SELECTION INDICES FOR COMBINING MARKER GENETIC DATA AND ANIMAL MODEL INFORMATION

by

Eduardo O. Romano

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

Masters of Science

in

Dairy Science

APPROVED:

Ina Hoeschele, Chairman

Ronald E. Pearson

Eric Hallerman

Ann Dunnington

David R. Notter

January, 1993
Blacksburg, Virginia

SELECTION INDICES FOR COMBINING MARKER GENETIC DATA AND ANIMAL MODEL INFORMATION

by

Eduardo O. Romano

Ina Hoeschele, Chairman

DAIRY SCIENCE

(ABSTRACT)

It was suggested that marker and phenotypic information be combined in order to obtain more accurate or earlier genetic evaluations. An improvement in accuracy or time of evaluation due to utilization of marker assisted selection (MAS) increases genetic progress. Fernando and Grossman (1989) suggested including marker information directly into the Animal Model, Best Linear Unbiased Prediction system, but several problems need to be solved before their approach becomes feasible. Other selection indices were suggested but either do not use all the available information or are suitable only for evaluation of the offspring of the sire from which the marker information was established.

A selection index combining marker and Animal Model information was developed to allow comparisons involving offspring, grandoffspring and great-grandoffspring of a sire. Marker information was assumed to be a least squares estimate of the difference between the average effects of the two quantitative trait loci (QTL) alleles present in a sire (D_p) and the standard error of this estimate ($SE(D_p)$). Estimates may have been obtained

from a daughter or granddaughter design. Comparisons among grandoffspring and great-grandoffspring also require an estimate of the recombination rate (r) between the marker and the QTL. The Animal Model information consists of predicted transmitting ability (PTA) and reliability of PTA. PTA was assumed not to include any marker information. The expected percentage of the gain in accuracy (PGA) due to the inclusion of marker information in the selection indices is affected by the degree of polymorphism at the marker locus. The polymorphism information content (PIC) of a marker locus was computed for the second and third generations and for mates genotyped or not. PGA increased with larger D_p, lower $SE(D_p)$, lower r, a smaller number of own and progeny records, and larger PIC. PGA and PIC reduce over generations. Marker information in dairy cattle is likely to be used in generations beyond offspring. Then, only the use of highly polymorphic markers with a large and accurately estimated effect may be economically justified.

Acknowledgements

I would like to thank Ina Hoeschele for serving as chairman of my committee and for sharing her knowledge of genetics with me. I also appreciate the time and effort Dr. Dunnington, Dr. Hallerman, Dr. Notter and Dr. Pearson contributed to my thesis.

Thanks to Alan Pasquino and Cindy Cassady for their computing assistance.

Thanks to all the members of the Department of Dairy Science for the encouragement and support they gave me.

Thanks to my parents, Pedro Hector and Elisa, to my sister Silvia and friends for being always with me.

Thanks to Fito Cantet and Eduardo Manfredi, two excellent animal breeders and better friends.

And my entire gratitude and love to my wife Daniela, who supported, encouraged and made me laugh and be happy all the way through.

Table of Contents

List of Illustrations

List of Tables

List of abbreviations

α_i .. average effect of QTL allele i

$\overline{\alpha}_d$ .. average effect of the QTL alleles in the mates of the sire

$\overline{\alpha}_{gd}$ .. average effect of the QTL alleles in the mates of the grandsire

$\overline{\alpha}_{ggd}$. average effect of the QTL alleles in the mates of the great-grandsire

BV ..... breeding value

DE ..... daughter equivalent

D_g ..... difference in the average effect of the two QTL alleles (1,2) in a sire ($D_g = (1-2r)(\alpha_1-\alpha_2)$)

D_p ..... observed or estimated phenotypic value of D_g

DYD .... daughter yield deviation

GC ..... genotype informative for the current generation

GGS_i ... great-grandsire i

GN ..... genotype informative for the next generation

GPA .... grandparent average

GS_i .... grandsire i

M_i ..... marker allele i

O_j ..... offspring j

PA ..... parent average

PGA .... percentage of gain in accuracy

PIC .... polymorphism information content

PTA .... predicted transmitting ability

Q_i ..... QTL allele i

QTL .... quantitative trait locus or loci

r ..... recombination rate between a marker and a QTL

$r^2_{(adj)}$... equilibrium squared accuracy of the selection index

Rel(PA) reliability of parent average

Rel(PTA) reliability of PTA

Rel_s^* .. selection adjusted (asymptotic) reliability of sires

Rel_d^* .. selection adjusted (asymptotic) reliability of dams

S_i ..... sire i

σ_a^2 .. additive genetic variance

σ_s^* .. adjusted (asymptotic) additive variance among males

σ_d^* .. adjusted (asymptotic) additive variance among females

S_g ..... mendelian segregation effect explained by the marker

S_p ..... observed mendelian segregation effect

TA ..... transmitting ability

x_1, x_2, x_3 weights for PA, YD and DYD used to compute PTA

YD ..... yield deviation

Introduction

The correct identification of animals of superior genetic merit is the basis of any successful breeding program. The accuracy of current genetic evaluations is based on the availability of performance records from closely related animals. The existence of genetic markers that identify genes affecting traits of interest may enhance the precision of the evaluations, especially when few records are available.

The search for genetic markers is not new, although until the late 1970's almost all the available markers were related to protein polymorphisms (Jeffreys, 1987). Since then, the DNA-marker technology has made the direct identification of polymorphisms possible (e.g., Soller and Beckmann, 1982).

Because DNA markers are not influenced by environmental effects (Hallerman et al., 1986), they are prime candidates for use in breeding programs (e.g., Soller, 1978; Soller and Beckmann,

1982, 1983; Hallerman et al., 1986). The potential for genetic evaluation free of environmental effects let some researchers envision future genetic evaluations based solely on marker information (e.g., Soller, 1978). Beckmann (1988) denoted this use of marker information as "genomic" genetics, in comparison with "mendelian" genetics based on phenotypic information.

Further research found some limitations to "genomic" genetics that dismissed such an optimistic forecast (e.g., Stam, 1987). Some of these limitations are likely to be reduced over time. For instance, the availability of highly polymorphic markers that allow the marker allele inheritance to be traced across generations should increase; and the high cost of genotyping will likely decline (Brascamp et al., 1992). However, there are other limitations to "genomic genetics" that will be more difficult to overcome.

The detection of linkage between a marker and a locus of interest involves the application of a statistical test that ensures that chance was not likely the cause of linkage among the studied loci (e.g., Soller, 1978). Broadly speaking, if some difference in performance between groups of animals receiving alternative marker alleles from a common parent is declared significant, then the marker is said to be linked to a locus affecting the animal's performance.

The power of this test (i.e., the probability of declaring a true marker-QTL association statistically significant) increases with the size of the effect of the locus of interest or the difference in performance between the groups of animals receiving alternative marker alleles). However, most of the effects of the loci affecting a quantitative trait are likely to be of small size (Hoeschele and Van Raden, 1993).

The availability of a large number of phenotypic records also increases the power of the test. Species with low reproductive rate and large generation interval may make the task of establishing linkage between a marker and a quantitative trait more difficult.

Although it is unlikely that "genomic genetics" will replace "mendelian genetics" for selection purposes, genetic markers can be used together with phenotypic records to increase the accuracy of the genetic evaluations. The utilization of both marker and phenotypic information in selection has been termed marker-assisted selection (MAS) (Geldermann, 1975).

Selection indices have been suggested in order to combine the information from markers and phenotypes (e.g., Soller, 1978; Kashi et al., 1990a). However, these indices did not include all the available phenotypic information (Kashi et al., 1990a), or were appropriate only for evaluations of the offspring of the animal

from which the marker information had initially been established (Lande and Thompson, 1990). Fernando and Grossman (1989) suggested the inclusion of marker information in the current Animal Model evaluations. However, further research is needed before this approach can be computationally feasible and safely implemented.

Many genetic markers have been found in dairy cattle (e.g., Georges and Massey, 1991) and biotechnology companies have started to advertise the use of some genetic markers (e.g., Genetic Visions, 1991). Questions about the use of these markers have arisen among farmers and artificial insemination (A.I.) companies. In order to answer some of these questions, an appropriate way to combine marker and performance information is needed.

The usefulness of the marker information is reduced by a limited polymorphism at the marker locus (e.g., Botstein et al., 1980). For low polymorphic markers, the inheritance of marker alleles may not be traceable across generations.

The fraction of marker genotypes in the offspring of a sire with traceable inheritance of the allele received from the sire is defined as the polymorphism information content (PIC) of a marker (Botstein et al., 1980). The expected gain in accuracy due to the use of such marker information is reduced by a factor (1-PIC). Botstein et al. (1980) and Dekkers and Dentine (1991) computed PIC for the offspring of an individual for which the marker information

had been established. Kashi et al. (1990a) extended the computation of PIC to the grandoffspring of a sire when mates are genotyped. However, they did not distinguish between marker genotypes that are informative with respect to marker allele inheritance in their own or the next generation.

This study includes the following objectives:

1. To develop selection indices that combine marker and all the available Animal Model information.

2. To assess the conditions under which the inclusion of marker information in the selection indices developed under objective 1. is useful.

3. To compute PIC for the grandoffspring and great-grandoffspring of a sire, distinguishing among cases where mates are genotyped or not genotyped.

Literature review

L.1. Genetic markers

L.1.1. Different types of genetic markers

Restriction fragment length polymorphism (RFLPs)

Among all the different types of DNA markers now available, restriction fragment length polymorphism (RFLPs; Botstein et al., 1980) were the first to be found. The development of RFLPs started with the discovery that bacteria utilize restriction endonucleases to cut foreign, invading DNA at particular base sequences (Smith, 1979). When added to DNA solutions from any species, these enzymes recognize those base sequences and cleave DNA into fragments of different lengths that can be separated by gel electrophoresis; the fragments run a different distance through the gel according to their length. These fragments can then be

hybridized with a radioactively-labeled probe and visualized as bands by autoradiography (Smith, 1979). This method allows distinction between fragments differing in as little as one nucleotide (Jeffreys, 1987). Fragments of different length are interpreted as different alleles (Stryer, 1988).

Alternative fragments or RFLP alleles at a marker locus may be linked to alleles at a locus affecting some quantitative trait of interest. If such a linkage is found, the inheritance of the respective alleles affecting the trait of interest can be monitored through tracing the RFLP's allelic inheritance (e.g., Hallerman, 1989).

However, monitoring inheritance at a marker locus sometimes may be impossible (e.g., Botstein et al., 1980). For instance, when only two marker alleles are segregating in the population (i.e, when the marker locus is dimorphic), all the heterozygous offspring of heterozygous parents will not be informative about the marker allele inheritance (e.g., Kashi et al., 1990a). The marker alleles present in this heterozygote offspring cannot be assigned to either of the parents (e.g., Botstein et al., 1980). Also, the smaller the number of alleles, the larger is the proportion of homozygous individuals in a population. A homozygous parent will not allow the tracing of transmission of its alleles to offspring (Soller, 1990). The traceability of a marker's inheritance increases with the number of alleles segregating at the marker

locus (i.e., with the degree of polymorphism of the marker locus). Dimorphic markers can show a maximum possible heterozygosity of 50% (Jeffreys, 1987). Many RFLPs at functional gene loci are dimorphic (e.g., Hallerman et al., 1988a), which reduces their usefulness as DNA markers.

RFLP haplotypes

The existence of closely linked RFLPs ("haplotypes") may allow researchers to use DNA markers with a higher degree of polymorphism (Soller and Beckmann, 1983). The usefulness of these haplotypes is related to the allelic frequencies of the individual RFLPs and the degree of linkage among them (Soller, 1990). Steele and Georges (1991) found a mean heterozygosity of 51.9% for the haplotypes that they identified in cattle. Haplotyping requires the utilization of several restriction enzymes, which increases cost and the amount of DNA required (Soller, 1990; Steele and Georges, 1991).

Variable number of tandem repeats (VNTRs)

Another type of DNA marker is the variable number of tandem repeats (VNTRs, Jeffreys, 1985). Here, repeated hypervariable sequences are used to probe DNA (Jeffreys, 1985). VNTRs differ from RFLPs in that they are not based on mutations of recognition sites of endonuclease, or deletions, insertions or duplications of unique sequences, but on variation in the number of tandem repeats

of short sequences (Fries et al., 1989).

The polymorphism of VNTRs is generally higher than that of RFLPs (Fries et al., 1989). Jeffreys (1987) pointed out that while dimorphic RFLPs have a maximum heterozygosity of 50%, VNTRs can approach almost 100% heterozygosity. However, Georges et al. (1991) found a mean heterozygosity of 59% for 36 VNTRs that they had isolated from U.S. Holstein cattle.

The usefulness of VNTRs in cattle would be reduced if they tended to show the proterminal chromosomal confinement found in humans (Steele and Georges, 1991). However, Georges et al. (1991) did not find this type of a non-random distribution of VNTRs in the bovine.

DNA fingerprints

DNA fingerprints are derived from certain VNTRs that, under special conditions of hybridization, produce a complex banding pattern (Jeffreys et al., 1985a). DNA fingerprints can be visualized as multilocus VNTRs. The banding pattern they produce is specific for each individual and is, hence, referred to as a "DNA fingerprint" (Jeffreys et al., 1985a).

Georges et al. (1988) estimated that the probability of two randomly selected individuals having the same DNA fingerprints is

1.4×10^{-11} in cattle.

DNA fingerprints have proven very useful for parental identification (Jeffreys et al., 1985a). Fries et al. (1989) pointed out that some DNA fingerprints do not allow a clear allelic interpretation of a banding pattern, reducing the usefulness of such markers (Fries et al., 1989).

Oligonucleotide polymorphism (OP)

OPs (Beckmann, 1988) use hybridization of synthetic oligonucleotides, which are sequences of 15-20 base pairs (Hallerman, 1989). If the target sequence is known, allele specific oligonucleotides can be constructed to detect single base changes. Several tightly linked OPs could be used to create "microhaplotypes" (Beckmann, 1988) of high heterozygosity. The requirement of known DNA sequence makes OPs appropriate for a targeted search for polymorphism (Fries et al., 1989).

Markers based on the polymerase chain reaction (PCR)

The polymerase chain reaction (PCR; Litt and Lutty, 1989) allows the amplification of targeted DNA sequences, increasing the amount of DNA available for analysis. PCR allows the performance of marker analysis on embryos (e.g., Kashi et al., 1990a, Georges and Massey, 1991). Amplified fragment length polymorphism

(AMP-FLP) is a technique developed to separate and detect PCR-amplified RFLPs. This technique does not require the use of radioactive labeling or molecular hybridization techniques, which reduces the cost of genotyping (Kirby, 1990).

L.1.2. Polymorphism at the marker locus

Botstein et al. (1980) defined the polymorphism information content (PIC) of a marker as the probability of finding individuals with a traceable marker inheritance. Markers with higher PIC allow the establishment of marker allele inheritance for a larger fraction of individuals in the population (Soller, 1990).

Dekkers and Dentine (1991) presented formulae to compute PIC in offspring of different parents. They differentiated among cases where either one or both parent were genotyped. They showed that, for dimorphic markers and both parents genotyped, maximum PIC is 37.5%, which occurs when both marker alleles have a frequency of .5 in the population. When only one parent is genotyped, PIC is further reduced, approaching zero for certain allelic frequencies (Dekkers and Dentine, 1991).

Dekkers and Dentine (1991) found that dimorphic markers show a larger PIC at intermediate frequencies, because the probability of finding animals which are heterozygous at the genetic marker is

larger. Individuals that are heterozygous at both marker and QTL are needed to establish linkage (Jacquard, 1970). However, for a dimorphic marker, given a marker heterozygous parent, PIC within the descendants of this parent is higher for markers with extreme allelic frequencies in the population, because then most of the mates of the heterozygous parent will be homozygous (e.g., Hallerman et al., 1988a).

The probability of finding individuals with traceable marker inheritance increases with the number of marker alleles segregating in the population (e.g., Botstein et al., 1980). Dekkers and Dentine (1991) showed that when highly polymorphic markers are available, genotyping both parents does not increase PIC significantly.

Kashi et al. (1990a) extended formulae for PIC given by Dekkers and Dentine (1991) to cases where marker alleles are traced back from a grandsire to his grandoffspring. They developed their formulae by assuming that both parents were genotyped.

Most RFLPs at functional gene loci are dimorphic (e.g., Hallerman et al., 1988a; Soller, 1990). Beckmann et al. (1986) found that 92% of the observed RFLPs in Holstein were dimorphic. This figure is similar in humans (Jeffreys, 1987). The dimorphism of RFLPs reduces their usefulness as genetic markers (Jeffreys, 1987).

Hallerman et al. (1988b), studying several RFLPs in the Israeli Holstein-Friesian dairy cattle breed, found that most RFLPs showed a frequency (q) of the less common marker allele of less than 0.10. Following Dekkers and Dentine (1991), first generation PIC of these markers is expected to be 0.16 and 0.09 when mates are genotyped and not genotyped, respectively.

L.1.3. Present availability of genetic markers

Molecular markers of different types have been established within the genome of many species, including humans (e.g., White et al., 1987), cattle (e.g., Steele and Georges, 1991), sheep (e.g., Sunden et al., 1992), swine (e.g., Zawadski and Johnson, 1992), horse (e.g., Ellegren et al., 1992), chicken (e.g., Dunnington et al., 1992), turkey (e.g., Foster and Foster, 1991), duck (e.g., Goddard and Boswell, 1991), Atlantic salmon (e.g., Knox and Verspoen, 1991) and mink (Xiong et al., 1992).

Most of the efforts in molecular genetics have been applied to the human genome (e.g., Steele and Georges, 1991). The availability of genetic markers for the human genome is no longer a limitation for linkage analysis (White et al., 1987). The current key limiting factor for genetic marker analysis is the availability of sufficient family material (Ploughman and Boehnke, 1989).

A project to map the human genome by using molecular markers is underway (White et al., 1987). Because of the conservation of linkage groups among across mammalian species, the development of the human genomic map will probably help in the development of genomic maps of other species (Hallerman, 1989; Womack and Moll, 1986; Steele and Georges, 1991; Kirkpatrick, 1992). Sunden et al. (1992) found that molecular markers obtained in sheep are useful in bovine, and vice versa, which will also enhance the development of genetic maps in both species. The development of a genetic map in swine (Rohrer and Beattie, 1992; Scheid et al., 1992) and in chicken (Crittenden et al., 1992) has been initiated.

The construction of a genetic map in bovine is underway (Fries et al., 1989; Laster and Beattie, 1992; Bishop et al., 1992). After man and mouse, the bovine is currently the mammalian species with the best characterized genetic map (Georges et al., 1990). Steele and Georges (1991) reported that their laboratory has isolated more than 225 DNA markers for cattle, including 82 multisite haplotypes, 40 VNTRs, and more than 50 microsatellites, with mean heterozygosities between 50% and 60%. These markers were estimated to cover around 70% of the bovine genome, assuming 30 Morgans as the size of the cattle genome (Steele and Georges, 1991).

As the marker coverage of the genome increases, efficiency of the search for additional random markers to fill the remaining gaps

decreases and targeted markers are needed (Steele and Georges, 1991). A targeted search requires prior development of the genetic map (Hallerman, 1989). The development of the human genomic map might help in this approach (e.g., Hallerman, 1989).

L.2. Marker applications

2.1. In humans

In humans, genetic markers have been suggested for use in paternity testing (Jeffreys et al., 1985b), forensic medicine (Gill et al., 1985), determination of twin zygosity (Hill and Jeffreys, 1988), and development of more precise knowledge of the genetic basis of diseases (e.g., Alzheimer's disease, St. George-Hyslop et al., 1990; atherosclerosis, Sing and Moll, 1990; certain types of tumors, Orkin, 1986).

L.2.2. In plants and animals

The experience gained in the study of humans is being adapted and used for other species, including plants (e.g., Paterson et al., 1990) and animals (e.g., Kennedy et al., 1990). The analysis of genetic markers in plants may be favored by the possibility of

selfing or the use of inbred lines (Soller, 1978; Soller and Genizi, 1978). Linkage disequilibria between pairs of loci may be produced by hybridization (Falconer, 1989). Crossing of inbred lines may be useful to create linkage disequilibria, since they are expected to differ at some of both the segregating quantitative trait loci (QTL, Geldermann, 1975) and marker loci (Soller et al., 1976). The crossing of inbred lines is a common technique in plants (van Arendonk and van der Beek, 1991) and in chickens (Soller et al., 1976).

Crossing of inbred lines followed by selection for the QTLs so identified was successfully performed in tomato by Nienhus and Helentjanis (1989). Smith (1991) pointed out however, that this success was in part due to a reselection of the favorable QTL alleles established by earlier selection. The detection of favorable QTL in different lines and subsequent combination of these QTL alleles in one line will be more useful (Smith, 1991).

Linkage disequilibria cannot be created by crossing lines in livestock, because highly inbred lines are not available (e.g., Beckmann and Soller, 1983; van Arendonk and van der Beek, 1991). In most cases, the analysis of genetic markers in livestock has to be restricted to segregating populations, as in dairy cattle (van Arendonk and van der Beek, 1991).

When the marker analysis is performed within a population,

analyses must be carried out within families (Smith and Simpson, 1986).

One suggested marker application is to monitor the introgression of desirable genes trait into a population (e.g., Beckmann and Soller, 1983; Lanneluc et al., 1992). Hillel et al. (1990) showed that marker information permits a reduction in the number of backcross generations required to introgress a favorable gene into a population. Animals with a desirable genome can be identified by genetic markers and selected as parents of the next backcross generation. Beckmann and Soller (1983) estimated that the number of backcross generations required to introgress a marked QTL in a population can be reduced from five to one or three by the application of marker information.

Marker information may also be useful in the analysis of inbreeding (Kuhnlein et al., 1990; Li et al., 1992). Kuhnlein et al. (1990) proposed to investigate the average sharing of DNA bands, denoted as "bandsharing", as a measure of the degree of inbreeding. Groen (1991) showed that, although a relationship between bandsharing and inbreeding exists, bandsharing can be an appropriate criterion to assess inbreeding levels only if knowledge about the allelic frequencies is available.

The use of bandsharing as an indicator of the relatedness of individuals makes DNA fingerprints suitable for the estimation of

genetic distances among breeds or species (Hillel et al., 1989; Gilbert et al., 1990; Haberfeld et al., 1992).

DNA fingerprints were also suggested for increasing the chances of survival of endangered species by choosing matings that reduce the level of inbreeding in small populations (Hillel, personal communication).

It has been suggested that marker information be used to predict heterosis, a phenomenon related to inbreeding, in crosses between breeds or individuals (Smith, 1991). Smith et al. (1990) attempted successfully to predict heterosis in corn, but Smith (1991) pointed out that this prediction was obtained from crosses where heterosis was well known which might have increased the precision of the prediction.

As in humans, DNA fingerprints were suggested for parental identification in animals (e.g., Kashi et al., 1990b, Yemm et al., 1992). Nichols and Baldy (1990) pointed out that this application of DNA fingerprints might lead to false conclusions if the population structure (i.e., the existence of heterogeneous subgroups in the population) is not taken adequately into account. Kennedy et al. (1990) suggested that parental identification by DNA fingerprints might increase genetic gains in largely maternal traits in swine, where DNA fingerprints may be used to clearly identify paternity from mixed inseminations.

Molecular markers were also found in mitochondrial DNA. Analysis of mitochondrial DNA is used for identification of certain species or subspecies groups of fish (e.g., Shields et al., 1992; Seyoum and Kornfield, 1992). Mitochondrial DNA is prefered to nuclear DNA in these cases, because its elevated mutational rate and maternal pattern of inheritance help in the differentiation of species or genetic stocks (Knox and Verspoen, 1991).

Genetic markers can also be applied to the study of the genetic basis of certain diseases. Genetic markers were found to be linked to genes causing the porcine stress syndrome (Smith and Bampton, 1992; Georges and Massey, 1991), the spider lamb syndrome (Shay et al., 1992), the Weaver locus in Brown Swiss (Hoeschele and Meinert, 1990), the bovine leukocyte adhesion deficiency (BLAD) (e.g., Shuster et al., 1992), and protoporphyria in cattle (Dean et al., 1992). DNA fingerprints were also used to characterize the different types of bacteria causing mastitis in dairy cattle (Jayarao et al., 1991, 1992) and other diseases (e.g., Sniper et al., 1992; Seyoum and Kornfield, 1992).

Molecular genetics is also used to achieve a better understanding of some physiological phenomena. Markers have been applied to study the variability of genes related to the expression of growth hormone (e.g., Guillemot and Juffrey, 1989; Kirkpatrick et al., 1990; Tuggle et al., 1992; Zhang et al., 1992b), the bovine major histocompatibility complex (e.g., Green et al., 1992); the

halothane resistance gene in pigs (e.g., Zhang et al., 1992a; Zawadski and Johnson, 1992); muscle protein genes that affect carcass characteristics of beef (Medrano and Famula, 1990; Winkelman and Schmutz, 1992); the genetics of the prolactin gene in cattle (e.g., Hallerman et al., 1988a) and genes regulating the secretion of pituitary glycoprotein hormones in turkeys (Foster and Foster, 1991).

Another suggested use for marker information is in genetic improvement of economically important quantitative traits of livestock through marker-assisted selection (MAS, Soller and Beckmann, 1982). Smith (1991) predicted that MAS will be the main use of marker information in dairy cattle.

MAS is already applied in specific instances, including sexing preimplantation embryos in cattle with Y-specific probes (e.g., Bishop and Woolliams, 1991), or reducing the frequency of the major gene causing porcine stress syndrome (e.g., Georges and Massey, 1991).

Georges and Massey (1991) suggested that MAS in dairy cattle may shorten generation intervals dramatically by application of PCR techniques at the embryo level and using this information in selection. They termed this process velogenetics.

L.3. MAS in dairy cattle

The use of genetic markers in dairy cattle was suggested more than a decade ago (Soller, 1978). He proposed that animals be selected based on "known loci". A locus affecting some trait of interest would become eventually "known" through closely bracketed markers and phenotypic information accumulated over years (Beckmann and Soller, 1983). However, Soller (1978) concluded that the probability of identifying a "known locus" is very small.

MAS implies selecting animals for a quantitative trait by combining production records assisted with marker information. MAS is expected to increase genetic gains by improving selection accuracies or by reducing generation intervals or both (Smith and Simpson, 1986).

Because marker information can increase the accuracy of genetic evaluations, it has been suggested that MAS be applied to young bulls before entering a progeny test (e.g.,Soller, 1978; Beckmann and Soller, 1983; Hallermann et al., 1986; Stam, 1986; Beckmann and Soller, 1987; Kashi et al., 1990a). Young bull selection in dairy cattle is presently a two-stage process. In the first stage, young bulls are selected on pedigree information. Then, the Mendelian component of each bull's genetic merit is estimated through a progeny test. It was suggested that marker

information be used in an intermediate stage of selection (e.g., Hallermann et al., 1986, Beckmann and Soller, 1987; Kashi et al., 1990a) in order to exploit early the Mendelian sampling variance (e.g., Dekkers and Dentine, 1991). The implementation of an intermediate stage of selection may reduce the response to selection in the final stage, but this reduction may be more than compensated by the earlier selection (Lande and Thompson, 1990). If markers can explain a sufficient amount of the Mendelian sampling variance, a reduction in the number of bulls to be progeny tested annually may be possible and a reduction in the cost of maintaining undesirable bulls may be achieved (Lande and Thompson, 1990).

Soller and Beckmann (1982) suggested that the full exploitation of 20 marked QTLs for milk production may, over a relatively short period of time, yield eventual production increases in the order of 2000 to 3000 kg per lactation. Stam (1986) estimated that MAS may allow up to 40% additional genetic gain per generation. This figure was obtained by assuming unrealistically that there is just one single multiallelic QTL locus which is responsible for all the quantitative variation. When more realistic assumptions are applied, this figure drops significantly (e.g., Stam, 1987). Kashi et al. (1990a) estimated that MAS of young bulls before progeny testing would be able to increase the rate of genetic gain by 15-25 % per generation if highly polymorphic markers were available. The authors assumed

that there were 5-20 segregating QTLs with similar effects and recombination rates, selection was on the number of favorable marker alleles inherited, the information coming from the markers was independent of the animal model information, and there was no reduction in the genetic variance among the selected bulls that enter MAS (Smith, 1991). Hence, Kashi et al.'s (1990a) figures may be considered as an upper limit in the gain that can be achieved by MAS in progeny testing schemes in dairy cattle.

Meuwissen and van Arendonk (1992) expectedly found negligible increases in gain when marker information was utilized together with progeny test results. In this case, Mendelian sampling variance is explained by both the progeny test and the marker information, and there is little additional gain due to markers.

Actual gains from MAS in dairy cattle will, likely, be somewhere between the two extremes. Gibson (1992) believes that marker information may help to increase genetic gains in livestock, but that this gain will likely be small under current breeding programs. Important genetic gains can be achieved only if QTLs of large effect are used (Smith, 1991). Larger gains may also be achieved with appropriate changes in breeding schemes, e.g., by increasing the number of progeny available for MAS (Gibson, 1992), or by applying MAS to a nucleus scheme (e.g., Meuwissen and van Arendonk, 1992).

Application of marker information to juvenile nucleus breeding schemes together with multiple ovulation and embryo transfer (MOET) may produce larger genetic gains than conventional progeny testing schemes (Woolliams and Smith, 1988; Meuwissen and van Arendonk, 1992; Dekkers, 1992b). Selection of young animals in juvenile nucleus schemes occurs before first breeding (Nicholas and Smith, 1983). The lack of production information increases the value of marker information in these schemes (e.g., Meuwissen and van Arendonk, 1992).

Smith and Simpson (1986) pointed out that utilization of marker information may depend on the underlying biological situation. Dominance and epistatic variation at QTL may exist. Dominance effects at a marked QTL were found by Dunnington et al. (1992) in chickens. Hyland and Quaas (1991) showed that epistasis reduces the response to MAS. Smith and Simpson (1986) pointed out that epistasis can occur between two marked QTLs, or between a marked QTL and other QTLs that have not been marked. Marked QTLs may have pleiotropic effects on several traits of economic importance. Genotype-environment interaction may cause marked QTL alleles to perform differently in different environments (Knapp, 1992). Smith and Simpson (1986) noted that all these potential complications may need to be considered and investigated in order to take full advantage of the marker information.

Several genetic markers are already used in dairy cattle. Some marked QTL are associated with diseases. Jayarao et al. (1991 and 1992) used genetic markers to identify different types of bacteria causing mastitis. Hoeschele and Meinert (1990) found the locus causing bovine progressive degenerative myeloencephalopathy (Weaver) in Brown Swiss has a positive effect on milk and fat yield. A marker for this locus has been found (M. Georges, personal communication).

Genes affecting the protein composition of milk have also been marked. The kappa casein locus affects protein yield and fat and casein content (e.g., Cowan et al., 1992b; Sabour et al., 1992). Gomez Raya and Gibson (1991) found that the average difference in percentage of protein in milk between marker genotypes AA and BB is around 3%. An RFLP marker for the prolactin locus was identified by Cowan et al. (1990) and Cowan et al. (1992a). There was a difference of 283 kg in milk yield among individuals bearing the alternative RFLP alleles.

Some markers are currently available and advertised by biotechnology companies. Genetic Visions (Madison, Wisconsin) advertises markers for the kappa casein, beta-lactoglobulin and prolactin loci (Chro-Mo-Probe) (e.g., Brown Swiss Bulletin, September 1991). Genmark (Salt Lake City, UT) also advertises kappa-casein and beta-lactoglobulin markers (Holstein World, August 1992), a marker for bovine leukocyte adhesion deficiency (BLAD)

(e.g., Holstein World, July 1992), and a marker for the Weaver defect, as well as markers for embryo sexing and parental fingerprinting.

L.4. Factors affecting the usefulness of the marker information

The usefulness of marker information is related to several factors, which are discussed below.

L.4.1. Heritability of the trait

Marker information tends to be more useful for traits of low heritability (Smith and Simpson, 1986; Lande and Thompson, 1990). The higher the heritability of the trait, the more accurate is the phenotypic information. In such cases, the availability of marker information may not add any significant information (e.g., Smith and Simpson, 1986). However, marker information is less likely to be found for traits of low heritability. A larger data set and larger QTL effect will be required to compensate for the reduced power of the test when the trait is largely affected by non-genetic effects.

L.4.2. Traits difficult to measure in the same individual

Marker information increases its value when it is applied to sex-limited traits (e.g., milk production) to juvenile individuals before the development of the adult phenotype, or to traits difficult to measure on the live individual (i.e., carcass quality) (Lande and Thompson, 1990).

L.4.3. Percentage of Mendelian sampling variance explained by the QTLs (size of the marked QTL effects)

Marker information becomes more valuable as more of the genetic variance is explained by the available marked QTLs (Lande and Thompson, 1990). The percentage of the genetic variance explained depends on the number of marker-QTL linkages found and on the size of the effects of individual QTLs (Lande and Thompson, 1990). It is very unlikely that all segregating QTL can be marked (Smith and Simpson, 1986).

To find a marked QTL, several conditions are to be met: an individual heterozygous at both the marker and the QTL must be found (Jacquard, 1970); this individual must have a large family that provides records for the analysis (Lander and Botstein, 1989); and the marked QTL must explain a minimum portion of the genetic variance (Lander and Botstein, 1989) to be detectable, given the

family size.

Marked QTLs with a major effect on the trait of interest would be very useful and easier to detect, but they may be not available or have deleterious pleiotropic effects (Zhang and Smith, 1992). The use of marker information on major genes will produce rapid genetic improvement in the short term, although in the long term genetic gain may be reduced due to a loss in selection effort in the polygenic variation that cannot be recouped later (Gibson and Jansen, 1990; Saeffundin and Gibson, 1991).

It is likely that some QTLs of small effect might be located close to each other within the genome, forming a cluster of loci or a superlocus (Smith and Simpson, 1986; Dentine and Cowan, 1990). This cluster of linked loci could explain a detectable portion of the genetic variance. A marker linked to a superlocus could monitor the inheritance of this segment of the chromosome. Dentine and Cowan (1990) termed the difference in genetic merit between those offspring receiving the alternative chromosome segments as "chromosome substitution effect". For purposes of selection, a marker locus may be useful regardless of its linkage to one or to a cluster of QTLs, but inferences on future generations may require differentiation among both cases (Dentine, 1990).

The actual value of the variance associated with marker information ranges from zero (no linkage at all) to half the

genetic variance when marked QTLs perfectly explain all the Mendelian variance in the offspring of two parents carrying marker information (Dekkers and Dentine, 1991). Dekkers and Dentine (1991) estimated that with one highly polymorphic marker on each of the 30 bovine chromosomes, 30-40% of the Mendelian sampling variance coming from both parents may be explained. For dimorphic markers this estimate reduced to less than 15% because of small PIC. The actual percentage depends on the other factors considered here.

L.4.4. Recombination

It is very unlikely that a marker locus has a direct effect on the quantitative trait (Beckmann and Soller, 1987). Thus, most cases will involve linkage between a marker and a QTL. Marker-QTL recombination reduces the effectiveness of MAS (Soller et al., 1976; Stam, 1986 and 1987; Smith and Simpson, 1986; Lande and Thompson, 1990). Stam (1986) estimated that with 15% recombination, approximately 70% of the genetic gain that could be obtained by MAS in the absence of recombination is retained. Recombination between marker and QTL reduces the size of the marker effect for a given QTL effect and, hence, decreases the chance of detecting a marker-QTL linkage (e.g., Beckmann and Soller, 1987). The case of flanking markers, i.e., markers situated on either side of the QTL, would reduce this problem (e.g., Kashi et al., 1990a).

L.4.5. Experimental designs for linkage detection

Linkage between a marker locus and a QTL can be tested by comparing the phenotypes for a trait of interest among the descendants of an individual heterozygous for the marker locus (e.g., Soller and Beckmann, 1982). Given that marker-QTL linkages are not expected to exist on a population level, linkage analysis must be performed within families (e.g., Smith and Simpson, 1986; Stam, 1987).

If performance differences among descendants that receive different marker alleles from the common parent are larger than expected by chance, then linkage between the marker locus and the QTL is declared (e.g., Soller and Beckmann, 1982). The sample size required to detect marker-QTL linkage may be a limiting factor for practical linkage analysis. In dairy cattle, only a limited number of bulls have sufficient number of descendants for the analysis. Soller (1990) computed that within the U.S. Holstein population there are only 46 sires with sufficient number of daughters for a linkage analysis. Romano (unpublished) estimated that this number has to be reduced to 13 if sons are used and the current availability of semen for the test is also considered.

To reduce the sample size required for linkage analysis, it was suggested that only individuals with extreme performance phenotypes be genotyped (Soller and Beckmann, 1983). Lander and

Botstein (1989) termed this approach "selective genotyping" and it is applied in chickens as "tail analysis" (e.g., Dunnington et al., 1992). Individuals with intermediate phenotypes contribute less information on marker-QTL linkage than those with extreme phenotypes. Lander and Botstein (1989) pointed out that individuals more than one standard deviation from the mean contribute 81% of the information, although they comprise only 33% of the population. Application of selective genotyping reduces the number of offspring to be genotyped, but increases the required total number of offspring per family needed. It is, therefore, of little value when offspring group sizes are more limiting than the cost of genotyping.

Plotsky et al. (1990) suggested comparison of mixtures of DNA from individuals at both extremes of the phenotypic distribution. Particular differences in the banding pattern between these two groups of mixtures may indicate the presence of one or more QTL associated with these bands.

In dairy cattle, two designs have been suggested for a linkage test. They are known as "daughter" and "granddaughter" designs (Weller et al., 1990). Their names refer to the descendants providing the records for linkage analysis. In the daughter design, the daughters of a sire are genotyped and grouped according to the marker allele they received. Differences in performance among the groups of daughters are compared. In the granddaughter

design, the sons of a sire are genotyped and grouped according to the marker allele they inherited. The production of the daughters of these sons (i.e., the granddaughters of the sire) provides the phenotypic records to test linkage between the marker and QTL. The power of the granddaughter design is larger than that of the daughter design (James, 1991). Also, the granddaughter design would require the collection of semen samples from bulls at AI centers, rather than of blood samples from cows scattered across many farms (Weller et al., 1990; Soller, 1990). Hoeschele and Van Raden (1993) suggested the use of progeny test or phenotypic averages of granddaughters of sons rather than individual records of granddaughters. Daughter averages are routinely computed by the U.S. Department of agriculture (USDA) as daughter yield deviations (DYD) of bulls (Van Raden and Wiggans, 1991). DYD are corrected for the effects of management and mates. The granddaughter design is an effective way to reduce the sample size for testing linkage with U.S. dairy cattle (Weller et al., 1990). However, the granddaughter design implies that by the time the performance records of granddaughters, i.e., progeny test of sons, are available for linkage analyses, the selection efforts of the dairy industry are aimed at the grandsons or great-grandsons of the sire for whom the marker-QTL linkages were established. Thus, the usefulness of the marker information from a granddaughter design is expected to be less than that of the daughter design (e.g., Hallerman et al., 1986).

I.4.6. Estimation of marker effects

Several procedures have been suggested for the estimation of marker effects. Soller (1978) used ordinary least squares (OLS) and Dentine and Cowan (1990) a generalized least squares procedure (GLS) including offspring with uncertain marker allele inheritance.

Maximum likelihood (ML) procedures have also been suggested (e.g., Weller, 1986; Jensen, 1989; Simpson, 1989; Lander and Botstein, 1989). Linkage between a marker and a QTL is tested by a likelihood ratio test and the value of this test statistic is denoted as lod score (Lander and Botstein, 1989). The ML approach allows simultaneous estimation of chromosome substitution effects and recombination rates (e.g., van Arendonk and van der Beek, 1991). Simpson (1989) showed that ML and OLS have similar power of detection when linkage between a marker and a QTL is tight.

All the approaches mentioned to this point treat the marker effect as fixed. Modeling the effect of marked QTLs as fixed leads to overestimation of the effects of the selected markers (Smith and Simpson, 1986; Hoeschele and Van Raden, 1993; Kennedy et al., 1992), similarly to treating sires or animals as fixed instead of random in genetic evaluations. Hoeschele and Van Raden (1993) proposed a Bayesian approach for linkage analysis. The Bayesian approach utilizes prior information on the probability of linkage and on the distribution of QTL effects (Hoeschele and Van Raden,

1993), whereas ML assumes that all QTL effects are equally linked a priori. QTLs of major effect are less likely to occur than QTLs of minor effect.

The level of significance used for testing linkage has also received some attention. As Lander and Botstein (1989) pointed out, testing many marker-QTL linkages for the same trait leads to a high probability of declaring at least one false linkage. For example, when ten unlinked markers are tested each at a 5% level of significance, the overall level of significance (i.e., the probability to detect falsely at least one linkage between any of the markers and a QTL) is $(1-.95^{10})$ and not .05. Smith (1991) suggested that, given this problem, the reliability of new markers should be tested with independent data before any application.

Lander and Botstein (1989) suggested following a classical statistical solution to this problem (e.g., Steel and Torrie, 1980) by setting a reasonable overall level of significance (e.g., 5 %) and then dividing it by the number of marker effects to test. The problem here is that if many markers are to be tested, the probability of detecting a single marker-QTL linkage is severely reduced. For instance, if the overall level of significance is 5%, then 10 markers are tested each at a 0.5% level of significance. Hoeschele and Van Raden (1993) pointed out that the prolactin locus linkage would have not been declared significant if Lander and

Botstein's (1989) approach had been applied instead of a type-I error of 5% (Cowan et al., 1990).

Hoeschele and Van Raden (1993) suggested that a Bayesian approach may deal with this problem also, since the posterior probability of linkage, which is a function of lengths of chromosomes, map function, total additive genetic variance, prior distribution of QTL effects and data, is a less stringent criterion for accepting linkage than that of Lander and Botstein (1989).

L.4.7. Economic analyses of genetic markers

The extent to which genetic markers will be used depends on the relationship between the benefits they produce and their costs. Brascamp et al. (1992) pointed out that the cost of MAS depends on the total number of markers to be evaluated, the number of animals to be genotyped, and the cost of genotyping for each marker. The cost of genotyping, however, is likely to be reduced over time (e.g., Hallerman et al., 1986). The development of new, cheaper laboratory techniques (e.g., PCR-based methods) is contributing to this reduction (E. Hallerman, Virginia Polytechnic Institute and State University, personal communication).

Genetic Visions (Madison, WI) advertised genetic markers for kappa casein, beta lactoglobulin and Chro-Mo-Probe at $40, $40, and $125 per sample, respectively (Brown Swiss Bulletin, 1991).

Brascamp et al. (1992) projected that the cost of genotyping may decline to as little as $1 per marker in the future.

MAS will be utilized more heavily not only if the cost of genotyping is reduced, but also if its application becomes more efficient. The availability of a few markers of moderate or large effects and tightly linked with the QTL, would make MAS more attractive (Smith, 1991). The implementation of new breeding strategies like juvenile nucleus schemes (e.g., Meuwissen and van Arendonk, 1992) or adaptation of progeny testing schemes to MAS, e.g., an increase in the proportion of animals selected in the first stage of selection (Gibson, 1992), will make MAS in dairy cattle more attractive.

Brascamp et al. (1992) estimated the financial returns from the genetic improvement of milk production due to MAS. They found returns from $7 to $21 per cow when marker information explaining different amounts of the Mendelian sampling variance was applied to progeny testing schemes. The authors also predicted that AI stud services will find the use of MAS economically attractive.

L.5. Evaluation of breeding values combining marker and phenotypic information.

Selection indices that include marker information can be used to estimate the genetic merit of animals (Soller, 1978; Soller and Beckmann, 1982; Smith and Simpson, 1986; Lande and Thompson, 1990; Kashi et al., 1990a; Meuwissen and van Arendonk, 1992). Kashi et al. (1990a) proposed a selection index based only on marker information to perform MAS of young dairy sires before progeny test. This index did not take advantage of all the existing information, i.e., did not consider the average predicted breeding value of the parents, and assumed markers to have equal effects. Soller (1978), Smith and Simpson (1986) and Lande and Thompson (1990) proposed selection indices combining molecular and phenotypic information. These indices were suitable for selection among offspring of an animal for which marker-QTL linkages were obtained.

Fernando and Grossman (1989) presented best linear unbiased prediction (BLUP) with an Animal Model including marker information. The authors partitioned the breeding value of an animal into the effect of the two alleles at the marker locus and the remaining breeding value due to the QTLs that were not marked.

By treating marker information as random, the problem of overestimating marker effects was reduced (e.g., Smith and Simpson, 1986; Goddard, 1992; James, 1991).

The approach of Fernando and Grossman (1989) required estimation of effects of marked QTL alleles for every animal in the population. For one marker, three genetic effects per animal were estimated, which are the QTL allelic effects plus the remaining breeding value. For m markers, (1+2m) times the number of animals is the number of genetic effects to be estimated. This implies a large increase in the size of the mixed model equations, and makes this approach computationally very demanding. To reduce this problem, Cantet and Smith (1991) proposed using a reduced animal model (RAM) where the QTL effects of non-parent individuals were absorbed. When the number of non-parent individuals in the population is not large, RAM does not reduce the number of equations substantially. Hoeschele (1993) suggested absorbing QTL effects for all animals that were not genotyped or did not provide relationship ties among genotyped descendants. For animals for whom there are marker data, no gain is expected from partitioning their breeding values into QTL allelic effect and residual breeding values (Hoeschele, 1993). Moreover, computing the inverse of the relationship matrix among the QTL effects of all animals becomes impracticable when many animals are not genotyped but have some genotyped relatives (Hoeschele, 1993). With a relatively high cost of genotyping, only animals of high value are expected to be

genotyped. Hence, with a small proportion of animals genotyped in the population, this approach may be computationally feasible (Hoeschele, 1993).

Although the Animal Model including genetic markers may become computationally feasible, there are other problems that must be considered prior to its adoption. Good estimates of variances at individual QTLs and of recombination rates between marker and QTL are needed (Fernando, 1990). Fernando (1990) pointed out that, given that recombination rates enter in the variances of QTLs, it might be difficult to estimate these parameters by classical animal breeding procedures. Also, because each animal is assumed to have unique effects of its two alleles at a marked QTL, repeated estimation of the same allelic effects may occur (Smith, 1991). This implies a loss of information which may be important if there are few alleles at a QTL in the population and limited number of records.

Goddard (1992) extended Fernando and Grossman's (1989) model to cases where flanking markers are available. Some reduction in the number of equations to be solved can be achieved if double recombination is ignored (Goddard, 1992).

Goddard (1992) suggested inclusion of a group effect in the Fernando and Grossman (1989) model as an ad hoc approach to deal with linkage disequilibrium in the population. The author pointed

out that in these cases, the alleles that come from each different lines have different expected effects. Group effects are expected to account for these differences.

Smith (1991) pointed out that the Animal Model, by taking all of the family information into account, produces more genetic gain than individual selection in the short run, but also more inbreeding. Marker information may reduce inbreeding rates in selected populations by reducing the correlation between estimated breeding values (BLUP) of individuals within families (Smith, 1991).

With an unlimited number of marker loci, markers close to the QTL will be surely found (Smith, 1991). Closer markers may help to analyze individual QTLs within a cluster (Smith, 1991). With a complete association between a QTL and a marker, the Animal Model may treat marker effects as fixed effects of known loci (e.g., Kennedy et al., 1990). Kennedy et al. (1990) suggested that effects of known loci can be included in a similar way as major gene effects (Kennedy et al., 1992). As Smith (1991) pointed out, "in the future, rather than devising and using complex low accuracy methods to exploit loose marker-QTL associations, the objective would be to locate the QTL. Then selection can be for the QTL themselves, which gives maximum selection accuracy and maximum selection response."

Finally, as Hill and Keightley (1988) and Kennedy et al., (1990) pointed out, constant development of quantitative, statistical techniques will be needed in order to cope with constant improvement in laboratory techniques, in such a way that a closing gap between molecular and quantitative genetics can be forecasted.

Methods

Marker and animal model information are combined into a single selection criterion using selection index theory. Marker information consists of the estimated unregressed difference D_p, the average effects of the two QTL alleles present in a sire., and the standard error of this estimate. Denote the true difference estimated by D_p as D_g , which is equal to $(1-2r)(\alpha_1-\alpha_2)$, where r is the recombination rate between marker and QTL, and α_i is the average effect of allele i (i=1,2) (e.g., Soller, 1978). The marker effect is assumed to have been estimated as a fixed effect within the family of a popular sire, which is heterozygous at both marker and QTL. The design used could have been either a "daughter" or a "granddaughter" design (Weller et al., 1990). The only difference between these two designs that is relevant to this study is the time when the marker information becomes available for commercial use.

It is assumed first that a single QTL is marked by one marker

locus. An estimate of the recombination rate between the marker and the QTL is assumed to be available. The estimated chromosome substitution effect (Dekkers and Dentine, 1991) may be due to a single QTL or a cluster of QTL. No distinction between these two cases is made here. The marked QTL is assumed to have an additive effect. Dominance or epistasis effects of marked QTL are not considered. The marker locus is assumed to be highly polymorphic (i.e., inheritance of marker alleles is assumed perfectly traceable). Later, the availability of marker loci with different polymorphism will be considered.

Marker information was suggested to be used in dairy cattle for comparisons among young bulls prior to progeny testing (e.g., Kashi et al., 1990a). The lack of phenotypic information beyond parent average (PA) on these animals justifies this suggestion. However, marker information could also be applied to bulls with initial progeny tests results, or to bull dams with own phenotypic records, although the availability of phenotypic data reduces the usefulness of the marker information. Selection indices for these different types of animals will be presented.

Marker information obtained from a sire through a "granddaughter design" will rarely be available for selection among the sire's offspring. Figure 1 shows that in theory, at the time when the marker information becomes available (i.e., when the sire has a sufficient number of progeny tested sons), selection among

the sire's sons and maternal grandsons has already occurred. Therefore, marker information from a granddaughter design is expected to be used mainly for selection among a sire's great-grandsons and paternal grandsons.

In practice, marker information will be obtained from very popular bulls. Figure 2, based on actual data of the Holstein sire Carlin-M-Ivanhoe Bell, shows that A.I. studs obtain offspring of the most popular bulls for a longer than expected period of time. Therefore, even if the marker information derives from a "granddaughter design", it will be used occasionally for selection among offspring, and mostly among grandoffspring and great-grandoffspring of a particular sire. Therefore, selection indices will be presented for offspring, grandoffpring and great-grandoffspring.

For each selection index, a range of possible gains in accuracy due to the inclusion of marker information is assessed. This gain is conditional on the sire being heterozygous at the marker and QTL, and known inheritance of the marker allele in the descendant to be evaluated.

M.1. First generation index

Here, candidates are offspring of a sire for whom the marker-QTL linkage was established. The situation is represented in

Figure 3. Figure 3 shows a sire **S**, heterozygous for both the marker and QTL, and its two offspring, O_1 and O_2. Sire **S** carries marker alleles M_1 and M_2 at the marker locus. Sire **S** is also heterozygous at the QTL. The two QTL alleles are denoted by Q_1 and Q_2 respectively. The QTL alleles Q_1 and Q_2 in the sire are linked with marker alleles M_1 and M_2 respectively. The average effect at the QTL in the gametes of the mates of **S** is represented by $\bar{\alpha}_d$. Alleles Q_d and M_d are the marker and the QTL allele coming from the sire's mate. Offspring O_1 received M_1 from **S**, and M_d from the dam. Offspring O_2 received M_2 and M_d. Offspring O_1 has genotype Q_1Q_d with probability (1-r), or Q_2Q_d with probability r. Offspring O_2 has genotype Q_2Q_d with probability (1-r) and genotype Q_1Q_d with probability r.

Given that O_1 received the marker allele M_1, its expected breeding value at the QTL (BV_{O1}) is:

$$BV_{01} = (1-r)\alpha_1 + r\alpha_2 + \bar{\alpha}_d \qquad [1]$$

Similarly, for offspring O_2,

$$BV_{02} = r\alpha_1 + (1-r)\alpha_2 + \bar{\alpha}_d \qquad [2]$$

The parent average breeding value at the QTL is

$$PA = 1/2(\alpha_1 + \alpha_2) + \bar{\alpha}_d \qquad [3]$$

Then, the expected Mendelian sampling effect at the QTL explained by segregation at the marker locus in the sire, for both

types of offspring, are

$$BV_{O1}-PA=1/2(1-2r)(\alpha_1-\alpha_2) \quad [4]$$

$$BV_{02}-PA=-1/2(1-2r)(\alpha_1-\alpha_2) \quad [5]$$

Denote the genetic difference among marker alleles in the sire, or the difference between **[4]** and **[5]**, by

$$D_g=(1-2r)(\alpha_1-\alpha_2) \quad [6]$$

Equation **[6]** was previously introduced by several authors as the chromosome substitution effect (e.g., Soller, 1978; Dekkers and Dentine, 1990).

Equations **[4]** and **[5]** can then be rewritten using **[6]** as

$$BV_{01}-PA=1/2D_g \quad [7]$$

$$BV_{02}-PA=-1/2D_g \quad [8]$$

Equations **[7]** and **[8]** can be expressed in the transmitting ability scale as

$$S_{g(01/02)} = (+/-)\ 1/4\ D_g \quad [9]$$

with S_g representing the Mendelian sampling effect explained by the marker-QTL linkage in the sire.

D_g and S_g pertain to the true portion of the Mendelian sampling effect explained by the marker. Denote the observed, phenotypic values of S_g and D_g by S_p and D_p, respectively, with

$$S_p = (+/-)\ 1/4\ D_p \qquad [10]$$

Then a selection index combining the animal model predicted transmitting ability (PTA) and the marker information S_p can be represented as:

$$I = b_1\ PTA\ + b_2\ S_p \qquad [11]$$

where b_1 and b_2 are the index weights. This index assumes that PTA and S_p were estimated separately, i.e., that PTA was computed with the current animal model system not including marker information. The selection index weights are:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} Var(PTA) & Cov(PTA, S_p|D_g) \\ sym & Var(S_p|D_p) \end{bmatrix}^{-1} \begin{bmatrix} Cov(PTA, TA) \\ Cov(S_p, TA|D_g) \end{bmatrix} \qquad [12]$$

where $Var(S_p/D_p)$, $Cov(PTA,S_p/D_g)$, $Cov(PTA,TA)$ and $Cov(S_p,TA/D_g)$ are, respectively, the variance of the marker effect given that D_p is fixed, the covariance between PTA and S_p given D_g, the covariance between PTA and transmitting ability (TA), and the covariance between the marker effect and TA given D_g. From **[10]**

$$Var(S_p|D_p) = .5(+.25\ D_p)^2 + .5(-.25\ D_p)^2 = 1/16\ D_p^2 \qquad [13]$$

$$Cov(S_p,TA|D_g) = Cov(S_p,S_g|D_g) = Var(S_g|D_g) \qquad [14]$$

$Var(S_p|D_p)$ and $Var(S_g|D_g)$ are the observed and true variance

at the QTL given D_p and D_g, which are usable for MAS (Dekkers and Dentine, 1990). Then,

$$\text{Var}(S_g|D_g) = 1/16\ D_g^2 \qquad [15]$$

$\text{Var}(S_g|D_g)$ is not known because D_g is not known, but only D_p, a least square estimate of D_g is observed. A potential estimator for $\text{Var}(S_g|D_g)$ based on D_p and its variance is presented in the appendix.

$\text{Cov}(\text{PTA}, S_p|D_g)$ depends on the sources of information contributing to PTA. PTA of any animal may be represented as the weighted average of three sources of information, pedigree, own and progeny records. From Van Raden and Wiggans (1991)

$$\text{PTA} = x_1\ \hat{PA} + x_2\ \text{YD}/2 + x_3\ \text{DYD} \qquad [16]$$

where x_1, x_2 and x_3 are weights summing to 1, $\hat{PA}$ is parent average, YD is yield deviation, the production of a cow corrected for environmental effects, and DYD is daughter yield deviation, which is defined as the average production of daughters adjusted for environmental effects and mate's genetic merit (Van Raden and Wiggans, 1991). For bulls, there is no YD and x_2=0.

M.1.1. Parent average as the sole source of information.

PTA of bulls prior to progeny testing contains only pedigree information. Then $x_3=x_2=0$ in **[16]** and PTA of these candidates is equal to their parent average ($\hat{PA}$). Then, equation **[12]** reduces to

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} Var(\hat{PA}) & 0 \\ 0 & Var(S_p|D_p) \end{bmatrix}^{-1} \begin{bmatrix} Cov(\hat{PA}, TA) \\ Cov(S_p, TA|D_g) \end{bmatrix} \qquad [17]$$

Note that in **[17]**, $Cov(PTA, S_p|D_g) = Cov(\hat{PA}, S_p|D_g) = 0$ due to the independence between parent average and mendelian sampling effect.

M.1.2. Daughter information available on bulls

In this case, PTA of a young candidate contains pedigree and progeny information. Then, PTA is no longer equal to parent average. From **[16]**, and because for bulls $x_2 = 0$, PTA of a young progeny tested sire is

$$PTA = x_1 \hat{PA} + x_3 DYD \qquad [18]$$

In this case, the covariance between PTA and S_p is no longer null as in **[17]**. Mendelian sampling effect is included in DYD, and

S_p estimates the fraction of the Mendelian sampling effect explained by the marker. Therefore, the covariance between PTA and S_p given D_g is

$$\begin{aligned} Cov(PTA, S_p|D_g) &= Cov(x_1 \quad \hat{PA} \quad + x_3 \ DYD, \ S_p|D_g) \\ &= x_3 \ Cov(DYD, \ S_p|D_g) \\ &= x_3 \ Cov(TA, \ S_p|D_g) \\ &= x_3 \ Var(S_g|D_g) \end{aligned} \qquad [19]$$

The required value for x_3 is not provided by the USDA evaluations. However, x_3 which weights the contribution of DYD to PTA may be approximated by the ratio between the daughter equivalent corresponding to DYD (DE_{DYD}) over the total daughter equivalent (DE_{TOT}) (Van Raden and Wiggans, 1991). DE_{DYD} and DE_{TOT} denote the contribution of progeny and all types of information to the reliability of PTA (Rel(PTA)), respectively. Then

$$x_3 = \frac{DE_{DYD}}{DE_{TOT}}$$

M.1.3. Own records available on dams

PTA of bull dams may include all the three sources of information. Mendelian sampling effect explained by the marker is included in both YD and DYD. Then,

$$\text{Cov}(\text{PTA}, S_p/D_g) = \text{Cov}(x_1 \quad \hat{PA} \quad + x_2 \text{ YD}/2 + x_3 \text{ DYD}, \; S_p|D_g)$$

$$= (x_2 + x_3) \text{ Cov}(\text{DYD}, \; S_p|D_g)$$

$$= (x_2 + x_3) \text{ Var}(S_g|D_g) \qquad \textbf{[20]}$$

Equation **[20]** differs from **[19]** only in the inclusion of x_2. An approximate value for x_2 can be obtained from the correspondant daughter equivalent (Van Raden and Wiggans, 1991).

$$x_2 = \frac{DE_{YD}}{DE_{TOT}}$$

M.1.4. Gain in accuracy (PGA) due to marker information.

From selection index theory, the adjusted squared accuracy of the selection index ($r^2_{(adj)I}$) in **[11]** is

$$r^2_{(adj)I} = \frac{b_1^2 Var_{adj}(PTA) + b_2^2 Var(S_p|D_p) + 2b_1 b_2 Cov(PTA, S_p|D_g)}{Var_{adj}(TA)} \qquad \textbf{[21]}$$

where the subscript **adj** indicates that a particular term in **[21]** was adjusted for selection. Bulmer (1971) showed that although the Mendelian sampling variance is not affected by selection for a polygenic trait, the additive genetic variance is reduced due to gametic phase disequilibrium (e.g., the Bulmer effect). The additive genetic variance will reach an equilibrium after 4 or 5 generations of a constant pattern of selection. $\text{Var}(S_p|D_p)$ and $\text{Cov}(\text{PTA}, S_p|D_g)$ in **[21]** are not adjusted because the marker effect

explains a fraction of the Mendelian variance, which is not altered by selection. Not only the genetic variance but also the accuracy of predicted breeding values are affected by selection (Bulmer, 1971)

Dekkers (1992a) presented formulae to compute accuracies and variances at the equilibrium for the different paths of selection. Following Dekkers (1992a) notation, denote by Rel_s^* and Rel_d^* the equilibrium reliabilities of the predicted transmitting abilities of the sire and the dam, respectively.

From Dekkers (1992a), these equilibrium variances are:

$$\sigma_s^{2*} = \sigma_A^2 \frac{(1-k_{ds})\,[k_{sd}(1-r_{sd}^2)+k_{dd}(1-r_{dd}^2)+2] + (3+k_{dd})\,[k_{ss}(1-r_{ss}^2)+k_{ds}(1-r_{ds}^2)+2]}{(3+k_{ss})(3+k_{dd}) - (1-k_{ds})(1-k_{ds})} \qquad [22]$$

and

$$\sigma_d^{2*} = \sigma_A^2 \frac{(1-k_{sd})\,[k_{ss}(1-r_{ss}^2)+k_{ds}(1-r_{ds}^2)+2] + (3+k_{ss})\,[k_{sd}(1-r_{sd}^2)+k_{dd}(1-r_{dd}^2)+2]}{(3+k_{ss})(3+k_{dd}) - (1-k_{ds})(1-k_{sd})} \qquad [23]$$

where the subscripts ds, dd, ss, and sd indicate the four paths of selection (i.e., dam of sire, dam of dam, sire of sire and sire of dam, respectively); $k_{xy} = i_{xy}(i_{xy}-t_{xy})$, where i_{xy} is the selection intensity in path xy, with xy = ds, dd, ss or sd, and t_{xy} is the standardized truncation point (Cochran, 1951).

Equations **[22]** and **[23]** were computed for the selection scheme shown in box.

Box. Parameters for a progeny tested population

	Path of Selection			
	ss	**sd**	**ds**	**dd**
% selected	.04	.20	.06	.90
i_{xy}	2.15	1.40	1.98	.195
t_{xy}	1.76	.84	1.56	-2.33
$r^2_{xy}=Rel_{xy}$	.72	.72	.42	.42

The proportion of animals selected and the accuracies in each path of selection are as in Lohuis et al. (1992). Selection intensities were obtained from tables in Van Vleck (1988). The standardized truncation point was obtained from tables in Steel and Torrie (1980).

Equations **[22]** and **[23]** can be used to compute the adjusted variances of PTA and TA. Then

$$Var_{adj}(PTA) = Var_{adj}(TA) \cdot Rel_{adj}(PTA)$$

where $Var_{adj}(TA)$ and $Rel_{adj}(PTA)$ are $1/4\ \sigma_i^{2*}$ and $1 - \frac{\sigma_a^2}{\sigma_i^{2*}}(1-Rel_{PTA})$ respectively, with i=(s,d) depending on whether the candidate is a young bull or a bull dam, respectively.

Equilibrium variances and accuracies were applied to estimate the gain in accuracy achieved by the selection index **[11]**. The gain in accuracy due to **[11]** is computed relative to the accuracy achieved without marker information, i.e., with respect to the

$$PGA = \frac{\sqrt{r^2_{(adj)I}} - \sqrt{Rel_{adj}(PTA)}}{\sqrt{Rel_{adj}(PTA)}} \qquad [24]$$

Where I denotes the index in **[11]**. When pedigree is the sole source of information used to estimate PTA, then **[24]** becomes

$$PGA = \sqrt{1 + \frac{b_2^2 (\frac{1}{4} D_p)^2}{\frac{\sigma_i^{2*}}{4} (1 - \frac{\sigma_a^2}{\sigma_i^{2*}} (1 - Rel(PTA)))}} - 1 \qquad [25]$$

Notice that PGA tends to 0 with little marker information (i.e., if D_p tends to 0). Also notice that less PGA is obtained with larger Rel(PTA), which in this case is also the reliability of parent average (Rel(PA)).

When information other than pedigree is used to estimate PTA, then Cov(PTA,S_p) is not null. The accuracy of the index is

$$r^2_{(adj)I} = \frac{1}{4}\sigma_i^{2*}\{b_1^2 \frac{1}{4}(\sigma_i^{2*} - \sigma_a^2(1 - Rel(PTA))) + b_2^2 (\frac{1}{4} D_p)^2 + 2 b_1 b_2 (x_2 + x_3)(\frac{1}{4} D_p)^2\} \qquad [26]$$

and **[24]** becomes

$$PGA=\sqrt{b_1^2+b_2^2\frac{(\frac{1}{4}D_p)^2}{\frac{1}{4}(\sigma_i^{2*}-\sigma_a^2(1-Rel(PTA)))}+2b_1b_2\frac{(x_2+x_3)(\frac{1}{4}D_p)^2}{\frac{1}{4}(\sigma_i^{2*}-\sigma_a^2(1-Rel(PTA)))}}$$

[27]

If $(x_2 + x_3) = 0$ (i.e., if only pedigree information was available), then equation **[27]** reduces to **[25]**.

M.2. Second generation index

Marker information is now passed from grandsire to a sire (or dam) and from the sire (or dam) onto offspring, as shown in Figure 4. The breeding value of a grandoffspring at the marked QTL (BV_{go}) may be partitioned into parent average (PA) and Mendelian sampling (BV-PA). PA may be partitioned into grandparent average (GPA) and a deviation of PA from GPA, (PA-GPA). GPA at the QTL is

$$GPA=\frac{1}{4}(\alpha_1+\alpha_2)+\frac{1}{2}\bar{\alpha}_{gd} \qquad \textbf{[28]}$$

or half the transmitting ability of the grandsire plus half the transmitting ability of the mates of the grandsire (grandams), where $\bar{\alpha}_{gd}$ is the average effect of the mates of the grandsire.

Marker information was independent of PA in the first generation. In the second generation, marker information coming from the grandsire is present in both PA and (BV-PA), because PA includes the Mendelian sampling effect that comes from the grandsire. The marker information included in PA has to be taken into account when building the second generation index.

The BV at the QTL in the grandoffspring can be partitioned as

$$BV_{go} = [BV_{go}-PA] + [PA-GPA] + GPA \quad [29]$$

Grandoffspring may be assigned to four groups according to the marker allele inherited from the grandsire or its mates, as shown in Figure 4. Equation **[29]** can be represented for each of these groups as

$$BV_{O1}-GPA = [BV_{O1}-PA_{S1}] + [PA_{S1}-GPA] \quad [30]$$

$$BV_{O2}-GPA = [BV_{O2}-PA_{S1}] + [PA_{S1}-GPA] \quad [31]$$

$$BV_{O3}-GPA = [BV_{O3}-PA_{S2}] + [PA_{S2}-GPA] \quad [32]$$

$$BV_{O4}-GPA = [BV_{O4}-PA_{S2}] + [PA_{S2}-GPA] \quad [33]$$

Figure 4 shows not only the marker genotypes of all individuals but also the possible QTL genotypes (two for the parents , three for the grandoffspring) inside the boxes and their probability of occurrence outside. Summing QTL effects for each

genotype within individual (e.g., $\bar{\alpha}_{gd} + \alpha_1$ for $Q_{gd}Q_d$), yields

$$BV_{01} = (1-r)^2\alpha_1 + r(1-r)\alpha_2 + r\bar{\alpha}_{gd} + \bar{\alpha}_d \quad [34]$$

$$BV_{02} = r(1-r)\alpha_1 + r^2\alpha_2 + (1-r)\bar{\alpha}_{gd} + \bar{\alpha}_d \quad [35]$$

$$BV_{03} = r(1-r)\alpha_1 + (1-r)^2\alpha_2 + r\bar{\alpha}_{gd} + \bar{\alpha}_d \quad [36]$$

$$BV_{04} = r^2\alpha_1 + r(1-r)\alpha_2 + (1-r)\bar{\alpha}_{gd} + \bar{\alpha}_d \quad [37]$$

$$PA_{S1} = 1/2\,[(1-r)\alpha_1 + r\alpha_2 + \bar{\alpha}_{gd}] + \bar{\alpha}_d \quad [38]$$

$$PA_{S2} = 1/2\,[r\alpha_1 + (1-r)\alpha_2 + \bar{\alpha}_{gd}] + \bar{\alpha}_d \quad [39]$$

where $\bar{\alpha}_d$ is the average effect of the QTL alleles present on the mates of S_1 and S_2 (see Figure 4); BV_{0i} (i=1,...,4) is the breeding value at the QTL in a grandoffspring from group i (Figure 4); PA_{Sj} (j=1,2) is the parent average for parent group S_j (Figure 4), and GPA is the grandparent average or the average transmitting ability of the grandsire and its mates.

The two components of equations **[30]** to **[33]**, using **[34]** to **[39]** and **[28]**, are

(i) The effect of the Mendelian sampling in the grandoffspring:

$$BV_{01}-PA_{S1}=1/2(1-2r)[(\alpha_1-\overline{\alpha}_{gd})-r(\alpha_1-\alpha_2)] \quad [40]$$

$$BV_{02}-PA_{S1}=-1/2(1-2r)[(\alpha_1-\overline{\alpha}_{gd})-r(\alpha_1-\alpha_2)] \quad [41]$$

$$BV_{03}-PA_{S2}=-1/2(1-2r)[(\alpha_2-\overline{\alpha}_{gd})+r(\alpha_1-\alpha_2)] \quad [42]$$

$$BV_{04}-PA_{S2}=1/2(1-2r)[(\alpha_2-\overline{\alpha}_{gd})+r(\alpha_1-\alpha_2)] \quad [43]$$

and

(ii) the deviation PA-GPA, which consists of Mendelian sampling in the sire due to segregation of the QTL in the grandsire plus the average transmitting ability of the mates of the sire, or

$$PA_{S1}-GPA = 1/4(1-2r)(\alpha_1-\alpha_2)+\overline{\alpha}_{gd} \quad [44]$$

$$PA_{S2}-GPA =-1/4(1-2r)(\alpha_1-\alpha_2)+\overline{\alpha}_{gd} \quad [45]$$

The Mendelian sampling effects represented in equations [30] to [33], or [40] to [43], will be denoted in the transmitting ability scale as:

$$S_{g(01)} = \frac{1}{2}(BV_{01}-PA_{S1}) \quad [46]$$

$$S_{g(02)} = \frac{1}{2}(BV_{02}-PA_{S1}) = -S_{g(01)} \quad [47]$$

$$S_{g(03)} = \frac{1}{2}(BV_{03}-PA_{S2}) \quad [48]$$

$$S_{g(04)} = \frac{1}{2}(BV_{04}-PA_{S2}) = -S_{g(03)} \quad [49]$$

$$S_{g(S1)} = \frac{1}{2}(PA_{S1}-GPA-\bar{\alpha}_{gd}) = (\frac{1}{8})D_g \quad [50]$$

$$S_{g(S2)} = \frac{1}{2}(PA_{S2}-GPA-\bar{\alpha}_{gd}) = -(1/8)D_g \quad [51]$$

Mendelian sampling effects $S_{g(01)}$ and $S_{g(02)}$ come from the half-sib family of sire S_1, and Mendelian sampling effects $S_{g(03)}$ and $S_{g(04)}$ come from the half sib family of sire S_2 (see Figure 4). Mendelian sampling variance at the QTL in each of the two families is

$$Var(S_{g(01/02)}|D_g) = S^2_{g(01)} = S^2_{g(02)}$$

and

$$Var(S_{g(03/04)}|D_g) = S^2_{g(03)} = S^2_{g(04)}$$

From equations **[46]** to **[49]** and **[40]** to **[43]**:

$$Var(S_{g(01|02)}|D_g,r) = \frac{1}{16}(1-2r)^2[(\alpha_1-\bar{\alpha}_{gd})^2+r^2(\alpha_1-\alpha_2)^2-2r(\alpha_1-\bar{\alpha}_{gd})(\alpha_1-\alpha_2)] \quad [52]$$

$$Var(S_{g(03/04)}|D_g,r) = \frac{1}{16}(1-2r)^2[(\alpha_2-\bar{\alpha}_{gd})^2+r^2(\alpha_1-\alpha_2)^2-2r(\alpha_2-\bar{\alpha}_{gd})(\alpha_1-\alpha_2)] \quad [53]$$

Mendelian sampling effects $S_{g(S1)}$ and $S_{g(S2)}$ come from the half sib family of the grandsire GS (see Figure 4). Mendelian sampling variance at the QTL in this family is:

$$Var(S_{g(S1/S2)}|D_g) = S^2_{g(S1)} = S^2_{g(S2)}$$

From **[50]** and **[51]**:

$$Var(S_{g(S1/S2)}|D_g, r) = (\frac{1}{8}(1-2r)(\alpha_1-\alpha_2))^2 = (\frac{1}{8}D_g)^2 \qquad \textbf{[54]}$$

Comparison among candidates can be made by using an index of the form:

$$I = b_1\ PTA + b_2\ S_{p(Si)} + b_3\ S_{p(Oj)} \qquad \textbf{[55]}$$

where b_1, b_2, and b_3 are index weights, $S_{p(Si)}$ with i=1,2 is the observed value of the true Mendelian effect $S_{g(Si)}$ defined in equations **[50]** and **[51]** with D_g replaced by D_p; and $S_{p(Oj)}$ with j=1,...,4 are the observed values of the Mendelian effects defined in equations **[46]** to **[49]**. As for the first generation, the selection index in **[55]** assumes that PTA was computed separately with an Animal Model, and includes either only pedigree information or also information from daughters or own production. Notice that the index for grandoffspring **[55]** differs from the index for offspring **[11]** in the inclusion of two segregation terms instead of one.

Segregation effects in the grandoffspring as defined in equations **[40]** to **[43]** depend not only on the difference between effects of the QTL alleles in the grandsire ($\alpha_1-\alpha_2$) as in the first generation, but also on the differences between average effects of QTL alleles in the dam and effects of the QTL alleles in the grandsire ($\alpha_1-\overline{\alpha}_{gd}$) or ($\alpha_2-\overline{\alpha}_{gd}$). This is a consequence of grandoffspring groups O_2 and O_3 (see Figure 4) not receiving any of

the grandsire's marker alleles.

The average value of the QTL alleles of the mates of the grandsire ($\bar{\alpha}_{gd}$) is unknown, and it is not obvious how to estimate it. One way of estimating segregation residuals **[40]** to **[43]** is to assume, as in Meuwissen and Van Arendonk (1992), that

$$\bar{\alpha}_{md} = \frac{1}{2}(\alpha_1+\alpha_2) \qquad \textbf{[56]}$$

Using **[56]** in equations **[40]** to **[43]** yields

$$BV_{01}-PA_{S1} = 1/4\ (1-2r)^2(\alpha_1-\alpha_2) = \frac{1}{4}(1-2r)D_g \qquad \textbf{[57]}$$

$$BV_{02}-PA_{S1} = -1/4\ (1-2r)^2(\alpha_1-\alpha_2) = -\frac{1}{4}(1-2r)D_g \qquad \textbf{[58]}$$

$$BV_{03}-PA_{S2} = -1/4\ (1-2r)^2(\alpha_1-\alpha_2) = -\frac{1}{4}(1-2r)D_g \qquad \textbf{[59]}$$

$$BV_{04}-PA_{S2} = 1/4\ (1-2r)^2(\alpha_1-\alpha_2) = \frac{1}{4}(1-2r)D_g \qquad \textbf{[60]}$$

Mendelian sampling variances at the QTL within the two grandoffspring families (i.e., equations **[52]** and **[53]**) expressed in a transmitting ability scale, reduce to

$$Var(S_{g(01/02)}|D_g,r) = Var(S_{(g(03/04)}|D_g,r) = \frac{1}{16}(1-2r)^2\frac{1}{4}(1-2r)^2(\alpha_1-\alpha_2)^2$$
$$= \frac{1}{64}(1-2r)^4(\alpha_1-\alpha_2)^2$$

$$= \frac{1}{64}(1-2r)^2D_g^2 \qquad \textbf{[61]}$$

If **[56]** holds, the weights b_1, b_2 and b_3 in **[55]** can be obtained by solving

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} Var(PTA) & Cov(PTA, S_{p(Si)}|D_g) & Cov(PTA, S_{p(Oj)}|D_g, r) \\ & Var(S_{p(Si)}|D_p) & 0 \\ sym. & & Var(S_{p(Oj)}|D_p, r) \end{bmatrix}^{-1} \begin{bmatrix} Cov(PTA, TA) \\ Cov(S_{p(Si)}, TA|D_g) \\ Cov(S_{p(Oj)}, TA|D_g, r) \end{bmatrix} \quad [62]$$

where

$$Var(S_{p(Si)}|D_p) = (1/8\ D_p)^2 \quad [63]$$

$$Var(S_{p(Oj)}|D_p, r) = (1/8\ (1-2r)\ D_p)^2 \quad [64]$$

$$Cov(S_{p(Si)}, TA|D_g) = Var(S_{g(Si)}|D_g) = (1/8\ D_g)^2 \quad [65]$$

$$Cov(S_{p(Oj)}, TA|D_g, r) = Var(S_{g(Oj)}|D_g, r) = (1/8\ (1-2r)\ D_g)^2 \quad [66]$$

The PTA covaries with $S_{p(Si)}$ only through parent average, because Mendelian sampling effects are independent. Then

$$Cov(PTA, S_{p(Si)}|D_g) = Cov(\ \hat{PA}_o\ , S_{p(Si)}|D_g)$$

The subscript "o" is added to $\hat{PA}$ to denote that $S_{p(Si)}$ varies with the second generation's parent average (i.e., parent average of animals O_i in Figure 4), but not with the first generation's parent average (i.e., parent average of S_1 and S_2 in Figure 4).

$Cov(\ \hat{PA}_o\ , S_{p(Si)}/D_g)$ can be derived by replacing $\hat{PA}_o$ with half the sum of breeding value estimates of the parents (i.e., 1/2 (BV_s + BV_d)) by realizing that only BV_s covaries with $S_{p(Si)}$. Then, from the USDA Animal Model evaluations (Van Raden and Wiggans, 1991),

$$PTA = x_1\ \hat{PA}_o + x_2\ YD/2 + x_3\ DYD \quad [67]$$

where x_1, x_2 and x_3 are weights that sum to 1. Then,

$$\begin{aligned} \mathrm{Cov}(PTA, S_{p(Si)} \mid D_g) &= \mathrm{Cov}(x_2\ YD/2 + x_3\ DYD,\ S_{p(Si)} \mid D_g) \\ &= (x_2 + x_3)\ \mathrm{Var}(S_{g(Si)} \mid D_g) \\ &= (x_2 + x_3)\ (1/8\ D_g)^2 \end{aligned} \qquad \textbf{[68]}$$

The PTA covaries with $S_{p(Oj)}$ only when own records from animals O_j or their daughters are available. If pedigree is the only information available for animals O_j, then

$$\mathrm{Cov}(PTA, S_{p(Oj)} \mid D_g) = 0 \qquad \textbf{[69]}$$

When a source of information other than pedigree is available, then

$$\begin{aligned} \mathrm{Cov}(PTA, S_{p(Oj)} \mid D_g, r) &= \mathrm{Cov}(x_2\ YD/2 + x_3\ DYD,\ S_{p(Oj)} \mid D_g, r) \\ &= (x_2 + x_3)\ \mathrm{Var}(S_{g(Oj)} \mid D_g, r) \\ &= (x_2 + x_3)\ (1/8\ (1-2r)\ D_g)^2 \end{aligned} \qquad \textbf{[70]}$$

When parent average is the only source of information then **[70]** reduces to **[69]**.

The system of equations **[62]** assumes that **[56]** holds. If there are only two alleles at the QTL or one of the alleles in the grandsire (e.g., allele 1) is considerably superior to any other allele in the population then the transmitting ability of granddam

may be defined as

$$\overline{\alpha}_{gd} = p\ \alpha_1 + (1-p)\ \alpha_2 \qquad [71]$$

where p is the frequency of QTL allele 1 in the granddam population. An estimate of p may be obtained by comparing the average transmitting ability at the QTL of all grandsire's mates (BV_{gd}^{QTL}) with the sire's transmitting ability at the QTL (BV_{gs}^{QTL}). BV_{gs}^{QTL} is equal to $.5(\alpha_1 + \alpha_2)$, and BV_{gd}^{QTL} is

$$\begin{aligned} BV_{gd}^{QTL} &= p\ \alpha_1 + (1-p)\alpha_2 \\ &= (p-1)(\alpha_1-\alpha_2)+\alpha_1 \end{aligned}$$

Then,

$$\begin{aligned} BV_{gs}^{QTL} - BV_{gd}^{QTL} &= .5(\alpha_1-\alpha_2) + \alpha_2 - (p-1)(\alpha_1-\alpha_2) - \alpha_1 \\ &= (.5-p)(\alpha_1-\alpha_2) \end{aligned}$$

Hence, p can be estimated from:

$$\hat{p} = .5 + \frac{BV_{gd}^{QTL} - BV_{gs}^{QTL}}{(\alpha_1-\alpha_2)} \qquad [72]$$

Estimates of BV_{gd}^{QTL} and BV_{gs}^{QTL} required in **[72]** can be obtained as the conditional expectation of BV^{QTL}, given total BV, or

$$\begin{aligned} BV_{gd}^{QTL} &= (\sigma^2_{QTL}/\sigma^2_a)\, BV_{gd} \\ BV_{gs}^{QTL} &= (\sigma^2_{QTL}/\sigma^2_a)\, BV_{gs} \end{aligned}$$

where σ^2_{QTL} is the additive genetic variance at the QTL. Estimates

of σ^2_{QTL} and $(\alpha_1-\alpha_2)$ can be obtained from D_g^2 and D_p, respectively, if r is known.

Then, the segregation residuals in **[40]** to **[43]** are

$$\frac{1}{2}(BV_{01}-PA_{S1}) = 1/4(1-2r)(1-p-r)(\alpha_1-\alpha_2) = (1-p-r)S_{g(01)} \quad \textbf{[73]}$$

$$\frac{1}{2}(BV_{02}-PA_{S1}) = -1/4(1-2r)(1-p-r)(\alpha_1-\alpha_2) = -(1-p-r)S_g \quad \textbf{[74]}$$

$$\frac{1}{2}(BV_{03}-PA_{S2}) = -1/4(1-2r)(p-r)(\alpha_1-\alpha_2) = -(p-r)S_g \quad \textbf{[75]}$$

$$\frac{1}{2}(BV_{04}-PA_{S2}) = 1/4(1-2r)(p-r)(\alpha_1-\alpha_2) = (p-r)S_g \quad \textbf{[76]}$$

Equations **[73]** to **[76]** can be used in **[62]** to find the index weights under assumption **[71]**. Under **[71]** the variances of the segregation effects for the two different offspring families, expressed in a transmitting ability scale, are

$$\begin{aligned} Var(S_{g(01/02)}|D_g, r) &= [1/4(1-2r)(1-p-r)(\alpha_1-\alpha_2)]^2 \\ &= [1/4(1-p-r)D_g]^2 \end{aligned} \quad \textbf{[77]}$$

$$\begin{aligned} Var(S_{g(03/04)}|D_g, r) &= [1/4(1-2r)(p-r)(\alpha_1-\alpha_2)]^2 \\ &= [1/4(p-r)D_g]^2 \end{aligned} \quad \textbf{[78]}$$

Note that **[77]** and **[78]** are generally not equal and become equal only if p=.5. Notice that if p=r, $Var(S_{g(03/04)})=0$. If p=r, the segregation effects of animals **03** and **04** are equal on average.

Observed segregation effects and variances can be defined as in **[77]** and **[78]**, but with subscript **g** replaced by **p**. These segregation effects require knowledge about **r** and **p**. Two different selection indices are needed for the two types of half-sib families of the grandoffspring (e.g., families of S_1 and S_2 in Figure 4), because of the difference in Mendelian sampling variance given in equations **[77]** and **[78]**.

M.3. Third generation index.

Figure 5 displays the eight different sets of marker genotypes (e.g., **O1-O8**) and associated probabilities of occurrence among great-grandoffspring. Figure 5 can be viewed as a continuation of Figure 4, displaying one additional generation. Figure 5 also indicates from which marker genotype these eights great-grandoffspring genotypes descended. For instance, genotypes **O1-O2** in Figure 5 descend from **S1** in Figure 4. Genotype **S1** in Figure 5 is identical to **O1** in Figure 4.

The offspring in Figure 4 (genotypes **O1-O4**) are the parents in Figure 5 (**S1-S4**). Similarly, dams and grandams in Figure 4 are grandams and great-grandams in Figure 5. Consequently, $\overline{\alpha}_d$ and $\overline{\alpha}_{gd}$ in Figure 4 are denoted as $\overline{\alpha}_{gd}$ and $\overline{\alpha}_{ggd}$ in Figure 5. The

term $\overline{\alpha}_d$ in Figure 5 denotes the average QTL effect in the dams of genotypes **O1-O8**.

Comparisons among great-grandoffspring (**O1-O8** in Figure 5) can be obtained as for the second generation, by partitioning the marker information into its components available at each generation. Marker information is now passed from great-grandsire onto grandsire (or grandam), onto sire (or dam), and onto offspring or great-grandoffspring. These three generations are represented in the selection index, or

$$I = b_1 PTA + b_2 S_{p(GS)} + b_3 S_{p(S)} + b_4 S_{p(O)} \qquad [79]$$

where S_p denotes the estimated Mendelian segregation effect at each generation (i.e., from greatgrandsire to grandsire (GS), from grandsire to sire (S) and from sire to offspring (O)).

Then, the breeding value of the great-grandoffspring at the QTL can be partitioned as:

$$BV_{01}-GGPA=[BV_{01}-PA_{S1}]+[PA_{S1}-.5PA_{GS1}]+[.5PA_{GS1}-GGPA] \quad [80]$$

$$BV_{02}-GGPA=[BV_{02}-PA_{S1}]+[PA_{S1}-.5PA_{GS1}]+[.5PA_{GS1}-GGPA] \quad [81]$$

$$BV_{03}-GGPA=[BV_{03}-PA_{S2}]+[PA_{S2}-.5PA_{GS1}]+[.5PA_{GS1}-GGPA] \quad [82]$$

$$BV_{04}-GGPA=[BV_{04}-PA_{S2}]+[PA_{S2}-.5PA_{GS1}]+[.5PA_{GS1}-GGPA] \quad [83]$$

$$BV_{05}-GGPA=[BV_{05}-PA_{S3}]+[PA_{S3}-.5PA_{GS2}]+[.5PA_{GS2}-GGPA] \quad [84]$$

$$BV_{06}-GGPA=[BV_{06}-PA_{S3}]+[PA_{S3}-.5PA_{GS2}]+[.5PA_{GS2}-GGPA] \quad [85]$$

$$BV_{07}-GGPA=[BV_{07}-PA_{S4}]+[PA_{S4}-.5PA_{GS2}]+[.5PA_{GS2}-GGPA] \quad [86]$$

$$BV_{08}-GGPA=[BV_{08}-PA_{S4}]+[PA_{S4}-.5PA_{GS2}]+[.5PA_{GS2}-GGPA] \quad [87]$$

where GGPA is the great-grand parent average or

$$GGPA=\frac{1}{8}(\alpha_1+\alpha_2)+\frac{1}{4}\overline{\alpha}_{ggd}$$

with $\overline{\alpha}_{ggd}$ as the average effect of the mates of the great-grandsire.

From Figure 5, the breeding values and parent averages of the great-grandoffspring at the QTL required in equations **[80]** - **[87]** are

$$BV_{01}=(1-r)^3\alpha_1+(1-r)^2r\alpha_2+(1-r)r\overline{\alpha}_{mgd}+r\overline{\alpha}_{md}+\overline{\alpha}_d \quad [88]$$

$$BV_{02}=(1-r)^2r\alpha_1+(1-r)r^2\alpha_2+r^2\overline{\alpha}_{mgd}+(1-r)\overline{\alpha}_{md}+\overline{\alpha}_d \quad [89]$$

$$BV_{03}=(1-r)^2r\alpha_1+(1-r)r^2\alpha_2+(1-r)^2\overline{\alpha}_{mgd}+r\overline{\alpha}_{md}+\overline{\alpha}_d \quad [90]$$

$$BV_{04} = (1-r)\,r^2\alpha_1 + r^3\alpha_2 + (1-r)\,r\bar{\alpha}_{mgd} + (1-r)\,\bar{\alpha}_{md} + \bar{\alpha}_d \quad [91]$$

$$BV_{05} = (1-r)^2 r\alpha_1 + (1-r)^3\alpha_2 + (1-r)\,r\bar{\alpha}_{mgd} + r\bar{\alpha}_{md} + \bar{\alpha}_d \quad [92]$$

$$BV_{06} = (1-r)\,r^2\alpha_1 + (1-r)^2 r\alpha_2 + r^2\bar{\alpha}_{mgd} + (1-r)\,\bar{\alpha}_{md} + \bar{\alpha}_d \quad [93]$$

$$BV_{07} = (1-r)\,r^2\alpha_1 + (1-r)^2 r\alpha_2 + (1-r)^2\bar{\alpha}_{mgd} + r\bar{\alpha}_{md} + \bar{\alpha}_d \quad [94]$$

$$BV_{08} = r^3\alpha_1 + (1-r)\,r^2\alpha_2 + (1-r)\,r\bar{\alpha}_{mgd} + (1-r)\,\bar{\alpha}_{md} + \bar{\alpha}_d \quad [95]$$

Expressions for PA_{GS1} and PA_{GS2} are similar to the second generation case (i.e., similar to the right hand side of equations **[44]** and **[45]** with $\bar{\alpha}_{gd}$ replaced by $\bar{\alpha}_{ggd}$).

Using equations **[88]** to **[95]** in **[80]** to **[87]** yields the three Mendelian sampling components, which are

(i) the Mendelian effect received by a great-grandoffspring from its parents:

$$\begin{aligned} BV_{01} - PA_{S1} &= (1-r)^3\alpha_1 + (1-r)^2 r\alpha_2 + (1-r)\,r\bar{\alpha}_{ggd} + r\bar{\alpha}_{gd} + \bar{\alpha}_d - \\ &\quad - \tfrac{1}{2}(1-r)^2\alpha_1 - \tfrac{1}{2}(1-r)\,r\alpha_2 - \tfrac{1}{2}r\bar{\alpha}_{ggd} - \tfrac{1}{2}\bar{\alpha}_{gd} - \bar{\alpha}_d \\ &= 1/2\,(1-2r)\,[(1-r)^2\alpha_1 + r(1-r)\,\alpha_2 + r\bar{\alpha}_{ggd} - \bar{\alpha}_{gd}] \end{aligned} \quad [96]$$

$$BV_{02} - PA_{S1} = -1/2\,(1-2r)\,[(1-r)^2\alpha_1 + r(1-r)\,\alpha_2 + r\bar{\alpha}_{ggd} - \bar{\alpha}_{gd}] \quad [97]$$

$$BV_{03} - PA_{S2} = 1/2\,(1-2r)\,[r(1-r)\,\alpha_1 + r^2\alpha_2 + (1-r)\,\bar{\alpha}_{ggd} - \bar{\alpha}_{gd}] \quad [98]$$

$$BV_{04} - PA_{S2} = -1/2\,(1-2r)\,[r(1-r)\,\alpha_1 + r^2\alpha_2 + (1-r)\,\bar{\alpha}_{ggd} - \bar{\alpha}_{gd}] \quad [99]$$

$$BV_{05} - PA_{S3} = 1/2\,(1-2r)\,[r(1-r)\,\alpha_1 + (1-r)^2\alpha_2 + r\bar{\alpha}_{ggd} - \bar{\alpha}_{gd}] \quad [100]$$

$$BV_{06}-PA_{S3}=-1/2(1-2r)[r(1-r)^2\alpha_1 + (1-r)^2\alpha_2 + r\overline{\alpha}_{ggd}-\overline{\alpha}_{gd}] \quad [101]$$

$$BV_{07}-PA_{S4}=1/2(1-2r)[r^2\alpha_1 + r(1-r)\alpha_2 + (1-r)\overline{\alpha}_{ggd}-\overline{\alpha}_{gd}] \quad [102]$$

$$BV_{08}-PA_{S4}=-1/2(1-2r)[r^2\alpha_1 + r(1-r)\alpha_2 + (1-r)\overline{\alpha}_{ggd}-\overline{\alpha}_{gd}] \quad [103]$$

(ii) the Mendelian effect passed to the sires from the grandsires

$$\begin{aligned} PA_{S1}-.5PA_{GS1} &= \frac{1}{2}(1-r)^2\alpha_1+\frac{1}{2}r(1-r)\alpha_2+\frac{1}{2}r\overline{\alpha}_{ggd}+\frac{1}{2}\overline{\alpha}_{gd}+\overline{\alpha}_d \\ &\quad -1/4(1-r)\alpha_1-1/4r\alpha_2-1/4\overline{\alpha}_{ggd}-1/2\overline{\alpha}_{gd} \\ &= 1/4(1-2r)[(1-r)\alpha_1+r\alpha_2-\overline{\alpha}_{ggd}]+\overline{\alpha}_d \end{aligned} \quad [104]$$

$$PA_{S2}-.5PA_{GS1}=-1/4(1-2r)[(1-r)\alpha_1+r\alpha_2-\overline{\alpha}_{ggd}]+\overline{\alpha}_d \quad [105]$$

$$PA_{S3}-.5PA_{GS2}= 1/4(1-2r)[r\alpha_1+(1-r)\alpha_2-\overline{\alpha}_{ggd}]+\overline{\alpha}_d \quad [106]$$

$$PA_{S4}-.5PA_{GS2}=-1/4(1-2r)[r\alpha_1+(1-r)\alpha_2-\overline{\alpha}_{ggd}]+\overline{\alpha}_d \quad [107]$$

and

iii) the Mendelian effect passed to the grandsires from the great-grandsire

$$\begin{aligned} .5PA_{GS1}-GGPA &= 1/4(1-r)\alpha_1+1/4r\alpha_2+1/4\overline{\alpha}_{ggd}+1/2\overline{\alpha}_{gd} \\ &\quad - 1/8(\alpha_1+\alpha_2)-1/4\overline{\alpha}_{ggd} \\ &= 1/8(1-2r)(\alpha_1-\alpha_2)+1/2\alpha_{gd} \end{aligned} \quad [108]$$

$$.5PA_{GS2}-GGPA=-1/8(1-2r)(\alpha_1-\alpha_2)+1/2\overline{\alpha}_{gd} \quad [109]$$

Similar to the second generation, the segregation terms for the great-grandoffspring are denoted in the transmitting ability

scale by

$$S_{g(Ok)} = \frac{1}{2}(BV_k - PA_{Sj}) \quad [110]$$

$$S_{g(Sj)} = \frac{1}{2}(PA_{Sj} - .5PA_{GSi} - \bar{\alpha}_d) \quad [111]$$

$$S_{g(GSi)} = \frac{1}{2}(.5PA_{GSi} - GGPA - 1/2\bar{\alpha}_{gd}) = \frac{1}{16}D_g \quad [112]$$

for i=1,2 , j=1,...,4 and k=1,...,8.

Variances explained by these segregation terms are

$$Var(S_{g(GSi)} | D_g) = (1/16\ D_g)^2 \quad [113]$$

$$Var(S_{g(S1)} | D_g, r) = [1/8(1-2r)[(1-r)\alpha_1 + r\alpha_2 - \bar{\alpha}_{ggd}]]^2$$
$$Var(S_{g(S2)} | D_g, r) = [1/8(1-2r)[(1-r)\alpha_2 + r\alpha_1 - \bar{\alpha}_{ggd}]]^2 \quad [114]$$

Equation **[113]** and **[114]** are equal to equation **[54]**, **[52]** and **[53]** times $(.5)^2$ and with $\bar{\alpha}_{gd}$ replaced by $\bar{\alpha}_{ggd}$. The term .5 reflects the additional generation considered.

$Var(S_{g(Ok)} | D_g)$ differs again among the different types of half-sibs families

$$Var(S_{g(O1/O2)} | D_g, r) = [1/4(1-2r)[(1-r)^2\alpha_1 + r(1-r)\alpha_2 + r\bar{\alpha}_{ggd} - \bar{\alpha}_{gd}]]^2 \quad [115]$$

$$Var(S_{g(O3/O4)} | D_g, r) = [1/4(1-2r)[r(1-r)\alpha_1 + r^2\alpha_2 + (1-r)\bar{\alpha}_{ggd} - \bar{\alpha}_{gd}]]^2 \quad [116]$$

$$Var(S_{g(O5/O6)} | D_g, r) = [1/4(1-2r)[r(1-r)\alpha_1 + (1-r)^2\alpha_2 + r\bar{\alpha}_{ggd} - \bar{\alpha}_{gd}]]^2 \quad [117]$$

$$Var(S_{g(O7/O8)}|D_g, r) = [1/4(1-2r)[r^2\alpha_1 + r(1-r)\alpha_2 + (1-r)\overline{\alpha}_{ggd} - \overline{\alpha}_{gd}]]^2 \quad [118]$$

Variances in **[114]** to **[118]** and segregation terms in **[110]** and **[111]** require knowledge about $\overline{\alpha}_{ggd}$ and $\overline{\alpha}_{gd}$. As in the second generation, one may assume that

$$\overline{\alpha}_{ggd} = \frac{1}{2}(\alpha_1 + \alpha_2)$$
$$\overline{\alpha}_{gd} = \frac{1}{2}(\alpha_1 + \alpha_2) \quad [119]$$

Under **[119]** the segregation terms in **[110]** and **[111]** and the variances in **[115]** to **[118]** reduce to

$$S_{g(O1)} = S_{g(O6)} = 1/8(1-2r)(1-r)D_g \quad [120]$$

$$S_{g(O2)} = S_{g(O5)} = -1/8(1-2r)(1-r)D_g \quad [121]$$

$$S_{g(O3)} = S_{g(O8)} = 1/8(1-2r)rD_g \quad [122]$$

$$S_{g(O4)} = S_{g(O7)} = -1/8(1-2r)rD_g \quad [123]$$

$$S_{g(S1)} = \frac{1}{16}(1-2r)D_g \quad [124]$$

$$S_{g(S2)} = -\frac{1}{16}(1-2r)D_g \quad [125]$$

$$Var(S_{g(Si)}|D_g, r) = (1/16(1-2r)D_g)^2 \quad [126]$$

$$Var(S_{p(O1/O2)}|D_g, r) = Var(S_{(O5/O6)}|D_g, r) = (\frac{1}{8}(1-2r)(1-r)D_g)^2 \quad [127]$$

$$Var(S_{g(O3/O4)}|D_g, r) = Var(S_{g(O7/O8)}|D_g) = \left(\frac{1}{8}(1-2r)\,rD_g\right)^2 \quad \textbf{[128]}$$

Notice that there are two different Mendelian sampling variances among the great-grandoffspring (i.e. among O1, O2, O5, O6 and O3, O4, O7, O8 in Figure 5) under **[119]**.

Equations **[120]** to **[125]** are the true segregation effects. The observed segregation effects needed in **[79]** are denoted as in the previous generations by replacing the suffix g by p in equations **[120]** to **[125]**.

M.4. Polymorphism information content (PIC)

The expected gain in accuracy that can be obtained from the selection indices developed in the previous sections was computed by assuming perfectly traceable inheritance of marker alleles. A marker genotype whose inheritance is not traceable is non-informative for selection purposes (e.g., Dekkers and Dentine, 1991). Traceability of the marker inheritance increases with the number of alleles (i.e., with the degree of polymorphism) of the marker loci. The fraction of marker genotypes in a population with traceable inheritance is called the polymorphism information content (PIC) of any particular marker locus (Botstein et al., 1980). The expected gain in accuracy due to the inclusion of marker information in a selection index is

reduced by the factor (1-PIC) if not all the genotypes are informative (i.e., if PIC is less than one).

Some homozygous marker genotypes that are informative for some particular generation may not be informative for the following generation. Denote **GC** and **GN** as the marker genotypes that are informative for the current and next generation, respectively. All heterozygous genotypes that are **GC** are also **GN**, but homozygous **GC** are not **GN**. Figure 6 illustrates this for a 5-allele marker locus.

Denote as $\mathbf{GN_s}^1$, $\mathbf{GN_o}^1$ and $\mathbf{GN_{go}}^1$ any **GN** (i.e., heterozygous) marker genotypes present among the sires (i.e., animals from where the marker information is obtained), offspring and grandoffspring of the sire, when mates are genotyped (1=G) or not (1=NG). Denote as $p(GN_s)$, $p(GN_o^1|GN_s)$ and $p(GN_{go}^1|GN_o)$ the probabilities of occurrence of genotypes $\mathbf{GN_s}$, $\mathbf{GN_o}$ given that its parent was $\mathbf{GN_s}$ (i.e., conditional on $\mathbf{GN_s}$), and $\mathbf{GN_{go}}$ given that its parent was $\mathbf{GN_o}$ (i.e., conditional on $\mathbf{GN_o}$). Denote $p(GC^1|GN_i^1)$ as the probability to have descendants with genotype **GC** from an animal with genotype $\mathbf{GN_i}$ (i.e., conditional on $\mathbf{GN_i}$), given mates genotyped (1=G) or not (1=NG).

Let PIC_v^1 denote the PIC at generation v=1,2,3 (i.e., offspring, grandoffspring and great-grandoffspring of an animal with genotype $\mathbf{GN_s}$) when mates are genotyped (1=G) or not (1=NG).

Then,

$$PIC_1^1 = \sum_{GN_s} p(GN_s)\, p(GC^1|GN_s) \qquad [129]$$

$$PIC_2^1 = \sum_{GN_s} \sum_{GN_o} p(GN_s)\;\; p(GN_o^1|GN_s)\;\; p(GC^1|GN_o^1) \qquad [130]$$

$$PIC_3^1 = \sum_{GN_s} \sum_{GN_o} \sum_{GN_{go}} p(GN_s)\, p(GN_o^1|GN_s)\, p(GN_{go}^1|GN_o^1)\, p(GC^1|GN_{go}^1) \qquad [131]$$

Equation **[129]** computes PIC at the first generation by multiplying the probability of a heterozygous sire in the population ($\mathbf{GN_s}$) times the probability of having descendants of these sires that have informative genotypes (**GC**), when mates are genotyped (1=G) or not (1=NG). Equations **[130]** and **[131]** compute PIC at the second and third generation. These two equations take the probability of **GN** genotypes at intermediate generations into account.

M.4.1. First generation PIC

The probability of finding a heterozygous sire ($p(GN_s)$) in equation **[129]** can be expressed as

$$p(GN_s) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i p_j \qquad [132]$$

where **i** and **j** denote two marker alleles, **n** is the total number of

marker alleles segregating in the population and p_i and p_j are the frequencies of alleles **i** and **j** in the population.

When mates are genotyped, the second term in equation **[129]**, $p(GC^1|GN_s)$ is obtained by considering that all genotypes of the offspring are informative for the current generation except for the case when the genotypes of sire, mate, and offspring are identical. Then

$$p(GC^G|GN_s) = 1 - 2p_ip_j\ \frac{1}{2}$$
$$= 1 - p_ip_j \qquad \textbf{[133]}$$

When mates are not genotyped, all offspring with the same genotype as the sire are non-informative for the current generation. Figure 6 represents these non-informative genotypes as "?". If the marker genotype of the sire is M_iM_j, an M_iM_j offspring (i.e., an offspring non-informative for the current generation) can come from mates that bear either allele **i** or **j**. Then

$$p(GC^{NG}|GN_s) = 1-\frac{1}{2}\sum_h (p_ip_h+p_jp_h)$$
$$= 1-\frac{1}{2}(p_i+p_j) \qquad \textbf{[134]}$$

where subscripts **i** and **j** denote the two marker alleles present in the sire, and **h** stands for any marker allele.

Then, from equations **[129]**,**[132]**,**[133]** and **[134]**.

$$PIC_1^G = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i p_j (1-p_i p_j) \qquad \textbf{[135]}$$

and

$$PIC_1^{NG} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i p_j (1-\frac{1}{2}(p_i+p_j)) \qquad \textbf{[136]}$$

Equations **[135]** and **[136]** can be rewritten as

$$PIC_1^G = 1 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2 - \sum_{i=1}^{n} p_i^2 \qquad \textbf{[137]}$$

$$PIC_1^{NG} = 1 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i p_j (p_i+p_j) - \sum_{i}^{n} p_i^2 \qquad \textbf{[138]}$$

Equations **[137]** and **[138]** were introduced by Dekkers and Dentine (1991) and are equivalent to formulae derived by Botstein et al. (1980).

It may be of interest to compute the conditional information content of a marker within a family, given that the sire is heterozygous M_iM_j. Denoted this PIC by $(PIC_1^I|GN_s)$. Then, equations **[135]** and **[136]** become

$$(PIC_1^G|GN_s) = 1 - p_i p_j \qquad \textbf{[139]}$$

$$(PIC_1^{NG}|GN_s) = 1 - \frac{1}{2}(p_i+p_j) \qquad [140]$$

When the marker is dimorphic, equations **[135]** and **[136]** reduce to

$$PIC_1^{G} = 2p_1p_2(1-p_1p_2) \qquad [141]$$

$$PIC_1^{NG} = p_1p_2 \qquad [142]$$

as was shown by Dekkers and Dentine (1991).

M.4.2. Second generation PIC

PIC in the second generation, or PIC_2, can be computed from **[130]**.

Computation of $p(GN_o^{NG}|GN_o)$ (mates not genotyped).

All GN_o^{NG} genotypes are heterozygous of the form M_iM_k or M_jM_k (see Figure 7). The probability of occurrence of a particular GN_o genotype of the form M_iM_k or M_jM_k, when mates are not genotyped, given a particular GN_s genotype M_iM_j is

$$p(GN_o^{NG}=M_iM_k|GN_s=M_iM_j) = p(GN_o^{NG}=M_jM_k|GN_s=M_iM_j) = \frac{1}{2}p_k \qquad [143]$$

Computation of $p(GN_o^G|GN_o)$ (mates genotyped).

The probability of a GN_o of the form M_iM_k or M_jM_k is identical to the case when mates are not genotyped (i.e., equation **[143]**). However, when mates are genotyped, GN_o includes also the genotype of the sire ($GN_s = M_iM_j$). Figure 7 illustrates this.

The probability of occurrence of a $GN_o = M_iM_j$ genotype is not that in **[143]** because $GN_o = M_iM_j$ genotypes cannot come from a mate with genotype identical to that of the sire (M_iM_j). The probability of finding a GN_o genotype M_iM_j given that $GN_s=M_iM_j$ is

$$p(GN_o^G=M_iM_j|GN_s^G=M_iM_j)=\frac{1}{2}[p_i(1-p_j)+p_j(1-p_i)] \qquad \textbf{[144]}$$

Figure 8 illustrates the computation of **[144]**.

Computation of $p(GC^1|GN_o^1)$.

The probability of finding any GC genotype given a particular parental genotype is identical for all GN_i (i.e., last terms in **[129]**, **[130]** and **[131]**). Therefore, equations **[133]** and **[134]** hold not only for the first but for the subsequent generations.

When mates are not genotyped, PIC_2^{NG} can be obtained from equations **[130]**, **[132]** and **[143]**

$$PIC_2^{NG}=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_ip_j\Big[\sum_{k\neq i,j}\frac{1}{2}p_kp(GC^{NG}|GN_o^{NG}=M_iM_k, GN_o^{NG}=M_jM_k)\Big] \quad \textbf{[145]}$$

and, from **[140]** and **[145]**

$$\begin{aligned} PIC_2^{NG} &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_ip_j\Big[\sum_{k\neq i,j}\frac{1}{2}p_k\Big(1-\frac{1}{2}(p_i+p_k)+1-\frac{1}{2}(p_j+p_k)\Big)\Big] \\ &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_ip_j\Big[\sum_{k\neq i,j}\frac{1}{2}p_k\Big(2-\frac{1}{2}(p_i+p_j)-p_k\Big)\Big] \end{aligned} \quad \textbf{[146]}$$

When mates are genotyped, the GC genotypes include all genotypes considered in **[146]** plus the GN_o^G genotype (M_iM_j). Then, from **[144]** and **[146]**,

$$\begin{aligned} PIC_2^{G}=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_ip_j\Big\{\sum_{k\neq i,j}\Big[&\frac{1}{2}p_k(2-p_k(p_i+p_j)\Big]+ \\ &+\frac{1}{2}\big[p_i(1-p_j)+p_j(1-p_i)\big](1-p_ip_j)\Big] \end{aligned} \quad \textbf{[147]}$$

Notice that for dimorphic markers, $p_k=0$, and therefore equation **[146]** yields $PIC_2^{NG}=0$. Hence, when mates are not genotyped, dimorphic markers are not informative if applied to any generation beyond the offspring. If mates are genotyped and the marker is dimorphic, then $p_k=0$ and only the last term in the inner summation in **[147]** contributes to PIC.

PIC_2^G in **[147]** differs from what Kashi et al. (1990a) obtained because they incorrectly assumed that every genotype that is informative for the current generation is also informative for the next.

M.4.3. Third generation PIC

PIC among the great-grandoffspring of a sire can be computed from equation **[131]**. Three of the four components of equation **[131]** have been previously introduced: $p(GN_s)$, $p(GN_o^1|GN_s)$ and $p(GC^1|GN_o^1)$.

Computation of $p(GN_{go}^{NG}|GN_o^{GN})$ (mates not genotyped).

Each GN_{go}^{NG} genotype that derived from a particular GN_o^{NG} genotype has one, but not two of the alleles present at the GN_o^{NG}. As in **[143]**, the probability of a particular GN_{go}^{NG} is equal to half the probability of the allele not shared by GN_{go}^{NG} and GN_{go}^{NG}. For instance,

$$p(GN_{go}^{NG}=M_iM_l|GN_o^{NG}=M_iM_k) = 1/2\ p_l$$
$$p(GN_{go}^{NG}=M_kM_l|GN_o^{NG}=M_iM_k) = 1/2\ p_l$$
$$p(GN_{go}^{NG}=M_jM_{l'}|GN_o^{NG}=M_jM_k) = 1/2\ p_{l'}$$
$$p(GN_{go}^{NG}=M_kM_{l'}|GN_o^{NG}=M_jM_k) = 1/2\ p_{l'}$$

where **l** is any allele other than **i** and **k** , and **l′** is any allele

other than **j** and k. Figure 9 illustrates the different paths of inheritance of the marker alleles.

Then, from **[131]**, **[132]**, **[143]**

$$PIC_3^{NG}=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_ip_j\Big[\sum_{k\neq i,j}\frac{1}{2}p_k\ p(GN_{go}^{NG}|GN_o^{NG})\ \ p(GC^{NG}|GN_{go}^{NG})\Big] \quad \textbf{[148]}$$

and, from **[144]** and **[148]**

$$\begin{aligned} PIC_3^{NG} &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_ip_j\Big[\sum_{k\neq i,j}\frac{1}{2}p_k\ \Big\{\sum_{l}\frac{1}{2}p_l\Big[1-\frac{1}{2}(p_i+p_l)+1-\frac{1}{2}(p_k+p_l)\Big] \\ &\qquad +\sum_{l'}\frac{1}{2}p_{l'}\Big[1-\frac{1}{2}(p_j+p_{l'})+1-\frac{1}{2}(p_k+p_{l'})\Big]\Big\}\Big] \\ &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_ip_j\Big[\sum_{k\neq i,j}\frac{1}{2}p_k\ \Big\{\sum_{l}\frac{1}{2}p_l\Big[2-\frac{1}{2}(p_i+p_k)-p_l\Big]+ \\ &\qquad +\sum_{l'}\frac{1}{2}p_{l'}\Big[2-\frac{1}{2}(p_j+p_k)-p_{l'}\Big]\Big\} \end{aligned} \quad \textbf{[149]}$$

Computation of $p(GN_{go}^{G}|GN_o^{G})$ (mates genotyped).

As in the second generation, GN_o^G includes the GN_s genotype (i.e., M_iM_j) which was not included among the GN_o^{NG}. And now, each GN_o^G also includes the genotype of the GN_o^G from whom it descended. Figure 10 illustrates this.

Finally,

$$PIC_3^a = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_ip_j \{ \sum_{k} \frac{1}{2}p_k [\sum_{l} \frac{1}{2}p_l(1-p_ip_l+1-p_kp_l) +$$

$$+ \frac{1}{2}[p_i(1-p_k)+p_k(1-p_i)](1-p_ip_k) +$$

$$+ \sum_{l'} \frac{1}{2}p_{l'}(1-p_jp_{l'}+1-p_kp_{l'}) +$$

$$+ \frac{1}{2}[p_j(1-p_k)+p_k(1-p_j)](1-p_jp_k)] +$$

$$+ \frac{1}{2}[p_i(1-p_j)+p_j(1-p_i)] [\sum_{l'} \frac{1}{2}p_{l''}(1-p_ip_{l''}+1-p_jp_{l''}) +$$

$$+ \frac{1}{2}[p_i(1-p_j)+p_j(1-p_i)](1-p_ip_j)] \}$$

which simplifies to

$$PIC_3^a = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_ip_j \{ \sum_{k} \frac{1}{2}p_k [\sum_{l} \frac{1}{2}p_l(2-p_l(p_l+p_k)) +$$

$$+ \frac{1}{2}[p_i(1-p_k)+p_k(1-p_i)](1-p_ip_k) +$$

$$+ \sum_{l'} \frac{1}{2}p_{l'}(2-p_{l'}(p_j+p_k)) +$$

$$+ \frac{1}{2}[p_j(1-p_k)+p_k(1-p_j)](1-p_jp_k)] +$$

$$+ \frac{1}{2}[p_i(1-p_j)+p_j(1-p_i)] [\sum_{l'} \frac{1}{2}p_{l''}(2-p_{l''}(p_i+p_j) +$$

$$+ \frac{1}{2}[p_i(1-p_j)+p_j(1-p_i)](1-p_ip_j)] \}$$

[150]

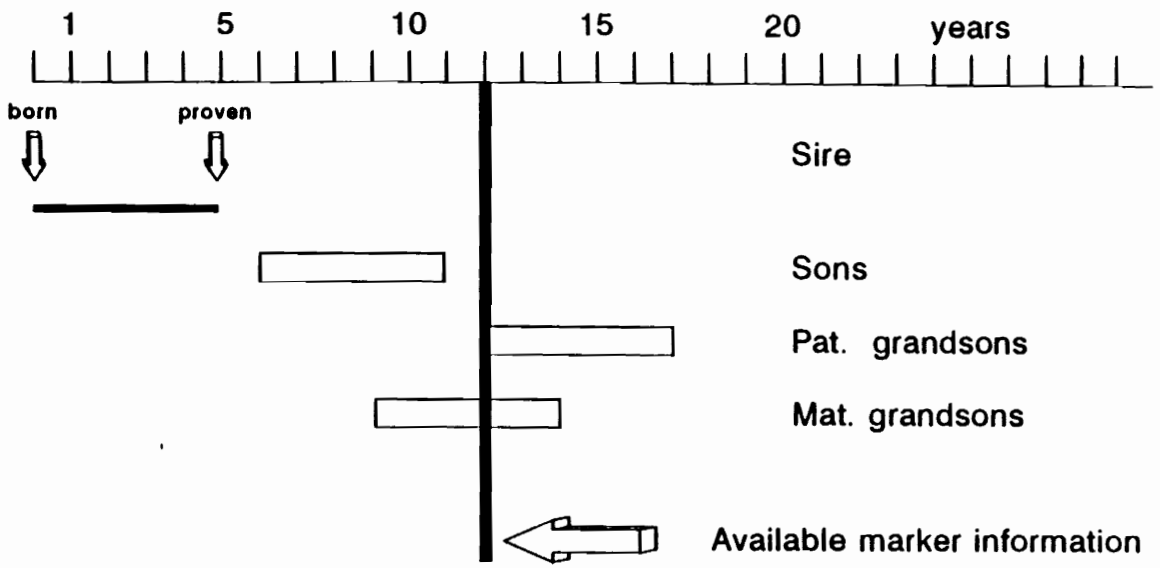

Figure 1

Granddaughter design, expected generations selected by MAS

Marker information from a granddaughter design is available 12 years after a sire is born. At this time, selection effort is dedicated to the sire's grandoffspring.

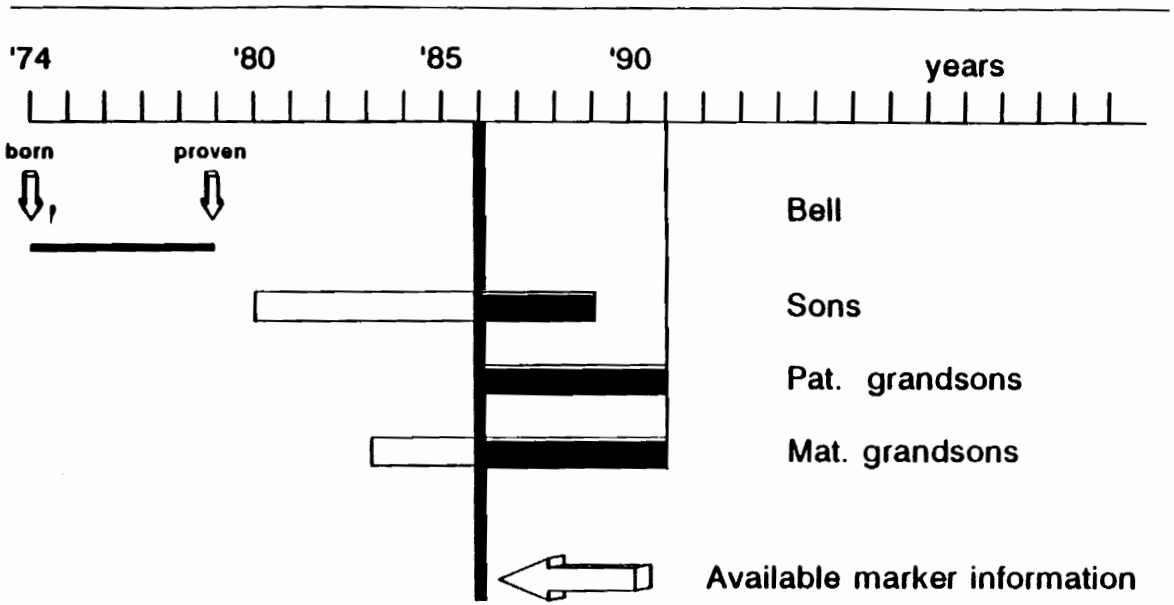

Figure 2

Descendants of Bell that could have been selected from a granddaughter design

Popular bulls like Bell have offspring for a longer period that it is shown in Figure 1. Hence, If marker technology were available in the early eighties, marker information could have been applied to ffspring, grandoffspring and great-grandoffspring of Bell

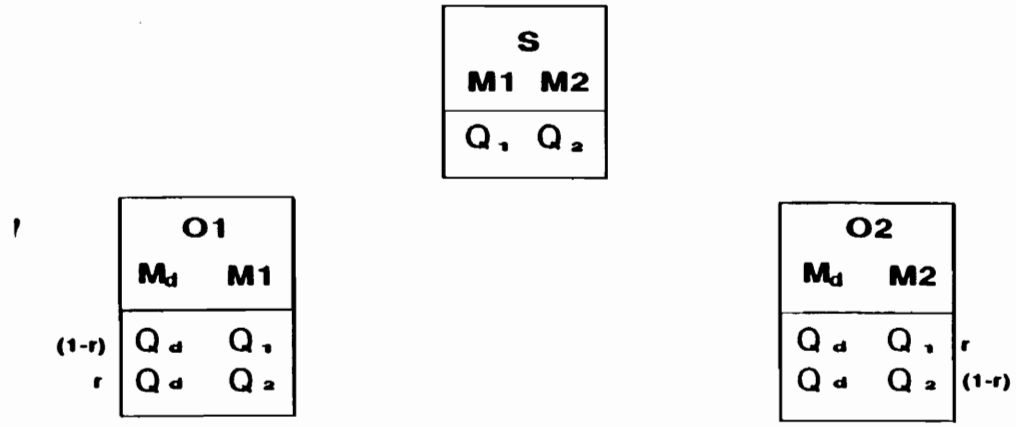

Figure 3

First generation marker inheritance

Boxes denote joint marker/QTL genotypes. The upper portion of each box contains marker genotype, and the lower portion QTL genotype. S, O1 and O2 denote sire and its offspring. O1 and O2 indicate the alternative marker genotypes that an offspring can have. M1, M2 and Md are marker alleles 1, 2 and coming from the sire's mate. Q1,Q2 and Qd denote QTL alleles, with their respective probabilities outside the box. r denotes recombination rate between marker and QTL alleles

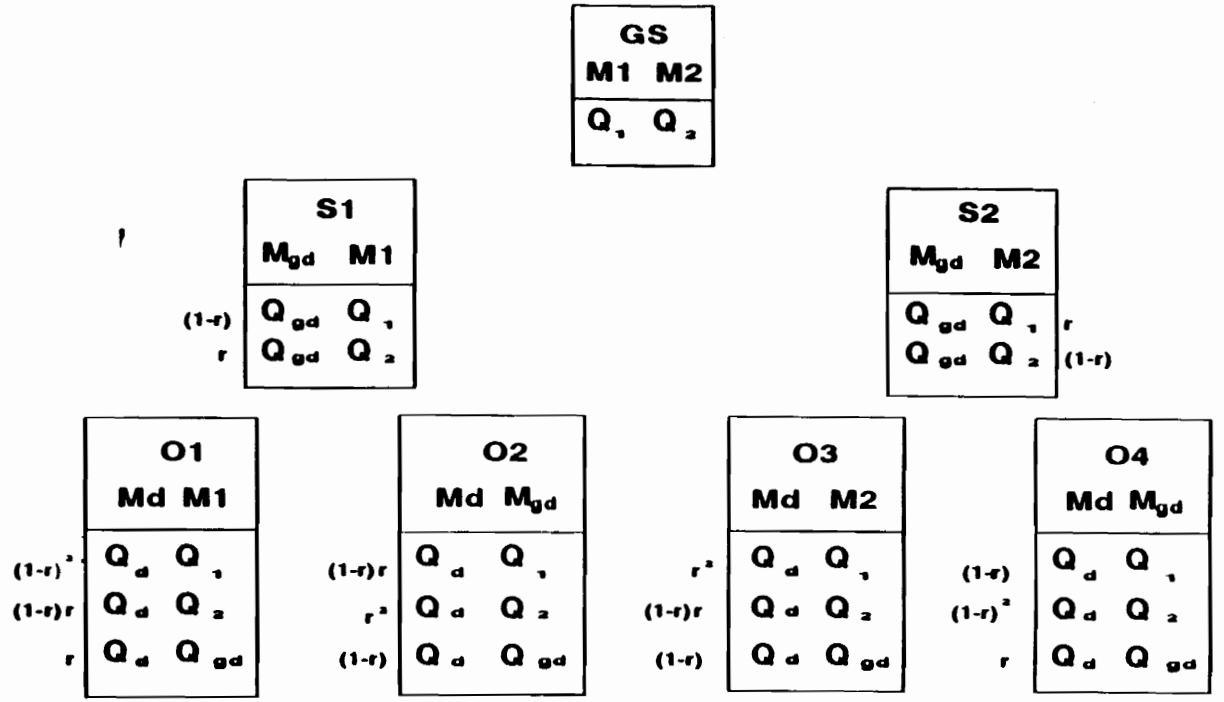

Figure 4

Second generation marker inheritance

This figure adds one generation to Figure 3. Gs denotes grandsire. Mgd and Md are the marker alleles inherited from the mates of the grandsire and sire, respectively.,

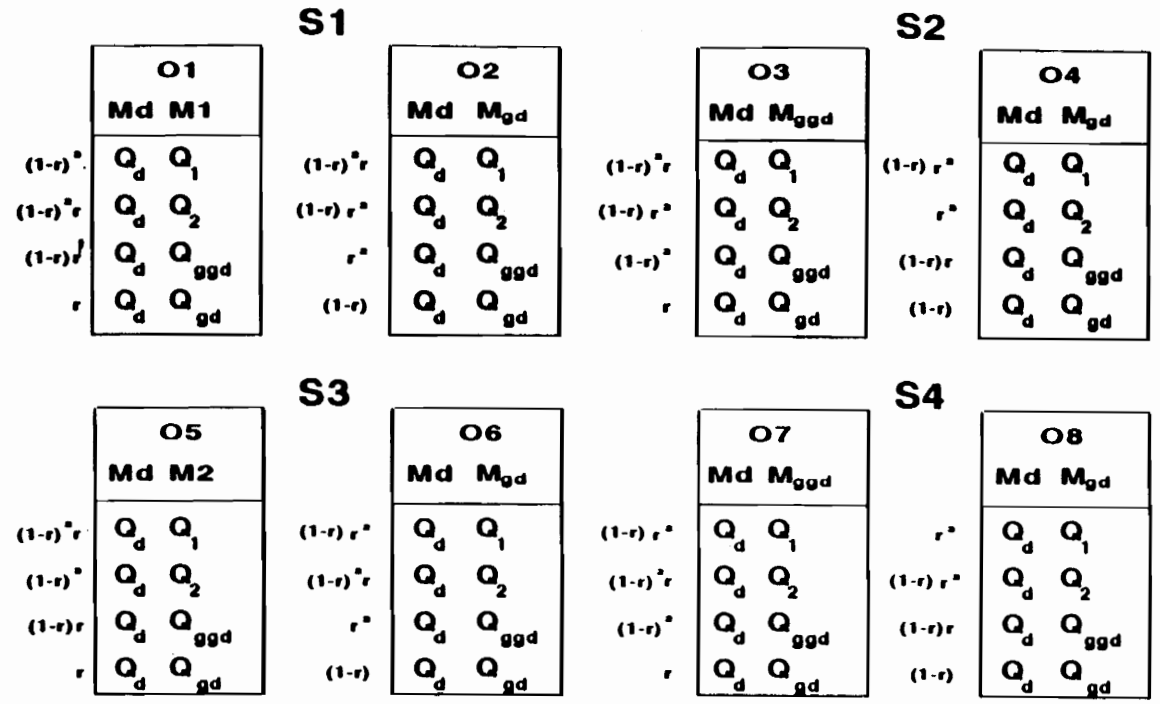

Figure 5

Third generation marker inheritance

Figure 5 continues Figure 4. S1-S4 in Figure 5 are O1-O4 in Figure 4. Mggd denotes the marker effect coming from the great-granddam

Mates	M1M1	M1M2	M2M2	M1M3	M1M4	M1M5	M2M3	M2M4	M2M5	M3M3	M3M4	M3M5	M4M4	M4M5	M5M5
1st gener.	M1M1	M1M1	M1M2	M1M1	M1M1	M1M1	M1M2	M1M2	M1M2	M1M3	M1M3	M1M3	M1M4	M1M4	M1M5
	M1M2	M1M2	M2M2	M1M3	M1M4	M1M5	M1M3	M1M4	M1M5	M2M3	M1M4	M1M5	M2M4	M1M5	M2M5
		M2M2		M1M2	M1M2	M1M2	M2M2	M2M2	M2M2		M2M3	M2M3		M2M4	
				M2M3	M2M4	M2M5	M2M3	M2M4	M2M5		M2M4	M2M5		M2M5	
Mates not genot.	GC	GC	?	GC	GC	GC	?	?	?	GN	GN	GN	GN	GN	GN
	?	?	GC	GN	GN	GN	GN	GN	GN	GN	GN	GN	GN	GN	GN
		GC		?	?	?	GC	GC	GC		GN	GN		GN	
				GN	GN	GN	GN	GN	GN		GN	GN		GN	
Mates genot.	GC	GC	GN	GC	GC	GC	GN	GN	GN	GN	GN	GN	GN	GN	GN
	GN	?	GC	GN	GN	GN	GN	GN	GN	GN	GN	GN	GN	GN	GN
		GC		GN	GN	GN	GC	GC	GC		GN	GN		GN	
				GN	GN	GN	GN	GN	GN		GN	GN		GN	

Figure 6

First generation marker genotypes that are informative for the current and next generation

Five marker alleles are assumed, denoted by M1 to M5. The sire is heterozygous M1M2. "?" denotes a non-informative genotype. "GN" and "GC" indicate genotypes that are informative for the present generation, but only genotypes "GN" will pass information to the following generation.

GN_S		M1M2
	NG	M1M3 M1M4 M1M5 M2M3 M2M4 M2M5
GN_O		M1M3 M1M4 M1M5 M2M3 M2M4 M2M5
	G	M1M2
GN_{GO}		M1M3 M1M4 M1M5 M2M3 M2M4 M2M5 M1M2 M3M4 M3M5 M4M5

Figure 7

Genotypes that are informative for the next generation

GNs denotes a particular heterozygous sire with alleles M1M2. The offspring of M1M2 with genotypes G information to the second generation. These genotypes are listed both for mates genotyped (G) or no and assuming that there are only five alleles segregating in the population. GNgo denote the genoty grandoffspring of M1M2 that are able to pass information to the next generation. The genotypes list mates genotyped and not genotyped, although at different probabilities.

Sire	Sire Allele	Offspring genotype	Dam genotype	Prob of dam genotype	Prob of offspring genotype
$M_i M_j$	M_i	$M_i M_j$	$M_j M_l$ $l \neq j$	$\sum 2 p_j p_l = 2 p_j (1 - p_j)$	$.5 p_j (1 - p_j)$
	M_j	$M_i M_j$	$M_i M_l$ $l \neq i$	$\sum 2 p_i p_l = 2 p_i (1 - p_i)$	$.5 p_i (1 - p_i)$

Figure 8

Probability of obtaining first generation genotypes identical to the sire when mates are genotyped

A sire with marker alleles Mi and Mj can have offspring with genotype MiMj from mates of genotype MiMl or MjMl. The probability of obtaining such offspring genotypes is a quarter the probability of dams.

GN_s	GN_o	$p(GN_o \mid GN_s)$	GN_{go}	$p(GN_{go} \mid GN_o)$
$M_i M_j$	$M_i M_k$	$.5p_k$	$M_i M_l$	$.5p_l$
			$M_k M_l$	$.5p_l$
	$M_j M_k$	$.5p_k$	$M_j M_{l'}$	$.5p_{l'}$
			$M_k M_{l'}$	$.5p_{l'}$

Figure 9

Inheritance of marker alleles through animals of GN genotype when mates are not genotyped

From GNs=MiMJ, GNo can be MiMk or MjMk (with k different from i andr j). From GNo=MiMk, GNgo can be MiMl or MkMl (with l different from i andr k). From GNo=MjMk, GNgo can be either MjMl' or MkMl' (with l' different from j and k)

GN_s	GN_o	$p(GN_o \mid GN_s)$	GN_{go}	$p(GN_{go} \mid GN_o)$
$M_i M_j$	$M_i M_k$	$.5p_k$	$M_i M_i$	$.5p_i$
			$M_k M_i$	$.5p_i$
			$M_i M_k$	$.5[p_i(1-p_k)+p_k(1-p_i)]$
	$M_j M_k$	$.5p_k$	$M_j M_i$	$.5p_i$
			$M_k M_i$	$.5p_i$
			$M_j M_k$	$.5[p_j(1-p_k)+p_k(1-p_j)]$
	$M_i M_j$	$.5[p_i(1-p_j)+p_j(1-p_i)]$	$M_i M_i$	$.5p_i$
			$M_j M_i$	$.5p_i$
			$M_i M_j$	$.5[p_i(1-p_j)+p_j(1-p_i)]$

Figure 10

Inheritance of marker alleles through animals of genotype GN when mates are genotyped

Figure 10 differs from Figure 9 because GNi (i=o,go) include genotypes identical to their sires.

Results and Discussion

PGA at equilibrium differs for males and females because of their different equilibrium variances. Results in this study refers only to PGA for males. PGA for females, ceteris paribus, differs only slightly from PGA for males.

First generation case

Figure 11 shows the percent gain in accuracy (PGA) obtained when the first generation selection index (equation **[12]**) is applied to selection for milk yield. PGA is shown for different sizes of marker effects, or for different D_p values, and for situations where all or half of the phenotypic information comes from pedigree ,i.e., when PTA=PA or $(x_2+x_3)=.5$, respectively. The values of D_p considered are 100, 200, 300, 400, 500 and 600 lbs, which correspond to a fraction of 2.56, 5.76, 10.24, 16.00 and 23.04 % of the total Mendelian sampling variance of milk yield in

dairy cattle.

As expected, PGA increases with D_p. Cowan et al. (1990) found a marked QTL (i.e., the prolactin locus) in an elite sire Holstein for which D_p was 624 lbs. This marker promised to be very useful for MAS. However, most of the marked QTLs will likely to be of smaller size. These QTLs are not only difficult to detect, but may also produce only small gains in accuracy of selection.

The usefulness of marker information is related to the availability of information other than pedigree. Figure 11 shows that PGA increases sharply with the size of D_p when parent average represents 100% of the Animal Model (AM) information, but only slightly when 50% of the AM information comes from the pedigree.

Figure 12 shows that for a given D_p=400 lbs, the usefulness of marker information diminishes when phenotypic information on Mendelian sampling effect is available. The maximum PGA is obtained when $x_2+x_3 = 0$ (i.e., when all the available information comes from the pedigree). PGA decreases significantly as initial phenotypic information on the Mendelian sampling effect becomes available, suggesting that marker information is useful only for cases when thre is little or no phenotypic information from offspring or own records. PGA tends to zero when the animal has a large amount of phenotypic information (i.e., when x_2+x_3 tends to 1). When x_2+x_3 approaches to 1, the reliability of the animal

evaluation also tends to 1. A reliability of 1 indicates that all the Mendelian sampling effect in an animal has been explained by phenotypic records and, therefore, the marker information is of no additional value. Table 1 illustrates the contribution of daughter information to a sire's PTA. Table 1 assumes that all mates are identified. Only eight daughters are needed for x_3 to reach a value of .5 (i.e., half of the information used to compute the sire PTA comes from its eight daughters). Table 2 shows that x_2 for a cow with 3 records and no progeny is almost equal to .5. From Tables 1 and 2 and Figure 12 it is likely that the use of marker information be confined to cases when the sole phenotypic information available is parent average (e.g., for young candidates prior to be progeny tested).

The usefulness of the marker information, especially when pedigree is the sole source of information, depends on the reliability of parent average (see Appendix A3). Figure 13 shows that the marker information is more useful as the reliability of parent average decreases. However, in most practical circumstances, the reliability of parent average will be above .3.

The usefulness of the marker information is also related to the accuracy of the estimated D_g. Figure 14 shows PGA in the first generation, as a function of $SE(D_p)$. PGA is reduced as the precision of D_p decreases. Marker effects of large size may be virtually useless when their effect was estimated with few data.

For instance, the phenotypic effect of variability at the prolactin locus was estimated from a granddaughter design with a data set of 26 sons (Cowan et al., 1990). Although the authors did not present information about $SE(D_p)$, it is possible, from the t-value that they used in the test, to infer that $SE(D_p)$ = 543 lbs. This large standard error is expected because of the small data set used to test the significance of variability at the prolactin locus. The usefulness of the marker at the prolactin locus is reduced because of the poor precision (i.e., large $SE(D_p)$). For instance, a PGA of 25.27% in the first generation, for D_p = 600 lbs and Rel(PA)=.375, is reduced when $SE(D_p)$ = 300 lbs to only 3.62% when $SE(D_p)$ = 543 lbs.

Second generation case

Figure 15 shows PGA as a function of different sizes of D_p when marker information is used for selection among the grandoffspring. As in the first generation, PGA increases with D_p, especially when pedigree is the only phenotypic information available. However, PGA in Figure 15 is lower than that in Figure 11 for all D_p values. Recombination between marker and QTL and covariance between marker information and parent average reduce the usefulness of the marker information for grandoffspring.

Although knowledge about the recombination rate (r) is not needed to compute the first generation selection index (i.e., r is

included in D_p), it is required for the second and third generations. Figure 15 shows that large recombination rates between marker and QTL reduce the usefulness of marker information. However, it is likely to expect that marked QTLs will generally show a small recombination rate with the marker locus.

Figures 15 and 16 were computed by assuming **[56]** (i.e., $\bar{\alpha}_{gd}=.5(\alpha_1+\alpha_2)$). Figure 17 assumes that **[71]** (i.e. $\bar{\alpha}_{gd}=p\alpha_1+(1-p)\alpha_2$) holds. When p in **[71]** is different from .5, two different selection indices must be used because of the different Mendelian sampling variances in **[77]** and **[78]**. The two curves in Figure 17, which converge on the right for p = .5, represent PGA for the two different indices. The two curves are not symmetric with respect to their common point p = .5. This is so because PGA is also a function of the recombination rate (r), and Figure 17 was computed by assuming that r=.1. It was shown in the Methods section that one of the variances of the segregation terms becomes null when p=r. Therefore, a symmetric set of curves in Figure 17 could only be obtained by assuming r=.5, which is impossible because for r=.5 D_p would be expected to be zero.

To compute the increase in accuracy due to the use of marker information across families in the second generation, a total PGA (PGA_t) was computed. If an equal probability of occurrence of both family types (see Figure 4) is assumed, then PGA_t is equal to the average of the PGAs obtained for each of the two selection indices

(i.e, the average of the two PGAs that correspond to a particular p in Figure 17). Given the lack of symmetry of the curves in Figure 17, the lowest total PGA_t is obtained when p = .5. Therefore, equation **[56]** provides a conservative approach when **[71]** holds but no knowledge about p is available.

Third generation case

Figure 18 shows the increase in accuracy due to the application of the marker information for selection in the third generation. Figure 18 assumes **[119]**. Under these assumptions, two different selection indices are also required. These two selection indices are displayed in Figure 18. The general shape of the curves in Figure 18 is, again, similar to the previous two generations, but with a generally reduced PGA.

Figure 18 shows that for MAS to be useful in the third generation, little or no phenotypic information on the Mendelian sampling effect should be available, and/or the marker effect must be large.

Polymorphism of the markers

All of the PGA values presented in Figures 11 to 18 were calculated assuming a perfectly traceable inheritance of the marker alleles (i.e., PIC=1). However, when the marker is not highly

polymorphic, PGA in generation i (PGA_i) is reduced to $PGA_i * PIC_i$ (i=1,2,3).

The need for highly polymorphic markers increases over generations, especially when mates are not genotyped. Figure 19 shows that PIC reduces over generations. The need for highly polymorphic markers is reinforced because not only PIC, but also PGA decreases over generations. A QTL of relatively small effect and marked with a dimorphic marker may not be useful in any generation beyond the first.

Figures 19 and 20 show how PIC increases with polymorphism of the marker locus. Also, the difference in PIC between the generations is smaller when highly polymorphic markers are available. The availability of highly polymorphic markers reduces the need to genotype mates in order to fully capture the PGA.

PIC values are particularly small when the marker is dimorphic. Figure 20 shows that PIC is zero for the second and third generations when the marker is dimorphic and mates are not genotyped. This implies that dimorphic markers (e.g., the prolactin locus, Cowan et al., 1990) require that mates be genotyped when the marker information is intended for use in the second or later generations. If mates are not genotyped, PIC and PGA are zero.

Even if mates are genotyped, traceability of the marker inheritance is poor for dimorphic markers . Figure 19 shows that for dimorphic markers and mates genotyped, PIC in the second and third generation are only .09% and .02% respectively.

Figures 19 and 20 assume that all marker alleles occur at identical frequencies. Figure 21 shows first generation PIC for different degrees of marker polymorphism when one of the marker alleles has a frequency of p=.50 or p=.90, and the remaining alleles each an identical frequency of (1-p)/(n-1), where n is the total number of marker alleles. Figure 21 illustrates that if one of the alleles is predominant in the population (p=.9), then the probability of occurrence of a heterozygous sire is reduced, and PIC is low irrespective of the polymorphism of the marker locus.

PIC in Figures 19 to 21 take into account the probability of finding a heterozygous sire or total PIC. Figures 22 to 26 show PIC within the family of a given heterozygous sire. Figure 22 gives PIC for the offspring of a particular heterozygous sire when mates are genotyped. Traceability of the marker information is generally very high for markers of high levels of polymorphism. Only dimorphic markers with alleles at intermediate frequencies show some important degree of reduction in PIC. This is because a larger number of first generation animals is heterozygous at intermediate frequencies and, therefore, not informative for their generation. However, when one of the marker alleles is rare,

traceability is improved and PIC is exceeds .90, even for dimorphic markers.

Figure 23 shows PIC among the grandoffspring of a particular grandsire when mates are genotyped. Here the allelic frequencies play a more important role in determining PIC as compared to the first generation. If one of the marker alleles present in the grandsire is common in the population (p=.9), then the proportion of homozygous individuals among the offspring of the sire increases. Homozygous offspring are informative for their generation, but not for the next. Therefore, PIC in the second generation tends to be small when one of the alleles present in the sire is common in the population. Only for dimorphic markers is PIC in the second generation smaller at intermediate frequencies. For dimorphic markers, the only possible heterozygous offspring, which is non-informative for either the first or second generation, occurs most frequently at intermediate allelic frequencies.

Figure 24 shows PIC among the great-grandoffspring of a great-grandsire. PIC is again reduced in the third generation. Note that markers of limited polymorphism, although possesing relatively high PIC in the first generation, lose a large amount of information when they are applied to the second and third generations. Dimorphic markers, even when linked to a QTL of large size, may be of little use when they are applied to the grand- or great-grandoffspring of a sire.

PIC in Figures 22 to 24 are reduced further when mates are not genotyped (Figures 25 to 27). Figure 27 shows an almost complete loss of marker information at the third generation, when one of the alleles present in the sire is common in the population (p=.9) and mates are not genotyped.

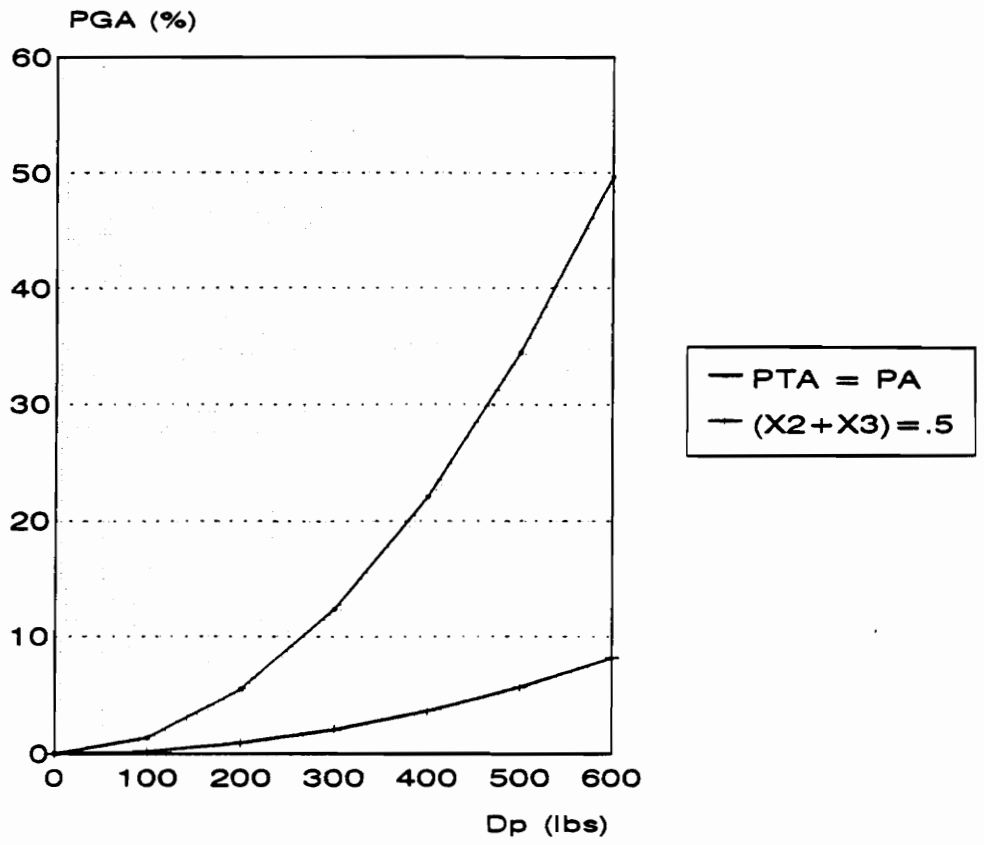

Figure 11

Percentage of gain in accuracy (PGA) in the first generation as a function of the marker effect (Dp)

Curves "PTA=PA" and "(x2+x3)=.5" show PGA when parent average (PA) is 100% or 50% of the information used to compute predicted transmitting ability (PTA). Assumptions: Rel(PA)=.375; SE(Dp) = .1 Dp.

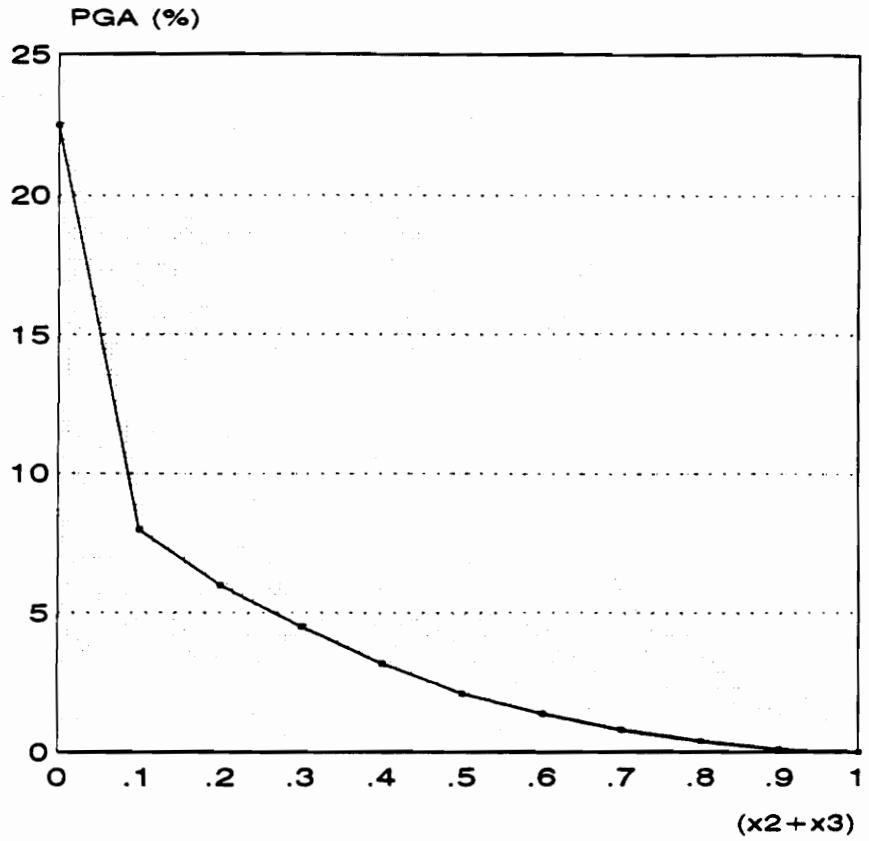

Figure 12

Percentage of gain in accuracy (PGA) in the first generation as a function of phenotypic information other than parent average (x2+x3)

(x2+x3) goes from 0 (only parent average available for PTA) to 1 (large number of records are available). Assumptions: Rel(PA)=.375; Dp= 400; SE(Dp) = 10.

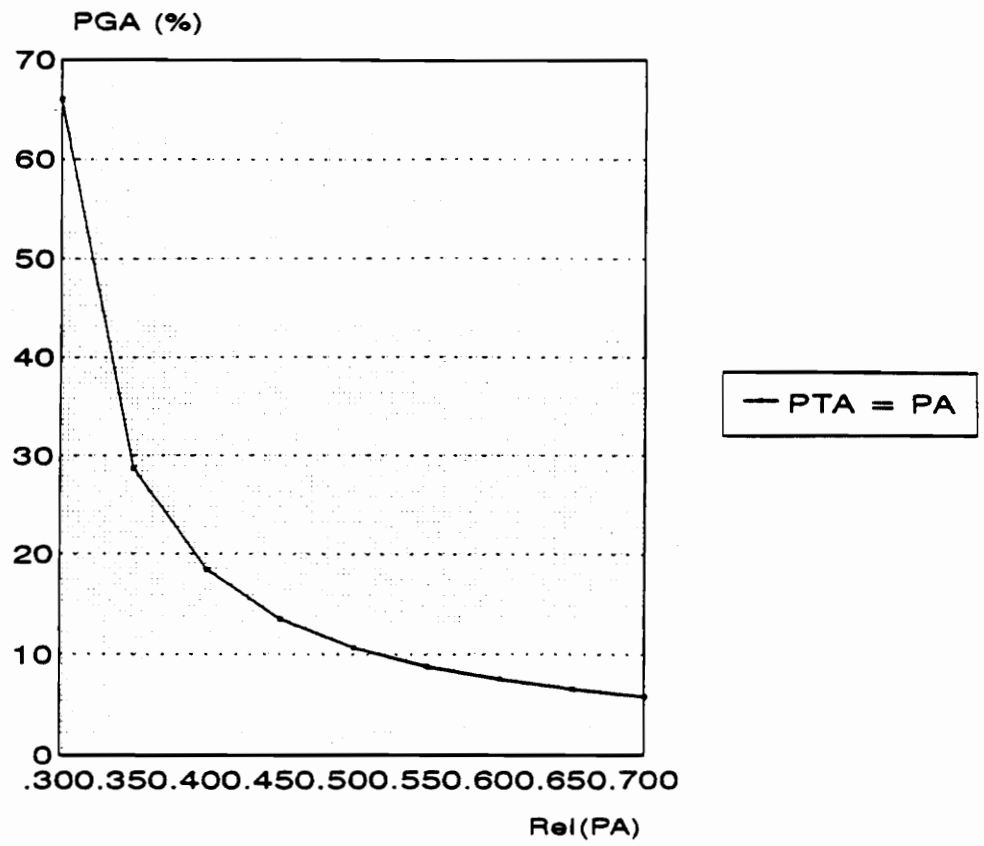

Figure 13

Percentage of gain in accuracy (PGA) for the first generation as a function of reliability of parent average (Rel(PA))

PGA is shown for "PTA=PA" (i.e., parent average (PA) is 100% of the information used to compute predicted transmitting ability (PTA). Assumptions: Dp=400; SE(Dp) = 10.

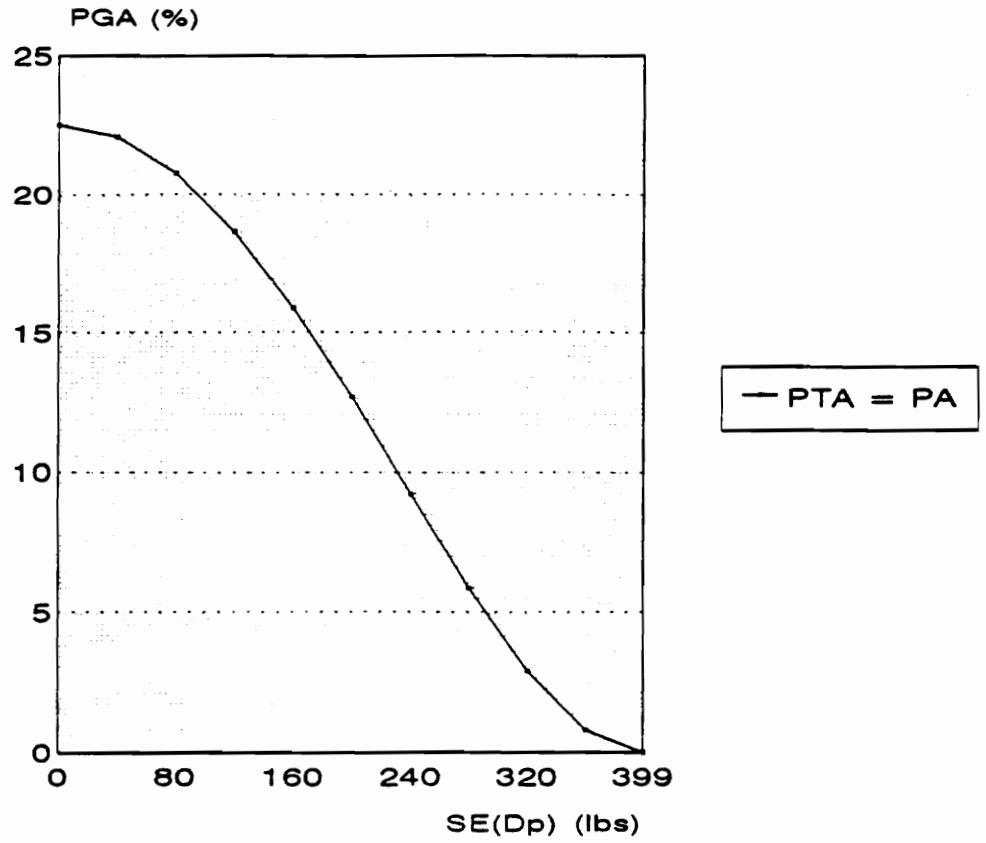

Figure 14

Percentage of gain in accuracy (PGA) in the first generation as a function of the standard error of the estimated marker effect (SE(Dp))

PGA is shown for "PTA=PA", i.e., when parent average (PA) is 100% of the information used to compute predicted transmitting ability (PTA). Assumptions: Rel(PA)=.375; Dp = 400.

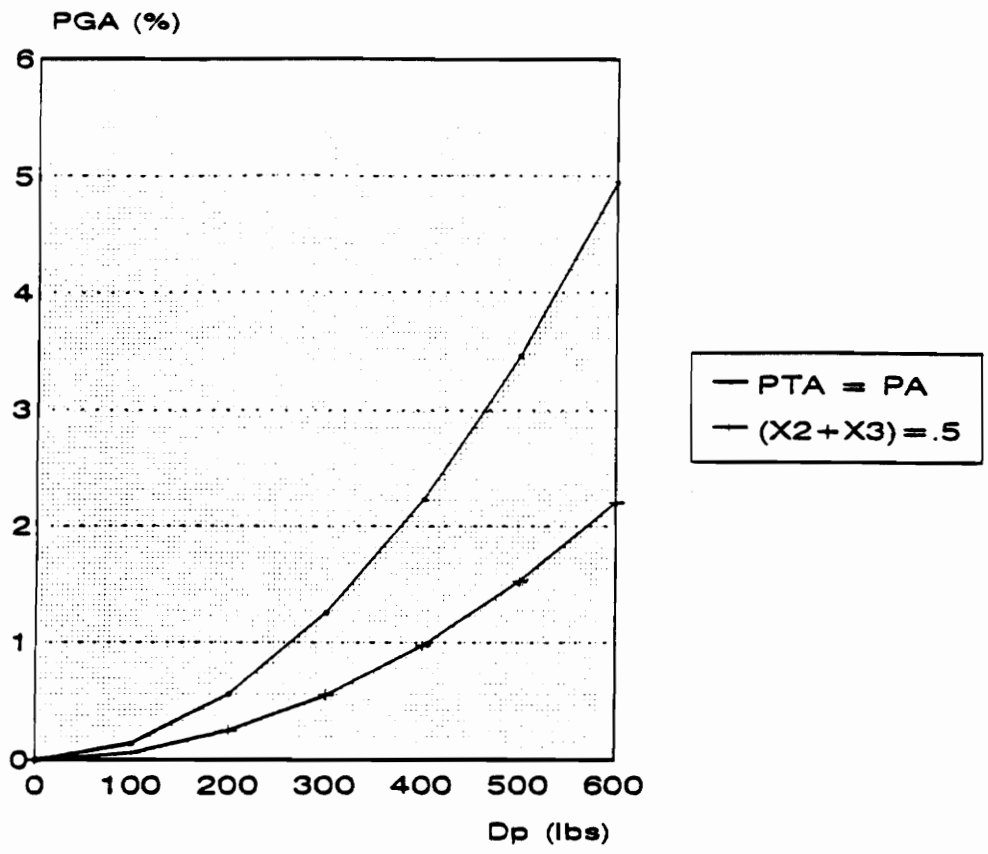

Figure 15

Percentage of gain in accuracy (PGA) in the second generation as a function of the marker effect (Dp)

Curves "PTA=PA" and "(x2+x3)=.5" show PGA when parent average (PA) is 100% or 50% of the information used to compute predicted transmitting ability (PTA). Assumptions: Rel(PA)=.375; SE(Dp) = .1 Dp; r=.1 ; p=.5.

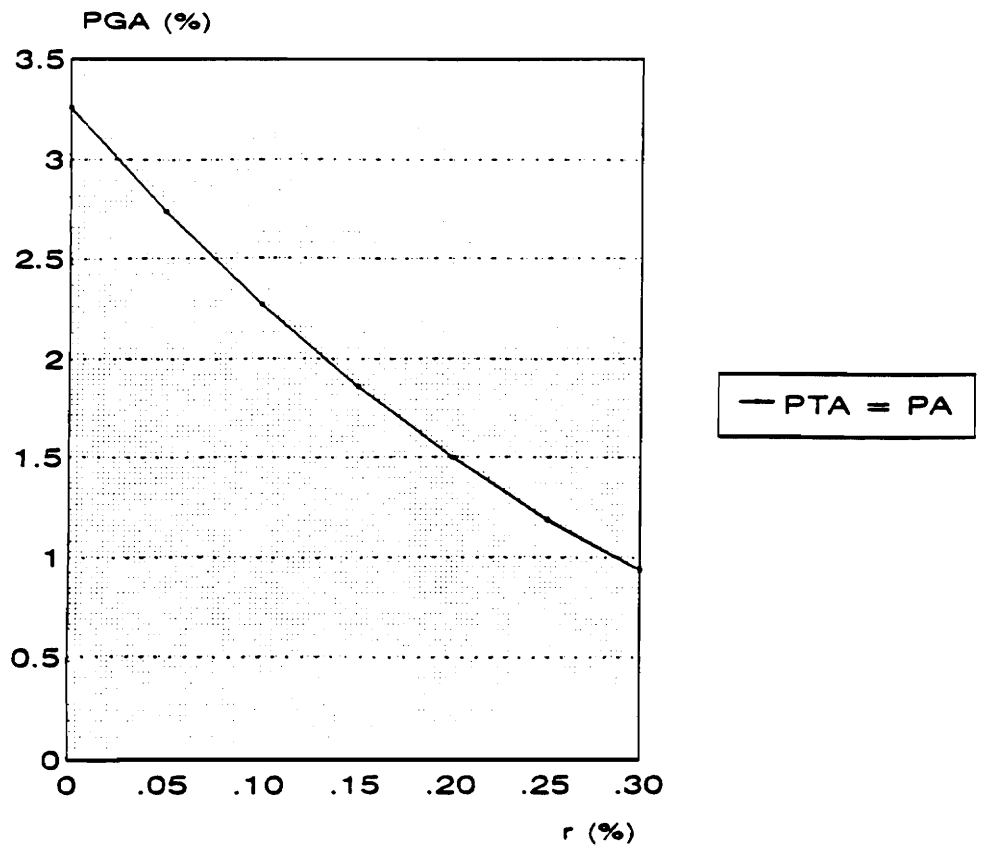

Figure 16

Percentage of gain in accuracy (PGA) in the second generation as a function of the recombination rate (r) between marker and QTL.

PGA is shown for "PTA=PA", i.e., when parent average (PA) is 100% of the information used to compute predicted transmitting ability (PTA). Assumptions: Rel(PA)=.375; SE(Dp) = 10; Dp = 400; p=.5.

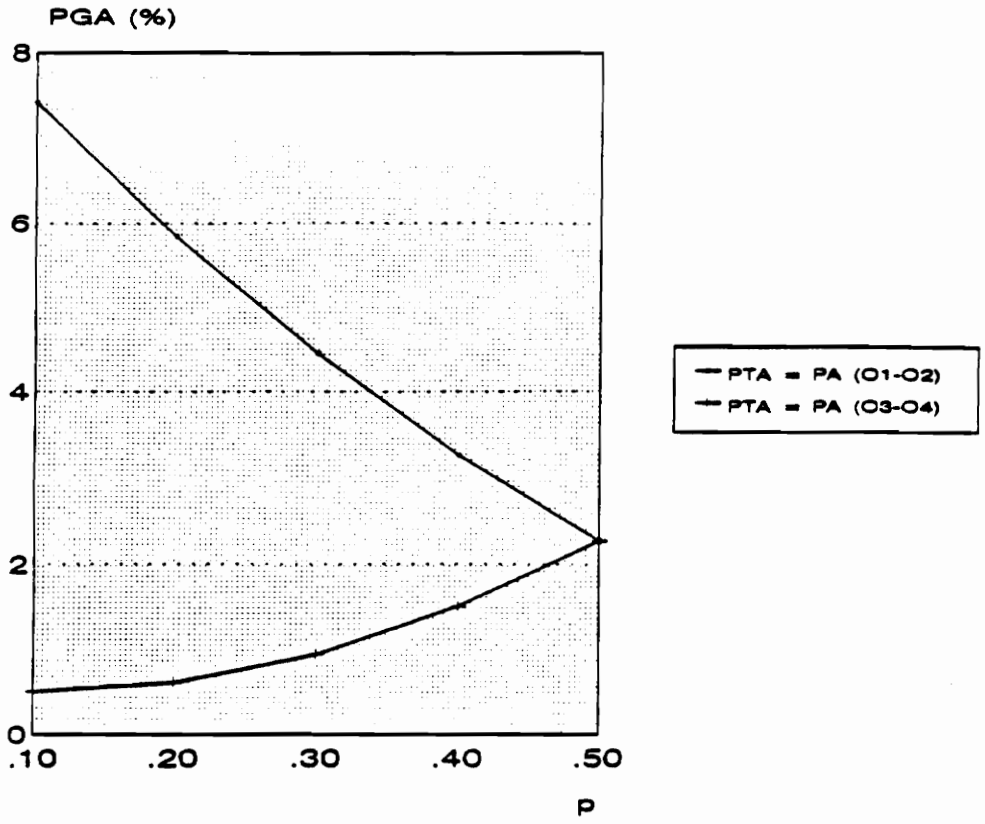

Figure 17

Percentage of gain in accuracy (PGA) in the second generation as a function of the proportion (p) of each QTL allele effect present in the sire that conforms the dam QTL allele (Qd).

PGA is shown for "PTA=PA", i.e., when parent average (PA) is 100% of the information used to compute predicted transmitting ability (PTA). Assumptions: Rel(PA)=.375; SE(Dp) = 10; Dp = 400; r=.1

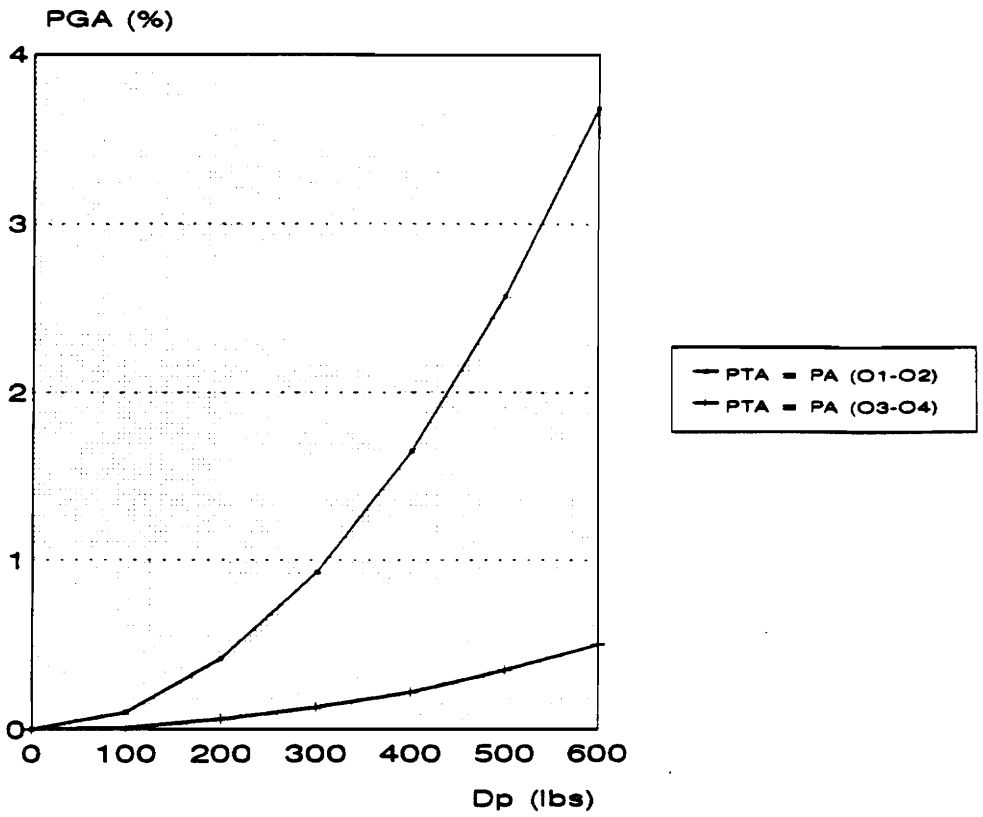

Figure 18

Percentage of gain in accuracy (PGA) in the third generation as a function of the marker effect (Dp)

PGA is shown for "PTA=PA" (i.e., when parent average (PA) is 100% of the information used to compute predicted transmitting ability (PTA). The two curves represent PGA within the families O1-O2 and O3-O4 (see Figure 5). Assumptions: r=.1, Rel(PA)=.375; SE(Dp) = .1 Dp., p=.5.

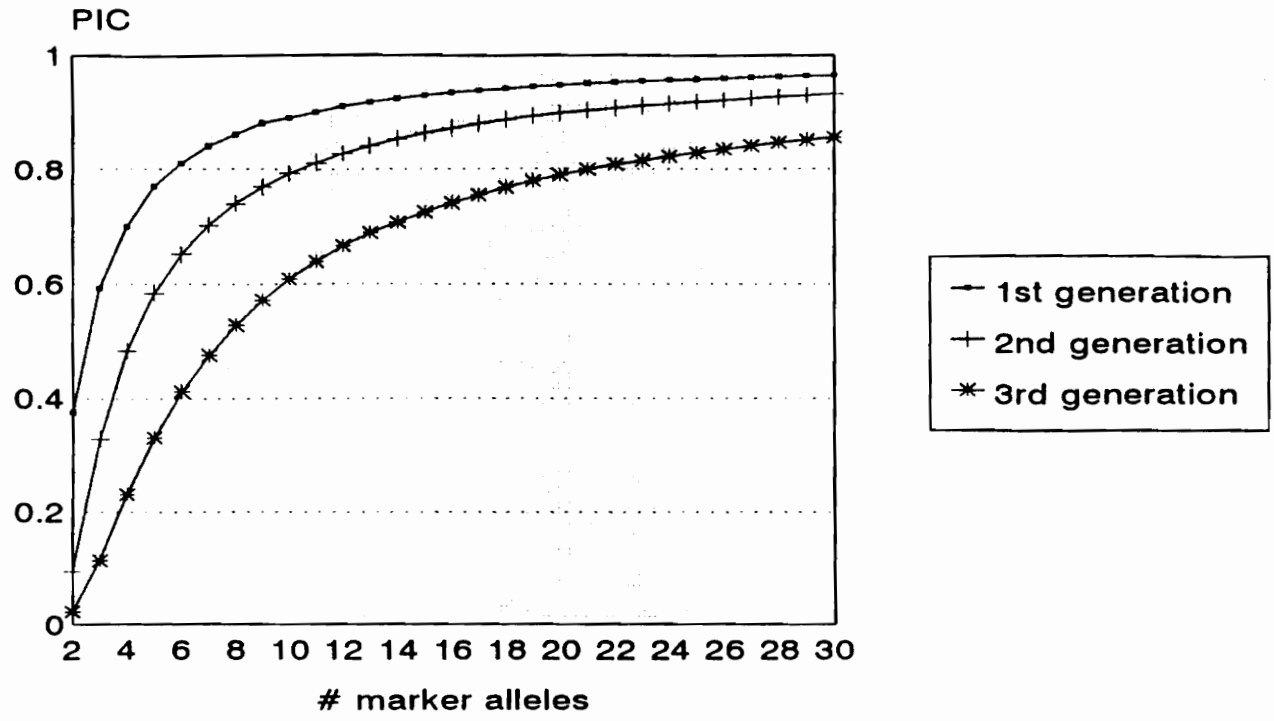

Figure 19

PIC in the population for the first three generations and mates genotyped.

PIC in the population takes the probability of finding an heterozygous sire into account. Assumption: all alleles are equally probable (i.e., when # of marker alleles (n) = 2, then p1=p2=.5; when n = 3, p1=p2=p3=.33, and so on).

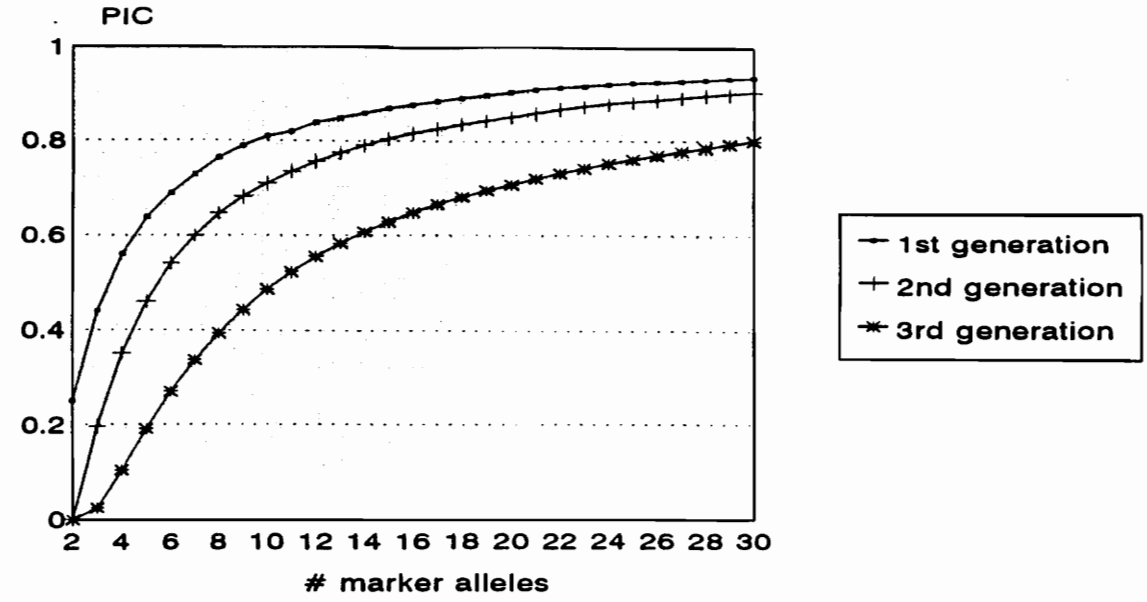

Figure 20

PIC in the population for the first three generations and mates not genotyped

PIC in the population takes the probability of finding an heterozygous sire into account. Assumption: alleles are equally probable (i.e., when # of marker alleles (n) = 2, then p1=p2=.5; when n = 3, p1=p2=p3=.33, and so on).

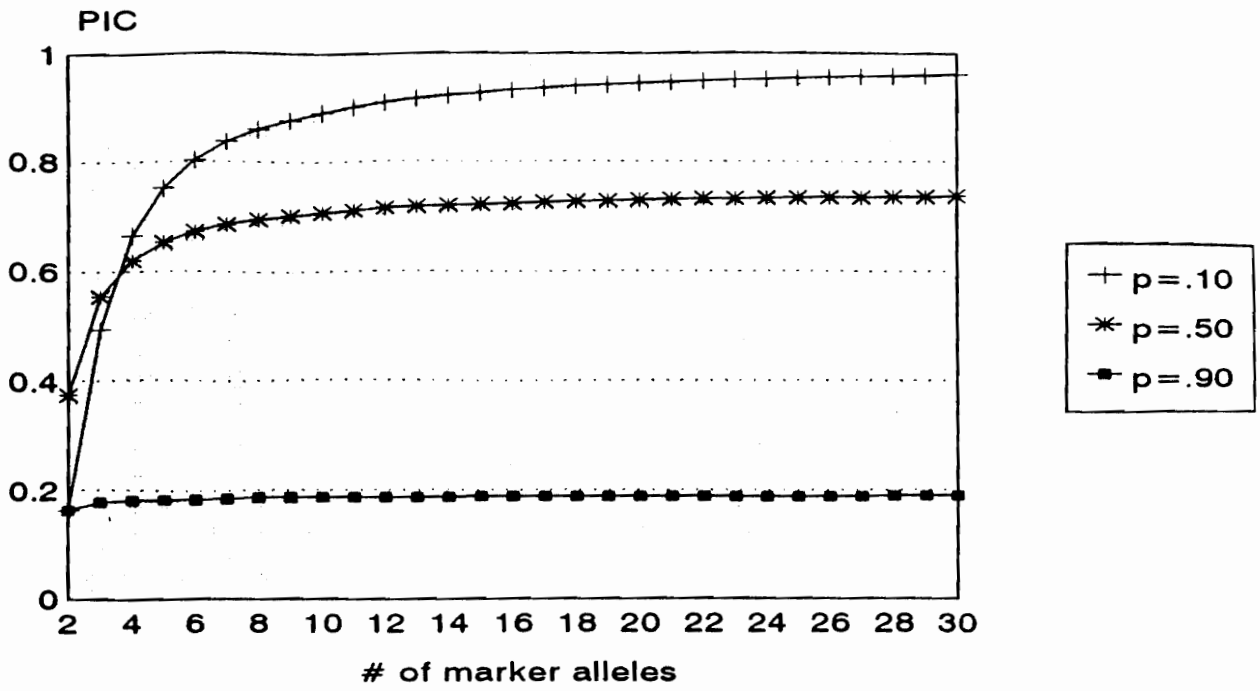

Figure 21

PIC in the population and in the first generation for different frequencies of marker alleles and different number of marker alleles

PIC in the population takes the probability of finding an heterozygous sire into account. p denotes the frequency of one allele. The remaining alleles are equally probable.

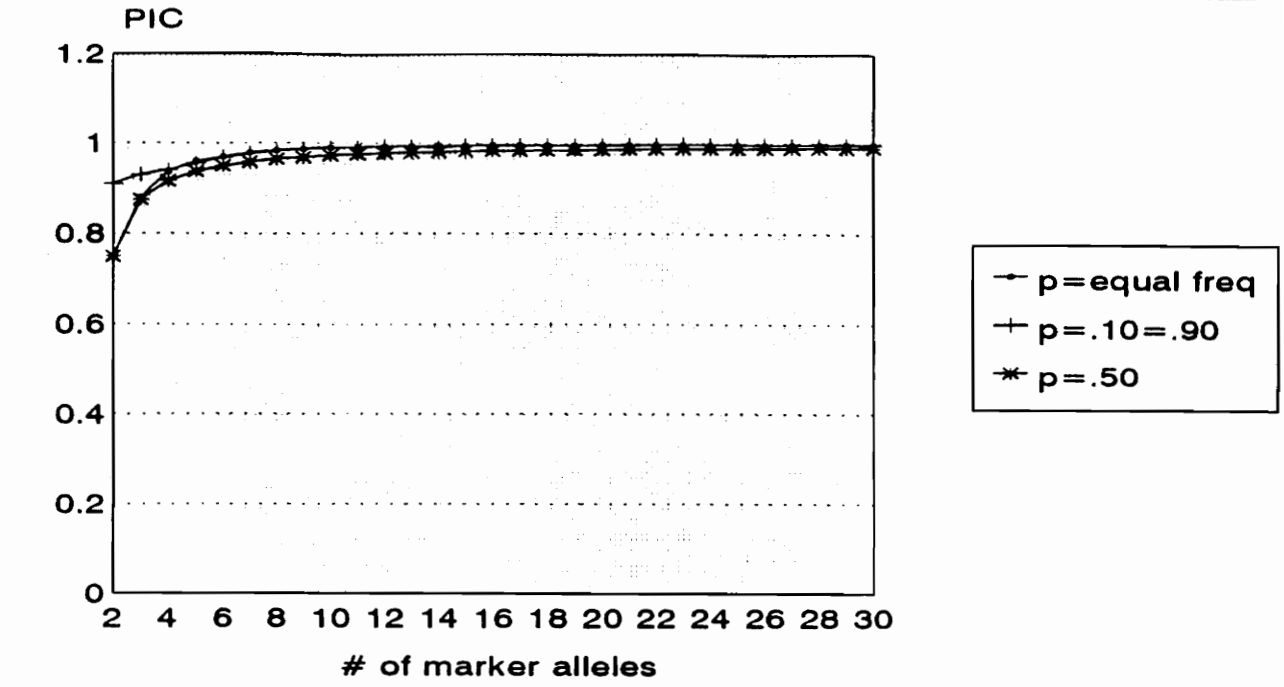

Figure 22

PIC within family in the first generation for different marker allelic frequencies and mates genotyped, as a function of the number of marker alleles

PIC is computed given that a heterozygous sire is found. p denotes the frequency of one of the marker alleles present in the heterozygous sire. The remaining alleles are equally probable

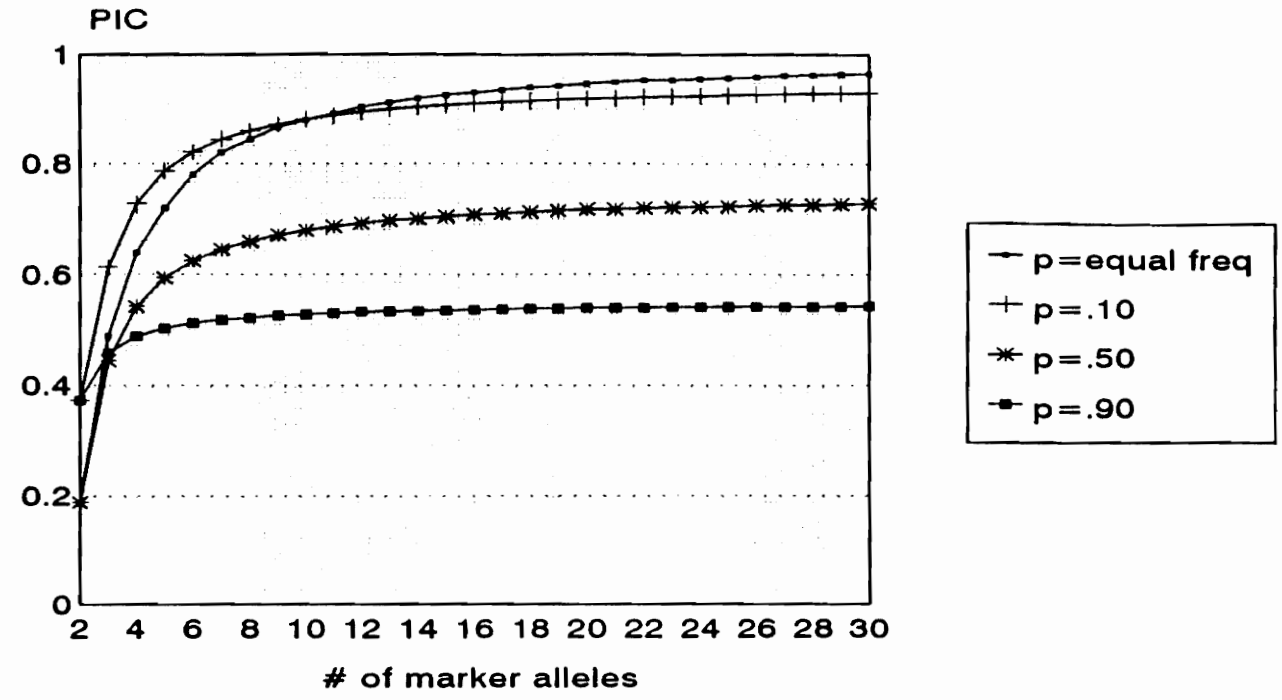

Figure 23

PIC within family in the second generation for different marker allelic frequencies and mates genotyped, as a function of the number of marker alleles

PIC is computed given that a heterozygous sire is found. p denotes the frequency of one of the marker alleles present in the heterozygous sire. The remaining alleles are equally probable.

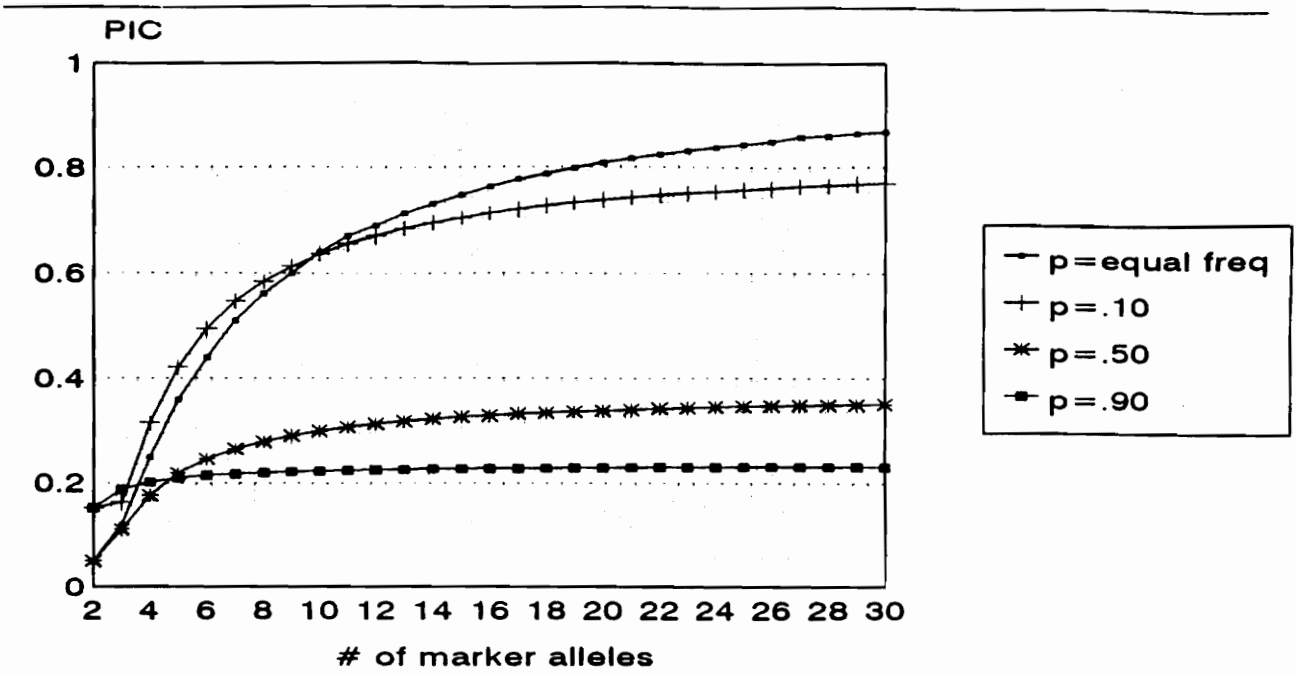

Figure 24

PIC within family in the third generation for different marker allelic frequencies and mates genotyped, as a function of the number of marker alleles

PIC is computed given that a heterozygous sire is found. p denotes frequency of one of the marker alleles present in the heterozygous sire. The remaining alleles are equally probable.

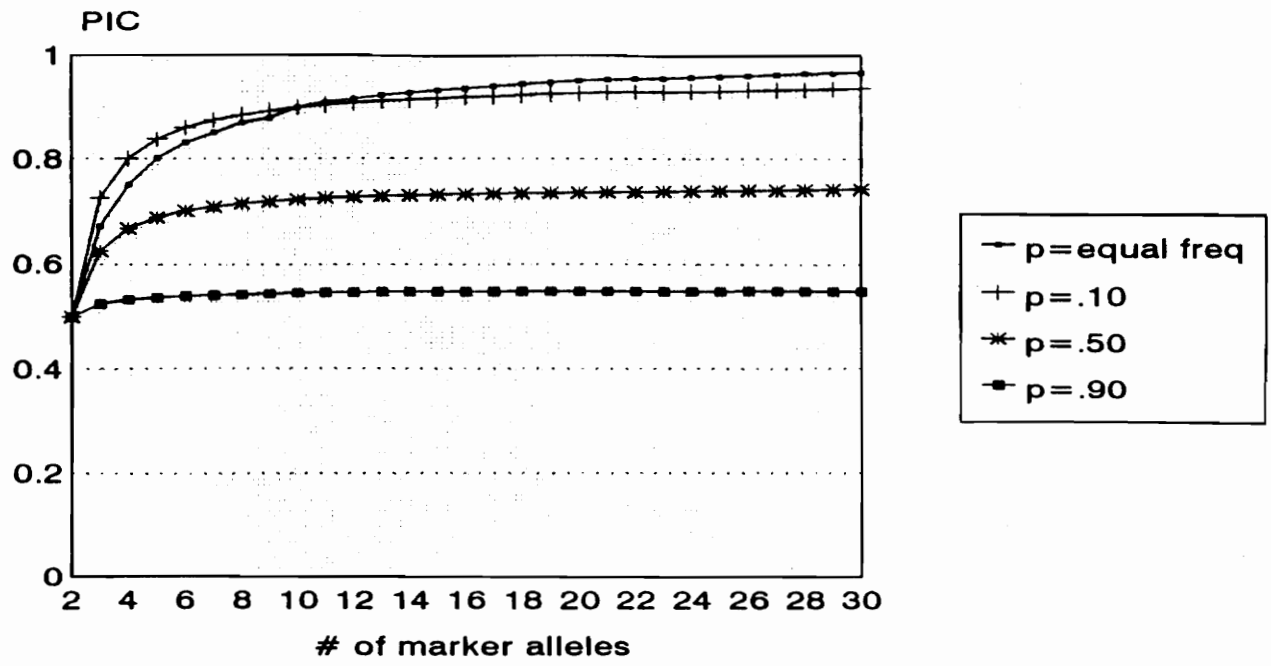

Figure 25

PIC within family in the first generation for different marker allelic frequencies and mates not genotyped, as a function of the number of marker alleles

PIC is computed given that a heterozygous sire is found. p denotes the frequency of one of the marker alleles present in the heterozygous sire. The remaining alleles are equally probable.

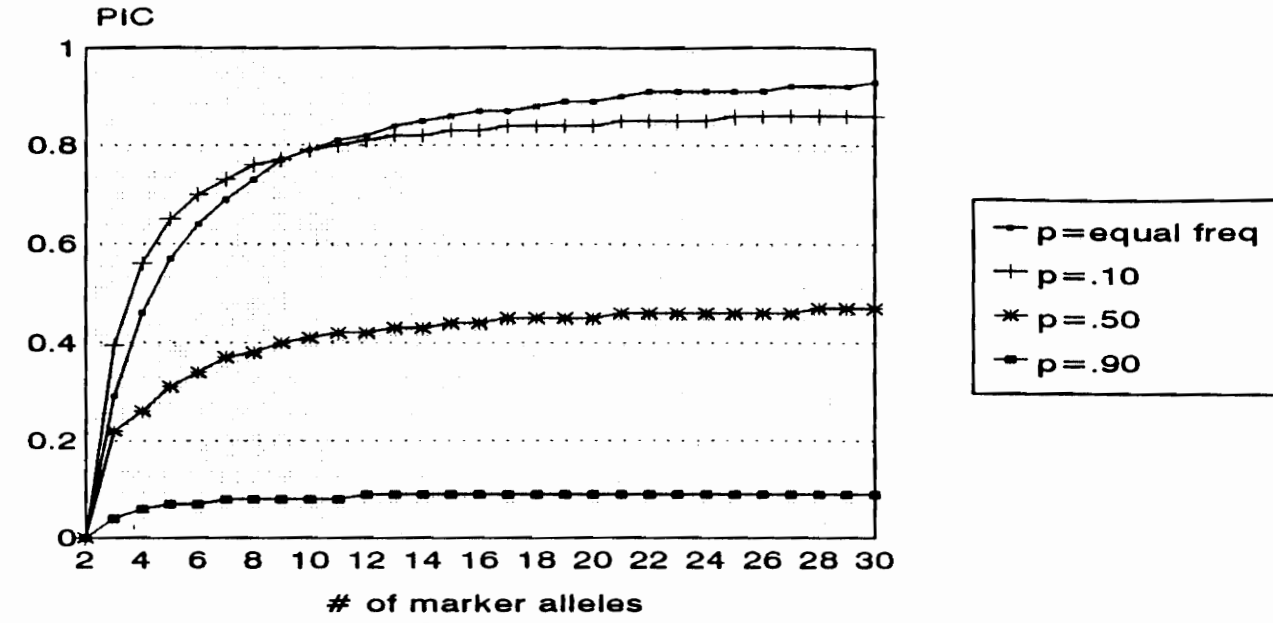

Figure 26

PIC within family in the second generation for different marker allelic frequencies and mates not genotyped, as a function of the number of marker alleles.

PIC is computed given that an heterozygous sire is found. p denotes the frequency of one of the marker alleles present in the heterozygous sire. The remaining alleles are equally probable.

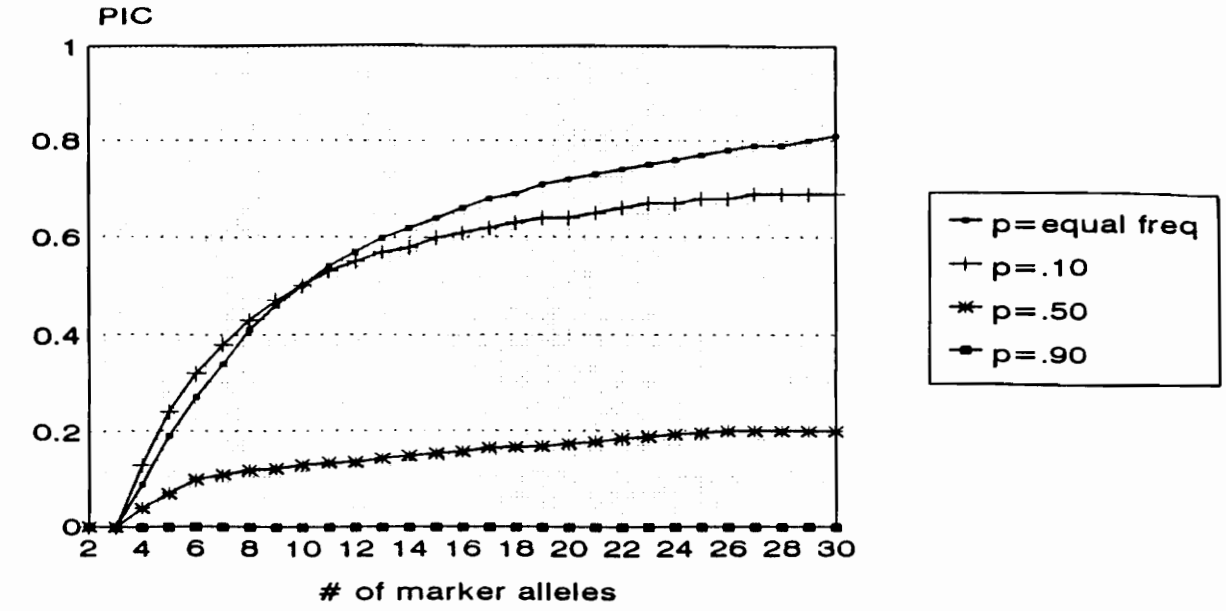

Figure 27

PIC within familyin the third generation for different marker allelic frequencies and mates not genotyped, as a function of the number of marker alleles.

Pic is computed given that an heterozygous sire is found. p denotes the frequency of the marker alleles present in the heterozygous sire. The remaining alleles are equally probable.

Table 1

Values of x1 and x3 for computing a bull's PTA

# daug.	x1	x3	# daug.	x1	x3
0	1.000	0.000	20	0.275	0.725
1	0.884	0.116	30	0.202	0.798
2	0.792	0.208	40	0.160	0.840
3	0.717	0.283	50	0.132	0.868
4	0.655	0.345	60	0.112	0.887
5	0.603	0.397	70	0.098	0.902
6	0.559	0.441	80	0.087	0.913
7	0.520	0.479	90	0.078	0.922
8	0.487	0.512	100	0.071	0.930
9	0.458	0.542	1000	0.007	0.992
10	0.432	0.568	5000	0.001	0.998
			10000	0.001	0.999

x1 and x3 are the weights for parent average and daughter yield deviation used to compute PTA. They vary with the number of daughters of the sire. It is assumed that all mates in the population are identified. Formulae used to compute x1 and x3 derived by Meinert (1990)

Table 2

Values of x1 and x3 for computing a cow's PTA

Information available	x1	x2	x3
PA, 3 records and no progeny	.545	.455	.000
PA, 5 records and no progeny	.419	.581	.000
PA, 5 records and 1 daughter	.396	.551	.052
PA, 5 records and 3 daughters	.359	.498	.142
PA, 5 records and 5 daughters	.328	.456	.216
PA, 5 records, 4 daughters and 6 sons	.270	.374	.354

x1, x2 and x3 are the weights for parent average, yield deviation and daughter yield deviation used to compute PTA. Weights vary with the number of own records and daughters of each cow. Formulae to compute x1, x2 and x3 derived by Meinert (1990) from data in Samuelson (1992)

Conclusions

This thesis provides selection indices that combine data on marked single genes and quantitative Animal Model information to increase genetic gain in livestock. Farmers and A.I. studs have a tool that allows them to make rational decisions about the application of marker information. A guidance for determining the usefulness of a particular genetic marker is also provided by this study.

The usefulness of marker information depends on several factors. Some of these factors are related to the type of marker information, e.g., size and accuracy of the estimated marker effect and recombination rate; some pertain to the moment in which the marker information is used (i.e., the generation of descendants), and other factors define the type and availability of the Animal Model information. All factors as well as the cost of marker genotyping (not considered in this study) determine the usefulness

of the marker information.

The criterion for evaluating the inclusion of marker information in this study has been the increase in accuracy achieved over the use of Animal Model information alone. In other words, the aim has been to assess the added value of the new technique of MAS over present techniques.

One important conclusion of this thesis is that the value of some particular marker information cannot be measured only by its effect (i.e., D_p). There is no guarantee that a marker of large effect is useful. Other factors need to be considered.

One of the factors to consider is the accuracy of the marker information. Marker information obtained from a small data set may not be of any use. It is likely to assume that markers of relatively small effect will not be found from small scale experiments. However, markers of large effect may be found from a relatively small number of animals. The value of this marker information will be severely reduced due to the poor quality of the estimation.

Another factor to consider is the availability and type of phenotypic information. Marker information applied to animals with a large number of records may be of no value. It is most likely to expect that only for young animals with no records (i.e., young

sires prior to progeny test) will breeders take advantage of marker information. Even when all the phenotypic information available is parent average, the quality of this information (i.e., the value of Rel(PA)) affect the usefulness of the marker information. However, it is likely to expect that Rel(PA) is relatively uniform for the young candidates.

It is well known that marker information is more valuable when the marker is closely linked to a QTL. The usefulness of marker information is reduced with an increase in recombination rate between marker and QTL. However, it is likely that only QTLs showing little recombination with the marker locus are able to be detected.

Having highly polymorphic marker loci is important. This is particularly true when marker information is used in later generations. Marker information that explains a large fraction of the Mendelian sampling variance may be of no value if it is based on a dimorphic marker, especially when used in generations beyond the first. Modestly polymorphic markers increase the need for genotyping mates. For instance, dimorphic markers lose all their information in the second generation if mates are not genotyped.

The population frequencies of the marker alleles present in the sire are also a factor to consider. Alleles present in the population at extreme frequencies reduce the probability of finding

heterozygous sires. However, if a sire carrying a rare marker allele (e.g., a mutation) is found, the inheritance of such allele among the sires's descendants will be facilitated.

The practical use of marker information is dependent upon the relationship between the monetary increase in genetic gain and the cost of genotyping. Although no economic analysis was considered in this study, the availability of inexpensive marker information may encourage its utilization.

In summary, the most important feature of this thesis is to provide the industry with a tool for deciding on the use of a particular genetic marker by comparing the expected genetic gain it produces with its economic cost.

Only marker information with additive effect was considered in this study. If dominance and epistasis are present at marked QTLs, the estimated increases in accuracy must be revised.

Selection indices were developed by combining Animal Model information with a single marked QTL. If more than one marked QTL is available, their inclusion in a selection index is straightforward as long as they are independent. For instance, marker information from QTLs allocated in different chromosomes can be treated as independent segregation terms. These segregation terms might covary only with PTA.

Finally, the most important feature of these indices is that they provide the industry not only with a tool for using the available marker information to evaluate genetic merit, but also a framework to make rational decisions about its use.

References

1. Beckmann, J. S. 1988. Oligonucleotide polymorphisms: a new tool for genomic selection. Biotechnology, 6:1061.

2. Beckmann, J. S., Y. Kashi, E. Hallerman, A. Nave and M. Soller. 1986. Restriction fragment length polymorphism among Israeli Holstein-Friesian dairy bulls. Animal Genetics, 17:25.

3. Beckmann, J. S. and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor. Appl. Genet., 67:35

4. Beckmann, J. S. and M. Soller. 1987. Molecular markers in the genetic improvement of farm animals. Bio/Technology 5:573.

5. Bishop, M. D. and J. A. Woolliams. 1991. Utilization of the sex-determining region Y gene in beef cattle breeding schemes. Anim. Prod., 53:157.

6. Bishop, M. D., R. Fries, R. T. Stone, J. W. Keele, S. F. L. Sunden and C. W. Beattie. 1992. Bovine gene mapping: definitions, approaches and consequences. 84th Annual Meeting of the American Society of Animal Science, J. Anim. Sci., 70, suppl. 1, 42.

7. Botstein, D., R.L. White, M. Skolnick, and R. Davi. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314.

8. Brascamp, E.W., J.A.M. van Arendonk and A.F. Groen. 1992. Economic appraisal of the utilization of genetic markers in dairy cattle breeding. Presented at the American Dairy Science Association Annual Meeting, Columbus, Ohio.

9. Bulmer, M. G. 1971. The effect of selection on natural variability. Am. Nat., 105:201.

10. Cantet, R. J. C. and C. Smith. 1991. Reduced animal model for marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. (in press).

11. Cochran, W. G. 1951. Improvements by means of selection. Proc. Second Berkeley Symp. Math. Stat. and Prob., 449.

12. Cowan, C. M., M. R. Dentine, R. L. Ax and L. A. Schuler. 1990. Structural variation among prolactin genes linked to quantitative traits in an elite holstein sire family. Theor. Appl. Genet., 79:577.

13. Cowan, C. M., T. Coyle and M. A. Dunkel. 1992a. Confirmation of quantitative trait differences identified by a bovine prolactin cDNA probe. J. Dairy Sci., 75 (Suppl.1): 286

14. Cowan, C. M., M.R. Dentine and T. Coyle. 1992b. Chromosome substitution effects associated with K-casein and B-lactoglobulin in Holstein cattle. J. Dairy. Sci. 75:1097.

15. Crittenden, L. B., I. Levin, L. Santangelo and J. Dogson. 1992. A population designed for molecular genetic mapping of the chicken genome. Proc 81st Annual Meeting of the Poultry Science Association. Poultry Sci., 71, suppl. 1, 57.

16. Dean, R., L. Cabinis and C. Carver. 1992. Cloning and sequencing of the bovine ferrochelatase gene. 84th Annual Meeting of the American Society of Animal Science, J. Anim. Sci., 70, suppl. 1, 149.

17. Dekkers, J.C.M. 1992a. Asymptotic response to selection on best linear unbiased predictors of breeding values. Anim. Prod., 54:351.

18. Dekkers, J.C.M. 1992b. Structure of breeding programs to capitalize on reproductive technology for genetic improvement. J. Dairy Sci. (subm.).

19. Dekkers, J. C. and M. R. Dentine. 1991. Quantitative genetics variance associated with chromosomal markers in segregating populations. Theor. Appl. Genet., 81:212.

20. Dentine, M.R. 1990. Using molecular biology to improve the accuracy of selection. Proc. of the IVth World Congress on Genet Appl to Livestock Prod., Edinburgh, 14:35.

21. Dentine, M.R. and C. M. Cowan. 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.

22. Dunnington, E. A., A. Haberfeld, L. C. Stallard, P. B. Siegel and J. Hillel. 1992. Deoxyribonucleic acid fingerprint bands linked to loci coding for quantitative traits in chickens. Poultry Sci., 71:1251.

23. Ellegren, H., L. Anderson, M. Johansson, and K. Sandberg. 1992. DNA fingerprinting in horses using a sample (TG)n probe and its application to population comparisons. Animal Genetics, 23:1.

24. Falconer. D. S. 1989. Introduction to quantitative genetics. 3rd. edition, John Wiley and Sons, New York.

25. Fernando, R. L. 1990. Statistical problems in marker assisted selection for QTL. Proc. 4th World Congress in Genetics Applied to Livestock Production, 13:436.

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

27. Foster, D. W. and L. K. Foster. 1991. Cloning and sequence analysis of the common a-subunit complementary deoxyribonucleic acid of turkey pituitary glycoprotein hormones. Poultry Sci., 70:2516.

28. Fries, R., J. S. Beckmann, M. Georges, M. Soller, and J. Womack. 1989. The bovine gene map. Animal Genetics, 20:3.

29. Gelderman, H. 1975. Investigations on inheritance of quantitative characters in animals by gene markers. I Methods. Theor. Appl. Genet., 46:319.

30. Genetic Visions. Brown Swiss Bulletin. 1991. September issue, page 7.

31. Genmark. Holstein World. July and August, 1992.

32. Georges, M., A.-S. Lequarre, M. Castelli, R. Hanset, and G. Vassart. 1988. DNA fingerprinting in domestic animals using four different minisatellite probes. Cytogenet. Cell Genet., 47:127.

33. Georges, M. and J. M. Massey. 1991. Velogenetics, or the synergistic use of marker assisted selection and germ-line manipulation. Theriogenology, 35:151.

34. Georges, M. A. J., M. Lathrop, Y. Bouquet, P. Hilbert, A. Marcotte, A. Schwers, J. Roupain, G. Vassart, and R. Hanset. 1990. Linkage relationships among 20 genetic markers in cattle. Evidence for linkage between two pairs of blood group systems: B-Z and S-F/V respectively. Animal Genetics, 21:95.

35. Georges, M. A. J., A. Gunawardana, D. W. Threadgill, M. Lathrop, I. Olsaker, A. Mishra, L. L. Sargeant, A. Schoeberlein, M. R. Steele, C. Terry, D. Threadgill, X. Zhao, T. Holm, R. Fries and J. E. Womack. 1991. Characterization of a set of variable number of tandem repeat markers conserved in bovidae. Genomics, 11:24.

36. Gibson, J. P. 1992. Using information on individual genes. 83rd Annual Meeting of the American Society of Animal science. J. Anim. Sci., 69, sup 1, 217.

37. Gibson, J. P. and J. Jansen. 1990. The effect of direct selection on a single gene on overall response to selection. Annual research report of the Centre for Genetic Improvement of Livestock, Univ. of Guelph, Guelph, Ontario, Canada, p.39.

38. Gilbert, D. A., N. Lehman, S. J. O'Brien, and R. K. Wayne. 1990. Genetic fingerprinting reflects population differentiation in the California island fox. Nature, 344:764.

39. Gill, P., A. J. Jeffreys and D. J. Warret. 1985. Forensic application of DNA "fingerprints". Nature, 318:577.

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

41. Goddard, C. and J. M. Boswell. 1991. Molecular biology and the growth of poultry. Crit. Rev. Poultry Biol., 3:325.

42. Gomez Raya, L. and J.P. Gibson. 1991. Genotype selection of young bulls at the embryo stage to increase the frequency of allele B of K-Casein in dairy cattle. The Centre for Genetic Improvement of Livestock. Univ. of Guelph., Guelph, Ontario, Canada. Annual research report. page 12.

43. Green, R. D., N. E. Ruggli-Cockett, S. J. Brinks and R. L. Henning. 1992. Association of the bovine major histocompatability system with various measures of immune function in beef cattle. 84th annual meeting of the American Society of Animal Science, J. Anim. Sci., 70, suppl. 1, 41.

44. Groen, A.F. 1991. A stochastic simulation study on the influence of inbreeding on DNA-fingerprint characteristics. 42nd Annual Meeting of the EAAP, September 1991, Berlin, Germany.

45. Guillemot, F. and C. Juffrey. 1989. Molecular biology of the chicken major histocompatability complex. Crit. Rev. Poultry Biol., 2:251.

46. Haberfeld, A., E. A. Dunnington, and P. B. Siegel. 1992. Genetic distances estimated from DNA fingerprinting in crosses of white plymouth rock chickens. Animal Genetics, 23:163.

47. Hallerman, E. M. 1989. Genetic improvement of livestock through utilization of DNA-level markers. Int. J. Anim. Sci., 4:60.

48. Hallerman, E. M., Y. Kashi, A. Nave, J. S. Beckmann and M. Soller. 1986. Restriction fragment length polymorphisms in dairy cattle and their utilization for genetic improvement. World Rev. of Anim. Prod., 22:31.

49. Hallerman, E. M., J. L. Theilmann, J. S. Beckmann, M. Soller and J. E. Womack. 1988a. Mapping of bovine prolactin and rhodopsin genes in hybrid somatic cells. Anim. Genet., 19:123.

50. Hallerman, E. M., A. Nave, M. Soller and J.S. Beckmann. 1988b. Screening of Israeli Holstein-Friesian cattle for restriction fragment length polymorphisms using homologous and heterologous deoxyribonucleic acid probes. J. Dairy Sci., 71:3378.

51. Hill, W. G. and P. D. Keightley. 1988. Interaction between molecular and quantitative genetics. In Adv. in Animal Breeding, Kuver et al., eds., Wageningen. The Netherlands.

52. Hillel, J., Y. Plotzy, A. Haberfeld, U. Lavi, A. Cahaner and A.J. Jeffreys. 1989. DNA fingerprints of poultry. Anim. Genet. 20:145.

53. Hillel, J., T. Schaap, A. Haberfeld, A. Jeffreys, Y. Plotzky, A. Cahaner and U. Lavi. 1990. DNA fingerprints applied to gene introgression in breeding programs. Genetics, 124: 783.

54. Hoeschele, I. 1993. Absorption of quantitative trait loci equations except for genotyped animals and tie ancestors. J. Dairy Sci. (in press).

55. Hoeschele, I. and T. Meinert. 1990. Association of genetic defects with yield and type traits: The Weaver Locus Effect on Yield. J. Dairy Sci., 73: 2503.

56. Hoeschele, I. and P. M. Van Raden. 1993. Bayesian analysis of linkage between genetic markers and quantitative trait loci. Theo. Appl. Gen., (in press).

57. Hyland, F. C. and R. L. Quaas. 1991. Benefits from marker-assisted selection under varying assumptions. J. Dairy Sci. 74, Suppl. 1: 2503.

58. Jacquard, A. 1970. The genetic structure of populations. Springler-Verlag, Berlin

59. James, J. N. 1991. Methods of analysis of associations between genetic markers and quantitative trait loci with special reference to cattle. Proc. 42nd Animal Meeting of the EAAP, Berlin, Germany.

60. Jayarao, B.M., S. P. Oliver, J. R. Tagg and K. R. Matthews. 1991. Genomic DNA restriction endonuclease. J. Dairy Sci. 74, suppl. 1, 294.

61. Jayarao, B.M., J. J. E. Dore Jr., and S. P. Oliver. 1992. 16 S rDNA fingerprinting: a new approach for the identification of Streptococcus and Enterococcus species. J. Dairy Sci. 75, suppl. 1, 308.

62. Jeffreys, A. J. 1985. DNA sequence variants in the C, A, alpha, and beta-globulin genes of man. Cell 18:1.

63. Jeffreys, A. J. 1987. Highly variable minisatellites and DNA fingerprints. Biochem. Soc. Trans., 15:309.

64. Jeffreys, A. J., V. Wilson and S. L. Thein. 1985a. Hypervariable 'minisatellite' regions in human DNA. Nature, 314:67.

65. Jeffreys, A. J., J. F. Y. Brookfield, and R. Semeonoff. 1985b. Positive identification of an immigration test-case using human DNA fingerprints. Nature, 317:818.

66. Jensen, J. 1989. Estimation of recombination parameters between a quantitative trait loci (QTL) and two marker gene loci. Theor. Appl. Genet., 78:613.

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

68. Kashi, Y., E. Lipkin, A. Darvasi, A. Nave, Y. Gruenbaum, J. S. Beckmann and M. Soller. 1990b. Parentage identification in the bovine using "deoxyribonucleic acid fingerprints". J. Dairy Sci., 73:3306.

69. Kennedy, B. W., A. M. Verrinder Gibbins, J. P. Gibson and C. Smith. 1990. Coalescence of molecular and quantitative genetics for livestock improvement. J. Dairy Sci., 73:2619.

70. Kennedy, B. W., Quinton, M. and J. A. M. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. J. Anim. Sci., 70:2000.

71. Kirby, L. T. 1990. DNA Fingerprints: An Introduction. Stockton Press, New York. 365 pages.

72. Kirkpatrick, B. W. 1992. Marker development by analysis of observed microsatellites. 84th Annual Meeting of the American Society of Animal Science, J. Anim. Sci., 70, suppl. 1, 143.

73. Kirkpatrick, B. W., C. W. Cowan and M. R. Dentine. 1990. Rapid detection of a restriction fragment length polymorphism at the growth hormone in Holstein bulls. Proc. 85th American Dairy Science Association. Annual Meeting. J. Dairy Sci., 73, suppl. 1, 252.

74. Knapp, S. 1992. Mapping genes underlying quantitative traits advances and methods for model organisms. J. Dairy Sci. 75, suppl. 1, 185.

75. Knox, D. and E. Verspoen. 1991. A mitochondrial DNA restriction fragment length polymorphism of potential use for discrimination of farmed Norwegian and wild Atlantic salmon populations in Scotland. Aquaculture, 98:249.

76. Kuhnlein, U., D. Zadworny, Y. Dawe, R.W. Fairfull and J.S Gavora. 1990. Assessment of inbreeding by DNA fingerprinting: development of a calibration curve using defined strains of chickens. Genetics, 125:161.

77. Lande, R. and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, 124:743.

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

79. Lanneluc, I., F. Hospital, C. Chevalet, J. M. Elsen and J. Gellin. 1992. Genetic analysis of fingerprints in Merinos d'Arles x Booroola Merino crossbred sheep. Animal Genetics, 23:239.

80. Laster, D. B. and C. W. Beattie. 1992. Issues affecting cattle genome mapping and plans for future research. Proc. American Dairy Science Association Annual Meeting. J. Dairy Sci. 75, suppl. 1, 185.

81. Li, S. Y., D. Zadworny and U. Kuhnlein. 1992. Establishment of an inbreeding index in Holstein dairy cattle using DNA fingerprints. J. Dairy Sci., 75, suppl. 1, 283.

82. Litt, M. and J. A. Luty, 1989. A hypervariable minisatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle active gene. Am. J. Hum. Genet., 44:397.

83. Lohuis, M. M., J. C. M. Dekkers and C. Smith. 1992. Probability of success and predicted returns of sires in progeny test programs. J. Dairy Sci., 75:1660.

84. Medrano, J. F. and T. R. Famula. 1990. Milk protein genotype in Holstein sires and genetic merit for production traits. Proc 85th Annual Meeting of the American Dairy Science Association, J. Dairy Sci., 73, suppl. 1, 252.

85. Meinert, T. 1991. Factors affecting the accuracy and stability of sire proofs from progeny test herds. Ph. D. dissertation. Virginia Polytechnic and State University, Blacksburg.

86. Meuwissen, T. H. E. and J. A. M. van Arendonk. 1992. Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. J. Dairy Sci., 75:1651.

87. Nicholas, F. W. and C. Smith. 1983. Increased rates of genetic change in dairy cattle by embryo transfer and splitting. Anim. Prod., 36:341.

88. Nichols, R. A. and D. J. Baldy. 1990. Effects of population structure in DNA fingerprint analysis in forensic science. Heredity, 66:297.

89. Nienhuis, J. and T. Helentjaris. 1989. Simultaneous selection for multiple polygenic traits through RFLP analysis. In T. Helentjaris and B. Burr, eds. Development and application of molecular problems in plant genetics. Cold Spring Harbour, New York.

90. Orkin, S. 1986. Reverse genetics and human disease. Cell, 47:845.

91. Paterson, A. H., J. W. De Verna, B. Lanini and S. D. Tanksley. 1990. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics, 124:735.

92. Plotzky, Y., A. Cahaner, A. Haberfeld, U. Lavi and J. Hillel. 1990. Analysis of genetic associations between DNA fingerprint bands and quantitative traits using DNA mixes. Proc. of the 4th World Congress in Genetics Applied to Livestock Production., 13:133.

93. Ploughman, L. M. and M. Boehnke. 1989. Estimating the power of a proposed linkage study for a complex genetic trait. Am. J. Hum. Genet., 44:543.

94. Rohrer, G. A., and C. W. Beattie. 1992. Mapping the porcine genome: An experimental approach. 84th Annual Meeting of the American Society of Animal Science, J. Anim. Sci., 70, suppl. 1, 41.

95. Sabour, M. P., C. Y. Lin and A. J. Lee. 1992. Effects of selection on the frequency of milk protein genotypes in Canadian AI bulls. J. Dairy Sci., 75, suppl. 1, 285.

96. Saeffundin, A. and J.P. Gibson. 1991. Selection response in populations with a transgene. 83th Annual Meeting of the American Society of Animal Science., J. Anim. Sci., 69, suppl. 1.

97. St. George-Hyslop, P. H., J. L. Haines, L. A. Farrer, R. Polinsky, C. Van Broeckhoven and other members of the F.A.D. Collaborative Study Group. 1990. Genetic linkage studies suggest that Alzheimer's disease is not a single homogeneous disorder. Nature, 347:194.

98. Scheid, P. J., B. A. Didion and M. A. Johns. 1992. Mapping molecular markers in the bovine genome. 84th Annual Meeting of the American Society of Animal Science., J. Anim. Sci., 70, suppl. 1, 149.

99. Seyoum, S. and I. Kornfield. 1992. Identification of the subspecies of Oreochromis niloticus (Pisces: Cichlidae) using restriction endonuclease analysis of mitochondrial DNA. Aquaculture, 102:29.

100. Shay, T. L., N. E. Mussli-Cockett, A. Maciulis and T. D. Bunch. 1992. Identification of a genetic marker for the spider lamb syndrome gene using RFLP analysis. 84th Annual Meeting of the American Society of Animal Science., J. Anim. Sci., 70, suppl. 1, 95.

101. Shields, B. A., A. R. Kapuscinski and K. S. Guise. 1992. Mitochondrial DNA variation in four Minnesota populations of lake whitefish: Utility as species and population markers. Trans. Am. Fish. Soc., 121:21.

102. Shuster, D. E., M. E. Kehrli (Jr)., R. Gonzalez, G. Rogers, J. Cullor and R. Gilbert. 1992. Molecular basis and prevalence of bovine leukocyte adhesion defiency (BLAD) among Holstein cattle. Proc. 87th Annual Meeting of the American Dairy Science Association., J. Dairy Sci., 75, suppl. 1, 286.

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

104. Sing, C. F. and P. P. Moll. 1990. Genetics of atherosclerosis. Annu. Rev. Genet., 24:171.

105. Smith, C. 1991. Possible uses of genetic markers in selection programs in cattle. Proc. 42nd Animal Meeting of the EAAP, Berlin, Germany.

106. Smith, C. and S. P. Simpson. 1986. The use of genetic polymorphisms in livestock improvement. J. Anim. Breed. Genet., 103:203.

107. Smith, C. and P. R. Bampton. 1992. Inheritance of reaction to halothane anaesthesia in pigs. Genet. Res., 29:287.

108. Smith, H. O. 1979. Nucleotide sequence specificity of restriction endonucleases. Science, 205:455.

109. Smith, O. S., J. S. C. Smith, S. L. Bowen, R. A. Tenberg and S. J. Wall. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F_1, grain yields, heterosis and RFLPs. Theor. Appl. Genet., 67: 25:33.

110. Sniper, K. P., R. W. Kaster, M. A. Wild, M. W. Miller, D. A. Jessup, R. L. Silflow, W. J. Foreyt and T. E. Carpenter. 1992. Using ribosomal RNA gene restiction patterns in distinguishing isolates of Pasteurella haemolytica for bighorn sheep (Ovis canadiensis). J. Wildlife Diseases, 28:347.

111. Soller, M. 1978. The use of loci associated with quantitative traits in dairy cattle improvement. Anim. Prod., 27:133.

112. Soller, M. 1990. Genetic mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. J. Dairy. Sci., 73:2628.

113. Soller, M., T. Brody and A. Genizi. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet., 47:35.

114. Soller, M. and A. Genizi. 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics, 34:47.

115. Soller, M. and J. S. Beckman. 1982. Restriction fragment length polymorphisms and genetic improvement. Proc. of the 2nd World Congr. Genet. Appl. Livest. Prod., Madrid.

116. Soller, M. and J. S. Beckman. 1983. Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. Theor. Appl. Genet., 76:228.

117. Stam, P. 1986. The use of marker loci in selection for quantitative characters. In: Smith, C., King, J.N.B., and McKay (eds). Exploiting new technologies in animal breeding genetic developments. Oxford Univ. Press, Oxford, page 170.

118. Stam, P. 1987. Marker genes in selection. Biochemical polymorphisms as markers in selection for quantitative traits. Anim. Genet., 18, suppl. 1, 97.

119. Steele, M. R. and M. Georges. 1991. Generation of bovine multisite haplotypes using random cosmid clones. Unpublished.

120. Steel, R. G. D. and J. H. Torrie. 1980. Principles and procedures of statistics. McGraw-Hill. New York.

121. Stryer, L. 1988. Biochemistry. 3rd ed., W.H.Freeman and Co., New York.

122. Sunden, S. F. L., A. M. Crawford, A. M., F. C. Buchanan, P. A. Swarbrick, R. T. Stone, J. W. Keele, M. D. Bishop and C. W. Beattie. 1992. Utility of sheep microsatellite markers for bovine gene mapping. 84th Annual Meeting of the American Society of Animal Science., J. Anim. Sci., 70, suppl. 1, 143.

123. Tuggle, C. K., J. M. Helm, and M. Rothschild. 1992. Identification and cloning of a swine growth hormone regulator gene. 84th Annual Meeting of the American Society of Animal Science, J. Anim. Sci., 70, suppl. 1, 41.

124. van Arendonk, J. A. M. and S. van der Beek. 1991. Estimation of recombination rates between markers from segregating populations. In: Proc. 9th Conference of Anim. Breed. and Genet., Melbourne, Australia.

125. Van Raden, P. M. and G. R. Wiggans. 1991. Derivation, calculation and use of national animal model information. J. Dairy Sci., 74, 2737.

126. Van Vleck, L. D. 1988. Notes on the theory and application of selection principles for the genetic improvement of animals. Cornell University, Ithaca, New York.

127. 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.

128. Weller, J.I., Y. Kashi, and M. Soller. 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.

129. White, R., J. M. Lalouel, P. O'Connell, Y. Nakamura, M. Leppert, and M. Lathrop. 1987. Current status in mapping the human genome: 440 RFLPs in 59 families and 690 new RFLPs. Cytogenet. Cell Genet., 46, 715.

130. Winkelman, D. C. and S. M. Schmutz. 1992. Muscle protein genes tested as candidate QTLs for carcass traits in beef cattle. 84th Annual Meeting of the American Society of Animal Science., J. Anim. Sci., 70, suppl. 1, 144.

131. Womack, J.E. and P. Moll. 1986. Gene map of the cow: conservation of linkage with mouse and man. J. Heredity, 77, 2.

132. Woolliams, J.A. and C. Smith. 1988. The value of indicator traits in the genetic improvement of dairy cattle. Anim. Prod., 46:333

133. Xiong, S., R. L. Park, R. P. Evans, R. W. Andersen and D. J. Fairbanks. 1992. RAPD analysis uncovers a male-specific marker in mink. 84th Annual Meeting of the American Society of Animal Science., J. Anim. Sci., 70, suppl. 1, 95.

134. Yemm, R. S., D. L. Kahrt, D. M. Kniffen, A. M. Swinker, and K. L. Hossner. 1992. DNA fingerprinting of sheep and horses for paternity testing using microsatellite probe (GTG)s. 84th Annual Meeting of the American Society of Animal Science., J. of Anim. Sci., 70, suppl. 1, 144

135. Zhang, W. and C. Smith. 1992. Computer simulation of marker assisted selection utilizing linkage disequilibrium. The Centre for Genetic Improvement of Livestock. Univ. of Guelph, Guelph, Ontario, Canada. Annual research report.

136. Zhang, W., D. L. Kuhlers and W. E. Manfield. 1992a. Halothane gene and swine performance. J. Anim. Sci., 70, 1307.

137. Zhang, H. M., S. K. DeNise, K. C. Maddock, M.E. Bellin and R. L. Ax. 1992b. Bovine somatotrophin (bST) gene polymorphism in Hereford bulls. 84th Annual Meeting of the American Society of Animal Science., J. Anim. Sci., 70, suppl. 1, 144.

138. Zawadski, S. M. and P. Johnson. 1992. The effect of the halothane gene on pork production and meat quality of pigs under commercial conditions. Can. J. Anim. Sci., 1, 959.

Appendix A1. An estimator of Var(S_g) when D_g is not known

Although $D_p = \hat{D}_g$ = BLUE (D_g) is an unbiased estimate of D_g, D^2_p is not an unbiased estimator of D^2_g, because

$$E(D_p^2) = E(\hat{D}_g^2) = D_g^2 + [SE(D_g)]^2$$

Hence, a possible estimator might be obtained by substituting $\hat{D}_g^2$ for $E(\hat{D}_g^2)$ and solving for D^2_g, or

$$\tilde{D}_g^2 = \hat{D}_g^2 - [SE(\hat{D}_g)]^2$$

Appendix A2. Selection index examples

A2.1. First generation

Marker and Animal Model information is combined in a selection index to estimate the genetic merit of a cow with one record and no progeny. Sources of information in the selection index are PTA = 600 lbs; $D_p = \pm 400$ lbs, which are the observed segregation effects ($\pm\frac{1}{2} D_p$) expressed in the transmitting ability scale; and $SE(D_p) = 100$ lbs. Parameter values used to compute index weights are Rel(PA) = .3725, which assumes that Rel(PTA) for the sire and dam of the cow are .99 and .5 respectively; Rel(PTA) for the cow with one record is .4815; the weights used for computing PTA, estimated as ratio of daughter equivalents, are $x_1 = .6385$ and $x_2 = .3615$ (Van Raden and Wiggans, 1991);

Selection index weights are computed from **[12]**. Then,

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sigma_a^2 Rel(PTA) & x_2 (\frac{1}{4})^2 (D_p^2 - SE^2(D_p)) \\ sym & (\frac{1}{4} D_p)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_a^2 Rel(PTA) \\ (\frac{1}{4})^2 (D_p^2 - SE^2(D_p)) \end{bmatrix}$$

$$= \begin{bmatrix} 1250^2 (.4815) & (.3615)(\frac{1}{16})(400^2 - 100^2) \\ sym & (\frac{1}{4} 400)^2 \end{bmatrix}^{-1} \begin{bmatrix} 1250^2 (.4815) \\ (\frac{1}{16})(400^2 - 100^2) \end{bmatrix}$$

$$= \begin{bmatrix} .9973 \\ .5995 \end{bmatrix}$$

and

$$I = b_1 \text{ PTA} \pm b_2 S_p$$

$$I = b_1 \text{ PTA} \pm b_2 .25 D_p$$

I = 658.33 lbs, if the cow received the favorable marker allele, and

I = 538.43 lbs, if the cow received the unfavorable marker allele.

The unadjusted reliability of the index (Rel(I)) is .581. Reliabilities and variances are adjusted for selection by computing [23], or

$$\sigma_d^{2*} = \sigma_a^2 \frac{(1-k_{sd})\,[k_{ss}(1-r_{ss}^2) + k_{ds}(1-r_{ds}^2) + 2] + (3+k_{ss})\,[k_{sd}(1-r_{sd}^2) + k_{dd}(1-r_{dd}^2) + 2]}{(3+k_{ss})\,(3+k_{dd}) - (1-k_{ds})\,(1-k_{sd})}$$

The equilibrium variance is computed from the parameters in box

Box. Parameters for a progeny tested population

	Path of Selection			
	ss	**sd**	**ds**	**dd**
% selected	.04	.20	.06	.90
i_{xy}	2.15	1.40	1.98	.195
t_{xy}	1.76	.84	1.56	-2.33
$r^2_{xy}=Rel_{xy}$	.72	.72	.42	.42

σ_d^{2*} = 1,190,749.3694 lbs^2 = $.763\sigma_a^2$

Rel(PTA) is adjusted for selection by using $1-\frac{\sigma_a^2}{\sigma_d^{2*}}(1-Rel_{PTA})$.

Rel(I) at the equilibrium is computed as in **[26]**. After adjusting for selection, Rel(PTA) and Rel(I) reduce to .319, and .339, which implies a gain in reliability of 6.10%. The gain in the accuracy scale (PGA) is 3%.

A2.2. Second generation

The example in **A2.1.** is now slightly altered: the cow is now the granddaughter of the grandsire for whom the marker information was obtained. Sources of information are PTA = 600 lbs, $S_{p(S)}$ = ± 1/8 D_p, and $S_{p(O)}$ = ± 1/8 (1-2r) D_p. Parameters are like in **A2.1**, only with the addition of assuming a recombination rate (r) equal .1, and the average additive effect of the QTL genes in the dams equals the average effect of the QTL

alleles in the grandsire. From **[62]**, the index weights are computed as

$$[b_1\ b_2\ b_3]' =$$

$$\begin{bmatrix} \sigma_a^2 Rel(PTA) & x_3^*(\frac{1}{8})^2(D_p^2-SE^2(D_p)) & (x_2+x_3)(\frac{1}{8})^2(1-2r)^2(D_p^2-SE^2(D_p)) \\ & (\frac{1}{8})^2 D_p^2 & 0 \\ sym. & & (\frac{1}{8})^2(1-2r)^2 D_p^2 \end{bmatrix}^{-1}$$

$$x \begin{bmatrix} \sigma_a^2 Rel(PTA) \\ (\frac{1}{8})^2(D_p^2-SE^2(D_p)) \\ (\frac{1}{8})^2(1-2r)^2(D_p^2-SE^2(D_p)) \end{bmatrix}$$

$$= \begin{bmatrix} 1250^2(.4815) & (.9)(\frac{1}{8})^2(400^2-100^2) & (.3615)(\frac{1}{8}.8)^2(400^2-100^2) \\ & (\frac{1}{8})^2 400^2 & 0 \\ sym. & & (\frac{1}{8})^2(.8)^2 400^2 \end{bmatrix}^{-1}$$

$$x \begin{bmatrix} 1250^2(.4815) \\ (\frac{1}{8})^2(400^2-100^2) \\ (\frac{1}{8})^2(.8)^2(400^2-100^2) \end{bmatrix}$$

where $x_3^* = .9$ is the weight for DYD in the PTA of the sire (S1 or S2 in Figure 4), which is assumed to be a proven bull. The weights x_2 and x_3 pertain to the PTA of the grandoffspring (O1-O4

$$= \begin{bmatrix} .999 \\ .094 \\ .599 \end{bmatrix}$$

in Figure 4) or the cow of interest here.

The index is then computed as

$$I = b_1 \text{ PTA} + b_2 S_{p(S)} + b_3 S_{p(O)}$$

$$I = (.999)\ 600 \pm .094\ (50) \pm .599\ (40)$$

$I = 628.252$, when both sire and granddaughter received the favorable marker allele M_1 from the grandsire,

$I = 580.346$, when the sire received M_1 and the granddaughter received a marker allele from the granddam,

$I = 618.816$, when the sire received M_2 and granddaughter received a marker allele from the granddam,

$I = 570.912$, when both sire and granddaughter received the unfavorable marker allele from the grandsire.

After adjusting for selection, Rel(PTA) and Rel(I) reduce to .319, and .327, which implies a relative gain in reliability of 2.56%. The gain in the accuracy scale (PGA) is 1.2%.

A2.3. Third generation

The example in **A2.2** is applied here where the cow is the

great-granddaughter of the great-grandsire for whom the marker information was obtained. Sources of information are PTA = 600 lbs, $S_{p(GS)} = \pm 1/16\ D_p$, $S_{p(S)} = \pm 1/16\ (1-2r)D_p$, and $S_{p(O)} = \pm 1/8(1-2r)(1-r)D_p$ or $\pm 1/8(1-2r)rD_p$ according to the family they pertain. Recombination rate is again known (r=.1), and the average additive effect of the QTL alleles in the granddams and great-granddams is assumed equal to the average effect of the alleles in the grandsire. The selection index weights are computed, for the **O1-O2** as

$$[b_1\ b_2\ b_3\ b_4]' =$$

$$\begin{bmatrix} \sigma^2_a Rel(PTA) & x_3^*(\frac{1}{16})^2(D_p^2-SE^2(D_p)) & x_3^*(\frac{1}{16})^2(1-2r)^2(D_p^2-SE^2(D_p)) & (x_2+x_3)(\frac{1}{8})^2(1-2r)^2(1-r)^2(D_p^2-SE^2(D_p)) \\ & (\frac{1}{16})^2 D_p^2 & 0 & 0 \\ & & (\frac{1}{16})^2(1-2r)^2 D_p^2 & 0 \\ sym. & & & (\frac{1}{8})^2(1-2r)^2(1-r)^2 D_p^2 \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \sigma^2_a Rel(PTA) \\ (\frac{1}{16})^2(D_p^2-SE^2(D_p)) \\ (\frac{1}{16})^2(1-2r)^2(D_p^2-SE^2(D_p)) \\ (\frac{1}{8})^2(1-2r)^2(1-r)^2(D_p^2-SE^2(D_p)) \end{bmatrix}$$

This system of equations can also be used to compute the index weights for the **O3-O4** family, by replacing the term (1-r) by r.

The index weights in this example for the **O1-O2** family are

$$[.999 \quad .094 \quad .094 \quad .599]'$$

and for the **O3-O4** family,

$$[.999 \quad .093 \quad .093 \quad .598]'$$

Hence, the selection index

$$I = b_1 \text{ PTA} + b_2 S_{p(GS)} + b_3 S_{p(S)} + b_3 S_{p(O)}$$

becomes for the **O1-O2** family

$$I = (.999)\ 600 \pm .094\ (25) \pm .094\ (20) \pm .599\ (36)$$

and for the **O3-O4** family

$$I = (.999)\ 600 \pm .093\ (25) \pm .093\ (20) \pm .599\ (4)$$

For the **O1-O2** family, the index values are

$I = 625.194$ (when + + +),

$I = 582.066$ (when + + -),

$I = 621.434$ (when + - +),

$I = 578.306$ (when + - -),

$I = 620.494$ (when - + +),

$I = 577.366$ (when - + -),

$I = 616.734$ (when - - +),

$I = 573.606$ (when - - -)

where the signs inside the parenthesis indicate the type of

algebraic sum that is performed.

For the **O3-O4** family the index values are

I = 605.981 (when + + +),

I = 601.189 (when + + -),

I = 602.261 (when + - +),

I = 597.469 (when + - -),

I = 601.331 (when - + +),

I = 601.331 (when - + -),

I = 597.611 (when - - +),

I = 592.819 (when - - -)

After adjusting for selection, Rel(PTA) and Rel(I) reduce to .319, and .321, which implies a gain in reliability of 0.78%. The gain in the accuracy scale (PGA) is 0.35%.

Appendix A3. Effect of parent average reliability (Rel(PA)) on the usefulness of marker information.

Marker information is more usefull when there is little or no phenotypic information explains Mendelian sampling variance. In the extreme case, parent average (PA) is the only phenotypic source of information. If PA is estimated with a low reliability, the usefulness of marker information, measured as percentage of gain in accuracy (PGA), increases.

For the first generation, as in **[24]** and **[25]**,

$$PGA = \frac{r_{I_{adj}} - r_{PTA_{adj}}}{r_{PTA_{adj}}}$$

$$= \frac{\sqrt{\frac{Var(PTA_{adj}) + b_2^2 Var(S_p)}{Var(TA)_{adj}}} - \sqrt{\frac{Var(PTA)_{adj}}{Var(TA)_{adj}}}}{\sqrt{\frac{Var(PTA)_{adj}}{Var(TA)_{adj}}}}$$

$$= \sqrt{1 + b_2^2 \frac{Var(S_p)}{Var(PTA)_{adj}}} - 1$$

$$= \sqrt{1 + \frac{b_2^2 (\frac{1}{4} D_p)^2}{\frac{\sigma_d^{2*}}{4} (1 - \frac{\sigma_a^2}{\sigma_d^{2*}} (1 - Rel(PTA))}} - 1$$

Assuming D_p = 400 lbs and $SE(D_p)$ = 10 lbs, then b_2 = .999.

If σ_d^{2*} = 1,190749.37, then

If Rel(PA) = .286 , then $Rel(PA)_{adj}$ = .062 and PGA = 23.8%

If Rel(PA) = .375 , then $Rel(PA)_{adj}$ = .179 and PGA = 8.9%

Vita

Eduardo O. Romano was born September 5, 1955 in Buenos Aires, Argentina. He graduated as Agronomical Engineer from the University of Buenos Aires in 1982. He pursued a M.S. degree in biometry at the University of Buenos Aires/INTA Castelar from 1985 to 1987. In 1989 he become a graduate student in the Dairy Science department of Virginia Polytechnic Institute and State University.