The use of genetic markers in poultry breeding

Promotor: dr. ir. E.W. Brascamp
hoogleraar in de veefokkerij
Co-promotor: dr. ir. J.A.M. van Arendonk
universitair hoofddocent veefokkerij

The use of genetic markers in poultry breeding

Sijne van der Beek

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen
op gezag van de rector magnificus,
dr. C.M. Karssen,
in het openbaar de verdedigen
op vrijdag 9 februari 1996
des namiddags te vier uur in de Aula
van de Landbouwuniversiteit te Wageningen.

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Beek, Sijne van der

The use of genetic markers in poultry breeding / Sijne van der Beek - [S.l.:s.n.]
Thesis Landbouwuniversiteit Wageningen. - With ref. - With summary in Dutch.
ISBN:90-5485-474-X
Subject headings: genetic markers; poultry; selection.; genetics.

Abstract

Beek, S. van der 1995. The use of genetic markers in poultry breeding. Doctoral thesis, Wageningen Agricultural University, P.O. Box 338, 6700 AH Wageningen, The Netherlands.

This thesis focuses on the design of linkage and quantitative trait locus (QTL) mapping experiments and on the use of genetic markers in poultry breeding schemes. Criteria to optimize the design of experiments that use outbred populations to create a linkage map were described. Those criteria were used to evaluate designs of experiments to map codominant and dominant genetic markers using half-sib of full-sib family structures. Once a linkage map is created, QTLs can be placed on the map. Deterministic computation methods were used to determine the power of two- and three- generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. Genetic markers that are linked to QTL can be used for selection purposes. The effect of using genetic markers to assist selection in an outbred poultry breeding nucleus was studied. The additional response to selection for a sex-limited trait was computed using deterministic simulation. The general discussion of this thesis addresses the status of the chicken linkage map, QTLs found in poultry, and discusses the usefulness of several applications of genetic markers in poultry breeding.

Stellingen

1. Om in een fokprogramma optimaal gebruik te kunnen maken van merkerinformatie dient de structuur grondig gewijzigd te worden.
Dit proefschrift

2. Bij het beoordelen van proeven voor het localiseren van genen moet meer nadruk gelegd worden op de verwachte nauwkeurigheid en minder op de power van de proef.
Dit proefschrift

3. Allelen die het nivo van een commerciële populatie kunnen verbeteren, kunnen het beste gezocht worden in een kruising tussen deze populatie en een genetisch ongerelateerde populatie van gelijkwaardig nivo.
Dit proefschrift

4. Voor het localiseren van genen is een drie generatie full-sib offspring, half-sib grandoffspring proefopzet optimaal.
Dit proefschrift

5. De door Gibson (1994) gevonden verlaging van lange termijn response door het gebruik van merkerinformatie treedt alleen op bij slecht gebruik van merkerinformatie.
(*JP Gibson, 1994. Proc 5th World Congr Genet Appl Livestock Prod 21:201-204.*)

6. Om informatie over gelocaliseerde genen juist te integreren in fokwaardeschatting, het schatten van genetische parameters en de evaluatie van fokprogramma's, is onderzoek naar het vervangen van het infinitesimal model nodig.

7. Voor een duurzame bijdrage van de moleculaire genetica aan de veefokkerij is het van belang dat effectieve technologie wordt ontwikkeld voor het opsporen van nieuwe mutaties in populaties.

8. Het inwegen van kenmerken in het fokdoel zou gebaseerd moeten worden op gewenste genetische vooruitgangen, en niet op economische waarden.

9. Techneuten pikken uit de duurzaamheidsdiscussie nieuwe vragen op, terwijl zij zouden moeten leren om anders tot antwoorden te komen.

10. Het grootste gemis in academische opleidingen is het niet leren omgaan met emotionele argumenten.

11. Een proefschrift dat zichzelf tegenspreekt duidt op ontwikkeling van de promovendus.

12. Soaps reinigen de academische geest.

Sijne van der Beek. The use of genetic markers in poultry breeding. Wageningen, 9 februari 1996.

'Ontdekkingen doen bestaat uit zien wat iedereen heeft gezien en denken wat niemand heeft gedacht.'

Albert Szent-Györgyi

Contents

Chapter 1	Introduction	**1**
Chapter 2	Criteria to optimize designs for detection and estimation of linkage between marker loci from segregating populations containing several families	**7**
Chapter 3	Evaluation of designs for reference families for livestock linkage mapping experiments	**31**
Chapter 4	Power of two- and three- generation QTL mapping experiments in an outbred population containing full-sib or half-sib families	**51**
Chapter 5	Marker assisted selection in an outbred poultry breeding nucleus	**77**
Chapter 6	General discussion	**97**
Summary		**119**
Waar gaat dit proefschrift over ?		**125**
Levensloop		**133**

Introduction

A genetic marker is a genomic locus for which the allele(s) of an individual can be identified (see Box 1). Genetic markers can be a powerful tool in breeding and genetics. Several applications of genetic markers have been described (Soller and Beckmann 1983; Smith and Simpson 1986) including: parentage control, varietal identification, identification of loci affecting quantitative traits (i.e., the mapping of Quantitative Trait Loci or QTL mapping), and marker assisted selection during introgression, in a crossbred population or in an outbred population.

Box 1: A genetic marker

In this thesis a genetic marker is defined as a parameter that has a direct relation with alleles at a genomic site. Given this definition a morphological trait or a biochemical polymorphism can be a genetic marker. For example, blood group polymorphisms have a direct relation to alleles at a blood group gene, and can be considered as genetic markers.

With the advent of molecular genetics new classes of genetic markers have been made available. One class of genetic markers that has been shown to be particularly useful is the class of the microsatellites. A microsatellite is a short DNA sequence consisting of di- tri or tetra- tandem repeats, e.g., DNA sequence TGTGTGTGTGTGTGTGTGTG or $(TG)_{10}$ is a microsatellite. Microsatellites have several appealing properties:

- Microsatellites are highly polymorphic. Different alleles vary in the number of repeats. For instance, one allele is $(TG)_{10}$ and a second allele is $(TG)_{12}$. These differences can be made visible in polyacrylamide gels.
- The alleles of a microsatellite marker can relatively easy be identified in a DNA sample using PCR technology and polyacrylamide gel electrophoresis. Differences between alleles are made visible due to the differences in length between the different alleles.
- Microsatellites have co-dominant inheritance. For heterozygous animals, both alleles can be identified.
- Many thousands of microsatellite markers exist in the mammalian genome. Microsatellites are distributed fairly random through the genome.

So, the microsatellites form a class of highly polymorphic, relatively easy to analyze, highly frequent genetic markers. Locus specific microsatellites can be developed by sequencing the DNA adjacent to the microsatellite, e.g., ACTGCCATGGAAC-$(TG)_{10}$-CCTGATGCATGCAAG will be a locus specific microsatellite marker. Currently, microsatellites are the predominant class of markers developed in farm animal molecular genetic research labs.

Most applications of genetic markers require the mapping of QTLs, i.e., the dissection of genetic variance into components due to individual QTLs. Sax (1923) was the first to show how genetic factors influencing quantitative traits can be identified using markers. In beans, Sax scored morphological traits with monogenic inheritance and found the seed weight of certain morphological variants to be significantly higher than the seed weight of other variants. He concluded that a size factor, which we would now call a QTL, was linked to the morphological marker studied. As a result of genetic linkage between the marker and the QTL, the size factors cosegregate with the genes underlying the morphological traits.

For decades the work of Sax was not followed upon by a systematic search for QTLs. In 1961, Neimann-Sørensen and Robertson, using blood groups in dairy cattle, had shown how associations between markers and quantitative traits can be studied in outbred populations. They found no significant associations, which they attributed to the fact that they had only a few markers available which gave them a low chance of having a marker close to a QTL. For a systematic search for QTLs, many genetic markers covering the whole genome are needed, information that was simply not available until recently. With the advent of molecular genetic technology, however, many genetic markers have become available. Botstein *et al.* (1980) realized the availability of many genetic markers facilitates QTL mapping. The construction of a linkage map with many genetic markers covering the whole genome allows for a systematic screening for genes or chromosomal regions influencing important traits (Botstein *et al.* 1980). After a successful search for QTLs, genetic markers can be applied in breeding programmes.

Conventional breeding methods have been well applied in poultry breeding leading to a substantial rate of genetic improvement (Smith 1985). In poultry, nucleus breeding schemes are applied in which selection is based on parental breeding values, sib performance, and, if available, own performance. The generation interval is close to minimum but selection accuracy is relatively low, especially for sex-limited and carcass limited traits. So, increasing the accuracy of selection is expected to result in an increase in the rate of genetic improvement in poultry breeding. The use of genetic markers might increase the rate of genetic improvement due to more accurate selection decisions. In addition, the detection of markers linked to QTLs will improve the understanding of the genetic architecture of poultry which will in particular be helpful for making selection decisions. For instance, undesired effects of selection may be controlled or anticipated upon if we detect a QTL that contributes to the unfavourable correlation between two important traits, e.g., between growth rate and viability.

Throughout this thesis it is assumed that the main application of genetic markers will be marker assisted selection. For efficient marker assisted selection, genetic markers first need to be placed on the linkage map. Subsequently, the markers will be used for QTL mapping. Finally, the markers, for which linkage to a QTL has been identified, will be applied in a breeding programme through marker assisted selection. Knowledge is limited on efficient experimental designs for linkage mapping and QTL mapping in outbred populations of livestock in general and poultry in particular. Information on optimum schemes to use genetic markers in poultry breeding is absent.

Aim

The aim of this thesis is to contribute to the efficient utilization of genetic markers in poultry breeding, and to investigate the potential increase in genetic improvement from the use of genetic markers.
This thesis studies the design of linkage and QTL mapping experiments, the use of marker assisted selection in an outbred population, and discusses several applications of markers in poultry breeding schemes.

Outline of this thesis

Construction of a linkage map is laborious, hence optimal experimental designs are important. The design of linkage mapping experiments that involve crosses between inbred populations is well documented. For outbred populations, however, little research has been aimed at specifying the optimal design. Most information on linkage mapping in outbred populations comes from human studies. The human population structure can not be manipulated, so human studies focus on deriving optimal methods to analyze data of a given design. In poultry, on the other hand, the reproductive capacity of the male and female animals allow for great flexibility in the design of experiments. Therefore, this thesis studies the design of linkage mapping experiments. Chapter 2 uses two criteria to evaluate the design of a linkage mapping experiment and describes the effect of several parameters on the value of those criteria. The criteria described in chapter 2 are used in chapter 3 to derive optimum designs for reference families for livestock linkage mapping experiments.

Once constructed, a linkage map can be used in a QTL mapping experiment. The design of QTL mapping experiments using half-sib families has been investigated by Weller *et al.* (1990). Designs for QTL mapping experiments using full-sib families or a combination of full-sib families and half-sib families have not been studied. These designs are, however, potentially more efficient than half-sib family designs and practically feasible in poultry. Therefore, in chapter 4 the efficiency QTL mapping experiments using outbred populations with half-sib or full-sib families is analyzed.

Once genetic markers have been identified that explain a significant amount of genetic variance, these genetic markers can be used in breeding programmes. For dairy cattle breeding schemes and schemes including crosses between inbred populations, the value of marker assisted selection has been studied. These studies show that the structure of the breeding population and the particular values of parameters such as heritability greatly affect the value of marker assisted selection. Therefore, conclusions on the use of markers in poultry breeding schemes can not be based on literature available on other species. In chapter 5, the value of genetic markers for selection for a sex-limited trait in an outbred poultry breeding nucleus is assessed.

In the general discussion, first the status of linkage and QTL mapping in poultry is described and uses of genetic markers are summarized. Then, at the end of the main part of this thesis, three potentially useful applications of genetic markers in poultry breeding are described utilizing the knowledge generated in chapter 5, and the characteristics of poultry breeding and genetic markers.

References

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331.

Neimann-Sørensen A, Robertson A (1961) The association between blood groups and several production characteristics in three Danish cattle breeds. Acta Agric Scan 11:163-196.

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in Phaseolus Vulgaris. Genetics 8:552-560.

Smith C (1985) Rates of genetic change in farm livestock. Res Develop Agric 1:79-85.

Smith C, Simpson SP (1986) The use of genetic polymorphisms in livestock improvement. J Anim Breed Genet 103:205-217.

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25-33.

Weller JI, Kashi Y, Soller M (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J Dairy Sci 73:2525-2537.

Chapter 2

Criteria to optimize designs for detection and estimation of linkage between marker loci from segregating populations containing several families

Sijne van der Beek and Johan A.M. van Arendonk

Published in Theoretical and Applied Genetics (1993) 86:269-280
Reproduced by permission of Springer-Verlag, Berlin, Heidelberg

Abstract

Construction of a genome map of highly polymorphic markers has become possible in the past decade. Establishing a complete marker map is an enormous task. Therefore, designs to map molecular markers should be optimal. Designs to detect and estimate linkage between markers from segregating populations were studied. Two measures of design quality were used. The expectation of the maximum lod score indicates the possibility of designs to detect linkage. The accuracy of estimating recombination rate was measured as the probability that the true recombination rate is in a specified interval given the estimate. Accurate approximate methods were developed for rapid evaluation of designs. Seven family types (e.g., double backcross) can be distinguished that describe all families in a segregating population. Family type influences expected maximum lod score and accuracy of estimation. Frequency of favourable family types increased with increasing marker polymorphism. At a true recombination rate of .20, 27 observations on offspring when five alleles were segregating, and 55 observations on offspring when two alleles were segregating, were necessary to obtain an expected maximum lod score of 3. The probability that the true recombination rate was between .15 and .25, given an estimate of .20, was about .85 for a design with 40 families with ten offspring and two alleles segregating and for a design with ten families with ten offspring and six alleles segregating. For smaller designs, accuracies were less, approximate evaluation of accuracy was not justified and, on average, true recombination rates were much greater than estimated, given a specified value for the estimated recombination rate.

Key words: gene mapping - design - segregating populations - detection - accuracy

Introduction

Construction of a genome map is in progress for several livestock species (e.g., Fries *et al.* 1989; Bitgood and Somes 1990; Georges *et al.* 1990; Haley *et al.* 1990; Brascamp *et al.* 1991). A map of marker loci, i.e., loci showing Mendelian inheritance, is of use in further mapping and utilizing loci affecting quantitative traits of economic importance and for introgression and isolation of genes (Soller and Beckmann 1983; Kennedy *et al.* 1990). Constructing a map of marker loci is laborious, hence optimal experimental designs and efficient statistical procedures are important.

Methods to detect and estimate linkage between loci based on completely inbred lines of plant and animal species have been extensively described (e.g., Mather 1951; Bailey 1961; Ritter *et al.* 1990). Availability of inbred lines provides a way to optimize design of experiments to map marker loci. Inbred lines are widely used in laboratory animals and plants. For livestock species completely inbred lines are not available. Methods using information from segregating populations have been developed in human genetics (Morton 1955; Ott 1991). Here, the influence of the researcher on the experimental design is limited. Therefore, emphasis has been on development of efficient estimation procedures given the data. In livestock species, however, experimental designs can be optimized. In most species many paternal half-sibs and full-sibs can be obtained in a short period.

In genome mapping experiments, a certain set of families is used to map marker loci. In a segregating population a family can be a backcross, an intercross or another type. Once families are selected they are used irrespective of their suitability for a specific pair of markers. Therefore, parameters are needed to measure the effectiveness for detecting and estimating linkage of different family types. Importance of a family type will be determined by its frequency rather than its suitability. An overview of family types, value per family type, frequencies of family types and other aspects of importance in designing an experiment to map loci in a segregating population, is currently unavailable.

This paper describes factors influencing the quality of designs to map marker loci. Family types will be described systematically. Experimental designs will be compared with respect to detection of linkage and accuracy of estimates in segregating populations. Accuracy will be the probability that true recombination rate is in specified interval given an estimated value for recombination rate. For detection of linkage an approximate algorithm is developed and evaluated. The accuracies of estimates obtained from two approximate methods are compared with results obtained from simulation. Optimal designs are determined by varying the number and size of full-sib families for different levels of recombination rate and polymorphism of marker loci. In addition, the importance of knowledge of the linkage phase of marker alleles in the parents is determined.

Notation and assumptions

The linkage relationship between two loci with completely codominant inheritance is studied. Loci are denoted as A and B with alleles $\{A_1,A_2 \ldots\}$ and $\{B_1,B_2,\ldots\}$ respectively. Genotypes are given as $A_1A_1B_1B_2$ when the linkage phase or simply phase is unknown and as A_1B_1/A_1B_2 when the phase is known ('/' separates the two haplotypes).

Recombination rate is denoted as θ, its maximum likelihood estimate as $\hat{\theta}$ and the true value as θ_t. The probability of a certain event x is denoted as $P(x)$.

Observations are from full-sib families with information for two generations; i.e., genotypes are known without error for parents and offspring. Full-sib families are assumed to be unrelated and of equal size.

Methods

Estimation of recombination rate. The recombination rate between two markers is estimated by maximum likelihood. The likelihood function for designs with genotype information on parents and offspring of unrelated full-sib families is:

$$L(\theta)=\prod_{f=1}^{N_f} l_f(\theta)=\prod_{f=1}^{N_f}\prod_{p=1}^{n(f)} P(g_{f(p)}\mid g_{sf},g_{df},\theta)$$
$$=\prod_{f=1}^{N_f}\sum_{i=1}^{2}\sum_{j=1}^{2}\prod_{p=1}^{n(f)} P(g_{f(p)}\mid h_{sfi},h_{dfj},\theta)P(h_{sfi}\mid g_{sf})P(h_{dfj}\mid g_{df}) \qquad (1)$$

where l_f is the likelihood for family f, g_{sf} is the genotype of the sire, g_{df} is the genotype of the dam, $g_{f(p)}$ is the genotype of the p^{th} offspring of family f, h_{sfi} is the haplotype of the sire given phase i, h_{dfj} is the haplotype of the dam given phase j, N_f is the number of full-sib families, $n(f)$ is the number of offspring in family f and θ is the recombination rate.

Three components determine the likelihood function: 1) information on the phase in parents; 2) information on the gamete a parent transmits to an offspring; 3) identification of parental gametes in offspring. These three factors will now be examined in more detail.

1) For a given genotype two phases are possible each with a probability of .5. When the phase for one or both parents is known, e.g., derived from genotypes of grandparents, the likelihood function can be simplified.
2) An animal with genotype A_iB_k/A_jB_l produces the two non-recombinant gametes A_iB_k and A_jB_l and the two recombinant gametes A_iB_l and A_jB_k. Probabilities are $.5\times(1-\theta)$ for the two non-recombinant gametes and $.5\times\theta$ for the two recombinant gametes. When an animal is homozygous for locus A, A_iB_k can not be distinguished from A_jB_k. The probability of a gamete which is either A_iB_k or A_jB_k is $.5\times(1-\theta)+.5\times\theta=.5$, i.e., the probability does not depend on θ. The type of a gamete which is either A_iB_k or A_jB_k is unknown and observing such a gamete provides no information about the recombination rate. An animal can produce 3 types of gametes that are denoted as non-recombinant (non), recombinant (rec) and unknown (un).
3) Information on genotype of an offspring and phases in parents is not always sufficient to decide which alleles an offspring inherited from a parent. For example, let A_iB_k/A_jB_l and $A_{i'}B_{k'}/A_{j'}B_{l'}$ be genotypes of two parents and $A_tA_uB_vB_w$ the genotype of an offspring. If $i\neq i'$, $j\neq j'$, $k\neq k'$ and $l\neq l'$ then any gamete the sire produces differs from the gamete the dam produces which enables identification of the parental gametes in offspring. However, if $i=i'$, $j=j'$, $i\neq j$ and $t\neq u$ for locus A then both A_t and A_u could be inherited from either parent and parental gametes cannot be identified.

Table 1 Information on gametes inherited and $P(g_{f(p)} \mid h_{sf}, h_{df}, \theta)$ for the possible genotypes of the offspring of parents with genotypes $A_1B_1/A_2B_2 \times A_1B_1/A_1B_2$

Genotype offspring	Type of gametes	$P(g_{f(p)} \mid h_{sf}, h_{df}, \theta)$
$A_1A_1B_1B_1 = A_1B_1/A_1B_1$	un,non	$\frac{1}{2}(1-\theta) \times \frac{1}{2}$
$A_1A_2B_2B_2 = A_2B_2/A_1B_2$	un,non	$\frac{1}{2}(1-\theta) \times \frac{1}{2}$
$A_1A_1B_2B_2 = A_1B_2/A_1B_2$	un,rec	$\frac{1}{2}\theta \times \frac{1}{2}$
$A_1A_2B_1B_1 = A_2B_1/A_1B_1$	un,rec	$\frac{1}{2}\theta \times \frac{1}{2}$
$A_1A_1B_1B_2 = A_1B_1/A_1B_2$ or A_1B_2/A_1B_1	un,un	$\frac{1}{4}\theta + \frac{1}{4}(1-\theta) = \frac{1}{4}$
$A_1A_2B_1B_2 = A_2B_1/A_1B_2$ or A_2B_2/A_1B_1	un,un	$\frac{1}{4}\theta + \frac{1}{4}(1-\theta) = \frac{1}{4}$

Table 2 Information on gametes inherited and $P(g_{f(p)} \mid h_{sf}, h_{df}, \theta)$ for possible genotypes of offspring of parents with genotypes $A_1B_1/A_2B_2 \times A_1B_1/A_2B_2$

Genotype offspring	Type of gametes	$P(g_{f(p)} \mid h_{sf}, h_{df}, \theta)$
$A_1A_1B_1B_1 = A_1B_1/A_1B_1$	non,non	$\frac{1}{2}(1-\theta) \times \frac{1}{2}(1-\theta)$
$A_2A_2B_2B_2 = A_2B_2/A_2B_2$	non,non	$\frac{1}{2}(1-\theta) \times \frac{1}{2}(1-\theta)$
$A_1A_1B_2B_2 = A_1B_2/A_1B_2$	rec,rec	$\frac{1}{2}\theta \times \frac{1}{2}\theta$
$A_2A_2B_1B_1 = A_2B_1/A_2B_1$	rec,rec	$\frac{1}{2}\theta \times \frac{1}{2}\theta$
$A_1A_2B_1B_1 = A_1B_1/A_2B_1$ or A_2B_1/A_1B_1	non,rec[a]	$2 \times \frac{1}{2}\theta \times \frac{1}{2}(1-\theta)$
$A_1A_2B_2B_2 = A_1B_2/A_2B_2$ or A_2B_2/A_1B_2	non,rec	$2 \times \frac{1}{2}\theta \times \frac{1}{2}(1-\theta)$
$A_1A_1B_1B_2 = A_1B_1/A_1B_2$ or A_1B_2/A_1B_1	non,rec	$2 \times \frac{1}{2}\theta \times \frac{1}{2}(1-\theta)$
$A_2A_2B_1B_2 = A_2B_1/A_2B_2$ or A_2B_2/A_2B_1	non,rec	$2 \times \frac{1}{2}\theta \times \frac{1}{2}(1-\theta)$
$A_1A_2B_1B_2 = A_1B_1/A_2B_2$ or A_2B_2/A_1B_1 or A_1B_2/A_2B_1 or A_2B_1/A_1B_2	2 non/2 rec	$2 \times \frac{1}{2}(1-\theta) \times \frac{1}{2}(1-\theta) + 2 \times \frac{1}{2}\theta \times \frac{1}{2}\theta$

[a] Both possible combinations of haplotypes, given the genotype of the offspring and linkage phases, include a non-recombinant gamete and a recombinant gamete; the factor in front of the probability of this class is due to the two possibilities. For matings with other parental haplotype combinations, it is possible that only one haplotype combination including one non and one rec can be derived from the genotype given the haplotypes (e.g., $A_1B_1/A_2B_2 \times A_3B_3/A_4B_4$ gives $A_1A_3B_2B_4$). $P(g_{f(p)} \mid h_{sf}, h_{df}, \theta)$ is then $\frac{1}{2}(1-\theta) \times \frac{1}{2}\theta$.

Offspring can be classified according to the type of gametes received from their parents. Within one class, offspring have equal probability given parental phases, $P(g_{f(p)} \mid h_{sfi}, h_{dfj}, \theta)$, in equation(1). The following seven classes can be distinguished:

un,un	two gametes of unknown type
un,non	one gamete of unknown type, one gamete non-recombinant
un,rec	one gamete of unknown type, one gamete recombinant

non,non	two gametes non-recombinant
non,rec	one gamete non-recombinant, one gamete recombinant
rec,rec	two gametes recombinant
2non/2rec	two gametes non-recombinant or two gametes recombinant.

The first class can result from the mating $A_1B_1/A_1B_1 \times A_2B_2/A_2B_2$ where recombinant gametes cannot be distinguished from non-recombinant gametes. Such a mating provides no information. Offspring with genotype $A_1A_2B_1B_2$ from mating $A_1B_1/A_2B_2 \times A_1B_1/A_2B_2$ either inherit two non-recombinant gametes or two recombinant gametes, i.e., class 2non/2rec. Tables 1 and 2 give offspring for two different matings that contain examples for all classes.

For given parental phases two offspring with different genotypes can be in the same class, i.e., have equal probability $P(g_{f(p)} \mid h_{sfi}, h_{dfj}, \theta)$ in equation (1). For other parental phases, these two offspring either have equal or different probabilities. Offspring which have equal probabilities independent of parental linkage phase can be grouped in likelihood calculations. All animals within a group have the same contribution to the likelihood function but the contribution of the group might differ between parental phases. Let w_{fk} denote the k^{th} group of offspring for family f. Without loss of information, equation (1) can be written as:

$$L(\theta)=\prod_{f=1}^{N_f}\sum_{i=1}^{2}\sum_{j=1}^{2}\prod_{k=1}^{nw_f}\left[P(w_{fk} \mid h_{sfi}, h_{dfj}, \theta)\right]^{n_{fk}} \quad P(h_{sfi} \mid g_{sf})P(h_{dfj} \mid g_{df}) \tag{2}$$

in which n_{fk} denotes number of offspring in group w_{fk} and nw_f number of groups in family f. Expectation of the likelihood can be calculated easier from this equation than from equation (1).

Family types. Families can be divided in seven groups according to possible classes of offspring (Table 3). The probability that a family from a segregating population is of certain type depends on polymorphism of the marker. For instance, any family will be of type I when a marker has only one allele. Assuming Hardy-Weinberg equilibrium and linkage equilibrium, probabilities can be calculated for different family types using the frequencies of the marker alleles.

For most family types, the parental phase does not affect the classes to which an offspring can be assigned. For family types IV and V, parents have the same genotypes but differ in phase. The classes to which an offspring can be assigned differ between parental phases for these two family types. When parental phases are unknown, family types IV and V can not be distinguished but have equal probability. All other family types can be distinguished without knowing parental phases.

Table 3 Gamete type inherited by offspring for the seven possible family types

	Type of gametes						
Family type	un,un	un,non	un,rec	non, non	non, rec	rec, rec	2non or 2rec
I	*						
II	*	*	*				
III		*	*				
IV					*		*
V				*	*	*	*
VI				*	*	*	*
VII				*	*	*	

I) none of the parents is heterozygous for both loci. This family type provides no information about linkage; II) single backcross, both parents have the same alleles for the intercrossed locus (e.g., $A_1B_1/A_2B_2 \times A_1B_1/A_1B_2$); III) double backcross (e.g., $A_1B_1/A_2B_2 \times A_1B_1/A_1B_1$) or single backcross in which the parents have at least one allele not in common for the intercrossed locus (e.g., $A_1B_1/A_2B_2 \times A_1B_1/A_1B_3$); IV) intercross between parents with the same alleles for both loci and unequal phase (e.g., $A_1B_1/A_2B_2 \times A_1B_2/A_2B_1$); V) intercross between parents with the same alleles for both loci and equal phase (e.g., $A_1B_1/A_2B_2 \times A_1B_1/A_2B_2$); VI) intercross between parents with the same alleles for one locus and at most one allele in common for the other locus (e.g., $A_1B_1/A_2B_2 \times A_1B_1/A_2B_3$); VII) intercross between parents in which both loci have at most one allele in common (e.g., $A_1B_1/A_2B_2 \times A_1B_1/A_3B_3$).

Detection of linkage. To determine strength of evidence in favour of linkage the lod score (Morton 1955) is commonly used. The lod score is defined as:

$$Z(\theta) = \log_{10}(L(\theta)/L(\tfrac{1}{2})) \tag{3}$$

The maximum value of $Z(\theta)$ is denoted by $Z(\hat{\theta})$.

A maximum lod score of 3 and larger is regarded as significant evidence for linkage. A lod score of 3 approximately equals a .05 probability of falsely positive linkage (Morton 1955; Ott 1991). A maximum lod score is not available at the time an experiment is planned. However, the expected maximum lod score can be calculated. This expectation provides a measure of the expected amount of evidence for linkage from a design. The expectation for the maximum lod score, $E[Z(\hat{\theta})]$, given parental genotypes and phases is:

$$E[Z(\hat{\theta})] = \sum_{x=1}^{ND} P(D_x)\ Z_x(\hat{\theta}) \tag{4}$$

in which D_x is the data in realization x, ND is the number of possible realizations of data, $P(D_x)$ is the probability for D_x given θ and phases, and $Z_x(\hat{\theta})$ is the maximum lod score for data set x.

The E[Z($\hat{\theta}$)] for a design with unknown phases is the weighted average of the E[Z($\hat{\theta}$)] for all the possible phases.

The number of data sets to be considered in exact calculation of E[Z($\hat{\theta}$)] increases rapidly with number of families and number of offspring per family. Computational requirements soon exceed a practical level. An approximate method to calculate E[Z($\hat{\theta}$)] based on distributional properties of the lod score, is used.

The expectation of Z($\hat{\theta}$) over all realizations of data can be written as :

$$\mathrm{E}[Z(\hat{\theta})] = \mathrm{E}\left(\log_{10}\frac{L(\hat{\theta})}{L(\theta_t)}\frac{L(\theta_t)}{L(\tfrac{1}{2})}\right) = \mathrm{E}\left(\log_{10}\frac{L(\hat{\theta})}{L(\theta_t)}\right) + \mathrm{E}\left(\log_{10}\frac{L(\theta_t)}{L(\tfrac{1}{2})}\right) \tag{5}$$

The term $2\ln(L(\hat{\theta})/L(\theta_t))$ has asymptotically a χ^2 distribution (Kendall and Stuart 1978) with an expectation over all data sets of 1. Equivalence between $.217 \times 2\ln(L(\hat{\theta})/L(\theta_t))$ and $\log_{10}(L(\hat{\theta})/L(\theta_t))$ leads to an approximation of the first part of (5):

$$\mathrm{E}\left(\log_{10}(L(\hat{\theta})/L(\theta_t))\right) = .217\,\mathrm{E}\left(2\ln(L(\hat{\theta})/L(\theta_t))\right) \approx .217 \tag{6}$$

The second part of (5) [$\mathrm{E}\left(\log_{10}(L(\theta_t)/L(\tfrac{1}{2}))\right)$], is equal to the expected lod score (Ott 1991). The expected lod score is additive over families (Ott 1991) and can be calculated as:

$$\mathrm{E}\left(\log_{10}\frac{L(\theta_t)}{L(\tfrac{1}{2})}\right) = \mathrm{E}\left(\log_{10}\prod_{f=1}^{N_f}\frac{l_f(\theta_t)}{l_f(\tfrac{1}{2})}\right) = \sum_{f=1}^{N_f}\mathrm{E}\left(\log_{10}\frac{l_f(\theta_t)}{l_f(\tfrac{1}{2})}\right) \tag{7}$$

where l_f represents the likelihood for family f as defined by (1).

The fact that families of the same type and with equal number of offspring have equal expectation can be used for further simplification:

$$\sum_{f=1}^{N_f}\mathrm{E}\left(\log_{10}\frac{l_f(\theta_t)}{l_f(\tfrac{1}{2})}\right) = \sum_{i=1}^{7} m_i \mathrm{E}\left(\log_{10}\frac{l_i(\theta_t)}{l_i(\tfrac{1}{2})}\right) = \sum_{i=1}^{7} m_i \sum_{y=1}^{nd_i} P(d_{iy})\,\log_{10}\left(\frac{l_i(\theta_t \mid d_{iy})}{l_i(\tfrac{1}{2} \mid d_{iy})}\right) \tag{8}$$

where i denotes family type, m_i number of families of type i, d_{iy} realization y of data for all families of type i, nd_i number of possible realizations of data for a family of type i, and $P(d_{iy})$ probability of realization of d_{iy}.

Combining (6) and (8) results in an approximate E[Z($\hat{\theta}$)], E[Z($\hat{\theta}$)$_{ap}$]:

$$E[Z(\hat{\theta})_{ap}] = .217 + \sum_{i=1}^{7} m_i \sum_{y=1}^{nd_i} P(d_{iy}) \log_{10}\left(\frac{l_i(\theta_t \mid d_{iy})}{l_i(\tfrac{1}{2} \mid d_{iy})}\right) \quad (9)$$

In segregating populations, family type of a family is unknown at the start of an experiment and the exact value of m_i is unknown. Expectation for m_i is calculated as the product of number of families and probability that a family is of type i.

Exact calculation of the E[Z($\hat{\theta}$)] involves the maximization of many ($\prod_{f=1}^{NF} nd_f$ where nd_f is the number of realizations of data for family f) likelihood functions. This number is reduced considerably with approximation (9), i.e., reduced to $2\sum_{i=1}^{7} nd_i$.

For families of type I,II,III,VI or VII, E[Z($\hat{\theta}$)] is independent of the phases. To calculate E[Z($\hat{\theta}$)] in these cases an arbitrary phase is assigned to a family when phases are unknown. When a family can be either type IV or V given its parental genotypes, E[Z($\hat{\theta}$)] is calculated for both cases. The average of both possibilities is used as an approximation to E[Z($\hat{\theta}$)].

Accuracy of estimation. Accuracy of estimated recombination rate ($\hat{\theta}$) is measured as the probability that true recombination rate (θ_t) is in a specified interval given the estimate ($\hat{\theta}=x$). This probability is calculated as:

$$P(y_1<\theta_t<y_2 \mid \hat{\theta}=x) = \frac{P(\hat{\theta}=x \mid y_1<\theta_t<y_2)}{P(\hat{\theta}=x)} \quad (10)$$

where

$P(\hat{\theta}=x \mid y_1<\theta_t<y_2) = \int_{y_1}^{y_2} P(\hat{\theta}=x \mid \theta_t) f(\theta_t) d(\theta_t)$ and

$P(\hat{\theta}=x) = \int_{0}^{.5} P(\hat{\theta}=x \mid \theta_t)\ f(\theta_t) d(\theta_t)$ with

y_1 and y_2 = lower and upper limits for θ_t, respectively and

$f(\theta_t)$ = prior density function of θ_t

Calculation of (10) involves the probability density of estimates given true recombination rate to calculate the probability that an estimate is in a certain range given θ_t and a prior probability density function of θ_t.

The maximum likelihood estimate is asymptotically normally distributed and has asymptotic variance equal to the inverse of the expected information (Kendall and Stuart 1978). Information is defined as the second derivative of the likelihood function. When

parental linkage phases are known for a family of given type, expected information is a linear function of the number of offspring in the family. For example, a family of family type III with n offspring has expected information $n/(\theta \times (1-\theta))$. The relation between number of offspring and expected information is nonlinear when parental phases are unknown.

The probability that an estimate is in a certain range for given θ_t is approximated assuming a normal distribution with variance equal to the inverse of the expected information. The same approximation is used for the probability that true recombination rate is in a certain range for a given value of the estimated recombination rate.

The prior density function of recombination rate between marker loci depends on several factors: number of chromosomes of the species, lengths of chromosomes, physical distribution of marker loci on chromosomes, and relation of distance and recombination rate between loci.

The following is assumed: loci have a probability of 1/20th to be located on the same chromosome, loci are uniformly distributed over chromosomes with a length of 1 morgan, and map distance and recombination rate are related by Haldane's mapping function (Haldane 1919). Following the approach of Morton (1955) the prior density function of θ_t is:

$$\begin{aligned} f(\theta_t) &= .05\left(\frac{2}{1-2\theta_t}(.5\ln(1-2\theta_t)+1)\right) && \text{for } 0 \le \theta_t < .432; \\ f(\theta_t) &= 0 && \text{for } .432 \le \theta_t < .5; \\ f(\theta_t) &= .95 && \text{for } \theta_t = .5 \end{aligned} \tag{11}$$

Simulation. The average value of the maximum lod score, the distribution of maximum likelihood estimator of θ and the distribution of θ_t for a given value of $\hat{\theta}$ were all obtained using Monte Carlo simulation.

The average maximum lod score was calculated for all seven family types for different values of θ and different number of offspring per family. In each simulated data set the number of offspring in each group w_{fk} was simulated using the probabilities of groups. The probability of genotype group w_{fk} for a given family type depends on θ. Values of .05 and .20 for θ and number of offspring of 2, 4, 8, 16 and 32 were used. For each alternative, 1000 data sets were simulated. In each data set $Z(\hat{\theta})$ was computed and averaged to obtain $E[Z(\hat{\theta})]$.

The distribution of the maximum likelihood estimator of θ was calculated from estimated recombination rates in different replicates. Recombination rate was estimated from data containing information from several families. The probability that a family is of a given type was calculated from frequencies of marker alleles assuming Hardy-Weinberg equilibrium for individual loci and linkage equilibrium between loci. These probabilities were used to simulate

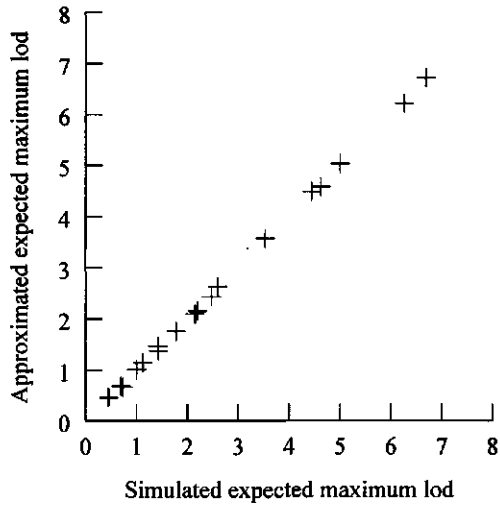


Figure 1 Relation between simulated expected maximum lod score and approximated expected maximum lod score based on designs with 4 to 16 unrelated full-sib families with 4 to 16 offspring where each (+) represents a design

Table 4 Probability for family types for various number of equiprobable alleles

No. alleles	%Hetero zygous	Type I	Type II	Type III	Type IV	Type V	Type VI	Type VII
2	50	.56	.25	.13	.03	.03	0	0
3	67	.31	.09	.41	.01	.01	.09	.09
4	75	.19	.04	.46	.00	.00	.09	.22
5	80	.13	.02	.44	.00	.00	.07	.33
6	83	.09	.01	.42	.00	.00	.06	.42
7	86	.07	.00	.39	.00	.00	.05	.49
8	88	.05	.00	.36	.00	.00	.04	.55
9	89	.04	.00	.33	.00	.00	.03	.59
10	90	.03	.00	.31	.00	.00	.03	.63

family types for a data set. The following values were used: for θ: .05, .20 and .50; for number of families: 1, 2, 4, 8, 16 and 32; for number of offspring per family: 2, 4, 8, 16 and 32. For each alternative, 10 000 data sets were simulated.

The distribution of θ_t for a given value of $\hat{\theta}$ was determined for given number of offspring per family and number of families. Simulations to obtain the distribution were as follows. First, two markers were randomly located on a chromosome of 1 morgan using a uniform distribution. The recombination rate between the markers was calculated from the distance between the markers assuming Haldane's mapping function (Haldane 1919). Family types and information on offspring were simulated for the simulated θ_t and $\hat{\theta}$ calculated from the data. This was repeated 100000 times. Number of realizations of ($\hat{\theta} = x$, $\theta_t = y$) were counted. Second, 100000 data sets were simulated with markers located on different chromosomes, i.e., for $\theta_t = .50$. Maximum likelihood estimates were calculated and number of realizations of ($\hat{\theta} = x$, $\theta_t = .50$) counted. The number of realizations of ($\hat{\theta} = x, \theta_t = .50$) were multiplied by 19 to take into account that the prior probability that two markers are on separate chromosomes is 19 times the prior probability that two markers are on the same chromosome. Obtained from counts of ($\hat{\theta} = x$, $\theta_t = y$) were: distribution of θ_t given $\hat{\theta} = x$; distribution of $\hat{\theta}$ given $\theta_t = y$; $P(y_1 < \theta_t < y_2 \mid \hat{\theta} = x)$; average value of θ_t given $\hat{\theta} = x$ and average value of $\hat{\theta}$ given $\theta_t = y$. The following values were used in simulations: for number of families: 10, 20 and 40; for number of offspring per family: 4, 10 and 46.

Results

Detection. Figure 1 shows the approximated expected maximum lod score, $E[Z(\hat{\theta})_{ap}]$, and the expected maximum lod score obtained by simulation, $E[Z(\hat{\theta})_{sim}]$, for alternatives with 4 to 16 families and 4 to 16 offspring per family. The $E[Z(\hat{\theta})_{ap}]$ agree with the expectation obtained by simulation. The $E[Z(\hat{\theta})_{ap}]$ will be used in this study and called $E[Z(\hat{\theta})]$ in the remainder of this paper.

Probabilities of family types were calculated assuming a population in Hardy-Weinberg equilibrium. Number of alleles per marker locus influenced the distribution of families over family types (Table 4). The probability of family type I (least favourable family type) decreased and the probability of type VII (most favourable family type) increased when number of alleles increased. Heterozygosity of marker loci increased with increasing number of alleles. The marginal change per additional allele decreased with increasing number of alleles for both heterozygosity and probabilities of family types. The results in Table 4 show a clear relation between the distribution over family types and heterozygosity. This relation is expected to hold when number of alleles differs per locus and when alleles have unequal frequencies.

Table 5 Expected maximum lod score ($E[Z(\hat{\theta})]$) for designs with one family for different family types and number of offspring per family, for two values of θ_t and when phase is known; $E[Z(\hat{\theta})]$ for phase known minus $E[Z(\hat{\theta})]$ when phase is unknown is given between brackets

No. offspring	Family type					
	II	III	IV	V	VI	VII
$\theta_t = .05$						
4	.46 {.19}	1.02 {.29}	.87 {.26}	1.40 {.54}	1.40 {.46}	1.93 {.59}
8	1.01 {.28}	1.93 {.30}	1.57 {.30}	2.62 {.60}	2.62 {.60}	3.68 {.60}
16	1.93 {.30}	3.68 {.30}	2.88 {.30}	5.00 {.60}	5.00 {.60}	7.12 {.60}
32	3.68 {.30}	7.12 {.30}	5.50 {.30}	9.74 {.60}	9.74 {.60}	13.98 {.60}
$\theta_t = .20$						
4	.31 {.14}	.59 {.22}	.35 {.14}	.65 {.35}	.65 {.20}	.92 {.41}
8	.57 {.19}	.92 {.25}	.46 {.17}	1.03 {.46}	1.03 {.31}	1.57 {.53}
16	.92 {.25}	1.57 {.28}	.68 {.21}	1.80 {.55}	1.80 {.43}	2.90 {.59}
32	1.58 {.29}	2.90 {.30}	1.14 {.26}	3.36 {.60}	3.36 {.52}	5.58 {.60}

Table 6 Additional number of observations[a] on offspring to compensate for smaller expected maximum lod score due to unknown phase for an average informative[b] family from populations with 2, 5 or 10 equiprobable alleles and θ_t of .05 or .20

No. alleles	Known - unknown[c]	θ_t	
		.05	.20
2	.141	1.9	5.5
5	.336	1.4	3.6
10	.462	1.4	3.6

[a] Average difference in $E[Z(\hat{\theta})]$ divided by the average $E[Z(\hat{\theta})]$ per observation on offspring. Average $EZ[(\hat{\theta})]$ per observation calculated as ($E[Z(\hat{\theta})]$ at 32 offspring - $E[Z(\hat{\theta})]$ at 16 offspring) / 16.

[b] An informative family is not of type I

[c] Average difference in $E[Z(\hat{\theta})]$ = difference in $E[Z(\hat{\theta})]$ for family type II (=.3) × probability family type II + difference in $E[Z(\hat{\theta})]$ for family type III (=.3) × prob. family type III + etc.

Table 5 gives the relation between $E[Z(\hat{\theta})]$ and family type, number of offspring, θ_t and knowledge of parental phases. When phases were known, $E[Z(\hat{\theta})]$ of family types II, III and VII were in the proportion of .5:1:2, independent of θ_t and number of offspring. This proportion corresponds to the number of informative gametes for these family types. The ratio between the $E[Z(\hat{\theta})]$ for family type IV and family type III was close to .8 when recombination rate was .05. The ratio was .4 when recombination rate was .20. For family types V and VI, ratios with family type III were 1.4 when recombination rate was .05 and

1.15 when recombination rate was .20. For family types IV, V and VI, there was no direct relation between proportional $E[Z(\hat{\theta})]$ and number of gametes.

The difference in $E[Z(\hat{\theta})]$ due to knowledge of phase approached to a constant for each family type with increasing family size (Table 5). As an explanation, consider the function for the lod score for one double backcross family (family type III):

$$\text{lod}_{\text{known}}(\theta) = \log_{10}\left(\frac{\theta^x(1-\theta)^{n-x}}{.5^n}\right) = x\log_{10}(\theta) + (n-x)\log_{10}(1-\theta) + n\log_{10}(2)$$

$$\text{lod}_{\text{unknown}}(\theta) = \log_{10}\left(\frac{.5\left(\theta^x(1-\theta)^{n-x} + (1-\theta)^x\theta^{n-x}\right)}{.5^n}\right)$$

$$= \log_{10}(.5) + \log_{10}(\theta^x(1-\theta)^{n-x}\left(1 + (\theta/(1-\theta))^{n-2x}\right)) + n\log_{10}(2)$$

$$= \log_{10}(.5) + \log_{10}(1 + (\theta/(1-\theta))^{n-2x}) + x\log_{10}(\theta) + (n-x)\log_{10}(1-\theta) + n\log_{10}(2)$$

$$= \text{lod}_{\text{known}}(\theta) - .3 + \log_{10}(1 + (\theta/(1-\theta))^{n-2x})$$

where x is the number of recombinant gametes, $\text{lod}_{\text{known}}$ is the function for lod score when phase is known, and $\text{lod}_{\text{unknown}}$ is the function for lod score when phase is unknown. The expectation for x is $n \times \theta_t$. The term $\log_{10}\left(1+(\theta/(1-\theta))^{n-2x}\right)$ goes to zero when n becomes large and θ_t is not close to .5. The difference between the lod score for phase known and phase unknown is then a constant (.3), which equals to the difference in $E[Z(\hat{\theta})]$. Similar relations occur for families of other types. The difference in $E[Z(\hat{\theta})]$ approached .3 for family types II, III and IV and .6 for family types V, VI and VII.

The additional number of observations on offspring needed to compensate for smaller $E[Z(\hat{\theta})]$ due to lack of knowledge of phases was dependent on family type and θ_t. Therefore, the additional number of observations on offspring needed for an average family depended on polymorphism of the marker and θ_t (Table 6). The maximum of 5.5 occurred when number of alleles was two and θ_t was .2. The alternative for typing additional offspring is typing grandparents (four additional observations). Partial or complete knowledge of phases can be obtained from grandparents.

The average number of offspring in one family needed to obtain an $E[Z(\hat{\theta})]$ of 3 is given in Table 7 for three levels of marker polymorphism and two values of θ_t. The required number of observations on offspring was 19 for θ_t of .05 and 55 for θ_t of .20 when number of alleles was two, which is twice the number required with five alleles.

Table 7 Number of observations[a] on offspring for an $E[Z(\hat{\theta})]$ of 3 for an average informative family from populations with 2, 5 or 10 alleles of equal frequency and a θ_t of .05 or .20[b]

No. alleles	θ_t	
	.05	.20
2	19	55
5	10	27
10	8	22

[a] Excluding observations on parents
[b] For each family type $E[Z(\hat{\theta})]$ per observation is calculated from the difference between $E[Z(\hat{\theta})]$ with 32 offspring and $E[Z(\hat{\theta})]$ with 16 offspring. A weighted average $E[Z(\hat{\theta})]$ per observation is calculated using the probabilities for the family types given the number of alleles

Table 8 Expected maximum lod score dependent on number of families and total number of observations[a] on offspring for designs with family types due to chance (both marker loci have two alleles), unknown phases and $\theta_t = .05$

No. observations	number of families				
	1	2	4	8	16
32	2.26	2.11	1.85	1.43	.90
64	4.44	4.30	4.01	3.48	2.64

[a] Excluding observations on parents

A given number of observations on offspring can be obtained by analysing different numbers of families. $E[Z(\hat{\theta})]$ decreased if number of families increased and phase was unknown (Table 8). For each family, information is used to estimate the phase. Therefore, required number of observations to obtain an $E[Z(\hat{\theta})]$ of 3 will increase when observations on offspring are divided over more than one family. Further, number of observations increases with number of families because for each family two parents must be genotyped.

Accuracy. In Table 9 the mean and standard error for $\hat{\theta}$ are given for different numbers of observations, θ_t and full-sib family size. Observed standard errors were obtained from replicated simulations. Standard errors were approximated using the second derivative of the correct likelihood function (σ_{ap2}) and the likelihood function where parental phases were assumed known (σ_{ap1}). Estimated recombination rates were biased upward for θ_t of .2. For θ_t of .05 an upward bias was observed when number of animals per family was 4. Downward bias was found for θ_t of .5 which is inevitable because estimated recombination rates are

restricted to be between 0 and .5. The bias diminished with increasing number of observations. For a given number of observations, bias was less when number of observations per family increased. Observed and approximated standard errors agree closely for designs with 120 or more observations when θ_t is .05 and 240 or more observations when θ_t is .20. With fewer observations both approximations underestimated standard error. Standard errors approximated using the correct likelihood function (σ_{ap2}) were closer to observed standard errors for θ_t of .2. Expected information calculated from the correct likelihood function is zero for unlinked loci and as a result σ_{ap2} does not exist for θ_t of .5.

Table 9 Average estimated recombination rate, observed standard error (σ_{obs}), standard error approximated using likelihood function assuming parental phases known (σ_{ap1}) and standard error approximated using correct likelihood function (σ_{ap2}) for designs varying in number of families, family size and $\hat{\theta}$; one marker has two alleles with equal frequency and one marker locus has six alleles with equal frequency, parental phases are unknown

θ_t	No. obs[a]	4 offspring per family, 6 observations per family				10[b] offspring per family, 12 observations per family			
		$\hat{\theta}$	σ_{obs}	σ_{ap1}	σ_{ap2}	$\hat{\theta}$	σ_{obs}	σ_{ap1}	σ_{ap2}
.05	30	.057	.083	.055	.057	.052	.067	.051	.051
	60	.053	.045	.039	.040	.050	.039	.035	.035
	120	.051	.030	.028	.028	.051	.027	.025	.025
	240	.051	.021	.020	.020	.049	.018	.017	.018
	480	.050	.014	.014	.014	.050	.013	.012	.012
	960	.050	.010	.010	.010	.050	.009	.009	.009
.20	30	.247	.166	.107	.124	.219	.136	.098	.102
	60	.228	.122	.076	.088	.213	.094	.064	.067
	120	.210	.084	.053	.062	.203	.055	.048	.049
	240	.203	.049	.038	.044	.202	.036	.034	.035
	480	.202	.032	.027	.031	.200	.025	.024	.025
	960	.201	.023	.019	.022	.200	.018	.017	.017
.50	30	.407	.140	.144	*	.430	.110	.131	*
	60	.414	.109	.102	*	.441	.086	.091	*
	120	.426	.093	.072	*	.447	.068	.064	*
	240	.436	.077	.051	*	.459	.054	.045	*
	480	.449	.062	.036	*	.463	.046	.032	*
	960	.457	.052	.025	*	.470	.038	.023	*

[a] No. obs = number of observations on parents and offspring

[b] For 30 observations, three families each with eight offspring were taken

* σ_{ap2} was undefined (1/0) for θ_t is .5

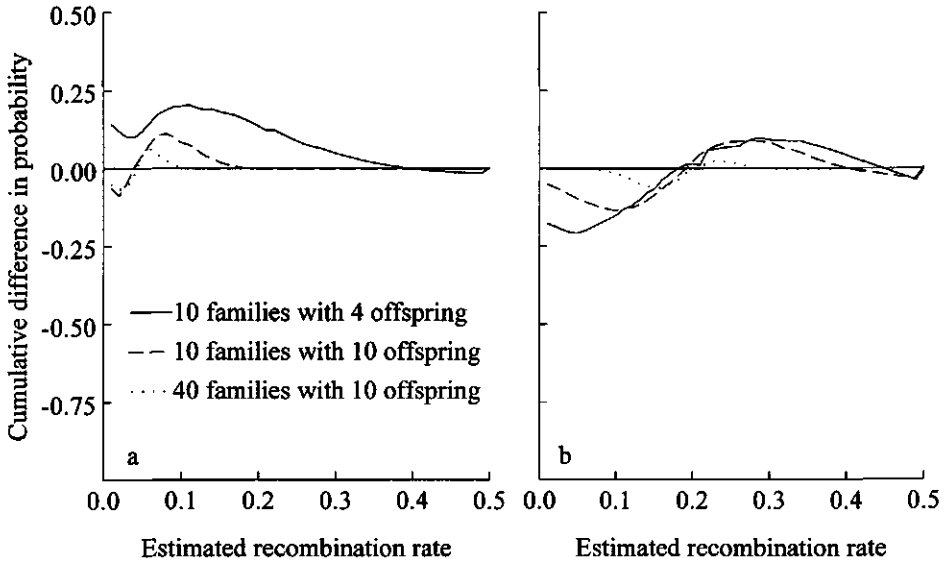


Figure 2 Cumulative difference between simulated and approximated probability of *estimated* recombination rate given *true* recombination rate of (a) .05 and (b) .20 for designs with two equiprobable alleles and three sizes

The observed cumulative probability distribution of the estimates is compared with the cumulative normal distribution in Figure 2. For ten families of four offspring and two alleles for each locus, observed probabilities for estimates of 0 or .5 were larger than probabilities calculated from the normal distribution. Differences might be expected since the normal distribution function is approximate for large numbers because of the central limit theorem. With increasing numbers, the deviation between the approximation and the observed distribution became smaller and was negligible for 400 observations. Probabilities that estimates were in a certain interval, given a true recombination rate, could be adequately approximated using the normal distribution for larger designs.

Gene maps or parts of gene maps are often evaluated based on spacing between considered loci. A logical assumption is that, on average, true recombination rate is equal to a given estimated recombination rate. It is not obvious whether or not this assumption is always correct. In Figure 3 average true recombination rates are plotted against estimated recombination rates. For alternatives with ten families and four or ten offspring per family, average true recombination rate deviated from given estimated recombination rate. This deviation can be explained by the .95 prior probability that θ_t is .5 and the large variance of the estimator when number of observations on offspring is small.

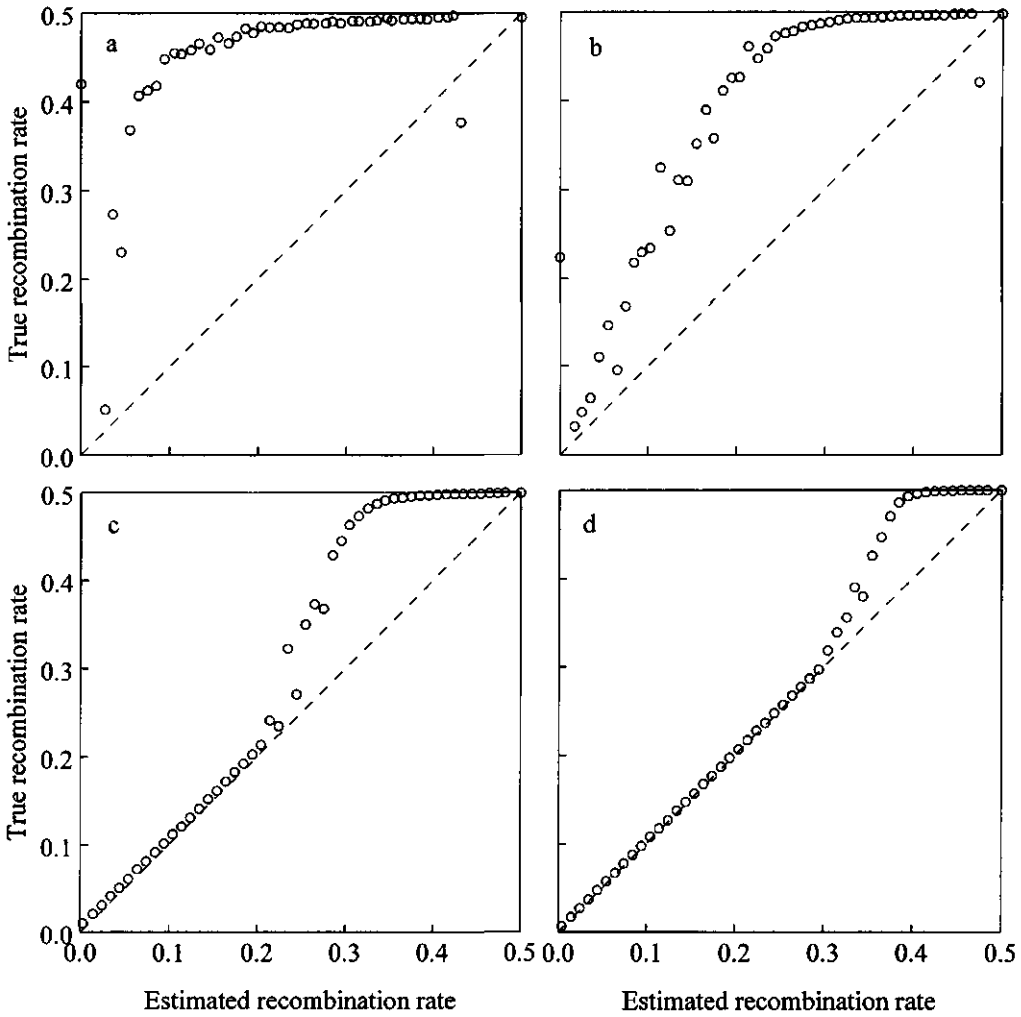


Figure 3 Average true recombination rate for given estimated recombination rate. For 100 000 replicates true recombination rate was sampled from zero to .5 and for 100 000 replicates true recombination rate was .5. Based on true recombination rate data were simulated. Replicates were classified in fifty classes according to estimated recombination rate. Because the prior probability that true recombination rate is .5 is 19 times the prior probability that true recombination rate is between 0 and .5, replicates with a true recombination rate of .5 were weighted by a factor 19. For each class average true recombination rate was calculated. Simulation was done for four designs with two alleles of equal frequency and (a) ten families with four offspring, (b) ten families with ten offspring, (c) 40 families with ten offspring, and (d) 20 families with 46 offspring.

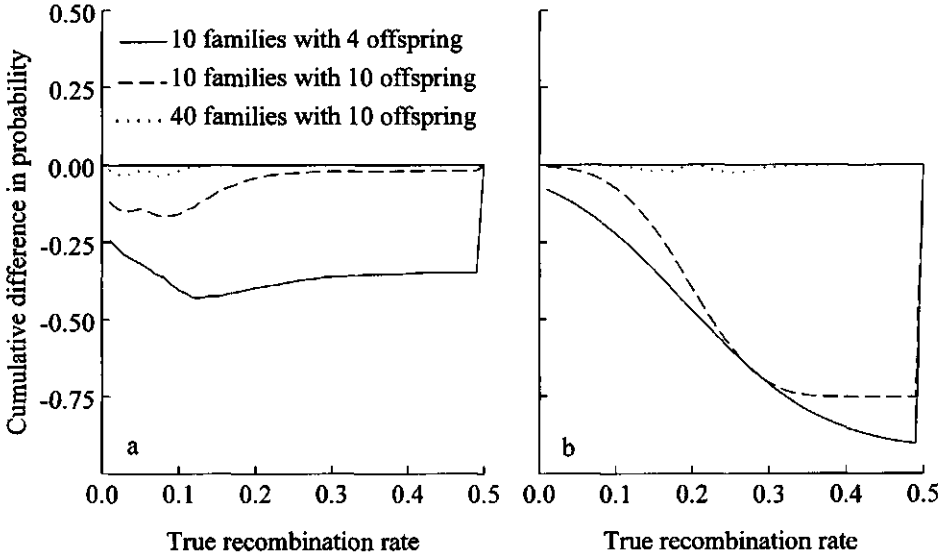


Figure 4 Cumulative difference between simulated and approximated probability of *true* recombination rate given *estimated* recombination rate of (a) .05 and (b) .20 for designs with two equiprobable alleles and three sizes

Most of the deviation between estimated and average true recombination rate disappeared when only replicates were considered where $\hat{\theta}$ was significantly different from .5, i.e., $Z(\hat{\theta})$ larger than 3. For the design with ten families and four offspring per family almost no replicates had a $Z(\hat{\theta})$ above 3. The latter observation is consistent with the fact that for large designs the deviation disappeared: recombination rates can only be significant if sufficient observations are available.

The difference between observed cumulative probability of θ_t given $\hat{\theta}$ and approximated normal probability is plotted in Figure 4. In all cases observed cumulative probability was smaller than approximated. With ten families, four offspring per family and $\hat{\theta}$ of .05, for θ_t of .49 approximated cumulative probability was .34 larger than observed cumulative probability. As expected, no difference was found for θ_t of .5. The observed probability that θ_t is .5 for $\hat{\theta}$ of .05 was underestimated by .34 using the approximate distribution function. For that design the probability θ_t is smaller than .2 for an $\hat{\theta}$ of .05 is overestimated with .42 by the approximate distribution. For larger designs this difference was negligible. For an $\hat{\theta}$ of .2 the probability of θ_t is .5 was underestimated by .91 using the approximate probability function for designs with ten families of four or ten offspring each (Figure 4). For the largest design, the approximated cumulative distribution was in good agreement with the observed distribution.

Table 10 Simulated and approximated probability that true recombination rate is in a specified interval ($.15 < \theta_t < .25$) given estimated recombination rate is .20

No. families	Family size	No. alleles	*P*(obs)	*P*(approx1)	*P*(approx2)
10	4	2	.0335	.1029	.3507
		6	.3013	.6438	.6359
	10	2	.1456	.4600	.5279
		6	.8786	.8543	.8487
40	10	2	.8496	.8529	.8495
		6	.9972	.9963	.9959
20	46	2	.9748	.9721	.9708
		6	1	.9999	.9999

P(obs) : Observed probability, calculated from simulation (100 000 replicates)
P(approx1): The term $P(\hat{\theta} \mid .15 < \theta_t < .25)$ from equation (10) is approximated using the normal distribution
P(approx2): $P(.15 < \theta_t < .25 \mid \hat{\theta})$ is entirely approximated using the normal distribution

In Table 10 observed and approximated $P(y_1 < \theta_t < y_2 \mid \hat{\theta} = x)$ are given. Two approximations were used. In both methods $P(\hat{\theta} = x)$ is calculated as $P(x - .005 < \hat{\theta} < x + .005)$. In the first approximation, probability on an estimate for a given value of θ_t, or $P(\hat{\theta} = x \mid \theta_t = y)$, is calculated assuming a normal distribution of $\hat{\theta}$ around θ_t. Multiplying $P(\hat{\theta} = x \mid \theta_t = y)$ by the prior probability of θ_t, integrating over θ_t and applying equation (10) completes the first approximation. In the second approximation, true recombination rate is falsely assumed to be normally distributed around $\hat{\theta}$ and an approximation of accuracy is directly obtained from the normal distribution. The second approximation is much more rigorous since the prior probability function of θ_t is ignored. However, Table 10 shows that both approximations worked equally well for the studied alternatives. The first approximation was only better for alternatives for which both approximations were bad. For designs larger than ten families with ten offspring and six equiprobable alleles per locus, approximations were similar and the deviation between observed and approximated probabilities was small.

Discussion

The expected value of maximum lod score and accuracy of an estimated recombination rate were used to describe and study quality of experimental designs. Ott (1991) argued that the expectation of maximum lod score is not additive over families and has no clear probabilistic interpretation. He concluded that expectation of lod score, for a given value of θ, should be preferred because this expectation is additive over families. The approximated expected maximum lod score, used in this study, was calculated as the sum of expected lod score and a constant. Taking the constant into account resulted in good approximation of the real situation where data of several families were used (Figure 1).

The approximate methods served two purposes. First, they simplified computations. Second, comprehension of the behaviour of estimators was enhanced.

Number of full-sib families, number of offspring per family and knowledge of phases were shown to affect $E[Z(\hat{\theta})]$ (Tables 5 and 8). $E[Z(\hat{\theta})]$ was larger when phase was known. The additional number of observations on offspring needed to compensate for lack of knowledge of phases was within a reasonable range (< 6 for $\theta_t \leq .2$, Table 6). Typing grandparents to determine the parental phase is not an alternative reducing the number of typings to be done for that range of θ_t. However, the additional number of observations on offspring will increase for larger θ_t. For θ_t larger than .20, obtaining information on parental phase might be worthwhile. The aim of most genome mapping projects is to create a map with markers spaced by no more than 20 centimorgans. In such projects DNA of grandparents is not really needed. Hetzel (1990) pointed out that typing grandparents provides a check for consistency of segregation. However, typing many offspring also provides a check. The possibility of typing errors emphasizes the necessity of typing many offspring per family rather than typing grandparents.

In this paper, designs with unrelated full-sib families were studied. Elements influencing the quality of designs are most clearly illustrated for this class of designs. Computations are simple. The results for these designs can be used for all other designs with information on parents and offspring when parental phase is known. In a hierarchical half-sib structure with equal number of offspring per dam and several dams per sire, fewer sires are used compared to a full-sib structure with the same number of dams and offspring. When parental phases are unknown, fewer sires means that less information will be used to infer parental phases from the data. As a consequence $E[Z(\hat{\theta})]$ will be larger for the hierarchical half-sib structure.

Results in Table 8 showed that $E[Z(\hat{\theta})]$ can be maximized by minimizing number of families. A minimum number of families is not necessarily optimal, however. With a minimum number of families the variation in realized maximum lod score is maximal and, as a consequence, the probability of having no information is maximal. The risk due to large variation in outcome of an experiment can be summarized by the probability of no information. Assume the probability of no information is to be less than .10. The necessary number of families can be calculated as $-1/\log_{10}(P(\text{typeI}))$ where $P(\text{typeI})$ is the probability that a family is of type I. For a design where marker loci have two equiprobable alleles, $P(\text{typeI})$ is .5625 and the probability of no information is .1 when number of families is four. Given this number of families, the number of offspring per family resulting in $E[Z(\hat{\theta})] = 3$ can be calculated. When marker loci have two equiprobable alleles, four families each with 33 offspring are needed for $E[Z(\hat{\theta})] = 3$. When marker loci have four equiprobable alleles the probability of no information is .04 with two families and in that case 21 offspring per family are needed for a $E[Z(\hat{\theta})]$ of 3. The number of offspring per full-sib family in these

examples is larger than that available in most livestock species. The restriction on probability of no information will not change the optimal design when number of offspring per family is less than 20. There is, however, still a risk that the realized maximum lod score in an experiment is lower than $E[Z(\hat{\theta})]$. A more general approach is to look at the power of a design.

The relation between marker polymorphism and distribution over family types demonstrated the advantage of highly polymorphic markers (Table 4). Research of Georges *et al.* (1990) in cattle showed an average heterozygosity of 51 % for VNTR markers and 65% for microsatellites. These heterozygosities correspond to about two or three alleles of equal frequency (see Table 4), or more alleles of varying frequencies. Consequently, on average, a considerable proportion of the families will provide no, or less than maximal, information on linkage.

Boehnke (1986) described a simulation approach by which average maximum lod scores and power can be obtained for any design. For plants, elements of the design of experiments are described in standard text books (e.g., Mather 1951; Bailey 1961; Green 1981). Restriction is usually made to designs with double backcrosses or intercrosses, family types III and V respectively, and known phases. This study described all possible family types in a segregating population and paid attention to families larger than those used for human linkage studies. The derived algorithm considers marker polymorphism, all family types and can be used for varying full-sib family sizes and number of families. The method of Boehnke (1986) is general and further calculates power but requires simulation of many replicates for each design to be evaluated.

Accuracy was calculated from the distribution function of θ_t for a given value of $\hat{\theta}$. Elements of the function were studied and compared with approximations. Bias in the estimate of recombination rate, given a θ_t for designs with few observations, could be explained by the observed distribution of the estimates. Bias of estimated recombination rates was studied in more detail by Bolling and Murphy (1979). For designs larger than or equal to ten families with ten offspring, use can be made of the normal distribution with approximated standard error to calculate the probability for an estimate given a true recombination value.

The use of the prior probability density of θ_t in linkage studies has been advocated by Smith (1959), Smith and Sturt (1976), Silver and Buckler (1985) and Neumann (1990,1991). This approach considers the .95 probability that loci are unlinked. The effect of prior density of true recombination rate is shown in Figure 3. The .95 probability of no linkage resulted in a large deviation between true and estimated recombination rate. The influence of the large probability of unlinked loci on average estimated recombination rate could be reduced by considering only replicates which had an estimate for θ significantly different from .5.

For large designs, inferences about true recombination rate can be made using the normal distribution and the approximated standard error. For small designs, a restriction needs to be made to significant recombination rates. Results of this study emphasize the necessity to use significant estimates because non significant estimates are not only inaccurate but, on average, are very different from true recombination rates.

Acknowledgements

The authors acknowledge financial support from Euribrid B.V.. The authors also thank two anonymous reviewers for their useful comments.

References

Bailey NTJ (1961) Introduction to the mathematical theory of linkage. Clarendon Press, Oxford.

Bitgood JJ, Somes RG (1990) Linkage relationships and gene mapping. In: Crawford RD (ed) Poultry breeding and genetics. Elsevier, Amsterdam.

Boehnke M (1986) Estimating the power of a proposed linkage study: a practical computer simulation approach. Am J Hum Genet 39:513-527.

Bolling DR, Murphy EA (1979) Finite sample properties of maximum likelihood estimates of the recombination fraction in double backcross matings in man. Am J Med Genet 3:81-95.

Brascamp EW, Haley CS, Fries R, Grosclause F, Hanset R, Womack J (1991) Genome mapping in farm animals, a survey of current research. Paper presented at the 42th annual meeting of the EAAP, Berlin.

Fries R, Beckmann JS, Georges M, Soller M, Womack J (1989) The bovine gene map. Anim Genet 20:3-29.

Georges M, Mishra A, Sargeant L, Steele M, Zhao X (1990) Progress towards a primary DNA marker map in cattle. Proc 4th World Congr Genet Appl Livestock Prod 13:107-112.

Green EL (1981) Genetics and probability in animal breeding experiments. Oxford Univ. Press, New York.

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. J Genet 8:299-309.

Haley CS, Archibald A, Andersson L, Bosma AA, Davies W, Fredholm M, Geldermann H, Groenen M, Gustavsson I, Ollivier L, Tucker EM, Van de Weghe A (1990) The pig gene mapping project - PiGMaP. Proc 4th World CongrGenet Appl Livestock Prod 13:67-70.

Hetzel DJS (1990) Reference families for genome mapping in domestic livestock. In: LB Schook, HA Lewin, DG McLaren (eds) Gene-mapping techniques and applications. Marcel Dekker, New York.

Kendall M, Stuart A (1978) The advanced theory of statistics, Vol 2. Griffin, London.

Kennedy BW, Verrinder Gibbins AM, Gibson JP, Smith C (1990) Coalescence of molecular and quantitative genetics for livestock improvement. J Dairy Sci 73:2619-2627.

Mather K (1951) The measurement of linkage in heredity. Mehtuen & co. Ltd., London.

Morton NE (1955) Sequential test for the detection of linkage. Am J Hum Genet 7:277-318.

Neumann PE (1990) Two-locus linkage analysis using recombinant inbred strains and Bayes' Theorem. Genetics 126:277-284.

Neumann PE (1991) Three-locus linkage analysis using recombinant inbred strains and Bayes' Theorem. Genetics 128:631-638.

Ott J (1991) Analysis of human genetic linkage. The Johns Hopkins University Press, London.

Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125:645-654.

Silver J, Buckler CE (1986) Statistical considerations for linkage analysis using recombinant inbred strains and backcrosses. Proc Natl Acad Sci USA. 83:1423-1427.

Smith CAB (1959) Some comments on the statistical methods used in linkage investigations. Am J Hum Genet 11:289-304.

Smith CAB, Sturt E (1976) The peak for the likelihood curve in linkage testing. Ann Hum Genet 39:423-426.

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25-33.

Chapter 3

Evaluation of designs for reference families for livestock linkage mapping experiments

Sijne van der Beek, Ab F. Groen and Johan A.M. van Arendonk

Published in: Animal Biotechnology 4: 163-182 (1993)
Reproduced by permission of Marcel Dekker, Inc.

Abstract

The development of dense linkage maps consisting of highly polymorphic loci for livestock species is technically feasible. However, linkage mapping experiments are expensive as they involve many animals and marker typings per animal. To minimize costs of developing linkage maps for livestock species, optimizing designs for mapping studies is necessary. This study provides a general framework for evaluating the efficiency of designs for reference families consisting of two- or three- generation full-sib or half-sib families selected from a segregating population. The influence of number of families, number of offspring per family, family structure (either half-sib or full-sib) and marker polymorphism is determined. Evaluation is done for two markers with a recombination rate of .20 and for a marker and a dominant single gene with a recombination rate of .20. Two evaluation criteria are used: expected maximum lod score for detection of linkage and accuracy of an estimated recombination rate defined as probability that the true recombination rate is in an interval around the estimated recombination rate. First, for several designs the contribution of reference families to expected maximum lod score and accuracy is given. Second, the required number of families in a design to obtain a certain value for the evaluation criteria is calculated, when number of offspring per family, family structure and marker polymorphism are specified. The required numbers increase when designs are optimized not only for expected maximum lod score but also for accuracy. The required number of animals to map a dominant single gene is very large. Therefore, a set of reference families should be designed for strictly mapping marker loci. Examples illustrate how tabulated results can be generalized to determine the values for a wide range of designs containing two- or three-generation full-sib or half-sib families.

Key words: gene mapping - design - segregating population - dominant loci - codominant loci

Introduction

The development of dense linkage maps consisting of highly polymorphic marker loci is technically feasible due to the discovery of classes of hypervariable loci and developments in molecular technology, especially PCR technology. This has been exploited fully for the human and mouse resulting in dense linkage maps (Dietrich *et al.* 1992; Weissenbach *et al.* 1992). The number of highly polymorphic markers mapped on public domain linkage maps in livestock is limited (poultry: Bumstead and Palyga 1992; cattle: Barendse *et al.* 1992; Womack 1993; pigs: Andersson *et al.* 1993), despite the many research groups working on mapping genes. Major efforts are still needed for (1) isolating markers and (2) mapping markers.

Mapping experiments are expensive and time consuming as they involve reference families with many animals and many marker typings per animal. Optimizing designs for reference families is necessary to minimize costs and maximize the efficiency of a design of a given size.

Methods to evaluate designs for human mapping experiments have been described and applied by Ott (1991). Polymorphism has a major influence on the efficiency of designs (e.g.,

Ott 1991). Polymorphism determines the probability that a parent is heterozygous at two loci, this is necessary for identification of the linkage state in the gametes transmitted by the parent. True recombination rate between loci is a second important factor. The number of individuals needed to significantly distinguish an estimated recombination rate from .5 increases with increasing true recombination rate. A third major factor is the ratio between the number of typed animals and the number of observed segregations of gametes, which depends on the number of offspring per parent. White *et al.* (1985) compared a design with several three-generation families to a design with one multigeneration family and found the three-generation family structure to be more efficient. Using the results of human genetic studies, general outlines for the use of reference families for mapping experiments in livestock were given by Hetzel (1991). However, no study is available that evaluates the design of reference families taking into account the locus polymorphism in the segregating population and the structure of the reference families. Furthermore emphasis has been on efficiency for detection of linkage and little attention has been paid to efficiency for accurate estimation of recombination rates.

Van der Beek and Van Arendonk (1993) determined the influence of some factors on the efficiency of a mapping experiment using full-sib reference families selected from a segregating population. They presented methods to calculate two criteria to evaluate the design of reference families: expected maximum lod score and accuracy. Expected maximum lod score assesses the efficiency of a design to detect linkage and is for example used to evaluate designs for linkage analysis between markers and a dominant locus for a quantitative trait (Boehnke 1990). Expected lod score, as used by Ott (1991) and Hodge (1992), and expected *maximum* lod scores differ by a constant value (Van der Beek and Van Arendonk 1993). Accuracy is defined as the probability that the true recombination rate is in a specific interval around the estimated recombination rate and can be calculated from the Fisher information used by Ott (1991).

This study provides a framework for evaluating and optimizing the design of reference families selected from a segregating population to estimate and detect linkage between either two marker loci or a marker locus and a dominant single gene. The expected maximum lod scores and Fisher information for two-generation full-sib and two-generation half-sib families with varying numbers of offspring are given for different values of polymorphism at marker loci and dominant single genes. Results for families of different sizes can be used to obtain expected maximum lod scores and accuracies for any design containing unrelated *two* and/or *three* generation full-sib and/or half-sib families of any size and with any degree of polymorphism at the marker loci. In addition, it is shown how to use these results to determine the number of families of a certain size required to obtain a specified expected maximum lod score or a desired accuracy. Various examples illustrate the calculations.

Methods

Reference families from segregating populations. Reference families are selected from a segregating population which is assumed to be in Hardy-Weinberg equilibrium and linkage equilibrium (Falconer 1989) for marker loci and single genes. Alleles of marker loci are assumed to be codominant. Four different types of reference families are considered:

(1) two-generation full-sib families with sire, dam and offspring typed for marker loci and (dominant) single genes;

(2) two-generation half-sib families where the sires are mated to many dams and all dams have one offspring; sires and offspring are typed, dams are not typed;

(3) three-generation full-sib families with all grandparents, parents and offspring typed;

(4) three-generation half-sib families with parents of the sire, sire and half-sib offspring typed.

For three-generation families it is assumed that linkage phase in parents can be inferred from the genotypes of the grandparents. Probability of a parent being double heterozygous for a pair of loci depends on the polymorphism at both loci.

Polymorphism of a locus will be defined as the number of alleles at a locus and their frequencies. Heterozygosity is another parameter that describes a locus but it is less specific because for a given heterozygosity the number of marker alleles and their frequencies can vary.

Expected maximum lod score. The lod score (Morton 1955) is commonly used to test significance of linkage. The lod score expresses the likelihood of observed data given the estimated recombination rate relative to the likelihood of observed data given a recombination rate of .5. Linkage is significant at a level of $p < .05$ when the lod score is larger than 3 (Morton 1955).

To determine the efficiency of a design for detecting linkage, the expected value of the maximum lod score is used. Expected maximum lod score is calculated according to Van der Beek and Van Arendonk (1993). Expected maximum lod score depends on the true recombination rate. Results in this study are given for a true recombination rate of .20 between either two marker loci or a marker locus and a dominant single gene. This recombination rate corresponds with the suggested recombination rate for marker loci (Botstein *et al.* 1980).

For two-generation designs the expected maximum lod score (E[Z($\hat{\theta}$)]) is equal to a constant (.217) plus the sum of the contributions of NF families (Van der Beek and Van Arendonk (1993)):

$$E[Z(\hat{\theta})] = .217 + \sum_{i}^{\#\ families} \text{contribution of family i}$$

The constant of .217 is equal to the difference between expected lod score and expected maximum lod score.

The contribution of a family depends on the number of offspring in the family and is, for a given size, calculated as the weighted sum over all possible parental genotype combinations. The probabilities for parental genotype combinations are calculated from the polymorphism at the loci.

The contributions of families to E[Z($\hat{\theta}$)] are calculated for two-generation full-sib and two-generation half-sib families in which the recombination rates between two markers are estimated and in which the recombination rates between a marker and a dominant single gene are estimated. The number of offspring per family was varied to include families representative of different livestock species and to provide situations which allowed for the calculation of the contribution of an additional offspring in an average family.

The contribution of a family to E[Z($\hat{\theta}$)] can be used to calculate the E[Z($\hat{\theta}$)] of a design and to calculate the number of families needed to obtain a desired level of E[Z($\hat{\theta}$)]. When desired E[Z($\hat{\theta}$)] is 3, the summed contribution over all families has to be 3 minus the constant .217. When the contribution of a family under the conditions of the design (family size, family structure and marker polymorphism) is Y, then the required number of families is (3-.217) / Y.

Accuracy. Estimates which are significantly different from .5 but which are inaccurate have limited value. Accuracy is estimated by the probability that the true recombination rate is in a specified interval around an estimated recombination rate which is significantly different from .5 (Van der Beek and Van Arendonk 1993). Accuracy can be approximated using Fisher information (Van der Beek and Van Arendonk 1993). Fisher information is a measure of the amount of information in a design for estimating a parameter (in this study the recombination rate) and is approximately equal to one over the variance of the estimated parameter (Kendall and Stuart 1978). Fisher information is calculated from the second derivative of the likelihood function which is used to estimate the recombination rate (Kendall and Stuart 1978). The approximate method assumes the true recombination rate to have a normal distribution with a mean equal to the estimated recombination rate and a variance of one over the Fisher information (Van der Beek and Van Arendonk 1993). For estimated

recombination rates which are not significantly different from .5, the assumption of a normally distributed true recombination rate is not justified (Van der Beek and Van Arendonk 1993). In the present study, the approximate method is used to calculate the accuracy of estimates which are significantly different from .5. When Fisher information for a design is known accuracy can be obtained from tables of the standard normal distribution.

In the remainder of this paper accuracy is defined in a more strict sense as the probability that the true recombination rate is between .15 and .25 when the estimated recombination rate is .20. The contribution of a family to Fisher information depends on family structure, the number of offspring in the family, marker polymorphism and recombination rate but does not depend on the other families. For more details on the methods used see Van der Beek and Van Arendonk (1993).

Results

Expected maximum lod score for linkage between two marker loci. Table 1 gives contributions to expected maximum lod score for two-generation full-sib families with different numbers of offspring per family and levels of marker polymorphism. For marker loci with two equiprobable alleles, the contribution of a family with 4 full-sib offspring was .029, while with 20 full-sib offspring the contribution was .365, i.e., contribution increased 12 times for a five fold increase in the number of offspring. The contribution to $E[Z(\hat{\theta})]$ per additional offspring was constant when the family had more than 20 offspring. The linear increase (dEZ) is given in the last column of Table 1 and can be used to calculate the contribution of full-sib families with 20 or more offspring.

Marker polymorphism (defined as the number of alleles and their frequencies) had a clear effect on the contribution to $E[Z(\hat{\theta})]$ of a full-sib family. Increasing the number of equiprobable alleles from 2 to 10 gave a five fold increase in the contribution to $E[Z(\hat{\theta})]$ of a full-sib family. The effect of marker polymorphism is caused by an increase in heterozygosity and in the number of marker alleles. The effect of number of alleles was largest for a heterozygosity of 50% (Table 1). The contribution to $E[Z(\hat{\theta})]$ almost doubled when the number of alleles increased from two to four. For a heterozygosity of 90% increasing the number of alleles from 10 to 20 had a marginal effect. The influence of the number of alleles at a constant level of heterozygosity can be explained as follows. If two parents have the same heterozygous genotype, then the alleles which they transmit to heterozygous offspring can not be identified. As a consequence these offspring contribute little or nothing to the estimate of the recombination rate. With an increasing number of alleles and a given level of heterozygosity, the probability of both parents having the same heterozygous genotypes decreases. The effect of the number of alleles on contribution to $E[Z(\hat{\theta})]$ for a given level of heterozygosity also applies when half-sib families are used to

estimate recombination between two markers and when recombination between a marker and a dominant single gene are estimated for full- or half-sib families. In the following tables results are given for one number of alleles per level of heterozygosity.

Half-sib family contributions to $E[Z(\hat{\theta})]$ for estimates of linkage between two marker loci are given in Table 2. For two equiprobable alleles, the contribution to $E[Z(\hat{\theta})]$ was .003 for 4 offspring and .050 for 20 offspring, i.e. a 16 fold increase in $E[Z(\hat{\theta})]$ for a five fold increase in the number of offspring. For 10 equiprobable alleles and 20 offspring the contribution was .858. Increasing the number of alleles from 2 to 10 caused a 16 fold increase in the contribution to $E[Z(\hat{\theta})]$ which was larger than the seven fold increase caused by the same increase in the number of alleles for a full-sib family (Table 1). For equal marker polymorphism and an equal number of offspring, the contribution of a half-sib family was always less than half that of a full-sib family (Table 1).

Tables 1 and 2 can be used to obtain the $E[Z(\hat{\theta})]$ of any design given a true recombination rate of .20. This is illustrated in example 1.

Table 1 Contributions to $E[Z(\hat{\theta})]$ of two-generation full-sib families with 4 to 40 offspring and contributions per additional offspring (dEZ) for two marker loci with a true recombination rate of .20 and varying polymorphism

No. alleles[a]	Hetero-zygosity	Number of offspring					dEZ
		4	6	10	20	40	
2e	50	.029	.060	.137	.365	.855	*.025*
3	50	.051	.101	.219	.547	1.23	*.034*
4	50	.056	.111	.240	.598	1.36	*.038*
3	60	.073	.144	.311	.776	1.74	*.048*
4	60	.081	.159	.344	.855	1.91	*.053*
3e	67	.093	.181	.387	.964	2.16	*.060*
4	70	.112	.220	.473	1.17	2.61	*.072*
5	70	.116	.229	.492	1.22	2.71	*.075*
4e	75	.131	.257	.551	1.36	3.04	*.084*
5e	80	.155	.306	.658	1.63	3.62	*.100*
10	80	.159	.318	.688	1.70	3.77	*.104*
20	80	.160	.322	.697	1.72	3.82	*.105*
10e	90	.204	.408	.884	2.18	4.85	*.134*
20	90	.204	.411	.891	2.20	4.88	*.134*

[a] Marker polymorphism is determined by number of alleles and level of heterozygosity. An 'e' is added to number of alleles when alleles are equiprobable; otherwise all but one alleles are equiprobable.

Table 2 Contributions to $E[Z(\hat{\theta})]$ for two-generation half-sib families with 4 to 40 offspring and contribution per additional offspring (dEZ) for two marker loci with a true recombination rate of .20 and 2 to 10 equiprobable alleles

No. alleles	Hetero-zygosity	Number of offspring					dEZ
		4	6	10	20	40	
2	50	.003	.006	.015	.050	.140	*.005*
3	67	.013	.029	.072	.210	.529	*.016*
4	75	.026	.056	.133	.370	.891	*.026*
5	80	.037	.079	.185	.500	1.18	*.034*
10	90	.071	.147	.332	.858	1.95	*.055*

Table 3 Contributions to $E[Z(\hat{\theta})]$ of two-generation full-sib families with 4 to 40 offspring and contributions per additional offspring (dEZ) when estimating recombination between a marker locus and a dominant single gene. The marker locus has 2 to 10 equiprobable marker alleles (No. alleles) and the dominant gene varying frequencies of the dominant allele (f_{dom}). The true recombination rate between the marker and the dominant gene is .20

No. alleles	f_{dom}	Number of offspring					dEZ
		4	6	10	20	40	
2	.2	.017	.037	.085	.223	.519	*.015*
	.5	.012	.028	.066	.181	.438	*.013*
	.8	.002	.005	.013	.037	.095	*.003*
3	.2	.030	.064	.144	.368	.837	*.023*
	.5	.022	.049	.113	.303	.705	*.020*
	.8	.004	.009	.022	.063	.152	*.004*
4	.2	.036	.077	.172	.439	.999	*.028*
	.5	.027	.057	.133	.355	.841	*.024*
	.8	.005	.010	.025	.072	.181	*.005*
5	.2	.041	.086	.192	.486	1.09	*.030*
	.5	.030	.066	.152	.402	.922	*.026*
	.8	.006	.012	.030	.084	.198	*.006*
10	.2	.047	.099	.221	.562	1.27	*.035*
	.5	.034	.074	.171	.455	1.07	*.031*
	.8	.006	.013	.032	.092	.232	*.007*

Example 1 *Design: both marker loci have six equiprobable alleles and three two generation full-sib families each with 22 offspring are used. The contribution to $E[Z(\hat{\theta})]$ of a family with 22 full-sib offspring is not given directly in Table 1. It can be obtained from Table 1 as follows. The contribution to $E[Z(\hat{\theta})]$ of a family with 20 offspring is 1.63 for five equiprobable alleles. The contribution per additional offspring (dEZ) is .10. Therefore, the contribution of a full-sib family with 22 offspring for five equiprobable alleles is 1.63 + (2 × .100) = 1.83. Likewise, the contribution of a full-sib family with 22 offspring for 10 equiprobable alleles is 2.18 + (2 ×.134) = 2.45. A linear relationship between heterozygosity and $E[Z(\hat{\theta})]$ is assumed in the determination of the $E[Z(\hat{\theta})]$ for 6 alleles from the $E[Z(\hat{\theta})]$ for 5 and 10 alleles. For 5, 6 and 10 equiprobable alleles, heterozygosities are .80, .8333, and .90, respectively. The contribution of a full-sib family with 22 offspring and 6 equiprobable alleles is 1.83 + [(.8333-.80)/(.90-.80)]×(2.45-1.83) = 2.04. Thus $E[Z(\hat{\theta})]$ for the recombination between two markers with a true recombination rate of .20 for this design is .217 + (3 × 2.04) = 6.34*

Table 4 Contributions to $E[Z(\hat{\theta})]$ of two-generation half-sib families with 4 to 40 offspring and contributions per additional offspring (dEZ) when estimating recombination between a marker locus and a dominant single gene. The marker locus has 2 to 10 equiprobable alleles (No. alleles) and the dominant gene varying frequencies of the dominant allele (f_{dom}). The true recombination rate between the marker and the dominant gene is .20

No. alleles	f_{dom}	Number of offspring					dEZ
		4	6	10	20	40	
2	.2	.002	.006	.012	.049	.129	*.0041*
	.5	.001	.003	.007	.025	.077	*.0028*
	.8	.000	.000	.000	.002	.008	*.0003*
3	.2	.006	.014	.033	.099	.249	*.0077*
	.5	.003	.006	.016	.054	.155	*.0054*
	.8	.000	.000	.001	.005	.018	*.0007*
4	.2	.008	.019	.046	.131	.324	*.0097*
	.5	.004	.009	.023	.073	.206	*.0067*
	.8	.000	.001	.002	.007	.025	*.0009*
5	.2	.010	.023	.054	.153	.373	*.0111*
	.5	.004	.010	.027	.086	.240	*.0080*
	.8	.000	.000	.002	.008	.029	*.0011*
10	.2	.015	.031	.074	.202	.483	*.0141*
	.5	.006	.014	.037	.117	.317	*.0100*
	.8	.000	.001	.003	.012	.040	*.0014*

Expected maximum lod score for linkage between a marker locus and a dominant single gene. In Tables 3 and 4 the contributions of families to $E[Z(\hat{\theta})]$ for estimating the recombination rate between a marker and a dominant single gene are given.

Table 3 gives full-sib family contributions to $E[Z(\hat{\theta})]$. For a dominant single gene, which had alleles D and d, different frequencies of the dominant allele were studied. For two marker alleles and a frequency of D of .5 the contribution to $E[Z(\hat{\theta})]$ was .012 for 4 offspring and .181 for 20 offspring, i.e., increasing number of offspring five times caused a 15 fold increase in $E[Z(\hat{\theta})]$. Full-sib family contributions to $E[Z(\hat{\theta})]$ for estimating recombination between a marker and a dominant single gene were smaller than contributions for estimating the recombination rate between two markers (Table 1). The reasons for this are that heterozygosity of a dominant gene is limited and that heterozygous Dd offspring can not be distinguished from homozygous DD offspring. A family is uninformative if either of the parents is homozygous DD. The probability that a parent is homozygous DD increases with the population frequency of D. Table 3 shows the effect of the frequency of the dominant allele on the contribution on $E[Z(\hat{\theta})]$. For two equiprobable marker alleles and four full-sib offspring, contributions to $E[Z(\hat{\theta})]$ were .017, .012 and .002 for frequencies of D of .2, .5 and .8, respectively.

Table 4 gives contributions to $E[Z(\hat{\theta})]$ for half-sib families. The contributions were low compared to full-sib family contributions. Full-sib family and half-sib family contributions are most easily compared by looking at the contribution per additional offspring (dEZ). For two marker alleles and a frequency of D of .2, dEZ was .0041 for a half-sib family (Table 4) and .015 for a full-sib family (Table 3). For two marker alleles and a frequency of D of .8, dEZ was .0003 for a half-sib family compared to a dEZ of .003 for a full-sib family (Table 3). For a half-sib family the effect of increasing the frequency of D was larger than for a full-sib family.

The use of Tables 3 and 4 to obtain $E[Z(\hat{\theta})]$ is illustrated in example 2.

Example 2 *Design: a marker has four equiprobable marker alleles and the frequency of the dominant allele D is .5. Four half-sib families with 37 offspring and three half-sib families with 61 offspring are used for estimating linkage between a marker and a dominant single gene. Given the marker polymorphism and the frequency of the single gene, a half-sib family with 37 offspring has a contribution of .073 (= contribution of 20 offspring) + (17 × .0067) = .187. A half-sib family with 61 offspring has a contribution of .206 (=contribution of 40 offspring) + (21 × .0067) =.347. Thus, for this design, $E[Z(\hat{\theta})]$ is .217 + (3 × .347) + (4 × .187) = 2.006.*

Required number of reference families for a specified expected maximum lod score. The number of families required for a specific $E[Z(\hat{\theta})]$ given marker polymorphism, frequency of D and number of offspring per family, can be calculated using Tables 1 - 4. Similarly, the required number of offspring per family for a specific $E[Z(\hat{\theta})]$ given marker polymorphism and frequency of the dominant allele and number of families can be calculated. This is illustrated in example 3.

Example 3 *Let desired $E[Z(\hat{\theta})]$ be 3. Full-sib families have 10 offspring. Marker loci have three equiprobable alleles. A family with 10 offspring contributes .387 to $E[Z(\hat{\theta})]$ when marker loci have three equiprobable alleles (Table 1). To obtain an $E[Z(\hat{\theta})]$ of 3, (3-.217)/.387 ≈ 7 families are necessary. This means 7 times 10 offspring and 2 parents or 84 animals are needed. Using half-sib families with 10 offspring, the number of families needed equals (3-.217)/.072 = 39 and the required number of animals is 429.*

Table 5 shows the required number of animals for an $E[Z(\hat{\theta})]$ of 3 in different situations. Marker polymorphism, frequency of the dominant allele, number of offspring per family and family structure were varied. For number of offspring, values of 10 and 100 offspring were used because for many species a full-sib family with 10 offspring or a half-sib family with 100 offspring are realistic. Values for four offspring were calculated to illustrate the effect of using small families. When families with four offspring were used the required number of animals was twice the number of animals required for families with 10 offspring.

Table 5 Required number of animals (parents + offspring) for $E[Z(\hat{\theta})] = 3$. Heterozygosity of marker loci, maximum number of offspring (No. offspring per family) per two-generation family and family structure are varied for (1) two marker loci with equal polymorphism, (2) a marker locus and a single gene with a dominant allele with a frequency of .2 and (3) a marker locus and a single gene with a dominant allele with a frequency of .8

Heterozygosity[a]	No. offspring per family[b]	full-sib structure		half-sib structure	
(1) two marker loci					
50 (2)	4	576		4640	
	10	240		2046	
	100	126	{61}	651	{92}
75 (4)	4	126		535	
	10	60		231	
	100	39	{37}	120	{59}
90 (10)	4	84		195	
	10	36		99	
	100	27	{25}	56	{55}
(2) one marker locus and one gene with a dominant allele with frequency .2					
50 (2)	10	396		2552	
	100	202		752	{93}
75 (4)	10	192		660	
	100	102		303	
90 (10)	10	156		418	
	100	85	{83}	202	
(3) one marker locus and one gene with a dominant allele with frequency .8					
50 (2)	10	2568		> 30000	
	100	1020		10807	
75 (4)	10	1332		15301	
	100	612		3535	
90 (10)	10	1044		10197	
	100	408		2222	

[a] Loci with equiprobable alleles; between parenthesis the number of equiprobable alleles which leads to specified heterozygosity

[b] Maximum number of offspring per family; when $E[Z(\hat{\theta})]$ is smaller than 3 for x families with the maximum number of offspring each but $E[Z(\hat{\theta})]$ is considerably larger than 3 for $x+1$ families with the maximum number of offspring, then $x+1$ families with each less than the maximum number of offspring are required for an $E[Z(\hat{\theta})]$ of 3. The number of offspring per family is given between braces after the required number of animals when it is less than the maximum number allowed.

The required number of animals varied widely, e.g., one full-sib family with 25 offspring was sufficient when two marker loci with 10 equiprobable alleles each were considered. For full-sib families with 10 offspring, a marker with two equiprobable alleles and a single gene with a dominant allele with frequency of .8, 2568 animals were needed. To compare a full-sib and a half-sib structure consider full-sib families with 10 and half-sib families with 100 offspring. Required number of animals was always lower with a full-sib structure (Table 5).

Contribution of families to Fisher information. Contributions to Fisher information of full-sib families or half-sib families for estimates of the recombination rate between two marker loci are shown in Tables 6 and 7. The full-sib family contribution to Fisher information was 6 for two alleles and four offspring, and 41 for two alleles and 20 offspring, i.e. Fisher information increased seven times for a five fold increase in the number of offspring. Values of 6 and 41 for Fisher information correspond to approximated standard errors on the estimated recombination rate of $1/\sqrt{6} = .408$ and $1/\sqrt{41} = .156$. These standard errors show that precision is low when Fisher information is 6 or 41. This is even more so if we take into account that this approximation underestimates the standard error for small values of Fisher information (Van der Beek and Van Arendonk 1993). Half-sib families with 4 or 20 offspring contributed .5 and 7 to Fisher information, i.e., Fisher information increased 14 times for a 5 fold increase in the number of offspring. Marker polymorphism had a distinct effect on Fisher information, e.g., the contribution to Fisher information per additional full-sib offspring was 2.1 for 2 marker alleles and 10.0 for 10 marker alleles; the contribution to Fisher information per additional half-sib offspring was .4 for 2 marker alleles and 4.1 for 10 marker alleles.

For estimates of recombination between a marker locus and a dominant single gene, full-sib and half-sib family contributions to Fisher information are given in Tables 8 and 9. For a frequency of the dominant allele D of .2, contributions per additional full-sib offspring were .7, 1.4 and 2.2 and contributions per additional half-sib offspring were .32, .70 and .99, for 2, 4 and 10 marker alleles, respectively. The contribution of a half-sib family was half the contribution of a full-sib family and the contribution for 2 alleles was one third of the contribution for 10 alleles. Frequency of D had a large effect, e.g., for two marker alleles, contributions per additional full-sib offspring were .7, .7 and .2 and contributions per additional half-sib offspring were .32, .26 and .05 for frequencies of D of .2, .5 and .8, respectively. The relative decrease in Fisher information resulting from an increase in the frequency of D from .5 to .8 was larger for a half-sib design than for a full-sib design. The effects of marker polymorphism, frequency of D and number of offspring on Fisher information were similar to the effects of these parameters on $E[Z(\hat{\theta})]$.

Table 6 Contributions to Fisher information of two-generation full-sib families with 4 to 40 offspring and Fisher information per additional offspring (dFI) for two marker loci with a true recombination rate of .20 and 2 to 10 equiprobable alleles (No. alleles)

No. alleles	Hetero-zygosity	Number of offspring					dFI
		4	6	10	20	40	
2	50	6	10	19	41	84	*2.1*
3	67	14	25	46	96	192	*4.8*
4	75	20	34	63	130	261	*6.5*
5	80	24	11	74	152	306	*7.7*
10	90	32	54	98	200	401	*10.0*

Table 7 Contributions to Fisher information of two-generation half-sib families with 4 to 40 offspring and Fisher information per additional offspring (dFI) for two marker loci with a true recombination rate of .20 and 2 to 10 equiprobable alleles (No. alleles)

No. alleles	Hetero-zygosity	Number of offspring					dFI
		4	6	10	20	40	
2	50	.5	1	3	7	15	*.4*
3	67	2.5	5	10	24	49	*1.3*
4	75	5	9	18	39	79	*2.0*
5	80	7	12	23	51	102	*2.5*
10	90	12	21	39	82	164	*4.1*

Table 8 Contributions to Fisher information of two-generation full-sib families with 4 to 40 offspring and Fisher information per additional offspring (dFI) when estimating recombination between a marker locus and a dominant single gene. The marker has 2 to 10 equiprobable alleles (No. alleles) and the dominant gene varying frequencies of the dominant allele (f_{dom}). The true recombination rate between the marker and the dominant gene is .20

No. alleles	f_{dom}	Number of offspring					dFI
		4	6	10	20	40	
2	.2	1	2	5	12	25	*.7*
	.5	1	2	5	12	26	*.7*
	.8	.2	.5	1	3	7	*.2*
4	.2	3	5	11	25	53	*1.4*
	.5	2	5	11	29	64	*1.8*
	.8	.5	1	3	8	19	*.6*
10	.2	3	7	15	33	70	*2.2*
	.5	3	7	16	39	87	*2.3*
	.8	.7	2	4	11	27	*.8*

Accuracy. The probability that the true recombination rate is between .15 and .25 given an estimated recombination rate of .20 was calculated for different values of Fisher information (Table 10). The relationship between Fisher information and probability is non-linear. This is not surprising since Fisher information has a non-linear relationship with the variance of the estimate. As a consequence, improving accuracy from .60 to .70 required a smaller increase in the number of animals than improving accuracy from .70 to .80.

Table 9 Contributions to Fisher information of two-generation half-sib families with 4 to 40 offspring and Fisher information per additional offspring (dFI) when estimating recombination between a marker and a dominant single gene. The marker has 2 to 10 equiprobable marker alleles (No. alleles) and the dominant gene has varying frequencies of the dominant allele (f_{dom}). The true recombination rate between the marker and the dominant gene is .20

No. alleles	f_{dom}	Number of offspring					dFI
		4	6	10	20	40	
2	.2	.5	1.1	2.3	5.6	11.9	*.32*
	.5	.2	.5	1.3	3.6	8.8	*.26*
	.8	.0	.0	.1	.4	1.4	*.05*
4	.2	1.5	2.9	5.9	13.2	27.1	*.70*
	.5	.8	1.6	3.6	9.3	21.4	*.61*
	.8	.1	.1	.4	1.3	3.7	*.12*
10	.2	2.5	4.6	8.9	19.3	39.1	*.99*
	.5	1.2	2.5	5.6	14.0	30.5	*.83*
	.8	.1	.2	.6	2.1	5.6	*.18*

Table 10 Accuracy [$P(.15 < \theta_t < .25 \mid \hat{\theta} = 0.20, Z(\hat{\theta}) > 3)$] and standard error ($\sigma$) for different values of Fisher information

Fisher information	σ	$P(.15 < \theta_t < .25 \mid \hat{\theta} = .20, Z(\hat{\theta}) > 3)$
50	.141	.27
100	.100	.38
182	.074	.50
200	.071	.52
282	.059	.60
400	.050	.68
430	.048	.70
657	.039	.80
800	.035	.84
1082	.030	.90
1600	.025	.95

Example 4 illustrates the use of Tables 6 - 10 to calculate the accuracy of a design.

Example 4 *Design: marker loci have 10 equiprobable alleles and three half-sib families with 100 offspring each are used. For this design the accuracy of an estimated recombination rate between two marker loci has to be obtained. To determine accuracy, first Fisher information is calculated. The contribution of one half-sib family with 100 offspring is 164 + (60 4.1) = 410 (Table 7). Fisher information for the design is 3 × 410 = 1230. The variance of an estimated recombination rate of .20 is, therefore, expected to be 1/1230 and the standard error on the estimated recombination rate is √(1/1230) = .028. Table 10 indicates that when Fisher information is 1230, accuracy is between .90 and .95.*

Number of animals required for a certain accuracy. The Fisher information required for a certain accuracy can be derived from Table 10. After that Tables 6-9 can be used to determine the required number and size of reference families. For a number of designs, Table 11 gives the number of animals required such that 80% of the significant estimated recombination rates of .20 will correspond to a true recombination rate with a value between .15 and .25. For two markers and a full-sib structure with 10 offspring (or a half-sib structure with 100 offspring), the required number of animals for an accuracy of .80 decreased from 420 (1717) for a heterozygosity of 50% to 84 (202) for a heterozygosity of 90%. For a full-sib structure with 10 offspring per family (or a half-sib structure with 100 offspring per family), the required number of animals for an $E[Z(\hat{\theta})]$ of 3 were 240 (651) for a heterozygosity of 50% and 27 (56) for a heterozygosity of 90%. The required number of animals were respectively 1.8, 2.6, 3.1 and 3.6 times higher for an accuracy of .80 than for an $E[Z(\hat{\theta})]$ of 3. The influence of family structure, marker polymorphism and frequency of the dominant allele on required number of animals for an accuracy of .80 was similar to the influence of these variables on required number of animals for an $E[Z(\hat{\theta})]$ of 3. However, as described above and as can be seen from Tables 5 and 11, two to four times more animals were required for an accuracy of .80 than for an $E[Z(\hat{\theta})]$ of 3.

Expected maximum lod score and Fisher information for three generation designs. In three-generation designs linkage phase in parents can be inferred from grandparental genotypes. When linkage phase in parents is known, increases in $E[Z(\hat{\theta})]$ and Fisher information per offspring are no longer dependent on the number of offspring in a family. As a consequence $E[Z(\hat{\theta})]$ and Fisher information can be derived from the dEZ and dFI values which have been tabulated (Van der Beek and Van Arendonk 1993). Example 5 illustrates the use of dEZ and dFI for a three-generation design.

Example 5 *Design: four three-generation full-sib families each with four grandparents, two parents and 25 offspring; marker loci with three equiprobable alleles. For marker loci with 3 equiprobable alleles, dEZ is .060 (Table 1) and dFI is 4.8 (Table 6). The number of offspring in the experiment is 100. E[Z(θ̂)] is .217 + 100 × .06 = 6.217 and Fisher information is 100 × 4.8 =480. Accuracy ≈ .725 (Table 10).*

Table 11 Required number of animals for $P(.15 < \theta_t < .25 \mid \hat{\theta} = .20, Z(\hat{\theta}) > 3) = .80$ for two-generation full-sib families with 10 offspring and two-generation half-sib families with 100 offspring, for (1) two markers with 2 to 10 equiprobable alleles, (2) a marker with two to 10 equiprobable alleles and a single gene with a dominant allele with a frequency of .2 and (3) a marker and a single gene with a dominant allele with a frequency of .8

Heterozygosity	Full-sib structure	Half-sib structure
	(1) two markers	
50 (2)	420	1717
75 (4)	132	404
90 (10)	84	202
	(2) one marker and one gene with a dominant allele with frequency .2	
50 (2)	1572	2121
75 (4)	720	1010
90 (10)	528	707
	(3) one marker and one gene with a dominant allele with frequency .8	
50 (2)	7884	15049
75 (4)	2628	6060
90 (10)	1968	4040

Discussion

Animal requirements for linkage studies can be evaluated on the basis of the results of this study. This is not only useful for designing new experiments but also for fast evaluation of currently used reference families. Evaluations were done for a true recombination rate of .20. To compare different designs, evaluation for one value of true recombination rate is sufficient because rankings are not influenced by true recombination rate. However, accuracy and $E[Z(\hat{\theta})]$ decrease when there is an increase in the true recombination rate. When a design is based on a true recombination rate of .20, the efficiency of the design for estimating and detecting linkage will be sufficient for true recombination rates of .20 and smaller. In order to be able to compare results, a true recombination rate of .20 between a dominant single gene and a marker was used. When a dominant single gene is mapped on a 20 centimorgan map, true recombination rate between the dominant single gene and most proximal marker will be smaller then .20. For a smaller true recombination rate, $E[Z(\hat{\theta})]$ and accuracy for a design will be larger, e.g., for a true

recombination rate of .10, contribution of a family to both E[Z($\hat{\theta}$)] and Fisher information is about twice the contribution of the same family for a true recombination rate of .20 (Van der Beek, unpublished results).

Tabulated results can be used to evaluate various designs using linear interpolation and extrapolation. The contribution of a family to E[Z($\hat{\theta}$)] and Fisher information can be calculated by linear extrapolation because of the linear increase in contribution to E[Z($\hat{\theta}$)] and Fisher information for families with more than 20 offspring. E[Z($\hat{\theta}$)] and Fisher information were given for several family sizes smaller than 20 which can be used to obtain the required E[Z($\hat{\theta}$)] and Fisher information in the area where the increase per additional offspring is not yet linear. A linear relation between heterozygosity and contribution of a family was assumed to obtain the contribution of a family selected from a population with a marker polymorphism unequal to the tabulated levels of marker polymorphism. The various values for the contribution to E[Z($\hat{\theta}$)] for one value of heterozygosity (Table 1) show that in reality the relation between heterozygosity and the contribution of a family is not completely linear. However, the error introduced by the assumption of linearity was small. This can be checked in the various tables. For instance, the contribution to E[Z($\hat{\theta}$)] of two-generation half-sib families with 20 offspring is .370 for four equiprobable alleles (Table 2). This contribution can also be approximated from the contributions for 3 equiprobable alleles (which is .210) and the contribution for 5 equiprobable alleles (which is .500): approximated contribution for 4 alleles is $.210 + (.75-.67)/(.80-.67) \times (.500-.210) = .388$.

Results are presented for two-point linkage analysis. However, the presented results can be used to make inferences about multi-point linkage analysis. First, a multi-point linkage analysis often starts with two-point linkage analysis. Second, variables which influence the efficiency of two-point linkage analysis (true recombination rate, family structure, number of offspring per parent, marker polymorphism), are expected to have the same influence on multi-point linkage analysis.

The number of animals required in an efficiently designed study for mapping codominant marker loci is reasonable. However, many more animals are needed for mapping a dominant single gene. Reference families are used to map many marker loci and genes. They are not selected because the parents are heterozygous for a particular locus. Therefore, reference families have a high probability of being uninformative for a dominant single gene and their use for mapping dominant single genes is therefore limited. Alternatively, a backcross of fully inbred lines can be used. When this backcross is used to map any dominant single gene, then the backcross can not be optimized for a specific dominant single gene. When the F1 animals are heterozygous Dd for a particular dominant single gene and the pure line animals crossed with the F1 are homozygous dd, then about 30 backcross animals are needed for a E[Z($\hat{\theta}$)] of 3. However, a backcross provides no information at all when the F1

animals are homozygous for the single gene or when the pure line animals crossed to the F1 animals are homozygous DD. There is no way to assure that this will not be the case. Thus, for a single backcross some dominant single genes can be mapped efficiently but probably the majority of them can not be mapped at all.

In conclusion, we suggest a set of reference families should be designed for strictly mapping marker loci. For mapping dominant single genes the reference families and additional families from the segregating population should be screened for families informative for the single gene. The procedure described by Van Arendonk *et al.* (1989) can be used for screening a population for animals being heterozygous for the single gene. Informative families should be typed for the single gene and a selected set of equidistant highly polymorphic marker loci.

Designing a mapping experiment is a decision process in which the designer has to: (1) specify the population(s) reference animals can be selected from, (2) specify (expected) marker polymorphism which is determined by the type of marker loci and the criteria used to select individual markers before mapping, (3) define desired values for evaluation criteria, and (4) evaluate the possible designs. It is important to realize that decisions about populations to be used, expected polymorphism and evaluation criteria will influence the design eventually chosen. This study showed that a fast evaluation of designs, taking into account major factors influencing the efficiency of a design, is possible.

Acknowledgements

The authors gratefully acknowledge financial support by Euribrid B.V. The authors thank Dr. C.J. Davies for useful comments.

References

Andersson L, Archibald A, Gellin J, Schook LB (1993) 1st pig gene mapping workshop (PGM1) 7th August, 1992, Interlaken Switzerland. Anim Genet 24:205-216.

Barendse W, Armitage SM, Moore SS, Georges M, Womack JE, Hetzel DJS (1992) Linkage relations of more than 100 genes and highly polymorphic DNA markers in cattle. Anim Genet 23 Suppl 1:82.

Boehnke M (1990) Sample-size guidelines for linkage analysis of a dominant locus for a quantitative trait by the method of lod scores. Am J Hum Genet 47:218-227.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331.

Bumstead N, Palyga J (1992) A preliminary linkage map of the chicken genome. Genomics 13:690-697.

Dietrich W, Katz H, Lincoln SE, Hee-Sup Shin, Friedman J, Dracopoli NC, Lander ES (1992) A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131:423-447.

Falconer DS (1989) Introduction to quantitative genetics. Longman scientific and technical, Essex.

Hetzel DJS (1991) Reference families for genome mapping in domestic livestock. In: Schook LB, Lewin HA, McLaren DG (eds) Gene-mapping techniques and applications. Marcel Dekker, New York.

Hodge SE (1992) Do bilineal pedigrees represent a problem for linkage analysis? Basic principles and simulation results for single-gene diseases with no heterogeneity. Genetic Epidemiology 9:191-206.

Kendall M, Stuart A (1978) The advanced theory of statistics, Vol 2. Griffin, London.

Morton NE (1955) Sequential test for the detection of linkage. Am J Hum Genet 7:277-318.

Ott J (1991) Analysis of human genetic linkage. The Johns Hopkins University Press, London.

Van Arendonk JAM, Smith C, Kennedy BW (1989) Method to estimate genotype probabilities at individual loci in farm livestock. Theor Appl Genet 78:735-740.

Van der Beek S, Van Arendonk JAM (1993) Criteria to optimize designs for detection and estimation of linkage between marker loci from segregating populations containing several families. Theor Appl Genet 86:269-280.

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second generation linkage map of the human genome. Nature 359:794-800.

White R, Leppert M, Bishop DT, Barker D, Berkowitz J, Brown C, Callahan P, Holm T, Jerominski L (1985) Construction of linkage maps with DNA markers for human chromosomes. Nature 313:101-105.

Womack JE (1993) The goals and status of the bovine gene map. J Dairy Sci 76:1199-1203.

Chapter 4

Power of two- and three- generation QTL mapping experiments in an outbred population containing full-sib or half-sib families

Sijne van der Beek, Johan A.M. van Arendonk and Ab F. Groen

Theoretical and Applied Genetics (In Press)

Abstract

QTL mapping experiments involve many animals to be genotyped and performance tested. Consequently experimental designs need to be optimized to minimize costs of data collection and genotyping. This study analyzed the power and efficiency of experiments with two- or three- generation family structures containing full-sib families, half-sib families or both. Focus was on data from one outbred population because the main interest is to locate genes that can be used for within line selection. For a two-generation experiment more animals had to be typed for marker loci to obtain a certain power than for a three-generation experiment. Fewer trait values, however, had to be obtained for a two-generation experiment than for a three-generation experiment. A two- or three- generation family structure with full-sib offspring was more efficient than a two- or three- generation family structure with half-sib offspring. A family structure with full-sib grandoffspring, however, was less efficient than a family structure with half-sib grandoffspring. For the most efficient family structure each pair of parents had full-sib offspring that were genotyped for the marker. For the most efficient family structure each full-sib offspring had half-sib grandoffspring for which trait values were obtained. For a heritability of .1 and 100 grandoffspring per full-sib offspring, for equal power 30 times less marker typings were required for this most efficient family structure than for a two-generation half-sib structure in which marker genotypes and trait values were obtained for half-sib offspring. The effect of heritability and type of analysis (single marker or interval analysis) on the efficiency of a family structure was described. The results of this study help to design QTL mapping experiments in an outbred population.

Key words: QTL mapping - experimental design - statistical power - outbred population

Introduction

Dense linkage maps consisting of highly polymorphic marker loci are available for most livestock species (Andersson *et al.* 1993; Barendse *et al.* 1994; Bishop *et al.* 1994; Crawford *et al.* 1994; Rohrer *et al.* 1994) and can be a powerful tool for mapping quantitative trait loci (QTL) (e.g. Patterson *et al.* 1988; Stuber *et al.* 1992; Andersson *et al.* 1994; Georges *et al.* 1995). QTL mapping experiments involve many animals to be genotyped and performance tested (Soller and Genizi 1978; Weller *et al.* 1990). Consequently, experimental designs need to be optimized to minimize costs of data collection and genotyping. In livestock, QTL mapping experiments can involve data from one outbred population or from a cross between populations (Soller 1991). An experiment involving a cross between populations, will reveal genes that explain the variance between populations, whereas an experiment within one population will reveal genes that explain the variance within a population. Within an outbred population, genetic markers and putative QTLs are expected to be in linkage equilibrium; therefore analysis has to be done within families. Weller *et al.* (1990) computed powers of experiments for balanced two- and three- generation half-sib designs for outbred populations. They quantified the influence on power of size of QTL effect, heritability, family size and number of half-sib families. They focused on dairy cattle breeding in which the half-sib family structure is predominant. In poultry and pigs,

however, full-sib family structures are feasible. For the mapping of markers, fewer animals are needed in an experiment with a full-sib family structure than in an experiment with a half-sib family structure. We expect that for QTL mapping a full-sib family structure has a higher efficiency than a half-sib family structure. The power of an experiment with a two-generation full-sib family structure or a three-generation full-sib family structure is unknown. This study analyzes power and efficiency of experiments with two- or three- generation family structures with full-sib families, half-sib families or both. We focus on data from one outbred population because the main interest is to locate genes that can be used for within line selection.

Methods

Outbred population. Consider an outbred population in genotypic equilibrium for individual loci and genetic equilibrium for any pair of loci. Genetic variance in the population is comprised of the variance due to one quantitative trait locus (QTL) with two codominant alleles, having frequencies p and $1\text{-}p$, and the variance due to polygenic effects. The QTL genotype effects are a, 0 and $-a$. Flanking markers are at .1 M (10 cM) from the QTL. Map distance between the two flanking markers is denoted d. The QTL is at the midpoint between two markers so map distance between marker and QTL is $.5d$. Recombination rate between two marker loci is denoted γ and recombination rate between marker and QTL is denoted r. Map distance and recombination rate are related by the Haldane mapping function:
$\gamma = .5(1-e^{-2d})$ and $r = .5(1-e^{-d})$.
Phenotypic expression of a trait is due to the effect of the QTL, a random normal polygenic effect and a random normal residual effect. We denote polygenic variance with σ_u^2 and residual variance with σ_e^2. Further, $\sigma_p^2 = \sigma_u^2 + \sigma_e^2$ and heritability is $h^2 = \sigma_u^2 / \sigma_p^2$. Note that QTL genotype effects are not included in σ_p^2 or h^2.

Family structures. In this paper, a parent is a first generation animal, an offspring is a second generation animal, and a grandoffspring is a third generation animal.
Five family structures are considered that differ in number of generations and relations between animals. For each family structure a single pedigree will be described. An experiment can involve one or several pedigrees.

HS2: Two-generation half-sib family structure. A sire has several unrelated mates and each mate has one offspring. Marker genotypes are obtained for the sire and the half-sib offspring but not for the mates. Trait values for half-sib offspring are obtained. Weller *et al.* (1990) named a design with this family structure the 'daughter design'.

FS2: Two-generation full-sib family structure. A pair of parents has several full-sib offspring. Each parent, male or female, has one mate. Marker genotypes are obtained for all

animals. Trait values for full-sib offspring are obtained.

HS3: Three-generation half-sib family structure. A sire has several half-sib offspring. Each half-sib offspring is mated to several unrelated animals to produce one half-sib grandoffspring per mate per half-sib offspring. Marker genotypes are obtained for the sire and the half-sib offspring, but not for the mates of the sire, the mates of the half-sib offspring and the half-sib grandoffspring. Trait values for the half-sib grandoffspring are obtained. Weller *et al.* (1990) named a design with this family structure the 'grand-daughter' design.

FS3: Three-generation full-sib family structure. A pair of parents has several full-sib offspring. Each full-sib offspring is mated to one unrelated animal to produce several full-sib grandoffspring per full-sib offspring. Marker genotypes are obtained for the parents and the full-sib offspring but not for the mates of the full-sib offspring and the full-sib grandoffspring. Trait values for the full-sib grandoffspring are obtained.

FSHS: Three-generation full-sib offspring, half-sib grandoffspring family structure. A pair of parents has several full-sib offspring. Each full-sib offspring is mated to several unrelated animals to produce one half-sib grandoffspring per mate per full-sib offspring. Marker genotypes are obtained for the parents and the full-sib offspring but not for the mates of the full-sib offspring and the half-sib grandoffspring. Trait values for the half-sib grandoffspring are obtained.

We consider balanced designs. All parents are heterozygous (Mm) for marker loci and 50% of the offspring inherit marker allele M and 50% of the offspring inherit marker allele m. Each family has the same number of offspring. In a three generation design each offspring has the same number of grandoffspring. Offspring are divided into two groups. In one group are the offspring that inherit marker allele M, and in the other group are the offspring that inherit marker allele m. Grandoffspring are also divided into two groups. In one group are the grandoffspring that descend from offspring that inherit marker allele M from the parent. In the other group are the grandoffspring that descend from offspring that inherit marker allele m from the parent.

Computation of power: single marker analysis. We assume that the QTL mapping experiment is analyzed with a linear model. In the model, the effect of the marker is nested within parent. The model for a design with a HS2 family structure is given as an example:

$$y_{ijk} = s_i + m_{ij} + e_{ijk} \quad (1)$$

where y_{ijk} is the trait value for the k-th offspring inheriting marker allele j of sire i, s_i is the

effect sire i, m_{ij} is the effect of marker allele j of sire i and e_{ijk} is the residual effect of offspring k. For a sire with marker genotype Mm, m_{i1} is the effect of allele M and m_{i2} the effect of allele m.

Let $(m_{i1}- m_{i2})$ be the marker contrast (MC) for sire i. Inferences about the presence of a QTL linked to the marker are based on the marker contrast. The marker contrast is expected to be zero if no QTL is linked to the marker or if a parent is homozygous for the linked QTL. The marker contrast is expected to be *nonzero* if a QTL is linked to the marker and the sire is heterozygous for the linked QTL. Thus, the presence of a linked QTL can be found by testing for significantly nonzero marker contrasts. The square of the marker contrast divided by the square of the standard error (SE) of the marker contrast is used to compute a test-statistic with value (Weller *et al.* 1990):

$$\sum_{i=1}^{n_p} \mathrm{MC}_i^2 / \mathrm{SE}_i^2 \tag{2}$$

where n_p is the number of parents for which a marker contrast is computed (for an experiment with a HS2 family structure, n_p is equal to the number of sires), MC_i is the marker contrast for the i-th parent, SE_i is the standard error of MC_i. If the standard error can be computed from a priori known phenotypic variance then this test statistic is a χ^2 statistic (Geldermann 1975). We assume that phenotypic variance is known and use the χ^2 statistic. Under the null hypothesis of no linked QTL the statistic has a central χ^2 distribution. The null hypothesis is rejected when the statistic is larger than threshold T. Threshold T is the $(1-\alpha)$ percentile of the central χ^2 distribution where α is the type I error. The power of a QTL mapping experiment is equal to the probability that the null hypothesis is rejected, i.e. the probability that the χ^2 statistic exceeds threshold T. Under the alternative hypothesis of a linked QTL, the χ^2 statistic has a noncentral χ^2 distribution. The noncentrality of this distribution depends on the expectation for the marker contrast, the standard error of the marker contrast and the number of parents that are heterozygous for the linked QTL.

Given the definitions and assumptions described above, power of an experiment is computed as (Weller *et al.* 1990):

$$\text{power} = \sum_{x=0}^{n_p} P(x) \times P[\chi^2(\mathrm{NC}(x), n_p) > \mathrm{T}] \tag{3}$$

where x is the number of parents that are heterozygous for the QTL, n_p is the number of parents for which a marker contrast is computed, $P(x)$ is the binomial probability that x out of n_p parents are heterozygous for the QTL, $\chi^2(\mathrm{NC}(x), n_p)$ is a noncentral χ^2 variable with n_p

degrees of freedom and with noncentrality parameter NC(x), NC(x) is the noncentrality parameter for the distribution under the alternative hypothesis given that x parents are heterozygous for the QTL, and $P[\chi^2(NC(x),n_p) > T]$ is the probability that the noncentral χ^2 variable exceeds threshold T. The noncentrality parameter is computed as:

$$NC(x) = x\,E^2(MC)/SE^2(MC) \tag{4}$$

where E^2(MC) is the square of the expectation of a marker contrast for a parent that is heterozygous for the QTL and SE^2(MC) is the square of the standard error of this marker contrast. E^2(MC) and SE^2(MC) depend on the design of an experiment as will be shown below.

If a parent is homozygous at the linked QTL then the marker contrast is expected to be zero. If a parent is heterozygous at the linked QTL then for a two-generation family structure (HS2, FS2) (Soller 1991):

$$E^2(MC\text{-}2) = a^2(1-2r)^2 \tag{5}$$

where E^2(MC-2) is the square of the expected marker contrast for a parent of a two-generation family that is heterozygous for the linked QTL, and a and r are as described earlier.

For a three-generation family structure (HS3, FS3 or FSHS) the marker contrast is the difference between the two groups of grandoffspring. The marker allele of the parent for which the marker contrast is computed is transmitted to 50% of the grandoffspring so the marker contrast of a three-generation design is expected to be half the marker contrast of a two-generation design (Weller *et al.* 1990). If a parent is heterozygous at the QTL then for a three-generation family structure:

$$E^2(MC\text{-}3) = 1/4\ a^2(1-2r)^2 \tag{6}$$

where E^2(MC-3) is the square of the expected marker contrast for a parent of a three-generation family that is heterozygous for the linked QTL. Note that E^2(MC) denoted the square of the expected marker contrast in general, where E^2(MC-2) and E^2(MC-3) are specific notations for two- and three- generation family structures.

Table 1 Squared standard error of the marker contrast (SE^2) for five family structures

Family Structure	SE^2
HS2	$(4-h^2)/n_o$
FS2	$(4-2h^2)/n_o$
HS3	$(.75h^2+(4-h^2)/n_{og})/n_o$
FS3	$(1.25h^2+(4-2h^2)/n_{og})/n_o$
FSHS	$(.25h^2+(4-h^2)/n_{og})/n_o$

h^2 is heritability; n_o is number of offspring; n_{og} is number of grandoffspring per offspring

SE^2(MC), the other parameter necessary to compute the noncentrality parameter, depends on heritability, family structure and number of animals. For a HS2 family structure:

$$SE^2_{HS2} = \text{var}\left(\frac{2}{n_o}\sum_{k=1}^{n_o/2} y_{i1k} - \frac{2}{n_o}\sum_{k=1}^{n_o/2} y_{i2k}\right) = \frac{4-h^2}{n_o} \qquad (7)$$

where n_o is the number of offspring per parent. A full derivation of SE^2_{HS2} is in appendix 1. Table 1 gives the SE^2 of the marker contrast for the five family structures.

Computation of power: interval analysis. Instead of performing single marker analysis, markers can be analyzed in pairs to detect a QTL in the interval between two markers (Lander and Botstein 1989). We assume that a QTL is located at the midpoint of the interval between marker loci M and N. Let a parent be heterozygous for marker loci M and N; the ordered genotype of the parent is MN/mn. Offspring can inherit four marker haplotypes (MN,Mn,mN or mn), two of which are non-recombinant (MN, mn) and two recombinant (Mn, mN), with respect to the two markers. Offspring inheriting a recombinant haplotype provide no information to detect a QTL if that QTL is at the midpoint of the interval (Lander and Botstein 1989). Therefore, information on offspring inheriting non-recombinant haplotypes is used only. This information is analyzed with a linear model. The model for a HS2 family structure is given as an example:

$$y_{ijl} = s_i + h_{ij} + e_{ijl} \qquad (8)$$

where y_{ijl} is the trait value of the l-th offspring inheriting non-recombinant haplotype j of sire i, s_i is the effect of sire i, h_{ij} is the effect of the j-th non-recombinant haplotype of sire i, and e_{ijl} is the residual effect of offspring l. Let $(h_{i1}-h_{i2})$ be the haplotype contrast (HC) for sire i. The presence of a linked QTL is tested by testing for significantly nonzero haplotype

contrasts using the square of the haplotype contrast divided by the square of the standard error of the haplotype contrast. The power of this test is computed using equation (3). For interval analysis, the noncentrality parameter is:

$$NC(x) = x\,E^2(HC)/SE^2(HC) \tag{9}$$

where $E^2(HC)$ is the square of the expected haplotype contrast for a parent that is heterozygous for the QTL and $SE^2(HC)$ is the square of the standard error of this haplotype contrast.

If a parent is homozygous at the linked QTL then the haplotype contrast is expected to be zero. If a parent is heterozygous at the linked QTL then for a two-generation family structure (see appendix 2):

$$E^2(HC\text{-}2) = a^2(1-2r)^2/(1-\gamma)^2 \tag{10}$$

where a, r and γ are described earlier. Using equation (5):

$$E^2(HC\text{-}2) = E^2(MC\text{-}2)/(1-\gamma)^2 \tag{11}$$

If a parent is heterozygous at the QTL then for a three-generation family structure:

$$E^2(HC\text{-}3) = 1/4\,a^2(1-2r)^2/(1-\gamma)^2 \tag{12}$$

and using equation (6):

$$E^2(HC\text{-}3) = E^2(MC\text{-}3)/(1-\gamma)^2 \tag{13}$$

The portion $(1-\gamma)$ of the offspring that inherits a non-recombinant marker haplotype are used to compute the haplotype contrast. All offspring are used to compute the marker contrast. In general, the squared standard error of a mean is inversely proportional to the number of observations used to compute the mean. The standard error of the haplotype contrast of a parent is therefore larger than the standard error of the marker contrast of that same parent. In particular:

$$SE^2(HC) = SE^2(MC)/(1-\gamma) \tag{14}$$

Table 2 Relative Efficiency (RE) between family structures HS2, FS2, HS3, FS3 and FSHS

RE(FS2/HS2)	$2 + h^2 / (2 - h^2)$
RE(HS3/HS2)	$(1-.25\ h^2) / (.75\ h^2 + (4-h^2)/n_{og})$
RE(FS3/HS2)	$(2-.5h^2) / (1.25\ h^2 + (4-2h^2)/n_{og})$
RE(FSHS/HS2)	$(2-.5h^2) / (.25\ h^2 + (4-h^2)/n_{og})$
RE(FS3/HS3)	$2 - (h^2 - 2h^2/n_{og}) / (1.25\ h^2 + (4-2h^2)/n_{og})$
RE(FSHS/HS3)	$2 + h^2 / (.25\ h^2 + (4-h^2)/n_{og})$
RE(FSHS/FS3)	$1 + (h^2 - h^2/n_{og}) / (.25\ h^2 + (4-h^2)/n_{og})$

h^2 is heritability; n_{og} is number of grandoffspring per offspring

Relative efficiency (RE) and relative effect of doubling (RED). Power is determined by n_p, $P(x)$, NC(x) and T (equation (3)). Parameter n_p is defined by the design of an experiment, $P(x)$ depends on n_p and the heterozygosity at the QTL, and T is directly related to α. Each other variable that determines the design of an experiment influences the power of an experiment via the influence it has on NC(x). To measure the relative effect on power of a variable, we define two parameters: Relative Efficiency (RE) and Relative Effect of Doubling (RED). RE is used to compare family structures, to compare interval analysis with single marker analysis or to determine the effect of changing the value of one variable. RE is defined only if designs A and B have equal number of offspring per family and is:

$$RE(A/B) = (NC^A(1) \times \#C^A) / (NC^B(1) \times \#C^B) \tag{15}$$

where RE(A/B) is the efficiency of an experiment with design A relative to the efficiency of an experiment with design B, $NC^Y(1)$ is the noncentrality per marker or haplotype contrast for an experiment with design Y, $\#C^Y$ is the number of marker or haplotype contrasts per family for an experiment with design Y. The use of RE can be illustrated by a simple example. If RE(A/B) is 2 and for experiments A and B the same number of marker contrasts are computed, then experiments A and B have equal power if the number of offspring per experiment is 2 times larger for experiment B than for experiment A.

First, the RE to compare family structures is described. To compare FS2 with HS2:

$$RE(FS2/HS2) = \left(\frac{a^2(1-2r)^2}{(4-2h^2)/n_o} \times 2\right) \Big/ \left(\frac{a^2(1-2r)^2}{(4-h^2)/n_o} \times 1\right) = 2 \times \frac{4-h^2}{4-2h^2} = 2 + \frac{h^2}{2-h^2} \tag{16}$$

Table 3 Efficiency of heritability at y relative to efficiency of heritability at z $\{RE(h^2=y/h^2=z)\}$ for five family structures

Family Structure	RE
HS2	$(4- z) / (4- y)$
FS2	$(2- z) / (2- y)$
HS3	$(.75\ z + (4-z)/n_{og}) / (.75\ y + (4-y)/n_{og})$
FS3	$(1.25\ z + (4-2z)/n_{og}) / (1.25\ y + (4-2y)/n_{og})$
FSHS	$(.25\ z + (4-z)/n_{og}) / (.25\ y + (4-y)/n_{og})$

n_{og} is number of grandoffspring per offspring

In Table 2 are RE among family structures HS2, FS2, HS3, FS3 and FSHS.

A second use of RE is to compare interval analysis (I) with single marker analysis (S). Using equations (4) and (9):

$$RE(I/S) = NC^I(1)/NC^S(1) = \frac{E^2(HC)}{SE^2(HC)} \Big/ \frac{E^2(MC)}{SE^2(MC)} \tag{17}$$

Using equations (11), (13) and (14):

$$\frac{E^2(HC)}{SE^2(HC)} \Big/ \frac{E^2(MC)}{SE^2(MC)} = \frac{E^2(MC)/(1-\gamma)^2}{SE^2(MC)/(1-\gamma)} \Big/ \frac{E^2(MC)}{SE^2(MC)} = \frac{1}{1-\gamma} \tag{18}$$

Third, RE is used to evaluate the effect of changing the value of a variable. We use RE to evaluate the effect of changing heritability and the effect of changing the map distance between two marker loci. For heritability, RE for a HS2 family structure is:

$$RE(h^2=y/h^2=z) = \left(\frac{E^2(MC)}{(4-y)/n_o}\right) \Big/ \left(\frac{E^2(MC)}{(4-z)/n_o}\right) = \frac{4-z}{4-y} \tag{19}$$

Table 3 gives $RE(h^2=y/h^2=z)$ for the five family structures.

RE for map distance between two marker loci depends on the type of statistical analysis. If the QTL is at the midpoint between two marker then for single marker analysis:

$$RE(d=y/d=z) = \left((1-2r_y)^2a^2/SE^2\right)/\left((1-2r_z)^2a^2/SE^2\right)=(1-2r_y)^2/(1-2r_z)^2=e^{-2y}/e^{-2z} \quad (20)$$

where r_y is the recombination rate between marker and QTL if y is the map distance between two marker loci. Recombination rate r_y is computed from map distance y using the Haldane mapping function. For interval analysis:

$$RE(d=y/d=z) = \left((1-2r_y)^2/(1-\gamma_y)\right)/\left((1-2r_z)^2/(1-\gamma_z)\right) = (e^{-2y}/e^{-2z})\times((1+e^{-2z})/(1+e^{-2y})) \quad (21)$$

Relative Effect of Doubling (RED) is a second measure to compare efficiency of designs. It is defined as:

$$RED(y,z) = \{NC(1)|2y,z\}/\{NC(1)|y,2z\} \quad (22)$$

where RED(y,z) is the change in NC due to doubling the value of variable y, relative to the change in NC due to doubling the value of variable z, $\{NC(1)|2y,z\}$ is the value of NC(1) for the design of a certain experiment if the value of y is doubled and $\{NC(1)|y,2z\}$ is the value of NC(1) for the design of the same experiment if the value of z is doubled. Two forms, RED(a^2,n_o) and RED(n_o,n_{og}), will be described. For a HS2 family structure:

$$RED(a^2,n_o) = \left(\frac{(2a^2)\ (1-2r)^2\ \times\ n_o}{4-h^2}\right)\Big/\left(\frac{a^2\ (1-2r)^2\ \times\ 2n_o}{4-h^2}\right) = 1 \quad (23)$$

Similarly, it can be shown that RED(a^2,n_o) is 1 for all five family structures. For all family structures, doubling a^2, and thus doubling the variance due to the QTL, has the same effect on the noncentrality parameter, and thus power, as doubling the number of offspring per family. The second form, RED(n_o,n_{og}) is defined for designs with a three generation family structure. For a HS3 family structure:

$$RED(n_o,n_{og}) = \frac{E^2(MC-3)\ \times\ 2n_o}{.75h^2+(4-h^2)/n_{og}}\Big/\frac{E^2(MC-3)\ \times\ n_o}{.75h^2+(4-h^2)/(2n_{og})} = 1+\frac{.75h^2}{.75h^2+(4-h^2)/n_{og}} \quad (24)$$

In Table 4 are the RED(n_o,n_{og}) for HS3, FS3 and FSHS family structures.

Table 4 Effect of doubling the number of offspring (n_o) relative to effect of doubling number of grandoffspring per offspring (n_{og}) {RED(n_o,n_{og})} for family structures HS3,FS3 and FSHS

Family Structure	RED
HS3	$1 + .75\ h^2 / (\ .75\ h^2 + (4 - h^2)/n_{og}\ )$
FS3	$1 + 1.25\ h^2 / (\ 1.25\ h^2 + (4 - 2\ h^2)/n_{og}\ )$
FSHS	$1 + .25\ h^2 / (\ .25\ h^2 + (4 - h^2)/n_{og}\ )$

h^2 is heritability

Results

Power of an experiment with a two-generation family structure. Power of experiments with a HS2 or a FS2 family structure for a QTL that explains 1% of the phenotypic variance and that has a heterozygosity of 50%, for various number of families, various number of offspring per family and for two heritabilities ($h^2 = .1$ and $h^2 = .4$) of the observed trait are in Table 5. First, Table 5 is used to compare HS2 and FS2 family structures. An experiment with a FS2 family structure and n families had about the same power as an experiment with a HS2 family structure and $2n$ families for all values of n. For a heritability of .1, an experiment with 5 full-sib families and 800 offspring per family had a power of .59 and an experiment with 10 half-sib families and 800 offspring per family had a power of .57. This example shows that an experiment with a FS2 family structure and n families had the same power as an experiment with a HS2 family structure and $2n$ families. RE(FS2/HS2) reflects this. RE(FS2/HS2) was close to 2: 2.06 for $h^2 = .1$ and 2.25 for $h^2 = .4$

Power of an experiment with a two-generation family structure increased with increasing number of families and with increasing number of offspring per family (Table 5). Increasing the number of offspring per family was more efficient than increasing the number of families, e.g. for a h^2 of .1, an experiment with 10 full-sib families and 200 offspring per family had a power of .43 whereas an experiment with 20 full-sib families and 100 offspring per family had a power of .27. The noncentrality parameter doubled by doubling the number of families and also doubled by doubling the number of offspring per family. Doubling the number of families, however, doubled the degrees of freedom, whereas doubling the number of offspring per family did not influence the degrees of freedom.

Power of an experiment with a two-generation family structure increased with increasing heritability if the effect of the QTL, expressed in phenotypic standard deviation units, remained constant (Table 5). For an experiment with 5 half-sib families and 800 offspring per family, power was .34 for $h^2 = .1$, and power was .38 for $h^2 = .4$. The effect of h^2 could be explained from RE($h^2 = .4/h^2 = .1$). RE($h^2 = .4/h^2 = .1$) was 1.08 for a HS2 family structure and 1.19 for a FS2 family structure.

Table 5 Power of experiments[a] with a two-generation half-sib (HS2) or two-generation full-sib (FS2) family structure for a QTL that explains 1% of the phenotypic variance and that has a heterozygosity of 50%, for various number of families (n_f), various number of offspring per family (n_o) and two heritabilities ($h^2=.1$ and $h^2=.4$)

n_f	n_o	$h^2=.1$		$h^2=.4$	
		HS2	FS2	HS2	FS2
5	50	.02	.02	.02	.02
	100	.03	.04	.03	.04
	200	.05	.09	.06	.11
	400	.13	. 24	.15	.31
	800	.34	.59	.38	.69
10	50	.02	.03	.02	.03
	100	.04	.05	.04	.07
	200	.08	.15	.09	.19
	400	.23	.43	.26	.54
	800	.57	.86	.62	.92
20	50	.03	.03	.03	.04
	100	.05	.09	.06	.11
	200	.14	.27	.16	.36
	400	.42	.72	.47	.83
	800	.85	.99	.88	.99
40	50	.03	.05	.04	.06
	100	.08	.15	.09	.21
	200	.26	.51	.30	.64
	400	.71	.95	.76	.99
	800	.99	.99	.99	.99

[a] For all experiments power is for interval analysis assuming that markers have a heterozygosity of 1.

Power of an experiment with a three-generation family structure. Power of experiments with a three-generation family structure (HS3, FS3 or FSHS) for a QTL that explains 1% of the phenotypic variance and that has a heterozygosity of 50%, for various number of families, for various number of offspring per family, for various number of grandoffspring per offspring, and for two heritabilities ($h^2=.1$ and $h^2=.4$) are in Table 6.

The power of an experiment with a HS3 family structure was similar to the power of an experiment with a FS3 family structure if n_f, n_o and n_{og} were equal for the two family structures (Table 6). For a heritability of .1, the power for a FS3 family structure was 1 to 1.2 times the power for a HS3 family structure whereas for a heritability of .4 power of a FS3 family structure was 0.85 to 1 times the power for a HS3 family structure. The power of an experiment with a FSHS family structure was higher than the power of an experiment

with either a HS3 or a FS3 family structure. For a heritability of .1, the power of an experiment with a FSHS family structure was 1 to 4 times the power of an experiment with a HS3 or a FS3 family structure and for a heritability of .4 power of an experiment with a FSHS family structure was 1 to 6 times the power of an experiment with a HS3 or a FS3 family structure. For example, for a heritability of .1, an experiment with 10 families, 50 offspring per family and 100 grandoffspring per offspring had a power of .82 for a FSHS family structure, a power of .26 for a HS3 family structure and a power of .27 for a FS3 family structure.

Table 6 Power of experiments[a] with three-generation family structures (HS3, FS3 or FSHS) for a QTL that explains 1% of the phenotypic variance and that has a heterozygosity of 50%, for various number of families (n_f), for various number of offspring per family (n_o), for various number of grandoffspring per offspring (n_{og}) and for two heritabilities ($h^2=.1$ and $h^2=.4$)

n_f	n_o	n_{og}	$h^2=.1$			$h^2=.4$		
			HS3	FS3	FSHS	HS3	FS3	FSHS
5	25	10	.02	.02	.02	.02	.02	.02
		50	.04	.04	.11	.02	.02	.05
		100	.06	.06	.22	.02	.02	.07
	50	10	.03	.03	.04	.02	.02	.04
		50	.10	.11	.30	.03	.03	.14
		100	.15	.15	.55	.04	.03	.14
	100	10	.06	.08	.11	.04	.04	.09
		50	.26	.31	.67	.08	.07	.38
		100	.37	.41	.88	.09	.07	.51
10	25	10	.02	.02	.03	.02	.02	.03
		50	.06	.07	.18	.02	.02	.08
		100	.09	.09	.39	.03	.02	.12
	50	10	.04	.05	.06	.03	.03	.06
		50	.17	.19	.53	.05	.04	.25
		100	.26	.27	.82	.05	.05	.36
	100	10	.09	.13	.18	.06	.06	.15
		50	.45	.54	.91	.12	.11	.65
		100	.62	.68	.99	.14	.12	.79

[a] For all experiments power is for interval analysis assuming that markers have a heterozygosity of 1.

Table 7 Relative Efficiency (RE) among experiments with three-generation family structures (HS3, FS3 or FSHS) for heritability (h^2) of 0, .1 or .4 and number of grandoffspring per offspring (n_{og}) of 10, 50 or ∞

h^2	n_{og}	RE(FS3/HS3)	RE(FSHS/HS3)	RE(FSHS/FS3)
0	10	2	2	1
	50	2	2	1
	∞	2	2	1
.1	10	1.84	2.24	1.22
	50	1.54	2.96	1.95
	∞	1.2	6	5
.4	10	1.6	2.86	1.78
	50	1.32	4.32	3.27
	∞	1.2	6	5

Table 8 Efficiency of heritability at .4 relative to efficiency of heritability at .1 {RE(h^2=.4/h^2=.1)} for experiments with three-generation family structures (HS3, FS3 or FSHS) with number of grandoffspring per offspring (n_{og}) of 10, 50 or 100

n_{og}	HS3	FS3	FSHS
10	.70	.62	.90
50	.41	.36	.60
100	.34	.31	.47

Table 9 Effect of doubling the number of offspring (n_o) relative to effect of doubling number of grandoffspring per offspring (n_{og}) {RED(n_o,n_{og})} for three-generation family structures with number of grandoffspring per offspring of 10, 50 or 100 and with heritability (h^2) of .1 or .4

n_{og}	h^2=.1			h^2=.4		
	HS3	FS3	FSHS	HS3	FS3	FSHS
10	1.16	1.25	1.06	1.43	1.65	1.22
50	1.49	1.62	1.24	1.79	1.90	1.58
100	1.66	1.77	1.39	1.89	1.95	1.74

RE for the HS3, FS3 and FSHS family structures depended on h^2 and n_{og} as is shown in Table 7 that gives the RE among HS3, FS3 and FSHS family structures for h^2 of 0, .1 or .4 and n_{og} of 10, 50 or ∞. RE(FS3/HS3), RE(FSHS/HS3) and RE(FSHS/FS3) were larger than 1 for all combinations of values of h^2 and n_{og}. RE(FS3/HS3) decreased and RE(FSHS/HS3) and RE(FSHS/FS3) increased with increasing h^2 and with increasing n_{og}. Maximum values were 2 for RE(FS3/HS3), 5 for RE(FSHS/FS3) and 6 for RE(FSHS/HS3).

Power of an experiment with a three-generation family structure increased with increasing number of families and with increasing number of offspring per family (Table 6).

Similar to two-generation experiments, power increased more by doubling the number of offspring per family than by doubling the number of families.

Power of an experiment with a three-generation family structure decreased with increasing heritability. If n_f was 10, n_o was 50 and n_{og} was 100, then for a FSHS family structure power of the experiment was .82 for $h^2 = .1$ and .36 for $h^2 = .4$. The effect of h^2 could be explained from RE($h^2 = .4/h^2 = .1$). Table 8 gives RE($h^2 = .4/h^2 = .1$) for experiments with three-generation family structures with 10, 50 or 100 grandoffspring per offspring. All RE($h^2 = .4/h^2 = .1$) in Table 8 were smaller than 1, i.e. power was lower for $h^2 = .4$ than for $h^2 = .1$ for all combinations of three-generation family structure and number of grandoffspring per family. RE($h^2 = .4/h^2 = .1$) was lowest for a FS3 family structure and highest for a FSHS family structure. RE($h^2 = .4/h^2 = .1$) was lowest for $n_{og} = 100$ and highest for $n_{og} = 10$.

Power of an experiment with a three-generation family structure increased more by doubling the number of offspring than by doubling the number of grandoffspring per offspring (Tables 6 and 9). Table 9 gives values for RED(n_o, n_{og}) for three-generation family structures with $n_{og} = 10$, 50 or 100 and with $h^2 = .1$ or .4. Table 8 shows that RED(n_o, n_{og}) increased with increasing heritability and with increasing n_{og}. RED(n_o, n_{og}) was lowest for a FSHS family structure and highest for a FS3 family structure.

Comparing experiments with two- and three- generation family structures. In Table 5 are powers of experiments with two-generation family structures and in Table 6 are powers of experiments with three-generation family structures. For a given number of families, more offspring per family are needed to obtain a certain power for experiments with two generations than for experiments with three generations. For example, if $h^2 = .1$ then an experiment with 10 two-generation full-sib families with 800 offspring per family had a power of .86, whereas an experiment with 10 families with a FSHS family structure, with 100 full-sib offspring per family and 50 half-sib grandoffspring per full-sib offspring had a power of .91.

Besides directly comparing power, experiments were also compared by Relative Efficiency (RE). In Figure 1 are the efficiencies of five family structures relative to the efficiency of a HS2 family structure for $h^2 = .1$ and $h^2 = .4$, for values of n_{og} ranging from 0 to 100. RE showed that three-generation experiments were more efficient than two-generation experiments, especially for traits with low heritability and if many grandoffspring per offspring are available. RE(FSHS/HS2) was 30 for $h^2 = .1$ and $n_{og} = 100$.

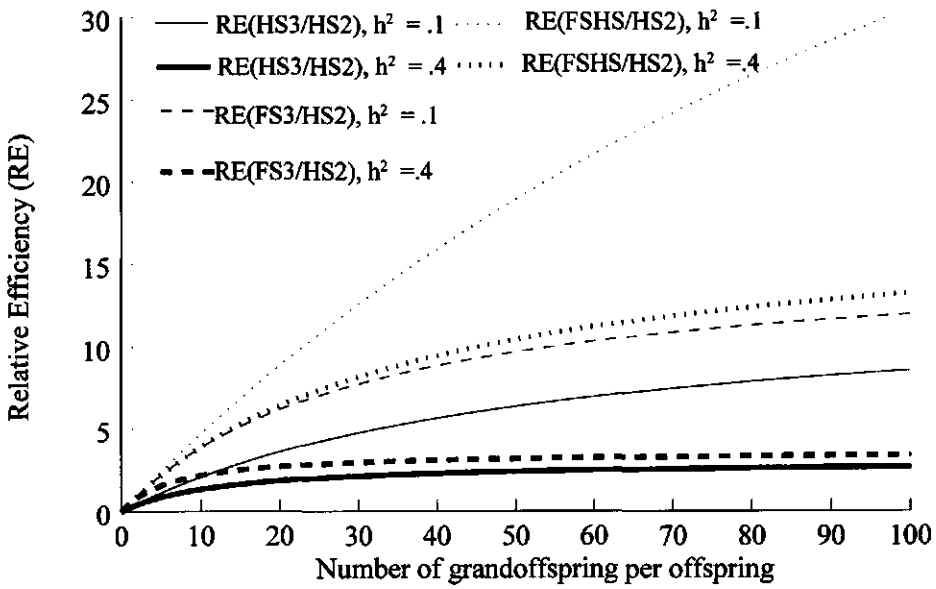


Figure 1 Efficiency of five family structures relative to the efficiency of a HS2 family structure for two heritabilities ($h^2 = .1$ and $h^2 = .4$)

Discussion

In this paper we used power and relative efficiency to compare the efficiency of designs of QTL-mapping experiments. Although power determines the value of a design, it has its limitations to compare designs because the relation between power and the size of an experiment is non-linear. Tables 5 and 6 showed that the effect on power of increasing the size of an experiment depended on the initial power. Relative efficiency is independent from initial power and size of an experiment. Relative efficiency can directly be translated to the relative number of marker genotypes that have to be determined in experiments whereas power cannot be used for this purpose.

We showed that family structure is an important factor in designing QTL mapping experiments. For a two-generation experiment, more offspring were required for a certain power than for a three-generation experiment. Consequently, more animals had to be typed for marker loci for a two generation experiment than for a three-generation experiment. Fewer offspring per two-generation experiment, however, were required than grandoffspring per three-generation experiment. Thus, fewer trait values had to be obtained for a two-generation experiment than for a three-generation experiment. In deciding on a design, the costs of typing a marker and the costs of obtaining a trait value have to be considered as well

as the time required to collect all information. The costs of obtaining a trait value are low if a trait is routinely collected for management or breeding purposes, such as milk production in dairy cattle. Furthermore, the family structure of the commercial dairy cattle population enables experiments with a three-generation half-sib family structure. The three-generation half-sib family structure is used in QTL mapping experiments in dairy cattle (Da *et al.* 1994; Georges *et al.* 1995). A trait like meat quality measured on carcasses, however, is expensive to measure and usually not collected routinely. In such a case cost of measuring meat quality should be balanced against costs of typing markers.

Relative Efficiency and comparison of power showed that a two- or three- generation family structure with full-sib offspring was more efficient than a two- or three- generation family structure with half-sib offspring. With full-sib offspring, two marker contrasts can be computed per family while with half-sib offspring only one marker contrast can be computed. A family structure with full-sib *grand*offspring, however, was less efficient than a family structure with half-sib grandoffspring. The standard error of a marker contrast was larger when full-sib grandoffspring were used than when half-sib grandoffspring were used.

Power of three-generation experiments decreased with increasing heritability if the effect of the QTL, expressed in *phenotypic* units, was constant. In a three-generation experiment, information comes from the average trait value on grandoffspring. This average represents the breeding value of the offspring, plus a residual. The part of the breeding value that is due to the QTL of interest decreases with increasing heritability if the QTL effect is constant in phenotypic units, as we used. So, those grandoffspring averages become less informative with increasing heritability and power decreases. In our statistical method only one QTL is considered. Recently, multi-marker methods have been developed that allow for several QTLs simultaneously (Zeng 1993; Jansen and Stam 1994). In theses methods a larger part of the breeding values of offspring is accounted for which results in a reduction of the residual genetic variance in grandoffspring averages and consequently in a higher power. Our results can be used to infer the effect of multi-marker methods on power of three-generation experiments. The effect of reduction of residual genetic variance is similar to that of reduction in heritability in a single QTL method. If markers explain, say, half the genetic variance then we could simply take the power for a situation where the heritability is halved.

For two-generation experiments the effect of heritability was small. For the range of heritabilities we studied, standard errors on marker contrasts in the two-generation designs are almost entirely determined by environmental errors. This suggests that for two-generation experiments multi-marker experiments are less beneficial.

So far, we compared full-sib families and half-sib families for a given family size. Due to limitations in female reproductive capacity, family size is more limited with full-sib than with half-sib families. This is particularly true in cows and pigs, but less in poultry and fish.

It would be fair to compare an experiment with a few large half-sib families to an experiment with many smaller full-sib families. Table 5 showed that an experiment with 5 half-sib families and 800 offspring per family had a higher power than an experiment with 40 full-sib families and 100 offspring per family. This illustrates that practical limitations have to be considered in designing an experiment.

Power of experiments were given for various family structures and also family sizes and heritability of the trait were varied. Other parameters that also determine power, like marker heterozygosity, were held constant. The effect on power of some other parameters will be discussed now.

The RE(I/S) in this study is for a linear model but was equal to the efficiency of interval mapping relative to the efficiency of single marker mapping for a likelihood model as given by Lander and Botstein (1989). The RE(I/S) was $1/(1-\gamma)$. This means that an experiment with x offspring per family that is analyzed by interval analysis, has the same power as an experiment with $x/(1-\gamma)$ offspring per family that is analyzed by single marker analysis, assuming that only the number of offspring per family and the type of analysis differ between the experiments. For example, a FS2 experiment with 5 families with 800 offspring per family had a power of .59 for $h^2 = .1$ and interval analysis. For single marker analysis, a FS2 experiment with 5 families with $800/(1-\gamma) = 800/(1-.165) = 958$ offspring per family would have had a power of .59 for $h^2 = .1$. The RE(I/S) was computed for a QTL at the midpoint of the interval between two markers. The difference between single marker analysis and interval analysis decreases if the QTL is closer to the bounds of the interval (Darvasi *et al.* 1993; Mackinnon *et al.* 1995).

Not only location of the QTL but also variance explained by the QTL is important. The variance depends on a^2 and heterozygosity of the QTL. The equation for RED(a^2,n_o) showed that doubling a^2 had the same effect on power as doubling n_o. Because the effect of doubling n_o is given in tables 5 and 6, also the effect of doubling a^2 can be derived from those tables. Heterozygosity at the QTL was .5 in this study. For a QTL with two alleles this is the maximum heterozygosity in a population that is in genetic equilibrium. For lower heterozygosity at the QTL, more families will be needed for equal power. The relation between heterozygosity at the QTL and required number of families is about linear (Weller *et al.* 1990).

We showed that map distance between the markers that flank the QTL determines the efficiency of interval analysis relative to the efficiency of single marker analysis. Map distance also has a direct influence on efficiency. The efficiency of a design with $d = .2$ relative to a design with $d = .1$ (RE($d = .2/d = .1$)) was .89. Thus, an experiment with $d = .2$ and n_o offspring per family has the same power as an experiment with $d = .1$ and $.89 \times n_o$ offspring per family.

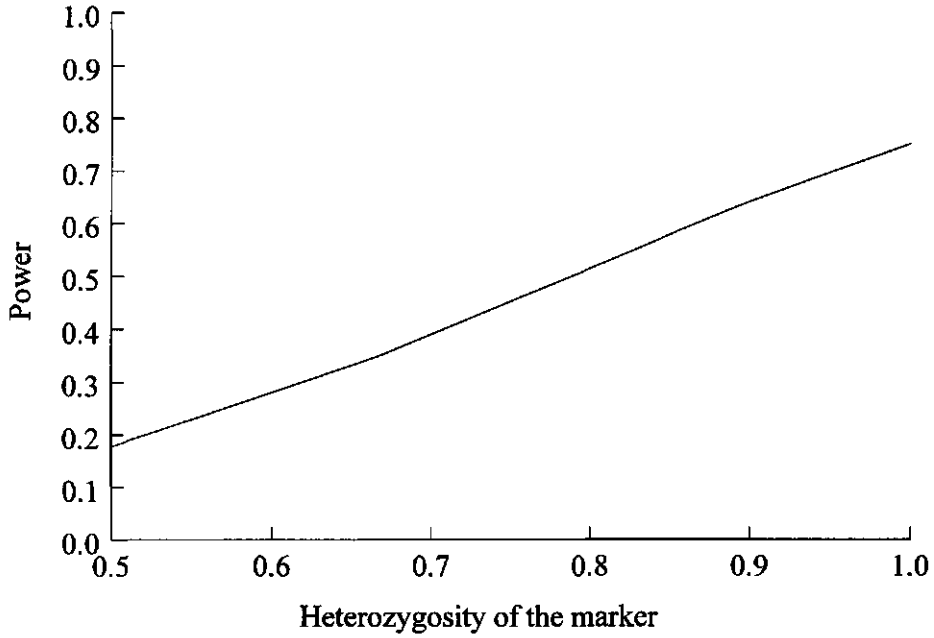


Figure 2 Effect of heterozygosity of the marker on power of an experiment with a two-generation half-sib family structure with 20 families and 800 offspring per family for a QTL that explains 1% of the phenotypic variance and that has a heterozygosity of .5, for a heritability of .1. Power is computed for single marker analysis as the weighted average of powers of experiments with 0 to 20 heterozygous parents. Heterozygosity of the marker is varied by the number of equiprobable marker alleles.

Heterozygosity at the marker was 1 in this study because due to the abundance of highly polymorphic microsatellite loci it is expected that a set of almost perfect markers will become available. A recent study in chicken (Groen *et al.* 1994), however, showed that the average heterozygosity of microsatellite markers was .28 in commercial layer populations and .55 in commercial broiler populations. At such a level of heterozygosity power of an experiment is lower than for a heterozygosity at the marker of 1. Figure 2 illustrates this for one design of an experiment. Power is decreased because a marker or haplotype contrast can only be computed for a heterozygous animal and further at lower marker polymorphism it will less often be possible to determine which marker allele is transferred from a parent to an offspring (Soller 1991). The latter problem will be larger in a half-sib family in which one parent is untyped than in a full-sib family in which both parents are typed for the marker (Soller 1991). If heterozygosity of the marker loci is not close to 1 then an alternative to analysing markers one by one or in pairs, is to simultaneously use many linked markers which increases the power of an experiment (Haley *et al.* 1994).

Power was computed for an experiment with a balanced design. The power of an experiment with a balanced design is, however, a good indicator for the power of an experiment with a design that is unbalanced with respect to the number of offspring per family, the number of grandoffspring per offspring or the number of offspring per marker allele per parent (Wang *et al.* 1995). The χ^2 method to compute power is approximate because error variance and heritability have to be known whereas in a real experiment these parameters have to be estimated from the data. Values computed with the χ^2 method are, however, close to exact values (Wang *et al.* 1995). For the situation studied by Wang *et al.* (1995) the maximum difference between the approximate and the exact value was .034.

We studied experiments within one outbred population. The efficiency of an experiment with a cross between populations is higher (Soller 1991). Genes of large effects are expected to segregate at higher frequencies in a cross than within an outbred population and also a high level of heterozygosity of the marker is more likely if a cross between populations is used. If in a cross a marker is found that explains between population variance, then possibly this marker also explains within population variance. Whether this is true or not remains uncertain until within population studies are performed. If the goal is to understand and exploit the variation within commercial populations, then QTL mapping experiments within the populations are necessary. This study will help to design such experiments.

Acknowledgements

The authors acknowledge financial support from Euribrid B.V. We thank M. Grossman for suggestions on the first version of the manuscript.

References

Andersson L, Archibald AL, Gellin J, Schook LB (1993) 1st pig gene mapping workshop (PGM1), August 7, 1992, Interlaken, Switzerland. Anim Genet 24: 205-216.

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

Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick BW, Ryan AM, Clayton D, Li L, Neibergs HL, Zhang N, Grosse WM, Weiss J, Creighton P, McCarthy F, Ron M, Teale AJ, Fries R, McGraw RA, Moore SS, Georges M, Soller M, Womack JE, Hetzel DJS (1994) A genetic linkage map of the bovine genome. Nature Genetics 6: 227-235.

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Hawkins GA, Solinas Toldo S, Fries R, Grosz MD, Yoo J, Beattie CW (1994) A genetic linkage map for cattle. Genetics 135: 619-639.

Crawford AM, Montgomery GW, Pierson CA, Brown T, Dodds KG, Sunden SLF, Henry HM, Ede AJ, Swarbrick PA, Berryman T, Penty JM, Hill DF (1994) Sheep linkage mapping: nineteen linkage groups derived from the analysis of paternal half-sib families. Genetics 137:573-579.

Da Y, Ron M, Yanai A, Band M, Everts RE, Heyen DW, Weller JI, Wiggans GR, Lewin HA (1994) The dairy bull DNA repository: a resource for mapping quantitative trait loci. Proc 5th World Congr Genet Appl Livestock Prod 21:229-232.

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting Marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943-951.

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

Geldermann H (1975) Investigations on inheritance of quantitative characters in animals by gene markers I. Methods. Theor Appl Genet 46:319-330.

Groen AF, Crooijmans RPMA, Van Kampen AJA, Van der Beek S, Van der Poel JJ, Groenen MAM (1994) Microsatellite polymorphism in commercial broiler and layer lines. Proc 5th World Congr Genet Appl Livestock Prod 21:95-98.

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

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455.

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

Mackinnon MJ, Van der Beek S, Kinghorn BP (1995) Use of deterministic sampling for exploring likelihoods in linkage analysis for quantitative traits. Theor Appl Genet: in press.

Paterson AH, Lander ES, Hewitt JD, Perterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Rohrer GA, Alexander LJ, Keele JW, Smith TP, Beattie CW (1994) A microsatellite linkage map of the porcine genome. Genetics 136:231-245.

Soller M, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34:47-55.

Soller M (1991) Mapping quantitative trait loci affecting traits of economic importance in animal populations using molecular markers. In: Schook LB, Lewin HA, McLaren DG (eds) Gene-mapping techniques and applications. Marcel Dekker, New York.

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823-839.

Wang Y, Dekkers JCM, Gibson JP (1995) Effect of data structure on statistical power of daughter and granddaughter designs for detecting marker-QTL associations. Theor Appl Genet:in press.

Weller JI, Kashi Y, Soller M (1990) Power of daughter and granddaughter designs for determining linkdage between marker loci and quantitative trait loci in dairy cattle. J Dairy Sci 73:2525-2537.

Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972-10976.

Appendix 1 Standard error of marker contrast

The linear model for an experiment with a HS2 family structure is:

$$y_{ijk} = s_i + m_{ij} + e_{ijk} \quad \text{(A1)}$$

where y_{ijk} is the trait value for the k-th offspring inheriting marker allele j of sire i, s_i is the effect sire i with $\sigma_s^2 = .25\ \sigma_u^2$, m_{ij} is the effect of marker allele j of sire i and e_{ijk} is the residual effect of offspring k with $\sigma_e^2 = \sigma_p^2 - \sigma_s^2$. Let σ_p^2 be 1 then $\sigma_s^2 = .25\ h^2$ and $\sigma_e^2 = 1 - .25\ h^2$

For a sire with marker genotype Mm, m_{i1} is the effect of allele M and m_{i2} the effect of allele m. Let the expected value of the marker contrast be $(m_{i1} - m_{i2})$ and the realized value:

$$MC_i = \frac{2}{n_o}\sum_{k=1}^{n_o/2} y_{i1k} - \frac{2}{n_o}\sum_{k=1}^{n_o/2} y_{i2k} \quad \text{(A2)}$$

where MC_i is the marker contrast for sire i, n_o is the number of offspring per sire and $n_o/2$ is the number of offspring per marker allele.
The squared standard error of the marker contrast is:

$$SE_{MC}^2 = E(MC - E(MC))^2$$

Rewriting A2 using A1 gives:

$$\begin{aligned} MC_i &= \frac{2}{n_o}\sum_{k=1}^{n_o/2}(s_i + m_{i1} + e_{i1k}) - \frac{2}{n_o}\sum_{k=1}^{n_o/2}(s_i + m_{i2} + e_{i2k}) \\ &= (s_i + m_{i1} + \frac{2}{n_o}\sum_{k=1}^{n_o/2} e_{i1k}) - (s_i + m_{i2} + \frac{2}{n_o}\sum_{k=1}^{n_o/2} e_{i2k}) \\ &= m_{i1} - m_{i2} + \frac{2}{n_o}\sum_{k=1}^{n_o/2}(e_{i1k} - e_{i2k}) \end{aligned}$$

We can derive that:

$$\begin{aligned} E(MC_i - E(MC_i))^2 &= E((m_{i1} - m_{i2} + \frac{2}{n_o}\sum_{k=1}^{n_o/2}(e_{i1k} - e_{i2k})) - (m_{i1} - m_{i2}))^2 \\ = E(\frac{2}{n_o}\sum_{k=1}^{n_o/2}(e_{i1k} - e_{i2k}))^2 &= \frac{4}{(n_o)^2}\sum_{k=1}^{n_o/2} 2\sigma_e^2 = \frac{4}{(n_o)^2} n_o \sigma_e^2 = \frac{4 - h^2}{n_o} \end{aligned}$$

Appendix 2 Expected haplotype contrast

Let M and N denote two markers and Q a QTL. The recombination rate between M and Q is r_1, between Q and N is r_2 and between M and N is γ. Let the ordered genotype of a parent be MQN/mqn. The haplotype contrast is defined as the average trait value of offspring that inherit nonrecombinant haplotype MN from the parent, minus the average trait value of offspring that inherit nonrecombinant marker haplotype mn from the parent. The expected haplotype contrast is the expected trait value of offspring that inherit nonrecombinant haplotype MN minus the expected value of offspring that inherit nonrecombinant haplotype mn from the parent.

Offspring that inherited MN, inherited marker-QTL haplopye MQN from the parent with probability $(1-r_1)\times(1-r_2)/(1-\gamma)$ and will have inherited marker-QTL haplotype MqN form the parent with probability $r_1\times r_2/(1-\gamma)$. The expected trait value of offspring that inherited MN is $(1-r_1)\times(1-r_2)/(1-\gamma) \times a/2 + r_1\times r_2/(1-\gamma) \times -a/2 = (1-r_1-r_2)/(1-\gamma) \times a/2$. Offspring that inherited mn have inherited marker-QTL haplotype mQn from the parent with probability $r_1\times r_2/(1-\gamma)$ and have inherited marker-QTL haplotype MqN from the parent with probability $(1-r_1)\times(1-r_2)/(1-\gamma)$. The expected trait value of offspring that inherited mn is $r_1\times r_2/(1-\gamma) \times a/2 + (1-r_1)\times(1-r_2)/(1-\gamma) \times -a/2 = (r_1+r_2-1)/(1-\gamma) \times a/2$. Thus, the expected haplotype contrast is $(1-r_1-r_2)/(1-\gamma) \times a/2 - (r_1+r_2-1)/(1-\gamma) \times a/2 = (1-r_1-r_2)/(1-\gamma) \times a$. For $r_1=r_2=r$ the expected haplotype contrast is $(1-2r)/(1-\gamma) \times a$.

Chapter 5

Marker assisted selection in an outbred poultry breeding nucleus

Sijne van der Beek and Johan A.M. van Arendonk

Animal Science (In Press)
Reproduced by permission of the British Society of Animal Science

Abstract

The value of using a marker for a Quantitative Trait Locus (QTL) affecting a sex limited trait in an outbred poultry breeding nucleus was studied. Marker and QTL were in linkage equilibrium in the base population. The recombination rate between marker and QTL was .05. A closed nucleus with 9000 chickens per generation was deterministically simulated. The genetic model contained polygenes and a QTL linked to a marker. Genetic effects explained 30% of the phenotypic variance before selection. Under selection, polygenic variance reached an equilibrium and QTL variance decreased continuously over time. Cocks were selected in two steps. First the best cocks of each full-sib family were selected (within family selection) while final selection took place after information on full-sibs was available. Hens were selected after they had completed production. The effect of using marker information in estimating breeding values was studied in an ongoing breeding programme. Transmission of marker alleles was always traceable. Cumulative response over five generations increased 6 to 13 % if a marker linked to a QTL that explained 20% of the genetic variance was used. Cumulative response increased up to 28% if the QTL explained 80% of the genetic variance. Additional response due to the use of a marker increased with increasing intensity of within family selection of cocks, increased with increasing variance explained by the QTL, and was higher if within family selection of cocks was carried out after rather than before their sibs had complete records.

Keywords: Marker assisted selection - poultry - selection response - genetic model

Introduction

In most livestock breeding schemes breeding values for quantitative traits are estimated using phenotypic information and additive genetic relationships among animals. The genetic model underlying the breeding value estimation assumes that an infinite number of genes each with an infinitely small effect influence the quantitative trait. An increasing number of genes or chromosome segments that explain a significant part of the variation in quantitative traits, however, are being identified. Use of this knowledge can improve current breeding value estimation procedures (Soller 1994). Once a gene is located and the alleles of the gene identified, the effects of those alleles can be incorporated in breeding value estimation (Smith 1967). If a marker is available that is linked to a locus explaining variation in a quantitative trait (i.e. a Quantitative Trait Locus: QTL) marker assisted breeding value estimation may be applied. Procedures for combining marker information and phenotypic information in selection index (Soller 1978) or best linear unbiased prediction (Fernando and Grossman 1989) have been described.

The use of markers can improve the selection response in a breeding programme (Soller 1978; Soller and Beckmann 1983; Smith and Simpson 1986; Stam 1986; Kashi *et al.* 1990; Dentine 1992; Meuwissen and Van Arendonk 1992). Factors that influence the additional selection response include: (1) amount of variance explained by identified QTL(s); (2) distance between a marker and a QTL; (3) marker polymorphism; (4) the structure of the

breeding programme; and (5) type of breeding goal trait (sex limited or not). Dairy cattle breeding was the subject of the studies mentioned above. The structure of a poultry breeding population differs from the structure of a dairy cattle breeding population and therefore we cannot predict from those dairy cattle breeding studies the value of marker assisted selection (MAS) in poultry breeding. Most studies computed additional selection response for one round of selection starting from an unselected base generation. Meuwissen and Van Arendonk (1992) computed equilibrium response assuming the variance accounted for by markers remained constant over generations. The latter assumption is unrealistic if a marker is linked to one or a few QTLs of large effect.

We study the value of using a marker for a QTL in a poultry breeding nucleus. We use a genetic model in which each round of selection reduces the variance explained by a QTL. The genetic model is used in deterministic simulations to quantify the additional response due to MAS and the change of this additional response over time.

Methods

Breeding programme. A poultry breeding nucleus with non-overlapping generations is simulated. In a hierarchical mating scheme each sire is mated to several dams and several offspring are produced by each dam. Selection is for egg production, a sex limited trait. Egg production is measured on all hens hatched. Each hen can be selected without a restriction on the number of full-sibs selected. Hens are selected after they have completed their first record. Selection of cocks differs among alternative breeding schemes. If no marker information is available then only the *max-c* first hatched cocks are candidates for selection, where *max-c* is the maximum number of cocks that may be selected from a full-sib family. This is random selection. With markers, this marker information is used to select *max-c* cocks from all available cocks in a full-sib family. This is called within family selection. Marker assisted within family selection is either immediately after hatching (juvenile within family selection) or after female sibs have complete records (adult within family selection). In both cases all cocks are kept until the time of selection.

Deterministic simulation. Response per generation is computed as:

$$R = (R_w + R_c + R_h)/2 \qquad (1)$$

where R is the response per generation, R_w is the response to within family selection of cocks, R_c is the response to final selection of cocks, and R_h is the response to selection of hens.

The deterministic simulation starts in generation zero, the base generation of unrelated animals. Each sire in the base generation has an estimated breeding value of zero. Breeding values for dams in the base generation are estimated using own performance. After five generations of selection, a marker is introduced. Introduction is not in generation zero to study the effect of introducing a marker in an ongoing breeding programme. One marker is introduced. Each animal in generation five and its parents are typed for this marker. Each parent is assumed to be heterozygous for the marker. Transmission of marker alleles can be traced without error. Let M1M2 be the marker genotype of a sire and M3M4 the marker genotype of a dam. The full-sib offspring are divided in four groups of equal size based on their marker genotype: M1M3, M1M4, M2M3 or M2M4.

Selection indices accounting for marker information are derived for cocks and hens. The selection index depends on the marker genotype of the candidate for selection. For each of the four possible marker genotype a selection index is derived. The mean and variance of each index are computed. Equation (2) gives the response per generation allowing for the different selection indices :

$$R_t = \left\{ \sum_{j=1}^{4} f_{wtj} \times \left(m_{wtj} - \bar{m}_{wt} + i_{wtj} \times \sigma(\hat{A}_{wtj}) \right) + \sum_{j=1}^{4} f_{ctj} \times \left(m_{ctj} - \bar{m}_{ct} + i_{ctj} \times \sigma(\hat{A}_{ctj}) \right) + \sum_{j=1}^{4} f_{htj} \times \left(m_{htj} - \bar{m}_{ht} + i_{htj} \times \sigma(\hat{A}_{htj}) \right) \right\} / 2 \qquad (2)$$

where R_t is the response in generation t, f_{wtj} is the fraction selected of cocks with within family selection index j in generation t, m_{wtj} is the mean genetic value before within family selection of cocks with within family selection index j in generation t, $\bar{m}_{wt}$ is the mean of all cocks before within family selection in generation t, i_{wtj} is the selection intensity within the group of cocks with within family selection index j in generation t, and $\sigma(\hat{A}_{wtj})$ is the standard deviation of within family selection index j in generation t. The appendix illustrates the use of (2) to compute within family selection response.

Effect of selection on genetic (co)variances and (co)variances between information sources is accounted for after each selection step (Cochran 1951). For hens there is one selection step in each generation. For cocks, (co)variances are corrected both after within family selection and after final selection. Before each selection step, genetic effects are assumed to follow a multivariate normal distribution.

The effect of inbreeding on (co)variances is ignored. Derivation of selection index and computation of selection intensities will be described in later sections.

Genetic model. Genetic variance is due to additive gene action at one QTL and polygenes unlinked to the QTL. The breeding value of an animal is the effect of the paternal allele at the QTL plus the effect of the maternal allele at the QTL plus the effect of the polygenes. The model for the breeding value of animal i is:

$$a_i = v_i^p + v_i^m + u_i \tag{3}$$

where a_i is the breeding value of animal i, v_i^p is the effect of the paternal QTL allele, v_i^m is the effect of the maternal QTL allele, and u_i is the polygenic effect. Genetic effects are random and multivariate normally distributed. In the base generation before selection, the variance of a QTL allele is $\sigma^2(v^0)$, polygenic variance is $\sigma^2(u^0)$, genetic variance is $\sigma^2(a^0) = 2\sigma^2(v^0) + \sigma^2(u^0)$, and covariance between the effect of a QTL allele and the polygenic effect is zero.

A marker linked to the QTL is available. The relation between the allelic effects at the QTL of animal i and the allelic effects of the parents depends on marker information (Fernando and Grossman 1989):

$$v_i^p = (1-q_s)v_s^p + q_s v_s^m + \epsilon(v_i^p) \tag{4}$$

$$v_i^m = (1-q_d)v_d^p + q_d v_d^m + \epsilon(v_i^m) \tag{5}$$

where s denotes the sire, d the dam, and $\epsilon(v_i^p)$ is the part of v_i^p not explained by the regression on the parental allelic effects. The term $(1-q_s)$ is the probability that the sire transmits its paternal QTL allele and q_s is the probability that the sire transmits its maternal QTL allele. Let r be the recombination rate between the marker and the QTL. The value of q_s is r if animal i inherits the paternal marker allele of the sire and q_s is $(1-r)$ if animal i inherits the maternal marker allele of the sire.

For polygenic effects the following relation holds:

$$u_i = 0.5\, u_s + 0.5\, u_d + \epsilon(u_i) \tag{6}$$

where $\epsilon(u_i)$ is the part of u_i not explained by the regression on the parental polygenic effects.

Equations (4), (5) and (6) are used to compute the covariance between genetic effects of different animals. The covariance between the effect of a QTL allele in animal j and the effect of the paternal QTL allele in animal i is computed using the covariance between the QTL allele of animal j and the effects of the QTL alleles of the sire of animal i (Fernando and Grossman 1989) :

$$\sigma(v_j^p, v_i^p) = (1-q_s)\sigma(v_j^p, v_s^p) + q_s\ \sigma(v_j^p, v_s^m) \tag{7}$$

where $\sigma(v_i^p, v_j^p)$ is the covariance between the paternal QTL allele of animal i and the paternal QTL allele of animal j. Equations similar to (7) were derived for the covariance between maternal QTL alleles or between a paternal and a maternal QTL allele. Fernando and Grossman (1989) describe a tabular method with repeated use of (6) to compute the covariance between any pair of QTL alleles. The covariance between the polygenic effect of animal j and the effect of the paternal QTL allele of animal i is:

$$\sigma(u_j, v_i^p) = (1-q_s)\sigma(u_j, v_s^p) + q_s\ \sigma(u_j, v_s^m) \tag{8}$$

In the base generation before selection the covariance between polygenic effects and QTL allelic effects is zero. Selection will introduce a non-zero covariance between polygenic effects and allelic effects at the QTL in later generations.

Each generation the variance before selection of allelic effects at the QTL is computed. We assume that the expectations and variances of the QTL allelic effects in the selected group of sires and dams completely determine the variance of QTL alleles in the next generation. In the next generation the variance of a paternal allelic effect at the QTL is:

$$\sigma^2(v_i^p) = (1-q_s)\ \sigma^2(v_s^p) + q_s\ \sigma^2(v_s^m) + (1-q_s)\ q_s \left(E(v_s^p - v_s^m)\right)^2 \tag{9}$$

where $\sigma^2(v_i^p)$ is the variance of the paternal QTL allele of animal i, $\sigma^2(v_s^p)$ is the variance after selection of the paternal QTL allele of the sire, $\sigma^2(v_s^m)$ is the variance after selection of the maternal QTL allele of the sire, $E(v_s^p)$ is the expectation after selection of the paternal QTL allele of the sire, and $E(v_s^m)$ is the expectation after selection of the maternal QTL allele of the sire.

The variance of polygenic effects before selection is:

$$\sigma^2(u_i) = 0.25\ \sigma^2(u_s) + 0.25\ \sigma^2(u_d) + 0.5\ \sigma^2(u^0) \tag{10}$$

Selection index. Selection is for an index including observations on full-sibs, half-sibs, own performance for hens, and estimated breeding values of parents. Estimated breeding values of parents are based on sibs of the parents and information from earlier generations but not on information from offspring. The selection index is computed by regression on groups of observations. The groups are defined such that some groups are absent in the conventional index without marker information and present in the marker assisted selection index. For conventional selection (CS) the index is:

$$\hat{A}_c = b_{c1}(\overline{DP}-\overline{SP}) + b_{c2}\overline{SP} + b_{c3}(\hat{A}_s+\hat{A}_d) \tag{11}$$

for cocks and

$$\hat{A}_h = b_{h1}(\overline{DP}-\overline{SP}) + b_{h2}\overline{SP} + b_{h3}(\hat{A}_s+\hat{A}_d) + b_{h4}X_h \tag{12}$$

for hens, where c denotes cock, h denotes hen, $\hat{A}$ is estimated breeding value, $\overline{DP}$ is the mean performance of all progeny of the dam, $\overline{SP}$ is the mean performance of all progeny of the sire, X_h is hen own performance, and b's are the regression coefficients.

The selection index is expanded to accommodate marker information. Regression on the difference in mean performance between groups of offspring that inherit alternative marker alleles from the parent is included as well as regression on estimated parental QTL effects:

$$\begin{aligned}\hat{A}_c^* = {} & b_{c1}^*(\overline{DP}-\overline{SP}) + b_{c2}^*(\overline{SP}) + b_{c3}^*(\hat{A}_s^* + \hat{A}_d^*) + \\ & b_{c5}^*(\overline{DP_d^p}-\overline{DP_d^m}) + b_{c6}^*(\overline{SP_s^p}-\overline{SP_s^m}) + b_{c7}^*(\hat{v}_s^p-\hat{v}_s^m) + b_{c8}^*(\hat{v}_d^p-\hat{v}_d^m)\end{aligned} \tag{13}$$

for cocks and

$$\begin{aligned}\hat{A}_f^* = {} & b_{h1}^*(\overline{DP}-\overline{SP}) + b_{h2}^*\overline{SP} + b_{h3}^*(\hat{A}_s^* + \hat{A}_d^*) + b_{h4}^* X_h + \\ & b_{h5}^*(\overline{DP_d^p}-\overline{DP_d^m}) + b_{h6}^*(\overline{SP_s^p}-\overline{SP_s^m}) + b_{h7}^*(\hat{v}_s^p-\hat{v}_s^m) + b_{h8}^*(\hat{v}_d^p-\hat{v}_d^m)\end{aligned} \tag{14}$$

for hens, where $\hat{A}^*$ is the estimated breeding value if marker information is used, $\overline{DP}_d^p$ is the average performance of progeny of the dam that inherit the paternal marker allele from the dam, $\overline{DP}_d^m$ is the average performance of progeny of the dam that inherit the maternal marker allele from the dam, $\overline{SP}_s^p$ is the average performance of progeny of the sire that inherit the paternal marker allele from the sire, $\overline{SP}_s^m$ is the average performance of progeny of the sire that inherit the maternal marker allele from the sire, $\hat{v}$ is the estimated effect of a QTL allele.

A within family selection index is defined that estimates the deviation of the breeding value of the selection candidate from the full-sib family mean. The juvenile within family selection index is:

$$\text{WF-ju} = b_{j1}(\hat{v}_s^p - \hat{v}_s^m) + b_{j2}(\hat{v}_d^p - \hat{v}_d^m) \tag{15}$$

and the adult within family selection index is

$$\begin{aligned}\text{WF-ad} = {} & b_{a1}(\overline{DP_d^p} - \overline{DP_d^m}) + b_{a2}(\overline{SP_s^p} - \overline{SP_s^m}) + \\ & b_{a3}(\hat{v}_s^p - \hat{v}_s^m) + b_{a4}(\hat{v}_d^p - \hat{v}_d^m)\end{aligned} \tag{16}$$

The selection indices do not allow for grandparental origin of parental marker alleles, i.e., animal *A* with a sire that inherited a paternal marker allele from the grandsire, is not distinguished from animal *B* with a sire that inherited a maternal marker allele from the grandsire. Therefore, for sire and dam effects (co)variances like $\sigma^2(v_s^p)$ are used that are independent from marker information. To compute unconditional (co)variances the general formula (Biswas 1991): $\text{cov}(X,Y) = E(\text{cov}(X,Y\,|\,Z)) + \text{cov}(E(X\,|\,Z),E(Y\,|\,Z))$ is used.

Computation of selection intensities. For each selection step we consider four indices, one for each marker genotype. Each index has an expectation and a variance. Due to differences in expectations and variances of indices, the proportions selected per index differ. For the final male selection step and the female selection step, the algorithm described by Ducrocq and Quaas (1988) has been used to compute the selection fractions and selection intensities for the four indices. These selection intensities are then corrected for finite population size (Burrows 1972).

For within family selection of cocks each index has only one realisation within a family. No deterministic method is available to compute selection intensities for this case. Therefore, a simple stochastic simulation was used. For each index one random realisation was simulated using the expectation and variance of the index. This realisation was assigned to a quarter of the cocks in the family. Then, the cocks with the best index were assumed to be selected. The marker genotype of the selected cocks and the index value were recorded. Each generation this was repeated 10 000 times. The proportion selected per index, the selection intensity per index, and the reduction of variance per index were computed from simulation results.

Table 1 QTL-, polygenic- and total- response per generation for standard breeding scheme with Conventional Selection (CS) or Marker Assisted Selection (MAS)

	QTL response			Polygenic response		Total response	
t[a]	CS	MAS[b]		CS	MAS	CS	MAS
0	.155	-		.621	-	.776	-
4	.110	-		.460	-	.570	-
5	.108	.178	(.034)	.461	.445	.569	.622
10	.100	.166	(.042)	.464	.450	.564	.616
15	.093	.141	(.038)	.467	.458	.560	.600
20	.087	.122	(.034)	.469	.465	.556	.587

[a] t is the generation. [b] Between brackets is the QTL response to within family selection of cocks.

Table 2 QTL variance ($\sigma^2(q)^b$), polygenic variance ($\sigma^2(u)$) and genetic variance ($\sigma^2(a)$) before selection of different genetic effects for standard breeding scheme with Conventional Selection (CS) or Marker Assisted Selection (MAS)

	$\sigma^2(q)$		$\sigma^2(u)$		$\sigma^2(a)$	
t[a]	CS	MAS	CS	MAS	CS	MAS
0	.060	-	.240	-	.300	-
4	.055	-	.198	-	.232	-
5	.054	.054	.198	.198	.231	.231
10	.050	.046	.197	.200	.228	.223
15	.046	.038	.197	.199	.225	.217
20	.043	.033	.196	.198	.223	.213

[a] t is generation. [b] $\sigma^2(q)$ is the variance of quantitative trait *locus* i.e. $\sigma^2(v^p+v^m)$.

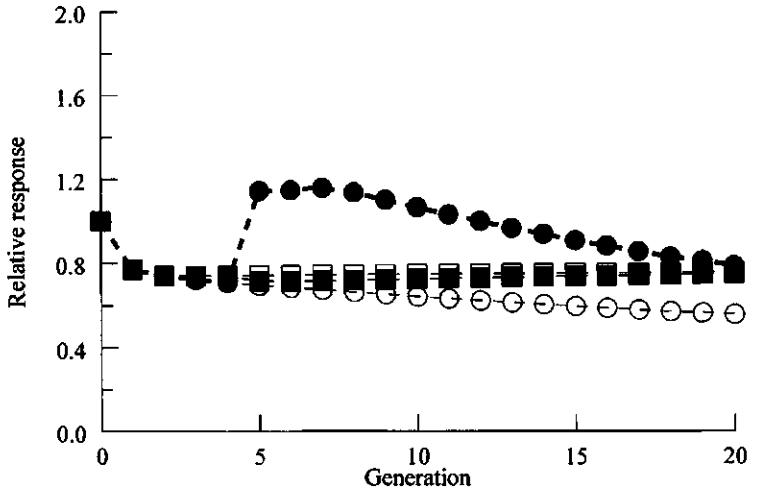


Figure 1 QTL response to conventional selection (○), polygenic response to conventional selection (□), QTL response to marker assisted selection (●), and polygenic response to marker assisted selection (■) in generations 0 to 20 for standard breeding scheme. All responses are relative to corresponding response in generation zero

Simulated schemes. In the standard scheme 50 sires were used per generation. Each sire was mated to six dams and each dam produced 30 offspring, 15 cocks and 15 hens. *Max-c* was 6. In the unselected and unrelated base generation, heritability was .3 and the proportion of the genetic variance due to the QTL was .2. The recombination rate between marker and QTL was .05. Population structure varied leaving genetic parameters constant. Number of sires (N_s) and number of offspring per sire (180) were constant. For number of dams per sire (N_d) the values 3, 6 and 9 were used. Number of offspring per dam (N_o) was $180/N_d$. For *max-c* the values 3, 6 and 9 were used. Schemes representing all nine combinations of N_d and *max-c* were simulated. In other alternative schemes, genetic parameters varied leaving population structure constant. For proportion of genetic variance due to the QTL the values 0, .1, .2, .4, .6, and .8 were used and for heritabilities the values .1, .3 and .5. Schemes representing all 18 combinations of proportion of genetic variance due to the QTL and heritability were simulated.

Results

Standard breeding scheme. Table 1 gives the responses to within family selection and total response per generation for the standard breeding scheme. The response per generation was highest in generation zero. After generation zero response decreased due to reduced genetic variances (Table 2). Response did not reach an equilibrium because QTL variance decreased continuously (Table 2). Response increased after introduction of a marker in generation 5. QTL response increased from .110 in generation 4 to .178 in generation 5 while polygenic response decreased from .460 in generation 4 to .445 in generation 5. The within family response in generation 5 was .034 which is .19 times the QTL response. After generation 5 within family response first increased and then gradually decreased, whereas QTL response due to final male selection and female selection constantly decreased. In generation 20 the within family response was .28 times the QTL response. Polygenic response increased after generation 5 both for conventional selection (CS) and for marker assisted selection (MAS). Over generations polygenic response increased more for MAS than for CS. QTL variance reduced more when MAS was applied. So, the proportion of genetic variance due to polygenes increased more for MAS than for CS. In absolute values polygenic variance hardly differed between MAS and CS.

Figure 1 gives the QTL response per generation relative to the QTL response in generation zero. Further, the polygenic response per generation relative to polygenic response in generation zero is given. Polygenic response was fairly constant after two generations with a slight decrease after the introduction of marker information. QTL response to conventional selection decreased constantly. QTL response increased sharply after introduction of a marker in generation 5. After generation 5 the QTL response to MAS decreased at a higher rate than for conventional selection. As a result, the difference in QTL response between MAS and conventional selection slightly decreased over generations.

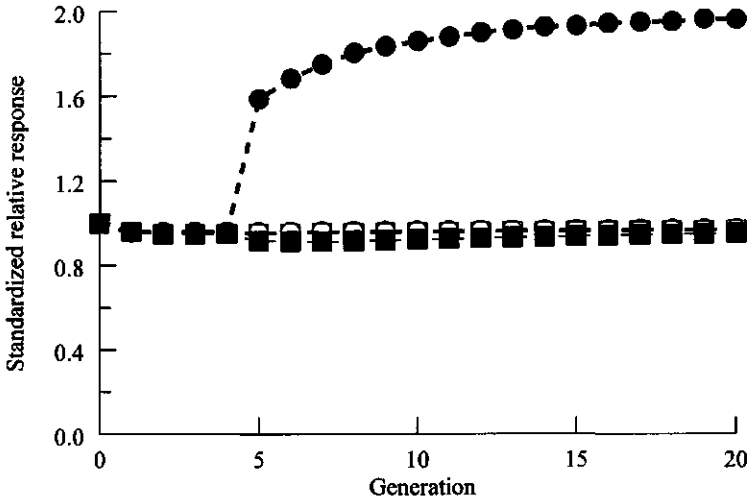


Figure 2 Standardized QTL response to conventional selection (○), standardized polygenic response to conventional selection (□), standardized QTL response to marker assisted selection (●), and standardized polygenic response to marker assisted selection (■) in generations 0 to 20 for standard breeding scheme. All standardized responses are relative to standardized response in generation zero

Table 3 Response per generation and cumulative response for conventional selection for various breeding schemes[a]

N_d	N_o	*max-c*	Generation 5	Generation 9	Cumulative over generations 5 to 9
3	60	3	.553	.548	2.751
		6	.589	.584	2.935
		9	.610	.604	3.034
6	30	3	.536	.531	2.668
		6	.569	.565	2.836
		9	.588	.584	2.928
9	20	3	.523	.520	2.608
		6	.555	.553	2.769
		9	.573	.571	2.859

[a] Varied are number of dams per sire (N_d), number of full-sib offspring per dam (N_o) and the maximum number of cocks selected per full-sib family (*max-c*)

Figure 2 gives the standardized QTL response and standardized polygenic response. In each generation, QTL response is divided by the covariance between QTL and breeding value ($\sigma(v_i^p+v_i^m,a_i)$) before selection in that generation. Polygenic response is divided by the covariance between polygenic effect and breeding value ($\sigma(u_i,a_i)$). Standardization was used to correct for the effect of variance reduction. Standardized polygenic response to MAS, standardized polygenic response to conventional selection, and standardized QTL response to conventional selection can hardly be distinguished. Standardized QTL response to MAS differed clearly from the other three. Standardized QTL response increased sharply after introduction of marker information in generation 5. After generation 5 standardized QTL response to MAS increased further while the other standardized responses remained constant. The results for CS in Figure 2 show that reduced variance influenced polygenic response and QTL response similarly although selection influences polygenic variance differently than QTL variance. Figure 2 further shows that standardized QTL response was constant without markers, but increased over generations when a marker was used. The decrease in absolute QTL response over generations (Figure 1) shows that reduced QTL variance had a larger effect than increased standardized QTL response.

Alternative schemes. Table 3 gives the response in generations 5 and 9, and cumulative response over generations 5 to 9 to conventional selection for various breeding schemes. Number of dams per sire (N_d), number of offspring per dam (N_o) and *max-c* varied between schemes. Number of sires (N_s) and number of offspring per sire ($N_d \times N_o$) were constant. Response per generation increased with increasing *max-c* and with decreasing N_d. The reason for this is that male final selection intensity increased with increasing *max-c* and female selection intensity increased with decreasing N_d.

Table 4 gives the additional response due to markers in percent of response to CS. Cumulative additional response was 6 to 13% of the cumulative conventional response. Additional response was higher for schemes with adult within family selection than for schemes with juvenile within family selection. This was especially true in generation 5 where additional response varied from 4.6 to 5.0% for juvenile schemes and from 5.7 to 12.2% for adult schemes. In generation 5 within family response is zero for juvenile schemes because there is no information to estimate within family deviations. For adult schemes in generation 5, sib information is available to estimate within family deviations. Cumulative over generations 5 to 9, additional response in juvenile schemes varied from 6.0% for N_o at 20 and *max-c* at 9 to 9.7% for N_o at 20 and *max-c* at 3. Cumulative additional response in adult schemes varied from 6.2% for N_o at 20 and *max-c* at 9 to 12.7% for N_o at 20 and *max-c* at 3. Additional response increased with decreasing *max-c*. With decreasing *max-c*, the within family selection intensity increased and therefore response to within family selection

increased. Additional response also increased with decreasing N_d at a given proportion of male offspring selected from within a family, i.e., additional response is higher with N_d at 6 and 3 out of 30 selected, than with N_d at 3 and 6 out of 60 selected. With increasing N_d the differences in precision of the effects of paternally and maternally derived QTL alleles increased. Differences between mean values of paternal and maternal alleles after selection increased with increasing N_d. Different expectations for maternal and paternal QTL alleles result in different expectations for the four selection indices. These differences can be exploited in selection.

Table 4 Additional response due to MAS as a percentage of conventional response for various breeding schemes

N_d	N_o	*max-c*	i_{rw}	Generation 5		Generation 9		Cumulative over 5 to 9	
				ju	ad	ju	ad	ju	ad
3	60	3	1.036	4.7	12.2	9.1	10.1	8.2	11.1
		6	1.039	4.9	11.6	9.2	10.1	8.1	10.8
		9	.916	5.0	10.6	8.4	9.1	7.5	9.8
6	30	3	1.048	4.6	11.6	11.1	12.6	9.2	12.4
		6	.774	4.8	9.3	9.1	9.6	7.8	9.7
		9	.518	4.9	7.6	7.3	7.3	6.6	7.5
9	20	3	.936	4.7	10.8	12.1	13.4	9.7	12.7
		6	.523	4.9	7.7	8.7	8.7	7.5	8.5
		9	.121	5.0	5.7	6.5	6.5	6.0	6.2

Varied are number of dams per sire (N_d), number of full-sib offspring per dam (N_o) and the maximum number of cocks selected per full-sib family (*max-c*)

i_{rw} is the realized within family selection, computed as the within family response divided by the standard deviation of the within family index, averaged over generations 5 to 9. The given values are for adult within family selection. Realized intensity of juvenile within family selection, computed over generations 6 to 9 because juvenile within family selection response is zero in generation 5, was between .1% and 1.8% higher than realized intensity of adult within family selection

ju is the additional response in % of conventional response for schemes with juvenile within family selection

ad is the additional response in % of conventional response for schemes with adult within family selection

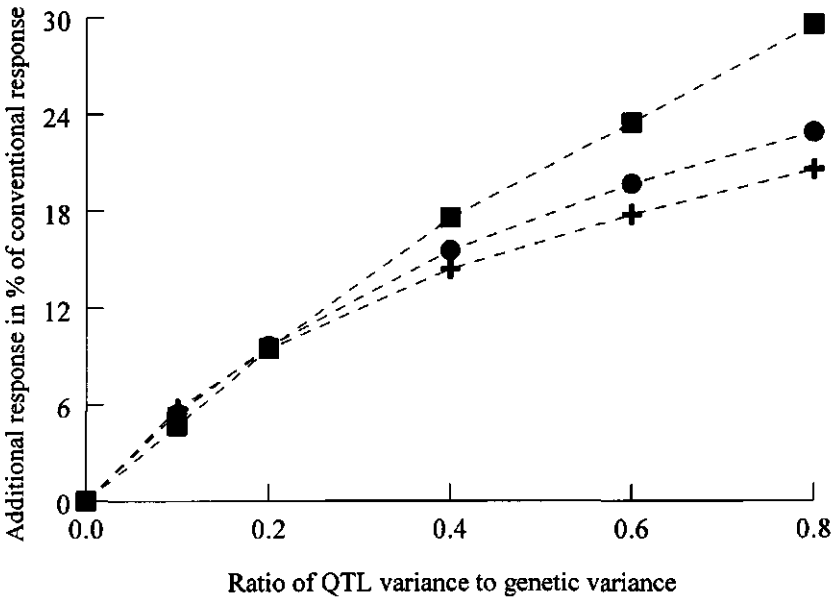


Figure 3 Effect of proportion of genetic variance due to QTL on additional response for heritabilities at .1 (■), .3 (●), and .5 (+)

Figure 3 shows how the proportion of genetic variance due to the QTL and heritability affect additional response. If the proportion of the genetic variance due to the QTL was .10 then additional response was highest for a heritability of .5 and lowest for a heritability of .1. If the proportion of the genetic variance due to the QTL was .20 then additional response hardly changed by changing heritability. For proportions due to QTL higher than .20, additional response was highest for a heritability of .1 and lowest for a heritability of .5. For a proportion due to the QTL of .20, as in the standard scheme, additional response was 9.5% for a heritability of .1, 9.7% for a heritability of .3, and 9.5% for a heritability of .5. For a proportion due to the QTL of .80, additional response was 27.8% for a heritability of .1, 22.9% for a heritability of .3, and 20.6% for a heritability of .5.

Discussion

Use of a marker linked to a QTL that explained a proportion of .2 of the genetic variance of a sex limited trait increased response with 6 to 12.7%. Increasing the proportion of the genetic variance explained by the QTL to .8 resulted in increases in response up to 27.8%. In other studies MAS resulted in 20 to 30% (Kashi *et al.* 1990), 10 to 25% (Meuwissen and van Arendonk 1992) and 40% (Stam 1986) increases in response. Stam (1986) studied sib selection of young bulls in a dairy cattle breeding scheme. The 40% increase was realized assuming that the true breeding value of a bull's sire was known with using markers and all genetic variance was due to one locus for which segregation could be observed. Kashi *et al.* (1990) and Meuwissen and Van Arendonk (1992) looked at the situation that the whole genome was covered with markers that were used to find associations in large groups of offspring of elite sires. Kashi *et al.* (1990) selected young bulls before entering the progeny test based on the marker alleles they inherited from their grandsires. In Meuwissen and Van Arendonk (1992) there was only one selection step. Marker associations found in grandsires were used to estimate within family deviations either in a progeny testing scheme or in a nucleus breeding scheme.

There are important differences between our study and the two last studies mentioned above. The within family selection of cocks in our study is similar to marker assisted selection of young bulls in Kashi *et al.* (1990). The final selection steps of cocks and hens in our study are similar to the use of markers in a nucleus herd as in Meuwissen and Van Arendonk (1992). So, we combined two means of getting additional response but our additional responses are lower than in either Kashi *et al.* (1990) or Meuwissen and Van Arendonk (1992). The reasons for this are: (1) we compute cumulative response over 5 generations using a model in which QTL variance and consequently QTL response declines over generations, where Kashi *et al.* (1990) compute response for one generation and Meuwissen and Van Arendonk (1992) compute equilibrium response assuming no decline in QTL response; 2) we consider one QTL explaining only part of the genetic variance and not the whole genome; 3) in dairy cattle breeding programmes more information is available to estimate marker allelic effects. In a poultry breeding programme only information from within the nucleus can be used for breeding value estimation, whereas in the dairy cattle breeding schemes marker allelic effects in elite sires are estimated using many offspring from outside the nucleus. Of course, the structure of a poultry breeding programme can be changed to estimate marker allelic effects more accurately. Because of such a change the polygenic part of the breeding value will, however, be estimated more accurately too. Consequently, additional response will not increase proportionally.

Additional response increased with increasing QTL variance. For proportions of genetic variance due to the QTL of above .20, additional response increased with decreasing

heritability. For proportions of genetic variance due to the QTL under .20, additional response decreased with decreasing heritability. This observed interaction is not fully understood.

We assumed that the transmission of all marker alleles is always traceable. For microsatellite markers, a highly polymorphic class of markers, on average a heterozygosity of .28 was observed in commercial layer lines (Groen *et al.* 1994). Allele frequencies were estimated for several microsatellite markers and heterozygosity was estimated assuming Hardy Weinberg equilibrium (Groen *et al.* 1994). Given this level of polymorphism, marker transmission will often be untraceable which will have a negative influence on the benefits of MAS (Kashi *et al.* 1990). The use of several closely linked markers largely solves the problem of untraceable marker transmission (Kashi *et al.* 1990). This, however, will generate additional costs of typing.

Inbreeding was ignored in this study. Without selection we can predict from the number of sires and dams used per generation that inbreeding would increase with a rate of .3% per generation. At this rate, inbreeding will hardly affect response. With selection, however, inbreeding will increase more rapidly since sibs will be co-selected, especially males. This will especially lower the level of response of schemes with a high *max-c* and high N_e. The exact impact of inbreeding on QTL response is hard to predict.

Responses were given per generation. The effect of generation interval was not included. For all schemes, either with or without the use of markers, final selection of males and females was after the females in a generation had completed their first record. The use of markers did not alter the generation interval and therefore additional response expressed in percentage of response to conventional selection was independent of generation interval. Comparisons over schemes were based on additional response. So, although generation interval might differ between schemes, comparisons were not affected by generation interval.

The genetic model in this study assumed a QTL with allelic effects that follow a normal distribution. Selection reduced the variance of the allelic effects. Recombination during the forming of new gametes did not counterbalance the reduction of QTL variance. In each round of selection, therefore, QTL variance was further reduced whereas for polygenic variance, recombination and selection reached an equilibrium. The proportion of the genetic variance due to the QTL, therefore, decreased over generations and consequently also the additional response due to MAS. Polygenic response, after an initial decrease due to a build up of linkage disequilibrium after selection in generation zero, gradually increased over time. This can be explained as follows: QTL variance decreases and therefore also negative covariance between polygenic and QTL effects. Therefore, the covariance between additive genetic effects and polygenic effects increases and also the polygenic response. Polygenic response was lower with MAS than with CS. This is due to the negative covariance

between QTL effects and polygenic effects. QTL response is higher with MAS than with CS and consequently also the correlated negative effect on polygenic response is higher with MAS than with CS.

We believe our genetic model is more realistic than a model that assumes a constant QTL response over generations. Our model, however, sets no limit to the cumulative QTL response. This is unrealistic for a single locus with a finite number of allelic effects in the base generation. A solution could be the introduction of a "best allele model" in which the population mean for the QTL cannot increase above the value of an animal homozygous for the best QTL allele in the base generation. A heuristic approach might be to let the QTL variance after selection depend on the cumulative QTL response. Alternatively the single locus model can be interpreted as a model for a cluster of many closely linked loci. For a given QTL variance in the base generation long term response will increase with increasing number of underlying genes and our model will become more realistic. The same argument applies for a situation with several independent marked QTLs.

We studied the additional response due to MAS for a poultry breeding nucleus assuming favourable conditions like a sex limited trait, a restriction on the number of cocks selected per full-sib family, and traceable marker alleles. For a commercial breeder increased response should lead to an improved market share or a higher price for its products. The costs are in collecting blood, isolating DNA and typing for marker loci. With the ongoing development of marker technology it is hard to quantify future costs of MAS. Predicting effects of additional response on market share is harder and will depend on the current position and strategies used by competitors. Given its position, the commercial poultry breeder can compare the costs and benefits of marker assisted selection. Results of this study can help in this decision process and in implementing the new technology efficiently.

Acknowledgements

The authors acknowledge financial support from Euribrid B.V. We thank T.H.E. Meuwissen and an anonymous reviewer for helpful comments.

References

Biswas S (1991) Topics in statistical methodology. John Wiley & Sons, New York.

Burrows PM (1972) Expected selection differentials for directional selection. Biometrics 28:1091-1100.

Cochran WG (1951) Improvements by means of selection. In: Neyman J (ed) Proceedings of the 2nd Berkeley symposium on mathematics, statistics and probability.

Dentine MR (1992) Marker-assisted selection in cattle. Anim Biotech 3:81-93.

Ducrocq V, Quaas RL (1988) Prediction of genetic response to truncation selection across generations. J Dairy Sci 71:2543-2553.

Fernando RL, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Genet Sel Evol 21:467-477.

Groen AF, Crooijmans RPMA, Van Kampen AJA, Van der Beek S, Van der Poel JJ, Groenen MAM (1994) Microsatellite polymorphism in commercial broiler and layer lines. Proc 5th World Congr Genet Appl Livestock Prod 21:94-97.

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

Meuwissen THE, Van Arendonk JAM (1992) Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. J Dairy Sci 75:1651-1659.

Smith C (1967) Improvement of metric traits through specific genetic loci. Anim Prod 9:349-358.

Smith C, Simpson SP (1986) The use of genetic polymorphisms in livestock improvement. J Anim Breed Genet 103:205-217.

Soller M (1978) The user of loci associated with quantitative traits in dairy cattle improvement. Anim Prod 27:133-139.

Soller M (1994) Marker assisted selection - an overview. Anim Biotech 5:193-207.

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25-33.

Stam P (1986) The use of marker loci in selection for quantitative characters. In: Smith C, King JWB, McKay J (eds) Exploiting New Technologies in Livestock Improvement: Animal Breeding. Oxford University Press, New York.

Appendix

In this appendix we will show how to compute the within family selection response in generation 10 of the standard scheme.

From equation (2), the contribution to selection response of within family selection is:

$$.5 \sum_{j=1}^{4} f_{wtj} \times \left(m_{wtj} - \bar{m}_{wt} + i_{wtj} \times \sigma(\hat{A}_{wtj})\right) = .5 \left(\sum_{j=1}^{4} f_{wtj} \times \left(m_{wtj} + i_{wtj} \times \sigma(\hat{A}_{wtj})\right) - \sum_{j=1}^{4} .25 \times m_{wtj}\right) \quad \text{(A1)}$$

where f_{wtj} is the fraction selected of cocks with within family selection index j in generation t, m_{wtj} is the mean genetic value before within family selection of cocks with within family selection index j in generation t, $\bar{m}_{wt}$ is the mean of all cocks before within family selection in generation t, i_{wtj} is the selection intensity within the group of cocks with within family selection index j in generation t, and $\sigma(\hat{A}_{wtj})$ is the standard deviation of within family selection index j in generation t. The group of offspring that inherit two paternal marker alleles have index 1, the group that inherit a paternal allele from the sire and a maternal from the dam have index 2, the group that inherit a maternal allele from the sire and a paternal from the dam have index 3, and offspring that inherit two maternal marker alleles have index 4. Before within family selection the four groups are of equal size.

The components required to compute within family selection response are given in Table A1. Before selection, offspring inheriting a paternal marker allele from the sire are on average better than offspring inheriting a maternal marker allele from the sire. For a given sire allele, animals inheriting a paternal allele from the dam are better than animals inheriting a maternal one. As a result, after selection animals with two paternal alleles are most frequent, and animals with two maternal alleles are least frequent. This means that groups with lowest mean were most intensely selected. Equation (A1) shows that the within family selection response can be computed from group means and fractions before and after selection as: $.5 \times (.397 \times 6.303 + .262 \times 6.278 + .221 \times 6.278 + .120 \times 6.264 - .25 \times (6.262 + 6.211 + 6.194 + 6.142)) = .042$, which is the within family response as also given in Table 1 for generation 10.

Table A1 Elements required to compute the within family selection response to marker assisted selection in the standard scheme

Index	Fraction before selection	Mean before selection	Standard deviation of index	Selection intensity	Fraction after selection	Mean after selection
1	.250	6.261	.096	.438	.397	6.303
2	.250	6.211	.098	.705	.262	6.278
3	.250	6.194	.102	.810	.221	6.278
4	.250	6.143	.104	1.135	.120	6.264

Chapter 6

General discussion

Research on and application of genetic markers in a poultry breeding programme involves several activities. In this chapter I discuss aspects of these activities. The chapter is divided into six sections: (I) the chicken linkage map, (II) identified QTLs in chicken, (III) the design of QTL mapping experiments, (IV) possible applications of genetic markers (V) three specific applications of genetic markers in a poultry breeding programme, and (VI) summary.

I The chicken linkage map

More than 460 genetic markers are placed on the chicken linkage map (Burt *et al.* 1995). The chicken linkage map is the composite of two maps; one based on female meioses in the Compton reference population (Bumstead and Palyga 1992) and one based on male meioses in the East Lansing reference population (Crittenden *et al.* 1993). The estimated length of the chicken linkage map based on meioses is 2500 to 3000 cM (Levin *et al.* 1994). The male and female linkage map do not seem to differ in length (Crittenden pers. com.). Based on chiasmata counts in oocytes a length of 2900 to 3200 cM (Rodionov *et al.* 1992) was estimated, corresponding well to the meioses based estimate. In this cytogenetic study of Rodionov *et al.* (1992), the male and female linkage maps were of equal length as well. Preliminary data from a study using markers covering only 400 cM of the chicken genome, but using many more informative meioses than available in either reference family, however, suggest that (parts of) the male map is larger than the female map (Groenen pers. com.). The chicken genome consists of 39 chromosome pairs that show large variation in size. The five macrochromosomes cover 40% of the linkage map (1200 cM), but represent 70% of the genomic DNA (Rodionov *et al.* 1992). The chicken genome contains many very small chromosomes, the microchromosomes. Compared to their physical length, microchromosomes account for a large part of the linkage map. A chromosome has a minimum recombination length of 50 cM irrespective of the physical chromosome length because in meiosis at least one crossover takes place on each chromosome (Rahn and Solari 1986; Burt and Bell 1987; Rodionov *et al.* 1992). This explains why the recombination length of the chicken genome is comparable to the genome length of other domesticated species while the physical length of the chicken genome is considerably smaller. The macrochromosomal crossing over rate is, however, similar to that of the female human genome (Rodionov *et al.* 1992).

From the linkage map markers are selected to be used for QTL mapping experiments. I will call this set of markers the reference map. Markers on the reference map have to meet three criteria. First, they have to be polymorphic in the population used for the QTL mapping experiment. This implies that each experiment requires a specific reference map. Second, markers should fit in groups for simultaneous typing. Third, the markers ought to cover the complete genome and for efficiency should be more or less equally spaced. A 20 cM distance

between adjacent polymorphic markers in the reference set is often used as a standard (Botstein *et al.* 1980). At this spacing 1200 cM / 20 cM = 60 markers are required to cover the macrochromosomes. At least one marker per microchromosome is also required, although two markers per microchromosome may be, however, required if microchromosomes contain recombination hot spots (Rodionov *et al.* 1992) that virtually divide the microchromosome in two independently segregating chromosome segments. So, a reference map with 20 cM spacing includes between 60 + 35 = 95 and 60 + 2×35 = 130 polymorphic markers. More markers are needed for experiments with populations in which the polymorphism of markers is moderate. This may be the case in chicken. In commercial outbred layer and broiler populations, microsatellites show a 30 to 60% heterozygosity (Groen *et al.* 1994).

Currently, most mapped chicken genetic markers are RFLP and RAPD type markers (Bumstead and Palyga 1992; Levin *et al.* 1994). The number of microsatellite markers placed on the linkage map is, however, rapidly increasing (Cheng and Crittenden 1994; Crooijmans *et al.* 1994a; Crooijmans *et al.* 1994b; Khatib *et al.* 1993; Crooijmans *et al.* 1995a). Microsatellite markers are moderately to highly polymorphic, a property required for efficient linkage mapping or QTL mapping. Around 150 microsatellites are at the linkage map and more than 200 unmapped microsatellite markers are available (Crooijmans pers. com.).
It is unknown how well the linkage map covers the genome. Currently, linkage groups are assigned to only three microchromosomes (Burt *et al.* 1995). To enable QTL mapping experiments in which potentially all segregating QTLs can be detected, a linkage map covering all microchromosomes is required. For coverage of microchromosomes, it is not necessary to know which linkage group belongs to which microchromosome. It is sufficient to know that a specific linkage group belongs to a different chromosome than the other linkage groups. The number of independent linkage groups will then equal the number of chromosomes covered. Currently, markers are assigned to different linkage groups if they are not significantly linked. For optimal use of the linkage map, determining that markers are significantly *un*linked is equally important. Linkage groups are independent if for all possible pairs of a marker from one linkage group and a marker from another linkage group, the markers are significantly unlinked.

II Identified QTLs in chicken

Few genes of economic importance have been identified in chicken (Merat 1990). Among these genes, the dwarf genes, the naked neck gene, and sex-linked genes used for sexing of chicks are of commercial interest. For the many biochemical polymorphisms studied, Merat (1990) concluded that estimated effects differed markedly over studies and were hardly ever significant. The same is true for blood group polymorphisms except for the

B locus (the chicken MHC) which has a well documented effect on susceptibility for Marek's Disease. Another, well studied, class of genes comprises the endogenous viral genes or *ev*-genes. *Ev*-genes are DNA sequences in the chicken genome that have a high degree of homology to exogenous avian leukosis viruses (Gavora *et al.* 1991). Two *ev*-genes from which complete endogenous viruses are transcribed, reduced annual egg production rate by 9% (Gavora *et al.* 1991). Another completely transcribed *ev*-gene is at the sex-linked slow feathering locus used for sexing chickens.

Recently, associations between anonymous marker loci and quantitative traits have been studied. A DNA fingerprint band with an effect of 0.9 standard deviations (not specified which standard deviation) on abdominal fat (Plotsky *et al.* 1993), and a DNA fingerprint band having an effect of more than two residual standard deviations on shank length and body weight (Dunnington *et al.* 1992) were found. Lakshmaman *et al.* (1994) identified 101 DNA fingerprint bands associated with 8 different traits. The practical use of these results is limited because the DNA fingerprint bands are not mapped on the linkage map. In another study, significant associations between morphological markers and age at first egg and body weight were found (Shoffner *et al.* 1993). From their tables I approximate that allelic effects were of the order of 0.2 residual standard deviations. The morphological markers used in the study of Shoffner and coworkers are placed on the linkage map. Khatib (1994) studied the association of 28 microsatellite markers, that covered over 50% of the chicken genome, with juvenile growth rate in a cross between White Leghorn layer females and a single White Rock broiler male. Eleven markers were significantly associated with growth rate, but sizes of effects were not given.

Khatib's study (1994) is the first step towards a whole genome QTL mapping experiment. A whole genome QTL mapping experiment uses markers covering the whole, or most of the, genome. From now on, the term 'QTL mapping experiment' will be used to refer to an experiment using a linkage map covering the whole genome.

Powerful QTL mapping experiments are required to answer questions such as: how many genes are responsible for genetic variation in the expression of economic traits, how are these genes spread over the genome, and what is the distribution of allelic effects. Information from experiments in poultry is currently limited but some inferences can be based on experiments in other species.

QTL mapping experiments in plants and Drosophila suggest that genetic variation is due to few genes with large effects and many genes with small effects (Shrimpton and Robertson 1988; Edwards *et al.* 1987; Paterson *et al.* 1988,1991; Stuber *et al.* 1992; deVicente and Tanksley 1993). The effects of spontaneous mutations in Drosophila show a similar distribution (Keightley 1994). Some QTLs have already been identified in chicken (Khatib 1994). QTL mapping experiments (plants: e.g., Edwards *et al.* 1987; Paterson *et al.*

1988,1991; Stuber *et al.* 1992; deVicente and Tanksley 1993; pigs: Andersson *et al.* 1994; dairy cattle: Georges *et al.* 1995) show that QTLs can be detected in many species.

The size of the effects of segregating QTL alleles depends on the population under study. Within a selected population alleles of large effect will quickly move to fixation. As a result, the largest effect of the remaining segregating QTL alleles will probably be smaller than the largest effect of the QTL alleles that segregate in an F2 between unrelated populations that differ widely for the studied trait. This expectation is unproven since only one QTL mapping experiment within a single selected population has been reported to date (Georges *et al.* 1995). Georges *et al.* (1995) identified QTLs explaining variance within dairy cattle elite sire families. In some families, the QTLs they identified explained 10 to 50% of the within family variance. Although this, as they point out, is an overestimate of the true values, it shows that QTL alleles of large effect segregate in a highly selected population. Additional within population experiments are needed to compare the effects of QTL alleles found within populations to effects of QTL alleles found in crosses.

Plant QTL mapping experiments identified QTLs in crosses between two commercial populations, between a commercial population and a wild type population, and between divergently selected populations. All studies that will be mentioned here, found QTLs for each trait examined. Identified QTLs together often explained more than 50% of the phenotypic variance in a cross. Individual QTLs could explain as much as 40% of the phenotypic variance. The type and direction of gene action depended on the studied trait and the population used in the cross. Stuber *et al.* (1992) crossed two elite maize inbred lines that produce superior hybrid performance. They observed overdominant gene action for all QTLs identified which agrees with the high heterosis observed. However, Cockerham and Zeng (1995) re-analyzed the data of Stuber *et al.* (1992) and concluded that overdominance is not necessary to explain the results of Stuber *et al.* (1992). Beavis *et al.* (1994), using a cross between the same populations, identified QTLs in other chromosomal regions than Stuber *et al.* (1992). Most QTLs showed dominant gene action and only few showed overdominant gene action (Beavis *et al.* 1994). Perhaps, the results differ because the studies were designed differently. Stuber *et al.* (1992) used a backcross whereas Beavis *et al.* (1994) used an F4; this could mean that genetic background influences gene action of QTLs (Beavis *et al.* 1994). The low power of the experiments is another plausible reason for differences in results of the experiments (Beavis *et al.* 1994). Paterson *et al.* (1988,1991) crossed a wild type tomato with a domesticated tomato. Gene action was additive or partially dominant. For two traits parental population means differed more than 10 phenotypic standard deviations. For all QTLs identified for those two traits the direction of the allelic effects agreed with the difference in parental means, i.e., the QTL allele with the higher effect always came from the population with the higher population mean. For a third trait, parental means hardly differed and

direction of QTL allelic effects was either positive or negative. Edwards *et al.* (1987, 1992) crossed two maize populations differing substantially for the traits measured in their experiments. The direction of QTL allelic effect was consistent with the difference in parental means. For gene action a continuum between additive and overdominant effects was found. Schön (1993) crossed two elite maize lines and identified QTLs of which the positive allele came from either parental line. Goldman *et al.* (1994) crossed divergently selected maize lines and for some identified QTLs, the direction of allelic effects was opposite to the difference between the parental means (Goldman pers. com.). DeVicente and Tanksley (1993) analyzed traits showing transgressive segregation in a tomato cross. They defined transgression as "the appearance of individuals in segregating populations that fall beyond their parental phenotypes". Seventy four QTLs were found for eleven traits. Thirty six percent of the QTLs had an effect opposite to the direction expected from the difference in parental means. The proportion of QTLs with opposite effect was significantly negatively related with the difference between parental means.

From these plant QTL mapping experiments I conclude:

- Few QTLs explain most of the phenotypic variance in a cross between relatively unrelated populations with largely different trait means.
- The proportion of QTLs with an effect opposite to the difference between parental means is negatively correlated to the size of this difference.

With respect to the interpretation of results of QTL mapping experiments it should be noted that resolution of QTL mapping experiments is often not sufficient to decide whether or not a single gene or several closely linked genes underlie an identified QTL. For example, one gene with overdominant gene action cannot be discriminated from two closely linked dominant genes as was shown by Cockerham and Zeng (1995) who re-analyzed the data of Stuber *et al.* (1992) and concluded that overdominance is not necessary to explain the results of Stuber *et al.* (1992). In tomato, isogenic lines were used for fine mapping of QTLs (Eshed and Zamir 1995). A first experiment identified QTLs (Eshed and Zamir 1994) that were further studied in a subsequent experiment. One QTL identified in the first experiment, could be resolved into three linked QTLs in the second experiment (Eshed and Zamir 1995).

QTL mapping experiments are interesting for poultry breeders, especially when QTL alleles can be identified that can improve the level of the populations used in the commercial breeding programme. I will call such QTL alleles 'favourable QTL alleles'. Even in a cross between lines with largely different trait means, within the low population QTL alleles of positive effect can be found that are not at fixation within the high population, as Khatib (1994) showed. Within a layer population he identified QTL alleles with a higher effect on

growth than the alleles of a single broiler male. The probability of finding favourable QTL alleles will be larger if a commercial population is crossed to an unrelated population with an equal trait mean. To enlarge the probability of finding favourable QTL alleles, crosses can first be screened based on the variance they show in an F2. Also, the heterosis level might be a predictor for the presence of favourable QTL alleles in case heterosis is due to dominant alleles that are in repulsion.

III The design of QTL mapping experiments

A QTL mapping experiment should be carefully designed. First, a QTL mapping experiment can only identify QTLs segregating in the population used for the experiment. So, the experimenter should decide which populations can provide useful answers to his questions. Second, a QTL mapping experiment is expensive and thus the gains of optimizing the design may be high.

Many factors influence the efficiency of a QTL mapping experiment, including the structure of the experimental population, the size of the experiment, the reference map density, and the ratio of the number of trait observations over the number of marker typings as discussed in Van der Beek *et al.* (1995). The optimization depends on several parameters including the desired type I and type II (power) errors for the experiment, the potential number of offspring per hen and per cock, the heritability of the trait(s) studied, the density of the available linkage map, the polymorphism of the available markers, the relative costs of marker typing and gathering phenotypic information, the practical setting of the experiment, and the available marker technology. Chapter 4 of this thesis covers several aspects of the design of QTL mapping experiments. Here, I shall first address the influence of the population used for QTL mapping on the QTLs likely to be identified. After that I will describe selective DNA pooling, a QTL mapping technique that can change the way QTLs are mapped and linked markers are used in breeding.

The properties of identified QTLs will depend on the populations used to identify them. I distinguish three QTL classes:

QTLs with alleles that explain the difference *between* two populations *not* used in the same breeding programme, e.g., a broiler population and a layer population. This difference can be studied in a cross between the two populations.

QTLs that explain the difference *between* two populations used in the same breeding programme, e.g., a broiler sire population and a broiler dam population. This difference can be studied in a cross between those populations.

QTLs that explain the differences *within* a population, i.e., between families or between animals. These differences can be studied in a within population experiment.

Eventually an experiment using animals from the population of interest is required to identify the QTLs accounting for the genetic variance in that population. However, first a cross between populations differing extremely for the trait of interest could be studied. In such a cross QTL alleles will segregate that have an effect that is larger than the alleles that segregate within the population of commercial interest. These larger QTL alleles can relatively easily be detected. Possibly, the QTLs with the alleles accounting for the differences between extremely different populations, also have alleles accounting for the differences between less divergent populations or the differences within populations (Edwards *et al.* 1992). This, however, is speculative and remains to be documented.

Starting with a cross between extremely different populations might not be efficient, even if the QTLs identified in such a cross also explain the differences between two comparable populations or the differences within a population. The QTLs identified in a cross between extremely different populations have to be confirmed in a subsequent experiment. This subsequent experiment has only value if QTLs are not only confirmed but if also QTL allelic effects are estimated accurately. For accurate estimates large experiments are required, as large as required for direct detection of QTLs. When only markers are typed for regions to which QTLs have been assigned in previous experiments, indeed less marker typings are required than for a whole genome mapping experiment. It is, however, hard to imagine that for a well designed large experiment, using a previously unstudied genetic resource, only part of the genome will be considered.

QTL mapping using selective DNA pooling. Selective DNA pooling combines selective genotyping with DNA pooling (Darvasi and Soller 1994). DNA is isolated from two blood pools: one pool includes blood from animals from the high tail of a phenotypic distribution and the other pool includes blood from animals from the low tail of a phenotypic distribution. As with normal selective genotyping (Lebowitz *et al.* 1987; Lander and Botstein 1989; Darvasi and Soller 1992), selection is from a group of animals expected to be in linkage disequilibrium. In an outbred population two pools are formed for each full-sib or half-sib family. In a crossbred F2 population two pools are formed for the whole population. The DNA pools are analyzed for markers; not only qualitatively to determine which marker alleles are in the DNA pool, but also quantitatively to estimate the relative marker allelic frequencies. The association between a marker and a quantitative trait can be inferred from differences in the frequencies of a marker allele in the low pool and in the high pool (Darvasi and Soller 1994).

The power of selective DNA pooling, compared with the power of selective individual genotyping, mainly depends on the technical error of estimating the marker allelic frequencies in the DNA pools (Darvasi and Soller 1994). Power of selective DNA pooling is only slightly lower than power of selective individual genotyping if there is no technical error, but decreases considerably with increasing technical error. Replicating the allelic frequency measurements reduces the error (Darvasi and Soller 1994). The reduction in technical error will be largest if the whole procedure starting from blood mixing is replicated.

Selective DNA pooling can considerably reduce the number of marker typings in a QTL mapping experiment. In an intercross between inbred lines only two pools have to be formed whereas for an outbred population with several families, more pools have to be formed that all have to be typed. The efficiency of DNA pooling further depends on the techniques used for individual and pooled typing, and the technical error on the allele frequency estimates. In particular, a multicolour fluorescent automatic sequencer can simultaneously type more markers for each individual sample than for each DNA pool. DNA fragments of the same size but labelled with different dyes, have a small influence on the signal of each other (read-through). For single sample typing this is not a problem, but for interpreting the results of pooled typing it is, because a signal due to read-through cannot be distinguished from a signal due to an allele of low frequency. Therefore, for pooled typing the regions to which alleles of different dyes migrate in a gel should not overlap, whereas for individual typing overlap is allowed. This effectively makes the maximum efficiency of pooled typing equal to the maximum efficiency of single colour typing. In addition, measuring allele frequencies from the output of the automatic sequences is much more labour intensive than scoring alleles. Replicating the measurements to decrease the error on the frequency estimates also reduces the advantage of selective DNA pooling. The advantage of selective DNA pooling further depends on the number of traits in the QTL mapping experiment. For each trait separate DNA pools have to be formed. So the advantage decreases with increasing number of traits in the experiment.

Selective DNA pooling is an efficient tool if the number of pools to be typed is many fold lower than the number of individual typings. The advantages are highest for experiments with crossbred populations and for experiments with two-generation half-sib family structures with very large half-sib families. With the availability of the technique, the optimal design of QTL mapping experiments might change. Several three generation QTL mapping experiments (see chapter 4 for a description of three generation experiments) are planned. In these experiments many more phenotypes than genotypes have to be scored. Perhaps, a two-generation experiment with large half-sib families can replace such a three-generation experiment if selective DNA pooling can be used and there are few quantitative traits to be scored.

IV Possible applications of genetic markers

In this section I describe how markers can be used to improve a poultry breeding programme. I address parentage control, varietal identification and marker assisted selection.

Parentage control. Markers can be used for parentage control. Parentage control requires typing the individual and the potential parents for as many markers as necessary to exclude all potential parents but one. A potential parent is excluded if the individual has inherited an allele not present in the potential parent, e.g. the individual has genotype AA and the potential parent has genotype BB. The number of markers needed for effective parentage control depends on four factors (Soller and Beckmann 1983): (1) whether both or one parent is unknown (often, the individuals' dam is known, but not the sire), (2) the degree of polymorphism at the marker, (3) the number of potential parents, and (4) the required probability of excluding all but one parent.

Assume the individuals' dam is known, that there are 10 potential sires, and that we require an 0.99 probability of excluding all but one sire. If markers have on average two equiprobable alleles, i.e., a 50% chance of heterozygosity, then 33 markers are needed. If markers have on average three equiprobable alleles, i.e., an expected heterozygosity of 67%, then only 14 markers are needed. For commercial layers a heterozygosity of around 50% and for broilers a heterozygosity of around 70% for a selected group of microsatellite markers is realistic (Groen *et al.* 1994).

Varietal identification. Varietal identification can be qualitative, i.e., using markers to decide if a population differs from another population, or quantitative when markers are used to estimate the genetic distance between populations. Efficient estimation of genetic distance is possible when DNA pooling and semi-automated microsatellite typing are used (Crooijmans *et al.* 1995b). Genetic distance estimates can be used to maximize the genetic variety in the gene pool for a given number of animals that form the pool. Genetic distance can also be used to predict heterosis. Groen and Crooijmans (1995) observed correlations between genetic distance and heterosis for economic traits ranging from 0.2 to 0.9 in layers. These correlations are such that heterosis prediction is indicative. A crossing experiment should verify the heterosis before introducing a cross in the breeding programme.

Marker assisted selection. In a recent review, Visscher and Haley (1995) distinguish marker assisted introgression, marker assisted selection in an intercross and marker assisted selection within an outbred population. I will describe these three applications.

Marker assisted selection during introgression

Following Groen and Smith (1995) I shall use donor population for the population containing a QTL allele to be introgressed, and recipient population for the population in which the QTL allele is to be introgressed. During introgression markers can be used to (1) select for the favourable QTL allele and (2) improve selection against the donor line background genome. If the QTL alleles can be identified, then selecting for the favourable QTL allele is straightforward. More likely, however, a QTL mapping experiment will reveal markers linked to a QTL, not the QTL itself. To efficiently introgress a QTL, bracketing polymorphic markers as close to the QTL as possible should be used. A dense linkage map and an accurate estimate of the QTL location are favourable. Because the QTL location is estimated with some error, use of only the markers closest to the estimated QTL location is not necessarily best; as closer the bracketing markers are to the estimated location, the more likely the true QTL location is outside the selected bracket. This risk is reduced when more distant markers are used. Within a wider bracket double crossovers are more probable than within a small bracket. Because a double crossover also can lead to losing the favourable QTL allele, several markers should be used that cover the whole region in which the QTL is likely located. Animals with no observable recombination in the whole region should be selected. (Enlarging the region in which no recombination is allowed leads to more stringent selection. It might be that the size of the experiments has to be adapted upwards). If the QTL location is inaccurate, the strategy to minimize the risk of losing the favourable QTL allele results in introgressing a large part of the donor genome into the recipient line. Introgressing unfavourable donor polygenes then lowers the benefits of introgressing the favourable QTL allele. An alternative strategy would be to maximize the value of the introgressed part of the donor genome, taking into account that the donor genome carries both the favourable QTL allele and unfavourable alleles of polygenes. The value of a location can be computed as the probability that the QTL is at that location times the value of the favourable QTL allele plus the polygenic value of that location in the donor genome minus the polygenic value of that location in the recipient genome. An estimate of the polygenic value of a location would be the population mean corrected for the QTL effect times the fraction of the genome at that location.

The use of markers to select against the donor population background genome can reduce the number of backcross generations needed for full recovery of the recipient genome by two generations (Hospital *et al.* 1992; Visscher *et al.* 1995). For recovery of the recipient genome, animals with fewest marker alleles originating from the donor population are selected.

Groen and Smith (1995) showed that full recovery of the recipient genome is not efficient when both the donor and the recipient line contain favourable alleles. For such a situation, selection for animals with the desired phenotype was superior to selection for animals with the highest proportion of recipient marker alleles (Groen and Smith 1995).

In an introgression programme, selection is within a cross. Marker assisted selection can exploit the linkage disequilibrium between markers and QTL alleles in a cross, resulting in genetic progress higher than by phenotypic selection (Lande and Thompson 1990; Zhang and Smith 1992,1993; Gimelfarb and Lande 1994a,b). So, selection should not be for the marker alleles coming from one population but for the marker alleles associated with highest phenotypic trait values. However, backcrossing is usually applied if the recipient population level is much higher than the donor population level. Then, for most markers the allele coming from the recipient population will be favourable.

Even if marker alleles from the donor population are associated with an increase in the traits that are measured, the breeder might prefer to select for the recipient genome to prevent introgression of donor alleles that have a negative influence on traits that are not observed in the introgression programme.

Marker assisted selection in a crossbred population

In a crossbred population, marker assisted selection can increase genetic gain up to 400% (Lande and Thompson 1990). A 400% increases was obtained under unrealistic assumptions, but illustrates the potential. Many factors influence the efficiency of marker assisted selection in a cross. I shall not address all these factors but try to outline optimal conditions for marker assisted selection in a crossbred population.

Marker assisted selection is beneficial if markers explain otherwise unexplained genetic differences among the candidates for selection. This is true both for selection within a crossbred population and for selection in an outbred population. To explain differences, a marker should be close to a QTL that accounts for a large genetic difference. The marker allelic effects have to be estimated. The more accurate these estimates of the marker allelic effects are compared with the accuracy of polygenic estimates, the more otherwise unexplained variance will be accounted for.

In an intercross, marker allelic effects can be estimated on a population basis because markers and QTLs are in linkage disequilibrium on a population basis. Consequently, many observations can be used to estimate the marker allelic effects. The number of observations used to estimate the allelic effects had the largest effect on the extra genetic gain (Gimelfarb and Lande 1994a; Zhang and Smith 1993). Increasing population size from 1000 to 3000 animals still greatly affected the accuracy of marker allelic effect estimates (Gimelfarb and Lande 1994a).

The work of Gimelfarb and Lande (1994a), and Zhang and Smith (1993) showed two aspects of the breeding value estimation procedure are very important: the model used in breeding value estimation, and whether or not maker allelic effects are re-estimated. Both studies simulated an F2 based on two inbred lines. Marker assisted selection in the F2 and later generations exploited the linkage disequilibrium between markers and QTLs in the F2. Gimelfarb and Lande (1994a) estimated marker allelic effects by standard multiple linear regression treating allelic effects as fixed. Zhang and Smith (1992,1993) used a mixed model with random marker allelic effects that was suggested by Goddard (1992). The random regression approach gave higher response than fixed regression (Zhang and Smith 1993). In an example, response using random regression was three times the response using fixed regression. Random regression is superior because it considers the differences in accuracy of marker haplotype effects (Goddard *et al.* 1995).

Zhang and Smith (1992,1993) estimated marker allelic effects once, in the F2 generation, and used these estimates to select in the F2 and all subsequent generations. Gimelfarb and Lande (1994a) showed response was higher if marker allelic effects were re-estimated each generation. Gimelfarb and Lande (1994a) based re-estimates on information from the latest generation and discarded information from earlier generations. So, both Zhang and Smith, and Gimelfarb and Lande show the amount of information to estimate marker allelic effects is crucial, but both do not use all information to estimate the effects. I suggest information from all generations should be used to obtain mixed model estimates of the marker allelic effects. The mixed model should allow for changing marker allelic effects over generations and differences over families.

For a cross between inbred populations using the model of Fernando and Grossman (1989), treating all F2-animals as offspring resulting from selfing one non-inbred animal, would be appropriate. This model links the gametic effect of all F2-animals to the same two parental gametes in the F1. Goddard (1992) derived an elegant simplified model appropriate for analysing F2 data, but not for data including several generations.

For a cross between outbred populations the approach suggested for a cross between inbred populations will not work. Within the F1 between outbred populations are many different gametes that can be divided into two groups of gametes, one for each outbred population. The model should allow for individual gametes and for grouping of gametes. This can be achieved for the model of Fernando and Grossman (1989) by including a random population effect for each QTL allele into the model similar to including random genetic group effects in a standard BLUP analysis (Goddard 1992). The effect of an individual QTL allele is modelled as the average of the QTL alleles in the group plus a deviation of the individual QTL allele from the group average.

Marker assisted selection in an outbred population

An outbred population is a collection of families whereas a crossbred population resembles one full-sib family. Linkage disequilibrium between markers and QTLs differs over families, i.e., the outbred population is expected to be in linkage equilibrium. So, marker allelic effects have to be estimated for each family separately.

For two reasons, estimated marker allelic effects in an outbred population explain less variance and therefore contribute less to genetic gain than estimated marker allelic effects in a crossbred population. First, a single family in an outbred population is generally smaller than an entire crossbred population. Therefore, the marker allelic effect estimates are regressed more towards zero in an outbred population. More speculatively, in an outbred population under selection, segregating large QTLs may be less likely than in a crossbred population.

The use of a cross is beneficial if two populations contain complementary QTL alleles. If two such populations are not available, genetic markers can only be applied for marker assisted selection within an outbred population. In chapter 5 we quantified the additional response due to marker assisted selection for a sex-limited trait in an outbred poultry breeding nucleus. Although population wide linkage disequilibrium could not be exploited, the use of genetic markers still increased the response in an outbred poultry population by 10% if a marker for a QTL explaining 20% of the genetic variance was available and by 20% if a QTL explaining 80% of the genetic variance was available (Van der Beek and Van Arendonk 1995). Van der Beek and Van Arendonk (1995) simulated a breeding programme in which the number of full-sib cocks selected was restricted to limit inbreeding. Additional response was mainly due to the marker based selection of cocks within a full-sib family; without markers these cocks all had the same estimated breeding value and selection was not possible. Additional response increased if fewer cocks were selected from each full-sib family.

Meuwissen and Van Arendonk (1992) studied marker assisted selection for a sex-limited trait in a dairy cattle breeding nucleus. They also found additional responses of 10 to 20% but in their study the number of male full-sibs selected was not restricted. In their study, however, information from outside the nucleus was used to estimate marker allelic effects, and the amount of information from outside the nucleus greatly affected additional response. If, similarly, for a poultry breeding programme, information from outside the breeding nucleus can be used to estimate marker allelic effects, then additional response in poultry breeding can be greater than the 10 to 20% we found.

V Three specific applications of genetic markers in a poultry breeding program

The density of the chicken linkage map allows powerful QTL mapping experiments. A QTL mapping experiment should be large and should concern the population used in the breeding programme to result in useful application of the acquired information. As long as QTL mapping does not lead to identified specific genes, information on QTLs has to be exploited via genetic markers. The benefits of using genetic markers depend on the traits considered and the breeding programme in which the genetic markers are used. Application of genetic markers is most promising for traits that are difficult or expensive to measure, expressed late in life, and traits for which mapped segregating QTLs explain much of the genetic variance. Application of genetic markers is most efficient in an intercross or in a backcross during introgression.

The conditions under which the use of markers is most beneficial can be inferred from the description in the previous section on the potential uses of markers. Current breeding practise does not necessarily meet those conditions. New breeding programmes or breeding strategies may need to be developed to fully exploit the potential of genetic markers. In this section I will try to describe breeding strategies that optimally exploit markers. I will describe three scenarios. In the first scenario, the emphasis is on difficult traits and on the use of introgression and selection in a cross. The second scenario describes a strategy to reduce the typing costs without affecting the benefits of marker assisted selection. In the third scenario markers are used primarily to restructure the breeding programme rather than for marker assisted selection directly.

Replacing a line. The final-product of a poultry breeding programme is a crossbred animal. All parental lines used in the cross are improved simultaneously. From time to time a parental line is replaced by a new line that either adds new characteristics to the final product, e.g. resistance against a specific disease, or simply is better for a trait such as growth rate.

The new characteristics of the line can involve traits that were previously not under selection. For unselected traits, QTLs explaining much variance are likely present. A logical approach is to first map QTLs for the unselected traits, and subsequently select for those QTLs via marker assisted selection.

The new characteristic of the line can be a specific gene. An example is a sex-linked slow feathering gene, or a dwarf gene. During the development of the new line often introgression or crossing is applied. Introgression will be used if a specific favourable allele is identified in a further inferior population. Markers will be used to introgress this favourable allele into the line that will be used to form the new line.

Crossing can be applied to combine the good qualities of two (or more) lines. Either two lines have a high level for different traits, e.g., one line excels in growth rate and a second line has good meat quality, or two lines are comparable for a certain trait, but different alleles account for the level of the two lines. Crossing two lines of comparable level but with different alleles will result in new variance available for selection, i.e., the crossbred line has more favourable alleles than either original line. Within the crossbred population linkage disequilibrium between markers and QTLs will exist which can be exploited via marker assisted selection.

Some observations indicate the existence of lines that exhibit complementary QTLs. For DNA fingerprints, Dunnington *et al.* (1994) observed much variation among lines of various commercial breeders. These commercial lines are expected to be of comparable level for commercial traits. In the section on design of QTL mapping experiments I pointed out that complementary QTLs are possible for unrelated lines of comparable level. A commercial breeder will therefore try to acquire genetic material unrelated to his own stocks. But also, the lines available within the breeding programme should be analyzed carefully. Some of these lines show good combining ability. Good combining ability is currently a reason for keeping lines separate to preserve the heterotic effect. However, the heterotic effect might *not* due to overdominance, but due to partial dominance. Kearsey and Pooni (1992) concluded in plants partial dominance usually accounts for heterosis. With partial dominance, crossing the lines and selecting within the cross will eventually result in a line with a higher level. To illustrate this, consider a trait determined by two dominant loci. Locus A has genotypes A1A1, A1A2 and A2A2 with levels +10, +7, and -10, Locus B has genotypes B1B1, B1B2 and B2B2 with levels +10, +8, and -10. Let line X be homozygous for A1 and B2, and line Y be homozygous for A2 and B1. The level of X is 10-10=0, the level of Y is -10+10=0. The F1 of X and Y is heterozygous A1A2 and heterozygous B1B2. The level the F1 is +8+7=15, i.e., higher than either parental line. Selection in the F1 would eventually lead to a line homozygous A1A1 and homozygous B1B1, i.e., with a level of +10+10=20, a higher level than a cross between X and Y would ever attain.

Breeding value estimation using selective DNA pooling. Marker assisted selection is only possible if candidates for selection are typed for the markers. The resulting marker typing costs will be high. For example, in the breeding programme described in chapter 5, each generation 9000 selection candidates had to be typed for markers. Although reducing the number of typings may be easy, this reduction can lead to reduced genetic response.

In some strategies using selective DNA pooling, however, fewer animals or fewer markers per animal might have to be typed without reducing the selection response. Selective DNA pooling can be applied if phenotypic trait values are measured for many individuals per

crossbred population or for many sibs per family in an outbred population. Marker allelic effects are computed using selective DNA pooling and used to select between selection candidates. Selection candidates have to be individually typed. Therefore, selective DNA pooling can only reduce the number of marker typings if either there are considerably fewer selection candidates than animals used for selective DNA pooling, or selection candidates are typed for only a few markers whereas selective DNA pooling is used to compute the effects for many more markers. In the latter case selective DNA pooling can be used to estimate allelic effects for many markers and selection candidates are only typed for those markers that explain most variance.

The (number of) selection candidates can differ from the animals used for selective DNA pooling:

- During multi-stage selection a phenotypic trait is measured in an early stage with many animals. In a later stage with few selection candidates this phenotypic trait is selected for using markers.
- Selection is for a sex-limited trait. For example, hens are used for selective DNA pooling to estimate marker allelic effects on egg production. Estimated marker allelic effects are used to select among the male sibs. For those males the use of markers has a much higher effect on selection response than for hens with own performance recorded (Chapter 5).
- A specific test population is formed in which selective DNA pooling is applied. The test population can consist of sibs of the selection candidates or sibs of the parents of the selection candidates.

The use of a test population has several advantages. The test population can be kept under specific conditions, e.g., a broiler *breeding* population can be kept under restricted feeding and the broiler *test* population under ad-lib feeding in a commercial or stressful environment. Another example is to challenge the test population with a parasite. A further example is to measure a trait like breast meat in the test population, such that no breeding animals have to be sacrificed. In this system crossbred test animals also could be used, so that crossbred performance is estimated.

The use of a test population is not restricted to a scheme with marker assisted selection. Without markers, however, selection based on information on sibs or grandparents is inaccurate. With markers, the relation between grandparents and grandoffspring or between sibs can be determined much more accurately resulting in more accurate breeding value estimation.

Combining the use of markers for selection and parentage control. The current poultry breeding programme has a structure that is optimized given current constraints. A major constraint is that hens are mated to one cock only (i.e., a hierarchical mating design is used) because if the hen would be mated to more cocks the sire of resulting offspring would be unknown. A hierarchical mating design results in less selection response than a (partly) factorial mating design (Woolliams 1989; Ruane 1991; De Boer and Van Arendonk 1994). Consequences of a hierarchical design are: (1) much emphasis is, especially for sex-limited traits, on full-sib family information, as a result of which related animals will be selected. Either inbreeding will be high or measures are taken to limit inbreeding, like restricting the number of male offspring selected per dam in a layer programme, which reduces selection intensity; (2) estimated breeding values for selection candidates are correlated, especially for sex-limited traits which reduces selection intensity; (3) the male generation interval is not minimized. Theoretically, a male could produce all his offspring in one week. But as a hen is mated to one cock, a cock is used during the whole breeding period of the hen.

The constraint of one cock per hen can be lifted if markers are used for parentage testing of pullets. Then, each week the breeding hens can be inseminated by sperm of a different cock, or by a mixture of sperm of several cocks. In case of mixed sperm, the relative fertility of the sires will determine the proportion of offspring of one sire (Martin and Dziuk 1977). So, if large differences in male fertility are expected this should be taken into account before deciding for the use of mixed sperm. A cock would be used one week only, during which he is mated to many hens. Hens are used for several weeks during which they are mated to many cocks. The result would be that (1) either inbreeding decreases or restrictions are no longer necessary resulting in higher selection intensity, (2) correlations between estimated breeding values are reduced, (3) if each week new young cocks are available then generation interval shortens, and (4) a reduction in pedigree errors.

The markers used for parentage control can also be used for marker assisted selection, leading to a further increase in genetic response.

For lines with little or no emphasis on reproduction traits, one could even consider keeping hens in groups. This would result in a significant cost reduction, which might balance the high costs of the scheme suggested.

VI Summary

At the end of 1995 the chicken linkage map will probably be sufficient to be used in whole genome QTL mapping experiments. Currently, there are few QTLs mapped. From QTL mapping experiments in plants and other livestock species it can be learned that usually a few QTLs explain a large part (>50%) of the genetic variance. The size of QTL allelic effects will depend on the magnitude of the genetic variance in the population used in the QTL mapping experiment.

Applications of genetic markers that were described include: parentage control, varietal identification, marker assisted introgression, marker assisted selection in a crossbred population, marker assisted selection in an outbred population. Three specific applications were described: marker assisted selection during the generation of a new line, marker assisted breeding value estimation when using selective DNA pooling, and the simultaneous use of markers for parentage control to allow a factorial mating design and for marker assisted selection. In these specific applications genetic markers are used under optimal conditions, the costs of use of genetic markers are reduced, or potential benefits of genetic markers are combined. These three applications further show that under optimal conditions, the use of genetic markers can improve poultry breeding schemes.

Acknowledgements

I would like to thank Drs Albers, Brascamp, Groen, and Van Arendonk for many helpful discussions, Drs Groenen and Crooijmans for many helpful comments, and Drs Haley, Visscher and especially Dr Soller for critical reading of an earlier version of the manuscript.

References

Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, Andersson K, Andersson Eklund L, Edfors Lilja I, Fredholm M, Hansson I, Hakansson J, Lundstrom K (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science 263:1771-1774.

Beavis WD, Smith OS, Grant D, Fincher R (1994) Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. Crop Sci 34:882-896.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331.

Bumstead N, Palyga J (1992) A preliminary linkage map of the chicken genome. Genomics 13:690-697.

Burt A, Bell G (1987) Mammalian chiasma frequencies as a test of two theories of recombination. Nature 326:803-805.

Burt DW, Bumstead N, Bitgood JJ, Ponce le Leon FA, Crittenden LB (1995) Chicken genome mapping: a new era in avian genetics. Trends Genet 11:190-194.

Cheng HH, Crittenden LB (1994) Microsatellite Markers for Genetic Mapping in the Chicken. Poultry Sci 73:539-546.

Cockerham CC, Zeng ZB (1995) Design III with marker loci. Genetics:submitted.

Crittenden LB, Provencher L, Santangelo L, Levin I, Abplanalp H, Briles RW, Briles WE, Dodgson JB (1993) Characterization of a red jungle fowl by White Leghorn backcross reference population for molecular mapping of the chicken genome. Poultry Sci 72:334-348.

Crooijmans RPMA, Kampen AJA, Van der Poel JJ, Groenen MAM, Van Kampen AJA (1994a) Highly polymorphic microsatellite markers in poultry. Anim Genet 24:441-443.

Crooijmans RPMA, Van Kampen AJA, Van der Poel JJ, Groenen MAM (1994b) New microsatellite markers

on the linkage map of the chicken genome. J Hered 85:410-413.
Crooijmans RPMA, Van Oers PAM, Brandts X, Dijkhof RJM, Strijk J, Van der Poel JJ, Groenen MAM (1995a) Preliminary linkage map of the chicken genome based on microsatellite markers: 75 new markers mapped. Poultry Sci:submitted.
Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor Appl Genet 85:353-359.
Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365-1373.
De Boer IJM, Van Arendonk JAM (1994) Additive response to selection adjusted for effects of inbreeding in a closed dairy cattle nucleus assuming a large number of gametes per female. Anim Prod 58:173-180.
De Vicente MC, Tanksley SD (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585-596.
Dunnington EA, Haberfeld A, Stallard LC, Siegel PB, Hillel J (1992) Deoxyribonucleic acid fingerprint bands linked to loci coding for quantitative traits in chickens. Poultry Sci 71:1251-1258.
Dunnington EA, Stallard LC, Hillel J, Siegel PB (1994) Genetic diversity among commercial chicken populations estimated from DNA fingerprints. Poultry Sci 73:1218-1225.
Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113-125.
Edwards MD, Helentjaris T, Wright S, Stuber CW (1992) Molecular-marker-facilitated investigations of quantitative trait loci in maize. 4. Analysis based on genome saturation with isozyme and restriction fragment length polymorphism markers. Theor Appl Genet 83:765-774.
Eshed Y, Zamir D (1994) Introgressions from Lycopersicon pennellii can improve the soluble-solids yield of tomato hybrids. Theor Appl Genet 88:891-897.
Eshed Y, Zamir D (1995) An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics:submitted.
Fernando RL, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Genet Sel Evol 21:467-477.
Gavora JS, Kuhnlein U, Crittenden LB, Spencer JL, Sabour MP (1991) Endogenous viral genes: association with reduced egg production rate and egg size in White Leghorns. Poultry Sci 70:618-623.
Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao XY, Womack JE, Hoeschele I (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920.
Gimelfarb A, Lande R (1994a) Simulation of marker assisted selection in hybrid populations. Genet Res 63:39-47.
Gimelfarb A, Lande R (1994b) Simulation of marker assisted selection for non-additive traits. Genet Res 64:127-136.
Goddard ME (1992) A mixed model for anlayses of data on multiple genetic markers. Theor Appl Genet 83:878-886.
Goddard ME, Zhang W, Smith C (1995) Factors affecting the accuracy of estimated breeding values calculated from genetic markers. Theor Appl Genet (Submitted).
Goldman IL, Rocheford TR, Dudley JW (1994) Molecular markers associated with maize kernel oil concentration in an Illinois High Protein X Illinois Low Protein cross. Crop Sci 34:908-915.
Groen AF, Crooijmans RPMA, Van Kampen AJA, Van der Beek S, Van der Poel JJ, Groenen MAM (1994) Microsatellite polymorphism in commercial broiler and layer lines. Proc 5th World Congr Genet Appl Livestock Prod 21:95-98.
Groen AF, Crooijmans RPMA (1995) Relationship between heterosis and genetic distance as measured by microsatellites in commercial layer lines. In preperation.
Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 132:1199-1210.
Kearsey MJ, Pooni HS (1992) The potential of inbred lines in the presence of heterosis. In: Darter Y, Dumas C, Gallais A (eds) Reproductive biology and plant breeding. Springer, Berlin .
Keightley PD (1994) The distribution of mutation effects on viability in Drosophila melanogaster. Genetics 138:1315-1322.
Khatib H, Genislav E, Crittenden LB, Bumstead N, Soller M (1993) Sequence-tagged microsatellite sites as markers in chicken reference and resource populations. Anim Genet 24:355-362.
Khatib H (1994) Mapping Quantitative trait loci in chicken populations. Phd thesis (Hebrew, with English summary) .

Lakshmanan N, Lamont SJ, Kaiser MG, Arthur JA, O'Sullivan NP (1994) Genetic markers linked with quantitative traits in chickens. Proc 5th World Congr Genet Appl Livestock Prod 21:205-208.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

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

Lebowitz RJ, Soller M, Beckmann JS (1987) Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor Appl Genet 73:556-562.

Levin I, Santangelo L, Cheng H, Crittenden LB, Dodgson JB (1994) An Autosomal Genetic Linkage Map of the Chicken. J Hered 85:79-85.

Martin PA, Dziuk PJ (1977) Assessment of relative fertility of males (cockerels and boars) by competitive mating. J Reprod Fertility 49:323-329.

Merat P (1990) Pleiotropic and associated effects of major genes. In: Crawford RD (ed) Poultry Breeding and Genetics. Elsevier, Amsterdam.

Meuwissen THE, Van Arendonk JAM (1992) Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. J Dairy Sci 75:1651-1659.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181-197.

Plotsky Y, Cahaner A, Haberfeld A, Lavi U, Lamont SJ, Hillel J (1993) DNA fingerprint bands applied to linkage analysis with quantitative trait loci in chickens. Anim Genet 24:105-110.

Rahn MI, Solari AJ (1986) Recombination nodules in the oocytes of the chicken, Gallus domesticus. Cytogenet Cell Genet 43:187-193.

Rodionov AV, Myakoshina YA, Chelysheva LA, Solovei IV, Gaginskaya ER (1992) Chiasmata on lampbrush chromosomes of Gallus Gallus Domesticus: A cytogenetic study of recombination frequency and linkage group lengths. Sov Genet 28:421-430.

Ruane J (1991) The effect of alternative mating designs and selection strategies on adult multiple ovulation and embryo transfer (MOET) nucleus breeding schemes in dairy cattle. Genet Sel Evol 23:47-65.

Schön CC (1993) RFLP mapping in maize (Zea mays L.): quantitative trait loci affecting testcross performance of elite European flint lines. Doctoral thesis, University of Hohenheim, Hohenheim.

Shoffner RN, Otis JS, Garwood VA (1993) Association of dominant marker traits and metric traits in chickens. Poultry Sci 72:1405-1410.

Shrimpton AE, Robertson A (1988) The isolation of polygenic factors controlling bristle score in Drosophila melanogaster. II. Distribution of third chromosome bristle effects within chromosome sections. Genetics 118:445-459.

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25-33.

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823-839.

Van der Beek S, Van Arendonk JAM, Groen AF (1995) Power of two- and three- generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. Theor Appl Genet:in press.

Van der Beek S, Van Arendonk JAM (1995) Marker assisted selection in an outbred poultry breeding nucleus. Anim Sci:in press.

Visscher PM, Haley CS, Thompson R (1995) Marker assisted introgression in backcross breeding programs. Genetics:submitted.

Visscher PM, Haley CS (1995) Utilizing genetic markers in pig breeding programmes. Anim Breed Abs 63:1-8

Woolliams JH (1989) Modifications to MOET nucleus breeding schemes to improve rates of genetic progress and decrease rates of inbreeding in dairy cattle. Anim Prod 49:1-14.

Zhang W, Smith C (1992) Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theor Appl Genet 83:813-820.

Zhang W, Smith C (1993) Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. Theor Appl Genet 86:492-496.

Summary

This thesis focuses on the detection and use of genetic markers linked to quantitative trait loci (QTL) in poultry breeding.

A genetic marker is a genomic locus for which the allele(s) can be identified. Genetic markers have become abundantly available due to the rapid development of molecular genetic technology and the identification of new classes of polymorphic genomic sites. A genetic marker can be used for parentage control and varietal identification. Genetic markers can also be used to identify QTLs affecting traits, and to select for favourable QTL alleles. The latter applications of genetic markers rely on the cosegregation between a genetic marker and the chromosomal segment surrounding the genetic marker.

Several steps have to be taken to exploit genetic markers to increase genetic improvement. Studies on three steps are reported in this thesis: the design of linkage mapping experiments, the design of QTL mapping experiments, and the use of markers linked to QTLs in selection.

To efficiently use genetic markers, their location at the genome has to be identified, i.e., a linkage map has to be constructed. Genotyping animals is laborious and consequently asks for optimization of the design to construct a linkage map. The information available to optimize linkage mapping experiments in poultry was inadequate; therefore chapters 2 and 3 deal with designing linkage mapping experiments that use an outbred population.

Chapter 2 reports on a study of designs to detect and estimate linkage between markers for outbred populations. Two parameters to measure design quality were used: expected maximum lod score and accuracy. The expected maximum lod score indicates the power of a design to detect linkage between markers. The accuracy was measured as the probability that the true recombination rate between markers falls in a specified interval. For expected maximum lod score the desired value was set to 3. For accuracy, the desired probability that a true recombination was between 0.15 and 0.25 for an estimated value of 0.20 was set to 0.85. Exact methods were described and used to evaluate approximate methods. These approximate methods were developed for rapid evaluation of designs. The approximation of expected maximum lod score was precise under all situations studied. The approximation of accuracy was only precise for experiments large enough to have an accuracy equal to or higher than the desired accuracy. For smaller experiments, the method for approximating accuracy was not valid. For those smaller experiments, true recombination rates were on average much higher than estimated recombination rates.

Chapter 3 provides a general framework for evaluating expected maximum lod score and accuracy of designs for reference families consisting of full-sib or half-sib families selected from an outbred population. Evaluation was both for two markers with a recombination rate of 0.20 and for a marker and a dominant single gene with a recombination rate of 0.20. The influence of number of families, number of offspring per family, family

structure (either half-sib or full-sib) and marker polymorphism was quantified. For mapping markers with a heterozygosity of 0.75 in a reference population containing full-sib families, at least 40 animals were required for an expected maximum lod score of 3, and at least 100 animals for an accuracy of 0.8, i.e., accuracy was a more stringent criterium than expected maximum lod score. The number of animals required for reference populations containing half-sib families was at least twice the number of animals required for reference families containing full-sib families. The number of animals required to map a dominant single gene was 3 to more than 10 times the number of animals required for mapping markers.

Chapter 3 showed how results can be generalized to determine the values for a wide range of designs containing two- or three- generations with full-sib or half-sib families.

After markers are located at the linkage map, a selected set of markers covering the whole genome can be used in QTL mapping experiments. QTL mapping experiments need to be optimized to minimize costs of data collection and genotyping. Chapter 4 analyzed the power and efficiency of experiments with two- or three- generation family structures containing full-sib families, half-sib families or a combination of both. Focus was on data from one outbred population. For a specified power, a two-generation experiment required more animals to be typed for marker loci than a three-generation experiment. Fewer trait values, however, had to be obtained for a two-generation experiment than for a three-generation experiment. A two- or three- generation family structure with full-sib offspring was more efficient than a two- or three- generation family structure with half-sib offspring. A three-generation family structure with full-sib grandoffspring, however, was less efficient than a family structure with half-sib grandoffspring. The most efficient family structure had full-sib offspring and half-sib grandoffspring. In that case marker genotypes were obtained for the animals in the first and second generation while trait values were obtained for the third generation animals. For a heritability of 0.1 and 100 grandoffspring per full-sib offspring, the most efficient family structure required 30 times less marker typings for a given power than a two-generation half-sib structure in which trait values were obtained for half-sib offspring. The effect of heritability on power differed between two- and three-generation designs. When the effect of the QTL was expressed in phenotypic standard deviation units, power increased with increasing heritability for a two-generation design and power decreased with increasing heritability for a three-generation design. The results described in chapter 4 can be used to design efficient QTL mapping experiments in an outbred population.

In chapter 5, the effect on selection response of using a marker for a QTL affecting a sex limited trait in an outbred poultry breeding nucleus was studied. A closed nucleus with 9000 chickens per generation was deterministically simulated. The genetic model contained

polygenes and a QTL linked to a marker. Marker and QTL were in linkage equilibrium in the base generation before selection. Genetic effects explained 30% of the phenotypic variance in the base generation before selection. Cocks were selected in two steps. First the best cocks of each full-sib family were selected (within family selection) while final selection took place after information on full-sibs was available. Hens were selected after they had completed production. The effect of using marker information in estimating breeding values was studied in an ongoing breeding programme. Cumulative response over five generations increased 6 to 13 % if a marker linked to a QTL that explained 20% of the genetic variance was used. Cumulative response increased up to 28% if the QTL explained 80% of the genetic variance. Additional response due to the use of a marker increased with increasing intensity of within family selection of cocks, increasing number of dams per sire, increasing variance explained by the QTL, and was higher if within family selection of cocks was after their sibs had complete records than if within family selection was before their sibs had complete records.

In the general discussion literature on linkage and QTL mapping in chickens is reviewed first. More than 460 genetic markers are currently at the chicken linkage map. Only few QTLs, however, are currently mapped on the chicken linkage map. Much effort in QTL mapping is required in the coming years. The population used for QTL mapping experiments may be a cross between inbred populations, a cross between outbred populations that differ much in population level for traits of interest, a cross between populations that are genetically distant but differ little for the traits of interest, or an outbred population. Advantages and disadvantages of populations to be used for QTL mapping experiments were described. Several general applications of markers in poultry breeding were discussed including: parentage control, varietal identification, marker assisted selection during introgression, marker assisted selection in a crossbred population, and marker assisted selection in an outbred population. Three specific potentially useful applications were described utilizing the knowledge generated in chapter 5 and the characteristics of poultry breeding and genetic markers. In the first application, genetic markers were used during the development of a new line to be used in a crossbreeding program. This application is promising for two reasons. First, markers are used in a situation where there is linkage disequilibrium between markers and QTLs. This linkage disequilibrium can be exploited during either introgression or advanced intercrossing. Second, during the development of a new line, markers for QTLs explaining a major part of the genetic variance in a certain trait are more likely to be found than for selection in an ongoing breeding programme.

The second application is the use of selective DNA pooling for marker assisted breeding value estimation. With selective DNA pooling, marker typing costs can be reduced significantly without affecting additional response to marker assisted selection.

In the third application markers are used for parentage control and marker assisted selection. In this application the restriction that a hen is mated to one cock only is lifted. With multiple sires mated to one hen, the sire of the offspring of the hen is initially unknown. Markers can, however, be used for parentage control. This facilitates a factorial mating design which results in higher genetic gains than a hierarchical mating design. In addition, the same markers used for parentage control can be used for marker assisted selection resulting in an increase in genetic gain that comes on top of the increase due to a change in mating design.

These three applications showed that under optimal conditions, the use of genetic markers can improve poultry breeding schemes.

Waar gaat dit proefschrift over ?

In dit hoofdstuk zal een algemene schets gegeven worden van de achtergrond, het doel, en de uitkomsten van het onderzoek dat beschreven is in dit proefschrift. Dit hoofdstuk is geen samenvatting van de resultaten; deze kan gevonden worden in 'Summary'.

Fokkerij beoogt het genetisch niveau van een populatie te verbeteren door de genetisch beste dieren uit een populatie te selecteren als ouders van een volgende generatie. De dieren met de beste genen worden indirect geïdentificeerd. Er wordt gebruik gemaakt van het feit dat die dieren die 'uiterlijk' het beste scoren voor een bepaald kenmerk (bijvoorbeeld door de meeste eieren te leggen) waarschijnlijk ook de beste genen hebben, met name als verwante dieren ook hoog scoren. In de loop van de jaren zijn efficiënte methoden ontwikkeld om op basis van een waarneming aan een dier en waarnemingen aan verwante dieren een zogenoemde fokwaarde te schatten. Op basis van deze fokwaarde worden dieren geselecteerd.

Fokkerij kan succesvol zijn zonder dat de genen waar het allemaal om draait direct waargenomen kunnen worden. Dit wil niet zeggen dat de fokkerij niet gebaat zou zijn bij directere informatie over de genen die verantwoordelijk zijn voor de verschillen die tussen dieren waargenomen worden voor economisch belangrijke kenmerken. De fokkerijwereld is dan ook ingesprongen op de snelle ontwikkeling van de moleculaire genetica. Deze ontwikkeling maakt het in theorie mogelijk om de genen te identificeren die verantwoordelijk zijn voor verschillen tussen dieren voor belangrijke kenmerken. Een gereedschap van de moleculaire genetica dat geschikt lijkt om meer informatie boven tafel te krijgen over belangrijke genen is de genetische merker.

Een genetische merker representeert een stukje erfelijk materiaal, ofwel een stukje DNA. DNA is gerangschikt op de zogenaamde chromosomen, waarvan de kip er 39 verschillende heeft. Alle chromosomen samen noemen we het genoom. Een voorbeeld van een genetische merker is veerkleur bij pluimvee. Door te kijken naar de kleur weten we iets over die specifieke plaats op het genoom waar het gen ligt dat verantwoordelijk is voor veerkleur. De waargenomen veerkleur geeft geen perfecte informatie over het veerkleurgen; veerkleur vererft dominant, waarbij wit dominant is over bruin. Een bruin dier heeft op beide chromosomen de bruine variant (de genetische term voor variant is allel) van het veerkleur gen. Een wit dier heeft in ieder geval één kopie van het witte allel, het andere allel kan wit of bruin zijn. Het veerkleurgen is derhalve een dominante merker. Het is dimorf: er komen twee allelen van voor.

'Moderne' klassen van genetische merkers zijn gestoeld op moleculair genetische technologie. De belangrijkste klasse van genetische merkers is de microsatelliet merker. Microsatelliet merkers hebben de volgende aantrekkelijke eigenschappen:

- de microsatelliet is vaak polymorf. In een populatie komen vele verschillende allelen voor. De merkers zijn daarom geschikt om verschillen tussen dieren aan te tonen en de overerving van allelen te volgen.

- de microsatelliet is codominant. Als een dier twee verschillende allelen bezit dan zijn altijd beide allelen zichtbaar. Dit in tegenstelling tot het veerkleurgen waarbij een witte kip, zonder dat we het zien, een bruin allel kan herbergen.
- microsatellieten komen overal in het genoom voor. Naar schatting zijn vele duizenden verschillende microsatellieten in het genoom aanwezig.

Waarom zijn genetische merkers nu interessant? De belangrijkste reden is dat genen die in de buurt van elkaar voorkomen op het genoom, gekoppeld overerven (zie intermezzo). Om te illustreren hoe een genetische merker gebruikt kan worden om genen op te sporen, wordt weer het veerkleurgen gebruikt. We nemen een witte haan waarvan bekent is dat hij een bruin allel heeft aangezien hij naast witte nakomelingen ook bruine nakomelingen heeft. We paren deze haan met louter bruine hennen. De helft van de nakomelingen zal wit zijn, en de andere helft bruin. De hennetjes laten we onder zo gelijk mogelijke omstandigheden eieren leggen. Vervolgens kunnen we onderzoeken of de witte dochters gemiddeld meer of minder eieren leggen dan de bruine dochters. Als de witte dochters significant meer of significant minder eieren leggen dan de bruine dochters, dan is een associatie tussen het veerkleurgen, de merker, en het kenmerk eiproduktie gevonden. Het veerkleurgen kan zelf de associatie veroorzaken. Anderzijds kunnen in de buurt van het veerkleurgen andere genen liggen die een rol spelen bij eiproduktie. Deze genen worden dan door hun ligging dichtbij het veerkleurgen gekoppeld overgeerfd. Wat nu het meest waarschijnlijk is kunnen we onderzoeken door de proef te herhalen voor andere witte hanen met een bruin allel. Als voor alle hanen witte dochters meer eieren leggen dan bruine dochters, dan is waarschijnlijk het kleurgen zelf of een zeer nauw gekoppeld gen verantwoordelijk voor het waargenomen verschil in eiproduktie. Echter, het kan ook zijn dat voor de ene haan wit samengaat met meer eieren en voor de andere haan bruin. In dat geval mogen we aannemen dat niet het veerkleurgen zelf maar een in de buurt gelegen gen invloed heeft op eiproduktie, waarbij de afstand tussen het veerkleurgen en het gen dat invloed heeft op eiproduktie dusdanig is dat overkruising plaatsvindt.

De literatuur beschrijft verschillende toepassingen van genetische merkers. Genetische merkers kunnen gebruikt worden voor ouderschapscontrole en voor het globaal bepalen van genetische verschillen tussen populaties. Genetische merkers kunnen gebruikt worden om genen op te sporen die genetische verschillen tussen dieren of populaties verklaren. Genetische merkers waarvan aangetoond is dat ze genetische verschillen verklaren, kunnen gebruikt worden tijdens selectieprocedures.

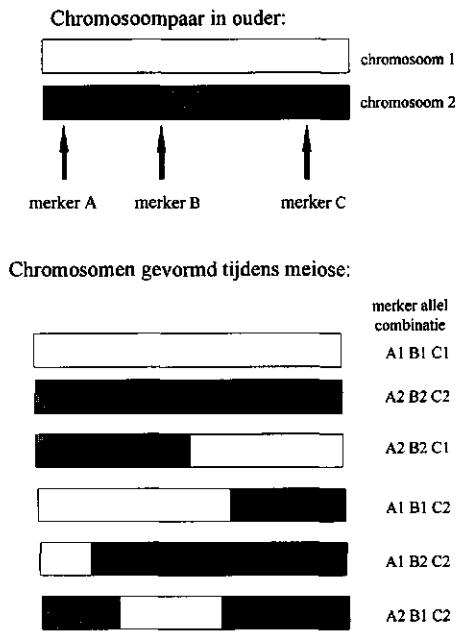
Voor alle toepassingen is het van belang dat eerst de positie van een genetische merker in het genoom wordt bepaald. De geschatte posities van genetische merkers in het genoom worden vastgelegd in een koppelingskaart. Deze koppelingskaart kan vervolgens gebruikt worden in proeven die tot doel hebben om genen op te sporen. Als we weten waar genetische

merkers in het genoom liggen, kunnen we vervolgens associaties tussen merkers en economisch belangrijke kenmerken als produktie en vitaliteit vaststellen. Hiermee kan geschat worden waar in het genoom de genen liggen die genetische verschillen in deze kenmerken bepalen.

Voor het opstellen van een koppelingskaart en het lokaliseren op deze kaart van genen die belangrijke kenmerken bepalen, zijn proeven nodig. Deze proeven kosten veel geld en moeten daarom zo efficiënt mogelijk opgezet worden. Bij aanvang van dit onderzoek was het niet duidelijk hoe gegeven de mogelijkheden van een pluimveefokprogramma deze proeven optimaal opgezet zouden moeten worden.

Na vaststelling van de ligging van genen op het genoom, kan deze informatie gebruikt worden tijdens het selecteren van dieren en bij het bepalen van de opzet van een fokprogramma. Voor melkvee is de optimale toepassing van merkers onderzocht. De optimale toepassing van merkers in een pluimveefokprogramma is echter onbekend.

Intermezzo:



Chromosoompaar (Chromosomen 1 en 2) van een ouder en de chromosomen die deze ouder doorgeeft aan nakomelingen. Met pijlen is op de chromosomen de plaats van drie merkers aangegeven. Tijdens de vorming van ei- of sperma-cellen treden overkruisingen tussen de twee ouderlijke chromosomen op. Een overkruising heeft tot gevolg dat een nakomeling een chromosoom krijgt dat samengesteld is uit stukken chromosoom 1 en stukken chromosoom 2. De kans dat tussen twee plaatsen op een chromosoom overkruising optreedt, hangt af van de afstand tussen de plaatsen. Hoe groter de afstand, hoe groter de kans op overkruising. Een overkruising kan resulteren in recombinatie tussen twee merkers: dat bijvoorbeeld op een chromosoom A1 samen voorkomt met B2 terwijl in de ouder A1 samenging met B1. Voor de situatie in dit intermezzo blijkt dat tussen merker A en merker B minder recombinaties voorkomen dan tussen de verder van elkaar afgelegen merker A en merker C.

Aan de relatie tussen afstand en recombinatie ontleent een merker zijn nut. Stel dat naast een merker een belangrijk polymorf gen ligt. Een ouder heeft op het ene chromosoom voor de merker het allel 1 en een + allel voor het belangrijke gen en op het andere chromosoom het merker allel 2 en een - allel voor het belangrijke gen. Nakomelingen van deze ouder die het merker allel 1 verkregen hebben, prefereren we boven nakomelingen die het merker allel 2 verkregen hebben omdat het waarschijnlijk is dat nakomelingen die merker allel 1 verkregen hebben ook het + allel verkregen hebben.

Doel

Het doel van dit onderzoek is om bij te dragen aan het efficiënt gebruiken van merkers in de pluimveefokkerij, en om te bepalen in welke mate merkers de genetische vooruitgang van een pluimveefokprogramma kunnen verhogen.
Om het doel te bereiken wordt onderzocht: de optimale opzet van proeven voor het maken van een koppelingskaart, de optimale opzet van proeven voor het plaatsen van belangrijke genen op de koppelingskaart, en de extra genetische vooruitgang die het gebruik van merkers bij het selecteren binnen een zuivere lijn van het pluimveefokprogramma met zich meebrengt.

Bij de opzet van proeven voor het maken van een koppelingskaart en het lokaliseren van genen, blijkt met name de familiestructuur van de populatie die in de proef gebruikt wordt van belang. Het aantal nakomelingen per ouder moet zo groot mogelijk zijn; hierbij kan optimaal gebruik worden gemaakt van de reproduktie-eigenschappen van de kip. Voor proeven om genen te lokaliseren is verder van belang welke informatie over de te onderzoeken kenmerken verzameld wordt. Gekozen kan worden voor een opzet zoals beschreven voor het bepalen van de associatie tussen veerkleur en eiproduktie. Hierbij wordt het kenmerk (in dat geval eiproduktie) gemeten aan dezelfde dieren als waarvoor de genetische merker (veerkleur) getypeerd is. Behalve genen, waarvan je er één hoopt op te sporen, bepalen ook omgevingsfactoren, zoals huisvesting en voeding, het kenmerk. Deze omgevingsfactoren hebben een storende invloed op de waarnemingen. Zonder omgevingsfactoren is het makkelijker om een verschil tussen groepen aan te tonen. De storing door omgevingsfactoren is minder als een kenmerk niet aan het dier zelf gemeten wordt, maar aan een groep nakomelingen van het dier. Het nakomelingengemiddelde kan gebruikt worden als waarneming voor het kenmerk. Hierdoor daalt het aantal benodigde merkertyperingen aanzienlijk ten koste van het aantal dieren waaraan het kenmerk gemeten moet worden. De optimale verhouding tussen het aantal merkertyperingen en het aantal kenmerk metingen hangt af van de kosten van merker typeren ten opzichte van de kosten van het meten van het kenmerk.

Bij het onderzoek naar het gebruik van een merker voor selectie binnen een zuivere lijn is gekeken naar een geslachtsgebonden kenmerk. Met name bij een geslachtsgebonden of karkasgebonden kenmerk is winst te verwachten omdat zo'n kenmerk niet aan alle selectiekandidaten gemeten kan worden. Voor eiproduktie bijvoorbeeld, hebben we voor alle volle broertjes dezelfde informatie, namelijk de prestaties aan half-zusjes en de geschatte fokwaarden van de ouders. Op basis van deze informatie kunnen we niet tussen deze volle broertjes selecteren. Een genetische merker voor eiproduktie zou hier uitkomst bieden. We hebben daarom een fokprogramma gesimuleerd voor een kenmerk dat alleen bij hennen tot expressie komt. Om inteelt te voorkomen wordt maar een beperkt deel van de volle broertjes

die uitkomen ook opgezet. Haantjes worden dus voorgeselecteerd. Als geen merkers gebruikt worden dan heeft deze voorselectie echter geen genetische vooruitgang tot gevolg. Er is maar één merker beschikbaar. Het effect van de merker op het kenmerk wordt gevarieerd. Het gebruik van de merker levert met name winst op tijdens voorselectie, waar een selectie op basis van de merker een willekeurige selectie vervangt. De winst die het gebruik van de merker oplevert, hangt met name af van de grootte van het effect van de merker, het aantal hennen waarmee een haan gepaard wordt en de scherpte van voorselectie (hoe scherper de voorselectie, hoe meer winst hier te boeken valt door het gebruik van de merker).

In de algemene discussie van het proefschrift is getracht het hele vakgebied op een rijtje te zetten. De status van de kippen-koppelingskaart, de genen die al gelokaliseerd zijn, de opzet van proeven en de verschillende gebruiksmogelijkheden van merkers worden besproken. Drie specifieke toepassingen worden extra toegelicht. De eerste toepassing is het gebruik van merkers tijdens de vorming van een nieuwe lijn voor een fokprogramma. Tijdens deze vorming van een nieuwe lijn worden vaak goede eigenschappen van verschillende kippen populaties bijeengebracht. Merkers kunnen gebruikt worden om de genen te volgen die verantwoordelijk zijn voor de goede eigenschappen zodat de nieuwe lijn ook inderdaad het goede verenigt. De tweede toepassing richt zich op een reductie van het aantal merkertyperingen. Van alle nakomelingen van een ouder wordt bloed getapt waarna bloed van goede nakomelingen gemengd wordt en bloed van slechte nakomelingen gemengd wordt. In de mengmonsters worden de frequenties van merkerallelen gemeten. Vervolgens wordt bekeken of bepaalde allelen in het goede mengmonster significant meer voorkomen dan in het slechte mengmonster. Het interessante van deze methode is nu dat op basis van een beperkt aantal merkertyperingen (2 per merker per ouder) associaties bepaald kunnen worden. De derde toepassing is het gecombineerde gebruik van merkers voor ouderschapscontrole en merker ondersteunde selectie. Als een hen met meerdere hanen wordt gepaard, is niet bekend wie de vader is van nakomelingen van de hen. Daarom wordt in het huidige fokprogramma een hen maar met één haan gepaard. Als we echter merkers gebruiken om ouderschap te controleren, dan kan een hen wel met meerdere hanen gepaard worden. Een fokprogramma waarin elke hen met meerdere hanen gepaard wordt, kan aanzienlijk efficiënter zijn dan een huidig programma. Het programma wordt extra interessant als de merkers die gebruikt worden voor ouderschapscontrole ook gebruikt worden voor het selecteren van dieren.

De drie bovenbeschreven specifieke toepassingen van genetische merkers geven aan dat onder optimale omstandigheden, het gebruik van genetische merkers een pluimveefokprogramma kan verbeteren. Genetische merkers worden nu al gebruikt voor ouderschapscontrole en het globaal bepalen van verschillen tussen populaties. Het hangt af van de associaties die men vindt tussen genetische merkers en economisch belangrijke kenmerken, of genetische merkers ook gebruikt gaan worden in selectieprocedures.

Levensloop

In 1985 verliet ik, Sijne van der Beek, het eiland Texel, alwaar ik op 15 juni 1967 geboren ben en mijn lagere en middelbare school opleiding genoten heb, om me te melden bij de Landbouw Hogeschool te Wageningen. In 1991 voltooide ik mijn opleiding Zoötechniek, specialisatie veefokkerij, middels afstudeervakken veefokkerij en agrarische bedrijfseconomie. Bij de vakgroep veefokkerij van de nu Lanbouwuniversiteit startte ik mijn promotie-onderzoek, waarvan u het resultaat in handen heeft. De laatste helft van 1995 was ik als post-doc werkzaam bij het Roslin Institute in Edinburgh, Schotland. Momenteel heb ik een hondenbaan: ik doe onderzoek naar de erfelijke achtergrond van ziekten bij honden.

Bedankt

Marry.
Pa en Ma, voor jullie genen en de omgeving waarin deze tot expressie kwamen.
Johan, voor je perfecte begeleiding die in losheid zeer wel strookte met mijn peuterachtige 'zelf doen' mentaliteit, en voor het leren maken van tussenstops.
Ab, voor het altijd klaarstaan als co-begeleider en je benadering die die van Johan aanvulde.
Pim, voor het vertrouwen in Johan en Ab, waardoor de begeleiding gewaarborgd was.
Euribrid, plat gezegd voor de centen, en voor het referentiekader.
Gerard, voor je betrokkenheid, en het overbrengen van een bewonderingswaardige mix van bedrijfsmatig realisme en wetenschappelijke nieuwsgierigheid.
Richard en Martien, voor het mogen exploiteren van jullie kennis.
Hendrik, Arnoud en Gerard, zonder wie fokkers hulpeloze zielen zouden zijn.
Imke en Luc, voor het leukste gewenningsproces van de jaren negentig.
Toin, vooral voor de laatste loodjes.
Ada, Addie, Ant, Carolien, Cees, Dom, Duncan, Geert, Henk, Jan, LAB-relatie-club, Marco, Margot, Mike, Peter, Richard, Ron, Ronald, Theo.
Marry.

Sijne.....