

24 Designs and Methods to Detect QTL for Production Traits Based on Mapped Genetic Markers

Johan A.M. van Arendonk and Henk Bovenhuis
*Animal Breeding and Genetics Group, Department of Animal Sciences,
Wageningen University, Wageningen, The Netherlands*

Introduction

Genetic improvement through artificial selection has been an important contributor to the enormous advances in productivity that have been achieved since the 1950s in plant and animal species that are of agricultural importance. Most of the traits that are selected for are complex quantitative traits, which means that they are controlled by several genes, along with environmental factors, and that the underlying genes have quantitative effects on the phenotype. So far, most selection has been on the basis of the observable phenotype, which represents the collective effect of all genes and the environment.

For many years, people have recognized the need for genetic evaluation of animals (e.g. Lush, 1931; Hazel, 1943; Henderson, 1949) and today best linear unbiased prediction (BLUP) has become the most widely accepted method for genetic evaluation of domestic livestock. Since Henderson (1949) first published the method for genetic evaluation, it has evolved from application to sire models in the early years to animal models and multivariate analysis more recently. These methods optimally combine phenotypic and pedigree information collected in a population. The application

of these methods has resulted in big changes in the performance of poultry. Besides selection within lines, the development of superior crossbreds through the combination of several improved lines or breeds plays an important role in poultry breeding.

The selection programmes were conducted without any knowledge of the genetic architecture of the selected trait. Dense linkage maps consisting of polymorphic marker loci are available for most livestock species and this is an important tool to unravel the genetic nature of quantitative traits in animals by identifying genes or chromosomal regions that affect the trait – so-called quantitative trait loci (QTL) (Groenen *et al.*, 2000; Anderson, 2001; see also Groenen and Crooijmans, Chapter 26).

Molecular genetic techniques can be used to collect more detailed genotypic information on individuals as well as information on which alleles or chromosomal segments are transmitted from parent to offspring. The purpose of this chapter is to provide background on the design and analysis of QTL mapping experiments in poultry. Mapping of QTL opens new opportunities for genetic evaluation and selection, as will be briefly described here and in other chapters of this book.

Phenotype/Genotype Relationship

The quantitative genetic approach to selection is based on knowledge of population genetic parameters for the traits of interest, such as heritabilities, genetic variances and genetic correlations. These parameters can be estimated using statistical analysis of phenotypic data from pedigrees (see Swaczkowski, Chapter 11). However, the genetic architecture of the trait itself is treated as a black box, with no knowledge of the number of genes that affect the trait, let alone of the effects of each gene or their locations in the genome. More specifically, quantitative genetic theory is based on Fisher's infinitesimal genetic model (Bulmer, 1980), in which the trait is assumed to be determined by an infinite number of genes, each with an infinitesimally small effect. Despite the obvious flaws of the infinitesimal model, the tremendous rates of genetic improvement that have been achieved attest to the usefulness of the quantitative genetic approach.

The prediction of an animal's breeding value is based on phenotypes of the animal itself and/or those of its relatives. When only observations on the trait of interest are considered, the contribution of observations on relatives to an animal's breeding value depends on the additive genetic relationship, i.e. the proportion of genes shared in common by descent, and the heritability of the trait. BLUP methods, by which estimates of fixed effects and predictions of breeding values are obtained simultaneously, require that the inverse of the numerator relationship matrix between animals (**A**), which is generally very large, be known. Henderson (1976) described a method for writing the inverse of **A** directly from pedigree records and inbreeding coefficients. This has enabled the use of improved methods for estimation of genetic parameters and prediction of the breeding value of animals. The concept of the numerator relationship matrix has been extended to the gametic relationship matrix (**G**), where paternal and

maternal gametes of an animal are considered separately (Smith and Allaire, 1985; Tier and Sölkner, 1993). The gametic relationship matrix has been used for constructing the relationship matrix due to dominance effects (Schaeffer *et al.*, 1989; Smith and Mäki-Tanilla, 1990) and for the analysis of gametic imprinting effects (Gibson *et al.*, 1988; Schaeffer *et al.*, 1989; Tier and Sölkner, 1993). In building the relationship matrix or its inverse at either the animal or gametic level, no knowledge of the actual contribution of a parent to its offspring is used. Instead, use is made of Wright's (1922) inbreeding coefficients and the coefficients of relationship between animals.

Traits

Observed scale

Many of the characters of interest to animal breeders are of a quantitative nature. On the other hand, there are other economically important traits that show a discrete phenotypic distribution, sometimes with as few as two classes. Many characters associated with reproduction and diseases are measured in a limited number of discrete classes. Lush *et al.* (1948), as one of the pioneers in genetic analysis of all-or-none traits in an animal breeding context, analysed mortality data in poultry. The essential difficulty posed by all-or-none traits, which these authors recognized and tried to accommodate, is that means and variances of the traits are correlated on the observed scale (McQuirk, 1989). In more recent analyses, the discrete phenotype is represented as a function of the animal's liability, which is modelled as a function of fixed effects and genetic and environmental effects (Gianola and Foulley, 1983). The underlying distribution of the genetic contribution is usually normal or logit of form. The remainder of this chapter will concentrate on phenotypes, which are measured on a continuous scale.

Importance

Poultry breeding made a substantial contribution to the increase of production volume and production efficiency per bird during the second half of the 20th century (Albers and Groot, 1998). In order to maintain an acceptable rate of progress, it is important that new tools are developed for effective selection on new traits such as leg weakness, cardio-respiratory problems, absorption problems, feather-pecking behaviour and disease resistance. These traits have in common that accurate recording of phenotypes on a large scale is difficult and that information on selection candidates is often not available prior to selection.

Selection for improved disease resistance in poultry is highly desirable for animal welfare, but also because reducing the losses due to disease is an excellent way of improving production efficiency. Selection for disease resistance is difficult to include in breeding programmes, because of the lack of accurate records on disease incidence in breeding flocks and the management practices that aim at minimizing exposure to pathogens. As a result there is considerable interest in finding genetic markers or diagnostic tests that can be applied to improve breeding programmes.

Dissecting the Genome

Molecular genetic techniques have made it possible to identify differences between individuals at the DNA level. In particular, the development of microsatellite markers as an abundant source of polymorphic and convenient markers has boosted the generation of linkage maps in the most important livestock species. These maps provide the basis for detection and exploitation of genes segregating at loci that have an effect on important traits. These are referred to as QTL. The principle of QTL mapping is not new; it was used by Sax (1923) to map a QTL associated with seed size by linkage

to a pigment locus in the bean *Phaseolus vulgaris*. The main practical limitation to implementing QTL mapping was the lack of a large number of evenly spaced Mendelian marker loci in most species. Two technical advances have opened the door for detailed characterization of the genetic architecture of quantitative traits: the discovery of abundant polymorphic neutral molecular markers and the development of sophisticated statistical methods for mapping QTL (MacKay, 2001).

QTL Mapping Methods

The principle of QTL mapping is simple. All that is required is two inbred strains in which different alleles at loci affecting the trait of interest are fixed, and a polymorphic molecular marker linkage map. In order to map the QTL, the inbred traits are mated to produce an F_1 population, which is subsequently used to produce a mapping population of backcross or F_2 individuals. The phenotype and multi-locus genotype of each individual in the mapping population are determined. In its simplest form, QTL mapping involves going through the genome, one marker at a time, dividing the individuals into marker genotype classes, and doing a statistical test to determine whether there is a significant difference in phenotype between the marker genotype classes. If there is such a difference, then the QTL is linked to the marker. This procedure, as described, underestimates the effect of the QTL by an amount that is proportional to the distance of the QTL from the marker locus, but this problem is readily overcome by mapping the QTL relative to two flanking markers (interval mapping).

If it is assumed that the inbred lines are fixed for alternative marker and QTL alleles, there are a number of important features of this design (Soller, 1991).

- All individuals in the F_1 are heterozygous for the marker as well as for the QTL.

- There is complete linkage disequilibrium between the marker and the QTL in the F_1 .
- There are only two marker alleles and two QTL alleles in the population under study, each allele with a frequency of 0.5.
- All individuals in the F_1 have the same genotype and the same linkage phase.

A design based on inbred lines is frequently used in laboratory animals (e.g. mice) and plants. For poultry and other farm animals, however, inbred lines are seldom available. In some experiments, crosses between lines with extreme phenotypes have been used instead. In addition, rearing large numbers of F_1 and F_2 animals is only practical for some farm animals (e.g. chickens) but not for others (e.g. cattle), due to the long generation interval and the costs of such an experiment (Bovenhuis *et al.*, 1997). As an alternative to a designed experiment involving crosses between lines or breeds, existing outcross populations can be used. In the latter approach the structure of breeding populations is exploited to map genes that are segregating within these elite populations. There are some important differences between analysing data from a cross between inbred lines and an outcross population.

- Only a fraction of the (grand)parents will be heterozygous for the marker as well as for the QTL.
- (Grand)parents might have different linkage phases and therefore observations in the mapping population cannot simply be pooled across families: marker effects need to be analysed within families.
- QTL might have more than two alleles in the population and allele frequencies are unknown.
- Linkage phase between marker alleles is unknown and needs to be inferred from the mapping population.

An experiment involving a cross between divergent populations is primarily focused on finding QTL that explain the variance between populations, whereas an

experiment within one outcross population will reveal QTL that explain the genetic variation within a population. It is still an open question whether QTL segregating in crosses between divergent lines also contribute to the genetic variance within breeds. De Koning *et al.* (1999) demonstrated that a cross between divergent lines can be used for the detection of QTL that explain between- and within-population genetic differences. A combination of both approaches is expected to provide information that is more informative for the dissection of the genome. Using divergent lines increases the probability that F_1 parents are heterozygous for the QTL and thereby increases the power of an experiment. At the same time, it will shift the emphasis towards finding genes that explain the differences between breeds. Emphasis on detecting genes that explain differences within a population offers the opportunity of marker-assisted selection within the breed. The choice of the population(s) involved in a genome mapping experiment depends largely on the goal of the experiment. When the emphasis is on finding genes that are involved in certain characteristics (i.e. genetic variation between lines), a cross between divergent lines is to be preferred. A single population or a cross between slightly divergent lines is to be preferred when the emphasis is on explaining within-population genetic differences or implementation of marker-assisted selection schemes.

In a number of poultry genome mapping experiments, a cross between different lines is used. In several cases, crosses between very divergent breeds have been used, such as the cross between Red Jungle Fowl and White Leghorn in Sweden (Andersson, personal communication) or the Sasumadori and White Plymouth Rock cross in Japan (Tatsuda and Fujinaka, 2001). In other cases the lines used were less divergent, such as the cross between lines divergently selected from a common base (Yonash *et al.*, 2001; Bovenhuis *et al.*, 2002) or the cross between commercial breeds used in Wageningen (Van Kaam *et al.*, 1998; Groenen *et al.*, 2001).

Principles of QTL Mapping

Inbred lines

To illustrate the principles of QTL mapping, let us first consider a cross between inbred lines in which the mapping population is genotyped for marker (with alleles *M1* and *M2*), which is linked to QTL (with alleles *Q1* and *Q2*). Inbred line 1 is homozygous for *M1* and *Q1* while inbred line 2 is homozygous for *M2* and *Q2*. These two lines are crossed to produce an F_1 population. Based on the F_1 , two types of design can be used for detecting QTL: the F_2 design and the backcross design. In a backcross design, F_1 individuals are crossed with one of the parental lines. Two types of backcross design are possible but consider a backcross with line 1. All F_1 individuals are double heterozygous and the *M1* and *Q1* alleles are on one chromosome and the *M2* and *Q2* alleles are on the other chromosome. An F_1 individual can produce four types of gametes: *M1Q1*, *M1Q2*, *M2Q1* and *M2Q2*. The *M1Q1* and *M2Q2* gametes are the parental types, i.e. this combination of alleles also occurred in one of the parents. Meiotic recombination generated the combination of alleles (*M1Q2* and *M2Q1*) which were not present in either of the parents and these gametes are referred to as recombinant. For genes close together on the same chromosome, the physical linkage of parental allele combinations makes independent assortment impossible and hence produces recombination frequencies significantly lower than 50%.

Table 24.1 gives the frequency of the gametes produced by an F_1 individual and

the corresponding genotypes and expected phenotypic values in the offspring. The expected difference in average phenotype ($\bar{y}$) between individuals in the backcross carrying the *M1M1* and *M1M2* marker genotype is equal to:

$$\bar{y}_{M1M1} - \bar{y}_{M1M2} = (1 - 2\theta)(a + d)$$

This equation illustrates that the expected difference depends on the recombination rate between *M* and *Q* (θ), the additive gene effect (*a*) and the dominance deviation (*d*) at the QTL.

An F_2 population is generated by crossing F_1 individuals. In the F_2 population, three marker genotypes can be distinguished: *M1M1*, *M1M2* and *M2M2*. The expected phenotypic difference between F_2 individuals carrying the *M1M1* and the *M2M2* genotype is equal to:

$$\bar{y}_{M1M1} - \bar{y}_{M2M2} = 2(1 - 2\theta)a$$

Compared with the backcross design, this contrast for a given recombination rate (θ) is only influenced by the additive gene effect at the QTL. Information on the size of the dominance deviation (*d*) can be obtained by including information on F_2 offspring with the *M1M2* genotype.

Outbred populations

Linkage disequilibrium in an outbred population between a genetic marker and a QTL can be found within families. This is due to the cosegregation of the marker and the QTL. The principle underlying the detection of a QTL in an outbred population will be illustrated for a two-generation design,

Table 24.1. Illustration of QTL mapping in a backcross population: frequency of gametes produced by F_1 parent and the resulting genotype and expected phenotype in the offspring.^a

Gamete F_1 parent	Type	Frequency	Genotype offspring	Phenotype offspring
<i>M1Q1</i>	Parental	$\frac{1}{2}(1 - \theta)$	<i>M1M1Q1Q1</i>	$\mu + a$
<i>M1Q2</i>	Recombinant	$\frac{1}{2}\theta$	<i>M1M1Q1Q2</i>	$\mu + d$
<i>M2Q1</i>	Recombinant	$\frac{1}{2}\theta$	<i>M1M2Q1Q1</i>	$\mu + a$
<i>M2Q2</i>	Parental	$\frac{1}{2}(1 - \theta)$	<i>M1M2Q1Q2</i>	$\mu + d$

^a θ is the recombination rate between *M* and *Q*, μ is the population mean, *a* is the additive effect at the QTL and *d* is the dominance effect at the QTL.

where sires have a large number of progeny and where the analysis is performed within half-sib families. The basic idea of the design is to trace marker alleles from the sire to his offspring and to determine whether offspring that have inherited alternative marker alleles from the sire differ with respect to the quantitative trait. When assuming a population with two marker alleles and two QTL alleles, in total ten different marker-QTL genotype combinations can be distinguished (Table 24.2).

Out of the ten possible sire genotypes in Table 24.2, only two are informative for the detection of the QTL, namely $M1Q1/M2Q2$ and $M1Q2/M2Q1$. For these two types of sires, the transmission of the chromosomal segment can be traced from sire to offspring (i.e. the sire is heterozygous for the marker) and a difference in the quantitative trait is expected when the QTL is linked to the marker (i.e. the sire is heterozygous for the QTL). For the $M1M1$ offspring it is certain that they inherited the $M1$ allele from the sire and similarly for the $M2M2$ offspring it is certain that they inherited the $M2$ allele from the sire. For the $M1M2$ offspring, however, it is not known which allele came from the sire and which from the dam. The expected difference in the quantitative trait for the $M1Q1/M2Q2$ sire between $M1M1$ and $M2M2$ offspring is equal to:

$$\bar{y}_{M1M1} - \bar{y}_{M2M2} = [1 - 2\theta] \times [a + (q - p)d]$$

where $\bar{y}$ is the mean phenotype, θ is the recombination rate between M and Q , a is the additive gene effect and d is the

dominance deviation at the QTL. The expected difference also depends on the allele frequency of $Q1$ (p) and $Q2$ (q) in the population. If it is assumed that the dominance deviation (d) is equal to zero, the expected difference between the groups of offspring inheriting a different marker allele from the sire reduces to $(1 - 2\theta)a$. This contrast is independent of the population QTL allele frequencies and is half the contrast observed for the cross between inbred lines.

- For an $M1Q2/M2Q1$ sire, the expected difference between $M1M1$ and $M2M2$ offspring is equal to $-[1 - 2\theta][a + (q - p)d]$. The absolute size of the expected difference is the same as for $M1Q1/M2Q2$ of sires but the sign differs.
- For sires homozygous at the QTL, no difference is expected between offspring inheriting alternative marker alleles. The proportion of sires homozygous at the QTL depends on the frequency of the QTL alleles and the number of QTL alleles. With only two QTL alleles, the proportion of sires with a homozygous QTL genotype is at least 50%. This minimum proportion reduces when more than two alleles are segregating at the QTL.

Design of Experiments

QTL mapping experiments involve many animals to be genotyped and performance tested (Soller and Genizi, 1978). Consequently, experimental designs need to be optimized to minimize costs of data collection and genotyping. Three different types of designs can be distinguished: (i) cross between divergent lines; (ii) within-family analysis where families could be full-sib or half-sib families; and (iii) full pedigree analysis exploiting all relations in a population.

Weller *et al.* (1990) computed the powers of experiments for balanced two- and three-generation half-sib designs for outbred populations. They quantified the influence on the power of size of QTL effect, heritability, family size and number

Table 24.2. All possible marker-QTL genotype combinations for a marker with two alleles ($M1$ and $M2$) and a QTL with two alleles ($Q1$ and $Q2$). The alleles on one chromosome are written adjacently and are separated from those on the other chromosome by a slash.

Marker geno- type	QTL genotype		
	$Q1Q1$	$Q1Q2$	$Q2Q2$
$M1M1$	$M1Q1/M1Q1$	$M1Q1/M1Q2$	$M1Q2/M1Q2$
$M1M2$	$M1Q1/M2Q1$	$M1Q1/M2Q2$	$M1Q2/M2Q2$
		$M1Q2/M2Q1$	
$M2M2$	$M2Q1/M2Q1$	$M1Q1/M1Q1$	$M2Q2/M2Q2$

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 structures are also feasible. Van der Beek *et al.* (1995) evaluated the powers of experiments with a full-sib structure. Some of their results are summarized here.

Inferences about the presence of a QTL linked to a 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 non-zero if a QTL is linked to the marker and the parent is heterozygous for the linked QTL. The square of the marker contrast divided by the square of the standard error of the marker contrast is used to compute a test statistic. When it is assumed that the phenotypic variance of the trait is known, this test statistic is a χ^2 statistic. Under the null hypothesis of no linked QTL, the statistic is a central χ^2 distribution. The null hypothesis is rejected when the statistic is larger than threshold T , the $(1 - \alpha)$ percentile of the central χ^2 distribution where α is the type I error rate. The power of a QTL mapping experiment is equal to the probability that the null hypothesis of no linked QTL is rejected, i.e. the probability that the test-statistic is larger than threshold T . The power of an experiment can be calculated (Weller *et al.*, 1990) as:

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

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, $\chi^2(\text{NC}(x), n_p)$ is a non-central χ^2 variable with n_p degrees of freedom and with a non-centrality parameter $\text{NC}(x)$, $\text{NC}(x)$ is the non-centrality parameter for the distribution under the alternative hypothesis given that x parents are heterozygous for the QTL and $P[\chi^2(\text{NC}(x), n_p) > T]$ is the probability that the non-central χ^2 variable exceeds the threshold T . The non-centrality of the distribution depends on the

expectation for the marker contrast, standard error of the marker contrast and the number of parents that are heterozygous for the linked QTL.

These parameters depend on the design of the experiment, as shown by Van der Beek *et al.* (1995), who looked at two- and three-generation designs. In both designs, marker genotypes are determined on parents and their offspring. In a half-sib design, marker genotypes are obtained for the sire and the half-sib offspring but not for the mates. In a full-sib design, marker genotypes are obtained for both parents and their full-sib offspring. In a two-generation design, the phenotypes are recorded on the offspring generation, whereas in a three-generation design each offspring is mated to produce grand-offspring. Phenotypes but not marker genotypes are recorded on these grand-offspring.

Table 24.3 presents the powers of two-generation experiments with full-sib and half-sib family structures for various numbers of families, various numbers of

Table 24.3. Power of experiments with a two-generation half-sib (HS2) or two-generation full-sib (FS2) family structure for a QTL explaining 1% of the phenotypic variance and with a heterozygosity of 50%, for various number of families (N_{fam}), various number of offspring per family (N_{off}) and two heritabilities ($h^2 = 0.1$ and $h^2 = 0.4$). (Source: Van der Beek *et al.*, 1995.)

N_{fam}	N_{off}	$h^2 = 0.1$		$h^2 = 0.4$	
		HS2	FS2	HS2	FS2
5	50	0.02	0.02	0.02	0.02
	100	0.03	0.04	0.03	0.04
	200	0.05	0.09	0.06	0.11
	400	0.13	0.24	0.15	0.31
	800	0.34	0.59	0.38	0.69
10	50	0.02	0.03	0.02	0.03
	100	0.04	0.05	0.04	0.07
	200	0.08	0.15	0.09	0.19
	400	0.23	0.43	0.26	0.54
	800	0.57	0.86	0.62	0.92
20	50	0.03	0.03	0.03	0.04
	100	0.05	0.09	0.06	0.11
	200	0.14	0.27	0.16	0.36
	400	0.42	0.72	0.47	0.83
	800	0.85	0.99	0.88	0.99

offspring per family and for two heritabilities of the observed trait. The results refer to a situation where interval mapping is used to detect a QTL that explains 1% of the phenotypic variance, that has a heterozygosity of 50% and that is located at the midpoint of a 20 cM interval flanked by two informative markers. An experiment with a full-sib family structure and n families has about the same power as an experiment with a half-sib family structure and $2n$ families for all values of n (Table 24.3). For a heritability of 0.1, an experiment with five full-sib families and 800 offspring per family had a power of 0.59 and an experiment with ten half-sib families and 800 offspring per family had a power of 0.57. The efficiency of an experiment with a full-sib family structure compared with a half-sib family structure was close to 2: 2.06 for $h^2 = 0.1$ and 2.25 for $h^2 = 0.4$. The power of an experiment with a two-generation family structure increased with increasing the number of families and with increasing the number of offspring per family (Table 24.3). Increasing the number of offspring per family was more efficient than increasing the number of families, e.g. for $h^2 = 0.1$, an experiment with ten full-sib families and 200 offspring per family had a power of 0.43, whereas an experiment with 20 full-sib families and 100 offspring per family had a power of 0.27.

The 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 24.3). For an experiment with five half-sib families and 800 offspring per family, the power was 0.34 for $h^2 = 0.1$ and 0.38 for $h^2 = 0.4$. The effect of h^2 on the power is slightly larger for a full-sib than for a half-sib family structure.

Powers of experiments with a three-generation family structure are given in Table 24.4 for three different family structures, namely HS3, FS3 and FSHS. In the three-generation half-sib structure (HS3), a sire has several half-sib offspring and each half-sib offspring is mated to several unrelated animals to produce one half-sib grand-offspring per mate per half-sib

Table 24.4. Power of experiments with a three-generation family structure (HS3, FS3 and FSHS) for trait with heritability of 0.10 and a QTL explaining 1% of the phenotypic variance and with a heterozygosity of 50%, for various number of families (N_{fam}), various number of offspring per family (N_{off}) and various number of grand-offspring per offspring (N_{goff}). (Source: Van der Beek *et al.*, 1995.)

N_{fam}	N_{off}	N_{goff}	HS3	FS3	FSHS
5	25	10	0.02	0.02	0.02
		50	0.04	0.04	0.11
		100	0.06	0.06	0.22
	50	10	0.03	0.03	0.04
		50	0.10	0.11	0.30
		100	0.15	0.15	0.55
	100	10	0.06	0.08	0.11
		50	0.26	0.31	0.67
		100	0.37	0.41	0.88
	25	10	0.02	0.02	0.03
		50	0.06	0.07	0.18
		100	0.09	0.09	0.39
10	50	10	0.04	0.05	0.06
		50	0.17	0.19	0.53
		100	0.26	0.27	0.82
	100	10	0.09	0.13	0.18
		50	0.45	0.54	0.91
		100	0.62	0.68	0.99

offspring. Weller *et al.* (1990) named a design with this family structure the 'grand-daughter design'. In the three-generation full-sib family structure (FS3), a pair of parents has several full-sib offspring. Each full-sib offspring is mated to one unrelated and ungenotyped animal to produce several full-sib grand-offspring per full-sib offspring. In the last family structure (FSHS), a pair of parents has several full-sib offspring as in FS3 but in FSHS each full-sib offspring is mated not to one but to several ungenotyped animals to produce one half-sib grand-offspring per mate per full-sib offspring. The power of an experiment with an HS3 family structure was similar to the power of an experiment with an FS3 family structure if the number of families (N_{fam}), number of offspring per family (N_{off}) and the number of grand-offspring per offspring (N_{goff}) were equal. For a heritability of 0.1, the power for

an FS3 family structure was 1 to 1.2 times the power for an HS3 family structure. For a heritability of 0.4, the power of an FS3 family structure was 0.85 to 1 times the power for an HS3 family structure (Van der Beek *et al.*, 1995). 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 grand-offspring. This resulted from the fact that the standard error of the marker contrast was larger when full-sib instead of half-sib grand-offspring were used.

The power of an experiment with an FSHS structure was higher than the power of an experiment with either an HS3 or an FS3 family structure. For a heritability of 0.1, the power of an experiment with an FSHS family structure was 1 to 4 times the power of an experiment with an HS3 or FS3 family structure (Table 24.4). This difference increased with an increase in heritability (Van der Beek *et al.*, 1995). The power of an experiment with a three-generation family structure increased with increasing the number of families and with increasing the number of offspring per family (Table 24.4). As with two-generation experiments, the power increased more by doubling the number of offspring per family than by doubling the number of families. The power increased more by doubling the number of offspring than by doubling the number of grand-offspring per offspring.

Factors influencing the power of the design

The results in Tables 24.3 and 24.4 show that family structure is an important factor in designing QTL mapping experiments. For a two-generation experiment, more offspring are 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 grand-offspring per three-generation experiment. This implies that fewer trait values have to be obtained for a two-generation experiment than for a three-generation experiment. For example, an experiment with ten two-generation full-sib families with 800 offspring per family had a power of 0.86, whereas a similar power (0.91) could be obtained with ten families in an FSHS design, with 100 full-sib offspring per family and 50 half-sib grand-offspring per full-sib offspring. The number of animals to be genotyped in the two-generation experiment is eight times the number in the three-generation design, whereas the number of phenotypes is 0.16 times the number of phenotypes in the three-generation experiment. In deciding on a design, the costs of typing the markers and the costs of recording trait values have to be considered as well as the time required to collect all the information. The costs of recording trait values are low if the 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 (e.g. Georges *et al.*, 1995; Spelman *et al.*, 1996). A three-generation design opens opportunities when interest lies in multiple trait and where each trait needs a specific environment. In the Wageningen poultry experiment, for example, groups of third-generation offspring were kept under different environmental conditions to measure individual feed intake (individual housing), carcass quality (group housing) and ascites (lower ambient temperature).

The variance explained by the QTL is important and depends on a^2 and heterozygosity of the QTL. The effect of doubling a^2 on power is the same as the effect of doubling the number of families (N_{fam}). As the effect of doubling N_{fam} is given in Tables 24.3 and 24.4, the effect of doubling a^2 can also be derived from those tables. Heterozygosity at the QTL was assumed to be 0.5. For a QTL with two alleles this is the maximum heterozygosity in a population that is in genetic

equilibrium. For a lower heterozygosity at the QTL, more families are needed for equal power. The relation between heterozygosity at the QTL and required number of families is about linear (Weller *et al.*, 1990). In the above power calculations, one outbred population was used; the efficiency of an experiment with a cross between divergent populations is higher (Soller, 1991). Genes with large effects are expected to segregate at higher frequencies in a cross than within an outbred population.

Selective genotyping

The number of marker genotypings in QTL mapping experiments is considerable, especially for the two-generation design. Selective genotyping may be a useful alternative to decrease the number of marker genotypings. The principle underlying selective genotyping (i.e. the genotyping of only those individuals that are extreme for a quantitative trait) was first described by Lebowitz *et al.* (1987). For crosses between inbred lines, Darvasi and Soller (1992) derived approximate formulae for the marker contrast and variance after selective genotyping has been applied. Selective genotyping in an outbred population should be applied within families. Bovenhuis and Spelman (2000) derived formulae for the marker contrast after selective genotyping in outbred populations. Darvasi and Soller (1992) and Bovenhuis and Spelman (2000) showed that selective genotyping can result in a considerable reduction of the number of genotypings for a given power. Selectively genotyping the 20% highest and 20% lowest animals resulted in a small reduction of the power compared with genotyping all animals.

Selective genotyping is often argued to have a major limitation if experiments are aimed at studying many traits. In such cases, selective genotyping could be applied to each of the traits, which would imply that most of the population needs to be genotyped and, thus, no reduction in genotyping can be obtained. In an experiment, one has to

choose to which trait selective genotyping should be applied. The highest power of detecting a QTL affecting a trait is achieved if selective genotyping is applied to the trait itself. However, the power of detecting a QTL affecting the trait is only slightly reduced if selective genotyping is applied to a highly correlated trait. The power is considerably reduced if selective genotyping is applied to traits with a correlation close to zero with the trait of interest. Bovenhuis and Spelman (2000) demonstrated that the power to detect a QTL affecting a correlated trait was at least as great as with random genotyping. In practice, this implies that selective genotyping should be applied to the trait of greatest interest, which might not be an optimal strategy for all traits.

Most QTL mapping experiments are not very large and will significantly detect genes with large effects, but for most QTL the evidence will not be conclusive. Those suggestive QTL need to be verified in further experiments (confirmation studies). A potentially interesting way of verifying marker-QTL associations is by using selective genotyping.

The high costs of screening large populations for marker allele frequencies can be markedly reduced by use of selective DNA pooling (Darvasi and Soller, 1994; Lipkin *et al.*, 1998). The marker contrast is determined by differences in allele frequency in pooled DNA samples of individuals taken from the extreme high and low tails of the phenotypic distribution of the population. Estimation of allele frequencies in pooled DNA samples is based on a linear relation between the allele frequency in the group of individuals making up the pool and the final allele band intensity after genotyping the pool, as determined by densitometry. In a retrospective study, Lipkin *et al.* (2002) successfully applied DNA pooling to an F_2 resource population created to map QTL affecting Marek's disease resistance. They observed very close correspondence between differences in allele frequency obtained on pools and differences obtained with individual genotyping. Their results support the proposition that selective DNA pooling can be combined with the large

half-sib families routinely produced in commercial genetic improvement programmes in layer and broiler chickens to provide high-power mapping of QTL affecting major production traits.

Choice of design

A QTL mapping experiment should be carefully designed. It is not possible to make a general statement on the design that is to be preferred. The experimenter needs to make a decision on the population(s) to be included in the QTL mapping experiment. This choice influences the probability that parents are heterozygous for QTL affecting the trait but it also influences the extent to which results can be applied in a commercial population. The heterozygosity at the QTL can be increased by using a divergent cross. However, it is still an open question to what extent these QTL also explain genetic variation within lines.

A QTL mapping experiment is expensive and the gains of optimizing the design may be high. Tables 24.3 and 24.4 demonstrate that more animals had to be typed for marker loci for a two-generation experiment than for a three-generation experiment. However, the number of animals to be phenotyped was larger in the three-generation design. In addition, the three-generation design enables the collection of information on a range of phenotypes and thereby offers the opportunity for information on interaction phenotypes. For example, production traits can be recorded for one group of offspring in a disease-free environment, while disease-resistance traits can be recorded for another group of offspring in a challenge experiment.

When the costs of recording phenotypes are low compared with the costs of genotyping, a three-generation design will result in lower total costs than a two-generation design. The costs of recording phenotypes also play an important role in deciding on the use of selective genotyping.

Methods of Analysis

Genetic mapping of a trait comes down to finding those chromosomal regions that tend to be shared among relatives performing well and tend to differ between the latter and poorly performing relatives. Conceptually, this mapping amounts to the following steps: scanning the entire genome with a dense collection of genetic markers; calculating an appropriate linkage statistic at each position along the genome; and identifying the regions in which the statistic shows a significant deviation from what would be expected under independent assortment (Lander and Kruglyak, 1995). A practical approach for estimating the position of a QTL is to perform a genome-wide scan for QTL using linkage analysis methods. This would result in a region of, say, 20 cM, that most likely contains the QTL. Next, this region is covered with additional markers, and linkage disequilibrium mapping methods can be used to narrow the position of the QTL. Finally, comparative mapping and/or positional candidate cloning may be used to identify the QTL (Anderson, 2001). Grisart *et al.* (2002) successfully identified a gene on bovine chromosome 14 in a QTL region that was initially identified using linkage analysis.

Different approaches can be used to analyse data from a QTL mapping study (Bovenhuis *et al.*, 1997; Hoeschele *et al.*, 1997; Doerge, 2002). This chapter will consider four different methods: variance analysis; regression; mixture models; and random effects models. Methods based on a variance component approach are described in another chapter. The first three methods are generally used for the analysis of an experiment with a two- or three-generation design in which transmission of marker alleles is traced from parent(s) to offspring. Methods are categorized based on the statistical-genetic model (number of QTL alleles, number of genetic effects) underlying the analysis and the information used in the analysis. For some of the models

different parameter estimation procedures can be used, e.g. maximum likelihood or Markov chain Monte Carlo; consequently, categorizing on the basis of parameter estimation procedure might be misleading (Bovenhuis *et al.*, 1997).

Analysis of variance using a single marker

Traditionally, detection of QTL has been performed by contrasting the phenotypes for marker genotype effects (e.g. Soller *et al.*, 1976). In this type of analysis (for example, using *t*-test, ANOVA and simple linear regression statistics), markers are analysed one at a time to assess the segregation of a phenotype with respect to a marker genotype. Typically, the null hypothesis tested is that the mean of the trait value is independent of the genotype at a particular marker. The null hypothesis is rejected when the test statistic is larger than a crucial value, and the implication is that a QTL is linked to the marker under investigation. The marker genotype effects can be tested using an *F*-test, as is common in analyses of variance. For an outcross population, significant marker effects at the population level are only expected if there is a linkage disequilibrium between the QTL and the genetic marker. Neimann-Sorensen and Robertson (1961) noted that for an outcross population, analysis of marker genotype effects should be performed within sire families. This can be done by contrasting progeny within a family that have inherited alternative parental alleles. This method of analysis yields estimates of marker allele substitution effects. However, the analysis does not provide any information about the location of the QTL, i.e. the model cannot distinguish between a tightly linked QTL with a small effect and a loosely linked QTL with a large effect. Another disadvantage of the method is that some of the progeny cannot be assigned to one of the two parental alleles. These animals have to be excluded from the analysis, resulting in reduced power.

Regression methods

Single-marker analysis investigates individual markers independently, and without reference to their position or order. When markers are placed in genetic (linear) map order, so that the relationships between markers are understood, the additional genetic information gained from knowing these relationships provides the necessary setting to address confounding between QTL effect and location. In addition to supplying the structure in which to search for QTL, the estimated genetic map benefits the estimation of missing marker information by using the surrounding marker genotypes to infer knowledge of the missing marker genotypes.

Haley and Knott (1992) and Martinez and Curnow (1992) independently introduced a regression method of analysis that is capable of utilizing information from flanking markers. When using information from a single marker only, regression is equivalent to analysis of variance. Instead of regression on marker genotypes, regression is performed on the probability of having a QTL genotype, given the genotypes for the flanking markers.

Line cross analysis

Under the line cross model, it is assumed that the two founder populations, although they may share alleles at the marker loci, are fixed for alternative alleles at the QTL affecting the trait of interest. The principles involved for an F_2 resource population are described here.

The analysis of an F_2 population can be considered in two stages. In the first stage, at locations throughout the genome, the probability of an F_2 offspring being each of three possible QTL genotypes (*QQ*, *Qq* and *qq*) is calculated at 1 cM intervals along the genome conditionally on the marker genotypes. In the second stage, these probabilities are used in a least-squares framework to investigate the genetic model underlying the trait of interest. Following Falconer and

Mackay (1996) and denoting the effect of QQ as a , the effect of Qq as d and the effect of qq as $-a$, the phenotype of an offspring at a given position in the genome can be written as a linear model in terms of the additive and the dominance contributions at a QTL:

$$y_i = \mu + c_{ai}a + c_{di}d + e_{ij}$$

where μ is the mean, c_{ai} is the coefficient for the additive component for individual i at the given position that, denoting the probability of an individual being genotype XX as $\text{prob}(XX)$, is equal to $\text{prob}(QQ) - \text{prob}(qq)$, c_{di} is the coefficient for the dominance component for individual i at the given position, which is equal to $\text{prob}(Qq)$, and e_{ij} is the residual effect. The calculation of these probabilities and the QTL effects are described by Haley *et al.* (1994).

The method developed by Haley *et al.* (1994) is suitable for crosses where the lines may be segregating at marker loci but can be assumed to be fixed for alternative alleles at the major QTL affecting the traits under analysis (e.g. crosses between divergent selection lines or breeds with different selection histories). The simultaneous use of multiple markers from a linkage group increases the sensitivity of the test statistic, and thus the power for the detection of QTL, compared with the use of single markers or markers flanking an interval. Use of multiple markers can also remove the bias in the estimated position and effect of a QTL that may result when different markers in a linkage group vary in their heterozygosity in the F_1 and are considered only singly or a pair at a time.

Half-sib and full-sib models for outcross populations

A multi-marker approach for interval mapping in a paternal half-sib design was described by Knott *et al.* (1996). In this design, the relationship between the sire and its offspring is used and all additional relationships between and within half-sib groups are ignored. For each offspring, the probability of inheriting the sire's first gamete of a chromosome is calculated at 1 cM intervals conditional on the most-linkage phase of the sire and the marker

genotypes of the offspring and its parent(s). A QTL with a gene substitution effect is fitted at 1 cM intervals along the chromosome:

$$y_{ij} = \mu + s_i + b_i x_{ij} + e_{ij}$$

where μ is the overall mean, s_i is the average effect of half-sib family i , b_i is allele substitution effect for the QTL in half-sib family i , x_{ij} is the conditional probability for individual j of inheriting the first parental allele and e_{ij} is the residual effect. The regression is nested within half-sib families because the assignment of the first gamete is random and not all sires are heterozygous for the QTL. Furthermore, the linkage phase between a marker and a QTL can differ between families. The number of QTL alleles is only constrained by the number of families. The test statistic is calculated as an F ratio for every map position within and across families. For details on the calculation of the test statistic, see Spelman *et al.* (1996).

The above method was extended to full-sib family structure by Van Kaam *et al.* (1998). In that case, marker information on both parents and their full-sib offspring is used to trace the transmission of QTL alleles from both parents to their offspring and an allele substitution effect is calculated for both parents.

Combination of approaches

De Koning *et al.* (1999) applied linkage analysis on Meishan crossbred pigs under two genetic models: (i) an outbred line cross model, where the founder lines were assumed to be fixed for different QTL alleles; and (ii) a within-family model, where a unique allele substitution effect was fitted within every paternal half-sib family. They demonstrated that the combined use of both models provides additional insight into the genes involved and their effects, compared with the use of one of these approaches. The latter is because animals within each divergent breed may also exhibit genetic variation for the QTL. The QTL effects estimated under a line cross model are biased downwards when QTL alleles are not fixed in the parental

lines, whereas a half-sib analysis is able to detect genes, albeit at a lower power, when alleles are segregating in one or both breeds.

Mixture model analysis

Weller (1986; Weller *et al.*, 1990) developed mixture model methods to detect QTL using single marker information in crosses between inbred lines and for outcross populations. Solutions are obtained by maximizing the likelihood function. In calculating an individual's QTL genotype probability, information from marker genotype as well as information from its phenotype is used. The mixture model approach suggested by Weller (1986) for crosses between inbred lines has been extended to take into account information from flanking markers (Lander and Botstein, 1989). A mixture model approach for the mapping of QTL in outbred populations was presented by Jansen *et al.* (1998). This method can be applied to situations in which information about the genotype of an individual is incomplete.

Random effects model

A QTL mapping experiment in an outbred population is typically based on genotypes for two generations of animals (parents and their offspring) and on phenotypes measured on animals of the second generation or their offspring. In most statistical analyses, the first-generation animals (parents) are assumed to be unrelated; the second-generation animals (offspring) are assumed to be related only through one common parent (usually the sire) in a half-sib analysis or through both parents in a full-sib analysis. This assumption does not hold for most commercial livestock populations, in which many additional genetic relationships are present. For example, in a dairy cattle population, bull dams may have multiple sons tested in the population, or bull

dams are related to grandsires. In these situations, a full pedigree analysis, accounting for all relationships, is more appropriate. Such an analysis cannot be performed by traditional analysis methods such as linear regression or maximum likelihood interval mapping, because these methods assume unrelated grandsire families and only two generations of genotyped individuals. Bink and Van Arendonk (1999) developed a Bayesian approach for a full pedigree analysis with incomplete marker data. Simulation results for a typical dairy cattle population indicated a significant increase in power of QTL detection when all relationships were included in the analysis.

In the genetic model, allelic effects at the QTL in an outbred population may be represented by normally distributed random effects, where covariances between allelic effects depend on the identity by descent probabilities that are derived from marker information (Fernando and Grossman, 1989; Van Arendonk *et al.*, 1994). Normality of QTL effects is a robust assumption to allow segregation of many alleles in the population and to allow changes in allelic effects over generations. Changes in allelic effects might originate from mutations, interactions between genes, dominance, and interaction with the environment. As an alternative to this so-called infinite-allele QTL approach, a single biallelic QTL can be modelled in which individuals are assigned to QTL genotypes based on information from flanking markers (Hoeschele *et al.*, 1997). This approach can be extended to situations with more alleles. When molecular genetic information on the mutations causing the allelic differences increases, it becomes conceptually more appealing to use a finite-allelic model for the QTL. However, even in cases where the actual number of alleles is known, this does not imply that the number of effects to be modelled is finite. The effects of a single mutation on the phenotype might differ depending on the genetic background (e.g. family) in which it occurs.

For more information on these types of models, see Yi and Xu, Chapter 25.

Correcting phenotypes

In the analysis of a two-generation design, the least-squares framework can be extended to incorporate fixed effects and covariates in the model. In a three-generation design, a two-step procedure is mostly used. In the first step, the trait values on the third-generation animals are corrected for fixed effects and covariates and subsequently accumulated for each second-generation animal. In the second step, the adjusted and accumulated trait values are used in the interval regression procedure. The number of offspring may differ between second-generation animals. This difference influences the residual variance of the average trait values and should be included in subsequent analysis. Van Kaam *et al.* (1998) described the weights in a regression method for a three-generation full-sib design.

Effects to be estimated

QTL mapping studies also provide a better insight into the mode of inheritance of traits. So far, the emphasis has been on QTL on the autosomes that show Mendelian inheritance (additive effects and dominance deviation). Recently, a few QTL have been described in livestock species that exhibit non-Mendelian inheritance. Cockett *et al.* (1996) found that the callipyge gene, which results in muscular hypertrophy in sheep, shows a non-Mendelian inheritance pattern, referred to as polar over-dominance, where only heterozygous individuals having inherited the callipyge mutation from their sire express the double-muscling phenotype.

Genomic imprinting is commonly regarded as a rare phenomenon and consequently is ignored in most studies. Nezer *et al.* (1999) and Jeon *et al.* (1999) identified an imprinted gene near the IGF-2 locus on porcine chromosome 2 that affects muscle mass and fat deposition. De Koning *et al.* (2000) indicated that genomic imprinting might be a more common phenomenon than

previously thought. They detected five QTL, of which four were subject to imprinting. For at least two of these regions, imprinting had not been reported before in pigs, and the known homologies to humans and mice did not reveal obvious positional imprinted candidates.

For detection of imprinting in a QTL analysis, it is essential that parental origin of marker alleles can be derived for the offspring (De Koning *et al.*, 2000). This requirement excludes studies based on F_2 crosses or a single backcross between inbred lines, as commonly used in mice and rats. These model species have contributed enormously to the current understanding of genetic variation. The inability to detect imprinting in the most commonly used mapping designs has certainly contributed to the current feeling that imprinting is a rare phenomenon.

Multiple QTL Analysis

Currently, single-QTL methods are still widely used in plant, animal and human genetics, but they are intrinsically inappropriate for complex traits affected by multiple QTL. In experimental plant applications, multiple-QTL models are now used more and more frequently; background QTL are taken into account by including them (via linked markers) as cofactors in the model (proposed by Jansen, 1992; see Jansen, 1996, for a review). This can be done in plants because complete marker maps are available for many plant species and also because experimental plant populations, e.g. F_2 or backcross, are easier to deal with from the analytical point of view. In animal and human applications, the effects of background QTL are often modelled by a single variance component term. In general populations, a marker can be segregating in some families whereas the QTL is not and vice versa. Then a marker linked to a putative QTL cannot be used as the cofactor in the (expectation or mixture) model, as also indicated by e.g. Spelman *et al.* (1996). In such cases, the QTL really should be included instead of the marker as cofactor

in the model, although the putative QTL can be put close to or even on top of a marker. Eventually dense marker maps may become available in human and animal applications, and, with cofactors for background QTL, it may not be necessary to include other parameters for genetic background control. Modelling via cofactors will also make it possible to explain differences in variances between families originating from segregation of different sets of genes and therefore residual variance may be assumed to be homogeneous; this cannot be achieved by a model term for polygenic background effect.

Selective DNA Pooling

DNA markers are now widely used for the detection and mapping of QTL. Selective genotyping – the marker assay of only individuals with the more extreme phenotypes for a quantitative trait – can provide considerable savings in genotyping costs while retaining most of the statistical power for detection of QTL affecting the trait on which the selection is based (Lebowitz *et al.*, 1987; Bovenhuis and Spelman, 2000). For single-trait studies it will almost never be useful to genotype more than 50% of the population (the high and low tail). However, linear model estimates of the QTL effect, for which individuals without genotypic information are excluded from the analysis, will be biased by the selective genotyping (Darvasi and Soller, 1992).

A mixture model approach for the mapping of QTL in outbred populations was presented by Jansen *et al.* (1998). This method can be applied to situations in which information about the genotype of an individual is incomplete. Incomplete information might be caused by the impossibility of tracing the inheritance of an allele at a locus in an individual, unknown linkage phases between loci, unknown QTL genotype and unknown genotypes for markers. The method can, therefore, be applied to selectively

genotyped data, but no information is available on the properties of the estimates.

A model can be defined to describe the relationship between phenotype and 'known' genotype. Since the genotype is in reality unknown, the possible genotype configurations that arise from this uncertainty then become the components of a mixture and this can be handled by an expectation maximization (EM) algorithm that yields maximum likelihood estimates of the model parameters (Jansen *et al.*, 1998). A simulation study is used to investigate the performance of this mixture model approach when selective genotyping is employed within a half-sib family structure (Johnson *et al.*, 1999). This approach was also compared with the multi-marker regression method (Knott *et al.*, 1994).

The major cost in the detection of QTL with the aid of genetic markers is that due to DNA collection and typing. Selective genotyping can provide considerable cost savings, particularly in those populations where recording of phenotypes is done on a routine basis, with little loss in accuracy of detection of QTL. Linear model regression estimates of allelic effects, such as those obtained from a multi-marker regression method, are biased upwards when selection is used to genotype only those individuals that are extreme for the quantitative trait. This is due to the positive correlation between residual effects and the QTL effect in the pooled tails population that magnifies the allelic effect.

On the basis of simulation work (Johnson *et al.*, 1999), the mixture model method would appear to yield estimates of gene substitution effects that are not biased by selective genotyping. This is the case not only for the primary trait on which selection is based but also for a correlated trait, provided that the latter is analysed jointly with the primary trait. Presumably critical to this result is the fact that the Monte Carlo (MC) EM method, when sampling missing marker genotypes, takes into account not only known marker information but also all phenotypic observations for the trait.

Linkage Disequilibrium Mapping

The quest for genetic loci corresponding to QTL will ultimately arrive at a point where the QTL is mapped to a small chromosomal region, in which all genes and predicted genes are known. Which of these loci is associated with the QTL? Having determined the genetic locus corresponding to the QTL, linkage disequilibrium mapping can be used to determine what molecular polymorphisms – single nucleotide polymorphisms (SNPs), small insertions/deletions or large insertions – define a functional QTL allele. Linkage disequilibrium is a measure of the correlation of allele frequencies at two polymorphic loci. Population genetic theory predicts that a molecular marker locus will be in linkage disequilibrium with a molecular variant affecting the quantitative trait (the quantitative trait nucleotide, or QTN) only if they are very tightly linked – provided that the population demography satisfies the theoretical assumptions.

The linkage disequilibrium between a single marker and a QTL can be measured by estimating the effect of the marker on the quantitative trait in a regression analysis. This approach is extended to multiple marker loci by estimating the effect of marker haplotypes on the quantitative trait. The marker haplotypes that have identical marker alleles in a region surrounding the QTL are expected to show similar haplotype effects, since the identical markers indicate that the region is identity-by-descent (IBD) and thus the haplotypes are expected to carry similar QTL alleles. In statistical terms, similar haplotype effects imply that the covariance between the haplotype effects is high. Whether two marker haplotypes have identical alleles in a region surrounding the QTL depends on the position of the QTL; hence, the covariance between the haplotype effects depends on the position of the QTL. Meuwissen and Goddard (2000) developed a method for the estimation of the QTL position using the linkage disequilibrium between several closely linked markers and the QTL.

Statistical Significance

Regardless of the method used to estimate and locate single or multiple QTL, once the test statistics are calculated the likelihood of the event is assessed. The statistical basis of these comparisons relies on model assumptions, the most common of which requires the quantitative trait values to be normally distributed (Doerge, 2002). If in fact these assumptions hold, many wonderful statistical properties follow and valid conclusions can be drawn from a range of powerful statistical tests. In reality, however, the distribution of the trait values is not normal, and needs to be considered as a mixture of (normal) distributions. Violating the normality assumption has an impact on the distribution of the statistic used to test for a QTL, which makes standard statistical procedures potentially inaccurate.

One approach to obtaining the distribution (or behaviour, in the long term) of the test statistic is to use a computer simulation to produce the data (Lander and Botstein, 1989; Jansen, 1994). Thousands of data sets, taken from the same statistical model, are simulated and analysed in order to represent the statistical distribution of the particular test statistic. From this distribution, one chooses the level of statistical significance or threshold above which results are considered statistically significant (valid). This approach assumes that the true model is used to simulate the data, which is rarely the case. Non-parametric resampling methods (Lander and Kruglyak, 1995) provided a useful alternative to simulation-based thresholds. In a large number of studies, the permutation resampling (Churchill and Doerge, 1994) is used as a means of randomizing the phenotypic (trait) data for the purpose of evaluating any test statistic under a null hypothesis that tests for a QTL. For a discussion of this and alternative methods, see Doerge (2002).

Mapping of QTL is plagued by the statistical problem of multiple tests. If only one test is done, 5% is conventionally accepted as a significance threshold. That

is, the result of a statistical test is deemed 'significant' if its probability of occurring by chance alone is 5% or less. However, if many tests are done on the same data, as occurs when associations between multiple markers and the quantitative trait are considered, 5% of the associations are expected to be significant by chance. So, the significance threshold for the entire experiment must be adjusted downwards to correct for the number of independent tests.

Chromosome-wise significance thresholds, which account for multiple testing on a single chromosome, can be determined empirically by permutation resampling. Lander and Kruglyak (1995) distinguished two significance thresholds: suggestive and significant linkage. The first level was suggestive linkage, where one false positive is expected in a genome scan. For claiming significant linkage a 5% genome-wide significance level can be applied which can be calculated from the empirically determined chromosome-wise significance level using a Bonferroni correction.

In comparing results from different QTL experiments, most attention is focused on significant results. However, it is important to pay attention to the significance thresholds that have been applied.

Molecular Tools

Molecular genetic analyses of quantitative traits lead to the identification of two broadly different types of genetic loci that can be used to enhance genetic improvement programmes: (i) causal mutations; and (ii) presumed non-functional genetic markers that are linked to QTL (indirect markers). Causal mutations for quantitative traits are hard to find and difficult to prove and few examples are available. By contrast, non-functional or anonymous polymorphisms are abundant across the genome and their linkage with QTL can be established by evidence of empirical associations of marker genotypes with trait phenotype. Two approaches are used to identify indirect markers: (i) directed searches using

candidate-gene approaches in unstructured populations; and (ii) genome-wide searches in specialized populations, such as F_2 crosses. Because candidate-gene markers focus on polymorphisms in a gene that are postulated to affect the trait, they are often tightly linked to the QTL. A candidate-gene marker can occasionally represent the functional variant itself, although this is difficult to prove. Genome scans, conversely, can only identify regions of chromosomes that affect the trait. The length of these regions is typically 10–20 cM, but the exact position and number of QTL in the region is unknown.

The candidate gene approach can be very powerful and can detect loci even with small effects, provided that the candidate gene represents a true causative gene. However, there are often many candidate genes and a major trait locus might fail to be identified simply because of the gaps in knowledge about gene function (Anderson, 2001). Candidate gene tests must also be interpreted with caution as spurious results can occur because of linkage disequilibrium to linked or non-linked causative genes or because the significance thresholds have not been adjusted properly when testing multiple candidate genes. A genome scan will always find the map location of a trait locus with a major effect, provided that an accurate genetic model has been postulated, a reasonable sample size has been used and the marker set provides full genome coverage. A genome scan will fail to detect trait loci with smaller effects if they do not reach the stringent significance thresholds that must be applied when doing a large number of tests in a full genome scan.

High-resolution mapping of trait loci can be carried out by IBD mapping (Anderson, 2001), which involves the collection of DNA samples from individuals that have inherited a certain allele at a trait locus from a common ancestor. The samples are screened with genetic markers to detect the minimum chromosomal region IBD among the individuals. A recent study of dairy cattle populations revealed fairly strong linkage disequilibrium for loosely linked loci and even weak linkage disequilibrium between

loci on different chromosomes (Farnir *et al.*, 2000). Linkage disequilibrium in farm animal populations is generated by genetic drift, the effects of which are more marked because of the limited number of breeding animals, migration and population stratification. The presence of linkage disequilibrium increases the probability of finding true associations between genetic markers and trait loci but also increases the risk of spurious associations in candidate gene studies. For example, admixture (interpopulation gene flow) between populations that have different gene frequencies at the marker and different values of the trait will cause spurious linkage disequilibrium, even between unlinked loci.

Relation to Marker-assisted Genetic Evaluation

Molecular genetic information can be used to enhance several breeding strategies through what is broadly referred to as marker-assisted selection (MAS). For an overview of different strategies for MAS in the improvement of agricultural populations, we refer to Dekkers and Hospital (2002). Genetic evaluation plays an important role in breeding programmes. The remainder of this section will discuss strategies to utilize QTL information in genetic evaluation schemes.

Genotypes at marker loci give information on transmission of genes from parents to offspring and that information can be used in predicting the individual's additive genetic value at linked QTL (Fernando and Grossman, 1989; Van Arendonk *et al.*, 1994). A mixed linear model exists for simultaneous evaluation of fixed effects, genetic effects at the QTL and the additive polygenic effect. These methods are efficient when marker data are complete but computations become intractable when the proportion of ungenotyped individuals increases. At present, the high costs for marker genotyping prohibit the routine genotyping of

large numbers of animals. Bink *et al.* (1998a) presented an approach for marker-assisted genetic evaluation allowing incomplete information on a single marker that is linked to the QTL. The efficiency of this procedure can be greatly improved by absorbing the genetic effects of ungenotyped offspring in a reduced animal model (Bink *et al.*, 1998b).

Marker-assisted genetic evaluation will first be introduced in the nucleus of the population, i.e. the group of animals with high genetic potential that are the candidate parents for the next generation. For poultry the nucleus might be a closed population, whereas for dairy cattle the nucleus has an open character. In an open nucleus system, parents of nucleus individuals may be present in the commercial population. Procedures to incorporate information from these parents (without including them explicitly in the genetic evaluation) need to be developed, similar to procedures that incorporate foreign information in national genetic evaluation of dairy cattle.

In prediction of genetic effects at marked QTL, there may be a desire to assume that there are accurate estimates for QTL location and QTL variance. This would greatly reduce computational requirements. Initially the number of animals within a breeding population with observed marker genotypes will probably be too small for accurate estimation of QTL parameters (Spelman and van Arendonk, 1997). In that case, parameter estimates obtained from a QTL mapping experiment seem most appropriate. The amount of information within the nucleus will increase over time and this enables estimation of the genetic parameters from the current breeding population and increase the accuracy of genetic evaluations. Information on multiple QTL can be combined into a single relationship matrix where each QTL is weighed according to its genetic variance. The weighting will differ between traits, which opens attractive opportunities for selection particularly in situations with antagonistic relationships between traits.

Some Results

Khatib (1994) studied the association of 28 microsatellite markers with juvenile growth rate in a cross between White Leghorn layer females and a single White Rock broiler male. He found 11 markers that were significantly associated with growth rate, but sizes of effects were not given. In a genome-wide scan with 437 microsatellite markers, Van Kaam *et al.* (1999a) reported four QTL that exceeded the significance threshold in a feed efficiency experiment. Microsatellite genotypes were determined on ten full-sib families with on average 45 offspring per family. Feed efficiency was recorded on over 2000 grand-offspring. The QTL analysis was undertaken by modelling the segregation from generation one to generation two, using a full-sib across-family regression interval mapping approach. Genome-wide significance thresholds were derived using permutation test and a Bonferroni correction. The most significant QTL was located on chromosome 1 and had a 4% genome-wide significance for feed intake between 23 and 48 days. Furthermore, this QTL exceeded suggestive linkage for growth between 23 and 48 days and body weight at 48 days. A second QTL was located on linkage group WAU26 and showed suggestive linkage for feed intake between 23 and 48 days. The same full-sib families were used to produce another set of 2000 grand-offspring to map QTL affecting carcass traits (Van Kaam *et al.*, 1999b). This experiment revealed a suggestive QTL affecting carcass percentage on chromosome 1 and a suggestive QTL affecting meat colour on chromosome 2.

Tatsuda and Fujinaka (2001) identified two QTL affecting body weight at 13 and 16 weeks in an F_2 cross between the indigenous Satsumadori and White Plymouth Rock chickens. The QTL were mapped to chromosomes 1 and 2 and the positions were in agreement with earlier studies.

In another study of Marek's disease (MD), searching a smaller resource population of egg-type chickens with fewer markers yielded 14 different genomic regions (QTL)

associated with the disease (Yonash *et al.*, 1999). The successful detection of QTL in this small population was probably due to the unique population derived from two inbred parental lines that differed in susceptibility to MD.

A population comprising F_2 and BC families, derived from a cross between two meat-type (broiler) lines that were divergently selected for early immune response, exhibited large genetic variation that facilitated the successful detection of RFLP markers linked to the selected trait (Yonash *et al.*, 2001). However, because of the outbred nature of the parental lines, the necessary linkage disequilibrium was obtained only within families (Hillel, 1997). These studies utilized a unique resource population, derived from a cross between two divergent lines, to search for linkages between microsatellite DNA markers and QTL associated with immune response in young broilers.

QTL on sex chromosome

Most of the emphasis in QTL mapping has been directed to autosomal chromosomes, but there might also be QTL located on the Z chromosome. These QTL might explain the observed difference in mean and variance of growth between males and females. Furthermore, Tixier-Boichard *et al.* (1995) found a genetic correlation between adult male and female body weight of 0.71. Hagger (1994) reported a genetic correlation of 0.84 between adult male and female body weights. These findings suggest that male and female body weights are genetically different traits, which might be due to genes located on the sex chromosome. In mice, a large QTL was detected on the sex chromosome affecting body weight (Rance *et al.*, 1997a,b). The estimated effect of the QTL was approximately 20% of the mean body weight in males and females at 10 weeks of age. On the sex chromosome of pigs, QTL were detected affecting backfat thickness and intramuscular fat content

(Harlizius *et al.*, 2000). These findings suggest that, in other species, genes affecting growth traits might also be detected on the sex chromosome. Detection of genes located on the Z chromosome differs from the detection of genes located on autosomal chromosomes. Hamoen *et al.* (2001) scanned the chicken Z chromosome for genes affecting growth traits and feathering. For this purpose parents and full-sib offspring were genotyped for 17 markers on the Z chromosome. Phenotypic data were only available for grand-offspring. Only the segregation of male chromosomes provides information on the presence of genes; therefore, a paternal half-sib interval mapping approach was used. The feathering gene was detected significantly and was located between markers ADL0022 and MCW0331. No significant indications were found for the presence of QTL affecting growth traits on the Z chromosome.

Utilizations

Phenotypic and genotypic information will be available for the selection of parents for the next generation. As pointed out by Lush (1933), there is no single desired genotype, i.e. the desired genotype cannot be produced by a single generation of selection. Combining genetic material from different individuals over a number of generations produces the desired genotype. Over generations, the selection scheme capitalizes on recombination and Mendelian sampling in bringing together a better combination of alleles. For longer-term progress, it is essential that sufficient genetic variation be maintained.

Woolliams *et al.* (1999) showed that rate of genetic gain is equal to the product of the Mendelian sampling term and the long-term genetic contributions summed over ancestors. This relationship explicitly shows that sustained genetic gain can only be achieved if animals with a non-zero long-term genetic contribution have an above-average Mendelian sampling term. This underlines the fact that selection programmes need to aim at

newly arising variation. Rates of inbreeding are proportional to the sum of squared long-term genetic contributions (Wray and Thompson, 1990).

When choosing between breeding schemes, one has to find a balance between short-term and intermediate/long-term response. In general, increased short-term response is accompanied by decreased long-term response. Long-term response can be safeguarded by maintaining genetic variation, i.e. by constraining the rate of inbreeding. It can be inferred from the genetic contributions theory that, given a restriction on the rate of inbreeding, genetic gain is maximized when the long-term contributions of ancestors increase linearly with their Mendelian sampling term once their breeding value is above a certain threshold (Woolliams and Bijma, 2000). This theoretical upper boundary is generally unattainable since it is not possible either to measure the true breeding value without error or to set genetic contributions exactly to their desired values. However, using a dynamic selection procedure, genetic gain can effectively be maximized while restricting the rate of inbreeding (Meuwissen, 1997; Grundy *et al.*, 1998). With the same rate of inbreeding, the dynamic rule obtained up to 44% more genetic gain than simple BLUP selection.

Optimizing the longer-term genetic gain results in putting a higher weight on the information from the selection candidate or its offspring, i.e. the Mendelian sampling term of the individual. Molecular genetic data plays an important role here because it provides information on genetic material that has been transmitted by the parent to its offspring. Molecular genetic information offers the opportunity for better exploitation of phenotypic information but collecting phenotypic information remains important in order to identify the genetically unique individuals.

For the practice of animal breeding, identification of QTL that are subject to imprinting has several implications (De Koning *et al.*, 2000). Imprinting calls for a revision of methods for genetic evaluation that currently ignore non-Mendelian expression. The net result of gametic imprinting

is a reduction of the expected phenotypic covariance between parents and offspring relative to that between sibs. Identification of imprinted loci opens new perspectives for crossbreeding, which is common practice in pig and poultry breeding. Imprinted genes could further accommodate differentiation between sire and dam lines.

Prospects

Genetic improvement programmes for poultry can be enhanced by the use of molecular genetic information in crossbreeding and selection programmes. The prospects for MAS are greatest for traits that are difficult to improve through conventional means, because of low heritability or the difficulty and expense of recording phenotypes. Selection using linked markers can be effective and does not require identification of functional mutations, although some level of fine mapping is required. Detection of functional mutation will improve the efficiency of selection and will increase understanding of quantitative genetic variation and the relationships between traits. The number of tools for studying quantitative genetic variation is expected to increase in the coming years.

Bolivar *et al.* (2001) explored the use of knockout/congenic mouse strains for isolating and mapping QTL. Because most knockout strains have been bred to be B6.129 congenic strains, they can be used to test for QTL in the targeted chromosomal area as long as there is a genetic difference between B6 and 129. Bolivar *et al.* (2001) tested a number of knockout/congenic strains in a series of behavioral tests in which mouse performance has a significant genetic component.

It seems only a matter of time before initiatives will be taken to sequence the genomes of farm animals. In combination with bioinformatic analysis, this will be an important step towards unravelling the molecular basis for a variety of phenotypic traits.

Advances in technology are providing tools that allow the simultaneous

quantification of both gene expression and protein expression. Specifically, gene-expression arrays have become more accessible in recent years, and are providing new and abundant levels of information on genome-wide patterns of transcription. It is expected that gene-expression and protein arrays will provide valuable information towards understanding gene function.

How many genes?

How many QTL affect variation in a quantitative trait? This simple question is not easy to answer. The number of QTL mapped in any one experiment is always a minimum number. It may be obvious, but is often forgotten, that QTL can only be mapped if there are allelic differences between the two parent strains used to construct the mapping population. To the extent that these strains are a limited sample of the existing genetic variation, it should not be surprising if different QTL are found in different studies. Designs that use parent strains derived from divergent artificial selection experiments will, however, contain a more representative fraction of segregating variation than will two random inbred lines. Furthermore, an axiom of QTL mapping is that the harder one looks, the more QTL one finds. There are two reasons why the number of QTL is expected to increase with the sample size (number of backcross or F_2 individuals). Firstly, the lower limit for the magnitude of QTL effect that is detectable in any mapping experiment is set by the sample size, and increasing the sample size allows mapping of QTL with smaller effects. Secondly, the precision of mapping depends on the ability to separate linked QTL by recombination. In general, the larger the sample size, the more recombinant events, and the more QTL that are detectable, given a sufficiently dense marker map.

Progress towards understanding the basis of quantitative genetic variation is likely to come from studying allelic

variation at specific QTL (Barton and Keightley, 2002). Cloning of QTL that cause phenotypic differences between selected populations should become more common, aided by the revolution in genomic technology, and should lead to the elucidation of the nature of such QTL. Achieving a satisfactory understanding of variation within populations is likely to be much more challenging. The information that we are seeking is the distribution of sizes of allelic effects at individual QTL and their frequencies within the population in question, as well as information on their effects on fitness. Highly replicated experiments that isolate QTL by fine-scale mapping could yield information on the allelic variants that segregate at specific regions of the genome or even specific loci, along with their frequencies within a population. Association studies should also reveal whether allelic variants repeatedly correlate with phenotypic variation for specific traits. Gaining such information will require very large experiments, larger perhaps than have previously been contemplated, but essential if we are to know the causes of variation among organisms – including our own species.

References

- Albers, G.A.A. and Groot, A. (1998) Future trends in poultry breeding. *World Poultry* 14, 42–43.
- Anderson, L. (2001) Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* 2, 130–138.
- Barton, N.H., and Keightley, P.D. (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics* 3, 11–21.
- Bink, M.C.A.M. and van Arendonk, J.A.M. (1999) Detection of quantitative trait loci in outbred populations with incomplete marker data. *Genetics* 151, 409–420.
- Bink, M.C.A.M., van Arendonk, J.A.M. and Quaas, R.L. (1998a) Breeding value estimation with incomplete marker data. *Genetics, Selection, Evolution* 30, 45–58.
- Bink, M.C.A.M., Quaas, R.L. and van Arendonk, J.A.M. (1998b) Bayesian estimation of dispersion parameters with a reduced animal model including polygenic and QTL effects. *Genetics, Evolution, Selection* 30, 103–125.
- Bolivar, V.J., Cook, M.N. and Flaherty, L. (2001) Mapping of quantitative trait loci with knock-out/congenic strains. *Genome Research* 11, 1549–1552.
- Bovenhuis, H. and Spelman, R.J. (2000) Selective genotyping to detect quantitative trait loci for multiple traits in outbred populations. *Journal of Dairy Science* 83, 173–180.
- Bovenhuis, H., van Arendonk, J.A.M., Davis, G., Elsen, J.M., Haley, C.S., Hill, W.G., Baret, P.V., Hetzel, D.J.S. and Nicholas, F.W. (1997) Detection and mapping of quantitative trait loci in farm animals. *Livestock Production Science* 52, 135–144.
- Bovenhuis, H., Bralten, H., Nieuwland, M.G.B. and Parmentier, H.K. (2002) Genetic parameters for antibody response of chickens to sheep red blood cells based on a selection experiment. *Poultry Science* 81, 309–315.
- Bulmer, M.G. (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.
- Churchill, G.A. and Doerge, R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.
- Cockett, N.E., Jackson, S.P., Shay, T.L., Farnir, F., Berghmans, S., Snowden, G.D., Nielsen, D.M. and Georges, M. (1996) Polar overdominance at the ovine callipyge locus. *Science* 273, 236–238.
- Darvasi, A. and Soller, M. (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theoretical and Applied Genetics* 85, 353–359.
- Darvasi, A. and Soller, M. (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics* 138, 1365–1373.
- Dekkers, J.C.M. and Hospital, F. (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* 3, 22–32.
- De Koning, D.J., Janss, L.L.G., Rattink, A.P., van Oers, P.A.M., De Vries, B.J., Groenen, M.A.M., van der Poel, J.J., De Groot, P.N., Brascamp, E.W. and van Arendonk, J.A.M. (1999) Detection of quantitative trait loci for backfat thickness and intramuscular fat content in Pigs (*Sus scrofa*). *Genetics* 152, 1679–1690.
- De Koning, D.J., Rattink, A.P., Harlizius, B., van Arendonk, J.A.M., Brascamp, E.W. and Groenen, M.A.M. (2000) Genome-wide scan for body composition reveals important role of imprinting. *Proceedings of the National Academy of Sciences USA* 14, 7947–7950.

- Doerge, R.W. (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics* 3, 43–52.
- Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*. Addison Wesley Longman, Harlow, UK.
- Farnir, F., Coppieters, W., Arranz, J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. and Georges, M. (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* 10, 220–227.
- Fernando, R.L. and Grossman, M. (1989) Marker assisted selection using best linear unbiased prediction. *Genetics, Selection, Evolution* 21, 467–477.
- Georges, M., Nielsen, D., Mackinnon, N., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Stelle, M.R., Zao, Z., Womack, J.E. and Hoeschele, I. (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139, 907–920.
- Gianola, D. and Foulley, J.L. (1983) Sire evaluation for ordered categorical data with a threshold model. *Genetics, Selection, Evolution* 15, 201–224.
- Gibson, J.P., Kennedy, B.W., Schaeffer, L.R. and Southwood, O.I. (1988) Gametic models for estimation of autosomally inherited effects that are expressed only when received from either male or female parent. *Journal of Dairy Science* 71, 143 (abstract).
- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. and Snell, R. (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research* 12, 222–231.
- Groenen, M.A.M., Cheng, H.H., Bumstead, N., Benkel, B.F., Briles, W.E., Burke, T., Burt, D.W., Crittenden, L.B., Dodgson, J., Hillel, J., Lamont, S., Ponce de Leon, A., Soller, M., Takahashi, H. and Vignal, A. (2000) A consensus linkage map of the chicken genome. *Genome Research* 10, 137–147.
- Groenen, M.A.M., Crooijmans, R.P.M.A., Vereijken, A.L.J., Albers, G., Veenendaal, T., Dijkhof, R.J.M., van der Poel, J.J., Pakdel, A., van Arendonk, J.A.M. and Bovenhuis, H. (2001) From genetics to genomics: the search for genes affecting ascites. In: *Proceedings of the 2nd Poultry Genetics Symposium*, Gödöllő, Hungary, pp. 21–33.
- Grundy, B., Villanueva, B. and Woolliams, J.A. (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genetical Research Cambridge* 72, 159–168.
- Hagger, C. (1994) Relationships between body weight of cocks and production traits in laying hens, and their possible use in breeding schemes. *Poultry Science* 73, 381–387.
- Haley, C.S. and Knott, S.A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69, 315–324.
- Haley, C.S., Knott, S.A. and Elsen, J.M. (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136, 1195–1207.
- Hamoen, F.F.A., Van Kaam, J.B.C.H.M., Groenen, M.A.M., Vereijken, A.L.J. and Bovenhuis, H. (2001) Detection of genes on the Z-chromosome affecting growth and feathering in broilers. *Poultry Science* 80, 527–534.
- Harlizius, B., Rattink, A.P., de Koning, D.J., Faivre, M., Joosten, R.G., van Arendonk, J.A.M. and Groenen, M.A.M. (2000) The X chromosome harbors quantitative trait loci for backfat thickness and intramuscular fat content in pigs. *Mammalian Genome* 11, 800–802.
- Hazel, L.N. (1943) The genetic basis for constructing selection indices. *Genetics* 28, 476–490.
- Henderson, C.R. (1949) Estimates of changes in herd environment. *Journal of Dairy Science* 32, 706.
- Henderson, C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics* 32, 69–83.
- Hillel, J. (1997) Map-based quantitative trait locus identification. *Poultry Science* 76, 1115–1120.
- Hoeschele, I., Uimari, P., Grignola, F.F.E., Zhang, Q. and Gage, K.M. (1997) Advances in statistical methods to map quantitative trait loci. *Genetics* 138, 205–211.
- Jansen, R.C. (1992) A general mixture model for mapping quantitative trait loci by using molecular markers. *Theoretical and Applied Genetics* 85, 252–260.
- Jansen, R.C. (1994) Controlling the Type I and Type II errors in mapping quantitative trait loci. *Genetics* 138, 871–881.
- Jansen, R.C. (1996) A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics* 142, 305–311.
- Jansen, R.C., Johnson, D.L. and van Arendonk, J.A.M. (1998) A mixture model approach to

- the mapping of quantitative trait loci in complex populations with an application to multiple cattle families. *Genetics* 148, 391–399.
- Jeon, T.T., Carlborg, Ö., Törnsten, A., Giuffra, E., Amarger, V., Chardon, P., Andersson-Eklund, L., Hansson, I., Lundström, K. and Andersson, L. (1999) A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the *IGF2* locus. *Nature Genetics* 21, 157–158.
- Johnson, D.L., Jansen, R.C. and Van Arendonk, J.A.M. (1999) Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach. *Genetical Research* 73, 75–83.
- Khatib, H. (1994) Mapping quantitative trait loci in chicken populations. PhD thesis. The Hebrew University of Jerusalem, Jerusalem, Israel. (In Hebrew with English summary.)
- Knott, S.A., Elsen, J.M. and Haley, C.S. (1994) Multiple marker mapping of quantitative trait loci in half sib populations. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* 21, 33–36.
- Knott, S.A., Elsen, J.M. and Haley, C.S. (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* 93, 71–80.
- Lander, E.S. and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199.
- Lander, E.S. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* 11, 241–247.
- Lebowitz, R.J., Soller, M. and Beckmann, J.S. (1987) Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theoretical and Applied Genetics* 73, 556–562.
- Lipkin, E., Mosig, M.O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A. and Soller, M. (1998) Mapping loci controlling milk protein percentage in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers. *Genetics* 149, 1557–1567.
- Lipkin, E., Fulton, J., Cheng, H., Yonash, N. and Soller, M. (2002) Quantitative trait locus mapping in chickens by selective DNA pooling with dinucleotide microsatellite markers by using purified DNA and fresh or frozen red blood cells as applied to marker-assisted selection. *Poultry Science* 81, 283–292.
- Lush, J.L. (1931) The number of daughters necessary to prove a sire. *Journal of Dairy Science* 14, 209.
- Lush, J.L. (1933) The bull index problem in the light of modern genetics. *Journal of Dairy Science* 16, 501.
- Lush, J.L., Lamoreux, W.F. and Hazel, L.N. (1948) The heritability of resistance to death in poultry. *Poultry Science* 27, 375–388.
- MacKay, T.F.C. (2001) Quantitative trait loci in *Drosophila*. *Nature Reviews Genetics* 2, 11–20.
- Martinez, O. and Curnow, R.N. (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics* 85, 480–488.
- McQuirk, B.J. (1989) The estimation of genetic parameters for all-or-none and categorical traits. In: Hill, W.G. and Mackay, T.F.C. (eds) *Evolution and Animal Breeding: Reviews on Molecular and Quantitative Approaches in Honour of Alan Robertson*. CAB International, Wallingford, UK, pp. 175–180.
- Meuwissen, T.H.E. (1997) Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science* 75, 934–940.
- Meuwissen, T.H.E. and Goddard, M.E. (2000) Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155, 421–430.
- Nezer, C., Moreau, L., Brouwers, B., Coppieters, W., Detilleux, J., Hanset, R., Karim, L., Kvasz, A., Leroy, P. and Georges, M. (1999) An imprinted QTL with major effect on muscle mass and fat deposition maps to the *IGF2* locus in pigs. *Nature Genetics* 21, 155–156.
- Neimann-Sorensen, A. and Robertson, A. (1961) The associations between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agriculturae Scandinavica* 11, 163–196.
- Rance, K.A., Hill, W.G. and Keightley, P.D. (1997a) Mapping quantitative trait loci for body weight on the X chromosome in mice. I. Analysis of a reciprocal F2 population. *Genetical Research* 70, 117–124.
- Rance, K.A., Heath, S.C. and Keightley, P.D. (1997b) Mapping quantitative trait loci for body weight on the X chromosome in mice. II. Analysis of congenic backcrosses. *Genetical Research* 70, 125–133.
- Sax, K. (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8, 552–560.

- Schaeffer, L.R., Kennedy, B.W. and Gibson, J.P. (1989) The inverse of the gametic relationship matrix. *Journal of Dairy Science* 72, 1266–1272.
- Smith, S.P. and Allaire, F.P. (1985) Efficient selection rules to increase non-linear merit: application in mate selection. *Genetics, Selection, Evolution* 17, 387–406.
- Smith, S.P. and Mäki-Tanilla, A. (1990) Genotypic covariance matrices and their inverses for models allowing for dominance and inbreeding. *Genetics, Selection, Evolution* 22, 65–91.
- Soller, M. (1991) Mapping quantitative trait loci affecting traits of economic importance in animal populations using molecular markers. In: Shook, L.B., Lewin, H. and McLaren, D.G. (eds) *Gene Mapping Techniques and Applications*. Dekker, New York, pp. 21–49.
- Soller, M. and 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., Brody, T. and Genizi, A. (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* 47, 35–39.
- Spelman, R.J. and van Arendonk, J.A.M. (1997) Effect of inaccurate parameter estimates on genetic response to marker-assisted selection in an outbred population. *Journal of Dairy Science* 80, 3399–3410.
- Spelman, R., Coppieters, W., Karim, L., van Arendonk, J.A.M. and Bovenhuis, H. (1996) Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144, 1799–1808.
- Tatsuda, K. and Fujinaka, K. (2001) Genetic mapping of the QTL affecting body weight in chickens using a F2 family. *British Poultry Science* 42, 333–337.
- Tier, B. and Sölkner, J. (1993) Analysing gametic variation with an animal model. *Theoretical and Applied Genetics* 85, 868–872.
- Tixier-Boichard, M., Boichard, D., Groeneveld, E. and Bordas, A. (1995) Restricted maximum likelihood estimates of genetic parameters of adult male and female Rhode Island Red chickens divergently selected for residual feed consumption. *Poultry Science* 74, 1245–1252.
- Van Arendonk, J.A.M., Tier, B. and Kinghorn, B.P. (1994) Use of multiple genetic markers in prediction of breeding values. *Genetics* 137, 319–329.
- Van der Beek, S., van Arendonk, J.A.M. and Groen, A.F. (1995) Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. *Theoretical and Applied Genetics* 91, 1115–1124.
- Van Kaam, J.B.C.H.M., van Arendonk, J.A.M., Groenen, M.A.M., Bovenhuis, H., Vereijken, A.L.J., Crooijmans, R.P.M.A., van der Poel, J.J. and Veenendaal, A. (1998) Whole genome scan for quantitative trait loci affecting body weight in chickens using a three generation design. *Livestock Production Science* 54, 133–150.
- Van Kaam, J.B.C.H.M., Groenen, M.A.M., Bovenhuis, H., Veenendaal, A., Vereijken, A.L.J. and van Arendonk, J.A.M. (1999a) Whole genome scan in chicken for quantitative trait loci affecting growth and feed efficiency. *Poultry Science* 78, 15–23.
- Van Kaam, J.B.C.H.M., Groenen, M.A.M., Bovenhuis, H., Veenendaal, A., Vereijken, A.L.J. and van Arendonk, J.A.M. (1999b) Whole genome scan in chicken for quantitative trait loci affecting carcass traits. *Poultry Science* 78, 1091–1099.
- Weller, J.I. (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* 42, 627–640.
- Weller, J.I., Kashi, Y. and Soller, M. (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science* 73, 2525–2537.
- Woolliams, J.A. and Bijma, P. (2000) Predicting rates of inbreeding in populations undergoing selection. *Genetics* 154, 1851–1864.
- Woolliams, J.A., Bijma, P. and Villanueva, B. (1999) Expected genetic contributions and their impact on gene flow and genetic gain. *Genetics* 153, 1009–1020.
- Wray, N.R. and Thompson, R. (1990) Prediction of rates of inbreeding in selected populations. *Genetical Research Cambridge* 55, 41–54.
- Yonash, N., Bacon, L.D., Witter, R.L. and Cheng, H.H. (1999) High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease. *Animal Genetics* 30, 126–135.
- Yonash, N., Cheng, H.H., Hillel, J., Heller, D.E. and Cahaner, A. (2001) DNA microsatellites linked to quantitative trait loci affecting antibody response and survival rate in meat-type chicken. *Poultry Science* 80, 22–28.