was half the son's breeding value, β was vector of fixed effects, $\mathbf{x}_i'$ was a row i of design matrix **X**, and **e** was a residual. Model [1] represents the AM incorporating marker data of Hoeschele (1993) (model H), where data

are linked to breeding values instead of to polygenic and QTL effects in the equivalent model of Fernando and Grossman (1989) and of Goddard (1992) (model FG-G). Hence, vector **s** was augmented by those s effects of ancestors and by those QTL gene effects of sons and ancestors needed to correctly account for additive genetic relationships among sons.

Variance-covariance matrices were $\mathbf{G} = \text{Var}(\mathbf{s})$ and $\text{Var}(\mathbf{e}) = \mathbf{D}\sigma_E^2$. Matrix **D** was diagonal with element $(1 - \text{rel}_{PT(i)})/\text{rel}_{PT(i)}$, where rel_{PT} was reliability of a son's genetic evaluations contributed by its progeny test. Rel_{PT} was computed by converting total rel to total number of daughter equivalents (*de*), subtracting *de* due to parent average to obtain *de* due to progeny test and converting these back to rel (VanRaden and Wiggans, 1991). Var(**G**) was not the usual additive genetic relationship matrix times σ_s^2, but rather a variance-covariance matrix of transmitting abilities and QTL gene effects in **s**, which depended on the marker haplotypes of sons inherited from their sires. Var(**G**) was a function of three parameters, σ_s^2, r_1 or the recombination rate between the left marker (assuming only one marker interval is considered) and the QTL, and σ_v^2 or the variance due to the QTL.

These three parameters and σ_E^2 were estimated via REML by modifying the DFREML package of Meyer (1989). Model [1] or model H was chosen over FG-G, because it allowed using the single record - AM option of DFREML and because it minimizes the number of QTL equations needed, which is important when many animals without marker genotypes must be included in the analysis. Inclusion of such animals is

necessary to provide relationship ties (ancestors) and to avoid selection biases, because in a real GDD (Georges et al., 1994), many sons culled after progeny testing were not genotyped due to unavailability of semen samples. The log likelihood was (in notation of Meyer, 1989, p. 320 & p. 323)

$$\log L = -.5\log|\mathbf{G}| - .5(N - NF^{*} - NR)\log(\sigma_E^2) - .5\log|\mathbf{C}^{*}| - .5\mathbf{y}'\mathbf{P}\mathbf{y}/\sigma_E^2 \qquad [2]$$

with $\log|\mathbf{C}^{*}|$ and $\mathbf{y'Py}$ evaluated using Gaussian elimination applied to augmented mixed model equations (AMME) as shown in Meyer (1989).

The main modifications required in DFREML were the inclusion of $\mathbf{D}$ when forming the AMME, the evaluation of $\mathbf{G}^{-1}$ and of $\log|\mathbf{G}|$ in each round of iteration using the algorithm of Hoeschele (1993), and the augmentation and reparameterization of the parameter list to include total narrow sense heritability h^2, $\{v_i, i=1,q\}$, $\{r_i, i=1,q\}$, and σ_E^2, where v_i is fraction of total additive genetic variance σ_A^2 explained by QTL i with position r_i within marker bracket.

The DFREML analysis was used to demonstrate that QTL location and variance are separately estimable from half-sib designs with a marker interval, although they are confounded with a single marker (van Arendonk et al., 1993). This result was expected because for a half-sib family and a marker interval, a reduced AM (RAM) requires fitting two random marker effects representing $[(1 - r_1)/r_m]\alpha$ and $[r_1/r_m]\alpha$ (Goddard, 1992), where r_m is marker recombination rate. However, for a single marker a RAM requires fitting only one marker effect, $m = (1 - 2r)\alpha$, indicating that r and α are not separately estimable.

RESULTS AND DISCUSSION

A GDD was simulated which contained 20 sires with 50 sons each. One biallelic QTL was simulated inside a marker interval, with gene frequency of .5 and α equal to one additive genetic standard deviation. Trait heritability was $h^2 = 0.3$, and ratio of QTL to total additive genetic variance was $v^2 = .1$. Value of r_m was 0.165 and r_l was 0.052. The dyd were generated assuming constant rel_{PT} of .7. Marker haplotype inherited from the sire was known for each son. For dyd, true parameter values were $h^2 = 0.7$, $v^2 = 0.1$, $\sigma_e^2 =$ 355., and $r_l = 0.052$; estimates were 0.721, 0.097, 421., and 0.04, and these were obtained with very different sets of starting values, using the Simplex algorithm of DFREML. The point was to verify estimability of all parameters with a half-sib design and marker bracket, not to evaluate accuracy achievable with existing designs at this stage.

Future work will expand this analysis to the simultaneous mapping of several QTL in different intervals and on flanks (chromosomal ends not bracketed by markers), as well as to the estimation of QTL variances by a Bayesian approach with an informative prior distribution of QTL variances rather than by REML with a uniform prior. A comparison of this empirical BLUP analysis with the Bayesian analysis of Hoeschele and VanRaden (1993a,b) and of Hoeschele (1994) and ML interval mapping will be conducted with simulated GDDs.

REFERENCES

Fernando, R.L. and Grossman, M. 1989. Genet. Sel. Evol.21:467-477.

Georges, M., Nielsen D, Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X., Womack, J.E., and Hoeschele, I. 1994. Genetics 139:907-920.

Goddard, M. 1992. Theor. Appl. Genet. 83:878-886.

Hoeschele, I. 1994. Bayesian QTL mapping via the Gibbs sampler. Proc. 5th World Congr. Genet. Appl. Livestock Prod.

Hoeschele, I. and VanRaden, P.M. 1993a. Theor. Appl. Genet. 85:953-960.

Hoeschele, I. and VanRaden, P.M. 1993b. Theor. Appl. Genet. 85:946-952.

Knott, S. and Haley, C.S. 1992. Genetics 132:1211-1222.

Lander, E.S. and Botstein, D.S. 1989. Genetics 121:185-199.

Meyer, K. 1989. Genet Sel Evol 21:317

Smith, C. and S.P. Simpson. 1986. J. Anim. Breed. Genet. 103:205-217.

Van Arendonk, J.A.M., Tier, B. and Kinghorn, B. 1993. Proc. 17th Int. Congr. Genet., Birmingham, p.192.

Weller, J.I., Kashi, Y. and Soller, M. 1990. J. Dairy Sci. 73:2525-2537.

APPENDIX TO CHAPTER 1

This appendix is intended to present and discuss details of the method, simulation and results not presented in Grignola, Hoeschele and Meyer (1994) due to the page limit constraint.

Simulation. A granddaugther design consisting of 20 sires, 50 sons per sire and 50 daughters per son was considered. Three different QTL models, the biallelic, multiallelic and normal-effects models, were studied. Absence of selection and inbreeding was assumed.

Biallelic model: 2 QTL alleles with gene frequency p = 0.5. Sire QTL genotypes, QQ, Qq, qQ, and qq, were generated according to Hardy-Weinberg equilibrium with probabilities p^2, p(1-p), (1-p)p and $(1-p)^2$, respectively. Sire QTL allelic effects, $v^k_{sire} = \alpha_l$ with k = 1,2 identifying the two sire alleles and l = 1,2 their corresponding allelic effects, were assigned according to his genotype and the corresponding additive effect of each allele, being the difference between homozygous $2\alpha = 2\sigma_a$. The QTL allelic variance according to Falconer (1989) was

$$\sigma_v^2 = \sum_k p_k \alpha_k^2$$

and the QTL allelic variance expressed as a fraction of the total additive genetic variance (v^2) was $v^2 = .25$ or $v^2 = 0.0625$

Multiallelic model: Same as the biallelic model, but considering 10 QTL alleles with equal gene frequencies. In this model, the maximum gene substitution effect was defined

as in the biallelic situation, and $v^2 = 0.102$ which corresponded to a $v^2 = 0.25$ in the biallelic model.

Normal-effects QTL model: The number of QTL alleles was twice the number of base animals (sires). The sire QTL allelic effects were drawn from a normal distribution as $v^k_{sire} \sim N(0, \sigma^2_v)$, k = 1,2, and $v^2 = 0.25$ or $v^2 = 0.0625$. With 20 unrelated sires, there were 40 QTL alleles.

For all models, the genetic value of a son was determined by the inheritance of his paternal QTL allele, with probability depending on his marker paternal haplotype, a QTL allele from the dam sampled according to the population gene frequency, and a residual polygenic effect, u, normally distributed with mean 0 and variance σ^2_u. The phenotype was simulated by adding a normally distributed residual with mean 0 and variance σ^2_e.

Variance-covariance matrix of transmitting abilities and QTL gene effects (v)

Consider a single marker bracket with a QTL inside,

M^1_L v^1_s M^1_R

---|----------|----------|---

---|----------|----------|---

M^2_L v^1_s M^2_R

<--r_1---><--r_2----> r_1 = recombination QTL-left marker

<--------r_m--------> r_m = recombination between markers

Then, by assuming no interference between crossovers (Haldane, 1919), r_2 can be calculated from the equality $r_m = r_1 + r_2 - 2r_1r_2$. For the two marker loci, a sire will produce four types of gametes with frequencies and expected QTL effects as shown in

Table 1. Double recombination was considered when computing the expected QTL effects for each haplotype, while Goddard (1992) used approximate means assuming double recombination equal to zero.

Table 1. Frequencies and Expectations of Marker Haplotypes

Haplotype	Frequency	Expectation
11	$.5\ (1\text{-}r_m)$	$(1 - r_1r_2)\ v^1_{sire} + (r_1r_2)\ v^2_{sire}$
12	$.5\ (r_m)$	$(1 - r_1)\ r_2/r_m\ v^1_{sire} + r_1/r_m\ (1 - r_2)\ v^2_{sire}$
21	$.5\ (r_m)$	$(1 - r_2)\ r_1/r_m\ v^1_{sire} + r_2/r_m\ (1 - r_1)\ v^2_{sire}$
22	$.5\ (1\text{-}r_m)$	$(r_1r_2)\ v^1_{sire} + (1 - r_1r_2)\ v^2_{sire}$

Recurrence equations. In order to compute $\mathbf{G}^{-1}$ for a simple GDD with unrelated sires, the following recurrence equations for BV and QTL effects were considered

$$a_{sire} = v^1_{sire} + v^2_{sire} + u_{sire} \qquad Var(u_{sire}) = \sigma^2_a - 2\sigma^2_v$$

$$a_{son} = 0.5a_{sire} - 0.5v^1_{sire} - 0.5v^2_{sire} + v_{son} + e_{son} \qquad Var(e_{son}) = 0.75\sigma^2_a - 0.5\sigma^2_v$$

$$v^k_{son} = b_1 v^1_{sire} + b_2 v^2_{sire} + e^k_{son} \qquad Var(e^k_{son}) = \left(1 - b_1^2 - b_1^2\right)\sigma^2_v,\ k = 1,2$$

where b_1 and b_2 are given explicitly in the last column of Table 1. Therefore, Var(**G**) was a function of three parameters, total additive genetic variance σ^2_a, r_1, and σ^2_v, which needed to be estimated.

Results

Results in Table 2 indicate that the REML analysis converged to the same estimates with very different starting values. This result also shows that with a half-sib family design and a marker interval, r_1 and σ^2_v can be estimated separately.

Table 2. Convergence of Marker-DFREML estimation with different starting values

Parameter	Start. value	Estimate	Start. value	Estimate
h^2	0.9	0.638	0.1	0.638
v^2	0.4	0.279	0.05	0.278
r	0.1	0.042	0.002	0.042
σ_e^2	200	542.65	1500	541.63
-2log L		-4331.32		-4331.32

The results for the biallelic model are presented in Table 3, where the true parameter values were used as starting values for estimation. None of the parameters were estimated very accurately, with h^2 and r_1 being underestimated, and v^2 and σ_e^2 over-estimated.

Table 3. Estimate and SE for h^2, v^2, r_1, and σ_e^2 in a biallelic QTL model (Mean estimates and SE from 45 replicates)

Parameter	True/Starting value	Estimate	SE
h^2	0.5	0.44	0.03
v^2	0.25	0.31	0.02
r_1	0.052	0.027	0.004
σ_e^2	750	884.99	53.97

In order to check accuracy of estimation with the marker-DFREML analysis, h^2 was estimated based on the same set of replicates, but by using a "standard" DFREML

analysis without including marker information (Table 4). The same estimates were obtained with both analyses. The main reason for the inaccurate estimation of h^2 and σ_e^2 is the small design with only 20 sires. Another possible reason for the inaccuracy in estimating h^2, v^2, r_l, and σ_e^2 is that the phenotypic data come from a mixture of normal distributions of 3 genotypes, which may inflate the error variance, as well as cause the inaccuracy in h^2, r_l and v^2 as well. It must be remembered that DFREML analyses assume normal distribution of the data; however, they are generally robust to deviation from normality.

Table 4. Heritability (h^2), additive (σ_a^2) and error variance (σ_e^2) estimated with DFREML and marker-DFREML from data generated with a biallelic QTL model. (Mean estimates and SE from 45 replicates)

Parameter	True value	DFREML	Marker-DFREML
h^2	0.5	.431±.03	0.441±.03
v^2	0.25	--	0.31±.02
r_l	0.052	--	0.027±.004
σ_a^2	750	651.97±34.66	655.03±28.05
σ_e^2	750	954.31±63.99	884.99±53.97

The estimates for the biallelic model with a smaller QTL variance (v^2=.0625) are shown in Table 5. Even though the parameters for this model were not estimated very accurately, they were much closer to the true values. In this situation the mixture distribution resembles more closely a normal distribution.

good parameter estimates, it was very demanding for computing time in particular when fitting two QTLs (Uimari and Hoeschele, 1996).

Therefore, Grignola et al. (1996a) developed a Residual Maximum Likelihood method, using a deterministic, derivative-free algorithm, to map a single QTL. Hoeschele et al. (1996) showed that this method can be considered as an approximation to the Bayesian analysis fitting a normal-effects QTL model. In the normal-effects QTL model postulated by the REML analysis, the vector of QTL allelic effects is random with a prior normal distribution. The REML analysis builds on earlier work by Fernando and Grossman (1989), Cantet and Smith (1991) and Goddard (1992) on Best Linear Unbiased Prediction of QTL allelic effects by extending it to the estimation of QTL, polygenic, and residual variance components and of QTL location.

Xu and Atchley (1995) performed interval mapping using Maximum Likelihood based on a mixed model with random QTL effects, but these authors fitted additive genotypic effects at the QTL, rather than allelic effects, with variance-covariance matrix proportional to a matrix of proportions of alleles identical-by-descent, and assumed that this matrix was known. Their analysis was applied to unrelated full-sib pairs. In order to account for several QTLs on the same chromosome, Xu and Atchley (1995) used the idea behind CIT and fitted variances at the two markers flanking the marker bracket for a QTL. This approach, however, is less appropriate for multi-generational pedigrees, as effects associated with marker alleles erode across generations due to recombination. It is also

Table 5. Estimates and SE for the parameters from REML analysis of data generated with a biallelic QTL model (Mean estimates and SE from 45 replicates)

Parameter	True/Starting value	Estimate	SE
h^2	0.5	0.455	0.03
v^2	0.0625	0.078	0.01
r_l	0.052	0.040	0.01
σ_e^2	750	837.26	61.71

Results for the multiallelic model are presented in Table 6. An improvement in the accuracy of estimation is observed due to an increase in the number of alleles from 2 to 10.

Table 6. Estimates and SE for the parameters from REML analysis of data generated with a multiallelic QTL model (Mean estimates and SE from 45 replicates)

Parameter	True/Starting value	Estimate	SE
h^2	0.5	0.524	0.03
v^2	0.102	0.113	0.008
r_l	0.052	0.078	0.006
σ_e^2	750	715.19	57.46

Results for the normal-QTL allelic effects model are presented in Tables 7 and 8, for v^2 = .25 and .0625 respectively. Because this model is very similar to the model of analysis (REML assumes normality), the parameters are estimated very accurately.

Table 7. Estimates and SE for the parameters from REML analysis of data generated with a normal-effects QTL model (Mean estimates and SE from 45 replicates)

Parameter	True/Starting value	Estimate	SE
h^2	0.5	0.545	0.03
v^2	0.25	0.249	0.02
r_1	0.052	0.053	0.004
σ_e^2	750	689.64	60.52

Table 8. Estimates and SE for the parameters from REML analysis of data generated with a normal-allelic QTL model (Mean estimates and SE from 45 replicates)

Parameter	True/Starting value	Estimate	SE
h^2	0.5	0.489	0.03
v^2	0.0625	0.074	0.007
r_1	0.052	0.053	0.008
σ_e^2	750	765.31	58.6

In summary, for the granddaughter design studied here, the REML analysis showed to be robust to the number of alleles at the QTL, regarding estimation of position and contribution to the QTL to the total additive variance, error variance and heritability.

Chapter 2

MAPPING QUANTITATIVE TRAIT LOCI VIA RESIDUAL MAXIMUM LIKELIHOO D: I. METHODOLOGY

F.E. Grignola, I. Hoeschele, and B. Tier[1]

Department of Dairy Science, Virginia Polytechnic Institute and State University Blacksburg, VA 24061-0315

[1]Animal Genetics and Breeding Unit, University of New South Wales Armidale 2351, Australia

Submitted for publication in **Genetics, Selection, Evolution**

Abstract

A Residual Maximum Likelihood method, implemented with a derivative-free algorithm, was derived for estimating position and variance contribution of a single QTL together with additive polygenic and residual variance components. The method is based on a mixed linear model including random polygenic effects and random QTL effects, assumed a priori to be normally distributed. The method was developed for QTL mapping designs in livestock, where phenotypic and marker data are available on a final generation of offspring, and marker data are also available on the parents of the final offspring and on additional ancestors. The coefficient matrix of Mixed Model Equations, required in the derivative-free algorithm, was derived from a Reduced Animal Model linking single records of final offspring to parental polygenic and QTL effects. The variance-covariance matrix of QTL effects and its inverse were computed conditional on incomplete information from multiple linked markers. The inverse is computed efficiently for designs where each final offspring has a different dam and sires of final offspring have many genotyped progeny such that their marker linkage phase can be determined with a high degree of certainty. Linkage phases of ancestors of sires do not need to be known. Testing for a QTL at any position in the marker linkage group is based on the ratio of the likelihood estimating QTL variance to that with QTL variance set to zero.

Keywords: quantitative trait loci / residual maximum likelihood / mapping

INTRODUCTION

Traditional methods for the statistical mapping of Quantitative Trait Loci (QTL) include ANOVA and (multiple) linear regression (e.g., Cowan et al., 1990; Weller et al., 1990; Haley et al., 1994; Zeng, 1994), Maximum Likelihood interval mapping (e.g., Lander and Botstein, 1989; Knott and Haley, 1992), or a combination of ML and multiple regression interval mapping (e.g., Zeng, 1994). These methods were developed mainly for line crossing and, hence, cannot fully account for the more complex data structures of outcrossed 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 information on different QTLs or in different families. The gene effects near markers selected based on a linkage test tend to be increasingly over-estimated with decreasing family size and true effect (Georges et al., 1995). Random treatment of QTL effects would cause shrinkage of estimates toward a prior mean in small families and for QTLs accounting only for a small portion of genetic variance.

Fernando and Grossman (1989) derived Best Linear Unbiased Prediction (BLUP) of QTL allelic effects, which are assumed to be normally distributed. For simple designs (e.g. (grand)daughter designs with unrelated sires), BLUP reduces to random linear regression (Goddard, 1992). Fernando and Grossman (1989) showed how to obtain BLUP estimates of additive allelic effects (v) at a QTL linked to a single marker and of residual polygenic effects (u), assuming that all individuals in a population are genotyped and that markers are fully informative. Subsequent developments allowed for multiple linked markers with a QTL in

each marker bracket (Goddard, 1992), multiple unlinked markers each associated with a QTL (van Arendonk et al., 1994a), incomplete marker information (Hoeschele, 1993; van Arendonk et al., 1994a; Wang et al., 1995), and for reductions in the number of equations by using a reduced animal model (RAM) (Cantet and Smith, 1991; Goddard, 1992), by including QTL gene effects only for genotyped animals and their tie ancestors (Hoeschele, 1993), or by estimating the sum of the effects at several unlinked, marked QTL (van Arendonk et al., 1994a). There are two linearly equivalent (Henderson, 1985) animal models incorporating marker information, the first linking an individual's phenotype to both of its marked QTL allelic effects and to its polygenic effect (Fernando and Grossman, 1989), and the other linking phenotypes to the total additive effects and linking total additive effects to QTL effects via the genetic covariance matrix (Hoeschele, 1993).

All methods described above are concerned with the prediction of genetic effects and assume that the dispersion parameters are known. These parameters include the additive polygenic variance, the variance contributed by a QTL, the QTL position, and the residual variance. A first attempt to estimate these parameters by Residual Maximum Likelihood (REML) methodology was undertaken by van Arendonk et al. (1994b) using a granddaughter design with unrelated sires and a single marker. These authors found that for this situation, QTL position and contribution to additive genetic variance were not separately estimable. Grignola et al. (1994) showed that for the same type of design, these parameters were estimable when performing interval mapping with flanking markers, with known linkage phases in the sires, and no relationships among sires.

Xu and Atchley (1995) performed interval mapping using Maximum Likelihood based on a mixed model with random QTL effects, but these authors fitted one additive genetic effect at the QTL rather than two allelic effects for each individual with variance-covariance matrix equal to a matrix of proportions of alleles identical-by-descent shared by any two individuals at the QTL and assuming that this matrix was known. These authors applied their analysis to full-sib pairs that were unrelated.

In this paper, we (i) apply the theory of Wang et al. (1995) for a single marker to compute the variance-covariance matrix among QTL effects conditional on incomplete information from multiple linked markers, (ii) use this covariance matrix in the estimation of position and variance contribution of a single QTL along with polygenic and residual variances and in testing for QTL presence in a marker linkage group via REML with a derivative-free algorithm, and (iii) include all known relationships between the parents (sires) of the final offspring in the analysis.

METHODOLOGY

Mixed Linear Model

The animal model including polygenic and QTL effects of Fernando and Grossman (1989) is:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{ZTv} + \mathbf{e} \qquad [1]$$

with

$$\text{Var}(\mathbf{u}) = \mathbf{A}\sigma_u^2,\ \text{Var}(\mathbf{v}) = \mathbf{Q}\sigma_v^2,\ \text{Var}(\mathbf{e}) = \mathbf{R}\sigma_e^2$$

where **y** is an Nx1 vector of phenotypes, β is vector of fixed effects, **X** is a design/covariate matrix relating β to **y**, **u** is an nx1 vector of residual additive (polygenic) effects, **Z** is an incidence matrix relating records in **y** to animals, **v** is a 2nx1 vector of QTL allelic effects, **T** is an incidence matrix relating each animal to its two QTL alleles, **e** is vector of residuals, **A** is the additive genetic relationship matrix, σ_u^2 is polygenic variance, $\mathbf{Q}\sigma_v^2$ s is the variance-covariance matrix of the QTL allelic effects conditional on marker information, being **Q** a matrix of probabilities of QTL alleles being identical by descent, σ_v^2 is half the additive genetic variance explained by the QTL (also referred to as the QTL allelic variance), **R** is a known diagonal matrix, and σ_e^2 is residual variance.

Matrix **Q** depends on one unknown parameter, the map position of the QTL relative to the origin of the marker linkage group (d_Q). For notational convenience, this dependency is suppressed in model [1] and below. Parameters related to the marker map (marker distances and allele frequencies) are assumed to be known.

The model is parameterized in terms of the unknown parameters narrow sense heritability ($h^2 = \sigma_a^2/\sigma_p^2$) with σ_a^2 being additive genetic and σ_p^2 phenotypic variance, fraction of the additive genetic variance explained by the QTL allelic variance or half of the additive variance due to the QTL ($v^2 = \sigma_v^2/\sigma_a^2$), residual variance σ_e^2, and QTL location d_Q.

Let there be phenotypes only on non-parents or final offspring which have single records. Furthermore, recurrence equations linking u and v effects of non-parents (n) to those of parents (p) are

$$\mathbf{u}_n = \mathbf{W}\mathbf{u}_p + \mathbf{m}, \quad \mathbf{v}_n = \mathbf{F}\mathbf{v}_p + \boldsymbol{\varepsilon} \qquad [2]$$

where the matrix $\mathbf{W}$ consists of rows with zero, one or two elements equal to .5 for none, one or two parents known, respectively, each row of the matrix $\mathbf{F}$ contains up to 4 nonzero coefficients as explained below, and $\mathbf{m}$ and $\boldsymbol{\varepsilon}$ are the corresponding residual terms for $\mathbf{u}_n$ and $\mathbf{v}_n$, respectively. With single records, $\mathbf{Z} = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix. Then, model [1] can be rewritten as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{u}_p + \mathbf{T}\mathbf{F}\mathbf{v}_p + \mathbf{m} + \mathbf{T}\boldsymbol{\varepsilon} + \mathbf{e} \qquad [3]$$

$$\text{Var}(\mathbf{u}_p) = \mathbf{A}_p\sigma_u^2, \ \text{Var}(\mathbf{v}_p) = \mathbf{Q}_p\sigma_v^2, \ \text{Var}(\mathbf{m}) = \Delta_u\sigma_u^2,$$

$$\text{Var}(\boldsymbol{\varepsilon}) = \Delta_v\sigma_v^2, \ \text{Var}(\mathbf{e}) = \mathbf{R}\sigma_e^2$$

The Reduced Animal Model (RAM) is obtained from [3] by combining $\mathbf{m}$ and $\boldsymbol{\varepsilon}$ into the residual. Mixed model equations (MME) for the RAM (Cantet and Smith, 1991) can be formed based on the RAM directly or by first forming MME based on [3] and subsequently absorbing the equations in $\boldsymbol{\varepsilon}$ and $\mathbf{m}$. The resulting MME for the RAM are

$$\begin{bmatrix} \mathbf{X}'\mathbf{D}_{uv}\mathbf{X} & \mathbf{X}'\mathbf{D}_{uv}\mathbf{W} & \mathbf{X}'\mathbf{D}_{uv}\mathbf{T}\mathbf{F} \\ \mathbf{W}'\mathbf{D}_{uv}\mathbf{X} & \mathbf{W}'\mathbf{D}_{uv}\mathbf{W} & \mathbf{W}'\mathbf{D}_{uv}\mathbf{T}\mathbf{F} \\ \mathbf{F}'\mathbf{T}'\mathbf{D}_{uv}\mathbf{X} & \mathbf{F}'\mathbf{T}'\mathbf{D}_{uv}\mathbf{W} & \mathbf{F}'\mathbf{T}'\mathbf{D}_{uv}\mathbf{T}\mathbf{F} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u}_p \\ \mathbf{v}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{D}_{uv}\mathbf{y} \\ \mathbf{W}'\mathbf{D}_{uv}\mathbf{y} \\ \mathbf{F}'\mathbf{T}'\mathbf{D}_{uv}\mathbf{y} \end{bmatrix} \qquad [4]$$

where the matrix resulting from absorption is

$$\mathbf{D}_{uv} = \mathbf{D}_v - \mathbf{D}_v(\mathbf{D}_v + \Delta_u^{-1}k_u)^{-1}\mathbf{D}_v = (\mathbf{D}_v^{-1} + \Delta_u\frac{1}{k_u})^{-1},$$

$$\mathbf{D}_v = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{T}(\mathbf{T}'\mathbf{R}^{-1}\mathbf{T} + \Delta_v^{-1}k_v)^{-1}\mathbf{T}'\mathbf{R}^{-1} = (\mathbf{R} + \mathbf{T}'\Delta_v\mathbf{T}\frac{1}{k_v})^{-1},$$

$$k_u = \frac{\sigma_e^2}{\sigma_u^2}, \text{ and } k_v = \frac{\sigma_e^2}{\sigma_v^2}$$

It can be easily verified that matrix $\mathbf{D}_v$ is diagonal even if Δ_v is not, i.e., $\mathbf{T}\Delta_v\mathbf{T}'$ is always diagonal. Inbreeding and unknown parental origin of marker alleles can give rise to some non-zero off-diagonal elements in Δ_v (Wang et al., 1995; Hoeschele, 1993). With $\mathbf{D}_v$ diagonal, matrix $\mathbf{D}_{uv}$ is also diagonal; hence, the MME are easily computed.

REML Analysis

The REML analysis was performed by maximizing the likelihood of error contrasts (LEC) (Patterson and Thompson, 1971) with respect to the parameters h^2, v^2, σ_e^2, and d_Q. The LEC was obtained under the assumption of a joint multivariate normal distribution of **y**, **u**, and **v**. For the full Animal Model (AM), the logarithm of the LEC (LLEC) can be expressed as (Meyer, 1989)

$$\text{LLEC}_{AM} = \log L(\mathbf{y};\theta) = \text{const} - .5\log|\mathbf{G}| - .5(N - NF - NR)\log\hat{\sigma}_e^2 - .5\log|C| - .5\mathbf{y}'\mathbf{P}\mathbf{y}\,\hat{\sigma}_e^2$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1},\ \hat{\sigma}_e^2 = \frac{\mathbf{y}'\mathbf{P}\mathbf{y}}{N - NF} \text{ and } \mathbf{V} = \text{Var}(\mathbf{y})\frac{1}{\sigma_e^2}$$

where θ is vector of parameters, **G** is variance-covariance matrix of the random effects (here, **u** and **v**), NF = rank(**X**), NR = dimension(**G**), and **C** is coefficient matrix of the MME for the AM (model [1] or [3]) re-parameterized to full rank and with σ_e^2 factored out. The estimate of σ_e^2 maximizes the likelihood for a given set of values for the other parameters (Graser et al., 1987). The terms **y'Py** and $\log|\mathbf{C}|$ are computed as in Meyer (1989) via Gaussian elimination applied to augmented MME, and $\log|\mathbf{G}|$ is obtained as $-\log|\mathbf{G}^{-1}|$ with $\mathbf{G}^{-1}$ computed directly. In the following, it is shown how to compute the AM likelihood in [5] when working with MME for the RAM.

When equation [5] is applied directly to the RAM, the result is

$$\text{LLEC}_{\text{RAM}} \propto -.5\log|\mathbf{G}_{\text{RAM}}| - .5(\text{N} - \text{NF} - \text{NR}_{\text{RAM}})\log(\hat{\sigma}_e^2)$$

$$- .5\log|\mathbf{C}_{\text{RAM}}| - .5\mathbf{y}'\mathbf{P}\mathbf{y}\hat{\sigma}_e^{-2} \qquad [6]$$

where all parts different from the AM LLEC are subscripted RAM. Let **G** be the variance-covariance matrix of the genetic effects (**u**, **v**) of the parents and of the Mendelian sampling effects for **u** and **v** of the non-parents or finals. Let **G** be partitioned accordingly. Then

$$\log|\mathbf{G}^{-1}| = \log\begin{vmatrix} \mathbf{G}_p^{-1} & \mathbf{0} \\ \mathbf{0} & \Delta^{-1} \end{vmatrix} = \log|\mathbf{G}_p^{-1}| + \log|\Delta^{-1}| \qquad [7]$$

where

$$\mathbf{G}_p = \begin{bmatrix} \mathbf{A}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_p \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_u & \mathbf{0} \\ \mathbf{0} & \Delta_v \end{bmatrix}$$

where Δ is block-diagonal with blocks of size ≤ 2. Similarly, partition the coefficient matrix of the MME for model [3], **C**, according to all other (1: β, $\mathbf{u}_p$, $\mathbf{v}_p$) and Mendelian sampling effects (2: **m**, ε). Then,

$$\log|\mathbf{C}| = \log|\mathbf{C}_{22}| + \log|\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}| \quad [8]$$

where $\mathbf{C}_{22}$ is diagonal or blockdiagonal with blocks of size ≤ 2. Hence, the RAM LLEC can easily be modified to yield the AM LLEC, or

$$\text{LLEC}_{AM} \propto \text{LLEC}_{RAM} - .5\log|\Delta| - .5\log|\mathbf{C}_{22}| + .5(NR - NR_{RAM})\log(\hat{\sigma}_e^2) \quad [9]$$

where NR is total number of random genetic effects while NR_{RAM} is number of genetic effects pertaining to parents.

The analysis was conducted in the form of interval mapping as in Xu and Atchley (1995), where d_Q was fixed at a number of successive positions (every centimorgan) along the chromosome, and at each position the likelihood was maximized with respect to h^2, v^2, and σ_e^2

Calculation of $\mathbf{Q}_p^{-1}$ and Δ_v

These matrices were computed by applying the theory presented in Wang et al. (1995) to marker information consisting of multiple linked markers rather than single markers. At a given QTL position, different markers were allowed to flank the QTL in different families due to some parents being homozygous at the closest flanking markers.

Notation. Let Q_i^k denote QTL allele k (k=1,2) in individual i and v_i^k the additive effect of this allele. Let ≡ denote identity by descent, let ⇐ stand for "inherited from", let $\mathbf{G}_{obs}$ represent the marker information observed on the pedigree, and let M_i^m denote a possible marker haplotype (m) of individual i at the closest pair of marker loci bracketing the QTL for which the parent of i is heterozygous. Furthermore, let **M** be a set of complete multi-locus marker genotypes for the entire pedigree. Finally, p denotes parent (p=s,d), s sire, d dam, and L_p denotes the linkage phase of the alleles at the narrowest marker bracket for which parent p is heterozygous.

Variance-covariance matrix of v effects $\mathbf{Q}_p$. In the presence of missing marker data and/or unknown linkage phases for parents, the variance-covariance matrix of the v effects is of the form

$$\mathbf{Q}_{G_{obs}} = \sum_{\mathbf{M}} \Pr(\mathbf{M} \mid \mathbf{G}_{obs})\, \mathbf{Q}_M \qquad [10]$$

where $\mathbf{Q}_M$ is conditional on a particular set of multi-locus marker genotypes (**M**). Equation [10] was given in Hoeschele (1993) and in Wang et al. (1995). The calculation of [10] is computationally very demanding for large pedigrees. The probability of a QTL allele in individual i being identical by descent (IBD) to a QTL allele in individual j (with j not being a direct descendant of i) in general cannot be computed recursively using IBD probabilities pertaining to the alleles in i and the parents of j when parental marker genotypes and/or linkage phases are unknown (Wang et al., 1995), hence there is no simple method to compute the inverse directly. A method for computing the inverse, which is more efficient than standard inversion, was derived by Van Arendonk et al. (1994a).

The variance-covariance matrix in [10] can, however, be computed by using Monte Carlo. The Monte-Carlo approximation of [10] is

$$\mathbf{Q}_{G_{obs}} \approx \frac{1}{S} \sum_{k=1}^{k=S} \mathbf{Q}_{M_k} \qquad [11]$$

where $\mathbf{M}_k$ is a particular realization of $\mathbf{M}$ from the probability distribution of $\mathbf{M}$ given $\mathbf{G}_{obs}$, and S is sample size. Note that [11] yields the exact variance-covariance matrix if sample size S is large. Samples from this distribution can be obtained by Gibbs sampling, which was implemented using blocking of the genotypes of parents and final offspring (finals) as in Janss et al. (1995). For a half-sib design (daughter or granddaughter design) with large family sizes (e.g., 50 to 100) and no relationships among final offspring (daughters or sons) through dams, the linkage phases of the parents of final offspring are "known", as always or most frequently (near 100%) the correct phase is sampled. Then, the inverse of the variance-covariance matrix of the QTL effects can be computed exactly (up to Monte Carlo error due to use of [11]) as follows. Equation [11] is employed to compute the sub-matrix pertaining to QTL effects of parents of finals and ancestors using marker information on the entire pedigree including final offspring. This sub-matrix is then inverted, and contributions of final offspring, computed with known parental linkage phases, are added into the inverse. Note that in the RAM in [4], offspring contributions appear in the least-squares part of the MME rather than in the inverse variance-covariance matrix of the QTL effects.

Recurrence equations for v effects. Recurrence equations for the v effects of the finals were required to compute the elements of $\mathbf{F}$ and Δ_v in [4]. The general recurrence equation for a QTL effect is

$$v_i^k = \sum_{p=s,d} \sum_{l=1,2} \Pr(Q_i^k \equiv Q_p^l | \mathbf{G}_{obs}) v_i^p + \varepsilon_i^k \qquad [12]$$

where

$$\Pr(Q_i^k \equiv Q_p^l | \mathbf{G}_{obs}) = \sum_{L_p} \sum_{m} \Pr(L_p, M_i^m \Leftarrow p | \mathbf{G}_{obs}) \Pr(Q_i^k \equiv Q_p^l | M_i^m \Leftarrow p, L_p) \quad [13]$$

The most likely linkage phase is assumed to be the true phase for the parents of final offspring. This assumption reduces the joint probability of parental linkage phase and offspring haplotype in [13] to the probability of the marker haplotype of an offspring. This probability is computed using the parental phase and the marker genotypes of an offspring at all linked markers. Alternatively, [13] could be used when parental linkage phases are not known by computing the joint probability of parent linkage phase and offspring haplotype for each interval as a frequency count across all Gibbs cycles after burn-in, using information from the entire marker linkage group and from all relatives in the pedigree. However, this approach would be only an approximation to calculating the variance-covariance matrix and its inverse based on [10] for the entire pedigree including the final offspring.

In [13], the $\Pr(Q_i^k \equiv Q_p^l | M_i^m \Leftarrow p, L_p)$ are $t_{11} = (1-r_L)(1-r_R)/(1-r_M)$ and $t_{12} = r_L r_R/(1-r_M)$ if M is a non-recombinant haplotype, or $t_{21} = (1-r_L)r_R/r_M$ and $t_{22} = r_L(1-r_R)/r_M$ if M_i^m is recombinant, where r_M is recombination rate for the marker bracket, r_L (r_R) is recombination rate between the QTL and the left (right) marker, and Haldane's no interference map function is employed. Here, we allow for double recombination while Goddard (1992) assumed it to be zero.

QTL alleles in final offspring are identified by parental origin, i.e. the two QTL alleles in an offspring are distinguished as the allele inherited from the sire (s) and the allele coming from the dam (d). This definition can be employed even if the parental origins of the alleles at the flanking markers are unknown, but it can be used only in the final generation. For illustration, consider a single parent p (here, p = s = sire) with genotype 12/12, linkage phase 1-1, and the worst case of an offspring with genotype 12/12 (inheritance unknown at both flanking markers). The possible marker haplotypes inherited from p are 1-1, 1-2, 2-1, and 2-2. Then, if the QTL alleles in i are identified by the alleles at the left marker (1,2):

$$v_i^1 = [\ \Pr(1\text{-}1 \Leftarrow s)t_{11} + \Pr(1\text{-}2 \Leftarrow s)t_{21}\ ]\ v_s^1 + [\ \Pr(1\text{-}1 \Leftarrow s)t_{12} + \Pr(1\text{-}2 \Leftarrow s)t_{22}\ ]\ v_s^2$$
$$+ [\ \Pr(2\text{-}1 \Leftarrow s) + \Pr(2\text{-}2 \Leftarrow s)\ ].5\left(v_d^1 + v_d^2\right) + \varepsilon_i^1$$

$$v_i^2 = [\ \Pr(2\text{-}1 \Leftarrow s)t_{22} + \Pr(2\text{-}2 \Leftarrow s)t_{12}\ ]\ v_s^1 + [\ \Pr(2\text{-}1 \Leftarrow s)t_{21} + \Pr(2\text{-}2 \Leftarrow s)t_{11}\ ]\ v_s^2$$
$$+ [\ \Pr(1\text{-}1 \Leftarrow s) + \Pr(1\text{-}2 \Leftarrow s)\ ].5\left(v_d^1 + v_d^2\right) + \varepsilon_i^2$$

whereas if the QTL alleles in i are identified by parental origin,

$$v_i^1 = [\ \Pr(1\text{-}1 \Leftarrow s)t_{11} + \Pr(1\text{-}2 \Leftarrow s)t_{21} + \Pr(2\text{-}1 \Leftarrow s)t_{22} + \Pr(2\text{-}2 \Leftarrow s)t_{12}\ ]\ v_s^1$$
$$+ [\ \Pr(1\text{-}1 \Leftarrow s)t_{12} + \Pr(1\text{-}2 \Leftarrow s)t_{22} + [\ \Pr(2\text{-}1 \Leftarrow s)t_{21} + \Pr(2\text{-}2 \Leftarrow s)t_{11}\ ]\ v_s^2 + \varepsilon_i^1$$

$$v_i^2 = .5v_d^1 + .5v_d^2 + \varepsilon_i^2$$

Note that summing the v_i^1 and v_i^2 equations yields the same result for both QTL identifications. Note also that the advantage of the identification by parental origin is that only v is linked to the v-effects of the dam (d), instead of linking both v_i^1 and v_i^2 to the dam effects required to include dam effects in the MME.

Hypothesis Test

The likelihood under the null hypothesis is evaluated at $v^2 = 0$. The distribution of the likelihood ratio statistic is not known exactly, regardless of the method used to locate QTL (Churchill and Doerge, 1994). For the null hypothesis postulating the absence of a QTL in a particular interval rather than in the entire genome, Xu and Atchley (1995) found the distribution to be between two chi-square distributions with degrees of freedom of one and two, respectively. Several factors may influence the distribution of a test statistic for QTL presence, e.g., the length of the genome, the marker density, the extent to which marker data are missing, segregation distortion, and the distribution of the phenotypes. Self and Liang (1987) derived analytical results for the asymptotic distribution of the likelihood ratio statistic for cases where the true parameter value may be on the boundary of the parameter space. However, with finite sample sizes and several factors influencing the distribution of the statistic, it is questionable whether their results can be utilized in QTL mapping.

When analyzing real data, the threshold value for significance can be determined empirically using data permutation (Churchill and Doerge, 1994). To obtain the threshold value for a genome-wide search, on the order of 10,000 to 100,000 permutations are necessary. As

these computations are unfeasible with the method presented here (see the companion paper by Grignola et al., 1996), one may resort to estimating thresholds for a number of less stringent significance levels and obtain the desired threshold by extrapolation (Uimari et al., 1996b).

CONCLUSIONS

The REML analysis described in this paper may be a useful alternative to other methods for the statistical mapping of QTL. The REML method is generally known to be quite robust to deviations from normality. When applied to QTL mapping, the REML analysis requires fewer parametric assumptions than ML (e.g., Weller, 1986) and Bayesian analyses (Thaller and Hoeschele, 1996a,b; Uimari et al., 1996a; Hoeschele et al., 1996) postulating a biallelic QTL with unknown gene frequency and substitution effect.

While Xu and Atchley (1995) estimate QTL, polygenic and residual variances by ML, we perform REML estimation. While REML should be preferred over ML in the presence of many fixed effects relative to the number of observations (Patterson and Thompson, 1971), a model for the analysis of QTL mapping experiments may only need to include an overall mean. In this case, the difference between the ML and REML analyses is negligible.

As the true nature of QTLs is unknown, it is important to evaluate the performance of this REML analysis and of other methods with data simulated under different genetic models (e.g., biallelic and multiallelic QTL models). In a companion paper (Grignola et al., 1996), we apply the REML analysis to granddaughter designs simulated with different models for the

additive variance at the QTL. Hoeschele et al. (1996) apply Bayesian analyses based on biallelic and multiallelic QTL models to data simulated under both models.

The REML analysis incorporates an expected variance-covariance matrix of the QTL allelic effects, which is equal to a weighted average of variance-covariance matrices conditional on all possible sets of multi-locus marker genotypes given the observed marker data. Alternatively, Schork (1993) formulated a likelihood for a mixture distribution which is a weighted average of REML likelihoods conditional on all possible sets of multi-locus marker genotypes given the observed marker data. He pointed out, however, that simulation results indicated that his modification may lead to a loss of power. In both approaches, the one considered in this paper and in equivalent form by Xu and Atchley (1995), and the approach of Schork (1993), probabilities of multi-locus marker genotypes are computed from the observed marker information. However, if markers are linked to QTLs, phenotypes also contain information about marker genotypes, and this information is ignored here (Van Arendonk, personal communication). In this regard, the REML analysis can be viewed as an approximation to the Bayesian analysis based on a multiallelic QTL model with QTL variance and allelic effects having a prior normal distribution (Hoeschele et al., 1996). The Bayesian analysis takes into account the joint distribution of the QTL and marker genotypes conditional on the phenotypic information.

We are currently extending our REML analysis to account for multiple linked QTLs. One way of approaching this problem was presented by Xu and Atchley (1995) and consisted of fitting variances associated with next-to-flanking markers. Disadvantages of this approach are that it is approximate, as effects associated with marker alleles identified within founders

erode over generations, and that it requires many additional parameters when the marker polymorphism is limited, causing the flanking and next-to-flanking markers to differ among families.

Finally, we plan to extend the REML analysis to other designs (e.g., full-sib designs), where the current computation of the inverse of the variance-covariance matrix becomes approximate due to uncertain linkage phases in parents of final offspring, and to other ways of computing this inverse exactly (e.g., Van Arendonk et al., 1994a).

ACKNOWLEDGEMENTS

This research was supported by Award No. 92-01732 of the National Research Initiative Competitive Grants Program of the US Department of Agriculture and by the Holstein Association USA. I. Hoeschele acknowledges financial support from the European Capital and Mobility fund while on research leave at Wageningen University, The Netherlands. B. Tier acknowledges financial support from the Australian Department of Industry, Training, and Regional Development while on research leave at Virginia Polytechnic Institute and State University.

REFERENCES

Cantet RJC, Smith C (1991) Reduced animal model for marker assisted selection using best linear unbiased prediction. Genet Sel Evol 23:221-233

Churchill G, Doerge R (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

Cowan CM, Dentine MR, Ax RL, Schuler LA (1990) Structural variation around prolactin gene linked to quantitative traits in an elite Holstein family. Theor Appl Genet 79:577-582

Fernando RL, Grossman M (1989) Marker-assisted selection using best linear unbiased prediction. Genet Sel Evol 21:467

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, Womack JE, Hoeschele I (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920

Goddard M (1992) A mixed model for analyses of data on multiple genetic markers. Theor Appl Genet 83:878-886

Graser H-U, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. J Anim Sci 64:1363-1370

Grignola FE, Hoeschele I, Meyer K (1994) Empirical best linear unbiased prediction to map QTL. Proc 5th World Congr Genet Appl Livest Prod, Vol 21:245-248

Grignola FE, Hoeschele I, Thaller G (1996) Mapping quantitative trait loci via residual maximum likelihood: II. A simulation study. Genet Sel Evol (in press)

Haley CS, Knott SA, Elsen J-M (1994) Mapping quantitative trait loci in crosses between outbred lines using least-squares. Genetics 136:1195-1207

Henderson CR (1985) Equivalent linear models to reduce computations. J Dairy Sci 68:2267
Hoeschele I (1993) Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. J Dairy Sci 76:1693-1713

Hoeschele I, Uimari P, Grignola FE, Zhang Q, Gage KM (1996) Statistical mapping of polygene loci in livestock. In: Proceedings of the International Biometric Society (in press)

Janss LLG, Thompson R, van Arendonk JAM (1995) Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. Theor Appl Genet 91:1137-1147

Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics 132:1211-1222

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

Meyer K (1989) Restricted Maximum Likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genet Sel Evol 21:317

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545-554

Schork NJ (1993) Extended multi-point identity-by-descent analysis of human quantitative traits: Efficiency, power, and modeling considerations. Am J Hum Genet 53:1306-1319

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio statistics under nonstandard conditions. J. Amer. Stat. Assoc. 82:605-610

Thaller G, Hoeschele I (1996a) A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. Methodology. Theor Appl Genet (in press)

Thaller G, Hoeschele I (1996b) A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. A simulation study. Theor Appl Genet (in press)

Uimari P, Thaller G, Hoeschele I (1996a) A Monte Carlo method for Bayesian analysis of linkage between multiple linked markers and a quantitative trait locus. Genetics (submitted)

Uimari P, Zhang Q, Grignola F, Hoeschele I, Thaller G (1996b) Analysis of QTL Workshop I granddaughter design data using least-squares, residual maximum likelihood and Bayesian methods. Annual Meeting of the International Society for Animal Genetics, Tours, France

Van Arendonk JAM, Tier B, Kinghorn BP (1994a) Use of multiple genetic markers in prediction of breeding values. Genetics 137:319-329

Van Arendonk JAM, Tier B, Kinghorn BP (1994b) Simultaneous estimation of effects of unlinked markers and polygenes on a trait showing quantitative genetic variation. Proc 17th Int Congr Genetics, Birmingham, UK, p. 192

Wang T, Fernando RL, van der Beek S, and M Grossman (1995) Covariance between relatives for a marked quantitative trait locus. Genet Sel Evol 27:251

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

Weller JI, Kashi Y, Soller M (1990) 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

Xu S, Atchley WR (1995) A random model approach to interval mapping of Quantitative Trait Loci. Genetics 141:1189-1197

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468

Chapter 3

MAPPING QUANTITATIVE TRAIT LOCI VIA RESIDUAL MAXIMUM LIKELIHOOD: II. A SIMULATION STUDY

F.E. Grignola, I. Hoeschele, Q. Zhang and G. Thaller[1]

Department of Dairy Science, Virginia Polytechnic Institute and State University Blacksburg, VA 24061-0315

[1] Technical University of Munich at Weihenstephan
85350 Munich, Germany

Submitted for publication in **Genetics, Selection, Evolution**

Abstract

Position and variance contribution of a single QTL together with additive polygenic and residual variance components were estimated using a Residual Maximum Likelihood method and a derivative-free algorithm. The variance-covariance matrix of QTL effects and its inverse were computed conditional on incomplete information from multiple linked markers. Simulation was employed to investigate the accuracy of parameter estimates and likelihood ratio tests. The design was a granddaughter design with 2000 sons, 20 sires of sons and 9 ancestors of sires. Designs with 1000 and 600 sons were also investigated. Data were simulated under three different genetic models for the QTL, a biallelic model, a multiallelic model with 10 alleles at equal frequencies, and a model with normally and independently distributed QTL allelic effects for base individuals. The trait analyzed was Daughter Yield Deviation or the daughter average adjusted for environmental effects and merits of mates of the sons. Genotypes for five markers situated on the same chromosome were generated for all sons and their ancestors. Data were analyzed with and without relationships among sires. Only small differences between analyses with known, most likely, or unknown linkage phases in the ancestors were detected. Parameters were estimated with good accuracy under all three simulation models. The REML method was fairly robust to the number of alleles at the QTL for the designs studied.

Keywords: quantitative trait loci / residual maximum likelihood / mapping / simulation / granddaughter design

INTRODUCTION

In a companion paper (Grignola et al., 1996), a Residual Maximum Likelihood (REML) method was derived for estimating position and variance contribution of a single QTL together with additive polygenic and residual variance components. The REML analysis was implemented with a derivative-free algorithm. The method overcomes shortcomings of the traditional methods of linear regression (e.g., Haley et al., 1994; Zeng, 1994) and Maximum Likelihood (ML) interval mapping (e.g., Weller, 1986; Lander and Botstein, 1989; Knott and Haley, 1992). ML and regression methods cannot fully account for the more complex data structures of outcrossed 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 QTLs or in different families.

The REML method is based on a mixed linear model including random polygenic effects and random QTL effects. Polygenic effects represent the sum of the additive effects at all loci not linked to the markers. The QTL allelic effects are assumed to have a prior normal distribution with variance-covariance matrix conditional on information from multiple linked markers. Because the true nature of the QTLs is unknown, i.e. the number of alleles at a QTL in the population studied is unknown, the robustness of the REML analysis to the number of alleles at the QTL must be evaluated.

One of the main experimental designs for QTL mapping in livestock is the half-sib design, used in cattle in the form of daughter or granddaughter designs (Weller, 1990).

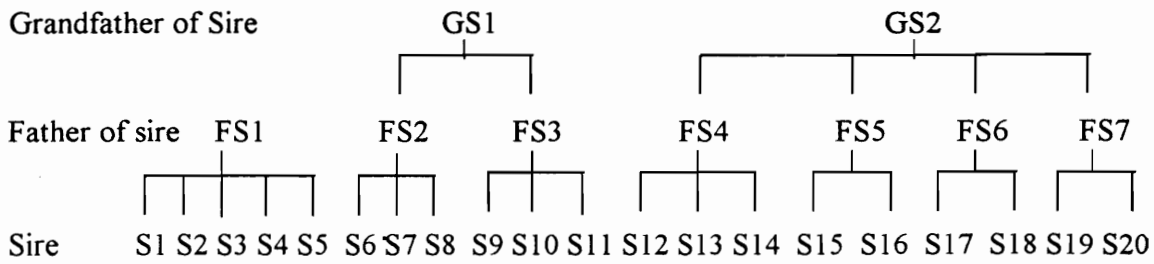
In this paper, we evaluate the accuracy of the REML analysis in QTL mapping using granddaughter designs. The simulated designs resemble actual designs for the U.S. Holstein population. Data are simulated under several genetic models differing in the number of QTL alleles. The analysis is carried out with and without consideration of relationships among sires. Here, we only present simulation results. The analysis is described in detail in the companion paper (Grignola et al., 1996).

SIMULATION

Design

The most frequently used design for mapping QTL in dairy cattle is the granddaughter design (GDD), where marker genotypes are collected on sons and phenotypes on daughters of the sons. A GDD was simulated with a pedigree structure resembling the real GDD of the US public gene mapping project for dairy cattle based on the Dairy Bull DNA Repository (Da et al., 1994). The simulated GDD consisted of 2000 sons, 20 sires, and 9 ancestors of the sires (Figure 1), and is identical to the design used in the Bayesian linkage analyses of Thaller and Hoeschele (1996a,b) and Uimari et al. (1996a). While this design had 100 sons per sire, designs with only 50 or 30 sons per sire also were simulated.

Figure 1: Pedigree of granddaughter design



The phenotype simulated was Daughter Yield Deviation (DYD) of sons (VanRaden and Wiggans, 1991). DYD is an average of the phenotypes of the daughters adjusted for systematic environmental effects and genetic values of the daughter's dams. Total variance of DYD equals $\text{Var(DYD)} = .25\sigma_a^2 / R$ and can be factored into

$$\text{Var(DYD)} = .25\sigma_a^2 + [(1-R)/R]*.25\sigma_a^2$$

where R is reliability or squared accuracy of the son's estimated breeding value contributed by its progeny test, and σ_a^2 is the additive genetic variance. In this factorization, the first component is the variance among the sons' transmitting abilities or half of their additive genetic values, and the second term is "residual variance" or the variance of the average dam and Mendelian genetic effect of the daughters and the average environmental effect. Variance of DYD can be rewritten as

$$\text{Var(DYD)} = .25\sigma_a^2 + w*\sigma_e^2$$

where $w = (1-R)/R$ and $\sigma_e^2 = .25\sigma_a^2$. Therefore, when analyzing DYD with the weight w, DYD has an expected heritability of .5, because the expected value of the estimate of σ_e^2 is

$.25\sigma_a^2$. Hence, there is the option of treating heritability as known in the REML analysis when phenotypic data are DYDs.

Marker and QTL genotypes were simulated according to Hardy-Weinberg frequencies and the map positions of all loci. One linkage group was considered which consisted of five marker loci and one QTL. Each marker locus had five alleles at equal frequencies, with the exception of one design where each marker had only three alleles at equal frequencies. The markers were spaced 20cM apart and, for the results presented here, the QTL was located in interval 3 at 5cM from the left marker (other QTL positions were simulated to verify that the analysis was working properly).

Polygenic and QTL effects were simulated according to the pedigree in Figure 1. Data were analyzed first (i) using full pedigree information, and alternatively (ii) assuming as is common practice that the 20 sires in Figure 1 were unrelated. The QTL contribution to the DYDs of sons was generated by sampling individual QTL allelic effects of daughters under each of the genetic models as described below. This sampling of QTL effects assures that DYD of a heterozygous son or of a son with substantial difference in the additive effect of its two QTL alleles has larger variance among daughters due to the QTL than a homozygous son or a son with similar QTL allelic effects.

Genetic models

Three different genetic models were used to simulate data. Common to all models were the parameters narrow sense heritability of individual phenotypes $h^2 = .3$, phenotypic SD $\sigma_p =$

100, and the order of loci and recombination rates among all loci. Under all three models, phenotypes were simulated as

$$DYD_i = \frac{1}{n_i}\sum_{j=1}^{n_i} g_{ij} + u_i + e_i; \quad Var(e_i) = \frac{1}{n_i}(.75\sigma_u^2 + \sigma_e^2)$$

where n_i was the number of daughters of son i, g was the sum of the v effects in daughter j of son i, u was a normally distributed polygenic effect, e was a normally distributed residual, polygenic variance (σ_u^2) was equal to the difference between additive genetic variance (σ_a^2) and the variance explained by the QTL ($2\sigma_v^2$), and σ_e^2 was environmental variance. Number of daughters per son was set to 50, corresponding to a Reliability (VanRaden and Wiggans, 1991) near .8. The ratio of the QTL allelic variance (σ_v^2) to the additive genetic variance (σ_a^2) is denoted by v^2 below.

Model 1: Normal-effects model. For each individual with both or one parent(s) unknown, both or one QTL effect(s), respectively, were drawn from N(0, σ_v^2). For the pedigree in Figure 1, there were 32 distinct base alleles, and the QTL was treated as a locus with 32 distinct alleles in passing on alleles to descendants. The parameter σ_v^2 was set to $.25\sigma_a^2$ or $.0625\sigma_a^2$, i.e., the simulated QTL accounted for 50% ($2v^2$=.5) or 12.5% ($2v^2$=.125) of the total additive genetic variance, respectively. In an additional simulation, v^2 was set to 0 to obtain the empirical distribution of the test statistic under the null hypothesis.

Model 2: Multiallelic model. The QTL had ten alleles with equal frequencies. For the biallelic QTL with $2v^2 = .5$ (see below), the difference among homozygotes was $2a = 2\sigma_a$, being a the genotypic value of one homozygote. For the multiallelic QTL, means of the ten homozygous genotypes ranged from $-\sigma_a$ to σ_a at equal intervals. Means of heterozygotes were calculated assuming additive gene action. Given these means (m), additive effects of alleles were determined as

$$\alpha_i = p\sum_{j=1}^{j=10}\mu_{ij}$$

where $p_1 = p_2 = \ldots = p_{10} = p = .1$. The variance at the QTL was

$$\sigma^2_{QTL} = 2v^2\sigma^2_a = 2p\sum_{i=1}^{i=10}\alpha_i^2$$

which yielded a value of $2v^2 = .204$.

Model 3: Biallelic model. The QTL was biallelic with allele frequency of .5. The variance at the QTL was

$$2\sigma^2_v = 2v^2\sigma^2_a = 2p(1-p)\alpha^2$$

where for $p = .5$ and $2v^2 = .5$ or $2v^2 = .125$, QTL substitution effect a was determined and used to compute the additive effects of the two QTL alleles as $-p\alpha$ and $(1-p)\alpha$.

RESULTS

The REML analysis for single QTL mapping using all markers on a chromosome is described in detail in the companion paper (Grignola et al., 1996). Analyses were performed

with and without considering relationships among sires and with the marker linkage phases of sires and ancestors known or unknown. For the designs considered here, the most probable (more than 90%) linkage phase of the sires was always the true phase, but phase probabilities calculated for the ancestors indicated more uncertainty about their true phases. When phases were treated as unknown, the analysis employed equation [11] of Grignola et al. (1996) to calculate the inverse of the variance-covariance matrix of the QTL effects of ancestors and sires. Contributions from sons were calculated assuming that the most likely sire phases equaled the true phases.

Analysis of a single data set took around eight minutes of computing time on an IBM SP2 system with RS6000 390 and 590 nodes. In a preliminary investigation, the REML analysis, using a derivative-free Simplex algorithm was started from very different initial values for the parameters to verify convergence to the same estimates. As an example, a particular data set simulated under the biallelic QTL model ($2v^2 = .5$) was analyzed using the two starting value sets [.9, .45, .6, 200.] and [.1, .05, .4, 1500.] for [h^2, v^2, d_Q, σ_e^2] with the parameters defined in Table 1. Parameter estimates and likelihood for the first starting value set were .6380, .2790, .4434, 542.6, and -4331.3. Corresponding figures for the second starting value set were .6380, .2780, .4440, 541.6, and -4331.3.

Figure 2 contains cumulative distribution functions with 90th percentiles of 2.706 and 4.605 for chi-square distributions with 1 and 2 degrees of freedom, respectively. The Figure also contains an empirical distribution function with 90th percentile of 3.746 for the likelihood ratio statistic from 1,000 replicates generated without a QTL. Xu and Atchley (1995) found

the distribution to be close to the 2 df chi-square, while here the empirical distribution was in between the 1 df and 2 df chi-square distributions and slightly closer to the 2 df distribution.

Parameter estimates and likelihood ratio statistics are defined in Table 1 and results are presented in Tables 2 to 6. All results are based on 50 replicates. In Table 2, results for the normal-effects QTL model with $2v^2 = .5$ are presented. Several different analyses, described in Table 2, were conducted. First, sires were simulated as unrelated and analyzed without relationships. Then, relationships among sires were simulated according to Figure 1. In analyses of these data, sires were treated as unrelated, treated as related with known marker linkage phases (i.e., the true phases were used for all sires and ancestors), treated as related with the

Figure 2: Cumulative distribution functions with 90th percentiles for chi-square distributions with 1 and 2 degrees of freedom, respectively, and empirical cumulative distribution function with 90th percentile of the likelihood ratio statistic from 1,000 replicates generated under the null hypothesis of no QTL.

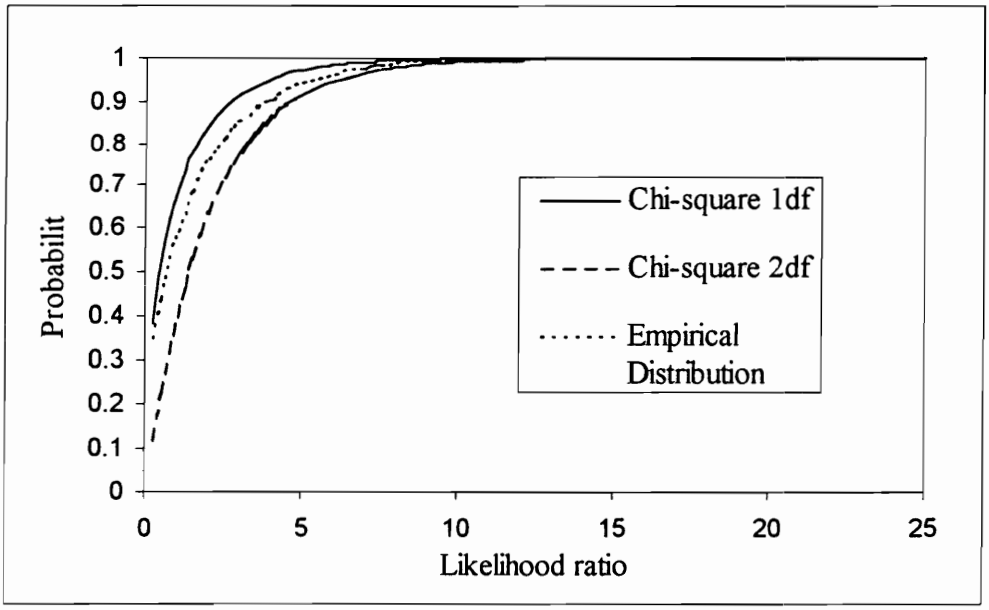


Table 1: Definition of model parameters and likelihood ratio

Symbol	Parameter
h^2	Narrow sense heritability
v^2	Ratio of QTL to additive genetic variance
d_Q	QTL location in cMorgan
s	Residual variance
LR	Likelihood ratio test statistic = $-2\log[L(H_0:v^2=0)/L(H_A:v^2\neq 0)]$

Table 2. Means and their SE for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the normal-effects QTL model explaining 50% of the additive genetic variance (50 replicates).

Parameter[1]	True value	Analysis I[2] Analysis II[3]	Analysis III[4] Analysis IV[5]	Analysis V[6]	Analysis VI[7] Analysis VII[8]
h^2	.5	.558 (.037) .537 (.039)	.558 (.039) .544 (.038)	.541 (.037)	.494 (.034) .5 (fixed)
v^2	.25	.243 (.015) .279 (.018)	.280 (.015) .279 (.015)	.280 (.015)	.288 (.016) .262 (.013)
d_Q	.4527	.461 (.004) .451 (.005)	.450 (.005) .450 (.005)	.449 (.005)	.444 (.005) .449 (.005)
σ_e^2	750.	748. (75.) 828. (86.)	767.4 (79.) 798.1 (79.)	802.3 (79.)	903.8 (79.) 759.1 (5.1)
LR	-	125.7 (6.8) 125.7 (6.8)	151.9 (9.6) 151.9 (9.6)	151.7 (9.6)	127.8 (7.8) 151.5 (9.6)

[1] Parameters and likelihood ratio are defined in Table 1.
[2] Analysis I: No relationships simulated, no relationships in analysis.
[3] Analysis II: Relationships simulated, no relationships in analysis.
[4] Analysis III: Relationships simulated and used in analysis with known phases.
[5] Analysis IV: Relationships simulated and used in analysis with most likely phases.
[6] Analysis V: Relationships simulated and used in analysis with unknown phases.
[7] Analysis VI: Relationships used in analysis and phases unknown; 3 alleles at markers.
[8] Analysis VII: Relationships used in analysis and phases unknown; h^2 fixed at .5.

most likely linkage phases used in place of the true phases, or treated as related with linkage phases considered as unknown (by using equation [11] in Grignola et al. (1996)).

Accuracy of parameter estimates (Table 2) was slightly higher when sires were simulated and analyzed as being unrelated (analysis I), compared to the case where sires were simulated and analyzed as related (analyses III to V). When sires were simulated related and analyzed unrelated, the estimate of heritability was slightly lower and the likelihood ratio statistic was lower than the corresponding values obtained with sires treated as related (analyses II and IV). When sires were analyzed using relationships, treating the most likely marker linkage phases of sires and ancestors as the true phases produced almost identical parameter estimates and likelihood ratio statistics as considering linkage phases as unknown (analyses III, IV, and V). Only small differences between analyses with known, most likely, or unknown linkage phases are to be expected for a granddaughter design with marker information on 100 sons per sire.

A further analysis (analysis VI in Table 2) was conducted on the same granddaughter design except that markers had three alleles at equal frequencies rather than the five alleles simulated for all other design variations. Parameter estimates were not noticeably affected by the decline in marker polymorphism, but the average likelihood ratio statistic was reduced.

In the last analysis of Table 2 (VII), heritability was fixed at .5. Accuracy of estimates of the QTL parameters and of residual variance was improved but the value of the likelihood ratio statistic was almost unchanged.

Table 3 contains results for data sets generated with $2v^2 = .125$ and with relationships among sires, and analyzed first by ignoring relationships among sires and secondly by

accounting for relationships with linkage phases treated as unknown (equation [11] in Grignola et al. (1996)). Analyses I and II in Table 3 were conducted for the granddaughter design with 100 sons per sire. The estimate of heritability and the likelihood ratio statistic were again lower when relationships among sires were ignored. Expectedly,

Table 3. Means and their SE for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the normal-effects QTL model explaining 12.5% of additive genetic variance (50 replicates).

Parameter[1]	True value	Analysis I[2] Analysis II[3]	Analysis III[4] Analysis IV[5]	Analysis V[6] Analysis VI[7]
h^2	.5	.617 (.042) .559 (.042)	.602 (.042) .472 (.040)	.538 (.041) .467 (.041)
v^2	.0625	.058 (.005) .063 (.006)	.069 (.006) .095 (.011)	.083 (.008) .112 (.013)
d_Q	.4527	.473 (.016) .434 (.017)	.435 (.016) .467 (.020)	.507 (.030) .526 (.030)
σ_e^2	750.	685.7 (97.) 810.2 (98.)	694.5 (83.) 994.6 (88.)	852.0 (95.) 1033.8 (99.)
LR	-	20.46 (1.9) 18.94 (1.7)	10.42 (1.0) 9.61 (.89)	6.56 (.74) 6.00 (.74)

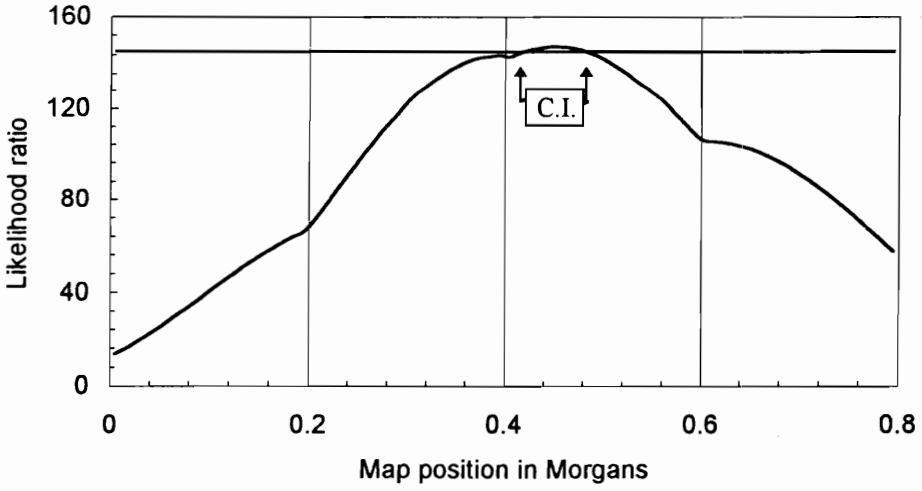
[1] Parameters and likelihood ratio are defined in Table 1.
[2] Analysis I: Relationships simulated, 100 sons/sire, relationships in analysis.
[3] Analysis II: Relationships simulated, 100 sons/sire, no relationships in analysis.
[4] Analysis III: Relationships simulated, 50 sons/sire, relationships in analysis.
[5] Analysis IV: Relationships simulated, 50 sons/sire, no relationships in analysis.
[6] Analysis V: Relationships simulated, 30 sons/sire, relationships in analysis.
[7] Analysis VI: Relationships simulated, 30 sons/sire, no relationships in analysis.

position of this smaller QTL was less accurately estimated, and the likelihood ratios were considerably lower than those in Table 2. In the analyses of the granddaughter designs with

only 50 (analyses III and IV) or 30 (analyses V and VI) sons, heritability estimates were also lower when sires were treated as unrelated, and QTL variance contribution was over-estimated. The differences between analyses with and without relationships, in the likelihood ratio statistics were rather small for the designs with 30 and 50 sons, where the likelihood ratio statistics were near the threshold values of 5.99 (.05 type-I error level) and 9.21 (.01 type-I error level) when assuming a chi-square distribution with 2 degrees of freedom.

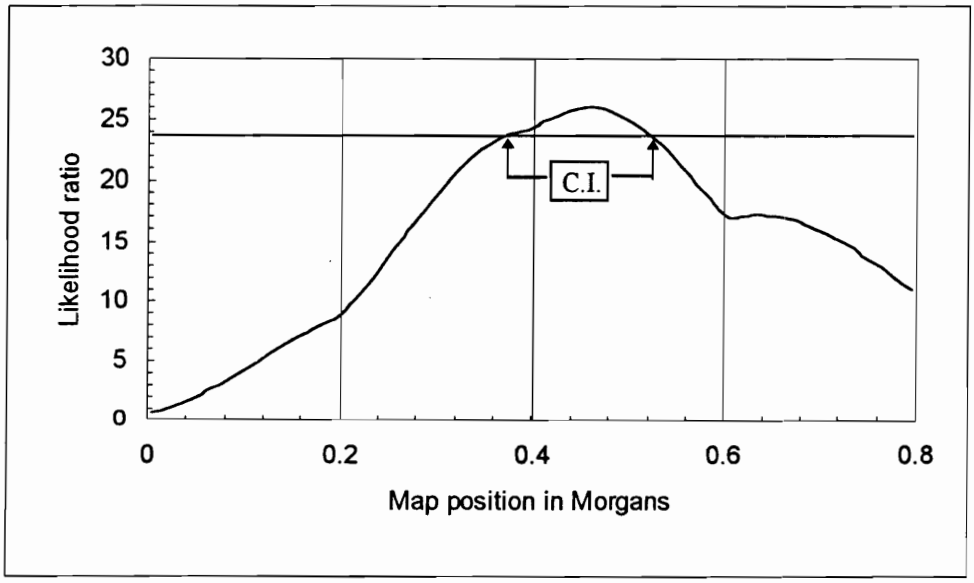
Figures 3 and 4 depict residual likelihood profiles for single replicates generated with $2v^2 = .5$ and $2v^2 = .125$, respectively, and obtained from analyses with relationships among sires and with linkage phases of sires and ancestors treated as unknown. Both figures display the marker positions, the most likely QTL location, and a confidence interval (CI) for the QTL

Figure 3: Likelihood profile for a single data set generated with $2v^2 = .5$ (see Table 1). Markers are located at 0cM, 20cM, 40cM, 60cM and 80cM. The QTL is located at 45.27cM. Significance threshold for log likelihood ratio is 5.991 from χ^2(2 df, α=.05)



position calculated by the LOD drop-off method of Lander and Botstein (1989). The limits of this CI were found by determining the map position at either side of the most likely position, where the LOD score had fallen by one unit (this calculation required converting natural logarithms to base 10 logarithms). As pointed out earlier, information from all markers was utilized, leading to smoother profiles (Knott and Haley, 1992; Georges et al., 1995) than the original interval mapping method of Lander and Botstein (1989).

Figure 4: Likelihood profile for a single data set generated with $2v^2 = .125$ (see Table 1). Markers are located at 0cM, 20cM, 40cM, 60cM and 80cM. The QTL is located at 45.27cM. Significance threshold for log likelihood ratio is 5.991 from χ^2(2 df, α=.05)



Results for the multiallelic model ($2v^2 = .204$) are presented in Table 4. Again, data sets were analyzed by first treating sires as unrelated and then by utilizing relationships among sires with linkage phases of sires and ancestors treated as unknown. Parameters were estimated

quite accurately, and likelihood ratios were significant and similar to those for the normal-effects QTL model with $2v^2 = .125$.

Table 4. Means and their SE for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the multiallelic QTL model explaining 20.2% of additive genetic variance (50 replicates). Relationships among sires simulated, no relationships considered in analysis or relationships used with unknown phases.

Parameter[1]	True value	Estimate[2]	SE
h^2	.5	.5718	.0443
		.5861	.0420
v^2	.102	.1123	.0103
		.1086	.0083
d_Q	.4527	.4404	.0080
		.4444	.0076
σ_e^2	750.	793.4	97.90
		741.9	91.00
LR	-	18.94	1.638
		20.47	1.934

[1]Parameters and likelihood ratio are defined in Table 1.

[2]First row: relationships in simulation and no relationships in analysis
Second row: relationships in simulation and in analysis with unknown linkage phase.

Results for the biallelic models with half of the homozygote difference equal to one additive genetic SD ($2v^2 = .5$) or to one half of it ($2v^2 = .125$) are presented in Tables 5 and 6, respectively. For both tables, sires were related in the simulation, and data sets were analyzed by ignoring relationships and by accounting for relationships with linkage phases unknown (equation [11] in Grignola et al. (1996)). Ignoring relationships among sires again decreased

the estimate of heritability but increased the estimate of v^2. The least accurate estimate of QTL position was obtained under the biallelic model with $2v^2 = .125$. Overall and at least when relationships among sires were considered in the analyses, parameter estimates were not noticeably inferior to those obtained under the corresponding normal-effects QTL models (Table 2), and likelihood ratios for the biallelic and normal-effects models were similar.

Table 5. Means and their SE for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the biallelic QTL model explaining 50% of additive genetic variance (50 replicates). Relationships among sires simulated, no relationships considered in analysis or relationships used with unknown linkage phases.

Parameter[1]	True value	Estimate[2]	SE
h^2	.5	.4688	.0360
		.5387	.0319
v^2	.25	.3042	.0177
		.2565	.0124
d_Q	.4527	.4498	.0031
		.4478	.0032
σ_e^2	750.	981.1	82.46
		770.4	68.39
LR	-	145.3	5.918
		150.6	6.072

[1]Parameters and likelihood ratio are defined in Table 1.

[2]First row: relationships in simulation and no relationships in analysis
Second row: relationships in simulation and in analysis with unknown linkage phase.

Table 6. Means and their SE for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the biallelic QTL model explaining 12.5% of additive genetic variance (50 replicates). Relationships among sires simulated, no relationships considered in analysis or relationships used with unknown linkage phases.

Parameter[1]	True value	Estimate[2]	SE
h^2	.5	.4789	.0378
		.5069	.0393
v^2	.0625	.0742	.0053
		.0713	.0049
d_Q	.4527	.4304	.0146
		.4284	.0152
σ_e^2	750.	960.0	86.57
		896.3	85.27
LR	-	22.33	1.754
		24.20	1.817

[1]Parameters and likelihood ratio are defined in Table 1.

[2]First row: relationships in simulation and no relationships in analysis
Second row: relationships in simulation and in analysis with unknown linkage phase.

Data under the biallelic QTL model were also simulated with a gene frequency of .2 for two designs. For design I, true values were v^2=.16, h^2=.5, σ_e^2=750 and d_Q=.4527, and estimated values v^2=.1803 ± .014, h^2=.5525 ± .05, σ_e^2=782.7 ± 103.62 and d_Q=.4467 ± .01. For design II, true values were v^2=.08, h^2=.5, σ_e^2=750 and d_Q=.4527, and estimated values v^2=.0454 ± .0005, h^2=.5429 ± .0557, σ_e^2=850.1 ± 126.22 and d_Q=.4583 ± .0213.

CONCLUSIONS

REML analysis based on a mixed linear model with random QTL allelic effects, having a priori normal distribution, provided quite accurate estimates of QTL location and of the variance components, in particular of the QTL variance contribution. Data were generated under three different genetic models for the QTL, the biallelic, the multiallelic (ten alleles) and the normal-effects (two effects per founder drawn from independent and identical normal distributions) model. While the normal-effects model is very similar to the model of analysis (REML requires the assumption of normality), the biallelic model with gene frequency p and substitution effect α deviates the most from the model of analysis. In the biallelic model, 50% of the individuals are expected to be homozygous for a gene frequency of p=.5, and variance among daughters of a homozygous son is equal to $p(1-p)\alpha^2 + .75\sigma_u^2 + \sigma_e^2$, while variance among daughters of heterozygous sons is higher by $.25\alpha^2$. Therefore, Thaller and Hoeschele (1996a,b) and Uimari et al. (1996a) fitted two different residual variances of DYD when performing Bayesian analysis of linkage of a biallelic QTL.

Despite the discrepancies between the biallelic model and the model of analysis, the REML analysis was quite robust to the number of alleles at the QTL, a result which is in agreement with findings of Xu and Atchley (1995); i.e., polymorphism at the QTL did not strongly affect parameter estimates or hypothesis tests. This finding confirms the usefulness of the REML analysis as an alternative method of analysis which, although not non-parametric, requires fewer parametric assumptions (number of QTL alleles and their frequencies) than

Maximum Likelihood and Bayesian analyses based on biallelic QTL models. Furthermore, REML is known in general to be quite robust to deviations from normality.

The REML analysis can be considered as an approximation to the Bayesian analysis of Hoeschele et al. (1996) fitting a normal-effects QTL model. The Bayesian analysis has the advantage of also being able to fit a biallelic model, but for the designs considered here, it does not give more accurate parameter estimates than the REML analysis and requires several hours of computing time. Although the REML analysis fitting a single QTL requires only around eight minutes of CPU time, this requirement still prohibits the calculation of genome-wide significance thresholds for this method using data permutation (Churchill and Doerge, 1994), unless a number of less stringent significance levels are chosen to obtain threshold values and these are used to extrapolate to the desired significance threshold (Uimari et al., 1996). Only the Least-Squares method allows to directly compute genome-wide significance thresholds from a large number of permutations (e.g., 10,000 to 100,000).

The power of the test likelihood ratio (LR) test was estimated as the number of replicates with LR > .99th percentile from Chi-square with 2 df. The power was 1.00 irrespective of the QTL model when true $v^2 = .25$. The power was .87 and .97 when true $v^2 = .0625$ for the normal-effects and biallelic QTL models, respectively. When the gene frequency at the biallelic QTL model was set to $p = .2$, the power was 1.00 for with $v^2 = .16$ and .60 for $v^2 = .04$.

The REML analysis has been implemented in the FORTRAN program Multiple QTL Residual Maximum Likelihood (MQREML), which is available from the authors upon request. The program currently is being extended to fit two linked QTLs per chromosome, rather than

using the approach of Xu and Atchley (1996) fitting the variances associated with the next-to-flanking markers to account for additional, linked QTLs. Their approach is only approximate as effects associated with marker alleles identified within founders erode over generations. Furthermore, it requires many additional parameters when the marker polymorphism is limited, causing the flanking and next-to-flanking markers to differ among families. In the near future, the program will be extended to other designs (e.g., full-sibships within half-sibships), where the current computation of the inverse of the variance-covariance matrix becomes approximate due to uncertain linkage phases in parents of final offspring, and other ways of computing this inverse exactly (e.g., van Arendonk et al., 1994) will be implemented.

ACKNOWLEDGEMENTS

This research was supported by Award No. 92-01732 of the National Research Initiative Competitive Grants Program of the US Department of Agriculture and by the Holstein Association USA. I. Hoeschele acknowledges financial support from the European Capital and Mobility fund while on research leave at Wageningen University, The Netherlands.

REFERENCES

Churchill G, Doerge R (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

Da Y, Ron M, Yanai A, Band M, Everts RE, Heyen DW, Weller JI, Wiggans GR, Lewin HA (1994) The Dairy Bull DNA Repository: A resource for mapping quantitative trait loci. Proc 5th World Congr Genetics Appl Livest Prod 21:229-232

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, Womack JE, Hoeschele I (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920

Grignola FE, Hoeschele I, Tier B (1996) Mapping quantitative trait loci via residual maximum likelihood: I. Methodology. Genet Sel Evol (in press)

Haley CS, Knott SA, Elsen J-M (1994) Mapping quantitative trait loci in crosses between outbred lines using least-squarés. Genetics 136:1195-1207

Hoeschele I, Uimari P, Grignola FE, Zhang Q, Gage KM (1996) Statistical mapping of polygene loci in livestock. Proc Int Biometric Soc (in press)

Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics 132:1211-1222

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

Thaller G, Hoeschele I (1996a) A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. Methodology. Theor Appl Genet (in press)

Thaller G, Hoeschele I (1996b) A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. A simulation study. Theor Appl Genet (in press)

Uimari P, Thaller G, Hoeschele I (1996a) A Monte Carlo method for Bayesian analysis of linkage between multiple linked markers and a quantitative trait locus. Genetics (in press)

Uimari P, Zhang Q, Grignola FE, Hoeschele I, Thaller G (1996b) Analysis of QTL workshop I granddaughter design data using least-squares, residual maximum likelihood and Bayesian methods. Annual Meeting of the International Society of Animal Genetics, Tours, France.

Van Arendonk JAM, Tier B, Kinghorn BP (1994a) Use of multiple genetic markers in prediction of breeding values. Genetics 137:319-329

VanRaden PM, Wiggans GR (1991) Derivation, calculation, and use of national animal model information. J Dairy Sci 74:2737-2746

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

Weller JI, Kashi Y, Soller M (1990) 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

Xu S, Atchley WR (1995) A random model approach to interval mapping of Quantitative Trait Loci. Genetics 141:1189-1197

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468

Chapter 4

MAPPING LINKED QUANTITATIVE TRAIT LOCI VIA RESIDUAL MAXIMUM LIKELIHOOD

F.E. Grignola and I. Hoeschele

Department of Dairy Science, Virginia Polytechnic Institute and State University Blacksburg, VA 24061-0315

Submitted for publication in **Genetics, Selection, Evolution**

ABSTRACT

A Residual Maximum Likelihood method is presented for estimation of the positions and variance contributions of two linked QTLs and of additive polygenic and residual variance components. The method also provides tests for zero versus one QTLs linked to a group of markers and for one versus two linked QTLs. A deterministic, derivative-free algorithm is employed. The variance-covariance matrix of the allelic effects at each QTL and its inverse is computed conditional on incomplete information from multiple linked markers. Covariances between effects at different QTLs and between QTL and polygenic effects are assumed to be zero. A simulation study was performed to investigate parameter estimation and likelihood ratio tests. The design was a granddaughter design with 2000 sons, 20 sires of sons and 9 ancestors of sires. Data were simulated under a normal-effects and a biallelic model for variation at each QTL. The trait analyzed was Daughter Yield Deviation or the son's daughter average adjusted for environmental effects and merits of mates. Genotypes at five or nine equally spaced markers were generated for all sons and their ancestors. Two linked QTLs accounted jointly for 50% or 25% of the additive genetic variance, and distance between QTLs varied from 10cM to 40cM. Power of detecting a second QTL was at least .5 when distance was 20cM and accounted for 50% of the additive variance, or 40cM and accounted for 25% of the additive variance. QTL parameters were estimated quite accurately except when QTLs were less than 20cM apart.

INTRODUCTION

A variety of methods for the statistical mapping of Quantitative Trait Loci (QTL) exist. While some methods analyze squared phenotypic differences of pairs related individuals (e.g., Haseman and Elston, 1972; Gotz and Ollivier, 1994), most methods analyze the individual phenotypes of pedigree members. The main methods applied to livestock populations are Maximum Likelihood (ML) (e.g., Weller, 1986; Lander and Botstein, 1989; Knott and Haley, 1992), Least-Squares as an approximation to ML (e.g., Weller et al., 1990; Haley et al., 1994; Zeng, 1994), and a combination of ML and LS referred to as Composite Interval Mapping (CIT) (Zeng, 1994) or Multiple QTL Mapping (MQM) (Jansen, 1993). These methods were developed mainly for line crossing and, hence, cannot fully account for the more complex data structures of outcrossed populations, e.g., data on several families with relationships across families, incomplete marker information, unknown number of QTL alleles in the population, and varying amounts of data information on different QTLs or in different families.

Recently, Thaller and Hoeschele (1996a,b) and Uimari, Thaller and Hoeschele (1996) implemented a Bayesian method for QTL mapping using single markers or all markers on a chromosome, respectively, via Markov chain Monte Carlo algorithms and applied the analyses to simulated granddaughter designs identical to those in the present study. Hoeschele et al. (1996) showed that the Bayesian analysis can accommodate either a biallelic or a normal-effects QTL model. While the Bayesian analysis was able to account for pedigree relationships both at the QTL and for the polygenic component and gave

good parameter estimates, it was very demanding for computing time in particular when fitting two QTLs (Uimari and Hoeschele, 1996).

Therefore, Grignola et al. (1996a) developed a Residual Maximum Likelihood method, using a deterministic, derivative-free algorithm, to map a single QTL. Hoeschele et al. (1996) showed that this method can be considered as an approximation to the Bayesian analysis fitting a normal-effects QTL model. In the normal-effects QTL model postulated by the REML analysis, the vector of QTL allelic effects is random with a prior normal distribution. The REML analysis builds on earlier work by Fernando and Grossman (1989), Cantet and Smith (1991) and Goddard (1992) on Best Linear Unbiased Prediction of QTL allelic effects by extending it to the estimation of QTL, polygenic, and residual variance components and of QTL location.

Xu and Atchley (1995) performed interval mapping using Maximum Likelihood based on a mixed model with random QTL effects, but these authors fitted additive genotypic effects at the QTL, rather than allelic effects, with variance-covariance matrix proportional to a matrix of proportions of alleles identical-by-descent, and assumed that this matrix was known. Their analysis was applied to unrelated full-sib pairs. In order to account for several QTLs on the same chromosome, Xu and Atchley (1995) used the idea behind CIT and fitted variances at the two markers flanking the marker bracket for a QTL. This approach, however, is less appropriate for multi-generational pedigrees, as effects associated with marker alleles erode across generations due to recombination. It is also

problematic for outbred populations, where incomplete marker information causes the flanking and next-to-flanking markers to differ among families.

In this paper, we extend the REML method of Grignola et al. (1996) to the fitting of multiple linked QTLs. While the extension is general for any number of linked QTLs, we apply the method to simulated granddaughter designs by fitting either one or two QTLs.

METHODOLOGY

Mixed Linear Model

The model is identical to that of Grignola et al. (1996a), except that it includes effects at several (t) QTLs, and it can be written as:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{Z}\boldsymbol{T}\left(\sum_{i=1}^{t} \boldsymbol{v}_i\right) + \boldsymbol{e} \qquad [1]$$

with

$$\mathrm{Var}(\boldsymbol{v}) = \mathrm{Diag}\{\mathrm{Var}(\boldsymbol{v}_i)\}, \quad \mathrm{Var}(\boldsymbol{v}_i) = \boldsymbol{Q}_i\sigma^2_{v(i)},$$

$$\mathrm{Var}(\boldsymbol{u}) = \boldsymbol{A}\sigma^2_u, \qquad \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{R}\sigma^2_e$$

where $\boldsymbol{y}$ is the vector of phenotypes, $\boldsymbol{X}$ is a design/covariate matrix, β is the vector of fixed effects, $\boldsymbol{Z}$ is an incidence matrix relating records to individuals, $\boldsymbol{u}$ is the vector of residual additive (polygenic) effects, $\boldsymbol{T}$ is an incidence matrix relating individuals to alleles, $\boldsymbol{v}_i$ is the vector of QTL allelic effects, $\boldsymbol{e}$ is the vector of residuals, $\boldsymbol{A}$ is the additive genetic

relationship matrix, σ^2_u is polygenic variance, $Q_i\sigma^2_{v(i)}$ is the variance-covariance matrix of the allelic effects at QTL i conditional on marker information, $\sigma^2_{v(i)}$ is the allelic variance at QTL i (or half of the additive variance at QTL i), $\boldsymbol{R}$ is a known diagonal matrix, and σ^2_e is residual variance. Each matrix Q_i depends on one unknown parameter, the map position of QTL i (d_i). Parameters related to the marker map (marker positions and allele frequencies) are assumed to be known. The model is parameterized in terms of the unknown parameters heritability ($h^2 = \sigma_a^2/\sigma_p^2$), fraction of the additive genetic variance explained by the allelic effects at QTL i ($v^2_i = \sigma^2_{v(i)}/\sigma^2_a$; i=1,...,t), residual variance σ^2_e, and QTL map locations $d_1,..., d_i, ..., d_t$.

A model equivalent to the animal model in [1] is (Grignola et al., 1996a):

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{W}\boldsymbol{u}_p + \boldsymbol{T}\sum_{i=1}^{t}\boldsymbol{F}_i\boldsymbol{v}_{p(i)} + \boldsymbol{m} + \boldsymbol{T}\sum_{i=1}^{t}\varepsilon_i + \boldsymbol{e} \qquad [2]$$

$$Var(\boldsymbol{u}_p) = \boldsymbol{A}_p\sigma^2_u,\ \ Var(\boldsymbol{v}_{p(i)}) = \boldsymbol{Q}_{p(i)}\sigma^2_u,\ \ Var(\boldsymbol{m}) = \Delta_u\sigma^2_u,$$

$$Var(\varepsilon_i) = \Delta_{v(i)}\sigma^2_{v(i)},\ \ Var(\boldsymbol{e}) = \boldsymbol{R}\sigma^2_e$$

where $\boldsymbol{W}$ has at most two non-zero elements equal to .5 in each row in columns pertaining to the known parents of an individual, $\boldsymbol{F}_i$ is a matrix with up to four non-zero elements per row pertaining to the QTL effects of an individual's parents (Wang et al., 1995; Grignola et al., 1996a), $\boldsymbol{A}_P$ and $\boldsymbol{Q}_{p(i)}$ are sub-matrices of $\boldsymbol{A}$ and $\boldsymbol{Q}$, respectively, pertaining to all animals that are parents, and $\mathbf{m}$ and ε_i are Mendelian sampling terms for polygenic and QTL effects, respectively, with covariance matrices as specified in equation [2]. While Var($\boldsymbol{m}$) is diagonal, Var(ε_i) can have some off-diagonal elements in inbred populations (Hoeschele, 1993; Wang et al., 1995).

Note that models [1] and [2] are conditional on a set of QTL map positions (and on marker positions which are assumed to be known). Dependent on the map positions are the matrices Q_i in model [1] and the matrices F_i and $Q_{p(i)}$ in model [2].

Note furthermore that models [1] and [2] assume zero covariances between effects at different QTLs, and between polygenic and QTL effects. However, selection tends to introduce negative covariances between QTLs (Bulmer, 1985).

A Reduced Animal Model (RAM) can be obtained from model [2] by combining $\boldsymbol{m}$, the ε_i (i=1,...,t) and $\boldsymbol{e}$ into the residual. Mixed Model Equations (MME) can be formed directly for the RAM, or by setting up the MME for model [2] and absorbing the equations in $\boldsymbol{m}$ and the ε_i (i=1,...,t). The resulting MME for the RAM and for t = 2 QTLs are:

$$\begin{bmatrix} X'DX & X'DW & X'DTF_1 & X'DTF_2 \\ W'DX & W'DW + A_p^{-1}k_u & W'DTF_1 & W'DTF_2 \\ F_1'T'DX & F_1'T'DW & F_1'T'DTF_1 + Q_{p(1)}^{-1}k_{v(1)} & F_1'T'DTF_2 \\ F_2'T'DX & F_2'T'DW & F_2'T'DTF_1 & F_2'T'DTF_2 + Q_{p(2)}^{-1}k_{v(2)} \end{bmatrix} \begin{bmatrix} b \\ u_p \\ v_{p(1)} \\ v_{p(2)} \end{bmatrix} =$$

$$\begin{bmatrix} X'Dy \\ W'Dy \\ F_1'T'Dy \\ F_2'T'Dy \end{bmatrix} \quad [3]$$

where $k_u = \sigma^2_e/\sigma^2_u$, $k_{v(i)} = \sigma^2_e/\sigma^2_{v(i)}$, and

$$D = D_1 - D_1(D_1 + D_u^{-1}k_u)^{-1}D_1, \quad D_1 = D_2 - D_2P(P'D_2P + D_{v(1)}^{-1}k_{v(1)})^{-1}P'D_2$$
$$D_2 = R^{-1} - R^{-1}P(P'R^{-1}P + D_{v(2)}^{-1}k_{v(2)})^{-1}P'R^{-1}$$

Matrix $\boldsymbol{D}$, which results from successive absorption of the Mendelian sampling terms for the polygenic component and the QTLs, can be shown to be always diagonal and very simple to compute, even when several (t > 2) QTLs are fitted. Let $\delta_{v(ijk)}$ represent the Mendelian sampling term pertaining to v effect k (k=1,2) of individual j at QTL i, and $\delta_{u(j)}$ Mendelian sampling for the polygenic effect of j. Then, the element of $\boldsymbol{D}$ pertaining to individual j (d_{jj}) is computed as follows:

$d_v = r^{jj}$

For i=1 to i=t compute

$$x = (\delta_{v(ij1)} + \delta_{v(ij2)}) / [(\delta_{v(ij1)} + \delta_{v(ij2)})d_v + \delta_{v(ij1)} \delta_{v(ij2)} k_v]$$

$$d_v = d_v - d_v^2 x$$

end

$d_{jj} = d_v - d_v^2 (1/d_v + \delta_{u(j)} k_u)$

where r^{jj} is the *j'th* diagonal element of $\boldsymbol{R}^{-1}$.

REML Analysis

The REML analysis was performed using interval mapping and a derivative-free algorithm to maximize the likelihood for any given set of QTL positions, as described by Grignola et al. (1996) for a single QTL model. The log residual likelihood for the animal model was obtained by adding correction terms to the residual likelihood formed directly from the RAM MME (Grignola et al., 1996a). The RAM residual likelihood is:

$$L_{RAM} \propto -.5log|G_{RAM}| - .5(N - NF - NR_{RAM})log(\hat{\sigma}_e^2) - .5log|C_{RAM}| - .5y'Py\hat{\sigma}_e^{-2}$$
$$where\ \hat{\sigma}_e^2 = y'Py/(N - NF)$$

where N is number of phenotypic observations, NF is number of fixed effects, NR_{RAM} is number of random genetic effects of the parents ((1+2t)*number of parents), C_{RAM} is the coefficient matrix in the left-hand-side of [3], $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$, $V = Var(y)/\sigma^2_e$, and G_{RAM} is block-diagonal with blocks $A_p\sigma_u^2$ and $Q_{p(i)}\sigma^2_{v(i)}$ for i=1,...,t (see also Meyer, 1989).

The RAM residual likelihood is modified to obtain the residual likelihood for the animal model as follows (Grignola et al.. 1996):

$$L_{AM} = L_{RAM} - .5log|\Delta^{-1}| - .5|C_{22}| + .5(NR - NR_{RAM})log(\hat{\sigma}_e^2)$$

where Δ is blockdiagonal with blocks Δ_u and $\Delta_{v(i)}$ (i=1,...,t) from [2], C_{22} is the part of the MME for model [2] pertaining to $\boldsymbol{m}$ and ε_i (i=1,...,t), and NR is total number of random genetic effects ((1+2t)*number of animals) in the animal model.

The analysis is conducted in the form of interval mapping as in Grignola et al. (1996b), except that now a t-dimensional search on a grid of combinations of positions of the t QTLs must be performed, which is feasible for t=2. More precisely, map position (in cM) of the first QTL (d_1) is moved in steps of, e.g., $2cM$ from the first toward the last marker position, and for any given d_1, map position of the second QTL (d_2) is moved from $d_1\ cM + d_{min}\ cM$ toward the last marker position, where d_{min} is the minimum distance allowed between the QTLs (see Results). At each combination of d_1 and d_2 values, the residual likelihood is maximized with respect to the parameters h^2, v_i^2 (i=1,...,t) and σ_e^2.

Matrices $Q_{p(i)}$, F_i and $\Delta_{v(i)}$ were calculated for each QTL as described in Grignola et al.(1996).

Hypothesis Testing

The presence of one or two QTLs on the chromosome carrying the marker linkage group is tested by first maximizing the likelihood under the one-QTL model and under a model with no QTL fitted. Distribution of the likelihood ratio statistic for these two models can be obtained via simulation or data permutation (Churchill and Doerge, 1994; Grignola et al., 1996a,b; Uimari et al., 1996). If the likelihood ratio statistic indicates that a QTL is present, then presence of a second QTL is tested by comparing the maximum likelihood under the two-QTL model with (i) the maximum likelihood under the one-QTL model, (ii) the likelihood maximized subject to variance at QTL 1 fixed at zero and position of QTL 2 fixed at its ML estimate from the two-QTL model, and (iii) the likelihood maximized subject to variance at QTL 2 fixed at zero and position of QTL 1 fixed at its ML estimate from the two-QTL model. The distribution of the likelihood ratio statistics is not known, and obtaining it via data permutation would be difficult computationally, as many permutations would need to be analyzed, and as the two-dimensional search took 3-4 hours of run-time for the design described below. The likelihood ratios corresponding to (ii) and (iii) above may have an asymptotic Chi-square distribution with 1 degree of freedom (df) (if we assume that only the variance at one QTL can vary), but to be on the conservative side, a 3 df Chi-square distribution may be assumed (if we consider that the positions d_1 and d_2 and the variance at one QTL can

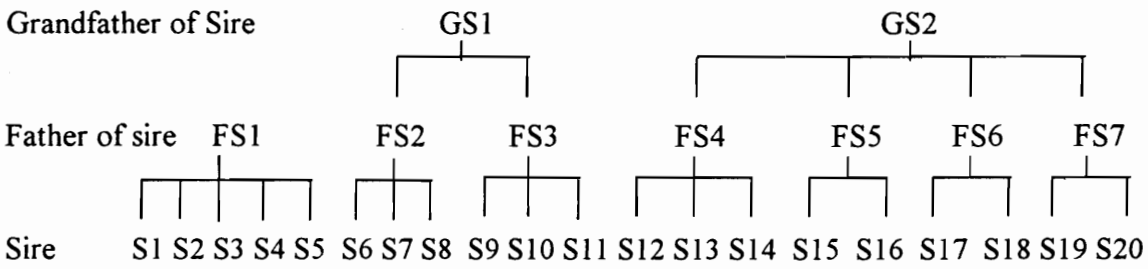
vary). For genome-wide testing, the significance level also should be adjusted for the number of independent tests performed (the number of chromosomes analyzed times the number of independent traits).

SIMULATION

Design

The design simulated was a granddaughter design (GDD) as in the single QTL study of Grignola et al. (1996b), where marker genotypes are available on sons and phenotypes on daughters of the sons. The structure resembled the real GDD of the US public gene mapping project for dairy cattle based on the Dairy Bull DNA Repository (Da et al., 1994). The simulated GDD consisted of 2000 sons, 20 sires, and 9 ancestors of the sires (Figure 1).

Figure 1: Pedigree of granddaughter design



The phenotype simulated was Daughter Yield Deviation (*DYD*) of sons (VanRaden and Wiggans, 1991). *DYD* is an average of the phenotypes of the daughters

adjusted for systematic environmental effects and genetic values of the daughter's dams. Total variance of *DYD* equals $Var(DYD) = .25\sigma_a^2 / R$ and can be factored into

$$Var(DYD) = .25\sigma_a^2 + [(1-R)/R]*.25\sigma_a^2$$

where R is reliability or squared accuracy of the son's estimated additive genetic effect or breeding value, and σ_a^2 is the additive genetic variance. In this factorization, the first component is the variance among the sons' transmitting abilities or half of their additive genetic values, and the second term is "residual variance" or the variance of the average dam and Mendelian genetic effect of the daughters and the average environmental effect. As in Grignola et al. (1996b), variance of *DYD* is rewritten as

$$Var(DYD) = .25\sigma_a^2 + w*\sigma_e^2$$

where $w = (1-R)/R$ and $\sigma_e^2 = .25\sigma_a^2$. Therefore, when analyzing *DYD* with weight w, *DYD* has an expected heritability of .5, because the expected value of the estimate of σ_e^2 is $.25\sigma_a^2$. Hence, there is the option of treating heritability as known in the REML analysis when phenotypic data are *DYD*s.

Marker and QTL genotypes were simulated according to Hardy-Weinberg frequencies and the map positions of all loci. One linkage group was considered, which consisted of five or nine marker loci and one or two QTLs. Each marker locus had five alleles at equal frequencies. The markers were spaced 20cM or 10cM apart, and for the results presented here, the QTL positions investigated were d_1 & d_2 = 30cM (interval 2) & 70cM (interval 4) with 5 markers, 25cM (interval 2) & 55cM (interval 3) with 5 markers, 30cM (interval 2) & 50cM (interval 3)

with 5 markers, 35cM (interval 2) & 45cM (interval 3) with 5 markers, and 35cM (interval 2) & 45cM (interval 3) with 9 markers.

Polygenic and QTL effects were simulated according to the pedigree in Figure 1. Data were analyzed by using the pedigree information on the sires. Note that in the simulation, no linkage disequilibrium (across families) was generated, i.e., covariances between pairs of effects at different QTLs or between QTL and polygenic effects were zero. Therefore, an additional design was simulated where linkage disequilibrium was generated by simulating *DYD*s also for sires, creating a larger number of sires and culling those sires with *DYD* lower than the 90^{th} percentile of the *DYD* distribution. QTL positions for this design were 30cM (interval 2) & 70cM (interval 3) with 5 markers, and the QTL model was the normal-effects model (see below). Estimates of the simulated correlations (SE in parentheses), across 30 replicates, were -.20 (.05), -.33 (.04), and -.32 (.04), between pairs of v effects at QTL 1 and QTL 2 , between pairs of v effects at QTL 1 and polygenic effects, and between pairs of v effects at QTL 2 and polygenic effects, respectively. The effects of one or several generations of phenotypic truncation selection on additive genetic variance in a finite locus model has been studied analytically by Hospital and Chevalet (1996).

QTL models

Two different QTL models were used to simulate data. Under both models, phenotypes were simulated as

$$DYD_j = \frac{1}{n_j}\sum_{i=1}^{2}\sum_{k=1}^{n_j} g_{ijk} + u_j + e_j; \quad Var(e_j) = \frac{1}{n_j}\left(.75s_u^2 + s_e^2\right)$$

where n_j was the number of daughters of son j, g_{ijk} was the sum of the v effects in daughter k of son j at QTL i, u_j was a normally distributed polygenic effect, e_j was a normally distributed residual, polygenic variance (σ_u^2) was equal to the difference between additive genetic variance (σ_a^2) and the variance explained by the QTL ($2\sigma_v^2$), and σ_e^2 was environmental variance. Number of daughters per son was set to 50, corresponding to a Reliability (VanRaden and Wiggans, 1991) near .8. Narrow sense heritability of individual phenotypes was $h^2 = .3$, and phenotypic SD was $\sigma_p = 100$.

Note that the QTL contribution to the DYDs of sons was generated by sampling individual QTL allelic effects of daughters under each of the two genetic models described below. This sampling of QTL effects assures that DYD of a heterozygous son, or of a son with substantial difference in the additive effects of the alleles at a QTL, has larger variance among daughters due to the QTL than a homozygous son or a son with similar QTL allelic effects.

Two different models were used to describe variation at the QTL, which are identical to two of the models considered by Grignola et al. (1996b):

Normal-effects model. For each individual with both or one parent(s) unknown, both or one effect(s) at QTL k (k=1,2) were drawn from $N(0,\sigma_{v(k)}^2)$. For the pedigree in Figure 1, there were 32 base alleles, and each QTL was treated as a locus with 32 distinct alleles in passing on alleles to descendants. The parameter $\sigma_{v(k)}^2$ was set to $.125\sigma_a^2$ or $.0625\sigma_a^2$, i.e.,

QTL k accounted for 25% ($2\nu_k^2$ =.25) or 12.5% ($2\nu_k^2$ =.125) of the total additive genetic variance, respectively. Consequently, the two QTLs accounted jointly for between 25% and 50% of the additive genetic variance.

Biallelic model. Each QTL was biallelic with allele frequency $p_1 = p_2 = p = .5$. The variance at QTL k was .

$2\sigma^2_{v(k)} = 2\nu_k^2\sigma_a^2 = 2p(1-p)a_k^2$ where for $p = .5$ and $2\nu_k^2 = .25$ or $2\nu_k^2 = .125$, half the homozygote difference at QTL k, a_k, and allelic variance $\sigma_{v(k)}^2$ were determined.

RESULTS

The first group of designs studied contains two QTLs of equal effect explaining jointly either 50% or 25% of the additive genetic variance. Parameters, likelihood ratio statistics and power tests are defined in Table 1. The designs are described in Table 2 and differ in the QTL positions and number of markers.

Overall, the QTL parameters were estimated quite accurately as in the single-QTL model analysis of Grignola et al. (1996), except that empirical standard errors and biases of QTL positions increased with decreasing true distance between the two QTLs.

Table 1: Definition of parameters and test statistics

Parameter	Definition
h^2	Narrow sense heritability
v_1^2	Ratio of QTL allelic to additive genetic variance at QTL 1 (QTL allelic variance equals half the additive variance due to the QTL)
v_2^2	Ratio of QTL to additive genetic variance at QTL 2
σ_e^2	Residual variance
d_1	Map position of QTL 1 in Morgan (for two-QTL model)
d_2	Map position of QTL 2 in Morgan (for two-QTL model)
d	Map position of QTL in Morgan (for one-QTL model)
LR_{d1}	Likelihood ratio statistic from the likelihood maximized under the two-QTL model and the likelihood maximized under the two-QTL model subject to variance at QTL 2 fixed at zero and position of QTL 1 fixed at its most likely value from the two-QTL model
LR_{d2}	Likelihood ratio statistic from the likelihood maximized under the two-QTL model and the likelihood maximized under the two-QTL model subject to variance at QTL 1 fixed at zero and position of QTL 2 fixed at its most likely value from the two-QTL model
LR_d	Likelihood ratio statistic from the likelihood maximized under the two-QTL model and the likelihood maximized under the one-QTL model
$PWR_{d1,d2}$	Power of the test of one versus two QTLs estimated as number of replicates with both LR_{d1} and LR_{d2} > .99th percentile from Chi-square with 1 df or 3df
PWR_d	Power of the test LR_d estimated as number of replicates with LR_d > .99th percentile from Chi-square with 1 df or 3df

When the QTLs jointly explained 50% of the additive genetic variance, the true QTL positions of 30cM & 70cM (design Ia) were estimated very accurately. For the true QTL positions of 25cM & 55cM (design IIa) and 30cM & 50cM (design IIIa), the distance between the two QTLs was slightly over-estimated (i.e., the position of the first QTL was slightly underestimated and the position of the other QTL slightly over-estimated). This small bias may be due to the search algorithm which fits the second QTL starting at a minimum distance of 10cM to the right of the first QTL. When analyzing design IVa with true positions of 35cM & 45cM with the same minimum distance, the distance between QTLs was greatly over-estimated. Therefore, the minimum distance between QTLs was reduced to d_{min} = 2cM. However, parameter estimates and likelihood

ratios remained unchanged, when the data were re-analyzed with d_{min} = 2cM. When the number of markers was increased from 5 to 9, with the same QTL positions of

Table 2: Parameter estimates (SE in parentheses) and likelihood ratio statistics for designs[1] with the normal-effects QTL model

Parameter[2]	Ia Ib	Iia Iib	IIIa IIIb	IVa IVb	Va Vb
h^2	.574 (.054) .551 (.053)	.570 (.046) .521 (.047)	.519 (.045) .591 (.056)	.589 (..051) .552 (.053)	.531 (.047) .596 (.054)
v_1^2	.142 (.012) .065 (.006)	.122 (.012) .072 (.010)	.144 (.016) .072 (.009)	.180 (.016) .088 (.010)	.152 (.019) .079 (.008)
v_2^2	.140 (.011) .069 (.008)	.124 (.013) .069 (.008)	.113 (.014) .068 (.010)	.110 (.018) .075 (.018)	.147 (.021) .060 (.006)
σ_e^2	750.6 (108.0) 802.9 (115.4)	712.5 (87.38) 834.5 (102.1)	819.5 (91.0) 737.4 (120.5)	705.3 (102.6) 793.2 (111.8)	831.9 (109.1) 714.5 (110.5)
d_1	.300 (.008) .308 (.018)	.224 (.014) .202 (.020)	.279 (.015) .282 (.020)	.324 (.021) .290 (.028)	.342 (.013) .290 (.021)
d_2	.702 (.011) .712 (.011)	.554 (.011) .582 (.017)	.559 (.019) .566 (.021)	.608 (.027) .626 (.029)	.501 (.019) .504 (.020)
d	.483 (.027) .488 (.027)	.416 (.020) .385 (.028)	.378 (.014) .400 (.015)	.382 (.009) .386 (.024)	.409 (.006) .388 (.010)
LR_{d1}	43.254 (4.5) 12.390 (1.7)	30.684 (3.9) 18.475 (3.5)	21.250 (4.9) 11.382 (2.6)	27.771 (7.5) 12.425 (3.1)	18.641 (4.1) 10.384 (2.5)
LR_{d2}	44.550 (5.1) 16.080 (2.7)	36.586 (6.2) 15.402 (2.4)	29.095 (5.3) 13.873 (2.7)	59.256 (9.58) 27.798 (4.3)	30.915 (7.1) 17.060 (4.1)
LR_d	24.368 (2.7) 7.048 (1.2)	13.730 (1.5) 7.089 (1.1)	6.191 (.9) 3.409 (.6)	10.878 (2.3) 3.835 (.51)	3.959 (.6) 3.017 (.4)
$PWR_{d1,d2}$[3]	.97 .93 .50 .17	.83 .70 .40 .27	.50 .27 .30 .10	.57 .37 .20 .00	.27 .17 .13 .00
PWR_d	.97 .93 .37 .13	.77 .60 .33 .23	.30 .10 .17 .03	.43 .33 .13 .00	.27 .00 .07 .00

1 Ia: d_1 = .3, d_2 = .7, 5 markers, $v_1^2 = v_2^2$ = .125
Ib: d_1 = .3, d_2 = .7, 5 markers, $v_1^2 = v_2^2$ = .0625
IIa: d_1 = .25, d_2 = .55, 5 markers, $v_1^2 = v_2^2$ = .125
IIb: d_1 = .25, d_2 = .55, 5 markers, $v_1^2 = v_2^2$ = .0625
IIIa: d_1 = .3, d_2 = .5, 5 markers, $v_1^2 = v_2^2$ = .125
IIIb: d_1 = .3, d_2 = .5, 5 markers, $v_1^2 = v_2^2$ = .0625
IVa: d_1 = .35, d_2 = .45, 5 markers, $v_1^2 = v_2^2$ = .125
IVb: d_1 = .35, d_2 = .45, 5 markers, $v_1^2 = v_2^2$ = .0625
Va: d_1 = .35, d_2 = .45, 9 markers, $v_1^2 = v_2^2$ = .125
Vb: d_1 = .35, d_2 = .45, 9 markers, $v_1^2 = v_2^2$ = .0625

[2] Parameters and likelihood ratios are defined in Table 1

[3] First number using 99th percentile from 1 df Chi-square; second number using 99th percentile from 3 df Chi-square.

35cM & 45cM, accuracy of the estimates of the QTL positions increased substantially, and accuracy of the QTL variance contributions also improved.

When the QTLs accounted jointly for only 25% of the additive genetic variance, precision of the estimates of QTL positions and variance contributions declined. However, QTL parameters were still estimated quite well if the QTLs were at least 20cM apart. At smaller QTL distances (designs IVb and Vb), QTL position estimates were biased with QTL distance over-estimated, similar to but somewhat more pronounced than the biases for designs IVa and Va with the larger QTLs.

When analyzing the designs in Table 2 with the single-QTL model, the most likely QTL position was always somewhere in between the QTL positions estimated under the two-QTL model. Averaged across replicates, the estimated QTL position was very near the mean of the true positions. This result was expected, as both QTLs had equal variance contributions and on average equally informative flanking markers.

The average values of the likelihood ratio statistics for testing between the single- and two-QTL models declined expectedly with decreasing distance between the two QTLs. For design IV, the average likelihood ratios were higher and more variable then for design V, which had nine markers instead of five. The reason for this finding was the occurrence of a few outlier values exceeding 100 for design IVa. Power of rejecting the single-QTL model or detecting a second QTL based on requiring both LR_{d1} and LR_{d2} to exceed the significance threshold was always higher than or equal to the power of the LR_d statistic. Use of LR_d is, therefore, not recommended. For the test based on LR_{d1} and LR_{d2},

power declined expectedly with decreasing distance among QTLs and with decreasing true QTL variance contribution. For joint QTL variance contribution of 50%, power was equal to or higher than .5 when the QTL distance was at least 20cM (1 df chi-square threshold) or 30cM (3 df chi-square). For joint QTL variance contribution of 25%, power of .5 was achieved only for the 40cM distance between QTLs and the 1 df chi-square significance threshold.

In Table 3, results for designs simulated under the biallelic QTL model are presented. Except for the QTL model, these designs are identical to designs Ia,b and IIIa,b in Table 2. Parameters were estimated with an accuracy not noticeably lower than for the normal-effects QTL model, an observation in agreement with the single-QTL study of

Table 3: Parameter estimates (SE in parentheses) and likelihood ratio statistics for designs[1] with the biallelic QTL model

Parameter[2]	Ia	Ib	Iia	IIb
h^2	.529 (.045)	.572 (.042)	.508 (.039)	.477 (.049)
v_1^2	.144 (.012)	.062 (.006)	.118 (.012)	.086 (.020)
v_2^2	.139 (.009)	.070 (.008)	.139 (.015)	.071 (.008)
σ_e^2	811.1 (91.6)	704.7 (90.3)	814.3 (79.6)	967.6 (110.5)
d_1	.305 (.008)	.269 (.014)	.267 (.014)	.278 (.020)
d_2	.698 (.006)	.683 (.012)	.522 (.014)	.561 (.022)
d	.505 (.027)	.487 (.029)	.412 (.014)	.419 (.015)
LR_{d1}	37.813 (2.6)	17.337 (1.8)	26.662 (5.0)	12.239 (2.8)
LR_{d2}	39.845 (3.2)	16.774 (2.1)	19.363 (3.1)	12.252 (3.2)
LR_d	24.319 (2.3)	8.266 (.9)	7.640 (1.1)	2.176 (.3)
$PWR_{d1,d2}$[3]	1.00 .97	.63 .40	.50 .33	.03 .00
PWR_d	.97 .93	.53 .23	.37 .20	.03 .00

[1] Ia: $d_1 = .3$, $d_2 = .7$, 5 markers, $v_1^2 = v_2^2 = .125$
Ib: $d_1 = .3$, $d_2 = .7$, 5 markers, $v_1^2 = v_2^2 = .0625$
IIa: $d_1 = .3$, $d_2 = .5$, 5 markers, $v_1^2 = v_2^2 = .125$
IIb: $d_1 = .3$, $d_2 = .5$, 5 markers, $v_1^2 = v_2^2 = .0625$

[2] Parameters and likelihood ratios are defined in Table 1.

[3] First number using 99th percentile from 1 df Chi-square; second number using 99th percentile from 3 df Chi-square.

Grignola et al. (1996b). Power of rejecting the single QTL hypothesis for the test based on LR_{d1} and LR_{d2} was similar to that for the corresponding designs in Table 2 from the normal-effects QTL model, or tended to be slightly higher.

The two-dimensional search required the evaluation of 630 QTL position combinations for the chromosome of 80cM length with 2cM step size and d_{min} = 10cM. Clock run-time of this analysis on a heavily used 21-processor IBM SP2 system was 3-4 hours for the two-QTL analyses, which likely would be reduced on a workstation dedicated to this analysis, while run-time for the single-QTL analysis was at most only 10 minutes.

For the designs in Tables 2 and 3, the two QTLs had equal variance contributions. Therefore, two additional designs with QTL positions of 30cM & 70cM and 30cM & 50cM, respectively, were simulated using the normal-effect QTL model, with QTL 1 explaining 25% and QTL 2 12.5% of the additive genetic variance. The average estimates (with SE in parentheses) of QTL position from the single QTL analysis were .396 ± .023 and .348 ± .010 for the 30cM & 70cM and 30cM & 50cM designs, respectively, being closer to the first locus with the larger variance contribution. Estimated QTL positions from the two-QTL analysis were 29.2 cM ± .8 and 69.9cM ± 1.0 for the 30cM & 70cM design, and 24.2 ± 2. and 49. ± 1.9 for the 30cM & 50cM design. Average v^2 estimates were .143 ± .013 and .072 ±.010 for the first design, and .126 ± .019 and .087 ± .012 for the second design, respectively. For the 30cM & 70cM design, power was .77 ± .57 for

the 1 (3) df chi-square threshold and the test based on LR_{d1} and LR_{d2}. For the 30cM & 50cM design, power was only .23 ± .10 for the same test.

When linkage disequilibrium was generated by phenotypic truncation selection of sires for the design with QTL positions of 30cM and 70cM and joint QTL genetic variance contribution of 50%, QTL parameters and their estimates clearly were affected. Heritability was under-estimated (.213 ± .020), and the v_k^2 (k=1,2) were over-estimated (.184 ± .014, .211 ± .014), except for QTL positions which were estimated accurately (.292 ± .010, .696 ± .007). Power appeared to be reduced somewhat compared to the same design without selection, and was estimated at .93 and .77 for the 1 and 3 df Chi-square distributions, respectively. Reduction in power was probably due to the high estimate of error variance (1737.9 ± 55.3).

CONCLUSIONS

The REML analysis of Grignola et al. (1996a), based on a mixed linear model with random and normally distributed QTL allelic effects and conditional on information from multiple linked markers, has been extended here to fit multiple linked QTLs. This extension is necessary to eliminate biases in the estimates of the QTL parameters position and variance, which occur when fitting a single QTL when other linked QTLs are present. At this time, the analysis has been implemented for two QTLs on the same chromosome using a two-dimensional search. When fitting more than two QTLs, a more efficient search strategy may be required. However, given the sizes of QTL variances that are needed to

detect a QTL, it seems rather unlikely that more than two QTLs would be detected on the same chromosome. As likelihood maximizations at different position combinations are independent of each other, use of multiple processors, if available, would be needed to reduce run times substantially.

REML analysis under.the two-QTL model yielded fairly accurate parameter estimates. Accuracy declined, in particular for QTL positions, with decreasing distance between the two QTLs. As in the single QTL study, the REML analysis was robust to the number of alleles at the QTLs, as there was little difference in parameter estimates and likelihood ratio statistics between designs generated with the normal-effects and biallelic QTL models. Previous linkage analyses (e.g., Knott and Haley, 1992; Haley et al., 1994) lead to the conclusion that a minimum distance of 20cM was required between linked QTLs for their separate detection. This result was confirmed in the present study for QTLs jointly explaining 50% of the additive genetic variance. A larger distance was required for QTLs jointly explaining only 25% of the additive genetic variance, to achieve a power of at least .5.

For the one-QTL model, relationships between the REML analysis, the equivalent method of Xu and Atchley (1995), the method of Schork (1993), and the Bayesian analysis of Uimari et al. (1996) were discussed in Grignola et al. (1996a). The method of Xu and Atchley (1995), fitting variances associated with the next-to-flanking markers to account for additional linked QTLs would not have worked well for the designs studied here. A first reason is the inclusion of ancestors of the sires in the analysis, as their method

fits random effects associated with the marker alleles in founders which erode across generations due to recombination. Another reason is the small number of families differing in the flanking and next-to-flanking markers, resulting in too little information for estimation of variances associated with the next-to-flanking markers in the method of Xu and Atchley (1995).

For testing the single- versus the two-QTL model, the test obtained from the likelihood maximized under the two-QTL model and the likelihoods maximized subject to the same positions and with either QTL variance contribution set to zero is a more appropriate and powerful criterion for detecting a second QTL on the same chromosome than the ratio derived from the maximized likelihoods under the single- and two-QTL models, respectively.

If there is linkage disequilibrium due to selection, QTL positions still will be estimated accurately while variance estimates and power may be affected. Accounting for disequilibrium in the analysis should be investigated.

Finally, the two-QTL analysis was applied to data generated with a single QTL explaining 50% of additive genetic variance and situated at 45 cM. Average estimates of QTL positions were 45.4 ± 0.4 cM from the single QTL analysis and 30.9 ± 3.2 cM and 56.3 ± 2.2 cM from the two-QTL analysis. Average estimates of QTL allelic variance as a fraction of additive genetic variance (v^2) were .279 ± .017 from the single-QTL analysis and .147 ± .027 and .139 ± .021 from the two-QTL analysis. Power of the test, or frequency of replicates with both LR_{d1} and LR_{d2} significant was .03 (.00) for the 1 (3) df Chi-square distribution, hence the

correct decision of not rejecting the single-QTL hypothesis was made most of the time or always.

Given the minimum distances reported here to detect two QTLs explaining jointly 50% of the additive variance (20cM) or 25% of the additive genetic variance (40cM) with a power of at least 50%, it seems unlikely that it will be possible to discriminate between higher numbers of linked QTLs based on data from livestock QTL mapping experiments.

ACKNOWLEDGEMENTS

This research was supported by Award No. 92-01732 of the National Research Initiative Competitive Grants Program of the US Department of Agriculture, by the Holstein Association USA, and by ABS Global Inc..

REFERENCES

Bulmer MG (1985) The mathematical theory of quantitative genetics. Clarendon Press, Oxford.

Cantet RJC, Smith C (1991) Reduced animal model for marker assisted selection using best linear unbiased prediction. Genet Sel Evol 23:221-233

Churchill G, Doerge R (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

Da Y, Ron M, Yanai A, Band M, Everts RE, Heyen DW, Weller JI, Wiggans GR, Lewin HA (1994) The Dairy Bull DNA Repository: A resource for mapping quantitative trait loci. Proc 5th World Congr Genetics Appl Livest Prod 21:229-232

Fernando RL, Grossman M (1989) Marker-assisted selection using best linear unbiased prediction. Genet Sel Evol 21:467

Goddard M (1992) A mixed model for analyses of data on multiple genetic markers. Theor Appl Genet 83:878-886

Gotz KU, Ollivier L (1994) Theoretical aspects of applying sib-pair linkage tests to livestock species. Genet Sel Evol 24:29-42

Grignola FE, Hoeschele I, Tier B (1996a) Mapping quantitative trait loci via residual maximum likelihood: I. Methodology. Genet Sel Evol (in press)

Grignola FE, Hoeschele I, Zhang Q, Thaller G (1996b) Mapping quantitative trait loci via residual maximum likelihood: I. A simulation study. Genet Sel Evol (in press)

Haley CS, Knott SA, Elsen J-M (1994) Mapping quantitative trait loci in crosses between outbred lines using least-squares. Genetics 136:1195-1207

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3-19

Hoeschele I (1993) Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. J Dairy Sci 76:1693-1713

Hoeschele I, Uimari P, Grignola FE, Zhang Q, Gage KM (1996) Statistical mapping of polygene loci in livestock. Proc Int Biometrics Soc (in press)

Hospital F, Chevalet C (1996) Interactions of selection, linkage and drift in the dynamics of polygenic characters. Genet Res Camb 67:77-87

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:252:324

Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics 132:1211-1222

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

Meyer K (1989) Restricted Maximum Likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genet Sel Evol 21:317

Schork NJ (1993) Extended multi-point identity-by-descent analysis of human quantitative traits: Efficiency, power, and modeling considerations. Am J Hum Genet 53:1306-1319.

Thaller G, Hoeschele I (1996a) A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. Methodology. Theor Appl Genet (in press)

Thaller G, Hoeschele I (1996b) A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. A simulation study. Theor Appl Genet (in press)

Uimari P, Thaller G, Hoeschele I (1996) The use of multiple linked markers in a Bayesian method for mapping quantitative trait loci. Genetics 143: (in press)

Uimari P, Hoeschele I (1996) Mapping linked quantitative trait loci with Bayesian analysis and Markov chain Monte Carlo algorithms. (In preparation)

VanRaden PM, Wiggans GR (1991) Derivation, calculation, and use of national animal model information. J Dairy Sci 74:2737-2746

Wang T, Fernando RL, van der Beek S, and M Grossman (1995) Covariance between relatives for a marked quantitative trait locus. Genet Sel Evol 27:251

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

Weller JI, Kashi Y, Soller M (1990) 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

Xu S, Atchley WR (1995) A random model approach to interval mapping of Quantitative Trait Loci. Genetics 141:1189-1197

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468

SUMMARY AND CONCLUSIONS

Statistical gene mapping methods including Least Squares (LS) and Maximum Likelihood (ML), were developed for line crossing and are unable to handle complex pedigree structures (e.g., relationships among families in an outbred population). ML and Bayesian methods can account for these type of pedigree structure, but they are highly demanding of computing time. Therefore, the aim of this dissertation research was to develop a Residual Maximum Likelihood (REML) approach based on a mixed linear model including polygenic effects and random QTL effects to map QTLs with multiple linked markers in outbred populations. The REML analysis can be considered as an approximation to the Bayesian analysis fitting a normal-effects QTL model with QTL variances having a normal distribution. The REML method can accommodate any pedigree structure, and is intermediate in computational requirements compared to the LS (low) and ML and Bayesian methods (high).

In Chapter 1, a half-sib design in the form of a granddaughter design (GDD), consisting of 20 unrelated sires with 50 sons each was simulated. Marker genotypes were available on sires and sons, and phenotypes on the daughters of the sons (50 daughters per son). The trait analyzed was daughter yield deviation. A QTL located between two markers was simulated, with each marker locus having five alleles. All sires were heterozygous at both markers, and linkage phases of sires and paternal marker haplotypes of sons were assumed to be known. Three different QTL models, assuming a biallelic, multiallelic (10 alleles) and normal-effects QTL (number alleles equal to twice the number of sires) were simulated. It was demonstrated that position and variance contribution of a single QTL are separately estimable from half-sib designs using a marker bracket instead

of a single marker. Regardless of the number of alleles at the QTL and the magnitude of the QTL variance contribution to additive genetic variance (0.125, 0.5), the three QTL models were able to estimate very accurately the parameters under consideration, including heritability ($h^2 = \sigma_a^2/\sigma_p^2$), ratio of QTL allelic variance to additive genetic variance ($v^2 = \sigma_v^2/\sigma_a^2$), QTL position ($d_Q$), and residual variance (σ_e^2). However, a slight improvement in the accuracy of estimation was observed as the number of alleles at the QTL increased. This was expected since the normal-effects QTL and the model of the REML analysis were very similar.

In Chapter 2, methodology for computing the variance-covariance matrix of the QTL allelic effects given multiple linked markers, incomplete marker information, and relationships among sires is presented. A REML method, using interval mapping with likelihood maximizations every cM along the chromosome via a derivative-free algorithm and a Reduced Animal Model is derived. A likelihood ratio test statistic (LR) was used to test for the presence or absence of a QTL on the chromosome under study. The LR is based on the maximum of the likelihood under the one-QTL model and under a model with no QTL fitted. The distribution of the test statistic was investigated by simulation and found to be between a 1 df and 2 df Chi-square distribution.

In Chapter 3, we performed a simulation study to investigate the methodology described in Chapter 2. Marker genotypes were simulated for a chromosome length of 80cM with 5 marker loci forming four 20cM intervals. The same QTL models as in Chapter 1 were considered, and the GDD consisted of 9 ancestors, 20 sires, 100 sons per

sire and 50 daughters per son. Overall, QTL location and allelic variance were well estimated under all three QTL models. For the GDD considered here and the normal-effects QTL model with $2v^2 = 50\%$, the results indicated that use of the true, most likely and unknown marker linkage phase in the sires affected neither the accuracy of the parameter estimation nor the Likelihood Ratios (LR). Ignoring relationships among sires or considering 3 alleles instead of 5 at each marker locus also did not affect accuracy of parameter estimation; however, a decrease in the LR was observed. To further investigate the effect of considering or ignoring relationships among sires on power of QTL detection, GDDs with 100, 50 and 30 sons per sire and a QTL with a smaller variance ($2v^2 = 12.5$) were simulated. The analysis was carried out by including or ignoring relationships among sires. In all designs, ignoring relationships led to a decrease in the heritability (h^2) estimate and to an over-estimate of the QTL variance contribution (v^2), in particular for the smaller designs. As the design became smaller, the LR decreased expectedly with or without use of relationships among sires However, superiority of the power of the LR with relationships over the power of the LR without relationships did not increase with decreasing size of the design.

Power calculations were carried out for the normal-effects and biallelic models considering relationships among sires and unknown linkage phases. The power was 1.0 irrespective of the model when the QTL explained 50% of the total additive variance; however, when the QTL explained only 12.5% of the total additive variance the power dropped to .87 and .97 for the biallelic and normal-effects model, respectively. When the

gene frequency in the biallelic model was fixed at p=.2, power was 1.0 and .60 if the QTL variance contribution was .32 and .08 of the additive variance, respectively.

For the multiallelic QTL model explaining 20.4% of the additive variance, and the biallelic QTL models explaining 50% and 12.5% of the additive genetic variance, location and dispersion parameters were estimated accurately. As in the normal-effects QTL model, ignoring relationships among sires increased v^2, and decreased the h^2 and LR. For the three QTL models simulated, and for otherwise identical designs and analyses, parameter estimates and LRs were very similar. Therefore, it can be concluded that the REML analysis is quite robust to the type of polymorphism at the QTL.

In Chapter 4, the REML method was extended to accommodate several QTLs on the same chromosome. The same GDD including relationships among sires as previously described, and only the normal-effects QTL and biallelic QTL models were considered. QTLs were simulated with equal variances (50% or 25% of the additive genetic variance was explained jointly), unequal variances (25% and 12.5% for each QTL), and different distances between QTLs (40, 30, 20 or 10 cM apart). Hypothesis testing was again based on the likelihood ratio statistic. First, a test for the presence or absence of at least one QTL was carried out as described in Chapter 2. If there was evidence for a single QTL, presence of a second QTL was tested by comparing the maximum of the likelihood under the two-QTL model with (a) the maximum of the likelihood under the one-QTL model (LR_d) and (b) the maximum of the likelihood subject to setting the variance of one QTL to zero and fixing the position of the other QTL at its maximum likelihood estimate from the two-QTL model (LR_{d1} and LR_{d2}). Power of the test requiring both LR_{d1} and LR_{d2} to be

significant was estimated as the number of replicates where both LR_{d1} and LR_{d2} exceeded the .99th percentile of the 1 df or 3 df Chi-square distribution.

As in the single QTL analysis, overall parameters were well estimated for all the designs considered, except when the QTLs were 10 cM apart. In the latter case, parameter estimation was improved by increasing the number of marker loci in the linkage group from 5 to 9. In general, it can be concluded that as the distance between QTLs and the contribution of the QTLs decreased, (i) bias and standard errors of position and QTL variances increased, and (ii) power of detecting two QTLs based on LR_d or LR_{d1} and LR_{d2} decreased. Power of rejecting the single QTL model based on LR_{d1} and LR_{d2} was always higher than power based on LR_d, indicating that the former test should be preferred. Finally, the effect of linkage disequilibrium generated by phenotypic truncation selection of sires was studied. Results showed that the position of the two QTLs was estimated accurately, however, heritability was underestimated, the QTL variance contribution over-estimated, and the power appeared to be somewhat reduced, compared to the same design without selection.

Main conclusions of this thesis

- The REML analysis using an expected variance-covariance matrix of random QTL effects and relationships among sires in a large granddaughter design (20 sires and 100 sons/sire) allowed us to estimate quite accurately the QTL allelic variance(s) and position(s).

- A single QTL explaining at least 12.5% of the total additive variance can be detected with a power of at least .87 for any of the genetic models considered here (biallelic, multiallelic and normal-effects QTL models).
- Two QTLs explaining jointly 50 % or 25% of the total additive genetic variance, can be detected with a power equal to or greater than .50 if the minimum distance between them is at least 20cM or 40cM respectively.
- The REML method proved to be robust to the number of alleles at the QTL. However, decreased polymorphism at the marker loci decreased the power of QTL detection.
- The REML method based on a Reduced Animal Model using a derivative-free algorithm and interval mapping is computationally efficient.

Areas of further research include analysis of other designs (e.g. full sibs), extension of the method to multiple trait analysis, consideration of unequal gene frequencies at the marker loci, and accounting for the effect of linkage disequilibrium.

VITA

Fernando Grignola was born on March 4, 1962 in Montevideo, Uruguay. He attended the Faculty of Agriculture of Uruguay, and received a Bachelor of Science degree in Agriculture and Livestock Production in 1988. Upon completion of his BS degree, he worked at the University of Agriculture of Uruguay as an assistant professor in the department of Animal Science, and as a technician at the National Institute for Agricultural Research of Uruguay until 1990. In 1991, he entered the Department of Animal Science at Michigan State University and received a Master of Science degree in Animal Science in 1993. Since that time he began to work on his doctorate program in the Department of Dairy Science at Virginia Polytechnic Institute and State University. He is currently a member of Gamma Sigma Delta and Sigma Xi.